Nucleic Acids Research, 2024, **52**, 9574–9585
https://doi.org/10.1093/nar/gkae616
Advance access publication date: 17 July 2024
Genome integrity, repair and replication

Instability throughout the *Saccharomyces cerevisiae* genome resulting from Pms1 endonuclease deficiency

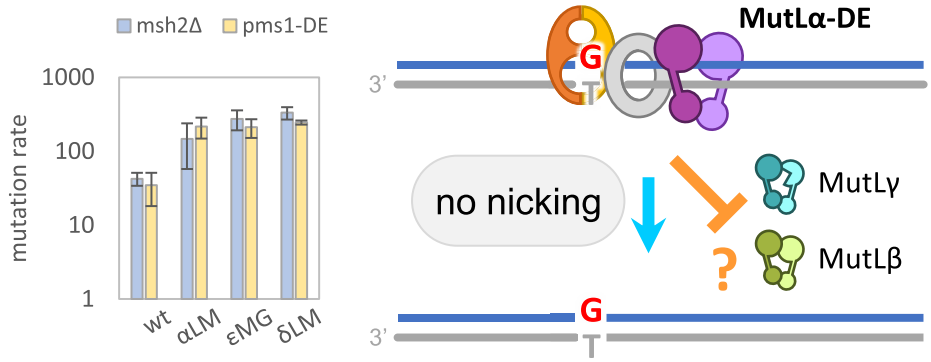
Scott A. Lujan [1], Marta A. Garbacz [1,2], Sascha E. Liberti [3], Adam B. Burkholder [4] and Thomas A. Kunkel [1,*]

[1]Genome Integrity & Structural Biology Laboratory, NIH/NIEHS, DHHS, Research Triangle Park, NC 27709, USA
[2]Currently Marta A. Garbacz works at Exact Sciences Corporation, Torrey Pines Science Park, La Jolla, CA 92037, USA
[3]Aidian Denmark ApS, Copenhagen, Denmark
[4]Office of Environmental Science Cyberinfrastructure, NIH/NIEHS, DHHS. Research Triangle Park, NC 27709, USA
[*]To whom correspondence should be addressed. Tel: +1 984 287 4281; Fax: +1 919 541 7613; Email: kunkel@nih.gov

Abstract

The endonuclease activity of Pms1 directs mismatch repair by generating a nick in the newly replicated DNA strand. Inactivating Pms2, the human homologue of yeast Pms1, increases the chances of colorectal and uterine cancers. Here we use whole genome sequencing to show that loss of this endonuclease activity, via the *pms1-DE* variant, results in strong mutator effects throughout the *Saccharomyces cerevisiae* genome. Mutation rates are strongly increased for mutations resulting from all types of single-base substitutions and for a wide variety of single- and multi-base indel mutations. Rates for these events are further increased in strains combining *pms1-DE* with mutator variants of each of the three major leading and lagging strand replicases. In all cases, mutation rates, spectra, biases, and context preferences are statistically indistinguishable from strains with equivalent polymerases but lacking initial mismatch recognition due to deletion of MSH2. This implies that, across the nuclear genome, strand discrimination via the Pms1 endonuclease is as important for MMR as is initial mismatch recognition by Msh2 heterodimers.

Graphical abstract



Introduction

In most bacteria and eukaryotes, mismatches made during DNA replication are corrected by DNA mismatch repair (MMR). Inactivation of this repair pathway strongly increases mutation rates, which can affect evolution and have several adverse consequences. In eukaryotic nuclei, mismatches are recognized by MutS family heterodimers of Msh2-Msh6 (MutSα) and Msh2-Msh3 (MutSβ). These recognize a wide variety of base-base mismatches and insertion-deletion loops made during nuclear DNA replication by any of the three major replicases, DNA Polymerases (Pols) α, δ and ε. Pol α is responsible for initiating genome replication via Okazaki fragment synthesis at replication origins and during lagging strand replication (1–4). Pol ε is thought to synthesize the bulk of the nascent leading DNA strand (5,6). Pol δ is posited to replicate Okazaki fragments during ongoing lagging strand replication and during initial leading strand synthesis at origins of replication (7–10). Pol δ is also thought to replicate both DNA strands during termination of replication (10), during Break Induced Repair (11), after replications stalling (12), and during post-repair patch resynthesis (13).

The second step of MMR involves MutL family heterodimers, either MutLα, comprised of hMlh1-hPms2 in humans and of Mlh1-Pms1 in *Saccharomyces cerevisiae*, MutLγ (Mlh1-Mlh3) or MutLβ (Mlh1-Mlh2). The MutL family heterodimers interact with the MutS family heterodimers and with proteins that operate downstream to remove DNA patches containing mismatches. MMR is then completed by

Received: February 21, 2024. **Revised:** June 25, 2024. **Editorial Decision:** June 29, 2024. **Accepted:** July 6, 2024
Published by Oxford University Press on behalf of Nucleic Acids Research 2024.

correct resynthesize of DNA and ligation of the resulting nick (14,15).

One key feature of this complex pathway is how MMR recognizes the newly synthesized DNA strand containing the mistaken nucleotide(s) rather than the template strand containing the correct nucleotide(s). In Nobel Prize winning studies, Paul Modrich and his colleagues were the first to implicate an endonuclease activity encoded by the product of the *PMS1* gene in generating nicks in DNA (16). The nicks serve as strand-discrimination signals, directing the MMR machinery to correct mismatches made in the nascent strand (17). These investigators introduced point mutations into yeast *PMS1* and its human homologue *hPMS2*. The point mutations inactivated the nuclease activity, inactivated MMR *in vitro,* and resulted in a strong mutator phenotype in a reporter gene in yeast. Similar mutations in other homologues of *PMS1/2* were subsequently shown to have similar, albeit lesser, consequences (16,18–20). These studies clearly implicate the endonuclease activity of these proteins in MMR of replication errors made during eukaryotic nuclear DNA replication.

The goal of this study is to examine the effects of loss of Pms1 endonuclease activity on MMR across the whole yeast nuclear genome. We introduced the *PMS1* mutant *pms1-D706N/E707Q* (abbreviated hereafter as *pms1-DE*), first studied by the Modrich group (17), into diploid *S. cerevisiae*. The mutation disrupts a conserved metal-binding site that is essential for Pms1 endonuclease activity. This endonuclease-deficient variant was introduced either alone or in combination with variants of Pols α, δ and ε. We used the *muver* whole genome mutation pipeline (21) to call mutations accumulated over hundreds of cellular generations in these strains. The rates, spectra, and contexts of these mutations demonstrate strong mutator phenotypes across the genome that allow several interesting interpretations (see below).

Materials and methods

Strain construction

Four diploid strains of *Saccharomyces cerevisiae* were constructed (see Methods). Each strain contains a mutation previously shown to inactivate the endonuclease activity of Pms1 (*pms1-D706N/E707Q*, abbreviated herein as *pms1-DE*), thereby generating a strong mutator phenotype in the *URA3* reporter gene (16). One strain contains only this mutation (Table 1). The other three strains each also contain a second mutation in a gene encoding the catalytic subunit of either DNA Polymerase (Pol) α, δ or ε (*pol1-L868M*, *pol3-L612M*, or *pol2-M644G*, or *pol3-L612M*, respectively). These active site mutations reduce the fidelity of DNA replication (5,22,23).

Mutation accumulation and genome sequencing

Whole genome mutation accumulation experiments were conducted as previously described (24), except that the *pms1-DE* strains were sequenced on the NovaSeq 6000 platform (Illumina Inc.; 2 × 150 bp paired end reads). Briefly, multiple independent diploid isolates were created for each strain. These isolates were passaged on solid media without phenotypic selection. At the beginning of the experiment (time zero samples), and again after either a set number of generations or for the previous time point if a growth defect was detected (outgrowth samples; approximately 30 per passage; see Table 1). Their DNA was extracted, sequencing libraries were prepared, and their genomes were sequenced.

Mutation calling

The *muver* pipeline (version 1.2.4) was then used to compare to the sequences of the initial samples to the sequences from the final samples (21). Significant differences between the initial and final sequences indicate mutations that arose in the endonuclease deficient *pms1-DE* strains during the intervening generations of growth. Mutation rates were calculated as previously reported (10) across the genome for each of the four *pms1-DE* strains, accounting for mutation type, chromosomal location, proximity to genomic features, and local sequence context (Supplemental Figures S2–S23; Supplemental File 1 for generations, queryable base counts, and annotated variant lists by sample and strain).

Calculations

Mutation calls and resulting rates may differ slightly for MMR-proficient and -deficient strains from previous reports (24), due to the inclusion of additional samples and the application of updated filtering protocols to all samples of all strains (muver version 1.2.4) (25). Previously reported strains are either wild type for MMR or lacked MMR due to deletion of either the *MSH2* gene or both the *MSH3* and *MSH6* genes (*pol2-M644G* only). The *pol2-M644G msh3*Δ *msh6*Δ strain was previously shown to be functionally indistinguishable from the *pol2-M644G msh2*Δ strain for the purposes of mutation rates and spectra (24). As was done previously, their data are herein combined and listed as *pol2-M644G* MMR— in figures and tables and occasionally *pol2-M644G msh**Δ in supplemental spreadsheets.

