Genome-wide model for the normal eukaryotic DNA replication fork

Andres A. Larrea[a,b], Scott A. Lujan[a,b], Stephanie A. Nick McElhinny[a,b], Piotr A. Mieczkowski[c], Michael A. Resnick[a], Dmitry A. Gordenin[a], and Thomas A. Kunkel[a,b,1]

[a]Laboratory of Molecular Genetics and [b]Laboratory of Structural Biology, Department of Health and Human Services, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709; and [c]Department of Genetics, Carolina Center for Genome Science, University of North Carolina, Chapel Hill, NC 27599

Edited* by Thomas D. Petes, Duke University Medical Center, Durham, NC, and approved September 9, 2010 (received for review July 13, 2010)

To investigate DNA replication enzymology across the nuclear genome of budding yeast, deep sequencing was used to establish the pattern of uncorrected replication errors generated by an asymmetric mutator variant of DNA polymerase δ (Pol δ). Sequencing of 16 genomes identified 1,206-bp substitutions generated over 33 generations by L612M Pol δ in a mismatch repair defective strain. Alignment of sequences flanking these substitutions identified "hotspot" motifs for Pol δ replication errors. The substitutions were distributed evenly across all 16 chromosomes. The vast majority were transitions that occurred with a strand bias that varied in a predictable manner relative to known functional origins of replication. This strand bias strongly supports the idea that Pol δ is primarily a lagging strand polymerase during replication across the entire nuclear genome.

DNA polymerase δ | lagging strand replication | mutational hotspot | replication fidelity | mutator

Replication of the eukaryotic nuclear genome is intrinsically asymmetric, with a continuously replicated leading strand and a discontinuously replicated lagging strand (1). DNA polymerase α (Pol α) initiates new DNA chains and DNA polymerases ε (Pol ε) and δ (Pol δ), then performs the bulk of chain elongation. Variants of Pol ε and Pol δ (Pol δ L612M) that have distinctive error signatures were used to infer which DNA strand(s) each of these enzymes replicates in yeast. The results (2–4) are consistent with a model wherein Pol δ is primarily responsible for copying the lagging strand template, and Pol ε is primarily responsible for copying the leading strand template. Those studies used an 804-bp reporter gene adjacent to a single replication origin on chromosome 3 that fires frequently in early S phase (5). This situation is akin to "looking under a lamp post," because the yeast genome is 15,000 times larger (12 million bp, 16 chromosomes) and contains hundreds of replication origins that fire with different efficiencies and at various times in S phase (6). The genome also varies widely in sequence composition (7), and it is highly organized with respect to transcriptional status and chromatin content. Each of these variables may influence which of the many replication proteins are operating at replication forks, either directly or indirectly by affecting susceptibility to DNA damage. Among many questions about replication enzymology raised by the size and complexity of the nuclear genome, here we examine whether the role of Pol δ at the replication fork is constant or variable across the genome. To do so, we use deep sequencing to establish the pattern of base substitution mutations arising in a *pol3-L612M* mutant that is deficient in Msh2-dependent mismatch repair.

Results and Discussion

Rationale. To determine whether Pol δ primarily copies the lagging strand template across the whole genome, we made use of the mutational asymmetry of Pol δ L612M, which has high error rates for only two of the four possible mismatches that give rise to transitions (3, 8). Thus, Pol δ L612M is more likely to generate A·T-to-G·C mutations by misincorporating dGMP opposite template T than by misincorporating dCMP opposite template A. Similarly, it is more likely to generate G·C-to-A·T transitions by misincorporating dTMP opposite template G than by misincorporating dAMP opposite template C. This specificity is illustrated in Fig. 1, where these preferred pathways are depicted in blue for forks moving to the right from a replication origin or in red for forks moving to the left from an origin. Using the upper strand as a point of reference, these asymmetric error rates predict that if L612M Pol δ preferentially copies the lagging strand template (colored in Fig. 1), then the highest proportion of T-to-C and G-to-A substitutions (Fig. 1, *Upper*, *Left,* in blue) should reside immediately to the right of functional origins, and the highest proportion of C-to-T and A-to-G substitutions (in red) should reside immediately to the left of functional origins. We tested these predictions by performing whole genome sequence analysis as follows.

