Specificity of the DNA Mismatch Repair System (MMR) and Mutagenesis Bias in Bacteria

Hongan Long,*[,1] Samuel F. Miller,[2] Emily Williams,[2] and Michael Lynch[2]

[1]Institute of Evolution & Marine Biodiversity, KLMME, Ocean University of China, Qingdao, Shandong, China

[2]Center for Mechanisms of Evolution, The Biodesign Institute, Arizona State University, Tempe, AZ

***Corresponding author:** E-mail: longhongan@gmail.com.

Associate editor: Iñaki Ruiz-Trillo

Abstract

The mutation rate of an organism is influenced by the interaction of evolutionary forces such as natural selection and genetic drift. However, the mutation spectrum (i.e., the frequency distribution of different types of mutations) can be heavily influenced by DNA repair. Using mutation-accumulation lines of the extremophile bacterium *Deinococcus radiodurans* Δ*mutS1* and the model soil bacterium *Pseudomonas fluorescens* wild-type and MMR$^-$ (Methyl-dependent Mismatch Repair-deficient) strains, we report the mutational features of these two important bacteria. We find that *P. fluorescens* has one of the highest MMR repair efficiencies among tested bacteria. We also discover that MMR of *D. radiodurans* preferentially repairs deletions, contrary to all other bacteria examined. We then, for the first time, quantify genome-wide efficiency and specificity of MMR in repairing different genomic regions and mutation types, by evaluating the *P. fluorescens* and *D. radiodurans* mutation data sets, along with previously reported ones of *Bacillus subtilis* subsp. *subtilis*, *Escherichia coli*, *Vibrio cholerae*, and *V. fischeri*. MMR in all six bacteria shares two general features: 1) repair efficiency is influenced by the neighboring base composition for both transitions and transversions, not limited to transversions as previously reported; and 2) MMR only recognizes indels <4 bp in length. This study demonstrates the power of mutation accumulation lines in quantifying DNA repair and mutagenesis patterns.

Key words: **DNA mismatch repair, mutation rate and spectrum, neutral evolution, spontaneous mutation.**

Article

Introduction

Studying mutations not only provides insight into the evolution of organisms but also into pathogenesis caused by either germline or somatic mutations. Previous studies have shown that mutations are nonrandom with respect to genomic regions and neighboring nucleotide contexts, as are the premutations that are fixed by repair systems (Benzer 1961; Jones et al. 1987; Friedberg et al. 2006). However, these biases have not been rigorously quantified because of the numerous assumptions and the limited genomic regions covered by reporter-construct methods applied in most earlier mutational studies. Quantifying such biases in a statistically rigorous manner requires a large number of mutations acquired in a nonselective fashion.

Fortunately, the mutation accumulation (MA) technique conquers these difficulties by imposing repeated single-individual bottlenecks on large numbers of parallel lineages over hundreds to thousands of cell divisions (Bateman 1959; Mukai 1964). During this process, the extreme population bottlenecks maximize random genetic drift, which greatly reduces the efficiency of selection, so that essentially all mutations except a tiny fraction with lethal fitness effects can proceed to fixation. Deep whole-genome sequencing and advanced bioinformatic tools further promote this experimental procedure by directly and accurately detecting mutations in the MA lines (Lynch et al. 2008). Furthermore, genome-wide mutagenesis prior to MMR involvement can be revealed in vivo by mutations accumulated in MMR-deficient (MMR$^-$) strains (Lee et al. 2012; Foster et al. 2013; Long, Kucukyildirim, et al. 2015; Long, Sung, et al. 2015). By comparing the molecular mutation spectra of wild-type and MMR$^-$ strains of phylogenetically diverse bacteria, general or lineage-specific patterns in mutagenesis and DNA repair can be directly and systematically studied.

MMR, one of the major DNA repair pathways, fixes mismatched base pairs in DNA (reviewed in Kunkel and Erie 2005). In this study, we ran MA experiments on an MMR$^-$ strain of *Deinococcus radiodurans* R1 (the most radiation-resistant organism known on the earth) and both wild-type and MMR$^-$ strains of another model bacterium *Pseudomonas fluorescens* SBW25 (Gram-negative). Together with recently published MA data sets for wild-type *Bacillus subtilis* subsp. *subtilis* NCIB 3610 (Gram positive, Sung et al. 2015), *D. radiodurans* BAA-816 (Gram positive, Long, Kucukyildirim, et al. 2015), *Escherichia coli* MG1655 (Gram negative, Lee et al. 2012; Long et al. 2016), *Vibrio cholerae* 2740-80, and *V. fischeri* ES114 (Gram negative, Dillon et al. 2017), we quantified the efficiency/specificity of the MMR system and mutagenesis bias in bacteria.

Results and Discussion

General information on all bacterial MA data sets is listed in table 1. Mutation details of wild-type, Δ*mutS*, and Δ*mutL*

Table 1. MA Data Sets.