Unless otherwise noted, wherever mutation rates are compared between strains, significance was tested using the heteroscedastic Welch's *t*-test (26). Degrees of freedom were approximated via the Welch–Satterthwaite equation (27). Significance thresholds were set by applying the Šidák correction for multiple hypothesis testing (28) for a family-wise error rate of 0.05 (see Supplemental File 1 for calculations). The family in each case was the total number of pairwise comparisons between mutation subtypes (27 per pair of strains).

Results and discussion

Total genomic mutation rates in the single- and double-mutant *pms1-DE* strains.

Mutations were compared between the four *pms1-DE* strains and eight previously reported strains: four MMR-proficient and four MMR-deficient strains (Table 1, Figure 1A, Supplemental Tables S1–S4). The 112 outgrowth isolates of the twelve strains collectively accumulated 183 391 bp substitution and indel mutations. Substitutions were the majority in ten of the mutation spectra, while insertions and deletions (indels) dominated in the *pms1-DE* and *msh2*Δ strains (Figure 1, Supplemental Tables S1–S4). Mutations were widely distributed across the sixteen nuclear chromosomes (Figure 2 for *pms1-DE* strains; otherwise, Supplemental Figure S1). Only six bins of 10 kb contained significantly more mutations than expected from random binomial distribution, after correcting for multiple hypothesis testing, all in the *pms1-DE* and *msh2*Δ strains. These potential mutation clusters consist of indels in rare, long homopolymers tracts and are fully explained by an

Table 1. Mutation rates, underlying counts, and rate comparisons. All *P*-values come from heteroscedastic Welch's *t*-tests (26) with degrees of freedom approximated via the Welch–Satterthwaite equation (27). Citation key: *a* = (25); *b* = this study; *c* = (21)

	Outgrowths	Ploidy (n)	Generations (total; 10^3)	Mean queryable (Mbp/ploidy)	Variants	Mutations	Mutation rate (Gbp^{-1} $gen.^{-1}$)	Standard error	Versus *wt* rate	*p*-value vs. *wt*	*p-value vs. pms1-DE*	Ref.
wild type	39	2	54	11.1	271	255	0.21	0.077	1			a
pms1-DE	8	2	6.4	10.9	5 277	4 549	34.7	16.6	170	0.038		b
msh2Δ	5	2	4.3	10.6	4 780	4 010	42.5	8.49	200	0.0038	0.34	c
pol1-L868M	6	2	5.4	11.1	122	111	0.928	0.221	4.4	0.014		c
pol1-L868M pms1-DE	8	2	7.4	11.1	39 453	35 770	217	68.5	1 000	0.0079		b
pol1-L868M msh2Δ	8	2	6.1	11.1	22 314	20 572	148	90.4	700	0.073	0.72	c
pol2-M644G	8	2	7.2	11.2	492	430	2.67	0.612	13	0.0026		c
pol2-M644G pms1-DE	8	2	6.5	11.0	36 573	30 100	212	60.4	1 000	0.005		b
pol2-M644G MMR-	5	2	3.8	11.1	25 305	22 221	275	83.4	1 300	0.0150	0.28	c
pol3-L612M	7	2	5.9	11.0	348	332	2.53	0.233	12	0.000 040		c
pol3-L612M pms1-DE	6	3	5.6	11.1	49 898	45 656	246	15.8	1 200	0.000 010		b
pol3-L612M msh2Δ	4	2	2.7	11.0	21 253	19 385	334	63.5	1 600	0.0067	0.14	c

exponential increase in rates with tract length, as shown previously in our MMR-deficient strains (21,29). Otherwise, at the 10 kb scale, the mutation distributions are indistinguishable from uniformity.

From these mutations, total genome-wide mutation rates were calculated for all point mutation types. In the *pms1-DE* strain, this rate is 34.7 ± 17 mutations per Gbp per generation (Table 1, Figure 1A). This is 170-fold higher than the rate in the wild-type yeast strain and very similar to that observed in for an *msh2Δ* yeast strain (42.5 ± 8.5 Gbp^{-1} $generation^{-1}$ Table 1, Figure 1A). This confirms the previous interpretation that was based on studies with reporter genes, i.e. that the endonuclease activity of Pms1 is important for MMR. In fact, the whole genome rate indicates that the endonuclease activity of Pms1 appears to be as important as the presence of Msh2 for MMR. Furthermore, the rates for the double-mutant strains encoding the three replicase variants are each elevated by several fold in comparison to the single *pms1* strain (Figure 1A). This fact indicates that in these strains, Pms1 is involved in MMR of errors made by all three variant replicases.

Rates for base substitution mutations

The mutation rates for base substitutions in all four *pms1* strains are strongly increased relative to the rates in the four equivalent MMR-proficient strains. This is so for total base substitutions (Figure 1B), and to different extents, it is also true for each of the six different single base substitutions (Supplemental Tables S1–S4, Supplemental Figure S1). These rates are similar to those observed in *msh2*Δ strains (Supplemental Tables S1–S4, Supplemental Figure S1), supporting the interpretation that the endonuclease activity of Pms1 is critical for repair of base-base mismatches made during replication of the yeast nuclear DNA genome. The data are consistent with the hypothesis that these mutator effects reflect loss of strand discrimination during MMR.

Multiple mismatch types may lead to the same mutagenic outcome, so a mutagenic process may alter mismatch preferences without changing mutation rates. To determine whether the *pms1-DE* introduces different mismatch preferences than complete removal of MMR, we compared complementary mutation fractions at varying distances from replication origins in strains with different mismatch preferences on different replication strands. Previous studies revealed that several types of base substitutions preferentially result from specific mismatches that can potentially distinguish whether the mistake was made during leading and lagging strand replication (5,22,23). For example, Pol δ-L612M misincorporates guanine opposite template thymine (T-dG) at a much higher rate than it makes the complementary error, cytosine opposite template adenine (A-dC, (23)). Thus, the majority of T-A to C-G transition substitutions in the *pol3-L612M msh2*Δ strain are likely to have been made through a T-dG intermediate. This bias allows the roles of each replicase variant to be inferred when replicating the two complementary DNA strands in various regions of the nuclear genome. For example, in the *pol3-L612M msh2*Δ strain, T-to-C substitutions vastly outnumber A-to-G substitutions collected from immediately to the right of each replication origin, whereas the inverse is true to the left of each origin (24). This pattern implies that DNA polymerase δ primarily replicates the lagging DNA strand. Here we show that a *pol3-L612M pms1-DE* strain has the same specificity (Figure 3). Human and yeast Pols α and ε are also biased towards T-dG (30–33), as are their respective mutator variants, Pol1-L868M and Pol2-M644G ((5) and Supplemental Tables S5–S6). Both *pol1-L868M* MMR- and *pol1-L868M pms1-DE* yeast strains have the same inter-origin T-dG bias as *pol3-L612M msh2*Δ, whereas *pol2-M644G* MMR- and *pol2-M644G pms1-DE* yeast strains have the opposite bias (Figure 3 and (24)). This implicates Pols α and ε primarily on the lagging and leading strands, respectively. The inter-origin biases are equivalent for the MMR- and *pms1-DE* backgrounds and are consistent with these strand assignments for all mismatch types given the *in vitro* mutation biases preferences of all three mutator polymerases (see Discussion and Supplemental Figure S3–S7). Identical biases, regardless mutator polymerase strand assignment, imply that loss of Pms1 endonuclease activity is as important as is loss of Msh2 for MMR of mismatches generated during replication of both DNA strands.

