

FILE S1. SUPPLEMENTAL MATERIAL AND METHODS

Bacterial strains and media: The strains used in this study and the methods of their construction are listed in Supplemental Table S1. All are derivatives of PFM2 (LEE *et al.* 2012). The *mutD5* allele, a gift from Roel Schaaper (National Institutes of Health), was moved into PFM2 by recombineering using *cat*-Sce I cassette (BLANK *et al.* 2011) with the I-SceI recognition sequence inserted into the *rnhA* gene, which is close to the *dnaQ* gene. Subsequently, the *mutD5* allele was moved by P1 phage transduction (MILLER 1992) into strains made auxotrophic for proline, selecting for *proA*⁺. The *lexA3* allele was moved by P1 phage transduction into strains auxotrophic for methionine, selecting for *metA*⁺. Deletion mutations from the Keio collection (BABA *et al.* 2006) were moved by P1 phage transduction and the Kn^R element removed using FLP recombination (DATSENKO AND WANNER 2001). The presence of the *mutD5* allele was confirmed by increased mutation rate measured by fluctuation assays (FOSTER 2006) and by sequencing. Sequencing was also used to confirm that the *dnaE* gene in all *mutD5* mutant strains was wild-type. The *lexA3* allele was confirmed by sequencing. Deletions of other genes were confirmed by PCR. The oligonucleotides used for sequencing and PCR are given in Supplemental Table S2.

The values for the “wild-type” strain given in the tables and figures are the combined data from 8 experiments previously reported (LEE *et al.* 2012; FOSTER *et al.* 2015). The strains used are PFM2, wild type (two data sets, 3K and 6K generations), PFM35, *uvrA*, PFM40, *alkA tagA*, PFM88, *ada ogt*, PFM91, *nfi*, PFM101, *umuDC dinB*, PFM133, *umuDC dinB polB*, all of which had nearly identical BPS rates and spectra. The values for “MMR” strain given in the tables and figures are the combined data from 10 experiments with *mutL*, *mutS*, and *mutH* mutant strains reported in LEE

et al., 2012 and Foster *et al.*, the accompanying paper. The strains used are PFM5, *mutL*, PFM144, *mutL*, PFM288, *mutL*, PFM304, *mutL mutS*, PFM342, *mutS*, PFM343, *mutS*, PFM555/556, *mutS*, PFM197, *mutH*, PFM567, *mutLSH*, PFM568, *mutLSH*, all of which had nearly identical BPS rates and spectra.

The sequence reads from some of our strains map to the genome of bacteriophage Phi80, a common contaminant of *E. coli* K12 laboratory strains (ROTMAN *et al.* 2010). Thus, these strains appear to be lysogens for bacteriophage Phi80. However, we have found that such strains do not differ in growth or mutational phenotypes from uninfected strains in our experiments (Foster *et al.*, accompanying paper). Strains reported here that are likely to be Phi80 lysogens are PFM479, PFM515/M517, and PFM686.

For both liquid and agar plates rich medium was Miller Luria Broth (LB) (BD Difco) and minimal medium was Vogel-Bonner minimal (VB min) with 0.2% glucose (MILLER 1992). When required the antibiotic concentrations were: anhydrotetracycline (AHT), 500 µg/mL; carbenicillin (Carb), 100 µg/mL; and kanamycin (Kn), 50 µg/mL; nalidixic acid (Nal), 40 µg/mL, chloramphenicol (Cam), 30 µg/mL, rifampicin (Rif), 100 µg/mL. Half of these concentrations was used in minimal medium.

Mutation accumulation experiments: The MA procedure has been described (LEE *et al.* 2012; FOSTER *et al.* 2015). The MA lines originated from single colonies isolated from two or more founder colonies. The founder colonies were generated by streaking from a freezer stock onto LB agar plates, or VBmin agar plates for MA experiments done on minimal medium, and incubating the plate overnight at 37°C. One or two well isolated colonies were then excised from the agar plate, soaked for 30 minutes in 0.85% NaCl + 0.01% gelatin, and vortexed for 60

seconds. Appropriate dilutions for obtaining well-isolated colonies were then plated onto fresh LB or VB min agar plates to start the MA lines.

