GENETICS


10.1126/sciadv.1501033

# Mutation at a distance caused by homopolymeric guanine repeats in *Saccharomyces cerevisiae*

**Michael J. McDonald,[1]* Yen-Hsin Yu,[1] Jheng-Fen Guo,[1] Shin Yen Chong,[2,3] Cheng-Fu Kao,[2] Jun-Yi Leu[1]***

**Mutation provides the raw material from which natural selection shapes adaptations. The rate at which new mutations arise is therefore a key factor that determines the tempo and mode of evolution. However, an accurate assessment of the mutation rate of a given organism is difficult because mutation rate varies on a fine scale within a genome. A central challenge of evolutionary genetics is to determine the underlying causes of this variation. In earlier work, we had shown that repeat sequences not only are prone to a high rate of expansion and contraction but also can cause an increase in mutation rate (on the order of kilobases) of the sequence surrounding the repeat. We perform experiments that show that simple guanine repeats 13 bp (base pairs) in length or longer ($G_{13+}$) increase the substitution rate 4- to 18-fold in the downstream DNA sequence, and this correlates with DNA replication timing ($R = 0.89$). We show that $G_{13+}$ mutagenicity results from the interplay of both error-prone translesion synthesis and homologous recombination repair pathways. The mutagenic repeats that we study have the potential to be exploited for the artificial elevation of mutation rate in systems biology and synthetic biology applications.**

## INTRODUCTION

Mutation is the source of both adaptive genetic variation and deleterious genetic load. As such, the mutation rates that we observe in nature represent a balance between opposing selective forces. Because extant organisms are relatively well adapted, most mutations that occur will be deleterious (*1*). It is this aversion to mutation that has driven the evolution of complex systems for high-fidelity replication and maintenance of genome integrity. However, mutation rates vary greatly; RNA viruses have an error rate of $10^{-4}$ substitutions per nucleotide per cell infection (*2*), whereas bacteria make only a single error for every billion nucleotides synthesized (*3*). Although it is perhaps not surprising that widely diverged species exhibit large differences in mutation rate, this observation raises questions about the degree of mutation rate variation within and across genomes.

Studies of clinical and experimental populations have revealed that mutation rates can evolve on time scales observable to experimentalists and provided insights into the dynamics of mutation rate evolution (*4*). In microbes, mutator strains arise via mutations that inactivate DNA repair enzymes, resulting in elevated mutation rates. These strains can become established in adapting populations because of the higher rate at which they produce beneficial mutations (*4*). However, after the population has become better adapted, the increased rate of deleterious mutation experienced by a mutator will be selected against, providing selective pressure to drive mutation rates back down (*5*, *6*).

The effects of mutator alleles that diminish polymerase accuracy extend to the whole genome. However, mutation rates can also vary within a single genome. Highly transcribed genes have been found to have elevated mutation rates proportional to the rate of transcription (*7*, *8*), probably because DNA that is highly transcribed is more often in a single-stranded state and thus more vulnerable to mutagens (*9*, *10*). Another well-established correlate of mutation rate variation is DNA replication timing. Experimental (*11*) and bioinformatic (*12*, *13*) studies have shown that later replicated DNA has a higher mutation rate than the earliest replicated regions. A proposed explanation for this is that DNA that is copied later during cell division will have less time for slower, high-fidelity repair mechanisms and will rely on error-prone DNA repair (*14*).

Primary DNA sequence can also influence mutation rates. Homopolymeric repeats of nucleotides are prone to increase and decrease in length at a high frequency (*15*, *16*) and have been found to play an important role in genome evolution (*17*, *18*), especially genetic switching mechanisms in pathogenic bacteria (*19*, *20*). Microsatellites or other short repeats have been implicated in several human diseases (*21*), and the high degree of polymorphism that results from their instability has been exploited as genetic markers (*22*). In some cases, the causes of genetic instability have been directly linked to physical properties of DNA sequence. For instance, the twist angle between two adjacent bases in a DNA double helix is predetermined by their identity: some combinations of nucleotides have twist angles that are fragile and thus more prone to breakage (*23*). The repeated evolution of mutations in the promoter of *pitX1* in multiple independent populations of sticklebacks has been hypothesized to be due to both the strong selection upon these mutations and the proclivity of this region to sustain double-strand breaks. The part of the *pitX1* promoter in which these adaptive mutations occur is the most fragile site in the stickleback genome, as predicted by DNA twist angles (*24*).

The mutagenic effects of some DNA sequences have been found to extend to flanking regions of DNA, with stretches of DNA that have the capacity to form secondary structures often being the culprit. Tang and colleagues (*25*) found that runs of 230 Friedreich's ataxia repeats (GAA•TTC) were able to induce large deletions [>50 bp (base pairs)] and point mutations in a reporter gene more than a kilobase downstream of the repeat itself. Others working with the same repeat unit found that this repeat sequence could induce mutagenesis in sequences up to 8 kb away (*26*). Directly inducing double-strand breaks using HO (*27*) or I-SCE1 (*28*) sharply increases not only the mutation rate in the surrounding 2 kb of sequence but also has a detectable

[1]Institute of Molecular Biology, Academia Sinica, Taipei 11529, Taiwan. [2]Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan. [3]Graduate Institute of Food Science and Technology, National Taiwan University, Taipei 10617, Taiwan.
*Corresponding author. Email: michaeljmcdonald@fas.harvard.edu (M.J.M.); jleu@imb.sinica.edu.tw (J.-Y.L.)

but weak long-range mutator effect that decays exponentially across 60 kb of DNA sequence (*28*). In all of these cases, a combination of double-strand break and translesion repair was directly implicated (*25*–*28*).