In a similar manner, we examined the effect of loss of MMR resulting from the loss of Pms1 endonuclease activity on other aspects of base substitution specificity. These aspects include across replication time (Supplemental Figure S8a), mutator effects in relation to nucleosome positions (Supplemental Figure S8b) and in protein coding genes and their flanking

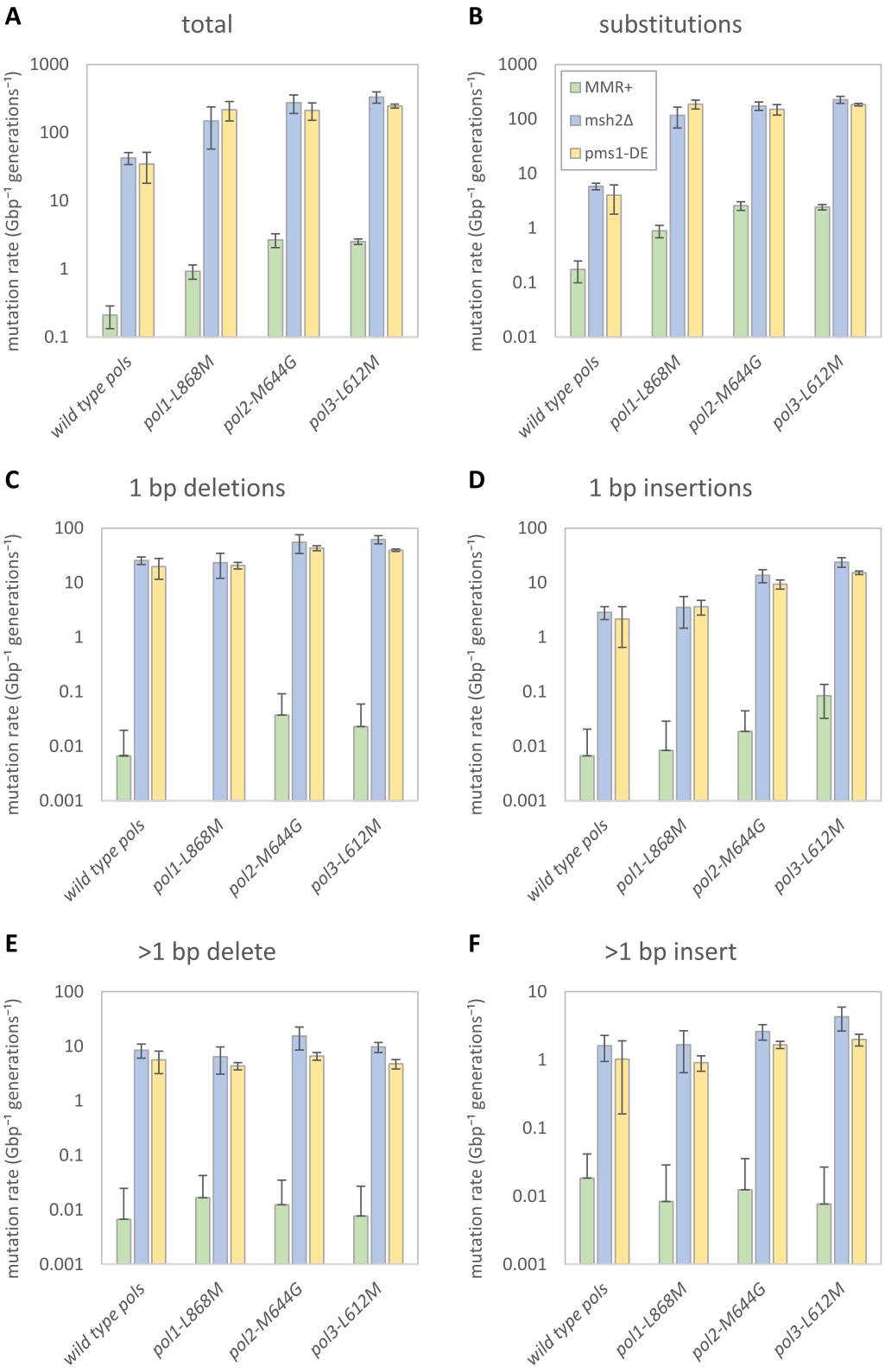


Figure 1. MMR- and *pms1-DE* rates are higher than MMR + but indistinguishable from one another. For each of four polymerase backgrounds, mutation rates in the MMR- (blue) and *pms1-DE* (yellow) strains are statistically indistinguishable from one another and higher than in the relevant MMR + strain (green). This is true for (**A**) overall mutation rates, (**B**) substitutions, (**C**) single-base deletions, (**D**) single base insertions, (**E**) multi-base deletions, and (**F**) multi-base insertions. Error bars = standard error.

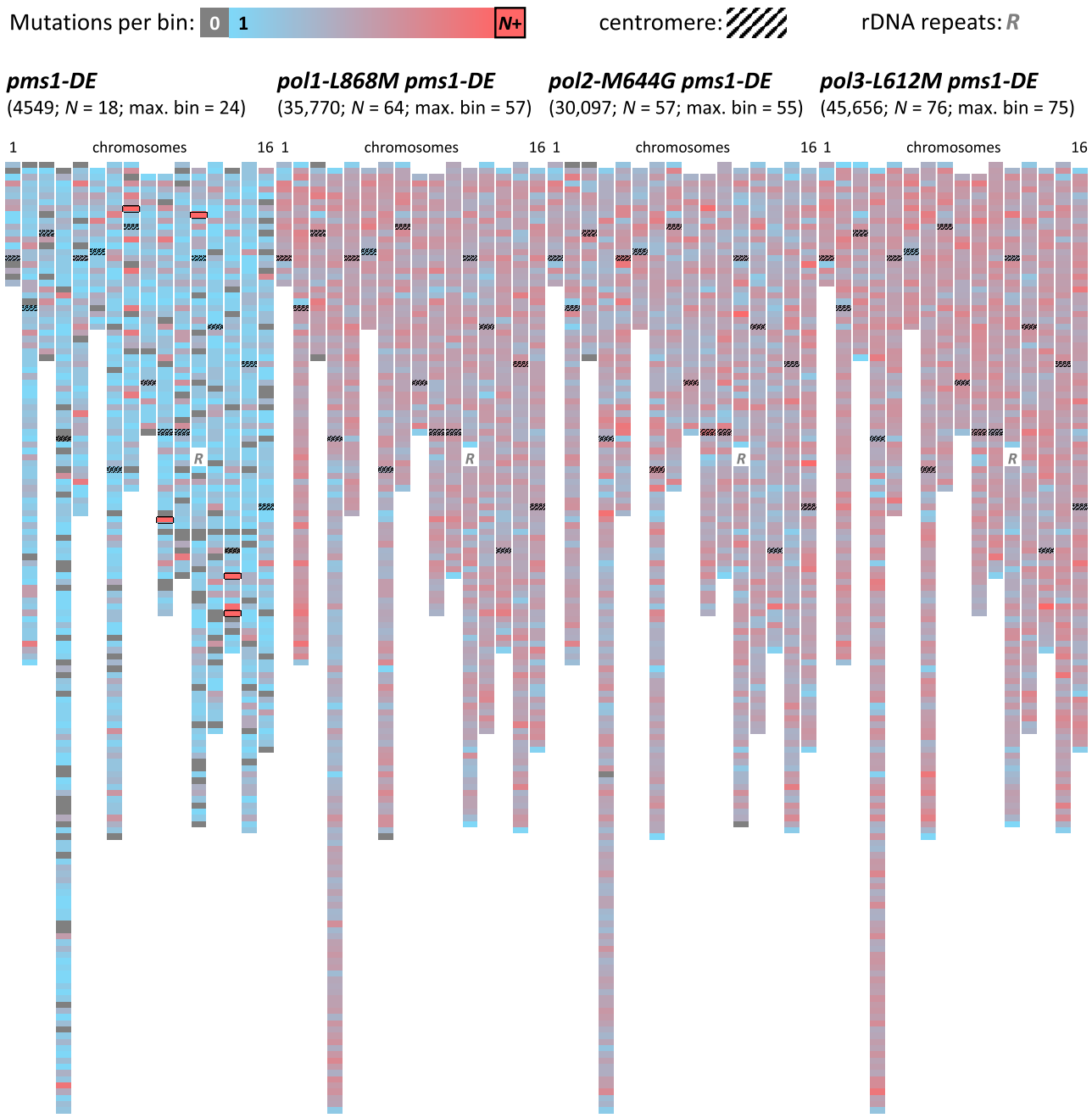


Figure 2. Mutation heatmaps at the 10 kbp scale show an even distribution of mutations across all chromosomes in *pms1-DE* strains. For the bulk of each genome, mutation frequencies were essentially uniform on this scale, as was found in the absence of MMR ((24); Figure S2). The 16 yeast chromosomes are shown left-to-right in ascending order from chromosome 1 to 16 for each *pms1-DE* strain. Below each strain name are the number of mutations mapped, the number of outgrowth isolates sequenced (*N*), and the maximum mutation count in any bin (max. bin). Bins represent 10 kb each. Bins with centromeres are (hashed black) or with no unique sequence (white) are indicated, including the rDNA loci (white with *R*). Bins with no mutations are grey (see scale). Bins with one mutation are blue and bins with more mutations become progressively redder until the threshold for a significance cluster is reached (*N* + on scale; see Materials and methods for calculations details). Bins that exceed this threshold, i.e. potential mutation hotspots, are red outlined black. These were only observed in the *pms1-DE* strain (and the *msh2*Δ strain; Figure S2). These are the only strains dominated by insertions and deletions in long homopolymers. All observed clusters corresponded with unusually high concentrations of very long homopolymer tracts.

