Supporting Information

Lee et al. 10.1073/pnas.1210309109

SI Materials and Methods

Bacterial Strains and Media. Genetic manipulations were as described (1). The strains used in this study are prototrophic derivatives of the *Escherichia coli* K12 reference strain MG1655 (2, 3), kindly supplied by R. Maurer, Hawken School, Gates Mills, OH 44040 (4). Because many laboratory substrains of MG1655 have a defective *rpoS* gene, we confirmed the functionality of this gene in our strain by assaying hydrogen peroxidase II activity with an H_2O_2 bubble test (5); the consensus sequence of the final mutation-accumulation (MA) lines further confirmed that the *rpoS* gene is wild type. The mutant *rph-1* allele in MG1655 was replaced with the wild-type *rph* allele by first moving the Δ*pyrE748::kan* allele from JW3617 (6) into MG1655 via P1 bacteriophage transduction, selecting for resistance to kanamycin (Kan^R). The *pyrE* gene is next to the *rph* gene in the *E. coli* genome. A Kan^R auxotrophic transductant then was transduced to prototrophy with a P1 bacteriophage lysate of P90C (7); loss of Kan^R was confirmed. The presence of the wild-type *rph* gene in the transductants was confirmed by sequence analysis. One rph^+ transductant was chosen to be the founder for our MA experiments and is designated "PFM2." Whole-genome sequencing found that PFM2 differs from the reference strain by only six point mutations, none of which are in genes, and has an additional insertion sequence element, IS186, inserted at nucleotide 187,786. To generate a mismatch repair (MMR)-deficient strain, the *ΔmutL720::kan* allele was transduced into PFM2 from JW4128 (6) by selecting for Kan^R. The Kan gene then was removed as described (8), leaving an in-frame scar sequence that encodes a 34-amino acid peptide. Deletion of the *mutL* gene and removal of the Kan cassette were confirmed by PCR and genomic sequencing. The $MutL^-$ founder strain is designated "PFM5." Cultures were grown in liquid Miller LB broth or on Miller LB agar plates (Difco, BD). When required, antibiotics were added at the following concentrations: carbenicillin, 100 μg/mL; kanamycin, 50 μg/mL; chloramphenicol, 10 μg/mL; nalidixic acid, 40 μg/mL; and rifampicin, 100 μg/mL. For strain construction, prototrophy was confirmed by growth on VB minimal glucose plates (1). For freezing, strains were grown in LB broth, DMSO was added to a final concentration of 9% (vol/vol), and the cultures frozen at −80 °C.

Estimation of Mutation Rates by Fluctuation Tests. Mutation rates to nalidixic acid resistance (Nal^R) or rifampicin resistance (Rif^R) were estimated using fluctuation tests as described (9). The tests consisted of multiple 0.1-mL LB cultures that were inoculated with a 10^{-6} dilution of a saturated culture of the appropriate strain and grown for 24 h at 37 °C with shaking. Then all or a portion of each culture was plated on LB agar plus rifampicin or nalidixic acid, except for a few cultures that were pooled and used to determine the total cell numbers by plating appropriate dilutions on LB agar. The results given in Table 2 are based on 40 cultures for each test of PFM2 and 70 cultures of PFM5. The mutation rate was calculated using the Ma–Sandri–Sarkar maximum likelihood method (10), and confidence limits (CL) were calculated as described (9). Both these methods were implemented by the FALCOR web tool found at www.mitochondria.org/protocols/FALCOR.html (11).

MA Procedure. After passage, plates were stored at 4 °C; if a well-isolated colony was not available on a particular day, a second attempt was made to streak from the same colony on the stored plates. However, if two such trials were unsuccessful, the line was dropped from the experiment. Initially wild-type MA lines were passaged 111 times, and $MutL^-$ MA lines were passaged 19 times; then a single colony from each line was used to establish a frozen stock. Subsequently, wild-type MA lines that had not been sequenced were streaked from the frozen stocks and passaged an additional 111 times; these lines are termed the "wild-type 6K lines," whereas the original set is termed the "wild-type 3K lines."

