Mutation Rates across Budding Yeast Chromosome VI Are Correlated with Replication Timing

Gregory I. Lang[1] and Andrew W. Murray*[,1,2]

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts

[2]FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts

*Corresponding author: E-mail: amurray@mcb.harvard.edu.

Abstract

Previous experimental studies suggest that the mutation rate is nonuniform across the yeast genome. To characterize this variation across the genome more precisely, we measured the mutation rate of the *URA3* gene integrated at 43 different locations tiled across Chromosome VI. We show that mutation rate varies 6-fold across a single chromosome, that this variation is correlated with replication timing, and we propose a model to explain this variation that relies on the temporal separation of two processes for replicating past damaged DNA: error-free DNA damage tolerance and translesion synthesis. This model is supported by the observation that eliminating translesion synthesis decreases this variation.

Key words: mutation rate, replication timing, DNA damage tolerance, budding yeast.

Introduction

The vast majority of mutations affecting fitness are deleterious; therefore, there is selection pressure to keep mutation rates low. In response, cells have evolved a number of mechanisms to avoid errors in DNA replication and correct them when they occur (Friedberg et al. 2005). Biases in the generation or repair of DNA damage can lead to variation in mutation rates across the genome. In the budding yeast, *Saccharomyces cerevisiae*, several experimental studies suggest that mutation rates across the genome are nonuniform.

One experiment looked at the frequency of mutations that convert tRNA-Tyr into ochre suppressor mutations. This change, a GC to TA transversion, converts the GTA tRNA-Tyr anticodon into TTA, enabling it to recognize the TAA ochre stop codon (Ito-Harashima et al. 2002). The yeast genome contains eight nearly identical tRNA-Tyr genes distributed between five chromosomes. If mutation rates are uniform across the yeast genome, each of the mutations that create ochre suppressors should occur with equal probability. However, the tRNA-Tyr genes do not mutate at equal frequency; mutations at one locus (*SUP6-o*) represent 31% of the ochre suppressors, whereas two other loci (*SUP2-o* and *SUP8-o*), each account for only 2% of the suppressors, suggesting that the rate of GC to TA transversions is nonuniform across the yeast genome (Ito-Harashima et al. 2002). The rate of tRNA-Tyr ochre suppressor mutations is uncorrelated with replication timing, the rate of fork movement, or proximity to centromeres, telomeres, Ty, or delta elements (Ito-Harashima et al. 2002).

Another experiment examined the effect of genome position on the stability of a microsatellite sequence. A synthetic microsatellite (16.5 copies of the GT dinucleotide) was placed in frame with the *URA3* gene and integrated at ten locations across the yeast genome and loss-of-function *ura3* mutants were selected by growth on 5-fluoro-orotic acid (5FOA) (Hawk et al. 2005). The construct was integrated near genomic features such as centromeres, telomeres, replication origins, and at the *SUP2-o* and *SUP6-o* loci, which were shown to mutate at different frequencies (Ito-Harashima et al. 2002). These ten strains show a 16-fold difference in the mutation rate to 5FOA resistance, and the majority of these mutations resulted from frameshift mutations within the polyGT tract (not mutations in the *URA3* coding sequence). Mismatch repair is responsible for correcting potential frameshifts that arise by slippage during DNA replication (Friedberg et al. 2005; Kunkel and Erie 2005). In order to determine if the varying mutation rate is due to varying production of replication errors or varying ability to correct errors, a key gene involved in mismatch repair, *MSH2*, was deleted in six of the strains. In the mismatch repair-deficient strains, the mutation rate variation is reduced from 16-fold to 2-fold, suggesting that the variation in microsatellite stability across the genome

is largely due to variation in the efficiency of mismatch repair (Hawk et al. 2005).

Although this study identified mismatch repair as the mechanism responsible for variation of microsatellite stability, it did not identify genomic features underlying the variation in the efficiency of mismatch repair. The rate of microsatellite frameshift mutations is not correlated with proximity to replication origins, orientation relative to replication origins, replication timing, rates of transcription, or GC content (Hawk et al. 2005). The authors propose that this variation may result from unknown factors that lead to differences in the ability of mismatch repair to recognize and/or access mismatched bases (Hawk et al. 2005).

In order to characterize mutation rate variation within the yeast genome and to determine genomic features correlated with mutation rate, we systematically integrated the URA3 gene across a single yeast chromosome. We have previously shown (Lang and Murray 2008) that spontaneous loss-of-function mutations in this gene occur at a wide variety of sites, ensuring that our assay would interrogate different types of mutations in different sequence contexts. Using the fluctuation assay (Luria and Delbrück 1943), we measured the rate at which each strain produced 5FOA-resistant ura3 mutations. We picked Chromosome VI for several reasons: it is the second smallest chromosome (270 kb, 40 kb larger than Chromosome I), it is close to being metacentric, two of the tRNA-Tyr ochre suppressor genes are on this chromosome, and none of the 30 known mutator alleles are on this chromosome. We created 43 strains with the URA3 gene integrated at a different location tiled across Chromosome VI. Using this collection of 43 strains, we show that mutation rate varies at least 6-fold across the yeast genome, that this variation exists on a length scale of 50–100 kb, and that mutation rate is correlated with replication timing, potentially as a consequence of the temporal separation of two mechanisms of DNA damage tolerance: error-free DNA damage tolerance and translesion synthesis.

Materials and Methods

Primers, Strains, and Media

The sequences of primers used for plasmid construction, gene replacement, verification, and sequencing are described in supplementary figure S1. The yeast strain yGIL066 (uracil prototroph, W303 background) was used a source for the URA3 gene used in this study. The yeast strains used in these experiments were derived from the Yeast MATa Knockout Strain Collection (Open Biosystems) and were modified by replacing the KanMX cassette with the URA3 gene (table 1). Yeast cultures were grown in either complete synthetic media (SC) or complete synthetic media without uracil (SC-Ura). Fluctuation assays were plated onto either 10× canavanine (complete synthetic media without arginine [SC-Arg], 0.6 g/l L-canavanine, Sigma-Aldrich, St Louis,

MO), or 5FOA (SC-Ura, 1 g/l 5FOA, Sigma-Aldrich). In preparation for plating several spots of mutant cultures on each plate, the plates were overdried by pressing a Whatman filter paper (Grade 3, 90 mm) onto the plates using a replica plating block and allowing the filter to remain in place for at least 30 min. The filters remove approximately 1 ml of liquid, and plates can be used for several days after filters have been removed.

Plasmid Construction

The plasmid pGIL001 was constructed to facilitate replacement of the KanMX4 cassette with the URA3 gene. The URA3 gene was amplified from a genomic preparation of the yeast strain yGIL066 using primers URA3extF_integration and URA3extR_integration. These primers amplify a 1.8-kb fragment containing the yeast URA3 promoter and coding sequence. In addition, these primers contain 60 bp of homology to the KanMX4 cassette. This polymerase chain reaction (PCR) fragment was used to transform the strain YEL020CΔ:KanMX from the Yeast Knockout Strain Collection. Transformants were sequenced using primers U1, D1, URA3intF2, and URA3intF3 to identify ones where no mutations were introduced into the URA3 gene during the construction. The kanMXΔ:URA3 cassette was amplified using primers U1 and D1, the universal upstream and downstream primers from the yeast deletion collection (Winzeler et al. 1999), digested with EcoRI and BamHI, and cloned into the plasmid pFA6a-KanMX4 (which was digested with EcoRI and BamHI to remove the KanMX4 gene and expose the corresponding restriction enzyme overhangs). Proper construction of the plasmid was verified by restriction enzyme digestion and sequencing. The resulting plasmid, pGIL001, is pFA6a-KanMX4 with a 1.8-kb URA3 fragment is inserted in the KanMX4 cassette. On either side of the URA3 fragment is 300 bp of homology to the KanMX4 cassette including a partial TEF promoter upstream, and some remaining KanMX4 coding sequence and the TEF terminator downstream. The URA3 sequence of pGIL001 differs from the published genomic sequence for URA3 by eight mutations. One mutation (an insertion of a T to a run of seven T's in the promoter region) was created during the construction of this plasmid. The other seven were present in the URA3 gene in our laboratory W303 background. Only one of these seven mutations is in the coding sequence and results in the substitution of serine for alanine at position 160.

