The symmetrical pattern of base-pair substitutions rates across the chromosome in *Escherichia coli* has multiple causes

Brittany A. Niccum[1], Heewook Lee[2], Wazim MohammedIsmail[3], Haixu Tang[3], and Patricia L. Foster[1]*

[1]Department of Biology, Indiana University, Bloomington, IN, USA, 47405

[2]School of Computer Science, Computational Biology Department, Carnegie-Mellon University

[3]School of Informatics and Computing, Indiana University, Bloomington, IN, USA, 47405

*Corresponding Author:

Patricia L. Foster

Department of Biology

Jordan Hall 447E

1001 East Third Street

Indiana University

Bloomington, IN 47405

812-855-4084

plfoster@indiana.edu

Author email addresses:

Patricia L. Foster, plfoster@indiana.edu

Brittany A. Niccum, brittanyniccum@gmail.com

Heewook Lee, heewookl@cs.cmu.edu

Wazim MohammedIsmail, wazimoha@indiana.edu

Haixu Tang, hatang@indiana.edu

Abstract

Background: Mutation accumulation experiments followed by whole-genome sequencing have revealed that for several bacterial species, the rate of base-pair substitutions (BPSs) varies across the chromosome in a wave-like pattern that is symmetrical about the origin of replication. The experiments presented here investigate the causes of this pattern in *Escherichia coli*.

Results: The timing of replication initiation, the progression of the replication fork, and the structure of the region where replication terminates all affect the mutational density pattern. As predicted previously, the activities of the nucleoid binding proteins, HU and Fis, are also important, suggesting that mutation rates increase when highly structured DNA regions are replicated. Biases in error-correction by proofreading and mismatch repair are major factors in establishing the pattern. A mutation accumulation experiment with mismatch-repair defective *Bacillus subtilis* revealed a similar symmetrical pattern of BPS rates.

Conclusions: The symmetrical pattern of BPS rates does not have a single cause, but is, instead determined by the initiation, progression, and termination of DNA replication, error-correction, and chromosome structure. These factors should apply to most bacterial, and possibly eukaryotic, genomes, and imply that different areas of the genome evolve at different rates.

Keywords *Escherichia coli*, mutation rate, mutation accumulation, whole genome sequencing, mismatch repair, replication timing, nucleoid associated proteins, proofreading

45 **Background**

46 Recent studies of mutations accumulated non-selectively across bacterial chromosomes have
47 revealed that rates of base-pair substitutions (BPSs) vary 2- to 4- fold in a wave like pattern that is
48 mirrored in the two independently replicating halves of the chromosome. Such symmetrical patterns
49 have been observed in mismatch-repair (MMR) defective strains of *Escherichia coli* (1), *Vibrio fischeri*,
50 *V. cholerae* (2-4), *Pseudomonas fluorescens* (5), and *P. aeruginosa* (6). Such variation in mutation rates
51 may affect the rate at which genes in different regions of the chromosome evolve, and may exert
52 selective pressure on gene placement. Yet the causes of this variation are not known.

53 The fidelity of DNA replication, which, in *E. coli*, is about 1 mistake in 1000 generations (7), is
54 determined by the intrinsic accuracy of the DNA polymerase and error correction by proofreading and
55 mismatch repair (reviewed in (8, 9). In *E. coli*, proofreading is performed by epsilon, a subunit of the
56 DNA polymerase III holoenzyme. If the polymerase inserts the incorrect base, epsilon's 5' to 3'
57 exonuclease activity degrades a few bases of the new strand and polymerase then re-synthesizes it.
58 The accuracy of DNA synthesis is improved about 4000-fold by proofreading (10). Mismatch repair is
59 performed by three proteins, MutS, MutL, and MutH. MutS recognizes a mismatch and recruits MutL.
60 Together they find a nearby GATC site in which, in *E. coli*, the A is methylated by the Dam methylase.
61 Because methylation lags behind replication, unmethylated As identify the "new", and presumably
62 error-containing, DNA strand. MutH is recruited by MutSL to the GATC site and activated to nick the
63 unmethylated DNA strand, which is then degraded past the mismatch by the concerted activity of the
64 UvrD helicase and one of four exonucleases. Pol III then re-synthesizes the strand. MMR improves the
65 accuracy of DNA replication 100 to 200 fold (11).

66 In our previous study of the BPS density pattern in MMR-defective *E. coli* (1), we correlated
67 mutation rates to the chromosomal sites that are affected by two nucleoid-associated proteins (NAPs),
68 HU and Fis. We suggested that when the replication fork encounters regions of the chromosome with
69 high superhelical density due to the binding of these NAPs, the mutation rate increases. An alternative
70 explanation ties mutation rates to replication timing (2, 3). An intriguing hypothesis is that mutation
71 rates vary in concert with fluctuations in dNTP concentration when the replication origin fires
72 repeatedly during rapid growth (2).
73 In the work presented here, we investigate further the causes of the wave-like pattern of BPS
74 rates. MMR-defective strains of *E. coli* additionally defective for other activities were used in mutation
75 accumulation (MA) experiments and the mutations identified by whole genome sequencing (WGS). We
76 also investigated the effects of different growth conditions on BPS rates. We conclude that the BPS
77 density pattern does not have a single cause, but is the result of several factors affecting DNA
78 replication, repair, and chromosome structure. In addition, we report that a MMR-defective *Bacillus*
79 *subtilis* also has a wave-like BPS density pattern that is symmetrical about the origin of replication.

80 ## Results

81 **The base-pair substitution (BPS) density pattern in mismatch repair-defective strains**

82 In a previous paper (1) we demonstrated that the density of BPS that accumulated across the
83 chromosome in a MMR-defective strain during an MA experiment fell into a wave-like pattern that was
84 symmetrical about the origin. We have since performed nine additional MA experiments with MMR
85 mutant strains, each experiment resulting in the accumulation of over 1000 BPSs, for a total of 30,061
86 BPSs (11). As shown in Fig. 1A, the wave-like BPS density pattern was reproducible among the 10
87 experiments. Also as shown in Fig. 1A, the mutational density pattern does not exactly match the non-

88 interacting chromosomal macrodomains (MCs) as defined by Valens et al, 2004 (12). In particular, the
89 wave pattern of BPSs is symmetrical about the origin whereas the Ori MC is not.

90 Following the analysis of Dillon et al, 2018 (2), we computed the wavelet coherence of the
91 collected MMR^{-} data taken in the clockwise and the counterclockwise directions around the
92 chromosome; Additional file 1: Fig. S1A shows that, except for some asymmetry at the midpoint, the
93 wave is symmetrical across the chromosome. The coherence is greatest between 800 to 1,600 Kb per
94 cycle (= 8 to 16 bins per cycle), which is a similar result to that found by Dillon et al, 2018 (2). We also
95 computed the wavelet coherence of the collected MMR^{-} data against each of the experimental results
96 discussed below, and these graphs are also given in the Additional file material. In the analyses below,
97 the combined data from the 10 experiments with MMR-defective strains are used as the standard to
98 which the results obtained in other genetic backgrounds are compared.

99 The spectrum of BPSs in the MMR-defective strains is dominated by A:T transitions at
100 5'NAC3'/3'NTG5' sites (11). To test if the mutational density pattern was simply due to the distribution
101 of these sites, we removed all the A:T transitions at 5'NAC3'/3'NTG5' sites from the data set. As shown
102 in Additional file1: Figs. S5A, S6A, and Tables 1 and 2, although the mutation rate was reduced by 60%
103 when these BPSs were removed, the remaining 12,542 BPSs fell into the same wave pattern. Thus, the
104 variation in BPS rates must be reflective of regions of the chromosome and not the distribution of the
105 hotspot DNA sequences.

106 **Transcription**

107 One obvious hypothesis is that the mutational density pattern reflects transcriptional patterns.
108 Mutation rates have been reported to be both increased (13) and decreased (14) by high levels of
109 transcription. In a previous paper (7) we reported that highly expressed genes had normal mutation

110 rates, a finding that was confirmed in a recent study using deep sequencing to detect mutations (15).
111 To determine if, nonetheless, transcription influences the wave pattern, we used RNA-Seq to
112 quantitate the RNA levels in a Δ*mutL* mutant strain in the lag, exponential, and stationary phases of
113 growth. These results should be representative of the cells in colonies during our MA experiments. We
114 then compared these results to the BPS density pattern by binning the RNA-Seq results into the same
115 bins as used for mutational data. As shown in Fig. 1B and Additional file 1: Fig. S1B, there was no
116 similarity in the two data sets.

117 Because ribosomal operons are homologous, we could not call SNPs in these genes. The RNA
118 reads from the genes in ribosomal operons were also removed from the RNA-Seq data, but their
119 positions have been indicated in Fig. 1B. Interestingly, bin 35, which includes the ribosomal *rrnG*
120 operon, has a high number of RNA reads even in the absence of reads from the *rrnG* genes. This high
121 level of expression is almost exclusively due to the *ssrA* gene, which encodes transfer-messenger RNA
122 (16), that is highly expressed under all three conditions (the RNA-Seq data will be further analyzed in a
123 subsequent paper). The BPS density pattern of several of the strains that will be discussed below tends
124 to reach a minimum in this general area, but it is as often at bin 33 as at bin 35. Given that the bin size
125 is 100 Kb, it seems unlikely that the high level of transcription level of *ssrA*, located in the middle of bin
126 35, is causing these patterns.

127 **Replication Initiation**

128 The BPS density pattern is centered on *oriC*, the origin of replication, and over many experiments
129 in different genetic backgrounds the pattern around the origin has proved to be stable. In both
130 replichores BPS rates decline to a minimum about 300 Kb from the origin and then increase until a

131 peak is reached about 900 Kb from the origin. Mutation rates then fall again and reach a minimum
132 about 3/5th of the distance along each replichore.
133 We tested whether replication initiation was responsible for the maintenance of this pattern by
134 performing MA experiments with strains with errant replication start sites. The *rnhA* gene encodes
135 RNase H1, which degrades RNA-DNA hybrids; in the absence of RNase H1, persistent R-loops can
136 initiate aberrant DNA replication and disrupt normal fork migration (17, 18). But an MA experiment
137 with a Δ*rnhA* Δ*mutL* mutant strain showed no difference in the BPS density pattern from that of the
138 Δ*mutL* mutant strain (Fig. 1C, Additional file 1: Fig. S1C, Tables 1 and 2), indicating that aberrant
139 replication Initiation does not influence where mutations occur, at least not when a powerful *oriC* is
140 present.
141 To further test the influence of replication initiation on the mutational density pattern, we
142 performed MA experiments with strains that have a 5.1 Kb region containing *oriC* moved to the
143 midpoint of the right replichore, where it is called *oriZ* (19). These strains are derived from *E. coli* K12
144 strain AB1157, instead of MG1655 strain, the ancestor of our MA strains, and have a large inversion in
145 the right replichore that relieves the head-on collision between replication initiating at *oriZ* and
146 transcription of the *rrnCABE* operon (20). As a control we created an AB1157 Δ*mutL* mutant strain,
147 which had the similar wave-like BPS density pattern as our MMR$^-$ strains, indicating that the pattern is
148 intrinsic to *E. coli* K12, not specific to sub-strain MG1655 (Additional file 1: Figs. S5B, S6B, Tables 1 and
149 2). We performed MA experiments on Δ*mutL* derivatives of two additional strains: one with both *oriZ*
150 and *oriC* (WX320 Δ*mutL*), and one with only *oriZ* (WX340 Δ*mutL*). The strain that contained two origins
151 had a similar BPS density pattern as the MMR$^-$ strains (Fig. 1D, Additional file 1: Fig. S1D, Tables 1 and
152 2), suggesting that, under our conditions, firing of *oriZ* could not overcome the influence of *oriC*.