In previous work, we found that short repeat sequences are positively correlated with the substitution rate in the surrounding DNA sequence (*29*), distinct from the well-known repeat length polymorphism associated with repetitive DNA sequences, and that the experimental insertion of repeat sequences in yeast could elevate mutation rates in the downstream sequence. We have proposed that repeat sequences are more likely than other sequences to recruit error-prone DNA repair polymerases, leading to an increased mutation rate in DNA sequence surrounding the repeat. Here, we test our hypothesis by investigating one particular type of repeat in more detail: homopolymeric runs of guanine nucleotides. We demonstrate which DNA replication repair pathways are necessary for mutagenesis and show that these sequences interact with other known causes of mutation rate variation.

## RESULTS

### Homopolymeric runs of guanines 13 bp or longer ($G_{13+}$) cause an increase in mutation rate

Using *Saccharomyces cerevisiae*, we engineered runs of 11 to 14 guanine nucleotides four bases upstream of the *URA3* coding region (Fig. 1A), and measured mutation rates (Fig. 2A). The mutation rate in the *URA3* coding region, downstream of a $G_{13}$ or $G_{14}$ repeat sequence, increased by up to

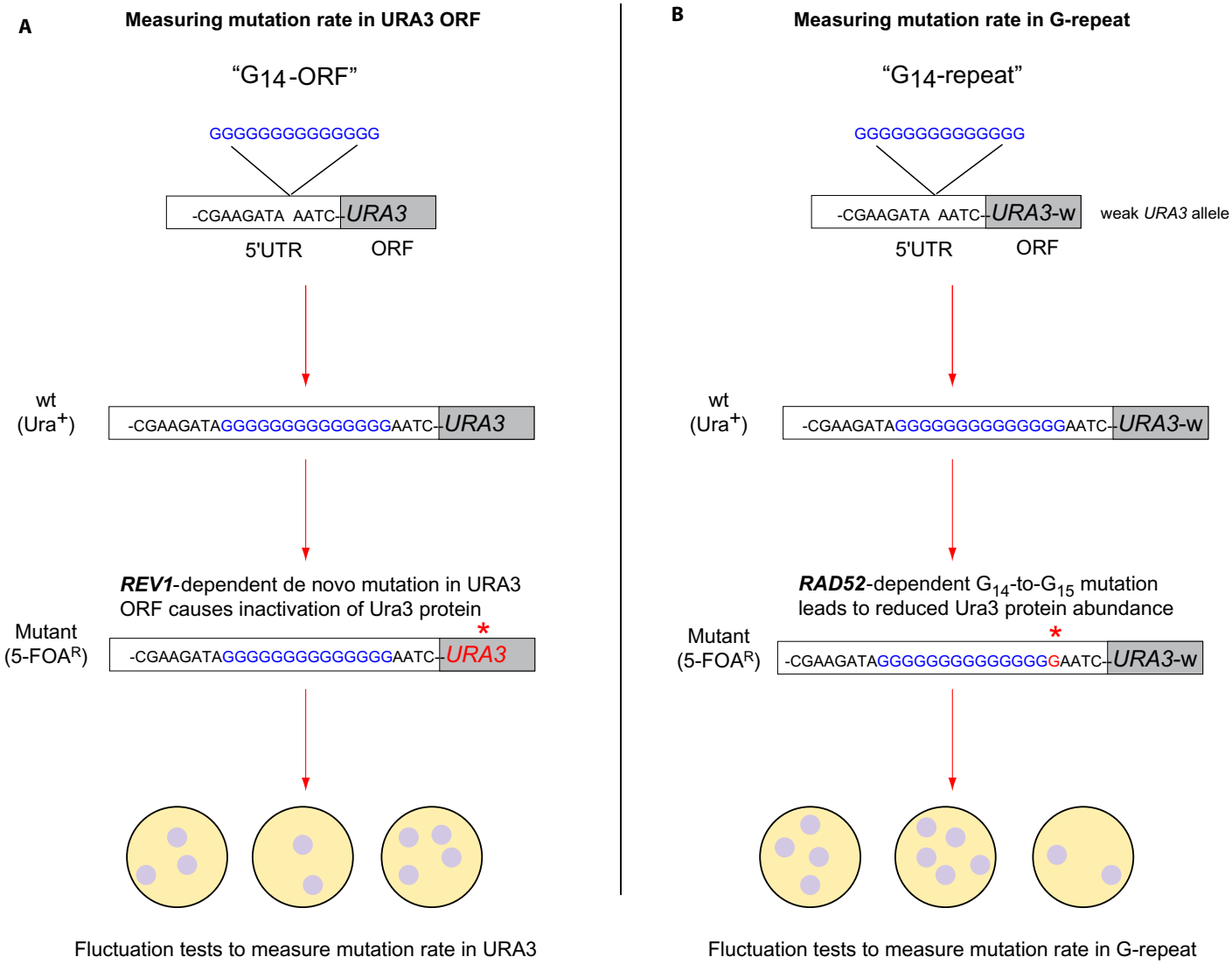


**Fig. 1. Experimental approach to quantifying mutagenicity of $G_{13+}$ DNA sequences.** (**A**) Poly-G sequences were engineered 4 bp upstream of the *URA3* translation start site (the 5′UTR region). The $G_{14}$-ORF (open reading frame) construct ($G_{14}$-ORF) allowed detection of loss-of-function mutations in the *URA3* reading frame by plating on medium containing 5-fluoroorotic acid (5-FOA), which selects for individual cells that contain mutations that inactivate *URA3*. The red asterisk indicates the mutation site. (**B**) Using a weak *URA3* allele (*URA3*-w), this construct ($G_{14}$-repeat) facilitated the detection of the polyguanine repeat expansion mutation, because the mutation, $G_{14}$ to $G_{15}$ or longer, results in the 5′-FOA–resistant phenotype. wt, wild type.

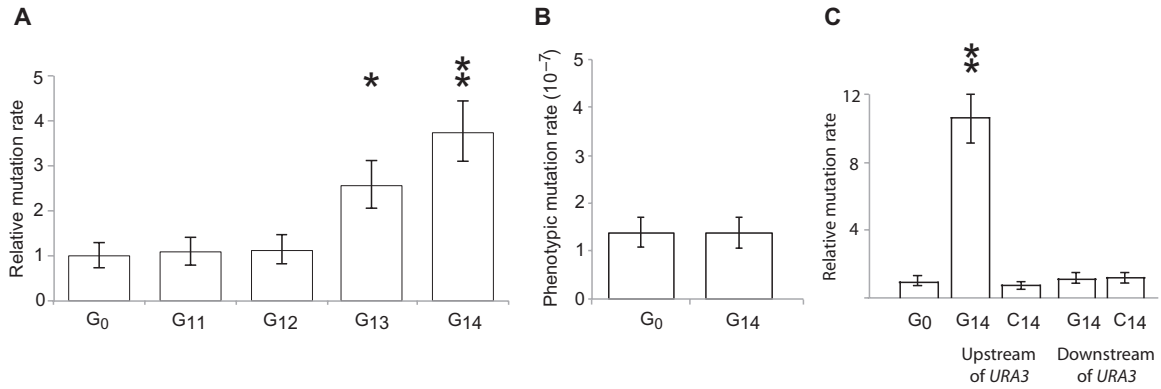