Species	GC%	μ_{BPS}	ts/tv	μ_{indel}	μ_{indel_CDS}	μ_{indel_NCDS}	in/del	N	G	References
Bacillus subtilis	43.35	3.28	3.23	1.18	0.69	2.16	0.32	50	5,078	Sung et al. (2015)
B. subtilis ΔmutS	43.35	101×[a]	3×	44×[b]	64×[b]	48×[b]	1.56[b]	19	2,000	Sung et al. (2015)
Deinococcus radiodurans[c]	66.61	4.99	1.71	0.22	0.21	0.26	1.13	43	5,961	Long, Kucukyildirim, et al. (2015)
D. radiodurans ΔmutL[c]	66.61	4×	12×	21×	18×	51×	0.84	24	993	Long, Kucukyildirim, et al. (2015)
D. radiodurans ΔmutS1[c]	66.61	4×	9×	15×	13×	31×	0.53	29	3,101	This study
Escherichia coli	50.79	2.02	1.28	0.37	0.28	1.84	0.60	59	4,246	Lee et al. (2012); Sung et al. (2016)
E. coli ΔmutL	50.79	136×	34×	141×	–	–	–	34	375	Lee et al. (2012)
E. coli ΔmutS[d]	50.79	115×	39×	119×	80×	91×	1.33	12	763	Long et al. (2016)
Pseudomonas fluorescens	60.50	0.93	1.72	0.12	0.11	0.19	0.39	80	5,141	This study
P. fluorescensΔmutL	60.50	279×	48×	176×	149×	332×	11.89	30	1,465	This study
P. fluorescens ΔmutS	60.50	309×	45×	206×	165×	389×	11.77	30	1,454	This study
Vibrio cholerae	47.57	1.07	1.56	1.71	0.15	0.31	0.29	49	6,453	Dillon et al. (2017)
V. cholerae ΔmutS	47.57	85×	29×	142×	116×	234×	1.07	22	1,254	Dillon et al. (2017)
V. fischeri	38.35	2.07	0.86	5.68	0.35	2.03	0.58	48	5,187	Dillon et al. (2017)
V. fischeri ΔmutS	38.35	317×	37×	102×	121×	83×	0.77	19	810	Dillon et al. (2017)

[a]Mutation rates in MMR^- strains are shown in fold-change relative to the wild-type rates. GC% is the G/C content of the genome; μ_{BPS} and μ_{indel} are genome-wide mutation rates of base substitutions and indels per site per cell division; *ts/tv* is the ratio of transition to transversion mutations; μ_{indel_CDS}, μ_{indel_NCDS} are conditional indel mutation rates per site per cell division in coding and noncoding sequences, respectively; in/del is the ratio of insertions to deletions; *N* is the number of MA lines; *G* is the average number of cell divisions that each MA line passed during the experiment. All mutation rates are in units of $\times 10^{-10}$ per nucleotide site per cell division.
[b]Reanalysis of published MA-line sequences using GATK.
[c]All indels and BPSs are chromosomal.
[d]Based on MA lines SA1–SA12 (Long et al. 2016).

P. fluorescens SBW25 and Δ*mutS1 D. radiodurans* R1, which are first reported in this study, as well as reanalysis of results based on published raw sequences, are in supplementary data set D1: supplementary tables S1—S10, Supplementary Material online.

Mutation Rates of *P. fluorescens* and *D. radiodurans* Strains

For *P. fluorescens*, we detected 253 base-pair substitutions (BPSs) in 80 SBW25 wild-type MA lines, 8,425 in 30 Δ*mutS* MA lines, and 7,529 in 30 Δ*mutL* MA lines. These yield BPS mutation rates of $9.31 \pm 0.66 \times 10^{-11}$ (SEM; one of the lowest mutation rates in bacteria), $2.87 \pm 0.12 \times 10^{-8}$ and $2.59 \pm 0.09 \times 10^{-8}$ per nucleotide site per cell division for the wild-type, Δ*mutS*, and Δ*mutL* strains, respectively. Knocking out *mutS* elevates the mutation rate ~309-fold, to a level that is highly similar to that of a naturally *mutS*-deficient *P. fluorescens* strain, 2.34×10^{-8} (Long, Sung, et al. 2015). Knocking out *mutL* inflates the *P. fluorescens* mutation rate ~278-fold. The extremely high mutation-rate elevation of MMR^- lines makes clear that this species is one of the most efficient among bacteria in repairing BPS premutations by MMR (table 2).

BPSs of the GC-rich *P. fluorescens* SBW25 wild-type strain do not show significant bias, with $\mu_{G/C\rightarrow A/T}$ (the mutation rate in the A: T direction) being 9.85×10^{-11} (95% Poisson confidence intervals: 8.30×10^{-11} to 11.59×10^{-11}), and $\mu_{A/T\rightarrow G/C}$ (the mutation rate in the G: C direction) being 8.39×10^{-11} (6.66×10^{-11} to 11.45×10^{-11}). Details of the mutation spectra for all *P. fluorescens* strains are shown in supplementary data set D1: supplementary tables S1—S3, Supplementary Material online and figure 1.

We also detected 32 small indels in the *P. fluorescens* wild-type MA lines, yielding a small-indel mutation rate of $1.18 \pm 0.21 \times 10^{-11}$ per nucleotide site per cell division. The insertion/deletion ratio is 0.39, showing that this strain has a deletion bias. Knocking out *mutS* or *mutL* increases the small-indel mutation rates to 2.47×10^{-9} and 2.11×10^{-9} per nucleotide site per generation, respectively, and the insertion/deletion ratio to 11.77 and 11.89 (table 1). The insertion-bias in MMR^- and deletion-bias in MMR^+ *P. fluorescens* shows that MMR preferentially repairs insertions over deletions.

For *D. radiodurans*, 418 BPSs accumulated in the chromosomes of 29 Δ*mutS1* MA lines (a full list of BPSs including plasmid BPSs are in supplementary data set D1: supplementary table S7, Supplementary Material online). The chromosomal BPS mutation rate for this strain is $1.84 \pm 0.08 \times 10^{-9}$ per nucleotide site per cell division, which is ~3.3× of the wild-type chromosomal mutation rate (Mennecier et al. 2004; Long, Kucukyildirim, et al. 2015). This low-level elevation of the mutation rate makes *D. radiodurans* MMR one of the least efficient among bacteria (table 2). This may be compensated by other highly efficient alternative DNA repair pathways, such as the oxidative damage repair pathways (Mennecier et al. 2004; Long, Kucukyildirim, et al. 2015), or by multiple copies of the genome enhancing gene conversion as a repair strategy. The $\mu_{A/T\rightarrow G/C}$ mutation rate 2.92×10^{-9} (95% Poisson confidence intervals: 2.54×10^{-9} to 3.33×10^{-9}) is significantly higher than $\mu_{G/C\rightarrow A/T}$ 1.25×10^{-9} (1.08×10^{-9} to 1.44×10^{-9}), yielding a GC-bias close to that in the wild-type strain (Long, Kucukyildirim, et al. 2015).