Whole Genome Sequence Analysis. A diploid strain was constructed that is homozygous for *pol3-L612M* (yeast *POL3* encodes the catalytic subunit of Pol δ) and heterozygous for a deletion of *MSH2* (3), a gene that is essential for repairing Pol δ replication errors (9). Tetrad dissection (Fig. 2) yielded two *pol3-L612M MSH2* single-mutant spores and two *pol3-L612M msh2Δ* double-mutant spores. All cells from each spore colony within a tetrad were suspended in rich yeast peptone dextrose adenine (YPDA) medium (Fig. 2, blue pathway) and grown to ≈10^{10} cells. This amount of growth corresponds to ≈33 generations during which L612M Pol δ replication errors that are not corrected by MMR result in mutations. The resulting populations of cells were used to obtain genomic DNA samples that serve as reference genomes. Single cells from these populations were then allowed to form single colonies (Fig. 2, red pathway). These colonies were grown in liquid medium to ≈10^{10} cells, and genomic DNA samples were isolated and sequenced to identify base substitutions that arose during the first cycle of growth.

As a master reference, we used the genome from passage 1 of a single *pol3-L612M* mutant (L03). This strain has a low spontaneous mutation rate (3×10^{-7} at *URA3*; ref. 8) because it is mismatch repair proficient and, therefore, corrects most replication errors generated by L612M Pol δ. The genomic DNA was sequenced on two lanes of a Genome Analyzer IIx (Illumina) and the data (22,500,098 paired-end and single reads) were pooled and aligned to a modified reference genome from strain S288c (7, 10). The resulting consensus genome (99.85% coverage relative to modified S288c) was annotated and served as the master reference for all other genome alignments. Relative to this master reference, 95% of the genome was covered by sequence analysis of the 39

Author contributions: S.A.N.M. and T.A.K. designed research; A.A.L., S.A.L., and P.A.M. performed research; A.A.L., S.A.L., S.A.N.M., P.A.M., and D.A.G. contributed new reagents/analytic tools; A.A.L., S.A.L., P.A.M., M.A.R., D.A.G., and T.A.K. analyzed data; and A.A.L., P.A.M., M.A.R., D.A.G., and T.A.K. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: kunkel@niehs.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010178107/-/DCSupplemental.

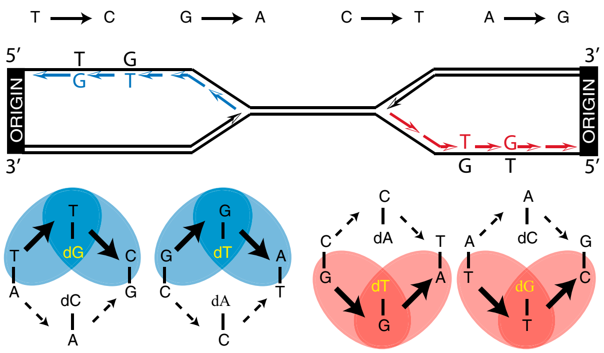

Fig. 1. Rationale to assign lagging strand replication errors to L612M Pol δ. This image depicts the predicted asymmetric distribution of the four transition mutations to the left and right of replication origins if L612M Pol δ replicates the lagging strand DNA template. See text for further description.

other genomes. Fig. 3*A* depicts the number of matched reads for each nucleotide in the 40 sequenced genomes, i.e., four reference and four outgrowth genomes for the *pol3-L613M* strain and 16 reference and 16 outgrowth genomes for the *pol3-L613M msh2Δ* strain. Base substitutions identified in more than one genome by pairwise comparisons with the master reference were then filtered out. This filtering was done to eliminate mutations that were not likely to have been generated by Pol δ L612M during outgrowth. As justification for this filtering, we calculated that the probability of the same mutation independently occurring in 2 of 16 sequenced genomes of double-mutant strains would require a hotspot whose mutation rate would need to be at least 800-fold higher than the hottest site for substitutions in our previous study with the *URA3* reporter gene (3). Additionally, 94% (767 of 813) of repeatedly identified base substitutions were found more than twice. This analysis further reduces the already low probability that repeatedly observed mutations originate from independent mutation events. Nonetheless, we cannot formally exclude the interesting possibility that extreme base substitution hotspot may exist in the genome.