Plasmid pGIL008 was constructed to facilitate deletion of ARS607. Primers ARS607_F5 and ARS607_R5 were annealed and extended, generating a 160-bp fragment corresponding to approximately 80 bp of homology to the regions flanking ARS607 but devoid of the 111-bp ARS607 sequence itself. This fragment was amplified using primers ARS607_F6 and ARS607_R6, which contain NsiI and EcoRI sites, respectively. The fragment was cut and cloned

Table 1

Strains Used for *URA3* Integration

Strain	ORF	Description	Position[a]	Strain	ORF	Description	Position
GL-0	YEL020C	Hypothetical	Chromosome V	GL-25	YFL003C	MSH4	134516
GL-1[b]	YFL063W	Hypothetical	5066	GL-26	YFL001W	DEG1	147126
GL-2	YFL056C	AAD6	14793	GL-27	YFR001W	LOC1	149105
GL-3	YFL055W	AGP3	17004	GL-28	YFR006W	Hypothetical	156139
GL-4	YFL054C	Hypothetical	20847	GL-29	YFR007W	Hypothetical	159293
GL-5	YFL052W	Hypothetical	28232	GL-30	YFR009W	GCN20	162482
GL-6	YFL050C	ALR2	33272	GL-31[b]	YFR012W	Hypothetical	167881
GL-7	YFL049W	Hypothetical	36803	GL-32	YFR014C	CMK1	172529
GL-8	YFL047W	RGD2	40421	GL-33	YFR016C	Hypothetical	177034
GL-9	YFL044C	YOD1	44655	GL-34	YFR017C	Hypothetical	182262
GL-10	YFL041W	FET5	49139	GL-35[b]	YFR019W	FAB1	184490
GL-11[b]	YFL036W	RPO41	58781	GL-36	YFR021W	ATG18	194800
GL-12	YFL034W	Uncharacterized	65475	GL-37	YFR023W	PES4	199862
GL-13	YFL032W	Hypothetical	74870	GL-38	YFR026C	Hypothetical	205736
GL-14	YFL027C	GYP8	80417	GL-39	YFR030W	MET10	213300
GL-15	YFL025C	BST1	84143	GL-40	YFR032C	Hypothetical	222078
GL-16	YFL023W	BUD27	90984	GL-41	YFR035C	Hypothetical	226109
GL-17	YFL021W	GAT1	95964	GL-42	YFR039C	Hypothetical	231999
GL-18	YFL019C	Hypothetical	100246	GL-43[b]	YFR043C	Hypothetical	239101
GL-19	YFL015C	Hypothetical	106463	GL-44	YFR045W	Hypothetical	242129
GL-20	YFL012W	Hypothetical	110641	GL-45[b]	YFR049W	YMR31	248510
GL-21	YFL011W	HXT10	112339	GL-46	YFR053C	HXK1	253579
GL-22	YFL010C	WWM1	115102	GL-47	YFR054C	Hypothetical	258842
GL-23	YFL007W	BLM3	123474	GL-48	YFR055W	Hypothetical	264191
GL-24	YFL004W	VTC2	131805	GL-49	YFR057W	Hypothetical	269048

[a] Position is the location at which the gene is first encountered moving across Chromosome VI starting at the left telomere.

[b] Strains omitted from analysis (see Materials and Methods).

into the *Nsi*I and *Eco*RI sites of pGIL001. The resulting plasmid, pGIL008, contains the *URA3* gene followed by a 160-bp fragment corresponding to approximately 80 bp of sequence from each side of *ARS607*.

Strain Construction

Forty-nine locations along Chromosome VI were selected for integration of the *URA3* gene (table 1). To aid in strain construction, we took advantage of the existence of the Yeast Knockout Strain Collection, where nearly every non-essential open reading frame (ORF) was systematically deleted and replaced with the *KanMX4* reporter, conferring resistance to the drug G418 (Winzeler et al. 1999). To integrate *URA3* at different locations, strains were pulled from the Yeast Knockout Strain Collection and the *KanMX4* cassette was replaced with the *URA3* gene. Our locations, therefore, are restricted to the locations of *KanMX4* in the Yeast Knockout Strain Collection and are enriched for protein coding sequences (although some "hypothetical" ORFs in table 1 are likely to be intergenic). Locations were chosen to avoid gene replacements that have fitness defects; therefore, many of the integrations were made in hypothetical ORFs (those that have no ascribed function and were identified by their likelihood of encoding protein). The coverage of Chromosome VI is shown in supplementary figure S1.

To replace the *KanMX4* cassette with the *URA3* gene, pGIL001 was digested with *Eco*RI and *Bam*HI, phenol chloroform extracted, ethanol precipitated, and used to transform each of the 49 strains. Transformants were subjected to three rounds of screening. First each was screened for the proper phenotype (Uracil prototrophy and G418 sensitivity). PCR, using primers U1 and D1, was used to verify integration in the correct genomic location (when necessary, ORF-specific primers were also used). The amplified *kanMX4Δ:URA3* cassettes were then sequenced using primers U1, D1, *URA3*intF2, and *URA3*intR2 to verify 1) that the strains were correct based upon the barcode used in the Yeast Knockout Strain Collection and 2) that no mutations were introduced in the *URA3* gene during transformation.

To manipulate replication timing, a two-step method was used in order to create a clean deletion of the early and efficient origin, *ARS607* (supplementary fig. S2). First, the *URA3* gene followed by approximately 80 bp of homology to the regions flanking *ARS607* (but devoid of the 111-bp *ARS607* sequence itself) was amplified from plasmid the pGIL008 using primers *ARS607*_F4 and *ARS607*_R7. This fragment was used to transform the strain YFR021WΔ:*KanMX* from the Yeast Knockout Strain Collection. The second step of strain construction was to select for

popout of the *URA3* gene. Following *URA3* integration, 12 transformants were grown overnight in SC-Ura and cells were plated on 5FOA to select for loss of *URA3*. Deletion of *ARS607* was determined by PCR using primers *ARS607*ext_F1 and *ARS607*ext_R1, which flank the *ARS607* sequence. Following popout of *URA3* at the deleted *ARS607* locus, the *URA3* gene was integrated in place of the *KanMX4* cassette to create the strain GL·36$^{ARS607\Delta}$.

To eliminate translesion synthesis, the *rev1Δ:KanMX4* cassette was amplified from the Yeast Knockout Strain Collection using primers *REV1*extF1 and *REV1*extR1, and this fragment was used to transform strains GL·3, GL·15, GL·24, and GL·37. Deletion of *REV1* was verified phenotypically by assaying for UV sensitivity and by PCR using primers *REV1*intF1/*REV1*extR3 and *KanMX*intF/*REV1*extR3.

Fluctuation Assays

Fluctuation assays were performed essentially as described previously (Lang and Murray 2008). For each strain, forty-eight 100 μl cultures and forty-eight 200 μl cultures of a 1:10,000 dilution of a saturated overnight culture were established in a 96-well plate. Twelve 100 μl cultures and twelve 200 μl cultures were pooled to determine the number of cells per culture. The remaining thirty-six 100 μl cultures were plated onto canavanine plates (0.6 g/l) and the remaining thirty-six 200 μl cultures were plated onto 5FOA plates. Mutants were counted after two (canavanine) or seven (5FOA) days of growth and mutation rates were calculated using the Ma–Sandri–Sarkar maximum likelihood method (Sarkar et al. 1992). Ninety-five percent confidence intervals were calculated using equations (24) and (25) from Rosche and Foster (2000).