75 small indels accumulated in the *D. radiodurans* Δ*mutS1* chromosomes, yielding a chromosomal small-indel mutation rate of $3.31 \pm 0.50 \times 10^{-10}$. Deletions are more abundant

Table 2. MMR Efficiency of Different Types of Mutations in Bacteria.

Mutations	*Bacillus subtilis*	*Deinococcus radiodurans*	*Escherichia coli*	*Pseudomonas fluorescens*	*Vibrio cholerae*	*V. fischeri*
G:C>A:T	0.9925 (0.0007)	0.7533 (0.0270)	0.9867 (0.0027)	0.9982 (0.0002)	0.9887 (0.0015)	0.9973 (0.0003)
A:T>G:C	0.9922 (0.0007)	0.8779 (0.0161)	0.9975 (0.0007)	0.9995 (0.0001)	0.9967 (0.0008)	0.9993 (0.0001)
A:T>T:A	0.9327 (0.0152)	–[a]	0.7407 (0.1399)	0.9935 (0.0046)	0.8691 (0.0845)	0.9935 (0.0034)
A:T>C:G	0.9051 (0.0230)	–[a]	0.7055 (0.1079)	0.8933 (0.0335)	0.5056 (0.3096)	0.7321 (0.1213)
G:C>T:A	0.7749 (0.0719)	–[a]	0.8622 (0.1022)	0.9049 (0.0204)	0.4547 (0.2937)	0.8582 (0.0291)
G:C>C:G	0.9579 (0.0168)	0.7119 (0.1399)	0.7659 (0.2034)	0.9658 (0.0072)	0.9128 (0.0504)	0.9485 (0.0221)
Total BPS	0.9901 (0.0005)	0.7294 (0.0196)	0.9906 (0.0011)	0.9968 (0.0002)	0.9882 (0.0011)	0.9968 (0.0002)
Insertions	0.9907 (0.0017)	0.8991 (0.0408)	0.9938 (0.0016)	0.9985 (0.0005)	0.9969 (0.0014)	0.9918 (0.0019)
Deletions	0.9710 (0.0051)	0.9524 (0.0192)	0.9862 (0.0029)	0.9558 (0.0110)	0.9888 (0.0029)	0.9893 (0.0019)
Total indels	0.9830 (0.0022)	0.9335 (0.0188)	0.9905 (0.0015)	0.9952 (0.0009)	0.9930 (0.0016)	0.9902 (0.0013)
Indels_SSR	0.9852 (0.0009)	0.9618 (0.0152)	0.9925 (0.0013)	0.9980 (0.0001)	0.9974 (0.0009)	0.9935 (0.0011)
Indels_non_SSR	0.7628 (0.1113)	0.8126 (0.0828)	0.4573 (0.5607)	0.7991 (0.0784)	0.3892 (0.4617)	0.7064 (0.1615)
Indels_coding	0.9844 (0.0020)	0.9233 (0.0234)	0.9876 (0.0028)	0.9940 (0.0012)	0.9914 (0.0022)	0.9917 (0.0016)
Indels_noncoding	0.9792 (0.0039)	0.9676 (0.0256)	0.9890 (0.0022)	0.9975 (0.0011)	0.9957 (0.0020)	0.9880 (0.0025)
Ne	6.119×10^{7}	–	1.796×10^{8}	–	4.783×10^{8}	–

NOTE.—SEs, calculated with equation A1.19b in Lynch and Walsh (1998) assuming no covariance between wild-type and Δ*mutS* MA lines, are in parentheses. Each measure of repair efficiency represents the proportion of mutations repaired by MMR (1.0 implying that all premutations are completely repaired) and is calculated by $(\mu_{\Delta mutS}-\mu_{+})/\mu_{\Delta mutS}$, where $\mu_{\Delta mutS}$ is the mutation rate of the Δ*mutS* MA lines and μ_{+} is the mutation rate of the wild-type MA lines.
[a]Negative number; Indels_SSR and Indels_non_SSR represent indel repair efficiencies in SSR and non-SSR regions; Indels_coding and Indels_noncoding are indel repair efficiencies in coding and noncoding regions; *Ne*, effective population size. Data sources: for *B. subtilis*, BPS data from Sung et al. (2015), indel data from this study; *D. radiodurans*, Long, Kucukyildirim, et al. (2015), this study; *E. coli*, (wild-type indel data were from Lee et al. 2012; all BPS and Δ*mutS* data were from Long et al. 2016); *P. fluorescens*, this study; *V. cholerae* and *V. fischeri*, Dillon et al. (2017); *Ne* data from Supplementary Information S1 of Lynch et al. (2016).