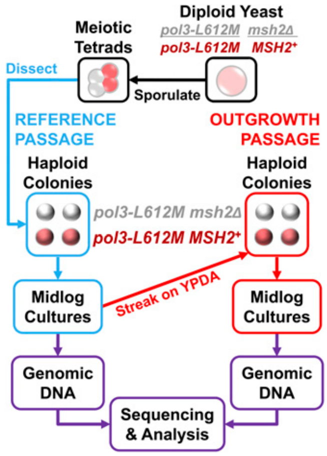

Fig. 2. Protocol to obtain genomic DNA for sequence analysis. A diploid strain homozygous for *pol3-L612M* and heterozygous for deletion of *MSH2* was sporulated to generate meiotic tetrads. These tetrads were dissected, and colonies resulting from the single-cell meiotic haploid products were grown overnight in 10 mL of YPDA medium. These cultures were added to 90 mL of YPDA medium and grown for 6 h to obtain $\approx 10^{10}$ cells, requiring ≈33 generations. This reference passage (blue path) is the period in which most or all of the mutations to be analyzed were generated. DNA obtained from this first passage, extracted from the whole population and, thus, representing the baseline haploid cells that emerged from tetrad dissection, served as the reference genome for each clone. Single colonies were obtained from these cultures by streaking out on YPDA plates, followed by a second round of growth in liquid YPDA medium. This outgrowth passage (red path) served to isolate and amplify genomes that were subject to mutation during the reference passage. DNA was extracted and sequenced to determine the uncorrected Pol δ L612M replication errors that had accumulated during the first round of growth.

When the genomes of the four *pol3-L612M* single-mutant outgrowths were sequenced, none had more than three substitutions when compared with their reference genome (Fig. 3*B*). In contrast, among the 16 genomes sequenced from outgrowths of *pol3-L612M msh2Δ* double mutants (Fig. 3*B*, filled bars), 13 contained between 37 and 129 substitutions, with 3 others having a smaller number. The difference in substitution density between the single- and double-mutant clones is highly significant (two-tailed Mann–Whitney, $P = 0.0014$).

In the genomes of the *pol3-L612M msh2Δ* double mutants that were sequenced after the outgrowth passage, we identified 1,206 unique single-base substitutions generated by L612M Pol δ during the reference passage in the absence of mismatch repair (Table S1). To quantify the extent of selective pressure during the reference passage, the 1,206 single base substitutions were subdivided into two classes. Of the 1,206 mutations, 883 (73%) were within an annotated gene. This fraction corresponds well with the amount predicted (75%), suggesting that there is little, if any, selective pressure against mutations in ORFs of genes. Among these 883 substitutions, only 600 (68%) lead to an amino acid change. This fraction is slightly less than predicted (689 substitution, 78%), suggesting that there is some selective pressure favoring silent mutations. This selective pressure makes sense given the relatively large portion of the yeast genome that is coding and the potential for synthetic lethality to arise from multiple, independently benign mutations.

The 1,206 substitutions were distributed uniformly along all 16 chromosomes (Fig. 4 *A* and *B*), with an average density of ≈1 substitution per 10,000 base pairs (Fig. 4*C*). This uniformity implies that Pol δ is a replicative polymerase for the vast majority of the nuclear genome. The density of mutations does not correlate with the distance from origins.

Strand Biases. More than 90% (1,099/1,206) of the base substitutions in the *pol3-L612M msh2Δ* double-mutant genomes were transitions (558 A·T to G·C and 541 G·C to A·T). Given L612M Pol δ's biased error rates, if L612M Pol δ preferentially copies the lagging strand template (Fig. 1, red or blue strand), then the highest proportion of T-to-C and G-to-A substitutions (in blue) should reside immediately to the right of functional origins, and the highest proportion of C-to-T and A-to-G substitutions (in red) should reside immediately to the left of functional origins. To determine whether this distribution is actually observed, we divided the distances between the 274 confirmed functional origins of replication in yeast (Table S2) into 20 equal intervals, each representing 5% of the distance between one origin and the next. When substitutions were binned based on their position relative to the nearest flanking origins, the proportions of each of the four transition mutations were biased exactly as predicted if Pol δ is primarily copying the lagging strand template during replication of the whole genome (Fig. 5 *A* and *B*). In other words, the highest proportion of T-to-C and G-to-A substitutions were to the right of functional origins, and the highest proportion of C-to-T and A-to-G substitutions were to the left of functional origins. These biases further imply that Pol δ is not contributing greatly to leading strand replication. By default, and supported by earlier results (2), our data suggest that Pol ε may be the primary leading strand polymerase for the genome. The resolution of the current analysis (one substitution per 10 Kb; Fig. 4*C*) does not exclude exceptions to this general model (see discussions in refs. 4 and 12), e.g., leading strand replication by Pol δ upon replication restart after encounters with DNA damage.