Computational Analysis

Mutation rates were calculated using the Matlab program *findMLm* described previously (Lang and Murray 2008). Mutation rates across Chromosome VI were compared with several other data sets to look for correlations; these include the production of double-strand breaks during meiosis (Gerton et al. 2000) and replication timing (Raghuraman et al. 2001). The Spearman rank correlation test was performed in Matlab and *P* values were determined by permutation. The sequences of RM11-1a and YJM789 were obtained from the Broad Institute Fungal Genome Initiative (http://www.broad.mit.edu/annotation/fungi/fgi/) and the Stanford Genome Technology Center (version 2, http://med.stanford.edu/sgtc/research/yjm789.html), respectively. Genes were identified by blasting the S288c sequences against these databases. Sequences were manually extracted and aligned to S288c and K_s (the number of synonymous substitutions per synonymous site) between S288c, RM11-1a, and YJM789 was calculated for each ORF. ORFs where S288c contained the allele of one of

the other strains (RM11-1a or YJM789) were excluded from the analysis. K_s values for *S. cerevisiae* versus *Saccharomyces paradoxus* were obtained from Kellis et al. (2003). Perl scripts written to calculate K_s and GC content are available by request.

Identification of Outliers

The original strain construction for this experiment involved integrating *URA3* at 49 locations across Chromosome VI. Fluctuation assays were performed on all 49 strains; however, six of the strains were eliminated from further analysis. Difficulties with three of the strains were apparent during construction. For two strains (GL·43 and GL·45), we were unable to generate a PCR product using either the universal primers or the ORF-specific primers, both of which were able to generate PCR products in a wild-type strain. Therefore, it is possible that a chromosomal rearrangement occurred in these strains. Interestingly, these two strains have the lowest mutation rates of the 49 measured strains (0.5×10^{-8} and 0.7×10^{-8}, respectively). For the strain GL·1, ORF-specific PCR shows that in the strain pulled from the deletion collection, the *KanMX4* is not integrated at the subtelomeric YFL063W locus. Phenotypically, we show that *URA3* successfully replaced the *KanMX4* cassette; however, because this strain is one where the universal primers fail to produce a PCR product, we were unable to determine the location of the *kanMX4Δ:URA3* cassette. Interestingly, this strain shows the highest mutation rate (46.8×10^{-8}, 5.3-fold higher than the second highest strain, which is also an outlier, described below), as one might expect for a subtelomeric reporter, which can be inactivated by silencing as well as mutation. Given the similarity of yeast telomeres, it is possible that this reporter is located in a subtelomeric region on a different chromosome.

In addition to the three outliers detected during strain construction, three outliers were detected during the experiment. As mentioned above, the strain with the second highest mutation rate at *URA3* (8.8×10^{-8}) is also an outlier. This is because this strain (GL·11) also has an elevated mutation rate at *CAN1* (4.5×10^{-7}, 4.8-fold higher than the median), indicating that this strain has a globally elevated mutation rate. None of the 30 known mutator alleles are found on Chromosome VI, and there is no reason to suspect that the gene deleted during construction of the strain (*RPO41*, encoding a mitochondrial RNA polymerase) is a mutator allele. Given that the yeast genome has been screened for mutator alleles (Huang et al. 2003), one of this strength is unlikely to have gone undetected; therefore, it is likely that this strain carries a spontaneous, transformation-induced mutation in one of the 30 genes that are known to be capable of giving rise to mutators. Two strains (GL·31 and GL·35) were eliminated from further analysis because they behave differently on 5FOA than the rest of the strains:
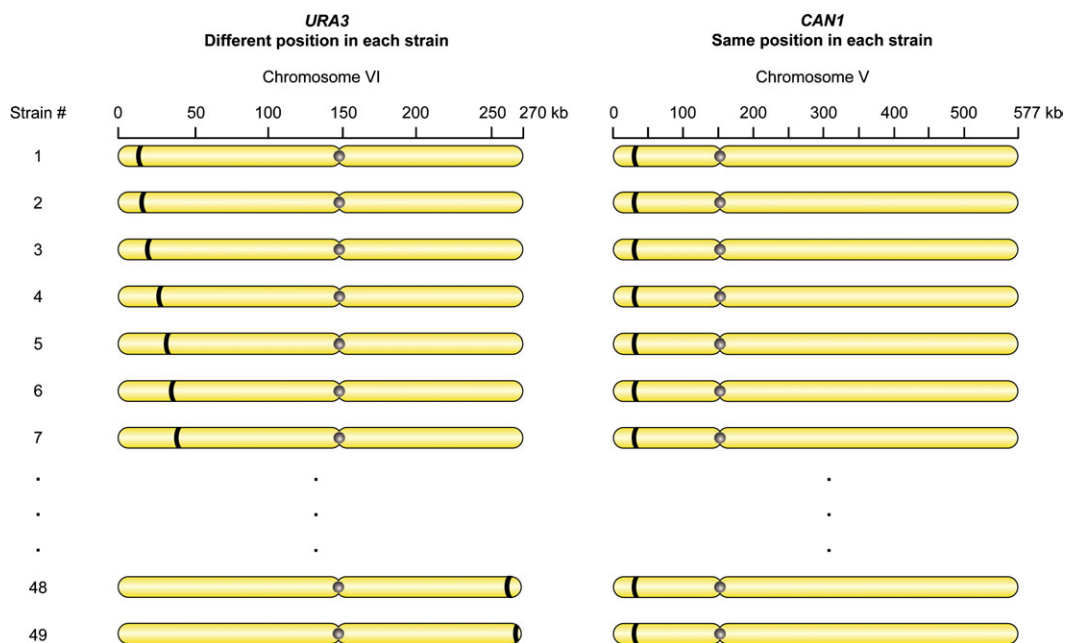
Fig. 1.—Schematic of strain construction. Forty-nine strains were constructed with the *URA3* gene integrated at a different location tiled across Chromosome VI. Each strain contains the wild-type *CAN1* gene at its endogenous location on Chromosome V. The different locations of the *URA3* gene are represented as bands on the chromosomes.

URA3 cells are more sensitive to 5FOA, resulting in a less background growth and larger *ura3* colonies. For fluctuation assays, these properties are desired, but because these were the only two strains behaving in this way, both were excluded. Both strains show a high mutation rate at *URA3*.

Notebook

The complete laboratory notebook describing these experiments is available at http://www.genomics.princeton.edu/glang/notebooks.htm.