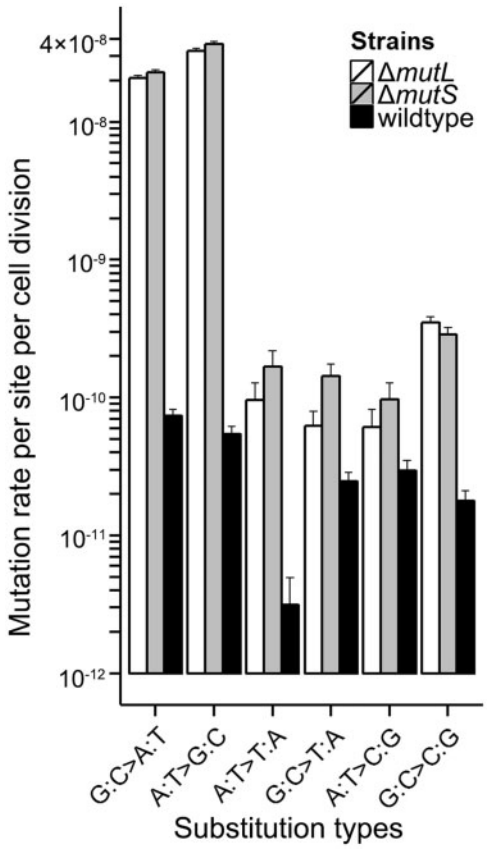

FIG. 1. Conditional mutation rates of *Pseudomonas fluorescens* SBW25 wild-type and MMR⁻ strains. Error bars are SEM.

than insertions, with an insertion/deletion ratio of 0.53, which is highly unusual in other MMR⁻ bacteria (table 1). The deletion bias in MMR⁻ *D. radiodurans* implies that DNA polymerases of *D. radiodurans* generally have more forward slippage errors than backward ones. This might be explained by insertions requiring melting and replication of a previously replicated DNA fragment, whereas only skipping of unreplicated fragments is needed for deletions (Petrov 2002), assuming that no other repair pathways are involved in this process in *D. radiodurans* MMR⁻ strains.

Mutation Spectrum

Mutations in MMR⁻ MA lines reveal the primary mutation spectrum of organisms, unmodified by the subsequent MMR pathway. Among MMR⁻ strains of all six bacteria—*B. subtilis*, *D. radiodurans*, *E. coli*, *P. fluorescens*, *V. cholerae*, and *V. fischeri*, transitions are ~16- to 82-fold more abundant than transversions, in contrast to only 0.86- to 3.23-fold in MMR-functional strains (table 1). The higher relative abundance of transitions in MMR⁻ strains reflects the fact that major DNA polymerases are prone to cause transition mutations (Curti et al. 2009).

Most indels in MMR⁻ strains are found in simple sequence repeats (SSR), which are tandem repeats of short DNA nucleotides, such as homopolymer runs, and microsatellite repeats, and are usually generated by unequal crossover or polymerase slippage during replication of SSRs (Levinson and Gutman 1987). For example, 81–99% of small-indels are in SSR motifs in *mutS*-inactivated MA lines (supplementary data set D1: supplementary tables S5, S8, and S10). In order to explore patterns of the indel mutation rate at SSR motifs, we analyzed the indel mutation rate at 3–9 bp homopolymer runs for each category (≥10 bp homopolymers are extremely rare or missing in the genomes of the four studied bacteria). As shown in figure 2, for either A/T or G/C homopolymers, the indel mutation rate per homopolymer per generation increases with homopolymer length, presumably due to a decline in proofreading efficiency of DNA polymerases as homopolymer-run length increases (Lujan et al. 2015). The indel mutation rate at long homopolymer motifs can be