Estimation of Generations in Colonies. The number of generations per passage was estimated from the diameter of the colonies, measured with a ruler. Each colony of the wild-type strain that was chosen for passage was measured, but colonies of the $MutL^-$ MA lines were measured only at the beginning and end of the experiment. Colony diameters ranged from 1–4 mm; most colonies were in the 2- to 3-mm range, and overall the mean (+SD) diameter was 2.57 ± 0.57 mm. The degree of crowding on the plate was the major determinant of the colony size; the diameters of 35 colonies chosen at random after streaking from a freezer stock of the $MutL^-$ strain ranged from 1–2.5 mm with mean (+SD) of 1.90 ± 0.47 mm.

The number of cells in colonies of different diameters was determined by excising colonies of various sizes from the agar plates, resuspending them in saline, and plating dilutions on LB agar plates. Colony diameters were converted to generations by taking the $\log_2$ of the number of cells in the colony. The coefficients of variation of the $\log_2$ values determined for each colony diameter ranged from 1–7%. For both the wild-type and $MutL^-$ strain, the number of cells in a colony ranged from $0.7–9.7 \times 10^8$, with a log-average of 3.8×10^8 cells, which converts to 28 generations.

Estimation of Cell Viability in Colonies. The fraction of dead cells in resuspended colonies was determined using the Live/Dead BacLight Bacterial Viability Kit (Invitrogen, Inc); with this kit dead cells stain with propidium iodide and fluoresce red. Microscopy was performed on a Nikon Eclipse 80i with an X-Cite-120 lamp; fluorescent and phase-contrast images were captured using Metamorph software and were analyzed with ImageJ (US National Institutes of Health, http://imagej.nih.gov/ij/, 1997–2011). To determine if cells lost viability during storage, we compared percentages of dead cells in fresh colonies and in colonies stored at 4 °C for 1–3 d. Based on three to four colonies and 12–24 microscope fields at each time point, the fractions (mean ± SEM) of dead cells in colonies of the wild-type strain were 0.05 ± 0.01, 0.07 ± 0.01, 0.09 ± 0.01, and 0.19 ± 0.01 at 0, 24, 48, and 72 h, respectively, at 4 °C. The corresponding numbers for the $MutL^-$ strain were 0.05 ± 0.01, 0.10 ± 0.03, 0.17 ± 0.02, and 0.31 ± 0.02.

SNP Calling. The reference genome sequence was NCBI Reference Sequence, NC_000913.2. For each sample, Illumina reads were aligned to the *E. coli* K12 (strain MG1655) genome with the short read alignment tool, BWA (ver. 0.5.9) (12). To increase the sensitivity, seeding was turned off, and a maximum of four edit distances was allowed; default values were used for other parameters. To detect mutations among MA lines accurately, the median and the median absolute deviation (MAD) of insert sizes were calculated for each MA line based solely on the read pairs with correct orientations and single-mapping loci. Any reads with MAD ≥3 (which equals approximately two SD for a large sample from a normally distributed population) were discarded before SNP calling. On average, the reads from each MA line covered 98% of the reference genome; uncovered regions consisted mainly of repetitive elements and insertion sequence elements in the reference sequence. For each position, the following procedures were performed to call SNPs: (*i*) independently in each line a consensus base was called requiring ≥20 reads covering the site (≥10 from

each DNA strand) and ≥80% of the reads indicating the same base; (*ii*) consensus bases across all MA lines were compared, and the SNP was called if it was observed only in that line but not in the rest. If a base was observed in all MA lines but not in the reference sequence, it was called as a fixed difference. Shared mutations were called if they were observed in some but not all of the lines.