Results

Mutation Rate Varies across Chromosome VI

To determine whether the mutation rate varies across the yeast genome, we created 43 strains, each of which has the *URA3* gene integrated at a different location tiled across Chromosome VI. In addition to the *URA3* gene, all of these strains contain the *CAN1* gene at its endogenous locus (fig. 1). Both genes confer sensitivity to a drug, allowing us to measure the rate at which they are inactivated by mutation: *URA3*, which encodes orotidine-5′-monophosphate decarboxylase, the last step in uracil biosynthesis, makes cells sensitive to 5FOA and *CAN1*, which encodes an arginine permease, makes cells sensitive to canavanine, an arginine analog. By measuring the mutation rates at both loci, we can control for any strain-specific effects that elevate or depress mutation rates across the genome (Lehner et al. 2007). Fluctuation assays were performed using these 43 strains to determine the mutation rate at the *URA3* and *CAN1* genes. Figure 2 shows the results from this experiment. The mutation rate at the *CAN1* locus varies between the 43 strains, but this variation is within the range that is expected by chance. For each strain, our estimate of the mutation rate has a 95% confidence interval, allowing us to ask if our estimate of the mutation rate lies outside the 95% confidence interval of the strain that has the median mutation rate of the 43 strains we tested (its estimated mutation rate and 95% confidence interval shown in red in fig. 2). For mutations at *CAN1*, only one of the strains has a mutation rate that lies outside this interval (fig. 2B). Because we examined 43 strains, the expectation is that roughly two strains our estimation of the mutation rate would lie outside this confidence interval, even if the actual mutation rate at *CAN1* was identical in all the strains. In contrast, the mutation rate at the *URA3* gene varies far more than expected by chance. There are 25 strains whose mutation rate lies outside the 95% confidence interval of the strain that has the median mutation rate (fig. 2A). The degree of variability is better illustrated by making all 903 pairwise comparisons between mutation rates in the 43 strains (fig. 3). For mutation rates at *CAN1*, there are only three significant pairwise comparisons (fig. 3B; the plot is symmetrical across the diagonal, thus every comparison is shown twice); for *URA3*, however, 262 of the 903 pairwise comparisons are significantly different (fig. 3A). From the pairwise comparisons, we identify three regions of Chromosome VI that have regionally different mutation rates, each 50–100 kb long: a region of high mutation rate on the left arm of the
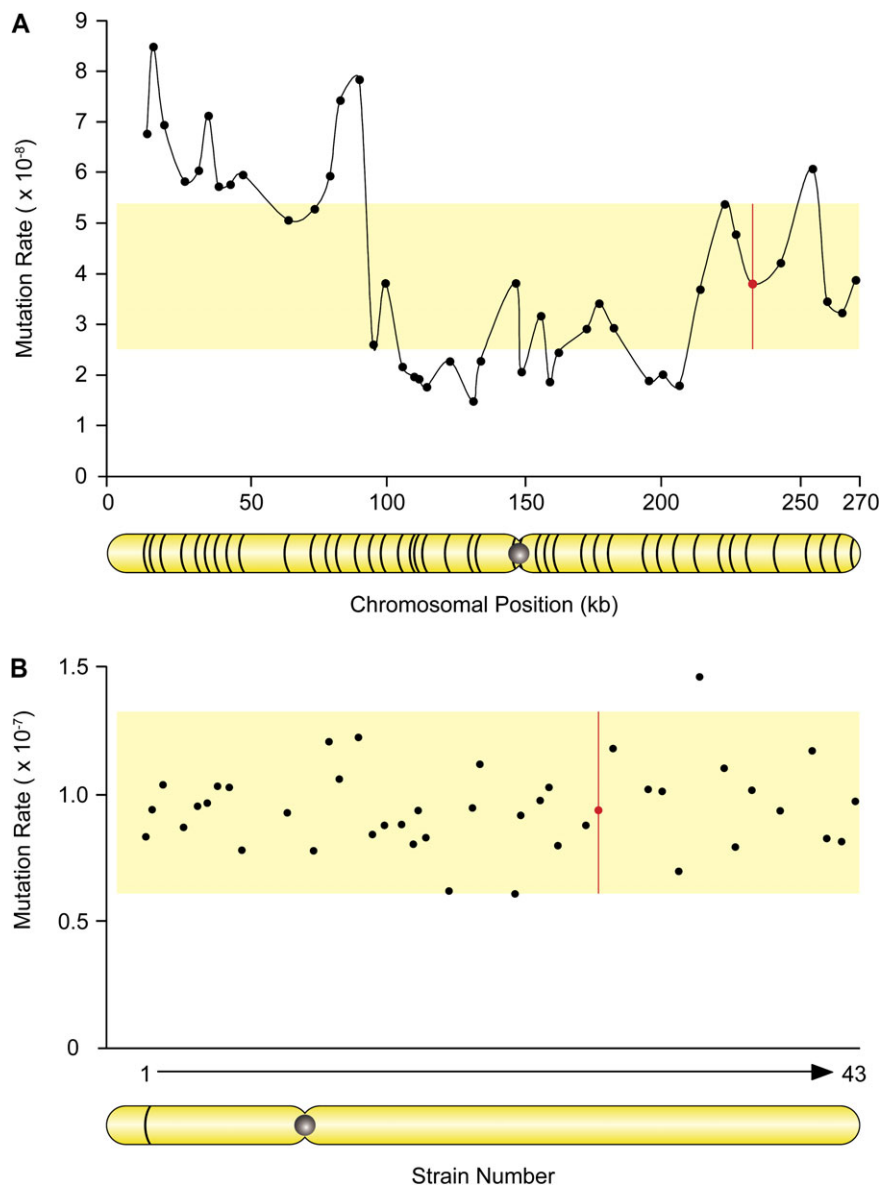
FIG. 2.—Mutation rate varies across Chromosome VI. (*A*) Mutation rate to 5FOA resistance from 43 strains (where *URA3* is integrated in a different location in each strain) reveals that mutation rate varies by an order of magnitude across Chromosome VI. The shaded region corresponds to the 95% confidence interval for the strain that had the median mutation rate (red data point). (*B*) There is less variation for mutation rates for canavanine resistance from the same 43 strains (where *CAN1* is in the same location in every strain). Only two of the 43 points lie outside of the 95% confidence interval for median mutation rate.

chromosome, a region of low mutation rate extending across the centromere, and a region of median mutation rate on the right arm of the chromosome.

Mutation Rate Is Correlated with Replication Timing

In order to determine the cause of mutation rate variation across Chromosome VI, we sought to determine if mutation rate is correlated to any other features of the chromosome. One possibility, which must be ruled out is that this variation

is not position dependent but rather strain dependent and that we do not detect this variation in the *CAN1* reporter because it may be less sensitive to this variation than the *URA3* reporter. This situation could arise if, for instance, the *URA3* gene contained mutational hotspots, which were missing (or underrepresented) in *CAN1*, and this experiment was really detecting strain-to-strain variation for one particular type of mutation. This situation is unlikely because both *URA3* and *CAN1* are large targets for mutation and do not contain any significant mutational hotspots (Lang and Murray 2008); therefore, there is no expectation that one
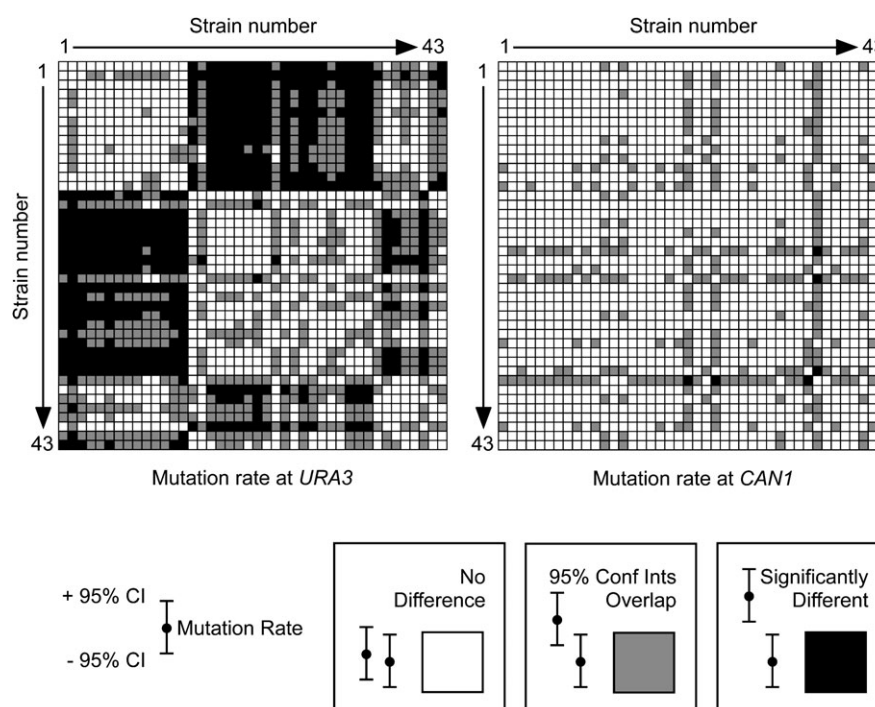
Fig. 3.—Pairwise comparisons of mutation rates at URA3 and CAN1 from the 43 strains used in this experiment. Ninety-five percent confidence intervals were generated for each point in figure 2. To determine if mutation rate varies significantly across Chromosome VI all pairwise comparisons of mutation rates from the 43 strains are shown for both URA3 and CAN1. The plots are symmetrical along the diagonal. For CAN1, there are only three significant differences in mutation rate, whereas for URA3, 262 of the 903 pairwise comparisons are significantly different. In particular, mutation rates cluster such that there are three regions of Chromosome VI with a relatively uniform mutation rate. These correspond to a region of high mutation rate along the left arm of Chromosome VI, a region of low mutation rate across the centromere, and a region of median mutation rate on the right arm (see fig. 2). The length of the regions in which the mutation rate is relatively constant is 50–100 kb.

of the two genes would be more sensitive to variation. If such a mechanism were acting in this experiment one would expect that this strain-to-strain variation would act in the same direction for both reporters, although the magnitude of the responses would be different. In other words, one would expect the mutation rates at CAN1 and URA3 to be correlated. We used two statistical tests to look for correlations: the Spearman rank correlation gives a probability that the rank order of two variables is correlated and the square of the Pearson correlation coefficient (R) measures the extent of the variance of one parameter (e.g., the mutation rate at CAN1) that can be explained by variation in the other parameter (in this case, the mutation rate at URA3). We find no correlation between mutation rates in the two reporters (fig. 4A, $P = 0.07$, Spearman rank correlation, $R^2 = 0.07$, Pearson correlation coefficient); therefore, the mutation rate variation at the URA3 gene in these strains is likely due to their position on Chromosome VI.