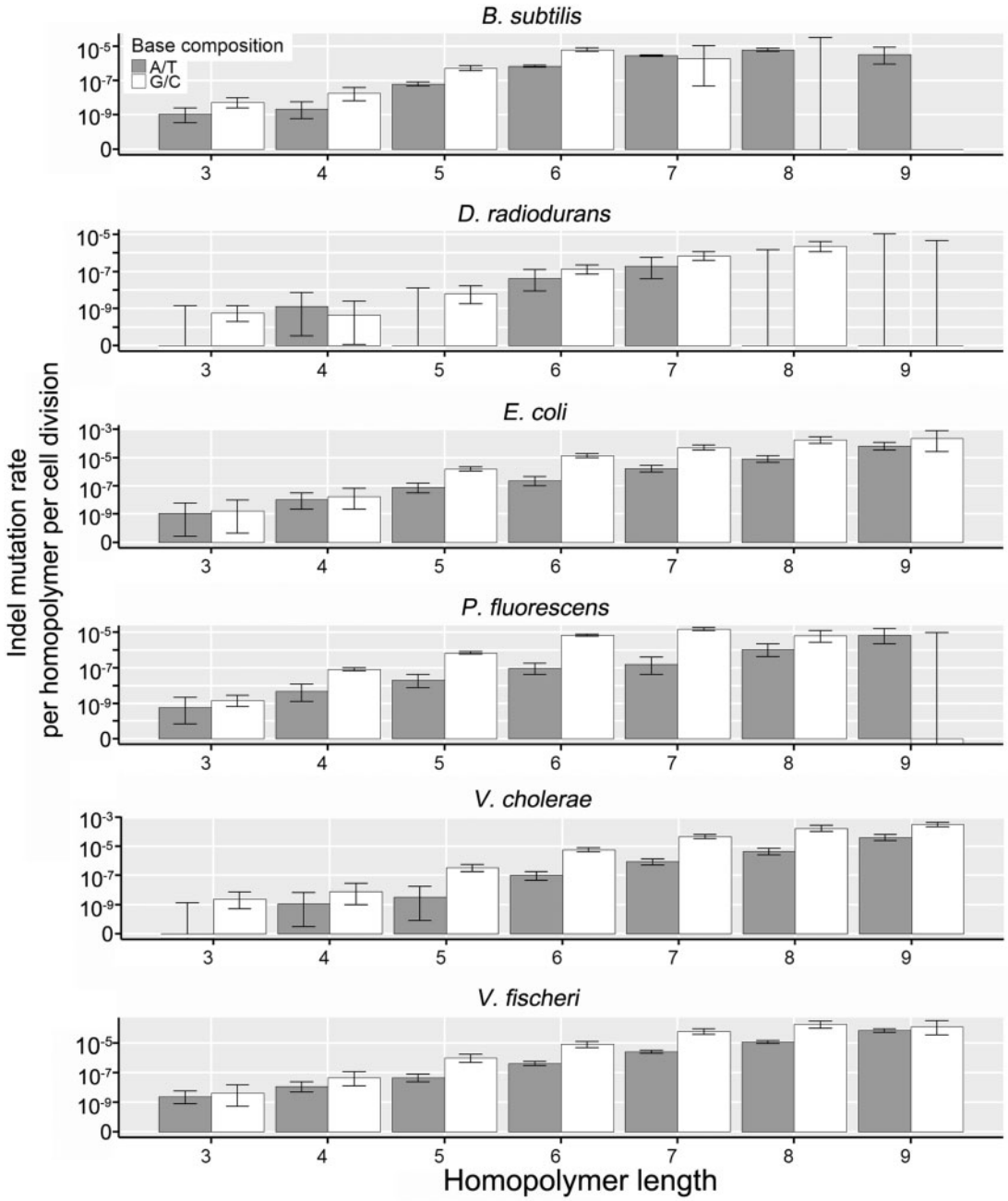

Fig. 2. Conditional indel mutation rates at different homopolymer-run lengths in MMR$^-$ strains. Error bars are 95% Poisson confidence intervals.

orders of magnitude higher than that of shorter ones; for example, the indel rate of 5′GGGGGGGG3′ and 5′AAA3′ motifs in *E. coli* differs by 200,861×, the same order of magnitude increase in the indel rate at homopolymer runs in yeast MMR$^-$ MA lines (Lujan et al. 2015). The very high indel mutation rate of long homopolymers could also explain the rarity of longer homopolymers in organisms' genomes, which may have been eroded by indels during long-term evolution.

The base composition of homopolymers also influences the indel mutation rate, with G or C homopolymers having a significantly higher indel mutation rate than A or T homopolymers (fig. 2). This may be explained by the stronger base-stacking interaction of G: C base pairs compared with A: T (Sagher et al. 1999), as well as by stronger base-pairing of G: C (Lujan et al. 2015). Thus, the relative abundance of G/C and A/T homopolymers in a genome influences the indel mutation rate.

MMR Repair Specificity

Using mutation rates of wild-type and MMR$^-$ MA lines, we calculated the repair efficiency of different types of mutations (table 2). The repair efficiency and effective population size (*Ne*) are not correlated with each other, though there are only three available data points. This may be a simple consequence of selection operating on the overall mutation rate, leaving degrees of freedom for individual components to evolve in mutually compatible upward/downward fashions, e.g., a local decrease in MMR efficiency may be balanced by an increased in accuracy at the polymerization and/or proofreading steps (Lynch 2012).

Although transitions are preferentially repaired over transversions in general, we find that MMR-repair specificity is highly variable among bacterial species and different types of base-pair substitutions (table 2). Our results suggest that A: T→T: A, A: T→C: G, and G: C→T: A premutations are not repaired by MMR in *D. radiodurans* (tables 1 and 2). MMR specificity for insertions/deletions in *D. radiodurans* is also unique: contrary to all other observed bacteria, deletions are repaired more efficiently than insertions in *D. radiodurans*. This is consistent with a prior argument that the MMR system in *D. radiodurans* is extremely divergent from that of most other bacteria (Mennecier et al. 2004). Considering the insertion/deletion ratio of 1.13 in the wild-type strain versus the above Δ*mutS1* ratio of 0.53 (table 1), MMR of *D. radiodurans* preferentially repairs deletions. The deletion-preference of MMR might be associated with recombination-bias caused by the polyploid nature of the *D. radiodurans* genome (four to ten ploidy). Alternatively, this might be an adaptive trait, which has evolved under the extreme radiation level that all other bacteria cannot tolerate, assuming *D. radiodurans* is geographically distributed in high-radiation habitats.

We also find some general features of MMR specificities in repairing indels across all six bacteria: only indel rates of 1–3 bp size are significantly lower in wild-type than in MMR^- strains, that is, these short indels are efficiently repaired by MMR, but larger ones are not (supplementary data set 1: supplementary tables S13 and S14, Supplementary Material online). This finding is highly consistent with previous in vitro and in vivo studies on *E. coli* MMR repair of indels (Learn and Grafstrom 1989; Parker and Marinus 1992; Carraway and Marinus 1993), which found that MMR only repairs insertion–deletion loops smaller than 5 bp and only 1–3 bp loops efficiently. Such indel-repair specificity of MMR appears to be explained by the biochemical property of MutS, which binds to heteroduplexes with 1–4 unpaired bases, but not to those with five or more unpaired bases (Learn and Grafstrom 1989; Parker and Marinus 1992; Carraway and Marinus 1993). Such specificity is not only influenced by indel size but also by the genomic regions within which indels occur, for example, indels occurring within SSR motifs are preferentially repaired relative to those at non-SSR genomic regions (table 2 and supplementary data set D1: supplementary table S11, Supplementary Material online), consistent with previous findings in eukaryotic organisms (Sia et al. 1997; Denver et al. 2005). The low MMR repair efficiency at non-SSR regions may be explained by the possibility that the key MMR protein MutS preferentially binds low-complexity DNA regions.