Short Indel Calling. Short indels (≤4 bp) were called based on the read mapping of the SNP calling procedures. Any position with at least one read with an indel event was considered initially a candidate site. In each line, an indel was called for these candidate sites if three or more reads from each strand of DNA covered the site and the frequency of the most abundant indel was at least 0.5, or if five or more reads covered the site and the frequency of the most abundant indel was at least 0.8. An approach similar to the SNP-calling procedures was used to differentiate the fixed differences from the indels that arose during the MA experiment. The depth values used to calculate the frequency of the most abundant indels were adjusted before the calculation; this adjustment was necessary because BWA disallows indels within 5 bp of the ends. The adjusted depth at site *i*, denoted as D_i' is calculated by the equation $D_i' = D_i - F_i$, where D_i is read depth at site *i*, and F_i is the number of reads whose mapping position at either end is within 5 bp of site *i*. The frequency then was calculated by the equation $Freq_i = n(A_i)/D_i'$, where $n(A_i)$ is the number of reads indicating the most abundant indel at site *i*.

Mutation Confirmation by Conventional Sequencing. The region to be sequenced was amplified by PCR directly from bacterial cells or from purified genomic DNA. The PCR fragment then was purified using either RapidTips (Diffinity Genomics, Inc.) or the QIAquick PCR purification kit (Qiagen, Inc.). DNA sequencing was performed using the ABI Big Dye sequencing kit (Applied Biosystems, Inc.) and analyzed on an Applied Biosystems 3730 automated sequencing system.

Shared Mutation Analysis. The probability of wild-type lines sharing mutations that arose before the first bottleneck is very small, but the high mutation rate of the MutL⁻ strain means there is a 95% probability that 2 of 34 MutL⁻ lines would share a mutation and a 32% probability that they would share two mutations (based on the binomial distribution). To determine the lineages of MutL⁻ MA lines, multiple sequence alignments were performed using CLUSTALW (http://www.genome.jp/tools/clustalw/) run with the default parameters (indels were converted to SNPs for this analysis). The lineages generated then were checked for two mutations that were deduced to have occurred early during growth before the first bottleneck: (*i*) a GC > AT transition at nucleotide 3,240,115 that appeared in 14 of the 34 lines, and (*ii*) a +T indel at nucleotide 1,727,596 that appeared in 10 of the 20 lines that retained a G at nucleotide 3,240,115. Lines with irreconcilable conflicts between the occurrence of these two mutations and the lineage analysis were eliminated. A shared base-pair substitution was assigned to only one line, but because indels occur at hotspots, it is possible that the same indel in two MA lines appeared independently instead of by descent. An indel was retained in the analysis if it met one of the following criteria: (*i*) it occurred in both of the lineages defined by the early-arising mutations, or (ii) it occurred in both a wild-type and a MutL⁻ line. In addition, different types of indels that occurred at the same site were retained.

SI Notes for Table S4

The analysis of indel hotspots was complicated by the high probability that some indels arose in the MutL⁻ line during growth before the first bottleneck (see *SI Materials and Methods*). To generate the conservative list of potential hotspots given in Table S4, we included only indels that were of independent origin based on one of the following criteria: (*i*) an indel occurred in the same run in both a MutL⁻ line and a wild-type line; (*ii*) different events occurred in the same run; or (*iii*) the same indel occurred in two clonal lineages that diverged early in the shared growth of the strain. By relaxing these criteria to include indels that appeared to have arisen independently based solely on lineage analysis (see *SI Materials and Methods*), we generated the additional liberal list given in Table S4. Also given in Table S4 is the probability of the listed number of events occurring at the same site, calculated from the Poisson distributions with a mean equal to the number of indels observed per run of a given length divided by the number of runs of that length in the genome. These probabilities are greater in the additional set because more indels were included. Because we are blind to indels that revert a previously mutated side, there may be more hotspots as well as more indels at the identified hotspots.

SI Notes for Table S5

Depurination and Deamination Rates. The depurination and deamination rates are from refs. 13–15. The depurination rates for bases in ssDNA are four times those in dsDNA; deamination rates for bases in ssDNA are 150–250 times those for bases in dsDNA, so we have taken 200 as the best estimate. 6meA depurinates at 2.5 times the rate of A. 5meC deaminates at three to four times the rate of C, so we have taken 3.5 as the best estimate. We assume a generation time of 20 min.

ssDNA. The amount of ssDNA in the cell was estimated as follows. Lagging-strand DNA replication creates about 2 kb of ssDNA (16); eight replication forks in a rapidly growing cell thus would create a total of 16 kb of ssDNA. [However, this DNA is coated immediately and presumably is protected from deamination by single-stranded binding protein (16)]. The transcription bubble is about 17 bp long (17); estimates of the number of RNA polymerase molecules engaged at any moment in *E. coli* cells growing in rich medium range from 200 (18) to 1,330 (19). Thus, at steady state 3,400–23,000 bases are exposed by transcription.