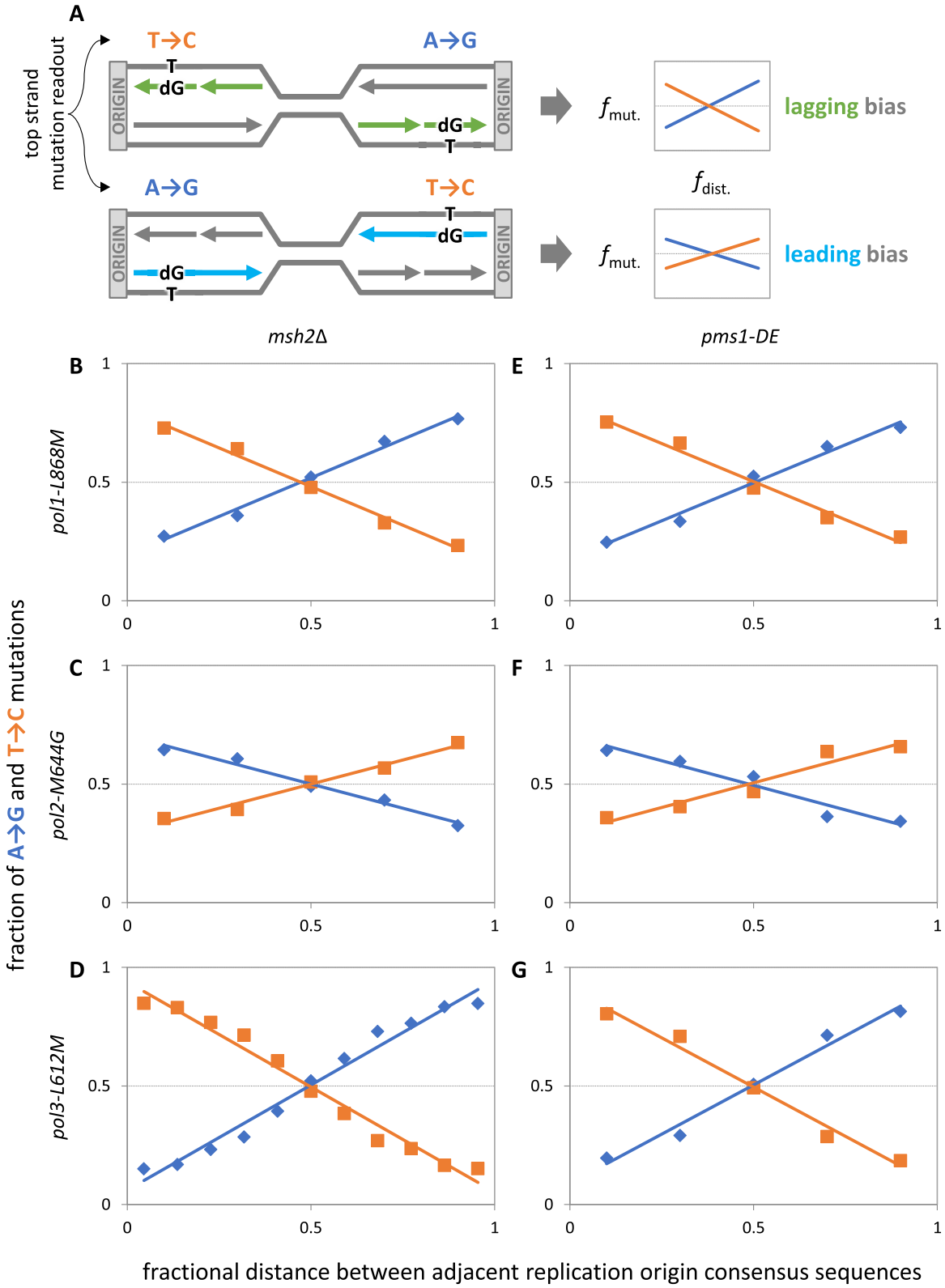


Figure 3. Complementary substitution biases between adjacent origins indicate identical polymerase strand specificity in *pms1-DE* and *msh2*Δ strains. AT-to-GC mutation fractions are plotted between adjacent origin positions (summed across all origin pairs). (**A**) If a mutagenic process works primarily on one strand and is biased for T-dG mispairs, then top strand A→G versus T→C mutation fractions will form a distinctive X-pattern. The three polymerase mutator variants prefer T-dG mispairs over complementary A-dC mispairs *in vitro* ([5,23] and Supplemental Tables S5–S6). (B–D) Absent MMR (*msh2*Δ or MMR–), AT-to-GC biases indicate (**B**) Polα-L868M and (**D**) Polδ-L612M activity on the lagging strand and (**C**) Polε-M644G activity on the leading strand, as previously reported ([5,23,24]). (**E–G**) AT-to-GC biases in the MutLα endonuclease-dead background (*pms1-DE*) are identical in both direction and magnitude to those in the absence of MMR. This is true for all substitution types (Supplemental Figure S3–S7).

untranslated sequences (Supplemental Figure S9–S16). The magnitude and patterns of mutation rates in the *pms1-DE* strains are similar to those seen in Msh2-deficient strains for each mutation class and each genomic feature. All these observations are consistent with the interpretation that loss of the endonuclease activity of Pms1, just like loss of Msh2, results in a strong mutator phenotype for base substitutions across the yeast nuclear genome.

Rates for insertion-deletion mutations

Loss of Pms1 endonuclease activity also yields strong mutator phenotypes for a variety of insertion-deletion (indel) mutations. These arise from mismatches containing one or more unpaired nucleotides in either the nascent DNA strand (insertions) or the template DNA strand (deletions). The rates for all indel types were again statistically indistinguishable from their equivalents in the *msh2*Δ strain in all four polymerase backgrounds (Figure 1C, D, Supplemental Figure S1, Supplemental Tables S1–S4). The indel rate patterns were essentially the same between *pms1-DE* and *msh2*Δ strains for replication time, nucleosome site proximity, and distance relative to protein coding sequences (Supplemental Figure S8–S16). This again implies that loss of the endonuclease activity of Pms1 inactivates repair of indel mismatches to roughly the same degree as does loss of Msh2.

Indel rates in endonuclease-deficient *pms1* strains are higher across the genome (Figure 1c-f) than in wild type *S. cerevisiae* (24,34–43), as was the case for indel rates in *msh2Δ* strains (21,24,29,36,38). These higher rates include both gain and loss of both A•T and G•C base pairs, as well as for indels of various lengths (Supplemental Figure S1, Supplemental Tables S1–S4). Indel rates increased as a function of the number of base pairs within repeat tracts, regardless of indel type, size, or tract composition: single-base deletions in A/T runs (Figure 4); single-base insertions in A/T runs (Supplemental Figure S17); single-base indels in G/C runs (Supplemental Figure S18); multi-base indels in A/T runs (Supplemental Figure S19); multi-base indels in AT/TA dinucleotide tracts (Supplemental Figure S20); and other indel sizes and tract compositions (Supplemental File 1). As originally proposed by Streisinger (44) and reported in *msh2*Δ strains (21,24,29,36), these data are consistent with DNA strand slippage as the primary mechanism underlying these mutations. Only minor deviations are observable between *pms1-DE* and MMR- curves and those are primarily where indel counts are very low, i.e. in short tracts where rates are low and in very long tracts where targets are rare.

Indel biases between adjacent replication origins suggest that replicases have preferences for inserting or deleting complementary base pairs and suggest which wild type replicases make more indel errors *in vivo*. The degree of bias may be estimated (Table 2; see Supplemental File 1 for calculations) from a linear regression of complementary fractions. Where sufficient mutation counts allow, higher resolution plots reveal that biased curves are somewhat sigmoid (e.g. Figure 3D), meaning that a linear regression will somewhat overestimate the bias. If a polymerase spends at least some of its time on each strand, like Pol δ (9,10), then this measure will underestimate the bias and the effects roughly cancel. Absent strong evidence for Pols α or ε switching strands, we *could* conclude, for example, that the *pol1-L868M* strain has at least a slight preference for looping out primer strand thymidine (+T) relative to primer strand deoxyadenosine (+A; Figure 5A versus 5B and 5E versus 5F; Table 2). However, *pol1-L868M* does not significantly increase + AT rates relative to a wild type polymerase background (Supplemental Figure S1a). Likewise, the +AT bias in the double-mutant *pol1-L868M msh2*Δ strain has the same magnitude and direction as the single-mutant *msh2*Δ strain. This suggests that indel biases exist with wild type replicases and begs the question of which replicases are responsible for which mutations. The +AT bias is unchanged, relative to wild type polymerases, in the *pol2-M644G* background (Figure 5c and g), even though *pol2-M644G* increases the + AT rate (Supplemental Figure S1b). In contrast, the +AT bias is reduced in the *pol3-L612M* strains (Figure 5D and H), which would require Pol δ-L612M to have the same + AT preference while working on the opposite strand. This suggests that Pols δ and ε prefer to loop out primer deoxyadenosine rather than primer thymidine. In general, the mutator polymerases increase rates and alter bias magnitude (lac Z mutation assay, (5,23) and Supplemental Table S5) but not bias direction (Supplemental Table S6, correcting for deamination-driven background U-dA pairs (45,46)). Thus, the observed whole genome biases also suggest that the majority of + AT mutations during wild type replication are due to Pol ε errors on the leading strand. Similar logic for -AT deletions implies that Pols δ and ε prefer to loop out template deoxyadenosine (–A) and that most wild type –AT mutations are due to Pol ε (Table 2 and Supplemental Figure S21).