Flanking nucleotides such as guanines or cytosines can elevate the base-substitution mutation rate by orders of magnitude via their differential base-stacking and base-pairing power (Yakovchuk et al. 2006; Lee et al. 2012, 2014; Long, Sung, et al. 2015; Sung et al. 2015). Nucleotide-context was also reported to influence MMR efficiency for repairing transversions in *E. coli* (Jones et al. 1987). Here, we explore such context-dependent MMR efficiency (context is defined as the content of the two flanking nucleotides on the same DNA strand) in the three non-*E. coli* bacterial species in this study. We observed higher repair efficiency in contexts with at least one adjacent G: C pair for some species. This trend applies to both transition and transversion mutations, and is not limited to transversions as previously reported (Jones et al. 1987). Contexts without G: C pairs are usually among the ones with the lowest repair efficiency, such as 5′AGA3′ (fig. 3 and supplementary data set D1: supplementary table S12, Supplementary Material online). By contrast, we did not observe context-dependency of the indel repair-efficiency, which may be convoluted by that most indels occur in SSRs.

Correlation between the Base-Substitution and Indel Mutation Rate

Because small indels and base substitution experience similar DNA replication and repair processes, their rates are likely not independent of each other. After pooling the *Pseudomonas fluorescens* data set in this study with another 23 published MA data sets of various MMR-functional bacteria (supplementary table S15, Supplementary Material online), we find a strong positive correlation between the base-substitution and small-indel mutation rates (Pearson's correlation test, $r = 0.97$, $P = 1.72 \times 10^{-14}$), consistent with another recent study using both eukaryotic and prokaryotic organisms (Sung et al. 2016). This correlation still holds even when MMR is inactive ($r = 0.84$, $P = 0.04$; supplementary table S15, Supplementary Material online). The ratios of small-indel to base-substitution mutation rates are similar in different bacteria, regardless of whether they are MMR functional, 0.22 ± 0.04 (SEM), or dysfunctional, 0.16 ± 0.03.

The aforementioned shared DNA-replication and repair machinery could be one cause of the correlation between base-substitution and indel mutation rates. Another potential explanation is an elevation of the base-substitution mutation rate around indel hotspots (McDonald et al. 2011), owing to indels stalling high-fidelity DNA polymerases and recruiting error-prone translesion DNA polymerases that increase the downstream mutation rate.

Conclusions

We report the genomic mutation rate and spectrum of wild-type and MMR^- *P. fluorescens*, and Δ*mutS1 D. radiodurans*. We find that *P. fluorescens* has one of the lowest mutation rates in bacteria and the highest MMR repair efficiency among the six studied bacteria. Using mutations in these original data sets and previously reported studies, for the first time, we quantify the efficiency and specificity of MMR over different types and genomic locations of mutations. We also discover that the MMR of *D. radiodurans* preferentially repairs deletions over insertions, thus causing a unique insertion-bias in this extremophile bacterium. Some features of mutagenesis and MMR are shared among all studied species: 1) the influence of the flanking nucleotide base-composition of a mismatch on MMR efficiency, with MMR preferentially repairing sites with at least one flanking G/C nucleotide; 2) G/C homopolymers having higher indel mutation rates than A/T homopolymers; and 3) MMR specifically repairing indels with $1-3$ base pairs. This research

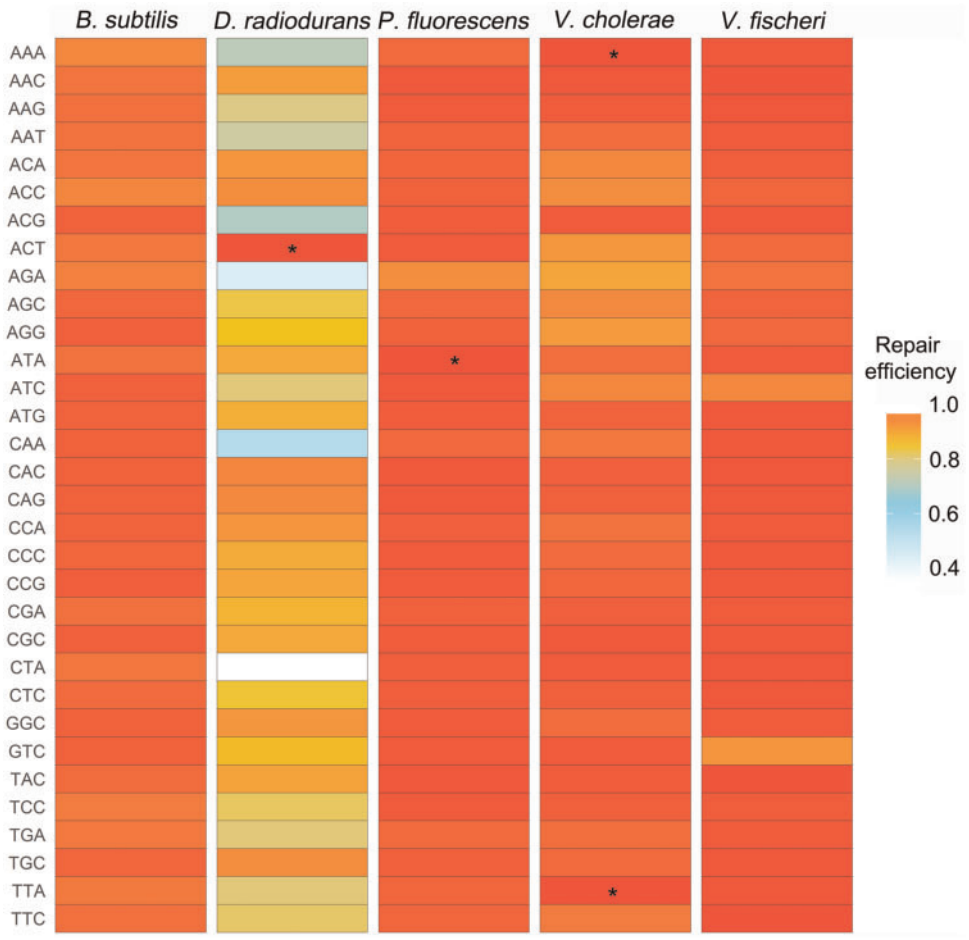

FIG. 3. Heatmap of MMR repair efficiency in different bacteria. Trinucleotides on the left show the nucleotide contexts in the 5′ to 3′ direction (reverse-complement of one nucleotide-context is deemed as the same context, due to the uncertainty of the DNA strand of each mismatch causing the base-substitution), with the focal bases in the middle. Red bars with stars denote 100% repair efficiency, an outcome of inadequate mutation sample size in the wild-type lines.

demonstrates the effectiveness of mutation accumulation techniques in studying the specificity of the DNA repair systems.