Genomic DNA preparation, library construction, sequencing, SNP and indel calling: Genomic DNA was isolated using PureLink Genomic DNA purification kit (Invitrogen). DNA concentrations were measured on an Epoch Microplate Spectrophotometer (BioTek Instruments, Inc.). Paired end libraries were made by the Indiana University Center for Genomics and Bioinformatics and were sequenced at the University of New Hampshire Hubbard Center for Genome Studies using the Illumina HiSeq 2500 platform. Insert size was between 350 and 600 base pairs.

SNP and indel calling were as described (LEE *et al.* 2012). *E. coli* MG1655 (NCBI reference sequence NC_000913.2) was used as the reference genome. The Illumina reads were aligned to the referenced genome using the Burrows-Wheeler short-read alignment tool, BWA version 0.7.9 (LI AND DURBIN 2009). Short indels (≤ 4) were called based on the read mapping of the SNP calling procedures.

Statistical Analysis of Mutation Rates: The complete method used is given in Foster *et al.*, accompanying paper. Briefly, m = mutations per MA line, g = generations per MA line, and the mutation rate of a MA line, r , is m/g . The overall mutation rate is $R = M/G$ where $M = \sum m$ and $G = \sum g$ over all MA lines. Calculated of the variance of R was adapted from (MANDANSKY 2010)

$$VarR = \left(\frac{\sum (R - r)^2}{N - 1} \right) \left(\frac{\sum g^2}{G^2} \right)$$

where N = number of MA lines being considered. To combine the results from multiple experiments, the same calculations were used except $(N-1)$ was replaced with $\sum (N-1)$, i.e. $(N-1)$ summed over all experiments. Then the Standard Deviation $= \sqrt{V}$ and 95% CL $= \sqrt{V} \times t_{0.05, \sum (N-1)}$

Mutation Annotation: Variants were annotated using custom scripts. Protein coding gene coordinates were obtained from the GenBank page of reference sequence, NC_000913.2. BPSs in coding sequences were determined to be synonymous or nonsynonymous based on the genetic code. Nonsynonymous BPSs were designated conservative or nonconservative based on the Blosum62 matrix (HENIKOFF AND HENIKOFF 1994) with a value ≥ 0 considered conservative.

Monte Carlo Simulations: For each strain a random distribution of BPSs corresponding to the observed mutational spectra was simulated using a custom script; the number of mutations was fixed at the observed numbers and 1,000 trials were simulated (LEE *et al.* 2012)

SUPPLEMENTAL REFERENCES

- Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura *et al.*, 2006 Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* 2: 2006.
- Blank, K., M. Hensel and R. G. Gerlach, 2011 Rapid and highly efficient method for scarless mutagenesis within the *Salmonella enterica* chromosome. *PLoS One* 6: e15763.
- Datsenko, K. A., and B. L. Wanner, 2001 One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences, USA* 97: 6640-6645.
- Foster, P. L., 2006 Methods for determining spontaneous mutation rates. *Methods in Enzymology* 409: 195-213.
- Foster, P. L., H. Lee, E. Popodi, J. P. Townes and H. Tang, 2015 Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proc Natl Acad Sci U S A* 112: E5990-E5999.

Henikoff, S., and J. G. Henikoff, 1994 Protein family classification based on searching a database of blocks. *Genomics* 19: 97-107.

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

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25: 1754-1760.

Mandansky, A., 2010 Weighted Standard Error and its Impact on Significance Testing,
www.analyticalgroup.com/download/WEIGHTED_MEAN.pdf

Miller, J. H., 1992 *A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Rotman, E., L. Amado and A. Kuzminov, 2010 Unauthorized horizontal spread in the laboratory environment: the tactics of Lula, a temperate lambdoid bacteriophage of *Escherichia coli*. *PLoS One* 5: e11106.