To look for features of the chromosome that are correlated with mutation rate, one should look for properties of the genome that vary on a similar length scale (50–100 kb). GC content is one such feature (Sharp and Lloyd 1993; Murakami et al. 1995). The average GC content for the 500 bp upstream and downstream of each gene does not correlate with its mu-

tation rate (fig. 4B, $P = 0.74$, Spearman rank test, $R^2 < 0.01$, Pearson correlation). We also looked for a correlation between the mutation rate and the production of double-strand breaks during meiosis. Gerton et al. (2000) measured binding of Spo11 during meiosis as a proxy the rate of production of double-strand breaks. It is possible that the same features that stimulate meiotic double-strand breaks also influence the mitotic mutation rate. We find a weak negative correlation between the production of double-strand breaks and the mutation rates on Chromosome VI (fig. 4C, $P = 0.02$, Spearman rank test, $R^2 = 0.09$, Pearson correlation). Another feature of the chromosome, which varies on a length scale of approximately 50–100 kb is replication timing. In yeast, replication of the genome is performed in a spatially and temporarily coordinated fashion, which is largely reproducible from cell cycle to cell cycle. The complete replication profile of the yeast genome has been determined (Raghuraman et al. 2001). There is a strong correlation between the time at which a region of the chromosome is replicated and its mutation rate (fig. 4D, $P < 10^{-4}$, Spearman rank test, $R^2 = 0.54$, Pearson correlation). This correlation is such that early-replicating regions have a low mutation rate and late-replicating regions have a high mutation rate. Repeating these calculations with a more recent data set for replication
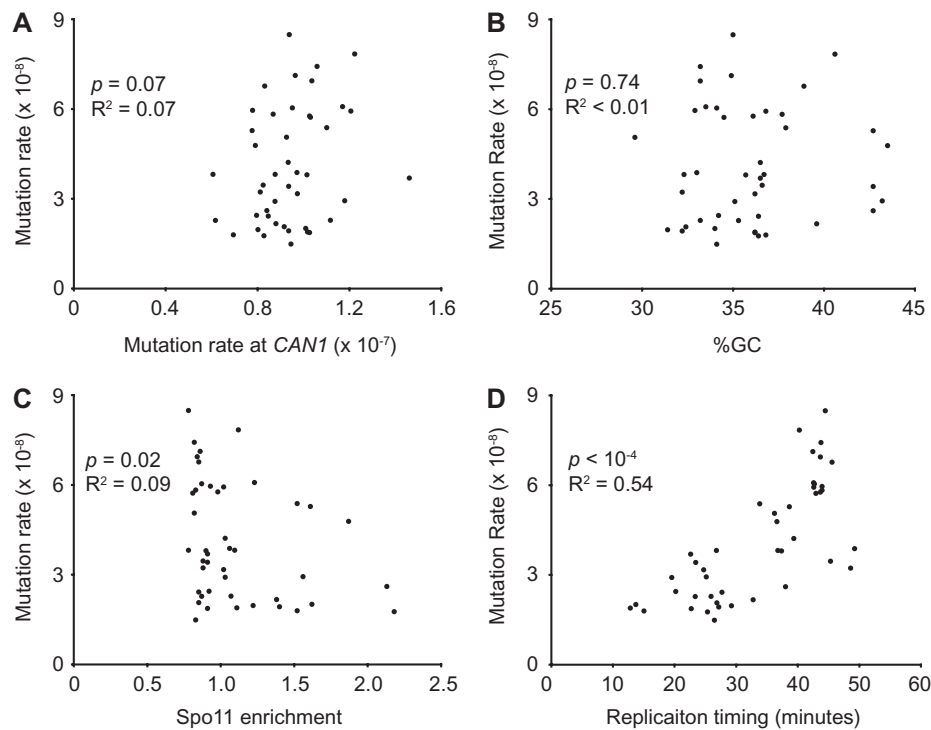
Fig. 4.—Mutation rate is correlated with replication timing. (*A*) To test the possibility that the observed mutation rate variation is strain-to-strain variation for which the *URA3* reporter is more sensitive, rates at *URA3* and *CAN1* were compared. They show no significant correlation. (*B*–*D*) To determine if mutation rate is correlated with a known property of the chromosome, mutation rate across Chromosome VI was compared with GC content, the production of double-strand breaks during meiosis, and replication timing. (*B*) Mutation rate is not correlated with the average GC content 500 bp upstream and downstream of the integrated *URA3*. (*C*) Mutation rate shows a weak negative correlation to the production of double-strand breaks during meiosis. Meiotic double-strand break data are from Gerton et al. (2000) and were estimated from a global analysis of Spo11 binding during meiosis. (*D*) Mutation rate is strongly correlated with replication timing. Replication timing is from Raghuraman et al. (2001). The correlation between mutation rate and recombination rate can be accounted for by a weak negative correlation between replication timing and recombination rate (*P* = 0.02). *P* values were determined using the Spearman rank correlation test. Data are available in supplementary table S2.

timing (Sekedat et al. 2010) gives a similar correlation between replication timing and mutation rate ($P < 10^{-4}$, Spearman rank test, $R^2 = 0.51$, Pearson correlation).

To determine if this mutation rate variation influences the pattern of synonymous substitutions, we calculated K_s (the number of synonymous substitutions per synonymous site) between S288c and two other *S. cerevisiae* strains (RM11-1a and YJM789) for each of the loci at which we measured mutation rate. We find that K_s within *S. cerevisiae* is correlated to our mutation rate estimates at these loci (fig. 5*A*, $P = 0.02$, Spearman rank test, $R^2 = 0.13$, Pearson correlation). For these same loci, however, we fail to find a correlation between mutation rate and K_s for *S. cerevisiae* and it is closest relative, *S. paradoxus* (fig. 5*B*, $P = 0.54$, Spearman rank test, $R^2 = 0.05$, Pearson correlation).

Increasing Mutation Rate by Manipulating Replication Timing

Figure 6*A* shows a comparison of the replication profile and the mutation profile of Chromosome VI. Chromosome VI contains 12 autonomous replicating sequences (*ARS*s) capa-

ble of initiating replication, each identified by the presence of a conserved *ARS* consensus sequence and by their ability to act as a replication origin on a plasmid (fig. 6*B*). Although Chromosome VI contains 12 *ARS* sequences, there are only seven prominent origins of replication (origins that fire in more than one quarter of cell cycles). Origins are classified by two measures: their efficiency (the number of cell divisions where the origin fires) and their timing of firing during S-phase. Based upon their timing, origins are classified as either early or late. Although the times at which origins fire lie on a continuum, early and late origins are distinct in terms of the proteins associated with preorigin complex and the genetic requirements for firing (Santocanale and Diffley 1998).