**Fig. 2. Polyguanine sequences cause a localized, directional effect on mutation rate.** (**A**) Mutation rates of homopolymeric guanine repeat sequences of increasing length. The estimated phenotypic mutation rate of $G_0$-ORF, $G_{13}$-ORF, and $G_{14}$-ORF is $5.4 \times 10^{-7}$, $13.5 \times 10^{-7}$, and $20.3 \times 10^{-7}$, respectively. $G_{11}$ and $G_{12}$ had no detectable increase in mutation rate. (**B**) Mutation rate was measured using *CAN1*, at a site distal from the *URA3* locus. (**C**) The $G_{14}$ repeat does not cause an increase in mutation rate if engineered on the template strand ($C_{14}$-*URA3*) or on either the coding strand (ORF-$G_{14}$) or the template strand (ORF-$C_{14}$) downstream of the *URA3* terminator sequence. Significant differences were calculated using *t* tests. *$P < 0.05$; **$P < 0.005$. Error bars show 95% confidence intervals.


**Fig. 3. Mutational spectrum of 183 5′-FOA–resistant mutants at the *URA3* locus.** 5′-FOA–resistant mutants were collected from $G_0$-*URA3* (brown) and $G_{14}$-ORF (blue) strains. Each point mutation is shown directly above the wild-type sequence, and indels and complex mutations are shown directly below.

fourfold. We measured the mutation rate at another locus encoding the *CAN1* gene (Fig. 2B), confirming that mutation rates obtained at a site that did not have a $G_{14}$ sequence upstream did not increase relative to the wild type. Moving $G_{14}$ from the coding strand to the template strand and from upstream of the *URA3* translation start site to downstream of the transcription termination site abolished the mutagenic effect of $G_{14}$ (Fig. 2C).

## Sequence analysis of $G_0$- and $G_{14}$-ORF mutants reveals similar spectrum of mutations

We sequenced 113 independent $G_{14}$-*ORF ura3* mutants and 101 $G_0$-*ura3* mutants and analyzed the distribution and identity of the mutations (Fig. 3). Although both sets of mutations were significantly different from a uniform distribution (Kolmogorov-Smirnov test, $G_{14}$, $P = 0.023$; $G_0$, $P = 0.002$), they were not different from each other (Kolmogorov-Smirnov test, $P = 0.99$).