Interestingly, the proportions of four different substitutions are most similar to each other at the midpoint between origins

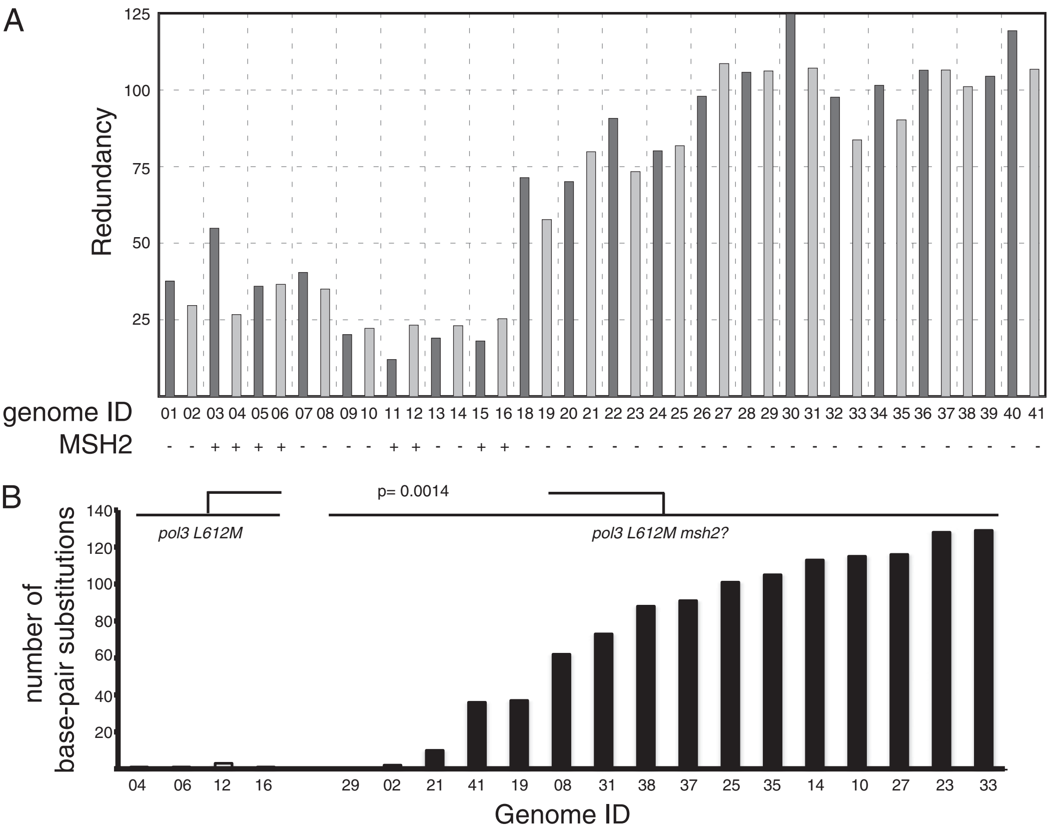

Fig. 3. Results for sequence analysis of 40 genomes. Four single-mutant clones (*pol2-L612M*, mismatch proficient) and 16 double mutants (*pol2-L612M msh2Δ*) were analyzed. In each case, one reference and one outgrowth genome were sequenced, representing a total of 40 genomes that are displayed side-by-side as pairs. The genome ID numbers range from 1 to 41; ID 17 is missing because it was not used for this study. (*A*) Plot showing the average number of reads per nucleotide (Redundancy) for each genome. The dark gray bars show redundancy for each reference genome, whereas the adjacent light gray bars show the redundancy for the paired outgrowth genome. (*B*) This graph depicts the number of single-base substitutions that accumulated in the genomes during the reference passage (Fig. 2, blue path), as detected by comparing the reference genome with the outgrowth genome for each clone.

(Fig. 4*B*), where replication forks converge. These data were then used to model the distribution of interorigin convergence points, as described in *SI Materials and Methods*. The results (Table S3 and Fig. S1) suggest considerable variability in replication fork convergence points, perhaps reflecting variations in the rate of fork movement, replication origin usage, the timing of origin firing, or some combination of these variables.