153 However, the strain containing only *oriZ* showed a decrease in BPS rate in the 200Kb-area surrounding
154 the new origin, similar to that normally observed about *oriC* (Fig. 1E, Additional file 1: Fig. S1E, Tables 1
155 and 2).

156 **SeqA**

157 As mentioned above, in *E. coli* the adenines in GATC sites are methylated by the Dam
158 methylase. The SeqA protein binds to hemimethylated GATC sites, many of which are clustered around
159 OriC, and, by so doing, SeqA occludes the replication initiation protein, DnaA, hindering origin firing.
160 Sequestering of the origin persists for about one third of a generation; however, the mechanism of
161 relief is not clear. In the absence of SeqA, unregulated initiation presumably results in over-replication,
162 at least when cells are rapidly growing in rich medium (21). Downstream events, such as replication
163 fork collapse, add to the phenotypes of *seqA* mutant cells (22).

164 Loss of SeqA affects chromosomal structure in areas distant from OriC. By binding to
165 hemimethylated DNA, SeqA forms complexes behind the replication fork as it progresses around the
166 chromosome (21). In addition, SeqA binds to areas of the chromosome with closely spaced GATC sites,
167 as well as to particular genes regulated by GATC methylation (23). In the absence of SeqA the
168 superhelicity of the chromosome increases, the nucleoid condenses (24), and transcription is altered
169 (25).

170 To test whether SeqA affects the BPS density pattern, we performed an MA experiment with a
171 Δ*mutL* Δ*seqA* mutant strain. As shown in Fig. 1F, Additional file 1: Fig. S1F, and Tables 1 and 2, loss of
172 SeqA somewhat amplified the BPS density pattern of the right replichore, but the pattern was still
173 highly correlated to that of right replichore of the MMR-defective strains. However, the pattern in the
174 left replichore was disrupted in the Δ*seqA* Δ*mutL* mutant strain. Based on chromatin

175 immunoprecipitation analysis, this area of the left replichore is not targeted by SeqA to a greater
176 extent than the same area of the right replichore (23, 26), suggesting that the disruption of the
177 mutational pattern is not due to loss of binding by SeqA. As shown in Fig. 1B, bins 44, 38, 37, and 35
178 contain a number of highly expressed genes; in addition, bins 42 and 35 contain highly transcribed
179 ribosomal RNA genes. Thus, we hypothesize that loss of SeqA makes the replication machinery
180 particularly susceptible to interference by transcription, disrupting the mutational pattern. If this
181 hypothesis is correct, the interference apparently makes replication more accurate, perhaps by slowing
182 the speed of DNA polymerase.

183 **Replication Fork Progression**

184 In a recent report, Dillon *et al*., 2018 (2), hypothesized that the periodic variation in mutation
185 rates across the chromosome was tied to the timing of DNA replication. Cellular levels of dNTPs are
186 controlled by ribonucleotide reductase (RNR), the expression of which increases when the origin fires
187 (27, 28) }. In fast-growing bacteria in which new rounds of replication are initiated before cell division,
188 the levels of dNTPs should be high when each origin fires, but then fall as progression of the multiple
189 forks dilute the dNTPs. High levels of dNTPs are predicted to increase the probability of
190 misincorporation, and thus the mutation rate, whereas low levels of dNTPs should slow replication and
191 improve fidelity. This systematic fluctuation in dNTP levels could account for the pattern of mutational
192 density across the chromosome (2).

193 This hypothesis can be tested by examining the wave pattern from several experiments that we
194 have already published (11). The number of replicating chromosomes is a positive function of the cell's
195 growth rate (29). When Δ*mutL* or Δ*mutS* mutant strains were grown on glucose minimal medium,
196 which reduces the growth rate about 2-fold relative to growth on LB, the BPS density pattern became

197 chaotic (Fig. 2A, Additional file 1: Figs. S2A, S5C, S6C, Tables 1 and 2). Supplementing the minimal
198 medium with just enough LB to increase the growth rate to normal restored the wave-like BPS density
199 pattern (Fig. 2B, Additional file 1: Fig. S2B, Tables 1 and 2). Growing the cells on diluted LB, on which
200 the growth rate was the same as on minimal glucose medium, also preserved the wave-like BPS density
201 pattern (Fig. 2C, Additional file 1: Fig. S2C, Tables 1 and 2). When the cells were grown at a lower
202 temperature, which also reduced the growth rate 2-fold, the overall shape of the BPS density pattern
203 was retained, but the increases in BPS rates that normally peak at about 900 Kb from each side of the
204 origin were shifted about 200 Kb further from OriC (Fig. 2D, Additional file 1: Fig. S2D, Tables 1 and 2).
205 In addition, the magnitude of the fluctuation of the BPS rate across the chromosome was doubled.
206 Thus, it appears that growth rate *per se* is not a major determinant of the BPS density pattern, but
207 other factors, such as the composition of the growth medium, are also important. For example, the
208 expressions and DNA binding characteristics of nucleoid associated proteins (NAPs) are different under
209 different growth conditions (30-32)

210 In *E. coli* growing aerobically dNTP levels are regulated by a Class Ia RNR encoded by the *nrdA*
211 and *nrdB* genes. Although most of this regulation appears to be due to DnaA interactions (28), the
212 NrdR repressor, which regulates a poorly expressed Class Ib RNR, also regulates *nrdAB* transcription;
213 loss of NrdR results in increased expression of RNR throughout the cell cycle (33). Increased RNR
214 should result in increased dNTP levels, and, indeed, the Δ*nrdR* Δ*mutL* mutant strain had twice the
215 mutation rate as the MMR^{-} strains, as expected when dNTP levels are high (Additional file 2: Table S1).
216 As shown in Fig. 2E, Additional file 1: Fig. S2E, Tables 1 and 2, loss of NrdR did not change the basic BPS
217 density pattern, but the peak rate was shifted away from the origin about 200 Kb on each side. This
218 pattern was similar to that observed when cells were grown at low temperature, as described above.

219 Replication fork progression is aided by the accessory replication helicase, Rep, and, in its
220 absence, the time required for chromosome duplication is doubled (34). Rep removes proteins bound
221 to the DNA in front of the fork (34, 35). While these nucleoproteins are primarily transcription
222 complexes (36, 37), Rep could also free the DNA of blocking NAPs. In addition, Rep aids in restarting
223 replication forks after they stall or collapse (38, 39). As shown in Fig. 2F, Additional file 1: Fig. S2F,
224 Tables 1 and 2, with the exception of a region close to the origin, loss of Rep disrupted the BPS density
225 pattern across the chromosome, suggesting that slowing or stalling the fork results in a distribution of
226 BPS that is essentially random.

227 **Replication Termination**

228 The results from almost all the strains tested show an increase in the BPS rate in the region
229 where replication terminates. The pattern of this increase varies somewhat among experiments.
230 Usually there are two unequal peaks, as shown in Fig.1A, but in some experiments these peaks are
231 better defined and of equal heights and occasionally there is just one peak. We do not know the source
232 of this variation, but suspect it is simply random noise.

233 Replication terminates approximately 180° from the origin in a 1200 Kb region bounded by
234 replication pause (Ter) sites; this region extends from bin 18 to bin 31 in our Figures. The anti-helicase
235 Tus protein binds to the Ter sites and allows each replication fork to enter but not to exit, creating a
236 replication fork "trap", within which the two forks fuse and the chromosome dimer is resolved (40). To
237 determine if the interaction of replication forks with Tus contributes to the increased mutation rate
238 within this region, we performed an MA experiment on a Δ*tus* Δ*mutL* mutant strain. As shown in Fig.
239 2G, Additional file 1: Fig. S2G, Tables 1 and 2, loss of Tus did not affect the BPS density pattern.

240 The Ter macrodomain (MD) extends from 1200 Kb to 2200 Kb (12), which is roughly from bin 20
241 to bin 28 in our Figures. The structure of the Ter MD is maintained by the MatP protein, which binds to
242 23 *matS* sites within this region (41). In the absence of MatP, the Ter MD is disorganized, the DNA is
243 less compact, and the Ter MD segregates too early in the cell cycle and fails to localize properly at
244 midcell (41, 42). Because the mobility of the Ter MD is increased in the absence of MatP, DNA
245 interactions across MD barriers can occur in Δ*matP* mutant cells (41).

246 As shown in Fig. 2H, Additional file 1: Fig. S2H, Tables 1 and 2, loss of MatP caused a severe
247 disruption of the BPS density pattern. The mutation rates in the Ter MD were depressed whereas new
248 peaks appeared on either side of the Ter MD in the Right and Left MDs. Interestingly, the BPS pattern
249 near the origin was maintained in the right but not in the left replichore.

250 **Recombination and the SOS response**

251 Homologous recombination is intimately connected to replication. As replication proceeds,
252 various blocks, such as DNA lesions, transcription complexes, and DNA secondary structures, can cause
253 the replisome to pause and to eventually disassemble. This potentially lethal event is prevented by
254 recombination, which can repair and restart the replisome (43). In addition, the termination region is
255 subject to hyperrecombination (44, 45) particularly in the region bounded by TerA and TerB (our bins
256 21 to 24), named the terminal recombination zone (TRZ) (46, 47).

257 Elimination of *E. coli*'s major recombinase, RecA, had a modest effect on the mutational density
258 pattern. As shown in Fig. 3A, Additional file 1: Figs. S3A, S5D, S6D, Tables 1 and 2, in the *recA mutL* and
259 *recA mutS* mutant strains, BPS rates declined in bins 21 to 25, corresponding fairly well to the TRZ
260 described above. However, the hyperrecombination within the TRZ is dependent on the recombination
261 pathway defined by participation of the RecBCD complex. When we eliminated RecB, the mutational

262 density pattern did not phenocopy that seen when RecA was absent (Additional file 1: Figs. S5E, S6E,
263 Tables 1 and 2). Either our protocol is not sensitive enough to detect an effect of loss of RecB, or
264 another recombination pathway, *e.g.* RecFOR (48), is sufficient to maintain the mutation rate in the
265 region.

266 In addition to its role in recombination, RecA is also a master regulator of the SOS response to
267 DNA damage, which includes the induction of two error-prone DNA polymerases, DNA Pol IV and V. To
268 test whether these polymerases are involved in determining the BPS density pattern, we performed an
269 MA experiment on a strain deleted for the genes that encode Pol IV, *dinB*, and PolV, *umuDC*. As shown
270 in Fig. 3B, Additional file 1: Fig. S3B, Tables 1 and 2, the BPS density pattern in the *mutL dinB umuDC*
271 mutant strain was not significantly different than the MMR^{-} pattern. The genes of the SOS response
272 are repressed by the LexA protein; the *lexA3* allele encodes a super-repressor LexA protein that
273 prevents the SOS genes from being induced (49). When this allele was present the BPS density pattern
274 was also unaffected (Fig. 3C, Additional file 1: Fig. S3C, Tables 1 and 2). Thus, the SOS response appears
275 to play no role in determining the pattern of BPSs across the chromosome.