Too few GC indels were found in the single-mutant strains to make a full determination (Supplemental Table S1) and -GC biases follow no obvious logic (Table 2 and Supplemental Figure S22). +GC biases run counter to the +AT observations, with identical bias direction and similar bias magnitude in the *pol2-M644G* and *pol3-L612M* backgrounds (Table 2 and Supplemental Figure S22). The simplest explanation is that Pols δ and ε have opposite + GC biases, with Pol δ preferring to loop out primer G and Pol ε preferring to loop out primer C. More data are needed to distinguish this explanation from more complicated hypotheses and to establish which replicase is responsible for more + GC mutations in the wild type background.

If Pol ε has AT indel biases and causes the majority AT indels in other organisms, then the implications for evolution and disease are obvious and potentially profound. After all, any frameshift indel in a protein coding gene alters all subsequent translation, and > 50% of in-frame β-lactamase indels decrease ampicillin resistance >100-fold (47) but only 16–18% of substitutions in saturated reporter genes or regulatory elements significantly altered phenotypes (*URA3* and *CAN1* in in our collections combined with (48); 20 disease-associated promotors and enhancers in (49)). The dangers of indels are illustrated by indel-derived sequence variants associated with at least 22% of severe human diseases, including many cancers (50), and stronger selection against them than against substitution-derived variants in coding regions (51). More work is required to clarify the mechanisms behind observed mutation biases and to expand these observations to other organisms and diseases.

Where indel counts are sufficient to make a determination, each *pms1-DE* strain has the same biases as the corresponding MMR- strains. When mutation fractions near origins (Table 2) are plotted, *pms1-DE* strains versus equivalent MMR- strains, both the slope and Pearson's correlation coefficient are near unity (0.986 and 0.991, respectively;

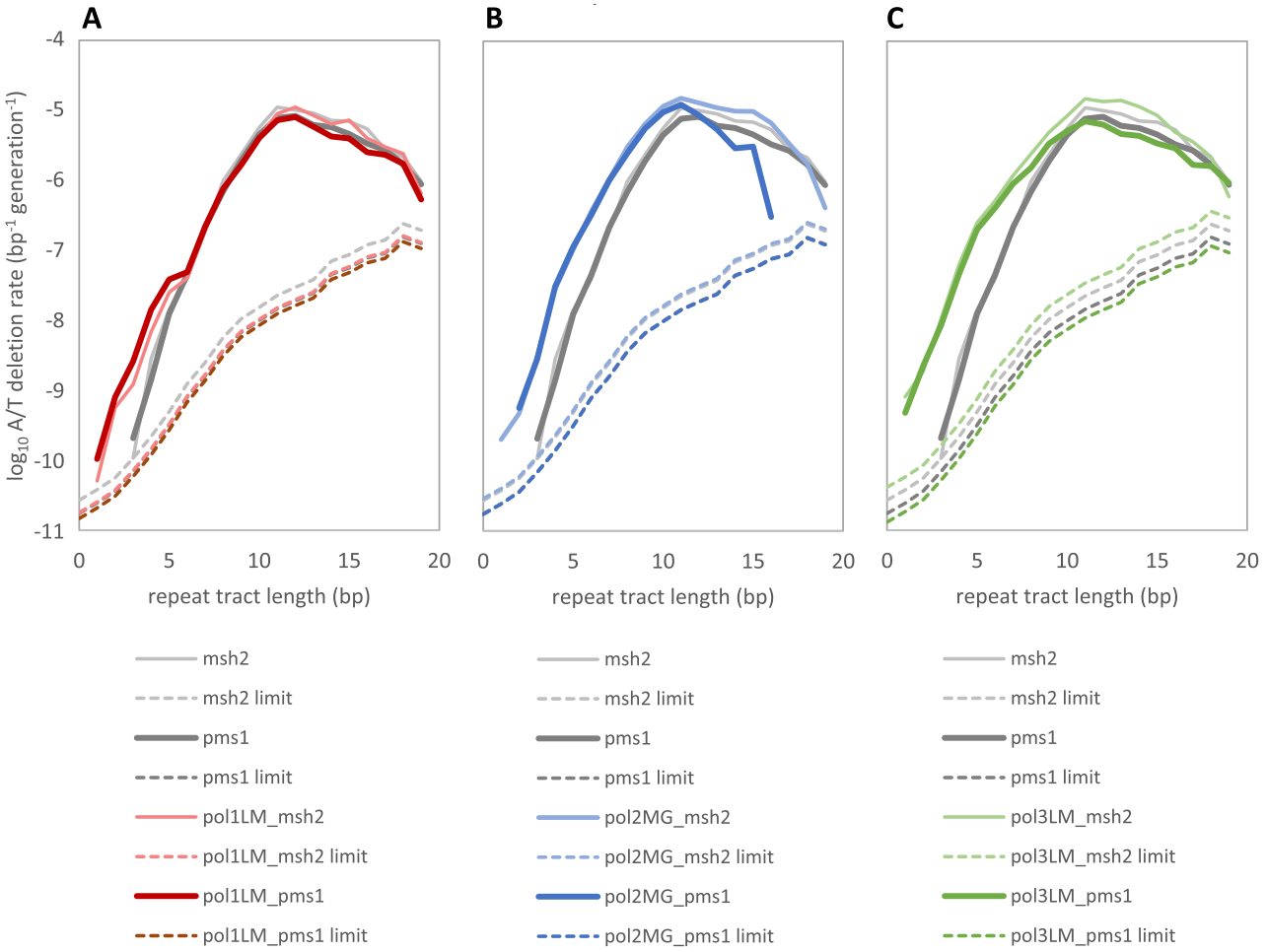


Figure 4. Deletion rates in A and T homopolymers vary with tract length and polymerase background, but not between *pms1-DE* and *msh2*Δ strains. Initially, deletion rates increase exponentially with increasing homopolymer tract length, but the rate of increase diminishes until a peak is reached at 11–12 bp tracts. Rates are shown as solid lines and minimum detection limits as dashed lines. Colors indicate strain, with MMR- strains in light shades and *pms1-DE* strains in dark shades. Wild type polymerase backgrounds are shown in shades of grey, (**A**) *pol1-L868M* backgrounds in shades of red, (**B**) *pol2-M644G* backgrounds in shades blue and (**C**) *pl3-L612M* backgrounds in green. Rates in mutator variants exceed rates in the wild type polymerase background in shorter homopolymers. With the possible exception of the longest tracts in the *pol2-M644G* background, rates in the MMR- and *pms1-DE* strains are indistinguishable.

Table 2. Mutation biases near replication origins

mutation class	AT→GC		GC→AT		AT→TA		AT→CG		GC→TA		GC→CG		-AT		-GC		+AT		+GC	
	Bias	to right	Bias	to right	bias	to right	bias	to right	bias	to right	bias	to right	bias	to right	bias	to right	bias	to right	bias	to right
pms1-DE	1.98	T→C	4.8	G→A					1.85	C→A			1.25	-A			2.16	+A		
msh2Δ	2.63	T→C	3.22	G→A					1.85	C→A			1.45	-A			2.04	+A		
pol1-L868M pms1-DE	4.58	T→C	2.64	G→A	1.1	T→A	10.9	T→G	14.6	C→A	13.1	G→C	1.31	-A	2.2	-C	2.04	+A		
pol1-L868M msh2Δ	4.74	T→C	2.58	G→A	1.03	T→A	6.45	T→G	11.3	C→A	39.4	G→C	1.45	-A	1.3	-C	2.02	+A		
pol2-M644G pms1-DE	2.42	A→G	1.17	C→T	37.8	A→T	6.45	A→C	7.53	G→T	5.44	C→G	1.79	-A	1.07	-G	2.01	+A	2.86	+C
pol2-M644G MMR-	2.37	A→G	1.18	C→T	div. 0	A→T	4.69	A→C	3.55	G→T	2.86	C→G	1.82	-A	1.7	-C	1.88	+A	3.14	+C
pol3-L612M pms1-DE	10.8	T→C	9.09	G→A	3.33	A→T	17.1	T→G	5.96	C→A			1.62	-T	1.97	-C	1.26	+A	8.98	+C
pol3-L612M msh2Δ	16.5	T→C	9.49	G→A	3.71	A→T	20.5	T→G	7.45	C→A	14	G→C	1.41	-T	1.11	-G	1.21	+A	7.68	+C

Mutation biases indicate a fold-preference for on mutation type versus the complementary mutation type (e.g. T→C versus A→G). See Methods and Supplemental File 1 for calculations. The preferred mutation to the right of origins is indicated. Blank cells indicate insufficient data. Div. 0 = divide by zero error, essentially perfect preference within the error of measurement.