Dam and Dcm Sites. There are 19,120 GATC sites in the genome. Each GATC has two potential 6meAs, giving 38,240 6meA per genome, or 3.3% of the total number of As. There are 12,045 CCWGG sites per genome. Each site has two potential 5meCs, giving 24,090 5meCs per genome, or 2% of the total number of Cs.

Noncanonical Dam Sites. The Dam methylase has some activity at noncanonical target sites, particularly GACC sequences but also CATC, TATC, AATC, and GATT sequences (20). There are 162,075 of these sites in the genome, and because only one A in each site can be methylated, 7% of the total number of As are potential targets for methylation. In the MutL⁻ strain, A:T mutations occurred at 107 such sites, but 104 of these mutations were transitions instead of the expected transversions. Seventy-seven of these transitions occurred at sites with the sequence 5′G**A**CC3′/3′C**T**GG5′ that includes the 5′**A**pC3′/3′**T**pG5′ sequence that is a hotspot for transitions in the MutL⁻ spectrum (Table S2). In the wild-type strain, 10 A:T mutations occurred at noncanonical Dam sites; six were A:T > C:G transversions, and four were A:T > G:C transitions. When both the wild-type and the MutL⁻ strain are considered together, 11% (9/79) of the A:T transversions occurred at these sites, a value not significantly different from the expected 7% ($\chi^2 = 0.9$, $P = 0.34$). Thus, these sites do not appear to be hotspots for transversions, suggesting that they are not methylated in vivo to the extent that they produce mutations via loss of the 6meA.

1. Miller JH (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria* (Cold Spring Harbor Laboratory, Cold Spring Harbor).
2. Blattner FR, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1462.
3. Riley M, et al. (2006) *Escherichia coli* K-12: A cooperatively developed annotation snapshot—2005. *Nucleic Acids Res* 34:1–9.
4. Slater SC, Lifsics MR, O'Donnell M, Maurer R (1994) *holE*, the gene coding for the theta subunit of DNA polymerase III of *Escherichia coli*: Characterization of a *holE* mutant and comparison with a *dnaQ* (epsilon-subunit) mutant. *J Bacteriol* 176: 815–821.
5. Zambrano MM, Siegele DA, Almirón M, Tormo A, Kolter R (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science* 259:1757–1760.
6. Baba T, et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol*, 2:2006, 0008.
7. Coulondre C, Miller JH (1977) Genetic studies of the lac repressor. IV. Mutagenic specificity in the *lacI* gene of *Escherichia coli*. *J Mol Biol* 117:577–606.
8. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97:6640–6645.
9. Foster PL (2006) Methods for determining spontaneous mutation rates. *Methods Enzymol* 409:195–213.
10. Sarkar S, Ma WT, Sandri GH (1992) On fluctuation analysis: A new, simple and efficient method for computing the expected number of mutants. *Genetica* 85:173–179.
11. Hall BM, Ma CX, Liang P, Singh KK (2009) Fluctuation analysis CalculatOR: A web tool for the determination of mutation rate using Luria-Delbrück fluctuation analysis. *Bioinformatics* 25:1564–1565.
12. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
13. Friedberg EC, et al. (2006) *DNA Repair and Mutagenesis*. 2nd Ed (ASM, Washington, D.C.).
14. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362: 709–715.
15. Lindahl T, Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11:3610–3618.
16. McHenry CS (2011) DNA replicases from a bacterial perspective. *Annu Rev Biochem* 80:403–436.
17. Berg JM, Tymoczko JL, Stryer L (2012) *Biochemistry* (W H Freeman, New York).
18. Bon M, McGowan SJ, Cook PR (2006) Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. *FASEB J* 20:1721–1723.
19. Ishihama A (2000) Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* 54:499–518.
20. Clark TA, et al. (2012) Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res* 40:e29.