The identities of the mutations were highly similar except for a slight enrichment of indels relative to substitutions (Fisher's exact test, $P = 0.048$) and transitions relative to transversions (Fisher's exact test, $P = 0.035$) in the set of $G_{14}$-*ORF ura3* mutants. Our sequence data incorporated the $G_{14}$ repeat as well as the entire *URA3* ORF, confirming that changes in guanine repeat length (repeat expansion or contraction) and large deletions were not responsible for any of the elevated mutation rate detected in this assay (table S1).

### DNA replication timing correlates with the degree of $G_{14}$ mutagenicity

To test whether genome position would influence $G_{14}$ mutagenicity, we engineered $G_0$-*URA3* and $G_{14}$-*ORF* genes into different positions on chromosomes XII and XV (Fig. 4 and table S2). We found that $G_{14}$-*URA3* inserts sustained a higher mutation rate than $G_0$-*URA3* inserts at the same site. The change in mutation rate associated with $G_{14}$ ranged from a 4-fold to an 18-fold increase, supporting the fact that $G_{14}$ mutagenicity occurs regardless of genome position.

The direction of DNA replication for the chromosome V version of the $G_{14}$-*ORF* construct indicates that the replication fork moves in the same direction as transcription for URA3. Because moving the $G_{14}$ sequence to the template strand or downstream of the transcription termination site abolishes the mutagenic effect, this suggests a relationship between $G_{14}$ mutagenicity and DNA replication timing. DNA replication timing correlates with mutation rate variation in organisms ranging from bacteria to humans (*11*, *12*). We calculated correlation coefficients for mutation rates of the $G_{14}$-*ORF* inserts and replication timing [from Nieduszynski *et al.* (*30*)] and found a significant positive correlation (Fig. 4A; Spearman's $R = 0.89$, $P = 0.007$). The $G_0$-*URA3* inserts in equivalent positions showed a nonsignificant positive correlation (Fig. 4B; Spearman's $R = 0.638$, $P = 0.086$) consistent with previous work (*11*).

### $G_{13}$ repeats and transcription promote the accumulation of replication fork intermediates

Repeat sequences are known to incur an increased risk of replication fork stalling (*31*). Upon fork stalling, replication reinitiates downstream, leaving a single-stranded gap that is filled in using either homologous recombination (HR) or translesion synthesis (TLS) (*32*–*34*), with a bias toward TLS for gaps requiring repair later in S phase (*14*, *34*). TLS, mediated by the Rev1/Polζ complex, often introduces errors, even when synthesizing undamaged DNA in vivo (*35*, *36*), and is responsible for approximately 50% of mutations in wild-type *S. cerevisiae* (*37*, *38*). Replication fork intermediates can be visualized using the two-dimensional (2D) gel technique (*39*). We made the doxycycline-repressible constructs *tet*-$G_{13}$-*URA3* and *tet*-$G_0$-*URA3* for 2D analysis (we were unable to generate a *tet*-$G_{14}$-*URA3* construct). These were designed so that when running the 2D gel, replication fork intermediates that stall at the beginning of *URA3* (Y molecules) should accumulate at the point indicated in Fig. 5B. We observed a 2.5-fold increase of Y molecules in $G_{13}$ constructs compared to $G_0$ constructs when cells were grown in the absence of doxycycline (Fig. 5, A to C).

Previous studies have shown that plasmid-based $G_{20}$ and $G_{32}$ repeats cause transcription-dependent replication fork stalling (*40*, *41*). When we repressed transcription by supplementing the growth media with doxycycline, we found that the enrichment of Y molecules in $G_{13}$ compared to $G_0$ was reduced compared to the treatment without doxycycline (Fig. 5C). The ratio of $G_{13}$ to $G_0$ decreased, whereas the amount of actual detected replication fork intermediates increased; possible explanations for this will be discussed below. The change in replication fork intermediates caused by repression of transcription raises the possibility that the mutagenic effect of $G_{13}$ is also changed. We measured the effect of transcriptional repression on mutation rates in *tet*-$G_{13}$-*URA3* and *tet*-$G_0$-*URA3*. The mutation rates of *tet*-$G_{13}$-*URA3* and *tet*-$G_0$-*URA3*