When designing this experiment, we did not anticipate that mutation rate would be correlated with replication timing, and because the strains were constructed such that *URA3* was integrated in place of an ORF, by chance three of these ORF deletions remove known yeast origins. *ARS605*, *ARS606*, and *ARS608* are deleted in strains GL-25, GL-31, and GL-39, respectively. Disruption of
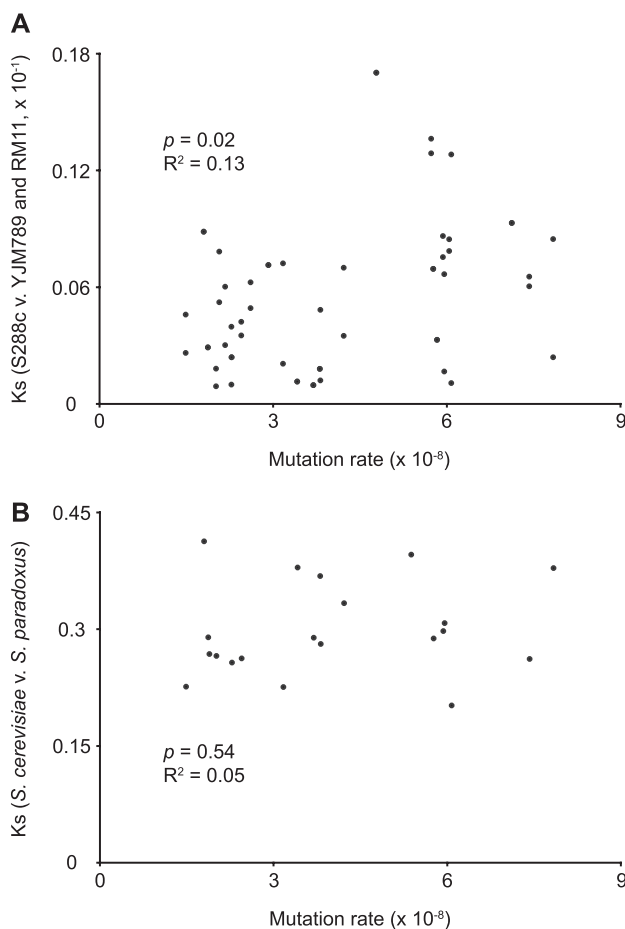
Fig. 5.—The number of synonymous substitutions per synonymous site, K_s, within *S. cerevisiae*, but not between *S. cerevisiae* and *S. paradoxus*, is correlated with mutation rate. The sequences of RM11-1a and YJM789 were obtained from the Broad Institute Fungal Genome Initiative (http://www.broad.mit.edu/annotation/fungi/fgi/) and the Stanford Genome Technology Center (version 2, http://med.stanford.edu/sgtc/research/yjm789.html), respectively. ORFs where S288c contains the allele one of the strains (RM11-1a or YJM789) were excluded from the analysis. K_s between *S. cerevisiae* and *S. paradoxus* were obtained from Kellis et al. (2003). *P* values were determined by permutation. Data are available in supplementary table S2.

ARS605 should have a small effect due to its close proximity to earlier firing *ARS603.5*. In addition, disruption of *ARS608* should have a negligible effect because it fires in only 10% of cell cycles. Disruption of *ARS606*, however, should affect the timing of replication because it is an early and efficient origin. Strain GL·31 was not used in the analysis because it affected growth on 5FOA (see Materials and Methods); however, interestingly, this strain had a high mutation rate (6.5×10^{-8}) compared with other *URA3* reporters in the same region, which may be partly attributable to disruption of *ARS606*.

To test if disruption of an origin of replication can increase the local mutation rate in an early-replicating/low mutation

rate region, the earliest and most efficient origin, *ARS607*, was deleted in strain GL·36, where the *URA3* gene is located 3 kb away from the origin. Deletion of *ARS607* increased the mutation rate at *URA3* by 30% (from 2.21×10^{-7} to 2.88×10^{-7}) without increasing the mutation rate at *CAN1* (0.81×10^{-7} in GL·36 and 0.76×10^{-7} in GL·36$^{ARS607\Delta}$). This slight increase in mutation rate is not significant given the error in fluctuation assays. It is possible that deletion of *ARS607* did not significantly delay replication timing in the region. The early but inefficient *ARS608* is 17 kb away. In the absence of *ARS607*, *ARS608* may fire in more cell cycles and allow for early replication of this region.

Mutation Rate Variation Is Dependent on Translesion Synthesis

We hypothesized that error-prone DNA synthesis could account for the higher mutation rate in late replication regions of chromosome VI. Cells can replicate past DNA lesions that block elongation by the normal replicative DNA polymerases (Polδ and ε) by two mechanisms: template switching and translesion synthesis. In contrast to the replicative polymerases, translesion polymerases have low processivity, high error rate, relaxed substrate specificity, and are employed to replicate damaged DNA templates (Friedberg et al. 2005). Rev1, which is both a translesion polymerase and helps to recruit other translesion polymerases, is not expressed until late S-phase (Waters and Walker 2006). Thus, the initial attempts to replicate past DNA lesions that occur early in S-phase must rely on template switching, which is not mutagenic, whereas attempts late in S-phase can rely on template switching and the mutagenic process of translesion synthesis. If this idea is correct, the increased mutation rate of late-replicating regions should depend on translesion synthesis, and eliminating translesion synthesis should reduce mutation in these regions: damaged DNA that would have been replicated by translesion polymerases (and would have given rise to mutations) remains single stranded, resulting in lethality.

To test this prediction, we investigated the effect of removing *REV1* on mutation rates in early- and late-replicating regions of Chromosome VI. We deleted the *REV1* gene from four strains (two early-replicating/low mutation rate and two late-replicating/elevated mutation rate). Strains GL·3, GL·15, GL·24, and GL·37 are replicated at 44.5, 43.8, 26.5, and 13.7 min, respectively. One of the two late-replicating regions has a very high mutation rate and the other is more similar to the mutation rate in early-replicating regions. Disruption of translesion synthesis results in a 4.8-fold reduction in the mutation rate at the late-replicating locus with the high mutation rate; for the late-replicating region with the lower mutation rate and the early-replicating regions with low mutation rates, there is no significant effect of *REV1* deletion (fig. 7).
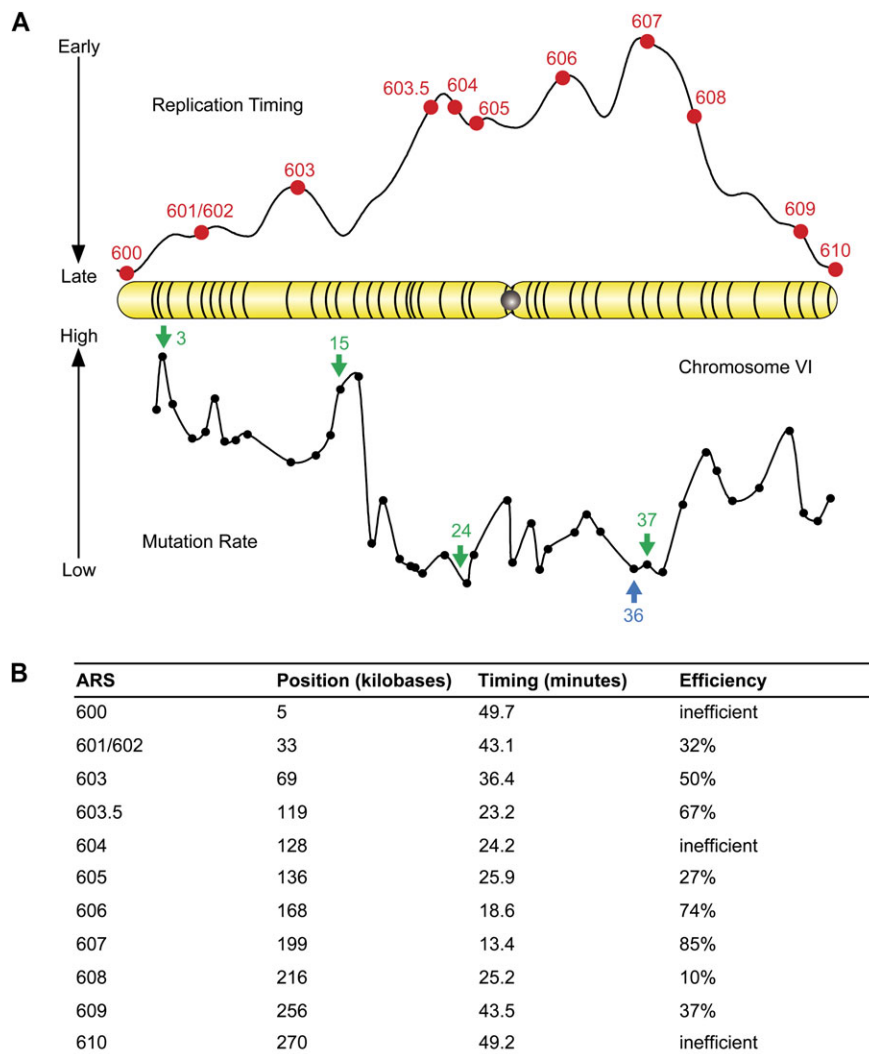
Fig. 6.—(A) Comparison of the replication profile and the mutation rate profile of Chromosome VI. Replication profile from Raghuraman et al. (2001). Both axes are linear and the range of mutation rates and replication times is the same as in figure 4D. A blue arrow below the mutation rate plot indicates the location selected to test mutation rate before and after ARS607 disruption. A green arrow above the mutation rate plot indicates the locations selected to test mutation rate following the deletion of *rev1* (see fig. 7) (B) Autonomously replicating sequences on Chromosome VI. Positions and efficiencies from the Saccharomyces Genome Database (http://www.yeastgenome.org). Position is the distance from the left telomere and efficiency is the fraction of cell divisions in which the origin fires. Replication timing from Raghuraman et al. (2001). ARS601 and ARS602 overlap and comprise one origin; ARS600 and ARS610 are subtelomeric.