Table S1. Synonymous codon changes in the wild-type MA lines

Wild-type triplet	Usage	Mutant triplet	Usage	Difference in usage	Gene	Protein
AAA	0.77	AAG	0.23	−0.54	*aspS*	Aspartate–tRNA ligase
CTG	0.50	CTA	0.04	−0.46	*dsdC*	D-serine dehydratase (deaminase) transcriptional activator
CTG	0.50	CTA	0.04	−0.46	*acnA*	Aconitase A
CTG	0.50	CTA	0.04	−0.46	*metQ*	DL-methionine transporter subunit
CTG	0.50	CTA	0.04	−0.46	*ybbW*	Probable allantoin permease,
CTG	0.50	CTA	0.04	−0.46	*hofO*	Protein involved in utilization of DNA as a carbon source
ATC	0.42	ATA	0.07	−0.35	*ytfN*	Conserved protein
CTG	0.50	TTG	0.13	−0.37	*glxR*	Tartronate semialdehyde reductase 2
CTG	0.50	TTG	0.13	−0.37	*ydjJ*	Predicted oxidoreductase
CTG	0.50	TTG	0.13	−0.37	*yhjJ*	Predicted zinc-dependent peptidase
CCG	0.53	CCA	0.19	−0.34	*phoP*	PhoP transcriptional regulator
CCG	0.53	CCA	0.19	−0.34	*pepQ*	Proline dipeptidase
CCG	0.53	CCA	0.19	−0.34	*gpt*	Guanine-xanthine phosphoribosyltransferase phosphotransferase
CCG	0.53	CCA	0.19	−0.34	*yehA*	Predicted fimbrial-like adhesin protein
CCG	0.53	CCA	0.19	−0.34	*mutS*	Methyl-directed mismatch repair protein
CAG	0.65	CAA	0.35	−0.31	*yhcM*	Conserved protein with nucleoside triphosphate hydrolase domain
CAG	0.65	CAA	0.35	−0.31	*mrdA*	Penicillin-binding protein PBP2
GGC	0.41	GGA	0.11	−0.30	*yicJ*	Putative sugar transporter
GAT	0.63	GAC	0.37	−0.25	*speE*	Component of spermidine synthetase
GAT	0.63	GAC	0.37	−0.25	casC	Predicted protein
GTG	0.37	GTA	0.15	−0.22	*yfaD*	Putative transposase
GCG	0.36	GCT	0.16	−0.20	*rcsD*	Phosphotransfer intermediate protein in two-component regulatory system
GGT	0.34	GGG	0.15	−0.19	*yhaO*	Putative amino acid:H+ symport permease
GCG	0.36	GCA	0.21	−0.15	*nanS*	Conserved protein
ACG	0.27	ACA	0.13	−0.14	*hyuA*	D-stereospecific phenylhydantoinase
AGC	0.28	AGT	0.15	−0.13	*yfaA*	Predicted protein
AGC	0.28	AGT	0.15	−0.13	*yfaA*	Predicted protein
TAT	0.57	TAC	0.43	−0.13	*ycfD*	Conserved protein
GTG	0.37	GTT	0.26	−0.12	*yjfF*	Putative ABC transporter permease protein
TGC	0.56	TGT	0.44	−0.12	*pepQ*	Proline dipeptidase
GCC	0.27	GCT	0.16	−0.11	*btuE*	Predicted glutathione peroxidase
GCG	0.36	GCC	0.27	−0.09	*yneF*	Predicted diguanylate cyclase
GGC	0.41	GGT	0.34	−0.07	*dgsA*	Transcriptional repressor
GCA	0.21	GCT	0.16	−0.05	*valS*	Valyl-tRNA synthetase
GGG	0.15	GGA	0.11	−0.04	*yhdY*	Putative ABC transporter permease protein
TCT	0.15	TCA	0.12	−0.02	*rfbB*	Component of dTDP-glucose 4,6-dehydratase
CGC	0.40	CGT	0.38	−0.02	*efeB*	Conserved protein
TCC	0.15	TCT	0.15	0.00	*fecR*	Regulator for fec operon
CGA	0.06	CGG	0.10	+0.03	*rhsB*	Function unknown, encoded within RhsB repeat
GGA	0.11	GGG	0.15	+0.04	*ascG*	Transcriptional repressor
GGA	0.11	GGG	0.15	+0.04	*zraP*	Zn-binding periplasmic protein
GCA	0.21	GCC	0.27	+0.06	*guaA*	Bifunctional GMP synthase/glutamine amidotransferase protein
GGT	0.34	GGC	0.41	+0.07	*ydjI*	Predicted aldolase
GGT	0.34	GGC	0.41	+0.07	*secY*	Preprotein translocase membrane subunit
ATC	0.42	ATT	0.51	+0.09	*ybbY*	Predicted uracil/xanthine transporter
ATC	0.42	ATT	0.51	+0.09	*dapA*	Dihydrodipicolinate synthase
GCC	0.27	GCG	0.36	+0.09	*ybjO*	Predicted inner membrane protein
GTT	0.26	GTG	0.37	+0.12	*tolQ*	Membrane spanning protein in TolA-TolQ-TolR complex
GCT	0.16	GCG	0.36	+0.20	*yeeR*	CP4-44 prophage; predicted membrane protein
GAC	0.37	GAT	0.63	+0.25	*panD*	Component of aspartate 1-decarboxylase
ACT	0.16	ACC	0.44	+0.28	*rhsC*	Function unknown, encoded within RhsC repeat
GAG	0.31	GAA	0.69	+0.38	*manX*	Mannose permease, EIIAB component
TTA	0.13	TTG	0.43	+0.30	*fepE*	Ferrienterobactin transport, membrane protein
ATA	0.07	ATC	0.42	+0.35	*yafD*	Conserved protein
AAG	0.23	AAA	0.77	+0.54	*csgD*	Two-component transcriptional regulator for second curli operon