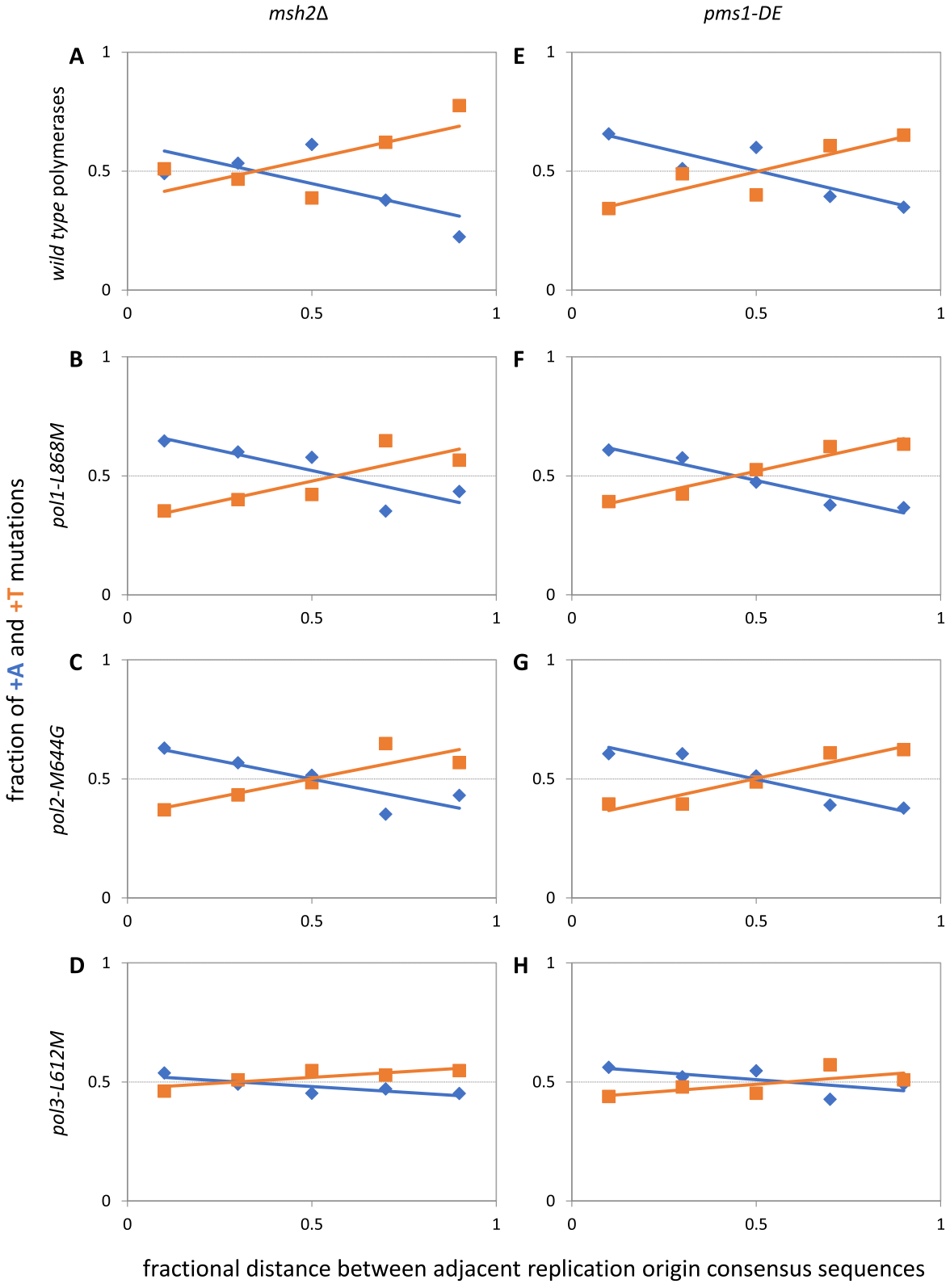


Figure 5. Complementary insertion biases between adjacent origins indicate identical polymerase strand specificity in *pms1-DE* and *msh2*Δ strains. +AT mutation fractions are plotted between adjacent origin positions (summed across all origin pairs). If a mutagenic process works primarily on one strand and is biased for either +A or the complementary +T primer loops, then +AT mutation fractions should make a distinctive X-pattern between origins. (A–D) Absent MMR (*msh2*Δ or MMR–), similar +AT biases indicate (**A**) the underlying wild type pattern (**B**) is not altered by Polα-L868M, which does not increase +AT rates (Supplemental Figure S1). (**C**) Polε-M644G also does not alter the patter, even though it does increase +AT rates. (**D**) Polδ-L612M raises +AT rates but reduces the bias. (**E–H**) +AT biases in the MutLα endonuclease-dead background (*pms1-DE*) are identical in both direction and magnitude to those in the absence of MMR. This is true for –AT and +GC indels, but not –GC indels (Supplemental Figure S21–S23).