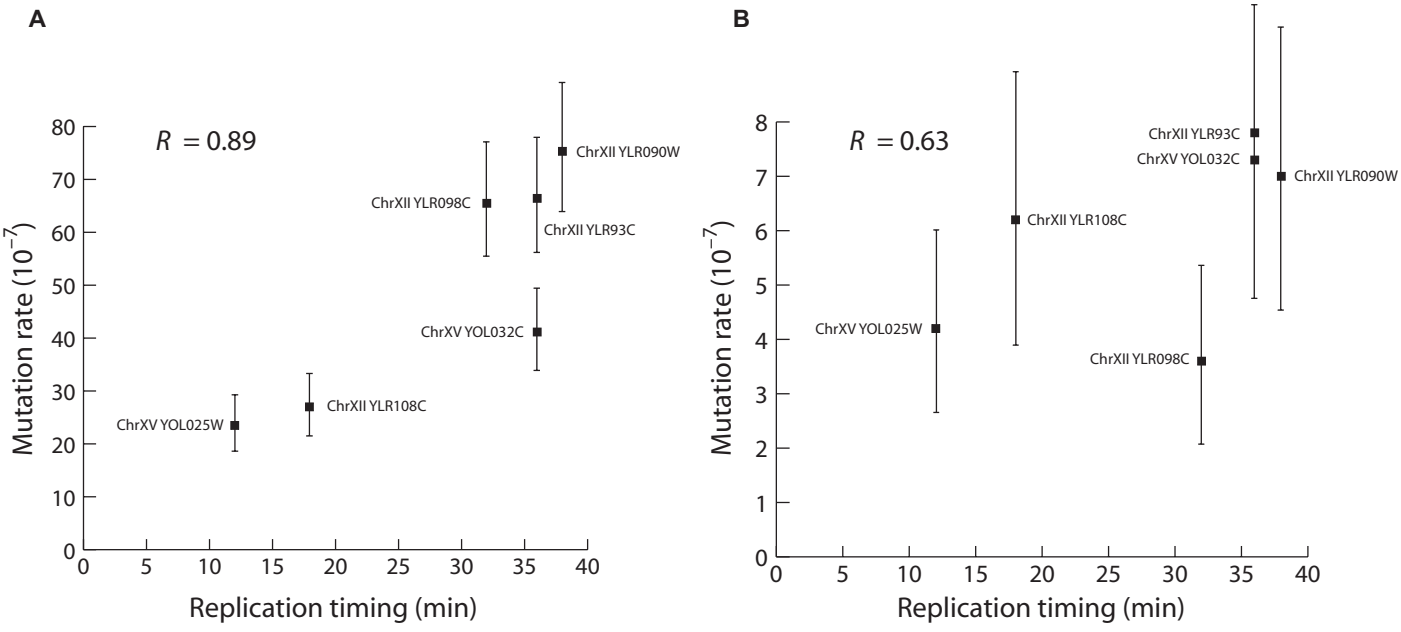


**Fig. 4. The effect of $G_{14}$ mutagenicity is correlated with DNA replication timing.** (**A** and **B**) The regression of mutation rate and replication timing for $G_{14}$-*ORF* alleles inserted at six different sites on chromosome XII and two sites on chromosome XV (Spearman's $R = 0.89$, $P = 0.007$) (A) and for insertion of $G_0$-*URA3* at the same sites (Spearman's $R = 0.63$, $P = 0.086$) (B). Replication timing is shown in minutes after the release of cells into synchronized S phase, as reported by Nieduszynski *et al.* (*30*). Error bars show 95% confidence intervals.