276 **Nucleoid Associated Proteins**

277 In a previous report (1), we found that the BPS density pattern of a Δ*mutL* strain was correlated
278 with the density of genes activated by the HU protein and repressed by the Fis protein. Combining
279 these two factors in a linear correlation model accounted for 33% of the variation in the mutational
280 data. HU constrains supercoils and compacts the DNA into nucleosome-like particles; Fis also
281 constrains supercoils but, in addition, bends the DNA (31). While both of these NAPs affect
282 transcription, the general lack of correlation of the BPS rate with transcriptional levels (7); also see
283 above) led us to hypothesize that mutation rates across the chromosome were correlated not with

284 transcription *per se*, but with areas of high DNA structure (1). To further test this hypothesis we
285 preformed MA experiments with MMR⁻ mutant strains also defective for each of a number of NAPs.
286 HU exists as a dimer of its two subunits, HUα and HUβ, encoded by the paralogous genes *hupA*
287 and *hupB*, respectively, in the three possible configurations. While loss of both subunits confers a
288 severe growth defect, loss of only one has little consequence during a normal growth cycle, suggesting
289 they can substitute for each other. HU$\alpha\beta$ is the dominant form over most of the cell cycle, but
290 significant amounts of HUα_2 are found during lag phase and early exponential phase, and HUβ_2 is
291 prominent in stationary phase (30). Chromatin immunoprecipitation sequencing (ChIP-Seq) results
292 revealed that HU binds non-specifically to the chromosome and the DNA binding patterns of the three
293 dimers appear to be identical (50).
294 We performed MA experiments with both Δ*hupA* Δ*mutL* and Δ*hupB* Δ*mutL* mutant strains. As
295 shown in Figs. 3D, 3E, Additional file 1: Figs. S3D, S3E, Tables 1 and 2, although loss of *hupB* appeared
296 to affect the BPS density pattern, the differences from the MMR⁻ pattern were not significant.
297 However, loss of *hupA* depressed mutation rates across the chromosome, particularly in the terminus
298 region, while creating new peaks on either side of the Ter MD.
299 Fis has many binding sites across the chromosome, but, based on data at RegulonDB, SOBETZKO *et*
300 *al.* (2012) (51) reported more Fis binding sites in the origin region. However both Chromatin
301 immunoprecipitation plus microarray analysis (ChIP-chip) and ChIP-Seq studies found the density of Fis
302 binding to be more-or-less constant across the chromosome (32, 52). Loss of Fis in both Δ*fis* Δ*mutL* and
303 Δ*fis* Δ*mutS* mutant strains tended to flatten the BPS density pattern across the chromosome except in
304 the region around the origin (Fig. 3F, Additional file 1: Figs. S3F, S5F, S6F, Tables 1 and 2).

305 The NAP HNS binds to DNA at its high-affinity binding sites and then spreads by oligomerization
306 along A:T rich regions of DNA. Bridging between HNS-DNA complexes condenses the DNA into a few
307 clusters per chromosome (53-55). However, as shown in Fig. 3G, Additional file 1: Fig. S3G, Tables 1
308 and 2), loss of HNS had little effect on the BPS density pattern, and, thus, the long-range structures
309 produced by HNS appear not to affect BPS rates.

310 The DPS protein accumulates in stationary phase cells, condenses the nucleoid into a crystalline-
311 like state, and protects the DNA from oxidative and other damage (56). Despite this radical physical
312 change, loss of DPS had little effect on the BPS density pattern Fig. 3H, Additional file 1: Fig. S3H, Tables
313 1 and 2). Of course, we do not know the degree to which cells in stationary phase contribute to the BPS
314 rates under our experimental conditions.

315 **Proofreading**

316 As mentioned above, epsilon is the proofreading subunit of the DNA polymerase III holoenzyme.
317 The *mutD5* allele encodes an epsilon protein that is inactive for proofreading, and strains carrying this
318 allele have a mutation rate 4000-fold greater than that of wild-type strains, and 35-fold greater than
319 that of MMR-defective strains (10). As shown in Fig. 4A, Additional file 1: Fig. S4A, Tables 1 and 2, when
320 proofreading was inactive but MMR was active, the BPS density pattern was less dramatic than when
321 MMR was inactive and proofreading was active; but, nonetheless, the wave pattern was basically the
322 same. When both MMR and proofreading were inactive, which reveals the mutations solely due to
323 replication errors, the BPS density pattern was nearly flat but around the origin it retained significant
324 correlations to the patterns of both the MMR-defective and *mutD5* mutant strains (Fig. 4B, Additional
325 file 1: Fig. S4B, Tables 1 and 2). Thus, the pattern of BPS on both sides of the origin appears to be

326 established by replication errors and then elsewhere across the chromosome the density pattern is
327 largely due to differential error-correction by both proofreading and MMR.

328 **Wild-type**

329 The mutational density wave patterns evident in our data, and in data from other bacteria (2, 3),
330 were obtained when MMR was inactive. Thus, the question arises: does the pattern appear in wild-
331 type strains? It is difficult to answer this question because mutation rates in wild-type strains are so
332 low (in *E. coli*, 120-fold lower than that of MMR-defective strains (7, 11) that enormous experiments
333 would have to be conducted in order to accumulate enough mutations to approach statistical
334 confidence. In a recent study we compared the mutation rates and spectra of a number of *E. coli*
335 strains defective in various DNA repair activities; of these, the results from seven strains were
336 indistinguishable from those of the wild-type parent (57). By combining the mutations from these
337 strains, we achieved 1933 BPSs (11), enough to expect to see a wave pattern if it existed. As is evident
338 in Figs. 4C, 4D, Additional file 1: Figs. S4C, S4D, Tables 1 and 2, these BPSs did not create fall into a
339 recognizable pattern. Indeed, the pattern from the wild-type strains appears to be random; the
340 variance to mean ratio of the binned mutations is 1.4, indicating the values are not disperse, and the
341 MatLab "runstest", a test for runs, returns a P value of 0.58, also indicating that the bin values are
342 random. In the wild-type strain both MMR and proofreading are active, and, while these two activities
343 have similar correction biases (Fig. 4A), proofreading is much more powerful, producing the pattern
344 seen in the MMR-defective strains. Although we cannot conclude that the mutational density pattern
345 in the wild-type strain is other than random, it does have similarity to both the patterns seen in the
346 MMR-defective strains and in the *mutD5* mutant strain, particularly around the terminus (Figs. 4C, 4D,
347 Additional file 1: Figs. S4C, S4D and Tables 1 and 2).

348 ***Bacillus subtilis***

349 In additional to *E. coli* strains, symmetrical mutational density patterns have been demonstrated
350 in MMR- derivatives of *Vibrio fischeri*, *V. cholera* (2-4), *Pseudomonas fluorescens* (5), and *P. aeruginosa*
351 (6). Here we add *Bacillus subtilis* to this list. As shown in Fig. 4E, Additional file 1: Fig. S4E, Tables 1 and
352 2, the BPS mutation rates in a *B. subtilis mutS::*Tn*10* mutant strain fell into a wave like pattern that was
353 symmetrical about the origin. Although similar in shape, the pattern was significantly different from
354 that of *E. coli* (Fig. 4E, Additional file 1: Fig. S4F, Tables 1 and 2). However, as in *E. coli*, the BPS rate
355 appeared to increase in the terminus region, which in *B. subtilis* is not 180° from the origin and
356 corresponds to bins 20-24 in Fig. 4E.

357 **Discussion**

358 In this report we have examined a number of factors that could be responsible for establishment
359 and maintenance of the symmetrical wave-like BPS density pattern across the chromosome. In broad
360 terms these factors were: transcription; DNA replication initiation, progression, and termination;
361 recombination and the SOS response to DNA damage; the binding of nucleoid-associated proteins;
362 and, error-correction by MMR and proofreading. As discussed above, we found that transcription and
363 the SOS response had little effect, and the effect of recombination was modest and confined to the
364 terminal region. We discuss the more significant, factors in greater detail here.

365 **DNA replication initiation, progression, and termination**

366 Providing additional replication origins, either by eliminating RNase H1 or by inserting an ectopic
367 *oriC* (*oriZ*), did not disrupt the wave (Fig. 1C and 1D). However, when *oriZ* was the only origin of
368 replication, the region of depressed BPS rate that surrounds *oriC* when it is in the normal position was
369 re-established about the new origin (Fig. 1E). Note that only 5.1 Kb of DNA containing *oriC* was

370 relocated (19), whereas the region of reduced mutation rate is about 200 Kb; thus, the mutation rate is
371 not determined just by the DNA sequence surrounding the origin. We hypothesize that the process of
372 replication initiation protects the DNA from damage and/or newly established replication forks have a
373 low error-rate. In addition, the BPS mutation rate was increased for about 1000 Kb (10 bins) on either
374 side the new origin so that it resembled the same region about the normal origin. The size of this area
375 is close to the same size detected as "interacting zones" around *oriZ* (58). However, the mutational
376 density pattern across the rest of the chromosome did not re-establish itself to be symmetrical about
377 *oriZ*, but remained symmetrical about the absent *oriC*. Thus, other factors must be important at distant
378 regions. We can also conclude that the overall structure of the BPS pattern across the chromosome is
379 not determined by active replication initiation *per se*¸ but may have evolved in response to replication
380 initiation.

381 Both *V. fischeri* and *V. cholera* have multiple circular chromosomes of different sizes; by
382 comparing the mutational density patterns of these chromosomes, Dillon et al, 2018 (2) identified the
383 timing of replication as a significant determinant of the mutational density patterns. As mentioned
384 above, they suggested that the pattern could be the result of variations in the levels of dNTPs as origins
385 fire during rapid growth. Our results provide partial support for this hypothesis, but also indicate that
386 growth on rich medium, not growth rate *per se*, is a significant determinant of the mutational density
387 pattern, possibly because of effects on the expression and DNA binding of HU and Fis (see below). Our
388 results with a Δ*nrdR* mutant strain (Fig. 2E) also show that dNTP levels are important, but, again, other
389 factors are driving the overall wave pattern. In addition, the results with the Δ*rep* mutant strain (Fig.
390 2F) indicate that radically interrupting the progression of the replication fork disrupts the mutational
391 density pattern.

392 After declining to a local minimum about 3/5th of the distance along each replichore, mutation
393 rates rise in the terminus region (Fig. 1A). We originally hypothesized that this increase was due to
394 collisions of the replication complexes with the Tus anti-helicase, most of which would take place in
395 the region between bins 18 and 31 (1). However, elimination of Tus had no effect on the wave pattern
396 (Fig. 2G, Additional file 1: Fig. S2G), refuting this hypothesis. But, eliminating MatP, the protein that
397 maintains the structure of the Ter MD, reduced mutation rates in the Ter MD, suggesting that
398 replication of highly structured DNA is error-prone (Fig. 2H, Additional file 1: Fig. S2H). Interestingly, in
399 Δ*matP* mutant cells mutation rates increased in the Right and Left MCs, suggesting that in the absence
400 of MatP, these adjacent regions of DNA gained structure. The mechanism by which MatP structures
401 the Ter MC is not clear (59, 60), but it probably involves supercoils which, when unconstrained, could
402 migrate into adjacent areas. However, two studies have found that loss of MatP increases the mobility
403 and long-range interactions of adjacent DNA (41, 61). Resolution of these conflicts will require further
404 experimentation.

405 *Nucleoid-associated proteins*

406 Our previous results predicted that the NAPs HU and Fis should play a role in establishing or
407 maintaining the wave-pattern of BPS, but HNS should not (1). The results presented here confirmed
408 that prediction. In addition, we also found that Dps had no effect on the BPS density pattern. Because
409 the NAPs affect gene expression in various ways, we cannot conclude that DNA binding by the NAPs
410 themselves is responsible for the mutational pattern. And, indeed, we found no significant positive
411 correlations between the BPS pattern and published binding sites of the NAPs, although the location of
412 the binding sites themselves vary widely among published results (e.g. see (32, 50) and the data in

413 RegulonDB (62). Nonetheless, we favor the hypothesis that structuring of the DNA by the NAPs,
414 directly or indirectly, contributes to the BPS pattern.

415 The local effect of HU binding is to bend the DNA, but dimer-dimer interactions produce higher-
416 order HU-DNA complexes that can constrain negative supercoils (63, 64). The analysis of the effect of
417 HU on the mutational density pattern is also complicated by the variation in the cellular concentrations
418 of the three forms with growth cycle (30). Although the two subunits can compensated for each other
419 for viability, the affinities of three forms of HU for various DNA structures (linear, nicked, and gapped)
420 differ (65). Both HUα_2 and HU$\alpha\beta$ can constrain supercoils, but HUβ_2 apparently cannot, at least *in vitro*
421 (30). Since HUβ is not stable (66), the phenotype of a Δ*hupA* mutant strain may simply be that of low
422 cellular HU concentration.