Table S2. Local sequence context of transition mutations

			G:C or A:T					G:C or A:T	
5′p3′		5′p3′	Number	Fraction	5′p3′		5′p3′	Number	Fraction
Wild type*									
Gp**G**	+	**C**pC	13	0.16	**G**pG	+	Cp**C**	22	0.27
Cp**G**	+	**C**pG	27	0.33	**G**pC	+	Gp**C**	35	0.43
Ap**G**	+	**C**pT	16	0.20	**G**pA	+	Tp**C**	10	0.12
Tp**G**	+	**C**pA	26	0.32	**G**pT	+	Ap**C**	15	0.18
Sum			82		Sum			82	
Gp**A**	+	**T**pC	19	0.39	**A**pG	+	Cp**T**	5	0.10
Cp**A**	+	**T**pG	6	0.12	**A**pC	+	Gp**T**	26	0.53
Ap**A**	+	**T**pT	8	0.16	**A**pA	+	Tp**T**	8	0.16
Tp**A**	+	**T**pA	16	0.33	**A**pT	+	Ap**T**	10	0.20
Sum			49		Sum			49	
MutL$^-$*									
Gp**G**	+	**C**pC	166	0.37	**G**pG	+	Cp**C**	78	0.17
Cp**G**	+	**C**pG	165	0.37	**G**pC	+	Gp**C**	238	0.53
Ap**G**	+	**C**pT	37	0.08	**G**pA	+	Tp**C**	75	0.17
Tp**G**	+	**C**pA	79	0.18	**G**pT	+	Ap**C**	56	0.13
Sum			447		Sum			447	
Gp**A**	+	**T**pC	304	0.27	**A**pG	+	Cp**T**	114	0.10
Cp**A**	+	**T**pG	281	0.25	**A**pC	+	Gp**T**	897	0.79
Ap**A**	+	**T**pT	277	0.24	**A**pA	+	Tp**T**	57	0.05
Tp**A**	+	**T**pA	279	0.24	**A**pT	+	Ap**T**	73	0.06
Sum			1,141		Sum			1,141	

*The two bases indicated are on the same DNA strand with the mutated base in bold face; the "p" represents the phosphate linking the two nucleosides. The dinucleotides connected by a "+" are the same configuration but with the mutated base on opposite DNA strands.