Mutable Motifs. Next, we addressed the extent to which Pol δ replication errors are sequence-context dependent. To increase confidence in assignment to lagging strand replication and the identity of the mismatch, we focused only on transition mutations whose position relative to an origin was <25% of the interorigin distance. Within this cutoff, roughly 96% of sequences for a given mutation type should have been generated by a fork moving from the nearest origin on the expected strand. This permitted alignment of sequences flanking 214 substitutions inferred to result from template G-dT mismatches and 242 substitutions inferred to result from template T-dG mismatches (Table S4). Alignment of five bases on either side of these mismatches revealed two short motifs. For transitions involving the template **G**-dT mismatch, the motif is template 5′-C(G/C/T)**G**C(C/A)G (Fig. 6*A*). This motif is consistent with our previous study (3) wherein the hottest G-to-A transition hotspot in the *URA3* gene occurred in a template sequence context comprising four of the same five flanking bases, (5′-CT**G**CAa). For transitions involving the **T**-dG mismatch (Fig. 5*B*), the motif is template 5′-(G/C)**T**(G/C)AG. A position weight matrix analysis (Fig. S2) revealed that the sequence motifs for transitions made by L612M Pol δ were substantially overrepresented in the genome-wide mutation set as compared with all G·C base pairs in the coding sequences of the *URA3* and *CAN1* genes, or all G·C base pairs in chromosome 3. These mutable motifs for L612M Pol δ are not the hard consensus motifs characteristic of sequence-specific binding proteins. A less well-defined motif is perhaps expected, because Pol δ must accurately copy a wide variety of sequences present in the nuclear genome. Additional biochemical and genetic studies will be required to quantify the contribution (or not) of each nucleotide to mutability and the biological relevance of these motifs. Nonetheless, the two motifs share the general characteristic of being rich in G·C base pairs immediately flanking both sides of the mismatch. This characteristic raises the possibility that these motifs may be more mutable than average because of increased duplex stability that facilitates mismatch extension at the expense of proofreading.

Summary and Implications. The widespread distribution and strand-biased patterns of mutagenesis observed here strongly support a model wherein Pol δ has a primary role in replicating the lagging strand template across the whole nuclear genome of budding yeast. The biased mutagenesis further suggests that Pol δ has a lesser role in leading strand replication, which by default supports the previous suggestion (4) that Pol ε, the other major yeast replicative polymerase, acts as the primary leading strand replicase. Whole genome sequencing is underway to further examine this latter possibility by using a M644G mutator derivative of Pol ε (2). We are also examining the feasibility of sequencing genomes of strains grown for many more generations than the ≈33 generations used here. In this way, it may be possible to identify large numbers of

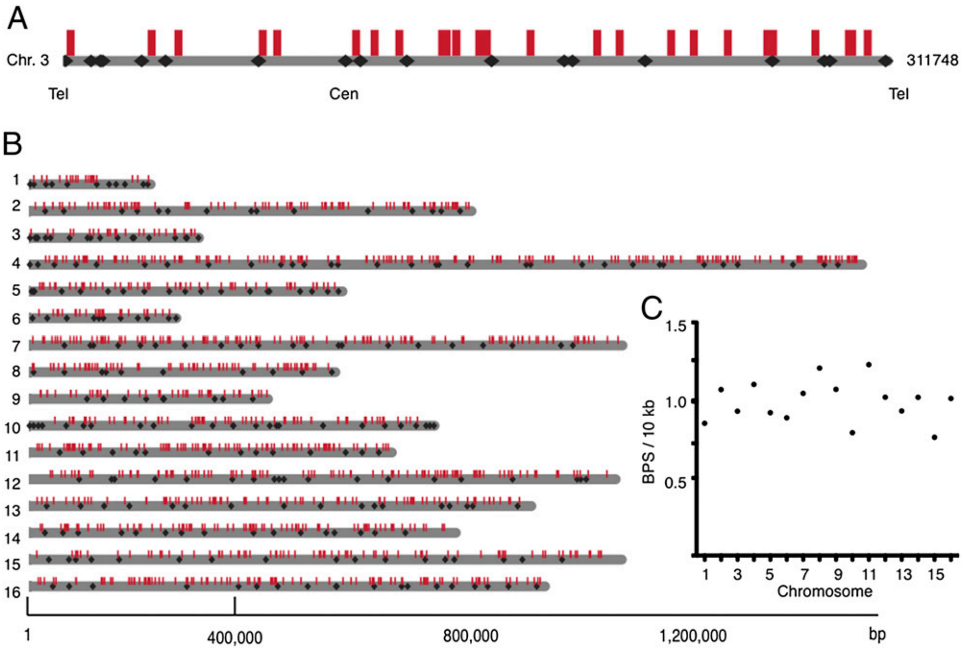

Fig. 4. Distribution of base-pair substitutions in the yeast genome. (*A*) *Saccharomyces cerevisiae* chromosome 3. Black diamonds represent confirmed replication origins, and red lines represent the locations of the base substitutions identified in the 16 outgrowth double-mutant genomes. (*B*) View of all 16 yeast chromosomes with 274 confirmed replication origins and the 1,206 identified base substitutions. (*C*) Density of substitutions per 10 Kb among the 16 chromosomes.