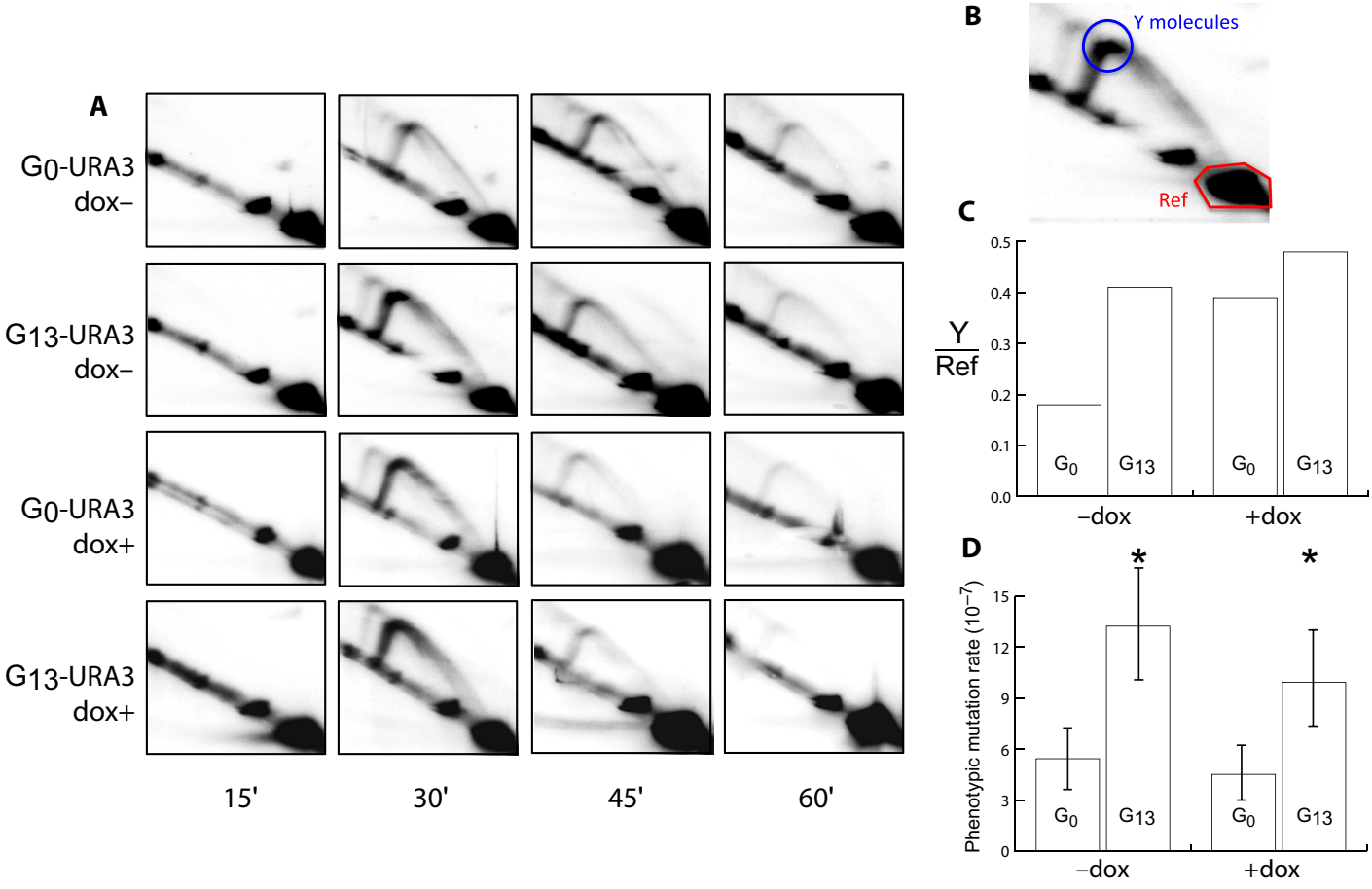


**Fig. 5. $G_{13}$ repeats are implicated in the accumulation of replication fork intermediates.** (**A**) 2D gels for tet-$G_0$-URA3 and tet-$G_{13}$-URA3 under the control of a repressible tetR promoter (*65*) for time points taken at 15, 30, 45, and 60 min after synchronized cells were released from $G_1$ cell cycle arrest. (**B**) The circled area show replication fork intermediates that are putatively stalled at the beginning of URA3 (Y molecules, blue circle). The reference region (red hexagon) is used to normalize between samples and determine the relative amounts of Y molecules in each treatment. (**C**) When transcription is enabled (−dox), the $G_{13}$ construct has ~2.5 as much putative fork stalling as the $G_0$ construct (calculated at the 30-min time point). When transcription is repressed, both $G_{13}$ and $G_0$ constructs appeared to have more Y molecules, although $G_{13}$ has only ~1.13 as much replication fork intermediates as $G_0$ (calculated at the 30-min time point). (**D**) Mutation rates for $G_0$-URA3 and $G_{13}$-ORF under the control of a repressible tetR promoter. −dox indicates the high-expression treatment, whereas +dox indicates repression of transcription. Significant differences were calculated using *t* tests; *$P < 0.05$. Error bars show 95% confidence intervals.

were reduced when *URA3* transcription was repressed, although not significantly. Even after repression of transcription by doxycycline, the mutation rate of *tet-$G_{13}$-URA3* remained significantly higher than that of *tet-$G_0$-URA3* (Fig. 5D).

### Mutagenesis downstream of $G_{14}$ repeats is Rev1-dependent

Previously, we had proposed that repeat sequence–mediated increases in downstream mutation rate were caused by frequent recruitment of error-prone translesion DNA polymerases by sequences prone to stall the high-fidelity, housekeeping DNA polymerase (*29*). To test this, we deleted *REV1*, which is required for all TLS in yeast (*42*, *43*). We found that ablation of *REV1* significantly reduced the mutation rate of both *$G_{14}$-ORF* (*t* test, $P < 0.005$) and *$G_0$-URA3* (*t* test, $P < 0.005$) (Fig. 6A). That deletion of *REV1* decreases mutation rate suggests that the replication fork interruptions that are typically accommodated by Rev1-mediated TLS DNA synthesis are lethal in the *rev1* mutant (*44*). If this is the case, $G_{13+}$ sequences would cause an increased likelihood of mutation in the surrounding DNA sequence by increasing the rate of error-prone translesion DNA synthesis in that region.

### Expansion of homopolymeric repeats also occurs at the $G_{14}$ repeat and is Rad52-dependent

It has long been established that homopolymeric repeat sequences are unstable, increasing and decreasing in repeat length at a high rate (*19*, *20*). In the experiment described above, only mutations that occur in the ORF of *URA3* can be recovered, even though mutations that change repeat length almost certainly occur in the $G_{14}$ sequences of some of the individuals within the populations of yeast cells used to measure mutation rate. This is because mutations changing the length of the $G_{14}$ repeat, which is in the 5′ untranslated region (5′UTR), do not cause the loss of *URA3* function that is selected in the assay.

To facilitate the capture of mutations that change the number of *G*'s in the $G_{14}$ repeat, we constructed a new strain containing the $G_{14}$ sequence, engineered upstream of an alternative *URA3* sequence