Table S3. Base changes on the transcribed and the nontranscribed DNA strands

	Number on strand		
Base change	TS†	NTS†	Ratio* TS†/NTS†
Wild type			
Transition	49	53	0.92
A > G	18	15	1.20
G > A	31	38	0.82
Transversion	37	40	0.93
A > T	7	6	1.17
A > C	15	14	1.07
G > T	10	11	0.91
G > C	5	9	0.56
MutL$^-$			
Transition	695	686	1.01
A > G	493	494	1.00
G > A	202	192	1.05
Transversion	13	18	0.72
A > T	5	8	0.63
A > C	3	6	0.50
G > T	5	2	2.50
G > C	0	2	–

NTS, nontranscribed strand; TS, transcribed strand.
*None of the ratios are different from 1 ($P > 0.45$ in all cases).
†Only the purines are shown; because the complementary pyrimidine is on the other strand, the TS/NTS ratio for the pyrimidine is the inverse of the ratio shown.

Table S4. Indel hotspots

Nucleotide	Run length	Event	Number of indels	Probability*	Gene
Hotspots based on conservative calls					
255,168	6	+C	2	0.0002	*pepD*
307,271	7	+C	2	0.0009	*yagX*
631,233	7	+G	2		—
2,680,172	7	+C	2		*yphG*
3,929,199	7	+G	2		—
867,637	8	−C	2	0.008	*gsiA*
1,211,303[†]	8	+C, +CC	2		—
1,592,148	8	+C, −C	3	0.0004	—
2,460,902	8	−C, +C	3		—
4,604,109	8	+G	3		—
4,604,345[‡]	8	+G, −G	3		—
1,712,341	9	−T	2	0.046	—
34,111	9	−T	2		—
379,236	10	−G, +G, −GG	5	NA	—
Additional hotspots based on liberal calls					
251,009	6	+G	2	0.0002	*dinB*
370,806	6	+G	2		*mhpC*
2,903,235	6	+C	2		*queE*
4,607,352	7	+C	2	0.001	—
1,592,272	8	+C	2	0.019	—
1,435,246	9	−T	2	0.11	—
4,408,067	9	−T	2		—

NA, not applicable; there is only one run of 10 in the genome, so the probability is not meaningful.
*The probability of the observed number of events occurring at a given site.
[†]+CC occurred in the MutL^{-} strain; +C occurred in the wild-type strain.
[‡]+G occurred in the MutL^{-} strain; −G occurred in the wild-type strain.

Table S5. Estimated maximum rates for depurination of adenines and deamination of cytosines

	Rate per base per hour	Rate per base per generation	No. bases per genome	Rate per genome per generation
dsDNA				
Depurination of A	1.0×10^{-7}	3.3×10^{-8}	1.1×10^{6}	3.8×10^{-2}
Depurination of 6meA	2.5×10^{-7}	8.3×10^{-8}	3.8×10^{4}	3.2×10^{-3}
Deamination of C	1.4×10^{-9}	4.8×10^{-10}	1.2×10^{6}	5.6×10^{-4}
Deamination of 5meC	5.0×10^{-9}	1.7×10^{-9}	2.4×10^{4}	4.0×10^{-5}
ssDNA				
Depurination of A	4.0×10^{-7}	1.3×10^{-7}	$2–4 \times 10^{4}$	$2.6–5.2 \times 10^{-3}$
Depurination of 6meA	1.0×10^{-6}	3.3×10^{-7}	$0.7–1.3 \times 10^{3}$	$2.2–4.4 \times 10^{-4}$
Deamination of C	2.9×10^{-7}	9.5×10^{-8}	$2–4 \times 10^{4}$	$1.9–3.8 \times 10^{-3}$
Deamination of 5meC	1.0×10^{-6}	3.3×10^{-7}	$4–8 \times 10^{2}$	$1.3–2.7 \times 10^{-4}$