unselected mutations throughout the genomes of strains with lower mutation rates. This XXX may have wide applicability. As one example, if this strategy works with a *pol3-L612M* strain that is proficient in mismatch repair, then comparisons of mutational patterns in that strain to those reported here may provide insights into genomic parameters that determine mismatch repair efficiency, such as replication timing, transcriptional status, or chromatin architecture (13). The identification of short mutable motifs for base substitutions made during lagging strand replication by Pol δ reveals the potential power of deep sequencing to provide new biomarkers for molecular defects that may be associated with disease states. The prototype for this concept is the renowned

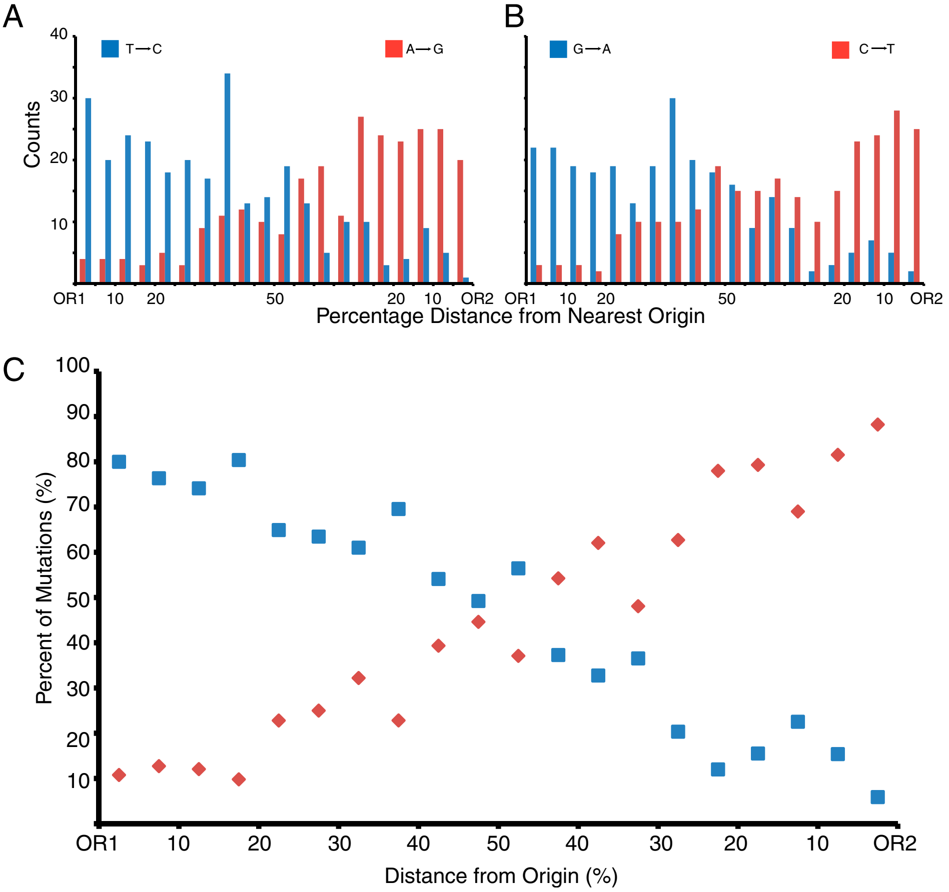

Fig. 5. Mutational asymmetry around replication origins. (*A*) T-to-C (blue) and A-to-G (red) mutations as a function of inter-origin distance. The bins represent percentages rather than absolute numbers of nucleotides, because origins are not equally spaced throughout the genome and strand bias depends on relative fork rates from adjacent origins. (*B*) G-to-A (blue) and C-to-T (red) mutations as a function of inter-origin distance. The analyses in *A* and *B* disregard variations in origin behavior. Accounting for these changes would only strengthen the conclusions by decreasing the background. (*C*) Percentage of all mutations in a bin that are G-to-A plus T-to-C (blue) compared with those that are A-to-G plus C-to-T (red).