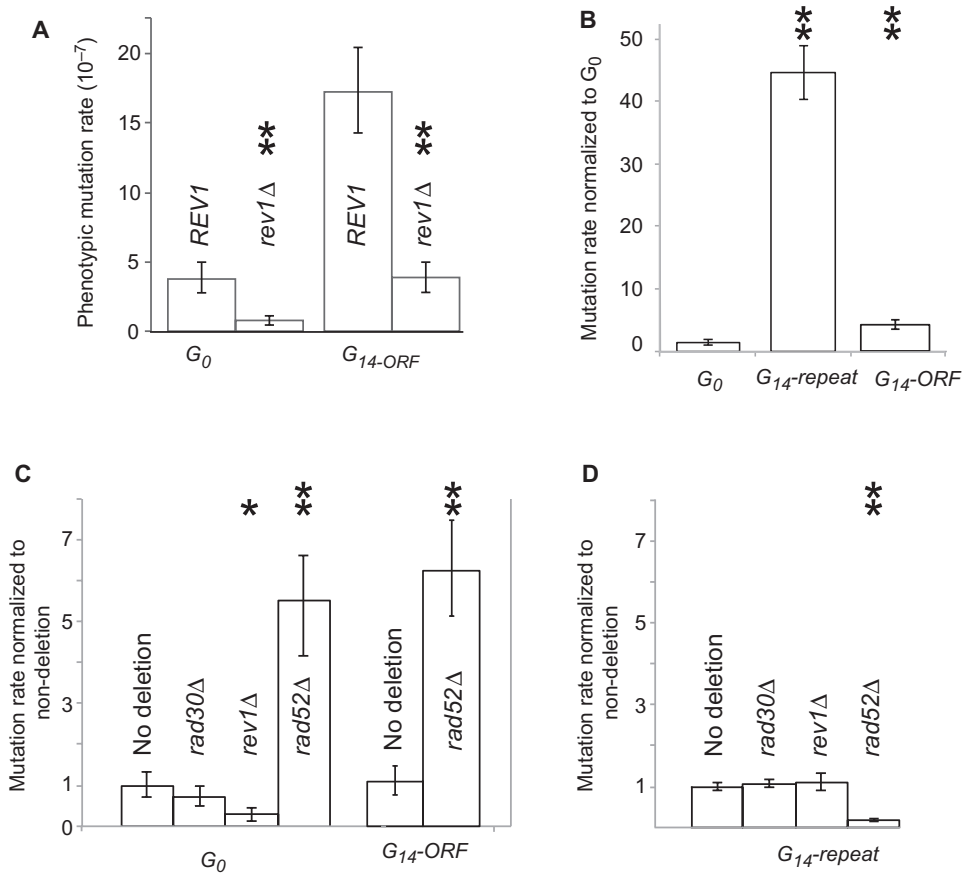


**Fig. 6. Expansion of polyguanine repeats occurs at a high rate and is RAD52-dependent.** (**A**) Mutation rates of $G_0$ and $G_{14}$-ORF strains compared to their respective rev1 deletion mutants. (**B**) Mutation rates of $G_0$, $G_{14}$-repeat, and $G_{14}$-ORF strains, shown relative to the $G_0$ mutation rate. (**C** and **D**) In each panel, the mutation rates of deletion mutants are shown relative to their respective nondeletion progenitor, either $G_0$ (C), $G_{14}$-ORF (C), or $G_{14}$-repeat (D). Significant differences calculated using *t* tests are indicated by asterisks: $*P < 0.05$ and $**P < 0.005$. Error bars represent 95% confidence intervals.

(*URA3-w*), whose function is mildly compromised (*$G_{14}$-repeat*, Fig. 1B). We had previously observed that a $G_{15}$ repeat in the 5′UTR region of *URA3-w* could cause a reduction in protein translation (fig. S1). Although *$G_{14}$-ORF-w* exhibits the $Ura^+$ phenotype of the wild-type allele, a mutation from $G_{14}$ to $G_{15}$ (or longer) results in the assayable loss of *URA3* function, probably due to a combined effect of impaired function and reduced translation (Fig. 1B). We used this construct (*$G_{14}$-repeat*) to directly measure the mutation rate of repeat length increase from $G_{14}$ to $G_{15}$ or longer (contraction of the G repeat would not generate the phenotype). We found that repeat length–dependent increases in mutation rate were higher than the downstream mutation rate, with a 45-fold difference between *$G_{14}$-repeat* and $G_0$ cells (Fig. 6B). Sequencing of 105 independent *ura3* mutant clones of *$G_{14}$-repeat* confirmed that most of them carried an increased polyguanine repeat ($G_{15}$) but there were no mutations in the coding region and no large deletions (table S1).

### Repeat expansion occurs independently of *REV1* and *RAD30*

Although deletion of *REV1* reduced the mutation rate within the *URA3* ORF (as detected by the *$G_{14}$-ORF* construct), we found that deletion of *REV1* had no effect on the mutation rate in the homopolymeric repeat as measured using the *$G_{14}$-repeat* construct (Fig. 6D). To confirm that another translesion DNA synthesis pathway was not involved, a gene essential for another translesion pathway, *RAD30*, was deleted and also had no effect on mutation rate. We next turned to the alternative mechanism for rescuing the stalled replication fork, HR. *Rad52* is essential for the annealing of DNA strands during HR (*45*), and its ablation causes an increase in mutation rate of approximately fivefold in $G_0$ cells (Fig. 6D). The reason for this increase is that *rad52* mutants depend on error-prone DNA polymerases to synthesize over the single-stranded gaps resulting from replication fork stalling (*32*, *33*). Conversely, we found that deletion of *RAD52* in the *$G_{14}$-repeat* strain markedly reduced *URA3* mutation rates (Fig. 6D), consistent with previous work examining recombination and frameshifts underlying "adaptive mutation" in *Escherichia coli* (*46*–*48*).