Discussion

Model for Replication Timing and Mutation Rate

We have shown that the mutation rate varies across yeast Chromosome VI and that earlier replicating regions have a lower mutation rate. This correlation between replication timing and mutation rate can be understood in terms of a model for how cells deal with damaged bases during replication (Waters and Walker 2006). The genome is subject to numerous types of DNA damage including alkylation, ionizing radiation, UV radiation, and oxidative damage, resulting in a variety of damaged bases (Friedberg et al. 2005). Prior to S-phase, damaged bases are corrected by base excision

repair and nucleotide excision repair; however, some damaged bases escape repair and interfere with DNA replication. The replicative DNA polymerases (Polδ and Polε in yeast) have a high processivity and a low error rate; however, they are unable to replicate past some types of damaged bases (Garg and Burgers 2005). Therefore, when a replication fork encounters a lesion, the leading and lagging strands decouple and replication resumes downstream of the lesion (Lopes et al. 2006). The result is a single-stranded region (including the damaged base) behind the replication fork, known as a daughter-strand gap. There are two ways a cell can fill in this gap: an error-prone method using a translesion polymerase to copy the damaged template or an error-free
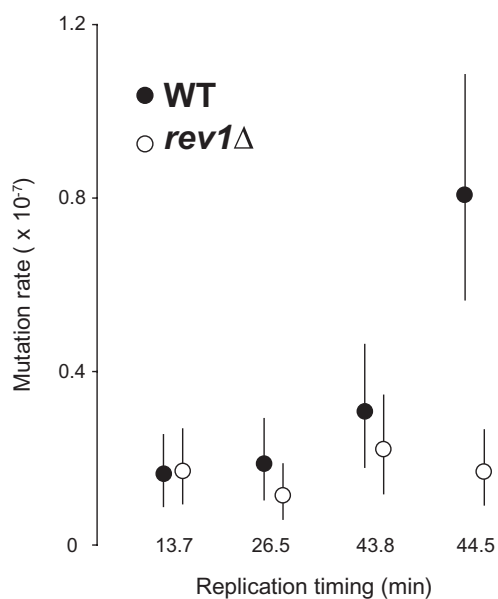
Fig. 7.—The varying mutation rate across Chromosome VI depends on translesion synthesis. To test the hypothesis that the correlation between replication time and mutation rate is due to the temporal separation of template switching and translesion synthesis, *REV1* (which encodes a translesion polymerase essential for translesion synthesis) was deleted in four strains that show variation in mutation rate and replication timing. Strains GL·3, GL·15, GL·24, and GL·37 (green arrows in fig. 6A) are replicated at 44.5, 43.8, 26.5, and 13.7 min, respectively. Disruption of translesion synthesis results in a strong reduction in the mutation rate in the late-replicating/high mutation rate region. For early-replicating regions with low mutation rates, there is no significant effect of *REV1* deletion.
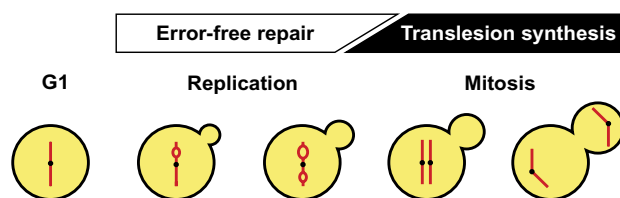

Fig. 8.—A model for the temporal separation of template switching and translesion synthesis. Damaged bases encountered by the replicative polymerase during S-phase result in single-strand gaps behind the replication fork. There are two ways a cell can fill in these gaps: a recombination-based approach (such as template switching) using the newly formed sister strand as a template or error-prone translesion synthesis. Template switching can occur as soon as the replication fork has passed and the sister sequence is available. Recent evidence suggests that translesion synthesis does not occur until the end of S-phase and into Mitosis (Waters and Walker 2006). Therefore, a damaged base in late-replicating regions is more likely to be subjected to translesion synthesis than the same lesion in an early-replicating region. This figure is adapted from Waters and Walker (2006).

method using the newly formed sister strand as a template (template switching). Error-free repair can occur as soon as the replication fork has passed and the homologous sequence is available. The work of Waters and Walker (2006) suggests that translesion synthesis is used only as a last-ditch effort to fill in these gaps and cannot occur until the end of S-phase (fig. 8). Therefore, regions of the genome that are replicated early in S-phase have longer to undergo error-free repair to replicate past lesions, whereas regions replicated late are more likely to require translesion synthesis.

It should be noted that the model of temporal separation of error-free repair and translesion synthesis is in contrast with an earlier model in which translesion synthesis occurs at the replication fork. The polymerase-switching model maintains that when a replicative polymerase encounters a lesion, the replication fork stalls leading to the dissociation of the replicative polymerase. A translesion synthesis polymerase could then replicate across the lesion, after which it dissociates, due to its low processivity, and the replicative polymerase can again take over. Although this model has not been disproven, recent evidence supports a model where translesion synthesis acts in late S-phase and not

at the replication fork. It has been observed that in an UV irradiation of an excision repair-deficient strain causes single-stranded regions to appear behind the replication fork (Lopes et al. 2006). The accumulation of single-stranded regions is increased in strains deficient in translesion synthesis, homologous recombination, or the DNA damage checkpoint (Lopes et al. 2006). Preventing translesion synthesis or inactivating the checkpoint only increases single-stranded regions late in S-phase, whereas loss of homologous recombination increases single-stranded regions throughout S-phase (Lopes et al. 2006). To test the model that translesion synthesis only occurs late in S-phase, expression levels of the three yeast translesion DNA polymerases were monitored during cell cycle progression (Waters and Walker 2006). Interestingly, Rev1, a translesion DNA polymerase essential for translesion synthesis, is not expressed until late in S-phase and into mitosis, after most of the DNA has been replicated (Waters and Walker 2006). These results support the model that translesion synthesis is used as a last resort to repair daughter-strand gaps in the genome. This model, in turn, provides an explanation for the observation that early-replicating regions have a low mutation rate and late replication regions have a high mutation rate: Damaged bases in late-replicating regions are more likely to be subjected to mutagenic translesion synthesis than similar lesions in early-replicating regions. In support of this model, we show that deleting the translesion polymerase *REV1* lowers the mutation specifically in late-replicating/high mutation rate regions.