423 While loss of HUβ had little effect on the mutational density pattern (Fig. 3D, Additional file 1:
424 Fig. S3D), loss of HUα reduced the overall BPS rate by 33% and also changed the pattern (Fig. 3E,
425 Additional file 1: Fig. S3E). In the Δ*hupA* Δ*mutL* mutant strain, mutation rates were low in the terminus
426 region but elevated in the Right and Left MCs, very similar to the pattern seen in the Δ*matP* Δ*mutL*
427 mutant strain (Fig. 2H). The low mutation rate in the Ter MD corresponds well with the density of
428 genes activated in wild-type cells but not in HU-deleted cells (67). HU regulates transcription by
429 modifying the DNA superhelical density, suggesting that, like MatP, HU increases the structure of the
430 DNA in the terminal region, causing DNA replication to become error-prone.

431 The mutational pattern in the Δ*hupA* Δ*mutL* mutant strain is also similar to that of the Δ*recA*
432 Δ*mutL* strain (Fig. 3A). An mutation accumulation experiment with a Δ*hupA* Δ*recA* Δ*mutL* mutant strain
433 resulted in a pattern similar to both single mutant strains, and so was not informative (data not
434 shown).

435 Fis is a major transcriptional regulator, either activating or repressing, directly or indirectly,
436 nearly a thousand genes (52). At some promotors Fis acts as a classic transcriptional regulator by
437 interacting with RNA polymerase, but at other promotors Fis affects gene expression by altering the
438 DNA superhelical density (68). The Fis DNA binding sequence is degenerate and the estimated number
439 of binding sites identified by sequence analysis across the genome varies widely among studies. The
440 results of ChIP analyses also vary: the number of DNA regions associated with Fis in different studies
441 ranges from 200 to 1500, which may reflect the effects of different growth conditions (32, 52, 69).
442 The loss of Fis nearly doubled the overall BPS rate, but flattened the wave pattern outside of the
443 origin region (Fig. 3F, Additional file 1: Figs. S5F, S6F). As mentioned above, the mutational density
444 pattern MMR^{-} strains is correlated to the density of genes activated in a Δ*fis* mutant strain as reported
445 by Blot *et al*, 2006 (70). This correlation is particularly strong in bins 14 to 33 (ρ = 0.62, P = 0.003),
446 which corresponds to the area flattened in the Δ*fis* Δ*mutL* and Δ*fis* Δ*mutS* mutant strains. However, no
447 correlation exists between our mutational data and genes found to be responsive to Fis in a recent
448 study (32). Clearly more studies are needed to resolve these conflicts.
449 *Error-correction*
450 Assuming that the BPS recovered from the *mutD5* Δ*mutL* mutant strain are due to intrinsic errors
451 made by DNA polymerase, we conclude that the polymerase is accurate close to the origin, then
452 becomes increasingly less accurate as replication proceeds to about 1/3 of the replichore, at which
453 point accuracy increases again (Fig. 4B). The mutational pattern in the MMR-defective strains must
454 reflect the biased ability of proofreader to correct these polymerase errors. Thus, proofreading is
455 effective close to the origin, but then declines for about 1/3 of the replichore, increases in the Right
456 and Left MCs, but declines again in the terminus region (Figs. 1A, 4B). Mutations in the *mutD5* mutant

457 strain are due to the failure of MMR to correct the polymerase errors that survive in the absence of
458 proofreading. Since the mutational pattern in the *mutD5* mutant strain mimics that of the MMR-
459 defective strains (Fig. 4A, Additional file 1: Fig. S4A, Tables 1 and 2), error correction by MMR, although
460 much less powerful, apparently has the same biases as error correction by proofreader.

461 Given the above considerations, we would expect the mutational density pattern in the wild-type
462 parental strain to mimic that of the MMR-defective strains. However, the nearly 2000 BPS accumulated
463 in eight strains with wild-type mutational phenotypes appear randomly distributed across the
464 chromosome (Figs. 4C, 4D). One explanation for this discrepancy is that when both proofreading and
465 MMR are active, the mutation rate is reduced to the point that other, weaker DNA repair and
466 mutagenic activities obscure the underlying pattern. Although not significantly correlated, the wild-
467 type BPS pattern is similar to the MMR^- pattern, particularly around the terminus, supporting this
468 hypothesis. Alternatively, although the eight strains that we combined to produce 2000 BPS have the
469 same mutation rates, they may have different wave patterns that tend to negate each other and
470 obscure the underlying wild-type pattern.

471 **Conclusions**

472 By applying mutation accumulation followed by whole-genome sequencing to bacterial strains mutant
473 in various activities we have determined that the most important factors determining the symmetrical
474 pattern of BPS rates across the chromosome are: the initiation, progression, and termination of DNA
475 replication, replication error-correction, and chromosome structure. Because of the conservation of
476 these factors, our results should apply to most bacteria, and possibly eukaryotes, and imply that
477 different regions of the genome evolve at different rates.

478 **Methods**

479 **Bacterial Strains and Media**

480 The strains used in this study are listed in Additional file 2: Table S1 and the methods of their
481 construction are in Additional file Materials and Methods. Standard media and antibiotics were used
482 (see Additional file 3: Methods).

483 **Mutation accumulation experiments**

484 The complete MA procedure has been extensively described (7, 11, 57). Details are given in the
485 Additional file 3: Methods

486 **Genomic DNA Preparation, Library construction, Sequencing, and Sequence Analysis**

487 Details are given in the Additional file 3: Methods.

488 **RNA sequencing**

489 The *E. coli* strain PFM144, which is PFM2 Δ*mutL* (11), was grown in LB and aliquots collected
490 during lag (OD = 0.022), log (OD = 0.3), and stationary (OD = 1.5) phase. The number of cells collected
491 was kept constant for each growth phase. Total RNA was extracted, DNA and rRNA were removed, and
492 libraries constructed and sequenced as detailed in the Additional file 3: Methods. Three biological
493 triplicates were prepared for each growth phase.

494 A complete analysis of the RNA-Seq results will be the subject of a subsequent report. For this report
495 the numbers of RNA-Seq reads for each condition were first normalized to the number of reads
496 mapped to the gene *holD*, which was determined by rtPCR to be expressed at the same level in all
497 phases of growth. The means of the normalized RNA reads from the triplicates were then binned into
498 the same bins used for the mutational analysis. A fourth-order Daubechies wavelet transform was
499 performed on the binned RNA-Seq reads as described for the mutational data (1).

500 **Statistical Analysis**

To obtain the BPS density patterns, the numbers of BPSs were binned into 46 bins, each 100 Kb long, as described (1). A fourth-order Daubechies wavelet transform was performed on the binned mutation data as described (1). For presentation in the Figures, these results were converted into rates by dividing the number of BPS by the appropriate number of generations. Pearson's product-moment correlation coefficient, ρ, was used to evaluate the correlations between the binned BPS data (Table 1). Spearman's nonparametric correlation coefficient was also computed for a few data sets, but gave similar results. To account for multiple comparisons, p values were adjusted using the Benjamini-Hochberg method (71) with the false discovery rate set at 25%, implemented with the MatLab R2018a "mafdr" command. Because comparisons using the data from the same strain are not independent, this adjustment was made separately for each column in Tables 1 and 2.

To further compare the BPS density patterns between two data sets, wavelet coherence was calculated and plotted using the MatLab R2018a "wcoherence" command. While the Daubechies wavelet provides a good visual representation of the binned data, it is not continuous and thus not easily adapted for wavelet coherence analysis. The MatLab program first converts the binned data to Morlet wavelets and then computes the coherence between two of these wavelets. We chose to analyze the data with wavelet coherence because it gives a measure of the correlation between the signals (displayed as colors in the Figures) (72, 73). In addition, the MatLab wavelet coherence plot indicates, as a dashed curve, the 'cone of influence' within which results are free of artifactual edge effects (72). The relative phase-lag between the two signals is indicated by small arrows: arrows pointing right indicate in-phase, arrows pointing left indicate 180° out-of-phase, and arrows pointing in other directions indicate the various degrees in between. Because the MatLab program assumes a

522 frequency-time series, the X axis of the plot is cycles/sample and the Y axis is time; we converted these
523 to bins/cycle (on an inverted scale) and bins, respectively.

524 Declarations

525 **Ethics approval and consent to participate**: Not applicable

526 **Consent for publication**: Not applicable

527 **Availability of data and material**

528 The sequences and SNPs reported in this paper have been deposited with the National Center for
529 Biotechnology Information Sequence Read Archive https://trace.ncbi.nlm.nih.gov/Traces/sra/
530 (accession no. in progress) and in the IUScholarWorks Repository (hdl.handle.net/2022/20340).
531 Bacterial strains are available upon request

532 **Competing interests**

533 The authors declare no competing interests

534 **Funding**

535 This research was supported by US Army Research Office Multidisciplinary University Research
536 Initiative (MURI) Award W911NF-09-1-0444 to P.L.F. and H.T., the National Institutes of Health T32
537 GM007757 to B.A.N., and the US Army Undergraduate Research Apprenticeship Program to J. H. and S.
538 R

539 **Authors' contributions**

B.A.N. and P.L.F. designed the research, analyzed the data, and wrote the paper. B.A.N executed the experiments. W.M., H.L., and H.T. performed the bioinformatic analyses. P.L.F. and H.T. supplied resources. All authors read and approved the final manuscript.

Acknowledgements

We thank H. Bedwell-Ivers, C. Coplen, M. Durham, J. Eagan, N. Gruenhagen, J. A. Healy, N. Ivers, C. Klineman., H. Lee, E. Popodi, I. Rameses, S. Riffert, H. Rivera, D. Simon, K. Smith, J. Townes, L. Tran, and L. Whitson for technical help. Bacterial strains were kindly provided by R. Schaaper, R. Reyes-Lamothe, D. Kearns, M. Konkol, and The National BioResource Project at the (Japanese) National Institute of Genetics. We also thank S. E.Bell and X. Wang for technical help and discussions.

References

1. Foster PL, Hanson AJ, Lee H, Popodi EM, Tang H. On the mutational topology of the bacterial genome. G3 (Bethesda). 2013;3(3):399-407.

2. Dillon MM, Sung W, Lynch M, Cooper VS. Periodic variation of mutation rates in bacterial genomes associated with replication timing. MBio. 2018;9(4).

3. Dillon MM, Sung W, Sebra R, Lynch M, Cooper VS. Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in *Vibrio cholerae* and *Vibrio fischeri*. Mol Biol Evol. 2017;34(1):93-109.

4. Wei W, Xiong L, Ye YN, Du MZ, Gao YZ, Zhang KY, et al. Mutation landscape of base substitutions, duplications, and deletions in the representative current cholera pandemic strain. Genome Biol Evol. 2018;10(8):2072-85.