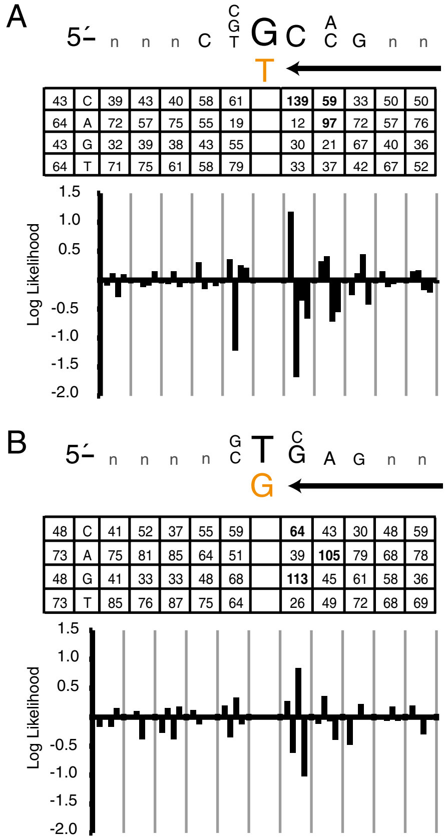

Fig. 6. Sequence motifs surrounding transition mutations. (*A*) The mutable motif depicted here is for the 214 G-to-A transitions inferred to result from misincorporation of dTMP opposite template G. In the inset table, the column of the far left shows the expected number of occurrences of C, A, G, and T, assuming a random distribution and given the base content of the yeast genome. The other columns list the observed numbers of occurrences of C, A, G, and T at each location among the 214 substitutions. The likelihood score contributions (natural log scale) for C, G, A, and T were plotted for the interval between −5 and +5, where position 0 is the error. The letters in the mutable motif correspond to positions where differences between expected and observed were statistically significant when calculated as described in *Materials and Methods*. (*B*) The same as in *A*, but for the 242 transitions inferred to result from misincorporation of dGMP opposite template T.

microsatellite instability used as a biomarker for replication errors that accumulate in tumors that are defective in DNA mismatch repair (14). The present study suggests that it may be possible to use deep sequencing, e.g., as in efforts to sequence cancer genomes (e.g., www.sanger.ac.uk/genetics/CGP; refs. 15 and 16, and references therein), to identify base substitution biomarkers that are signatures for defects in DNA replication or repair. By using mutational signatures to strongly support a genome-wide model for eukaryotic DNA replication, this study also reveals the power of deep sequencing to test important hypotheses in the fields of DNA replication, DNA repair, and mutagenesis, with many applications (e.g., ref. 17) anticipated as the technology continues to evolve (18).

Materials and Methods

Strains. All sequenced clones were derived from tetrad dissections of the diploid SNM1037 strain described in ref. 3.

Preparing Genomic DNA. Cells were cultured by following the protocol (Fig. 2). Cells were harvested from 100-mL cultures, resuspended in 10 mL of 1 M sorbitol, and split into 1-mL aliquots. DNA was isolated from the cells by using the Epicentre MasterPure Yeast DNA Purification Kit (MPY80200).

Library Preparation and Whole Genome Sequencing. Two to 5 μg of total DNA was fragmented to sizes between 200 and 800 bp by using a "Bioruptor" sonicator. DNA was purified on PCR purification columns (Qiagen) and separated on a 2% agarose gel. DNA fragments of 300 bp were extracted from the gel (Qiagen) and used for library preparation by using the protocol for Genome Analyzer IIx recommended by Illumina. This library was analyzed on an Expirion Automated Electrophoresis System (Bio-Rad) and quantified by using a Qubit fluorometer (Invitrogen). DNA was diluted to 15 nM, loaded on the flow cell, and sequenced on a Genome Analyzer IIx instrument (Illumina). We performed paired-end sequencing (2 × 36 cycles). The first step of data analysis was done with an Illumina analysis pipeline that delivers reads in fastq format that are then used in subsequent steps.

Master Reference Assembly. All sequence analyses, including reference sequence generation, sequence alignment (both reference and outgrowth), and SNP identification, were done by using the CLC Genomics Workbench (CLC Bio). L03, a single mutant bearing only the *pol3L612M* mutation, was used as the master reference. Two lanes of data were pooled and aligned against the S288C reference from the yeast genome database in "random" mode. This S288c reference was first modified to include large changes that had been made during strain construction (e.g., insertion of *URA3* near ARS306, deletion of *MSH2,* and insertion of Hyg^R cassette in its place). The annotations from the corrected S228c genome were then transferred onto the consensus sequence from the alignment. The two lanes of L03 data were then aligned against this new consensus genome in "ignore" mode. SNPs from this alignment were then identified and manually incorporated into the reference sequence used for this alignment. This newly modified sequence is the L03 master reference.