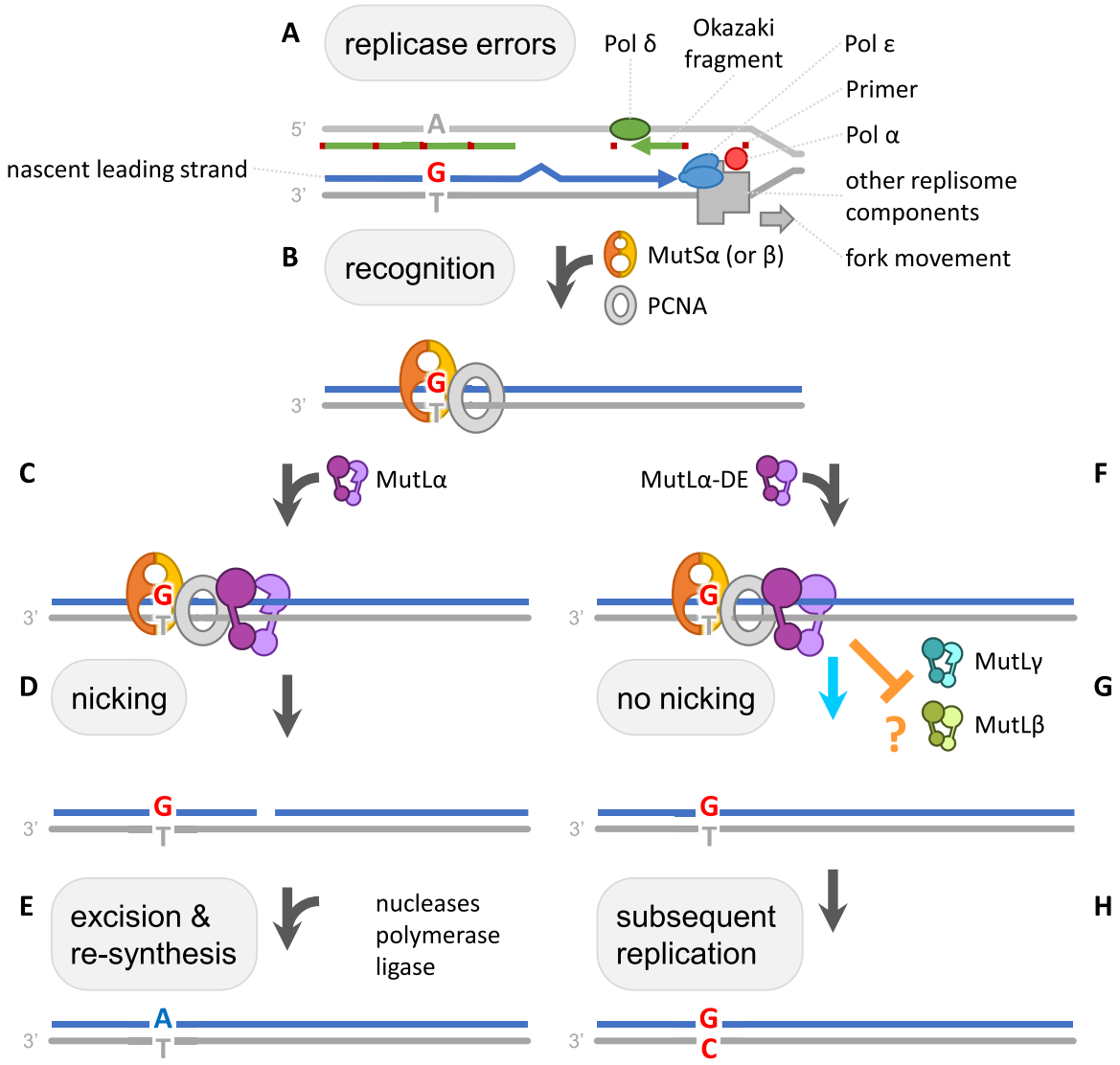


Figure 6. Hypotheses to explain the equivalent effects of complete MMR removal and loss of MutLα endonuclease activity. Enzymes are represented as colored shapes. Template strands are colored grey, nascent strands are colored to match the polymerase that synthesized them. (**A**) A polymerase makes an error. In this example the leading strand T-dG mispair made by Pol ε is a stand-in for all point mutation types made by all replicases on both nascent strands. (**B**) A heterodimer of MutS homologues, in conjunction with PCNA, recognize and bind to the mispair. (C-E) Standard MMR. (**C**) The MutS homologues help recruit MutL homologues, here represented by MutLα. Here we take no position on the stoichiometry of MutL homologues at on mispair. (**D**) The exonucleolytic subunit nicks near the mispair. Here we take no position on the number of nicks nor on the distance from the mispair. (**E**) The nick initiates patch removal, which is followed by resynthesis and strand ligation. (F-H) Aberrant MMR. (**F**) The MutS homologues help recruit MutLα-DE, which lacks endonuclease activity. (**G**) MMR will fail for all errors that require MutLα nicks (blue arrow). The apparent equivalence between MMR- and *pms1-DE* strains could mean that nearly all MMR can go through MutLα, leaving few roles for MutL heterodimers β and γ. Alternatively, if a subset of errors do indeed require MutLβ or γ activity, then either MutLα-DE must inhibit their involvement (orange T), or their contributions are within the error range of the rate measurements presented here. (**H**) Regardless, the next round of replication converts the mispair into a mutation.

Supplemental Figure S24). These indicate no systematic bias and near perfect correlation between strains, respectively. This again implies removing Pms1 endonuclease activity ablates MMR to the same degree as does loss of Msh2, regardless of indel type (insertion or deletion) or repeat tract composition (A, T, G or C).

Summary

The whole genome data presented here reinforce reporter gene studies demonstrating that the endonuclease activity in the *PMS1* gene in budding yeast is critical for MMR. The similarity in the rates and specificity of mutagenesis due to loss of Msh2 and Pms1 endonuclease implies that these two proteins act similarly to reduce mismatch rates. These include the rates of all six types of base substitutions and most, if not all, indel mutations. That said, because loss of the endonuclease activity of Pms1 (*pms1-DE*) results a much stronger mutator effect on indels in reporter genes (12 000× wild type in the *lys2::insE-A14* reporter (16)) than does loss of either Msh6 or Msh3 alone (190× and 6×, respectively (52)), we do not yet exclude the possibility that repair of certain indels may require other MutL proteins that also have endonuclease activity, including Mlh3 (53). This possibility could be examined by whole genome sequence analysis of strains harboring deletions of Msh3 and Msh6, to determine the effects of loss of MutSα and MutSβ on the specificity of indel mutagenesis across the genome.

These results are surprising because MutLβ and MutLγ have been shown to affect mutagenesis even during vegetative (non-meiotic) growth (54), so there must be some interplay between MutLα and its paralogs. For instance, MutLγ is important for deletion repair in the presence of *PMS1* mutants (*pms1-G882E* and *pms1-H888R*) which repair deletion but not insertion mispairs (55). It is also theoretically possible that Pms1-DE could have a dominant negative effect, perhaps by binding to certain indel mismatches and physically blocking MutLβ/γ access (Figure 6). Such MMR inhibition was suggested by Smith and collaborators based on strong evidence in wild type yeast transformed with vectors expressing other *PMS1* mutations (elevated reporter mutation rates despite endogenous wild type *PMS1* and persistent nuclear MMR foci (19)). An analogous situation exists for missense mutants of *MLH2*, constructed based on cancer-derived variants of the human homologue (*hPMS1*), which cause dominant mutator phenotypes and persistent nuclear MMR foci in yeast (55). If real, a dominant mutator *pms1-DE* should also cause persistent foci and mutation rates should be higher in a *pms1-DE* strain than in a *pms1*Δ strain. This latter appeared to be the case in *pms1-DE* than *pms1*Δ in *lys2::insE-A14*, *his7-2* and *CAN1* reporter genes, but the difference was only significant in *lys2::insE-A14* (16). To date, it is unknown how the DE mutation affects Pms1 expression, Pms1/Mlh1 heterodimer formation, or the formation of foci. Regardless, only the complete absence of Pms1 would falsify the MMR inhibition hypothesis. Larger *pms1*Δ mutation spectra in more diverse sequence contexts are needed to determine whether MutLα-DE prevents repair by MutLβ/γ.

Data availability

Sequencing data may be found at the Sequence Read Archive under BioProject PRJNA245050.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Youri Pavlov from the University of Nebraska Medical Center for preserving M13 *lacZ* data. We thank all Kunkel group members for their thoughtful comments on the manuscript. We thank P. Mieczkowski and others from the High Throughput Sequencing Facility of UNC Chapel Hill for performing Illumina sequencing. Sequencing data may be found at the Sequence Read Archive under BioProject PRJNA245050.

Funding

Project Z01 ES065070 to T.A.K from the Division of Intramural Research of the National Institute of Environmental Health Sciences, National Institutes of Health. Funding for open access charge: National Institute of Environmental Health Sciences.

Conflict of interest statement

None declared.

References

1. Bell,S.P. and Labib,K. (2016) Chromosome duplication in Saccharomyces cerevisiae. *Genetics*, **203**, 1027–1067.
2. Lujan,S.A., Williams,J.S. and Kunkel,T.A. (2016) DNA polymerases divide the labor of genome replication. *Trends Cell Biol.*, **26**, 640–654.
3. Stodola,J.L. and Burgers,P.M. (2017) Mechanism of lagging-strand DNA replication in eukaryotes. *Adv. Exp. Med. Biol.*, **1042**, 117–133.
4. Burgers,P.M.J. and Kunkel,T.A. (2017) Eukaryotic DNA replication fork. *Annu. Rev. Biochem.*, **86**, 417–438.
5. Pursell,Z.F., Isoz,I., Lundstrom,E.B., Johansson,E. and Kunkel,T.A. (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, **317**, 127–130.
6. Clausen,A.R., Lujan,S.A., Burkholder,A.B., Orebaugh,C.D., Williams,J.S., Clausen,M.F., Malc,E.P., Mieczkowski,P.A., Fargo,D.C., Smith,D.J., *et al.* (2015) Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat. Struct. Mol. Biol.*, **22**, 185–191.
7. Nick McElhinny,S.A., Gordenin,D.A., Stith,C.M., Burgers,P.M. and Kunkel,T.A. (2008) Division of labor at the eukaryotic replication fork. *Mol. Cell*, **30**, 137–144.
8. Yeeles,J.T., Janska,A., Early,A. and Diffley,J.F. (2017) How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol. Cell*, **65**, 105–116.
9. Garbacz,M.A., Lujan,S.A., Burkholder,A.B., Cox,P.B., Wu,Q., Zhou,Z.X., Haber,J.E. and Kunkel,T.A. (2018) Evidence that DNA polymerase delta contributes to initiating leading strand DNA replication in Saccharomyces cerevisiae. *Nat. Commun.*, **9**, 858.
10. Zhou,Z.X., Lujan,S.A., Burkholder,A.B., Garbacz,M.A. and Kunkel,T.A. (2019) Roles for DNA polymerase delta in initiating and terminating leading strand DNA replication. *Nat. Commun.*, **10**, 3992.
11. Donnianni,R.A., Zhou,Z.X., Lujan,S.A., Al-Zain,A., Garcia,V., Glancy,E., Burkholder,A.B., Kunkel,T.A. and Symington,L.S. (2019) DNA polymerase delta synthesizes both strands during break-induced replication. *Mol. Cell*, **76**, 371–381.
12. Daigaku,Y., Keszthelyi,A., Muller,C.A., Miyabe,I., Brooks,T., Retkute,R., Hubank,M., Nieduszynski,C.A. and Carr,A.M. (2015) A global profile of replicative polymerase usage. *Nat. Struct. Mol. Biol.*, **22**, 192–198.
13. Prindle,M.J. and Loeb,L.A. (2012) DNA polymerase delta in DNA replication and genome maintenance. *Environ. Mol. Mutagen.*, **53**, 666–682.
14. Kunkel,T.A. and Erie,D.A. (2005) DNA mismatch repair. *Annu. Rev. Biochem.*, **74**, 681–710.
15. Kunkel,T.A. and Erie,D.A. (2015) Eukaryotic mismatch repair in relation to DNA replication. *Annu. Rev. Genet.*, **49**, 291–313.
16. Kadyrov,F.A., Holmes,S.F., Arana,M.E., Lukianova,O.A., O'Donnell,M., Kunkel,T.A. and Modrich,P. (2007) Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease. *J. Biol. Chem.*, **282**, 37181–37190.
17. Kadyrov,F.A., Genschel,J., Fang,Y., Penland,E., Edelmann,W. and Modrich,P. (2009) A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 8495–8500.
18. Hall,M.C., Shcherbakova,P.V., Fortune,J.M., Borchers,C.H., Dial,J.M., Tomer,K.B. and Kunkel,T.A. (2003) DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair. *Nucleic Acids Res.*, **31**, 2025–2034.
19. Smith,C.E., Mendillo,M.L., Bowen,N., Hombauer,H., Campbell,C.S., Desai,A., Putnam,C.D. and Kolodner,R.D. (2013) Dominant mutations in S. cerevisiae PMS1 identify the Mlh1-Pms1 endonuclease active site and an exonuclease 1-independent mismatch repair pathway. *PLoS Genet.*, **9**, e1003869.
20. Torres,K.A., Calil,F.A., Zhou,A.L., DuPrie,M.L., Putnam,C.D. and Kolodner,R.D. (2022) The unstructured linker of Mlh1 contains a motif required for endonuclease function which is mutated in cancers. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2212870119.