Mutation Rate Variation on Multiple Scales

The correlation between replication timing and mutation rate in this work raises the question why this relationship was not identified in previous experimental studies. Two

earlier experiments showed that mutation rate varies across the genome for ochre suppressor mutations and frameshifts at microsatellite repeats. In the latter experiment, the 16-fold difference in mutation rates in a wild-type strain is reduced to 2-fold in an $msh2\Delta$ strain, indicating that the observed variation is due to differential ability of mismatch repair across the genome (Hawk et al. 2005). The variation in mutation rate for the tRNA suppressor mutations can also be explained as variation in the effectiveness of mismatch repair. Further analysis of the data suggests that much of the observed variation can be attributed to the orientation of the tRNA gene relative to the nearest origin of replication. The three tRNAs with the lowest mutation frequencies are transcribed in the direction of fork progression, whereas the other five tRNAs are transcribed in the opposite direction (Ito-Harashima et al. 2002). Ochre suppressors arise by a GC to TA transversion in the anticodon of tRNA-Tyr. Therefore, this could be either by the incorporation of an adenine opposite guanine on one strand or by the incorporation of a thymine opposite cytosine on the opposite strand. A common type oxidative DNA damage is 8-oxo-guanine, which can pair with adenine causing a GC to TA transversion (Friedberg et al. 2005). Mismatch repair is more efficient at correcting 8-oxo-guanine-adenine base pairs on the lagging strand than the leading strand, possibly due to the presence of more nicks on the lagging strand (Pavlov et al. 2003). The tRNA-Tyr alleles with low mutation rates to ochre suppressors are oriented such that adenine incorporation opposite 8-oxo-guanine will occur on the lagging strand, whereas for the tRNA-Tyr alleles with high mutation rates this will occur on leading strand and have a greater potential of escaping mismatch repair.

This result shows that orientation with respect to the replication fork can have an impact on mutation rate for a single base-pair substitution; however, this is unlikely to impact mutation rates in our experiment because we are detecting loss-of-function mutations over an entire gene, which will average out these small-scale effects. Classifying the strains based upon the orientation of URA3 with respect to the most likely direction of fork movement does not reveal an orientation bias in our results ($P > 0.05$, Wilcoxon rank-sum). Additionally, orientation relative to the replication fork is not responsible for variation of mutation rate observed for microsatellite frameshift mutations (Hawk et al. 2005).

Different genes detect different mechanisms that cause mutation rates to vary across the genome. Variation in the rate of frameshift mutations is largely due to variation in the efficiency of mismatch repair across the genome, although the genomic feature responsible for this variation is unknown. Variation in the rate of tRNA-Tyr ochre suppressor mutations is associated with the orientation of the gene with respect to the nearest replication origin and may result from differential efficiencies of mismatch repair on the leading and lagging strands. In the experiment described

here, mutation rate variation is shown to correlate with replication timing and we argue that it results from the temporal separation of error-free repair (template switching) and translesion synthesis. Therefore, the replication profile can impact mutation rate in two ways, by determining the direction of replication fork movement and the timing of replication. Although the mechanism for variation in microsatellite mutations is unknown, neither replication timing nor orientation can account for it, suggesting that other aspects of genome structure can influence the mutation rate.

Evolutionary Consequences of Mutation Rate Variation

Spatial clustering of mutation rates is likely to have significant evolutionary consequences in shaping patterns of synonymous substitutions and the location of essential genes. Synonymous substitutions are largely unaffected by selection; therefore, the number of synonymous substitutions per synonymous site (K_s) provides a measure of the accumulation of neutral mutations. K_s between S. cerevisiae strains, but not between S. cerevisiae and S. paradoxus, is correlated with mutation rate (fig. 5). The lack of a correlation between mutation rate and K_s and between S. cerevisiae and S. paradoxus is consistent with previous work showing that the rate of synonymous substitutions between these species does not vary across the genome (Chin et al. 2005), although it does correlate with the strength of gene expression (Drummond and Wilke 2008). There are two possible explanations for the lack of correlation between replication timing and sequence divergence between the two related yeast species: at longer times, other features exert stronger control over which mutations can survive or replication timing may change rapidly on an evolutionary time scale. A survey of nine origins on Chromosome VI shows strain-to-strain variation in the efficiency of at least one origin within S. cerevisiae (Yamashita et al. 1997). Centromeres, however, are consistently early replicating, and in yeast, it has been observed that essential genes tend to be located near centromeres (Taxis et al. 2005). Taxis et al. (2005) suggest that linking essential genes to centromeres may mask recessive deleterious mutations by restoring heterozygosity during intraascus mating because the MAT locus itself is weakly centromere-linked. Alternatively, centromere-proximal positioning of essential genes may have been selected in order to keep essential genes in regions of low mutation rate.

In summary, we show that mutation rates vary within the yeast genome and correlate with replication timing such that early-replicating regions have a low mutation rate. We interpret this observation in terms of a model in which temporal separation between two types of DNA damage tolerance: recombination-based template switching and mutagenic translesion synthesis. A correlation between replication timing and synonymous substitution has been

demonstrated for phylogenically diverse organisms: *Escherichia coli* (Sharp et al. 1989), humans (Stamatoyannopoulos et al. 2009), and the Archaeon *Sulfolobus islandicus* (Flynn et al. 2010) raising the possibility that the mechanisms underlying mutation rate variation are highly conserved.

Supplementary Material

Supplementary tables S1 and S2 and figures S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Chin CS, Chuang JH, Li H. 2005. Genome-wide regulatory complexity in yeast promoters: separation of functionally conserved and neutral sequence. Genome Res. 15:205–213.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134:341–352.

Flynn KM, Vohr SH, Hatcher PJ, Cooper VS. 2010. Evolutionary rates and gene dispensability associate with replication timing in the archaeon Sulfolobus islandicus. Genome Biol Evol. 2:859–869.

Friedberg EC, et al. 2005. DNA repair and mutagenesis. Washington (DC): ASM Press.

Garg P, Burgers PM. 2005. DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol. 40:115–128.

Gerton JL, et al. 2000. Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 97:11383–11390.

Hawk JD, et al. 2005. Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. Proc Natl Acad Sci U S A. 102:8639–8643.

Huang ME, Rio AG, Nicolas A, Kolodner RD. 2003. A genomewide screen in Saccharomyces cerevisiae for genes that suppress the accumulation of mutations. Proc Natl Acad Sci U S A. 100:11529–11534.

Ito-Harashima S, Hartzog PE, Sinha H, McCusker JH. 2002. The tRNA-Tyr gene family of Saccharomyces cerevisiae: agents of phenotypic variation and position effects on mutation frequency. Genetics 161:1395–1410.

Kellis M, et al. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254.

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

Lang GI, Murray AW. 2008. Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. Genetics 178:67–82.

Lehner KR, Stone MM, Farber RA, Petes TD. 2007. Ninety-six haploid yeast strains with individual disruptions of open reading frames between YOR097C and YOR192C, constructed for the Saccharomyces genome deletion project, have an additional mutation in the mismatch repair gene MSH3. Genetics 177:1951–1953.

Lopes M, Foiani M, Sogo JM. 2006. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. Mol Cell. 21:15–27.

Luria S, Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511.

Murakami Y, et al. 1995. Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae. Nat Genet. 10:261–268.

Pavlov YI, Mian IM, Kunkel TA. 2003. Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. Curr Biol. 13:744–748.

Raghuraman MK, et al. 2001. Replication dynamics of the yeast genome. Science 294:115–121.

Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations. Methods 20:4–17.

Santocanale C, Diffley JF. 1998. A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. Nature 395:615–618.

Sarkar S, Ma WT, Sandri GH. 1992. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. Genetica 85:173–179.

Sekedat MD, et al. 2010. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. Mol Syst Biol. 6:353.

Sharp PM, Lloyd AT. 1993. Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. Nucleic Acids Res. 21:179–183.

Sharp PM, Shields DC, Wolfe KH, Li WH. 1989. Chromosomal location and evolutionary rate variation in enterobacterial genes. Science 246:808–810.

Stamatoyannopoulos JA, et al. 2009. Human mutation rate associated with DNA replication timing. Nat Genet. 41:393–395.

Taxis C, et al. 2005. Spore number control and breeding in Saccharomyces cerevisiae: a key role for a self-organizing system. J Cell Biol. 171:627–640.

Waters LS, Walker GC. 2006. The critical mutagenic translesion DNA polymerase Rev1 is highly expressed during G(2)/M phase rather than S phase. Proc Natl Acad Sci U S A. 103:8971–8976.

Winzeler EA, et al. 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285:901–906.

Yamashita M, et al. 1997. The efficiency and timing of initiation of replication of multiple replicons of Saccharomyces cerevisiae chromosome VI. Genes Cells 2:655–665.

Associate editor: Esther Betran