Sequence Alignments and Identification of Base-Pair Substitutions. All other genomic DNA samples were run on a single lane. Data from each lane were aligned to the L03 master reference in ignore mode. Once each genome had been aligned, a SNP analysis was run where a variant was considered a SNP only if the coverage was greater than eightfold and the variant was present in at least 80% of the reads. All SNPs were then filtered out and only unique mutations that were observed in only one isolate were kept for subsequent analysis.

Identifying Mutable Motifs. To identify mutable motifs, pairwise alignments (A vs. C/G/T, C vs. A/G/T, G vs. C/A/T, T vs. A/C/G, C/A vs. G/T, C/G vs. A/T, or C/T vs. A/G) were performed at each position between −10 and +10 (where 0 is the position of the mutagenized template position). Using transitions present only within the first and last 25% of each interorigin region, a position weight matrix was generated for mismatches inferred to be either G-dT or T-dG. For each pulse-width modulation, the contribution to the overall position weight score (PWS) of a given sequence is the natural log of the likelihood ratio (Fig. 6, black bars), based on the observed and expected occurrences for each base at each position (Fig. 6, inset tables). Within each motif, bases were assigned at each position based on three tiered criteria. First, the distribution of all four bases was required to differ from the expected distribution by more than an amount likely to be due to random chance (χ^2 test; $P < 0.05$). Second, each possible pairwise division of the observed distribution was tested against its expected equivalent by χ^2 analysis and used if the *P* value was <0.01. Third, the identity that was ultimately assigned in the motif had the best χ^2 score among all candidates at that position. Thus, both the PWS and the assigned identity represent deviation from expectation beyond random chance, rather than representing overall preponderance of a certain base or bases.

ACKNOWLEDGMENTS. We thank Karen Adelman, Jessica Williams, and Steven Roberts for their thoughtful comments on this manuscript. This work was supported by Division of Intramural Research of the National Institutes of Health, National Institute on Environmental Health Sciences Project Z01 ES065089 (to T.A.K.) and Project ES065073-19 (to M.A.R.); and by National Institutes of Health Grant RC1 ES018091-02 and the University Cancer Research Fund (both to P.A.M.).

1. Burgers PM (2009) Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem* 284:4041–4045.
2. Pursell ZF, Isoz I, Lundström EB, Johansson E, Kunkel TA (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* 317:127–130.
3. Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA (2008) Division of labor at the eukaryotic replication fork. *Mol Cell* 30:137–144.
4. Kunkel TA, Burgers PM (2008) Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol* 18:521–527.
5. Poloumienko A, Dershowitz A, De J, Newlon CS (2001) Completion of replication map of Saccharomyces cerevisiae chromosome III. *Mol Biol Cell* 12:3317–3327.
6. Nieduszynski CA, Hiraga S, Ak P, Benham CJ, Donaldson AD (2007) OriDB: A DNA replication origin database. *Nucleic Acids Res* 35(Database issue):D40–D46.

7. Cherry JM, et al. (1997) Genetic and physical maps of Saccharomyces cerevisiae. *Nature* 387(6632, Suppl):67–73.
8. Nick McElhinny SA, Stith CM, Burgers PM, Kunkel TA (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.
9. Hsieh P, Yamane K (2008) DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mech Ageing Dev* 129:391–407.
10. Cherry JM, et al. (1998) SGD: Saccharomyces Genome Database. *Nucleic Acids Res* 26: 73–79.
11. Wyrick JJ, et al. (2001) Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: High-resolution mapping of replication origins. *Science* 294:2357–2360.
12. Pavlov YI, Shcherbakova PV (2010) DNA polymerases at the eukaryotic fork-20 years later. *Mutat Res* 685:45–53.
13. Hawk JD, Stefanovic L, Boyer JC, Petes TD, Farber RA (2005) Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. *Proc Natl Acad Sci USA* 102:8639–8643.
14. Umar A, et al. (2004) Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96:261–268.
15. Fox EJ, Salk JJ, Loeb LA (2009) Cancer genome sequencing—an interim analysis. *Cancer Res* 69:4948–4950.
16. Friedberg EC (2010) A comprehensive catalogue of somatic mutations in cancer genomes. *DNA Repair (Amst)* 9:468–469.
17. Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA (2008) Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast Saccharomyces cerevisiae. *PLoS Genet* 4:e1000264.
18. Check Hayden E (2009) Genome sequencing: The third generation. *Nature* 457: 768–769.