560 5. Long H, Sung W, Miller SF, Ackerman MS, Doak TG, Lynch M. Mutation rate, spectrum, topology,
561 and context-dependency in the DNA mismatch repair-deficient Pseudomonas fluorescens ATCC948.
562 Genome Biol Evol. 2014;7(1):262-71.
563 6. Dettman JR, Sztepanacz JL, Kassen R. The properties of spontaneous mutations in the
564 opportunistic pathogen Pseudomonas aeruginosa. BMC Genomics. 2016;17:27.
565 7. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in
566 the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proceedings of the
567 National Academy of Sciences of the United States of America. 2012;109(41):E2774-E83.
568 8. Marinus MG. DNA Mismatch Repair. EcoSal Plus. 2012; doi:10.1128/ecosalplus.7.2.5
569 9. Ganai RA, Johansson E. DNA replication-a matter of fidelity. Mol Cell. 2016;62(5):745-55.
570 10. Niccum BA, Lee H, MohammedIsmail W, Tang H, Foster PL. The spectrum of replication errors in
571 the absence of error correction assayed across the whole genome of *Escherichia coli*. Genetics.
572 2018;209(4):1043-54.
573 11. Foster PL, Niccum BA, Popodi E, Townes JP, Lee H, MohammedIsmail W, et al. Determinants of
574 base-pair substitution patterns revealed by whole-genome sequencing of DNA mismatch repair
575 defective *Escherichia coli*. Genetics. 2018;209(4):1029-42.
576 12. Valens M, Penaud S, Rossignol M, Cornet F, Boccard F. Macrodomain organization of the
577 *Escherichia coli* chromosome. EMBO J. 2004;23(21):4330-41.
578 13. Jinks-Robertson S, Bhagwat AS. Transcription-associated mutagenesis. Annu Rev Genet.
579 2014;48:341-59.
580 14. Rasouly A, Pani B, Nudler E. A magic spot in genome maintenance. Trends Genet. 2017;33(1):58-
581 67.

582 15. Zhang X, Zhang X, Zhang X, Liao Y, Song L, Zhang Q, et al. Spatial vulnerabilities of the *Escherichia*
583 *coli* genome to spontaneous mutations revealed with improved duplex sequencing. Genetics.
584 2018;210(2):547-58.
585 16. Withey JH, Friedman DI. A salvage pathway for protein structures: tmRNA and trans-translation.
586 Annu Rev Microbiol. 2003;57:101-23.
587 17. Kogoma T. Stable DNA replication: interplay between DNA replication, homologous
588 recombination, and transcription. Microbioland MolecBiolRev. 1997;61:212-38.
589 18. Maduike NZ, Tehranchi AK, Wang JD, Kreuzer KN. Replication of the *Escherichia coli* chromosome
590 in RNase HI-deficient cells: multiple initiation regions and fork dynamics. Mol Microbiol. 2014;91(1):39-
591 56.
592 19. Wang X, Lesterlin C, Reyes-Lamothe R, Ball G, Sherratt DJ. Replication and segregation of an
593 *Escherichia coli* chromosome with two replication origins. Proceedings of the National Academy of
594 Sciences of the United States of America. 2011;108(26):E243-50.
595 20. Ivanova D, Taylor T, Smith SL, Dimude JU, Upton AL, Mehrjouy MM, et al. Shaping the landscape
596 of the *Escherichia coli* chromosome: replication-transcription encounters in cells with an ectopic
597 replication origin. Nucleic Acids Res. 2015;43(16):7865-77.
598 21. Waldminghaus T, Skarstad K. The *Escherichia coli* SeqA protein. Plasmid. 2009;61(3):141-50.
599 22. Sutera VA, Jr., Lovett ST. The role of replication initiation control in promoting survival of
600 replication fork damage. MolMicrobiol. 2006;60(1):229-39.
601 23. Sanchez-Romero MA, Busby SJ, Dyer NP, Ott S, Millard AD, Grainger DC. Dynamic distribution of
602 SeqA protein across the chromosome of *Escherichia coli* K-12. MBio. 2010;1(1):e00012-10.

603 24. Weitao T, Nordstrom K, Dasgupta S. *Escherichia coli* cell cycle control genes affect chromosome
604 superhelicity. EMBO Rep. 2000;1(6):494-9.
605 25. Lobner-Olesen A, Marinus MG, Hansen FG. Role of SeqA and Dam in *Escherichia coli* gene
606 expression: a global/microarray analysis. Proceedings of the National Academy of Sciences of the
607 United States of America. 2003;100(8):4672-7.
608 26. Waldminghaus T, Skarstad K. ChIP on Chip: surprising results are often artifacts. BMC Genomics.
609 2010;11:414.
610 27. Sun L, Fuchs JA. *Escherichia coli* ribonucleotide reductase expression is cell cycle regulated. Mol
611 Biol Cell. 1992;3(10):1095-105.
612 28. Gon S, Camara JE, Klungsoyr HK, Crooke E, Skarstad K, Beckwith J. A novel regulatory mechanism
613 couples deoxyribonucleotide synthesis and DNA replication in *Escherichia coli*. EMBO J.
614 2006;25(5):1137-47.
615 29. Cooper S, Helmstetter CE. Chromosome replication and the division cycle of *Escherichia coli* B/r. J
616 Mol Biol. 1968;31(3):519-40.
617 30. Claret L, Rouviere-Yaniv J. Variation in HU composition during growth of *Escherichia coli*: the
618 heterodimer is required for long term survival. J Mol Biol. 1997;273(1):93-104.
619 31. Dorman CJ. Function of nucleoid-associated proteins in chromosome structuring and
620 transcriptional regulation. J Mol Microbiol Biotechnol. 2014;24(5-6):316-31.
621 32. Kahramanoglou C, Seshasayee AS, Prieto AI, Ibberson D, Schmidt S, Zimmermann J, et al. Direct
622 and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. Nucleic Acids
623 Res. 2011;39(6):2073-91.

624 33. Torrents E, Grinberg I, Gorovitz-Harris B, Lundstrom H, Borovok I, Aharonowitz Y, et al. NrdR
625 controls differential expression of the *Escherichia coli* ribonucleotide reductase genes. J Bacteriol.
626 2007;189(14):5012-21.
627 34. Lane HE, Denhardt DT. The rep mutation. IV. Slower movement of replication forks in *Escherichia*
628 *coli* rep strains. J Mol Biol. 1975;97(1):99-112.
629 35. Guy CP, Atkinson J, Gupta MK, Mahdi AA, Gwynn EJ, Rudolph CJ, et al. Rep provides a second
630 motor at the replisome to promote duplication of protein-bound DNA. Mol Cell. 2009;36(4):654-66.
631 36. Boubakri H, de Septenville AL, Viguera E, Michel B. The helicases DinG, Rep and UvrD cooperate
632 to promote replication across transcription units in vivo. EMBO J. 2010;29(1):145-57.
633 37. Myka KK, Hawkins M, Syeda AH, Gupta MK, Meharg C, Dillingham MS, et al. Inhibiting translation
634 elongation can aid genome duplication in *Escherichia coli*. Nucleic Acids Res. 2017;45(5):2571-84.
635 38. Heller RC, Marians KJ. Non-replicative helicases at the replication fork. DNA Repair (Amst).
636 2007;6(7):945-52.
637 39. Sandler SJ, Marians KJ. Role of PriA in replication fork reativation in *Escherichia coli*. J Bact.
638 2000;182:9-13.
639 40. Duggin IG, Wake RG, Bell SD, Hill TM. The replication fork trap and termination of chromosome
640 replication. MolMicrobiol. 2008;70(6):1323-33.
641 41. Mercier R, Petit MA, Schbath S, Robin S, El Karoui M, Boccard F, et al. The MatP/matS site-specific
642 system organizes the terminus region of the *E. coli* chromosome into a macrodomain. Cell.
643 2008;135(3):475-85.
644 42. Espeli O, Borne R, Dupaigne P, Thiel A, Gigant E, Mercier R, et al. A MatP-divisome interaction
645 coordinates chromosome segregation with cell division in *E. coli*. EMBO J. 2012;31(14):3198-211.

646 43. Syeda AH, Hawkins M, McGlynn P. Recombination and replication. Cold Spring Harb Perspect
647 Biol. 2014;6(11):a016550.
648 44. Louarn JM, Louarn J, Francois V, Patte J. Analysis and possible role of hyperrecombination in the
649 termination region of the *Escherichia coli* chromosome. J Bacteriol. 1991;173(16):5097-104.
650 45. Bierne H, Ehrlich SD, Michel B. The replication termination signal terB of the *Escherichia coli*
651 chromosome is a deletion hot spot. EMBO J. 1991;10(9):2699-705.
652 46. Louarn J, Cornet F, Francois V, Patte J, Louarn JM. Hyperrecombination in the terminus region of
653 the *Escherichia coli* chromosome: possible relation to nucleoid organization. J Bacteriol.
654 1994;176(24):7524-31.
655 47. Corre J, Cornet F, Patte J, Louarn JM. Unraveling a region-specific hyper-recombination
656 phenomenon: genetic control and modalities of terminal recombination in *Escherichia coli*. Genetics.
657 1997;147(3):979-89.
658 48. Cox MM. Regulation of bacterial RecA protein function. Crit Rev Biochem Mol Biol.
659 2007;42(1):41-63.
660 49. Little JW, Harper JE. Identification of the lexA gene product of *Escherichia coli* K-12. Proceedings
661 of the National Academy of Sciences of the United States of America. 1979;76(12):6147-51.
662 50. Prieto AI, Kahramanoglou C, Ali RM, Fraser GM, Seshasayee AS, Luscombe NM. Genomic analysis
663 of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in
664 *Escherichia coli* K12. Nucleic Acids Res. 2012;40(8):3524-37.
665 51. Sobetzko P, Travers A, Muskhelishvili G. Gene order and chromosome dynamics coordinate
666 spatiotemporal gene expression during the bacterial growth cycle. ProcNatlAcadSciUSA.
667 2012;109(2):E42-E50.

668 52. Cho BK, Knight EM, Barrett CL, Palsson BO. Genome-wide analysis of Fis binding in *Escherichia*
669 *coli* indicates a causative role for A-/AT-tracts. Genome Res. 2008;18(6):900-10.
670 53. Wang W, Li GW, Chen C, Xie XS, Zhuang X. Chromosome organization by a nucleoid-associated
671 protein in live bacteria. Science. 2011;333(6048):1445-9.
672 54. Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, et al. High-affinity DNA binding
673 sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. Nucleic
674 Acids Res. 2007;35(18):6330-7.
675 55. Bouffartigues E, Buckle M, Badaut C, Travers A, Rimsky S. H-NS cooperative binding to high-
676 affinity sites in a regulatory element results in transcriptional silencing. Nat Struct Mol Biol.
677 2007;14(5):441-8.
678 56. Wolf SG, Frenkiel D, Arad T, Finkel SE, Kolter R, Minsky A. DNA protection by stress-induced
679 biocrystallization. Nature. 1999;400(6739):83-5.
680 57. Foster PL, Lee H, Popodi E, Townes JP, Tang H. Determinants of spontaneous mutation in the
681 bacterium *Escherichia coli* as revealed by whole-genome sequencing. Proceedings of the National
682 Academy of Sciences of the United States of America. 2015;112(44):E5990-E599.
683 58. Duigou S, Boccard F. Long range chromosome organization in *Escherichia coli*: The position of the
684 replication origin defines the non-structured regions and the Right and Left macrodomains. PLoS
685 Genet. 2017;13(5):e1006758.
686 59. Dupaigne P, Tonthat NK, Espeli O, Whitfill T, Boccard F, Schumacher MA. Molecular basis for a
687 protein-mediated DNA-bridging mechanism that functions in condensation of the *E. coli* chromosome.
688 Mol Cell. 2012;48(4):560-71.