Materials and Methods

MA Lines Information

The *Pseudomonas fluorescens* SBW25 wild-type strain was provided by the Vaughn Cooper lab, University of Pittsburgh, and we created 80 MA lines from a single-cell ancestor. We single-colony transferred MA lines every other day for 244 times per line. The *P. fluorescens* SBW25 Δ*mutS* (30 MA lines) and Δ*mutL* (30 MA lines) strains were provided by Angus Buckling, University of Exeter Cornwall Campus, UK (Pal et al. 2007; O'Brien et al. 2013), and MA lines from these mutator strains were transferred 71 times on average. The *Deinococcus radiodurans* R1 Δ*mutS1* (we initiated 30 MA lines, line D33 was removed from the final analysis due to cross contamination detected after genome sequencing) was from the Suzanne Sommer lab, Université Paris-Sud 11, Orsay, France and MA lines from this strain were transferred ~130 times. We cultured all *P. fluorescens* MA lines on nutrient agar (Becton, Dickinson and Company, Sparks, MD) at 25°C; we grew and single-colony transferred *D. radiodurans* Δ*mutS1* MA lines at 30°C every other day on nutrient agar supplemented with 1% glucose. About every five weeks, we estimated cell divisions that the MA lines had passed by CFU (colony forming units) of diluted single colonies from ten randomly selected MA lines; the mean cell divisions from a single cell to a colony were estimated by $\log_2$(CFU). The total cell division number of each MA line is the product of the grand mean (21.1 cell divisions for *P. fluorescens* wild-type, 20.5 for *P. fluorescens* Δ*mutS*, 20.7 for *P. fluorescens* Δ*mutL*; 23.8 for *D. radiodurans* Δ*mutS1*) of all cell division estimates and the total transfer number for each line.

Genome Sequencing

We extracted DNA of the final MA lines using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI). DNA libraries were constructed using the Nextera DNA Library Preparation Kit (Illumina). We then size-selected the libraries for an insert size of 300 bp and sequenced them using HiSeq2500 2 × 150 rapid run at the Hubbard Center for Genome Studies, University of New Hampshire. Median depths of coverage were 70× (*P. fluorescens* wild-type), 62× (*P. fluorescens* Δ*mutS*), 78× (*P. fluorescens* Δ*mutL*) and

172× (*D. radiodurans* Δ*mutS1*). All raw sequence reads first reported in this study and those used but not published in previous studies (Long, Kucukyildirim, et al. 2015; Long, Sung, et al. 2015) were deposited in NCBI SRA (BioProject No. PRJNA397203).

Mutation Analyses

We trimmed off adaptors in raw reads with Trimmomatic 0.32 (Bolger et al. 2014), and then mapped reads to the reference genome using BWA-0.7.10 mem (GenBank genome accession numbers: NC_012660.1 for *P. fluorescens* SBW25; NC_000958.1—plasmid MP1, NC_000959.1—plasmid CP1, NC_001263.1—chromosome 1, NC_001264.1—chromosome 2 for *D. radiodurans* R1; GCA_000186085.1 for *B. subtilis* subsp. *subtilis* NCIB 3610) (Li and Durbin 2009). Duplicate reads were removed using picard-tools-1.141 and reads mapping around indels were realigned using GATK-3.5; SNP and indel discovery was performed with standard hard filtering parameters described by GATK Best Practices recommendations (except that Phred-scaled quality score QUAL >100 and RMS mapping quality MQ >59 for both variant and invariant sites) (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). Base-pair substitutions and small indels were called using the UnifiedGenotyper in GATK. We also required >99% of reads in a line to determine the line-specific consensus nucleotide at a candidate site—1% was set to allow for aberrant reads originated from sequencing errors, not absolutely pure indexes during library construction, or barcode degeneracy during sequence demultiplexing. Mutation rate μ was calculated by $\mu = \frac{m}{\sum_{1}^{n} N \times T}$, where m is the total number of mutations pooled from all MA lines, n is the total number of lines, N is the analyzed sites in one line and T is the number of cell divisions the MA line passed.