21. Burkholder,A.B., Lujan,S.A., Lavender,C.A., Grimm,S.A., Kunkel,T.A. and Fargo,D.C. (2018) Muver, a computational framework for accurately calling accumulated mutations. *Bmc Genomics [Electronic Resource]*, **19**, 345.
22. Pavlov,Y.I., Frahm,C., Nick McElhinny,S.A., Niimi,A., Suzuki,M. and Kunkel,T.A. (2006) Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta. *Curr. Biol.*, **16**, 202–207.
23. Nick McElhinny,S.A., Stith,C.M., Burgers,P.M. and Kunkel,T.A. (2007) Inefficient proofreading and biased error rates during inaccurate DNA synthesis by a mutant derivative of Saccharomyces cerevisiae DNA polymerase delta. *J. Biol. Chem.*, **282**, 2324–2332.
24. Lujan,S.A., Clausen,A.R., Clark,A.B., MacAlpine,H.K., MacAlpine,D.M., Malc,E.P., Mieczkowski,P.A., Burkholder,A.B., Fargo,D.C., Gordenin,D.A., *et al.* (2014) Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.*, **24**, 1751–1764.
25. Zhou,Z.X., Lujan,S.A., Burkholder,A.B., St Charles,J., Dahl,J., Farrell,C.E., Williams,J.S. and Kunkel,T.A. (2021) How asymmetric DNA replication achieves symmetrical fidelity. *Nat. Struct. Mol. Biol.*, **28**, 1020–1028.
26. Welch,B.L. (1947) The generalization of `Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28–35.
27. Satterthwaite,F.E. (1946) An approximate distribution of estimates of variance components. *Biometrics*, **2**, 110–114.
28. Zbyněk,Š. (1967) Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Statist. Assoc.*, **62**, 626–633.
29. Lujan,S.A., Clark,A.B. and Kunkel,T.A. (2015) Differences in genome-wide repeat sequence instability conferred by proofreading and mismatch repair defects. *Nucleic Acids Res.*, **43**, 4067–4074.
30. Kunkel,T.A., Hamatake,R.K., Motto-Fox,J., Fitzgerald,M.P. and Sugino,A. (1989) Fidelity of DNA polymerase I and the DNA polymerase I-DNA primase complex from Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **9**, 4447–4458.
31. Thomas,D.C., Roberts,J.D., Sabatino,R.D., Myers,T.W., Tan,C.K., Downey,K.M., So,A.G., Bambara,R.A. and Kunkel,T.A. (1991) Fidelity of mammalian DNA replication and replicative DNA polymerases. *Biochemistry*, **30**, 11751–11759.
32. Shcherbakova,P.V., Pavlov,Y.I., Chilkova,O., Rogozin,I.B., Johansson,E. and Kunkel,T.A. (2003) Unique error signature of the four-subunit yeast DNA polymerase epsilon. *J. Biol. Chem.*, **278**, 43770–43780.
33. Korona,D.A., Lecompte,K.G. and Pursell,Z.F. (2011) The high fidelity and unique error signature of human DNA polymerase epsilon. *Nucleic Acids Res.*, **39**, 1763–1773.
34. Lynch,M., Sung,W., Morris,K., Coffey,N., Landry,C.R., Dopman,E.B., Dickinson,W.J., Okamoto,K., Kulkarni,S., Hartl,D.L., *et al.* (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9272–9277.
35. Nishant,K.T., Wei,W., Mancera,E., Argueso,J.L., Schlattl,A., Delhomme,N., Ma,X., Bustamante,C.D., Korbel,J.O., Gu,Z., *et al.* (2010) The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLoS Genet.*, **6**, e1001109.
36. Lang,G.I., Parsons,L. and Gammie,A.E. (2013) Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3*, **3**, 1453–1465.
37. Stirling,P.C., Shen,Y., Corbett,R., Jones,S.J. and Hieter,P. (2014) Genome destabilizing mutator alleles drive specific mutational trajectories in Saccharomyces cerevisiae. *Genetics*, **196**, 403–412.
38. Serero,A., Jubin,C., Loeillet,S., Legoix-Ne,P. and Nicolas,A.G. (2014) Mutational landscape of yeast mutator strains. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 1897–1902.
39. Zhu,Y.O., Siegal,M.L., Hall,D.W. and Petrov,D.A. (2014) Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E2310–E2318.
40. Dutta,A., Lin,G., Pankajam,A.V., Chakraborty,P., Bhat,N., Steinmetz,L.M. and Nishant,K.T. (2017) Genome dynamics of hybrid Saccharomyces cerevisiae during vegetative and meiotic divisions. *G3*, **7**, 3669–3679.
41. Sharp,N.P., Sandell,L., James,C.G. and Otto,S.P. (2018) The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E5046–E5055.
42. Williams,J.S., Lujan,S.A., Zhou,Z.X., Burkholder,A.B., Clark,A.B., Fargo,D.C. and Kunkel,T.A. (2019) Genome-wide mutagenesis resulting from topoisomerase 1-processing of unrepaired ribonucleotides in DNA. *DNA Repair (Amst.)*, **84**, 102641.
43. Sui,Y., Qi,L., Wu,J.K., Wen,X.P., Tang,X.X., Ma,Z.J., Wu,X.C., Zhang,K., Kokoska,R.J., Zheng,D.Q., *et al.* (2020) Genome-wide mapping of spontaneous genetic alterations in diploid yeast cells. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 28191–28200.
44. Streisinger,G., Okada,Y., Emrich,J., Newton,J., Tsugita,A., Terzaghi,E. and Inouye,M. (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harb. Symp. Quant. Biol.*, **31**, 77–84.
45. Kunkel,T.A. and Alexander,P.S. (1986) The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation. *J. Biol. Chem.*, **261**, 160–166.
46. Bebenek,K., Joyce,C.M., Fitzgerald,M.P. and Kunkel,T.A. (1990) The fidelity of DNA synthesis catalyzed by derivatives of Escherichia coli DNA polymerase I. *J. Biol. Chem.*, **265**, 13878–13887.
47. Gonzalez,C.E., Roberts,P. and Ostermeier,M. (2019) Fitness effects of single amino acid insertions and deletions in TEM-1 beta-lactamase. *J. Mol. Biol.*, **431**, 2320–2330.
48. Lang,G.I. and Murray,A.W. (2008) Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. *Genetics*, **178**, 67–82.
49. Kircher,M., Xiong,C., Martin,B., Schubach,M., Inoue,F., Bell,R.J.A., Costello,J.F., Shendure,J. and Ahituv,N. (2019) Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat. Commun.*, **10**, 3583.
50. Stenson,P.D., Mort,M., Ball,E.V., Howells,K., Phillips,A.D., Thomas,N.S. and Cooper,D.N. (2009) The Human Gene Mutation Database: 2008 update. *Genome Med.*, **1**, 13.
51. Chen,J.Q., Wu,Y., Yang,H., Bergelson,J., Kreitman,M. and Tian,D. (2009) Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. *Mol. Biol. Evol.*, **26**, 1523–1531.
52. Tran,H.T., Gordenin,D.A. and Resnick,M.A. (1999) The $3'\rightarrow5'$ exonucleases of DNA polymerases delta and epsilon and the $5'\rightarrow3'$ exonuclease Exo1 have major roles in postreplication mutation avoidance in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **19**, 2000–2007.
53. Rogacheva,M.V., Manhart,C.M., Chen,C., Guarne,A., Surtees,J. and Alani,E. (2014) Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. *J. Biol. Chem.*, **289**, 5664–5673.
54. Romanova,N.V. and Crouse,G.F. (2013) Different roles of eukaryotic MutS and MutL complexes in repair of small insertion and deletion loops in yeast. *PLoS Genet.*, **9**, e1003920.
55. Reyes,G.X., Zhao,B., Schmidt,T.T., Gries,K., Kloor,M. and Hombauer,H. (2020) Identification of MLH2/hPMS1 dominant mutations that prevent DNA mismatch repair function. *Commun. Biol.*, **3**, 751.

Received: February 21, 2024. **Revised:** June 25, 2024. **Editorial Decision:** June 29, 2024. **Accepted:** July 6, 2024
Published by Oxford University Press on behalf of Nucleic Acids Research 2024.