689 60. Lioy VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espeli O, et al. Multiscale structuring
690 of the *E. coli* chromosome by nucleoid-associated and condensin proteins. Cell. 2018;172(4):771-83
691 e18.
692 61. Thiel A, Valens M, Vallet-Gely I, Espeli O, Boccard F. Long-range chromosome organization in *E.*
693 *coli*: a site-specific system isolates the Ter macrodomain. PLoSGenet. 2012;8(4):e1002672.
694 62. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muniz-Rascado L, Garcia-Sotelo JS, et
695 al. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated
696 gold standards and more. Nucleic Acids Res. 2013;41(Database issue):D203-13.
697 63. Guo F, Adhya S. Spiral structure of *Escherichia coli* HUalphabeta provides foundation for DNA
698 supercoiling. Proceedings of the National Academy of Sciences of the United States of America.
699 2007;104(11):4309-14.
700 64. van Noort J, Verbrugge S, Goosen N, Dekker C, Dame RT. Dual architectural roles of HU:
701 formation of flexible hinges and rigid filaments. Proceedings of the National Academy of Sciences of
702 the United States of America. 2004;101(18):6969-74.
703 65. Pinson V, Takahashi M, Rouviere-Yaniv J. Differential binding of the *Escherichia coli* HU,
704 homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA. J Mol Biol.
705 1999;287(3):485-97.
706 66. Bonnefoy E, Almeida A, Rouviere-Yaniv J. Lon-dependent regulation of the DNA binding protein
707 HU in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of
708 America. 1989;86(20):7691-5.

709 67. Berger M, Farcas A, Geertz M, Zhelyazkova P, Brix K, Travers A, et al. Coordination of genomic
710 structure and transcription by the main bacterial nucleoid-associated protein HU. EMBO Rep.
711 2010;11(1):59-64.
712 68. Browning DF, Grainger DC, Busby SJ. Effects of nucleoid-associated proteins on bacterial
713 chromosome structure and gene expression. CurrOpinMicrobiol. 2010;13(6):773-80.
714 69. Grainger DC, Hurd D, Goldberg MD, Busby SJ. Association of nucleoid proteins with coding and
715 non-coding segments of the *Escherichia coli* genome. Nucleic Acids Res. 2006;34(16):4642-52.
716 70. Blot N, Mavathur R, Geertz M, Travers A, Muskhelishvili G. Homeostatic regulation of
717 supercoiling sensitivity coordinates transcription of the bacterial genome. EMBO Rep. 2006;7(7):710-5.
718 71. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach
719 to multiple testing. J Roy Stat Soc Ser B (Stat Method). 1995;57(1):289-300.
720 72. Grinsted A, Moore JC, Jevrejeva S. Application of the cross wavelet transform and wavelet
721 coherence to geophysical time series. Nonlinear Processes in Geophysics. 2004;11:561-6.
722 73. Maraun D, Kurths J. Cross wavelet analysis: significance testing and pitfalls. Nonlinear Processes
723 in Geophysics. 2004;11:505-14.

724

725

Table 1. Pearson's correlation, ρ, of each strain versus MMR⁻

		Whole Chromosome Bins 1 to 46		Right Replichore Bins 1 to 23		Left Replichore Bins 46 to 24		Origin Bins 1 to 13, 46 to 34		Terminus Bins 14 to 33	
Strain	Description	ρ	P[a]	ρ	P[a]	ρ	P[a]	ρ	P[a]	ρ	P[a]
Collective[b]	MMR⁻	—	—	—	—	—	—	—	—	—	—
Collective	MMR⁻ minus A:T ts at NAC/NTG	0.94	<0.001	0.93	<0.001	0.96	<0.001	0.97	<0.001	0.71	0.004
PFM421	Δ*rnhA* Δ*mutL*	0.74	<0.001	0.63	0.002	0.84	<0.001	0.81	<0.001	0.47	0.075
PFM669	Δ*mutL* (AB1157)	0.59	<0.001	0.40	0.064	0.75	<0.001	0.71	<0.001	0.45	0.079
PFM430/ 431	*oriC*⁺ *oriZ*⁺ Δ*mutL* (AB1157)	0.75	<0.001	0.71	<0.001	0.81	<0.001	0.83	<0.001	0.46	0.075
PFM426	Δ*oriC oriZ*⁺ Δ*mutL* (AB1157)	0.39	0.008	0.21	0.34	0.62	0.003	0.40	0.044	0.41	0.11
PFM533/ 534	Δ*seqA* Δ*mutL*	0.49	0.001	0.71	<0.001	0.27	0.23	0.53	0.006	0.38	0.14
PFM5[b]	Δ*mutL* on minimal medium	0.50	0.001	0.40	0.069	0.61	0.003	0.57	0.003	0.58	0.029

PFM343[b]	Δ*mutS* on minimal medium	0.18	0.22	0.18	0.41	0.18	0.41	0.21	0.31	0.22	0.38
PFM343[b]	Δ*mutS* on diluted LB	0.84	<0.001	0.84	<0.001	0.85	<0.001	0.87	<0.001	0.71	0.004
PFM343[b]	Δ*mutS* on supplemented minimal medium	0.81	<0.001	0.82	<0.001	0.81	<0.001	0.83	<0.001	0.65	0.008
PFM342[b]	Δ*mutS* at low temperature	0.57	<0.001	0.60	0.003	0.53	0.013	0.69	<0.001	-0.02	0.95
PFM799	Δ*nrdR* Δ*mutL*	0.49	0.001	0.58	0.005	0.39	0.080	0.69	<0.001	-0.31	0.24
PFM677	Δ*rep* Δ*mutL*	0.48	0.001	0.49	0.022	0.48	0.027	0.62	0.001	-0.12	0.67
PFM256	Δ*tus* Δ*mutL*	0.82	<0.001	0.79	<0.001	0.85	<0.001	0.87	<0.001	0.70	0.004
PFM257	Δ*matP* Δ*mutL*	0.40	0.007	0.60	0.004	0.21	0.34	0.70	<0.001	-0.47	0.075
PFM422	Δ*recA* Δ*mutL*	0.71	<0.001	0.70	<0.001	0.73	<0.001	0.85	<0.001	0.06	0.84
PFM424	Δ*recA* Δ*mutS*	0.80	<0.001	0.78	<0.001	0.83	<0.001	0.91	<0.001	0.44	0.082
PFM456	Δ*recB* Δ*mutL*	0.82	<0.001	0.87	<0.001	0.79	<0.001	0.89	<0.001	0.39	0.12
PFM118	Δ*umuDC* Δd*inB* Δ*mutL*	0.59	<0.001	0.62	0.003	0.55	0.010	0.61	0.001	0.69	0.004
PFM120	*lexA3* Δ*sulA* Δ*mutL*	0.82	<0.001	0.82	<0.001	0.83	<0.001	0.83	<0.001	0.80	0.001
PFM259	Δ*hupB* Δ*mutL*	0.73	<0.001	0.70	<0.001	0.77	<0.001	0.81	<0.001	0.57	0.031
PFM258	Δ*hupA* Δ*mutL*	0.49	0.001	0.43	0.053	0.58	0.005	0.67	<0.001	-0.23	0.38
PFM317/ 318	Δ*fis* Δ*mutL*	0.51	<0.001	0.61	0.003	0.42	0.060	0.67	<0.001	-0.30	0.25
PFM482	Δ*fis* Δ*mutS*	0.49	0.001	0.31	0.15	0.62	0.003	0.65	<0.001	-0.28	0.27

PFM741	Δ*hns* Δ*mutL*	0.81	<0.001	0.81	<0.001	0.81	<0.001	0.88	<0.001	0.46	0.075
PFM713	Δ*dps* Δ*mutL*	0.80	<0.001	0.84	<0.001	0.78	<0.001	0.89	<0.001	0.47	0.075
PFM163[c]	*mutD5*	0.76	<0.001	0.79	<0.001	0.74	<0.001	0.82	<0.001	0.78	0.001
PFM165/397/399[c]	*mutD5* Δ*mutL*	0.64	<0.001	0.76	<0.001	0.52	0.014	0.86	<0.001	-0.46	0.075
Collective[b]	Wild type	0.36	0.014	0.41	0.061	0.31	0.17	0.37	0.065	0.51	0.060
Collective[d]	*Bacillus subtilis* *mutS::*Tn*10*	0.52	<0.001	0.63	0.003	0.40	0.068	0.74	<0.001	-0.52	0.055

[a] P values were adjusted for multiple comparisons using the Benjamini-Hochberg method (71) with the false discovery rate set at 25%

[b] Data are from (11)

[c] Data are from (10)

[d] Data are combined from four MA experiments with DK2140 (NCIB3610 *mutS*::Tn*10*), DK2141 (NCIB3610 *mutS*::Tn*10*), DK2142 (NCIB3610 *mutS*::Tn*10* Δ*comI*), and DK2143 (NCIB3610 *mutS*::Tn*10* Δ*com*I). There were no differences in mutation rates or spectra among these strain

Table 2. Pearson's correlation, ρ, of right versus left replichore for each strain

		Whole Chromosome		Origin		Terminus	
		Bins 1 to 23, 46 to 24		Bins 1 to 13, 46 to 34		Bins 14 to 23, 33 to 24	
Strain	Description	ρ	P[a]	ρ	P[a]	ρ	P[a]
Collective[b]	MMR⁻	0.91	<0.001	0.95	<0.001	0.66	0.23
Collective	MMR⁻ minus A:T ts at NAC/NTG	0.89	<0.001	0.94	<0.001	0.42	0.60
PFM421	Δ*rnhA* Δ*mutL*	0.49	0.026	0.58	0.051	0.24	0.82
PFM669	Δ*mutL* (AB1157)	0.52	0.018	0.64	0.026	0.29	0.77
PFM430/ 431	*oriC*⁺ *oriZ*⁺ Δ*mutL* (AB1157)	0.60	0.005	0.80	0.003	0.19	0.84
PFM426	Δ*oriC oriZ*⁺ Δ*mutL (AB1157)*	0.28	0.22	0.33	0.31	-0.06	0.91
PFM533/ 534	Δ*seqA* Δ*mutL*	0.41	0.062	0.53	0.079	0.32	0.71
PFM5[b]	Δ*mutL* on minimal medium	0.27	0.22	0.19	0.53	0.05	0.91
PFM343[b]	Δ*mutS* on minimal medium	-0.02	0.93	0.32	0.31	-0.53	0.39
PFM343[b]	Δ*mutS* on diluted LB	0.87	<0.001	0.93	<0.001	0.66	0.23
PFM343[b]	Δ*mutS* on supplemented minimal medium	0.74	<0.001	0.79	0.003	0.54	0.39
PFM342[b]	Δ*mutS* at low temperature	0.84	<0.001	0.89	<0.001	0.73	0.19
PFM799	Δ*nrdR* Δ*mutL*	0.79	<0.001	0.78	0.003	0.84	0.07
PFM677	Δ*rep* Δ*mutL*	0.43	0.053	0.58	0.051	-0.16	0.84
PFM256	Δ*tus* Δ*mutL*	0.70	0.001	0.88	<0.001	0.26	0.81
PFM257	Δ*matP* Δ*mutL*	0.45	0.042	0.40	0.20	0.72	0.19

PFM422	Δ*recA* Δ*mutL*	0.66	0.001	0.81	0.003	0.11	0.88
PFM424	Δ*recA* Δ*mutS*	0.64	0.002	0.80	0.003	0.07	0.91
PFM456	Δ*recB* Δ*mutL*	0.67	0.001	0.78	0.003	0.03	0.94
PFM118	Δ*umuDC* Δ*dinB* Δ*mutL*	0.39	0.075	0.68	0.018	-0.13	0.87
PFM120	*lexA3* Δ*sulA* Δ*mutL*	0.51	0.019	0.57	0.055	0.42	0.60
PFM259	Δ*hupB* Δ*mutL*	0.64	0.002	0.78	0.003	0.35	0.65
PFM258	Δ*hupA* Δ*mutL*	0.54	0.014	0.69	0.015	0.15	0.84
PFM317/ 318	Δ*fis* Δ*mutL*	0.61	0.004	0.67	0.019	0.58	0.34
PFM482	Δ*fis* Δ*mutS*	0.34	0.13	0.49	0.108	0.19	0.84
PFM741	Δ*hns* Δ*mutL*	0.81	<0.001	0.85	0.001	0.60	0.33
PFM713	Δ*dps* Δ*mutL*	0.52	0.017	0.77	0.004	-0.20	0.84
PFM163[c]	*mutD5*	0.75	<0.001	0.77	0.004	0.52	0.39
PFM165/397/399[c]	*mutD5* Δ*mutL*	0.78	<0.001	0.88	<0.001	0.39	0.62
Collective[b]	Wild type	0.17	0.46	0.24	0.45	0.17	0.84
Collective[d]	*Bacillus subtilis mutS::*Tn*10*	0.73	<0.001	0.87	0.000	0.37	0.65