Context-Dependency of MMR Repair

For each bacterial species, a 3-bp sliding window with a 1-bp step size was used to parse out the contexts in trinucleotides in the entire genome. Mutation rate at a certain context—$\mu_{5'\mathrm{X}\underline{\mathrm{X}}\mathrm{X}3'}$ was calculated by:

$$\mu_{5'\mathrm{X}\underline{\mathrm{X}}\mathrm{X}3'} = \frac{n}{N \times G \times L}$$

where n is the pooled number of mutations in both the 5′X$\underline{\mathrm{X}}$X3′ context and its reverse complement context across all MA lines of the species, N is the total number of both the trinucleotide 5′X$\underline{\mathrm{X}}$X3′ and its reverse complement in the same strand across the entire genome, G is the average number of generations passed in each MA line, and L is the total number of MA lines. We combined nucleotide triplets that are reverse complemented, for example, 5′G$\underline{\mathrm{A}}$T3′ is taken as the same context as 5′A$\underline{\mathrm{T}}$C3′, due to the uncertainty of the DNA strand of each mismatch causing the base-substitution.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Angus Buckling, Vaughn Cooper, Thomas G. Doak, Jacqueline Hernandez, Chloe Strauss, and Way Sung for technical assistance. This research was supported by Fundamental Research Funds for the Central Universities of China (201822020, to H.L.), National Natural Science Foundation of China (31741071, to H.L.), the Multidisciplinary University Research Initiative awards W911NF-09-1-0444 (Patricia Foster, Michael Lynch, Haixu Tang, Steven Finkel) and W911NF-14-1-0411 (Michael Lynch, Patricia Foster, Jay Lennon, Jake McKinlay) from the US Army Research Office, National Institutes of Health awards R01-GM036827 and R35-GM122566 to M.L.

References

Bateman AJ. 1959. The viability of near-normal irradiated chromosomes. *Int J Radiat Biol Relat Stud Phys Chem Med*. 1(2):170–180.

Benzer S. 1961. On the topography of the genetic fine structure. *Proc Natl Acad Sci USA*. 47(3):403–415.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Carraway M, Marinus MG. 1993. Repair of heteroduplex DNA molecules with multibase loops in *Escherichia coli*. *J Bacteriol*. 175(13):3972–3980.

Curti E, McDonald JP, Mead S, Woodgate R. 2009. DNA polymerase switching: effects on spontaneous mutagenesis in *Escherichia coli*. *Mol Microbiol*. 71(2):315–331.

Denver DR, Feinberg S, Estes S, Thomas WK, Lynch M. 2005. Mutation rates, spectra and hotspots in mismatch repair-deficient *Caenorhabditis elegans*. *Genetics* 170:107–113.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43:491–498.

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

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

Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T. 2006. DNA repair and mutagenesis. Washington, D.C: ASM Press.

Jones M, Wagner R, Radman M. 1987. Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* 115(4):605–610.

Kunkel TA, Erie DA. 2005. DNA mismatch repair. *Annu Rev Biochem*. 74:681–710.

Learn BA, Grafstrom RH. 1989. Methyl-directed repair of frameshift heteroduplexes in cell extracts from *Escherichia coli*. *J Bacteriol*. 171(12):6473–6481.

Lee H, Popodi E, Foster PL, Tang H. 2014. Detection of structural variants involving repetitive regions in the reference genome. *J Comput Biol*. 21(3):219–233.

Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci USA*. 109(41):E2774–E2783.

Levinson G, Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*. 4(3):203–221.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25(14):1754–1760.

Long H, Kucukyildirim S, Sung W, Williams E, Lee H, Ackerman M, Doak TG, Tang H, Lynch M. 2015. Background mutational features of the radiation-resistant bacterium *Deinococcus radiodurans*. *Mol Biol Evol*. 32(9):2383–2392.

Long H, Miller SF, Strauss C, Zhao C, Cheng L, Ye Z, Griffin K, Te R, Lee H, Chen CC, et al. 2016. Antibiotic treatment enhances the genome-wide mutation rate of target cells. *Proc Natl Acad Sci USA*. 113(18):E2498–E2505.

Long H, Sung W, Miller SF, Ackerman M, Doak TG, Lynch M. 2015. Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair (MMR) deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol*. 7(1):262–271.

Lujan SA, Clark AB, Kunkel TA. 2015. Differences in genome-wide repeat sequence instability conferred by proofreading and mismatch repair defects. *Nucleic Acids Res*. 43(8):4067–4074.

Lynch M. 2012. Evolutionary layering and the limits to cellular perfection. *Proc Natl Acad Sci USA*. 109(46):18851–18856.

Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet*. 17(11):704–714.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA*. 105(27):9272–9277.

Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits. Sunderland (MA): Sinauer Associates. p. 818.

McDonald MJ, Wang WC, Huang HD, Leu JY. 2011. Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLoS Biol*. 9(6):e1000622.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20(9):1297–1303.

Mennecier S, Coste G, Servant P, Bailone A, Sommer S. 2004. Mismatch repair ensures fidelity of replication and recombination in the radioresistant organism *Deinococcus radiodurans*. *Mol Genet Genomics* 272(4):460–469.

Mukai T. 1964. Genetic structure of natural populations of *Drosophila melanogaster*. 1. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50:1–19.

O'Brien S, Rodrigues AM, Buckling A. 2013. The evolution of bacterial mutation rates under simultaneous selection by interspecific and social parasitism. *Proc Biol Sci*. 280(1773):20131913.

Pal C, Macia MD, Oliver A, Schachar I, Buckling A. 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450(7172):1079–1081.

Parker BO, Marinus MG. 1992. Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc Natl Acad Sci USA*. 89(5):1730–1734.

Petrov DA. 2002. Mutational equilibrium model of genome size evolution. *Theor Popul Biol*. 61(4):531–544.

Sagher D, Hsu A, Strauss B. 1999. Stabilization of the intermediate in frameshift mutation. *Mutat Res*. 423(1–2):73–77.

Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD. 1997. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol*. 17(5):2851–2858.

Sung W, Ackerman M, Dillon MM, Platt TG, Fuqua C, Cooper VS, Lynch M. 2016. Evolution of the insertion-deletion mutation rate across the tree of life. *G3 (Bethesda)* 6(8):2583–2591.

Sung W, Ackerman MS, Gout JF, Miller SF, Williams E, Foster PL, Lynch M. 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. *Mol Biol Evol*. 32(7):1672–1683.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11 10 11-33.

Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*. 34(2):564–574.