[a] P values were adjusted for multiple comparisons using the Benjamini-Hochberg method (71) with the false discovery rate set at 25%

[b] Data are from (11)

[c] Data are from (10)

[d] Data are combined from four MA experiments with DK2140 (NCIB3610 *mutS*::Tn*10*), DK2141 (NCIB3610 *mutS*::Tn*10*), DK2142 (NCIB3610 *mutS*::Tn*10* Δ*comI*), and DK2143 (NCIB3610 *mutS*::Tn*10* Δ*com*I). There were no differences in mutation rates or spectra among these strain

Figure Legends

Fig. 1. The BPS density pattern is not due to transcription but is affected by moving the origin of replication. For each plot the data were collected into 100 Kb bins starting at the origin of replication on the left and continuing clockwise around the chromosome back to the origin on the right. **Fig. 1A**. The BPS density pattern of MMR-defective strains. Bars represent the mean mutation rate in each bin calculated from the BPSs collected from 10 experiments with MMR-defective strains. Error bars are the 95% CLs of the means. The grey line represents the Daubechies wavelet transform of the binned data. The chromosomal macrodomains (MCs) are indicated at the top of the plot: green, Ori MC; blue, Right and Left MC; red, Ter MC. **Fig. 1B**. The Daubechies wavelet transform of the binned reads from RNA-Seq samples taken during lag, log, and stationary growth phases of the Δ*mutL* mutant strain, PFM144. The data used for the wavelet transforms were the means of three biological replicates. Arrows indicate the positions of the rRNA operons, but the reads from the ribosomal genes in those operons were not included in the plot. **Fig. 1C, D, E, and F**. The BPS density patterns (bars) and Daubechies wavelet transforms (pink line) of the indicated mutant strains compared to the Daubechies wavelet transform of the combined MMR^{-} strains (grey line). See text for a description of the strains. In each plot the left-hand scale has been adjusted to bring the wavelets close together for comparison. The red arrow in 1E indicates the position of *oriZ*. Strains: 1C, PFM421; 1D, PFM430/431; 1E, PFM426; 1F, PFM533/534.

Fig. 2. Altering the progression of DNA replication affects the BPS density pattern. For each plot the data were collected into 100 Kb bins starting at the origin of replication on the left and continuing clockwise around the chromosome back to the origin on the right. The BPS density

23 patterns (bars) and Daubechies wavelet transforms (pink lines) of the indicated strains are
24 compared to the Daubechies wavelet transform of the combined MMR⁻ strains (grey line). In
25 each plot the left-hand scale has been adjusted to bring the wavelets close together for
26 comparison. **Fig. 2A, B, C, and D**. Altering the cellular growth rate by growing a Δ*mutS* mutant
27 strain on different medium and at different temperatures differentially affects the BPS density
28 pattern. In the experiments shown in Plot A, C, and D, the cells grew at about half the normal
29 rate, whereas in the experiment shown in plot B they grew at the normal rate. Strains: 2A, 2B,
30 and 2C, PFM343; 2D, PFM342. **Fig. 2E and F**. Dysregulation of dNTP levels and loss of Rep, the
31 auxiliary replication helicase, affect the BPS density pattern. 2E, PFM799; 2F, PFM677.
32 Fig. 2G and H. Loss of the Tus antihelicase has no effect, but loss of the terminus organizing
33 protein, MatP, changes the BPS density pattern across the chromosome. The arrows in Fig. 2G
34 mark the major Ter sites where Tus binds. The bar in Fig. 2H shows the region in which MatP
35 binds. Strains: 2G, PFM256; 2H, PFM257.
36 **Fig. 3**. Loss of RecA, HUα, or Fis changes the BPS pattern, but the SOS response and loss of
37 other NAPs do not. For a description of the plots, see the legend to Fig. 2. **Fig. 3A, B, and C.** Loss
38 of RecA alters the BPS density pattern in the terminus region, but not through its role as a
39 master regulator of the SOS response. RecA (Fig. 3A) is *E. coli*'s major recombinase. The *dinB*
40 and *umuDC* genes (Fig. 3B) encode the error-prone DNA polymerases, Pol IV and Pol V, that are
41 induced as part of the SOS response. The *lexA3* allele encodes a non-inducible repressor of the
42 SOS genes. Deletion of *sulA* prevents lethal filamentation after induction of the SOS response
43 (not relevant to this study). Strains: 3A, PFM422; 3B, PFM118; 3C, PFM120. **Fig. 3D, E, F, G, and**
44 **H.** Loss of the HUα subunit of HU or of Fis changes the BPS density pattern, but loss of the HUβ

45 subunit of HU, HNS, or DPS has only minor effects. Strains: 3D, PFM259; 3E, PFM258; 3F,

46 PFM317/318; 3G, PFM741/742; 3H, PFM713.

47 **Fig. 4.** The BPS density pattern is a result of biased correction by replication proofreading and

48 MMR. For a description of the plots, see the legend to Fig. 2. **Fig. 4A and B**. A strain with

49 deficient proofreading but active MMR yields a BPS pattern similar to a MMR-defective strain,

50 but a strain with neither proofreading nor MMR does not. The *mutD5* allele encodes an

51 exonuclease deficient proofreader. Strains: 4A, PFM163; 4B, PFM165/397/399. **Fig. 4C and D.**

52 The BPS density pattern in the wild-type strain does not match the pattern of either the MMR-

53 defective strain, or the *mutD5* mutant strain. The green line in Fig. 1D is the Daubechies

54 wavelet transform of the *mutD5* mutant strain (pink line in Fig. 4A). Strains: 4C and D, eight

55 strains with wild-type mutational phenotypes (see text). **Fig. 4E**. MMR-defective *B. subtilis* also

56 has a symmetrical BPS density pattern, but it is different than *E. coli*'s pattern. Strains: three *B.*

57 *subtilis* mutant strains with the same mutational phenotype (see Additional file 2: Tables S1 and

58 S3).

Figure 1

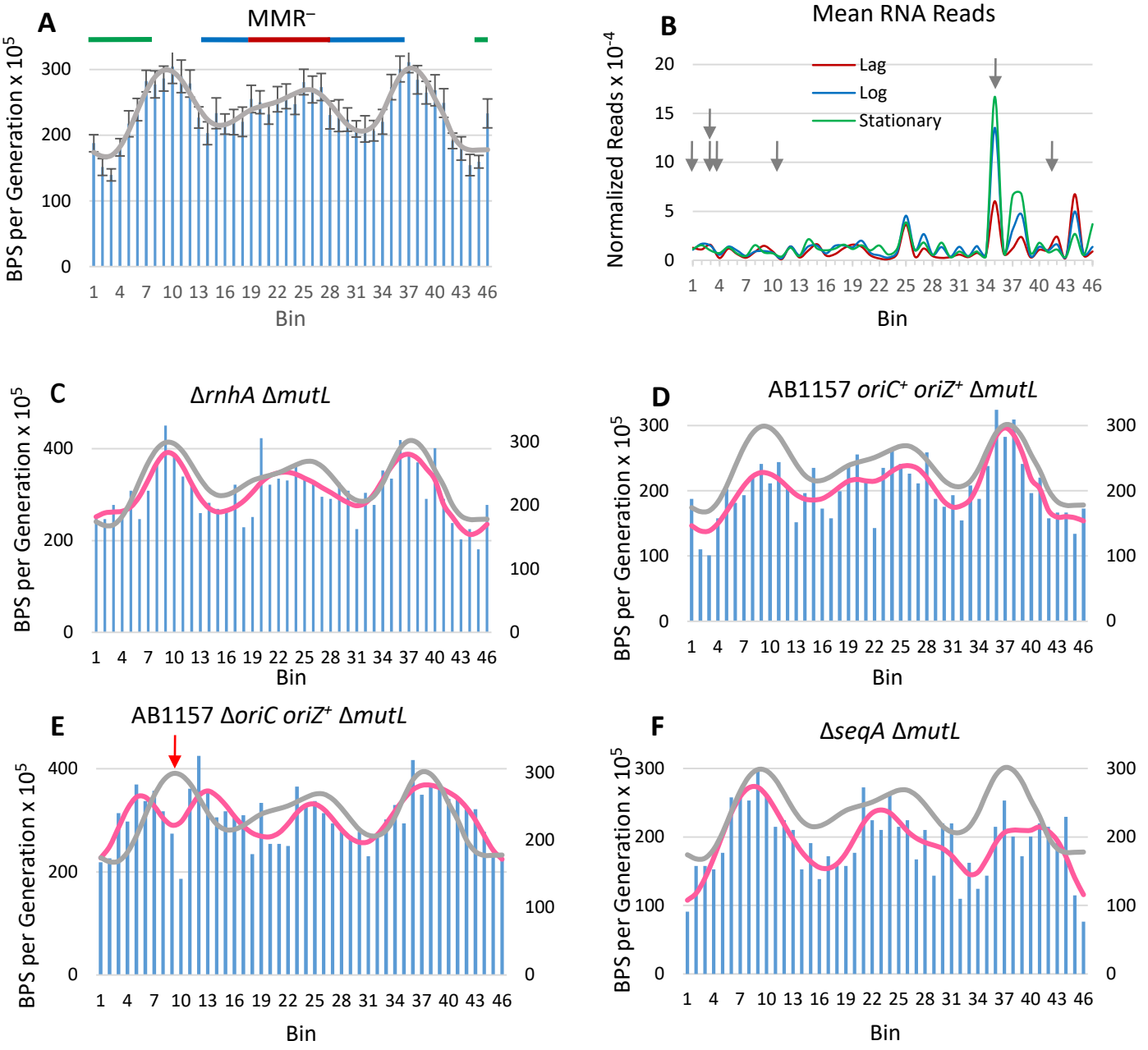

A
MMR⁻
BPS per Generation x 10^5
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
B
Mean RNA Reads
Normalized Reads x 10^-4
20
15
10
5
0
Lag
Log
Stationary
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
C
ΔrnhA ΔmutL
BPS per Generation x 10^5
400
200
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
D
AB1157 oriC⁺ oriZ⁺ ΔmutL
BPS per Generation x 10^5
300
200
100
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
E
AB1157 ΔoriC oriZ⁺ ΔmutL
BPS per Generation x 10^5
400
200
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
F
ΔseqA ΔmutL
BPS per Generation x 10^5
300
200
100
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin

Figure 2

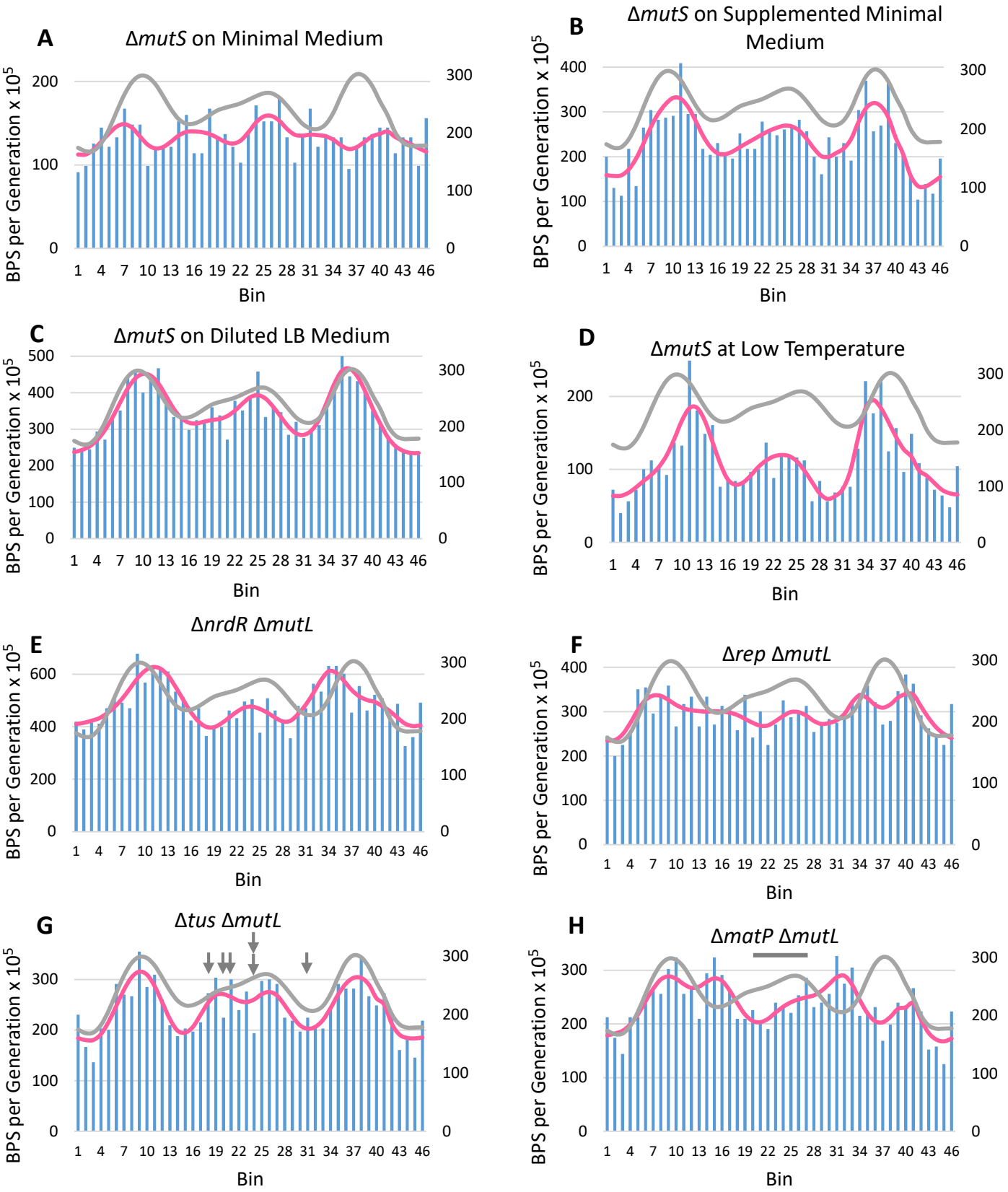

A
ΔmutS on Minimal Medium
B
ΔmutS on Supplemented Minimal Medium
C
ΔmutS on Diluted LB Medium
D
ΔmutS at Low Temperature
E
ΔnrdR ΔmutL
F
Δrep ΔmutL
G
Δtus ΔmutL
H
ΔmatP ΔmutL
BPS per Generation x 10^5
Bin
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46

Figure 3

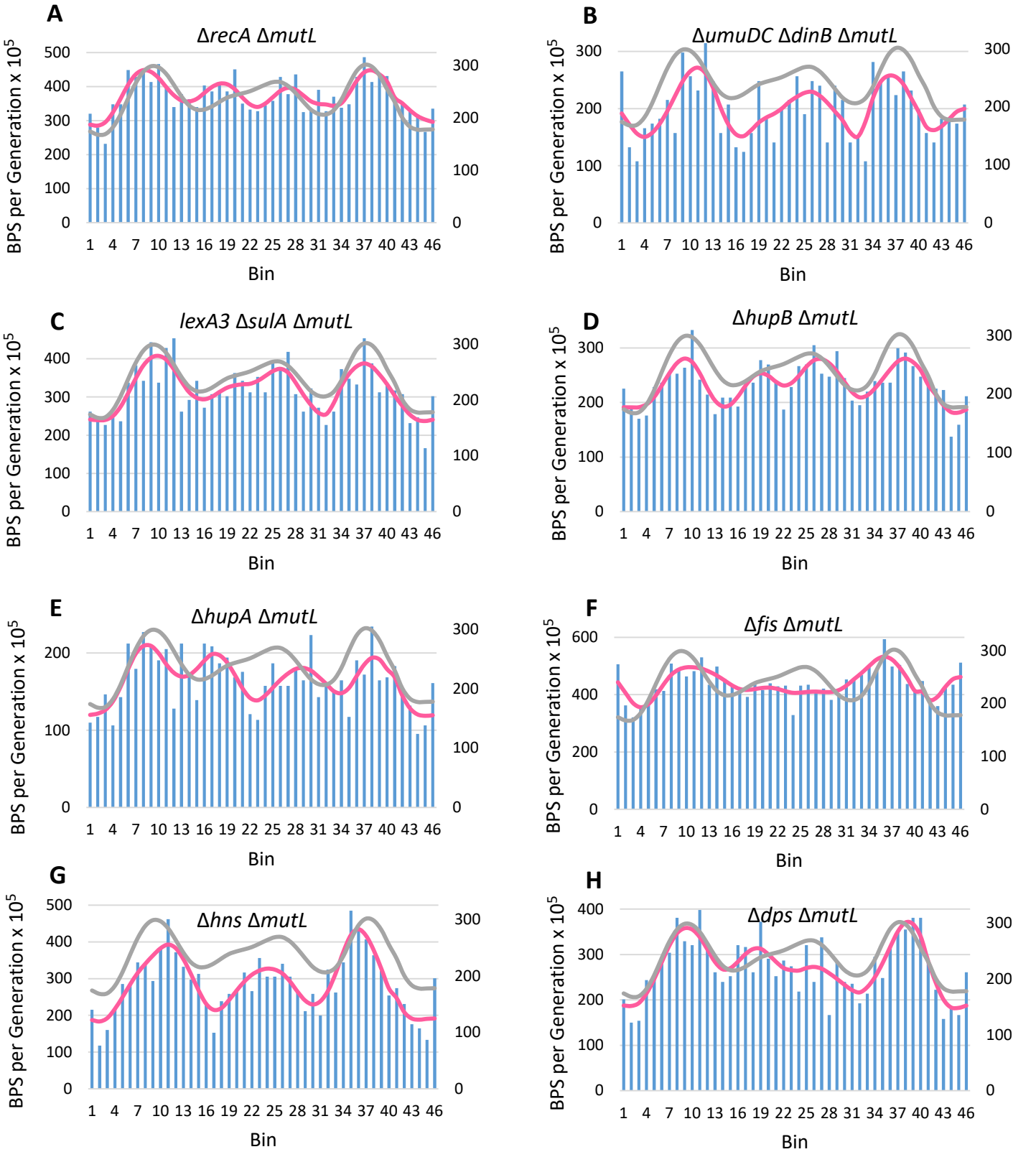
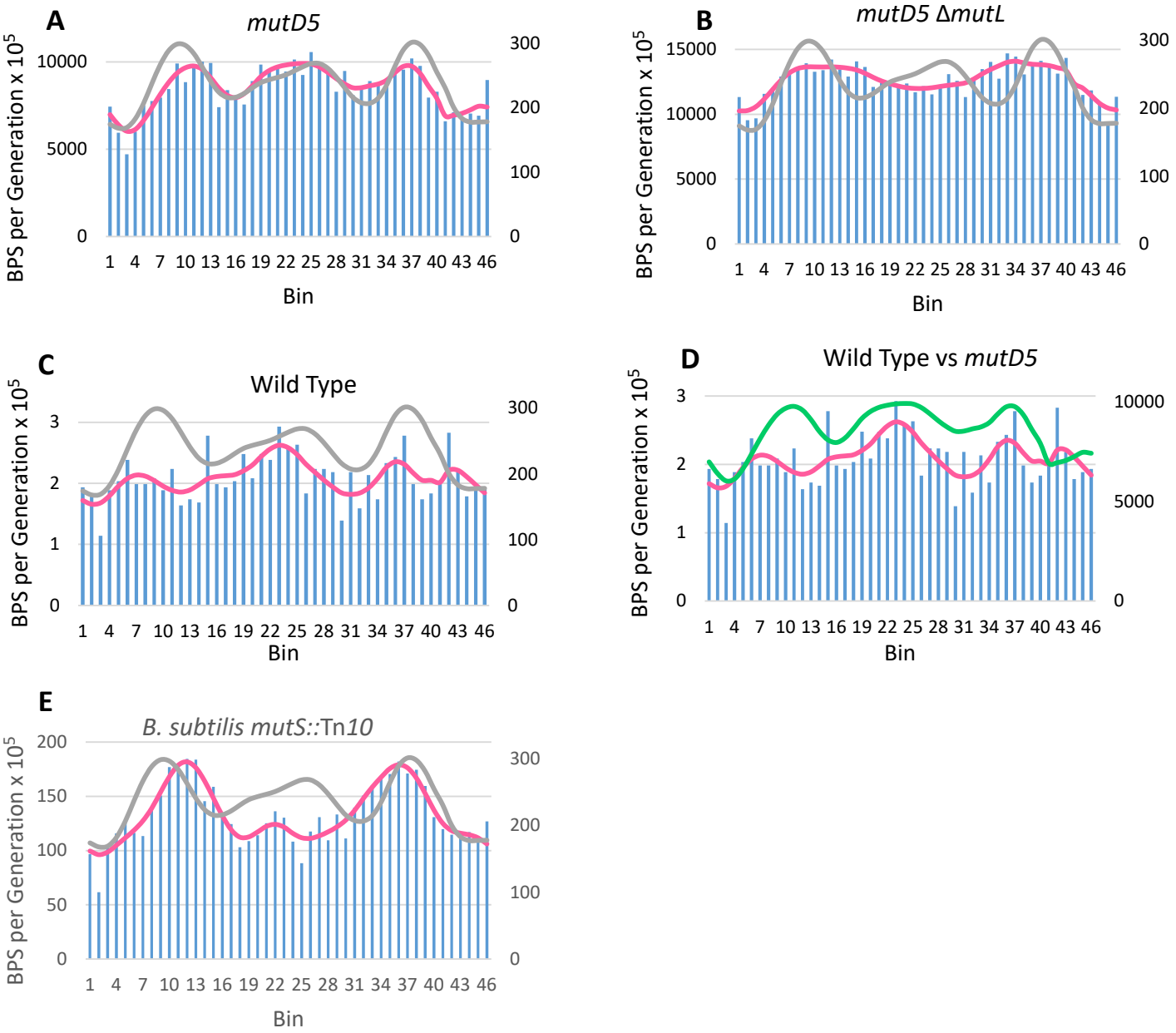

A
ΔrecA ΔmutL
B
ΔumuDC ΔdinB ΔmutL
C
lexA3 ΔsulA ΔmutL
D
ΔhupB ΔmutL
E
ΔhupA ΔmutL
F
Δfis ΔmutL
G
Δhns ΔmutL
H
Δdps ΔmutL
BPS per Generation x 10^5
Bin

Figure 4


A
mutD5
BPS per Generation x 10^5
10000
5000
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
B
mutD5 ΔmutL
BPS per Generation x 10^5
15000
10000
5000
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
C
Wild Type
BPS per Generation x 10^5
3
2
1
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
D
Wild Type vs mutD5
BPS per Generation x 10^5
3
2
1
0
10000
5000
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin
E
B. subtilis mutS::Tn10
BPS per Generation x 10^5
200
150
100
50
0
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46
Bin