We checked whether *rad52* deletion was able to reduce the mutation rate in *$G_{14}$-ORF* cells. However, similar to $G_0$ cells, the mutation rate was increased approximately fivefold (Fig. 6C), showing that Rad52-mediated HR affects only the change of $G_{14}$ length, not the downstream mutagenic effect of $G_{14}$.

### $G_{14}$ mutagenesis is not caused by formation of a G-quadruplex structure

G-quadruplex structures, which can induce double-strand breaks and replication fork pausing in Yeast *pif1* (G-quadruplex resolvase)

deletion mutants, are another potential cause of DNA replication stress (*49*). To test whether our polyguanine sequences could form G-quadruplex structures, we compared a known G-quadruplex–forming sequence from *Tetrahymena* (*50*) to $G_{11}$, $G_{12}$, $G_{13}$, and $G_{14}$ sequences. We designed five oligomers (given in Materials and Methods) that included either the G-quadruplex control sequence or 11 to 14 guanines in a row, each integrated into the same sequence context as the *URA3* constructs used for fluctuation tests in this study. We first performed circular dichroism analysis of the oligos in ionic solutions that promote G-quadruplex formation. Circular dichroism analysis showed that the control was able to form a G-quadruplex, whereas the $G_{11}$ to $G_{14}$ repeat sequences could not (fig. S2, A to E). We then used DNA polymerase stop assays to test whether DNA polymerase could synthesize the complementary DNA across the single-stranded template, based on the principle that a stable secondary structure should inhibit DNA synthesis. The results show that the G-quadruplex control blocked the polymerase, whereas $G_{11}$ to $G_{14}$ sequences did not have the same effect (fig. S2F).

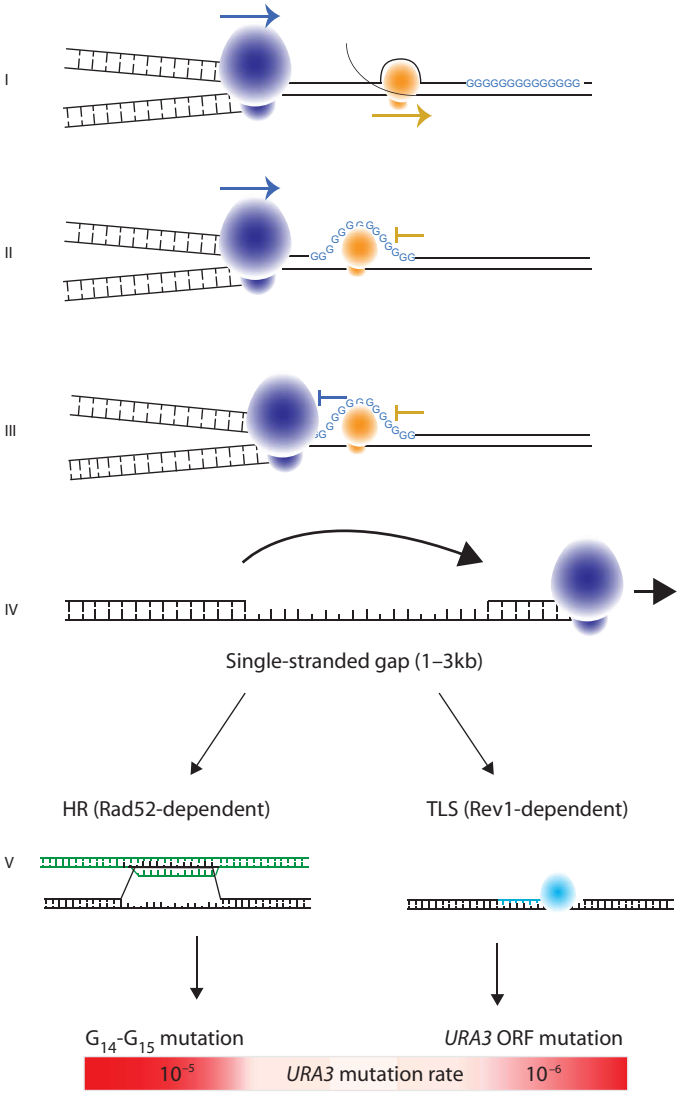


**Fig. 7. Model for the outcome of $G_{13+}$-induced replication fork stalling.** (**I**) The replication fork can proceed in either direction, but the transcription complex must encounter the $G_{13+}$ repeat before the transcribed gene. (**II**) Transcription stalls at the $G_{13+}$ sequence. (**III**) The replication fork stalls at the $G_{13+}$ sequence; stalling is more likely if there is a stalled transcription complex already present. (**IV**) The replication fork detaches from the template and reinitiates replication downstream, leaving a patch of single-stranded DNA that is 800 to 3000 bp in length. (**V**) The DNA complementary to the single-stranded gap is synthesized using either Rad52-dependent HR (detected using the $G_{14}$-repeat construct) or Rev1-dependent TLS (detected using the $G_{14}$-ORF construct) to bypass the difficult-to-replicate region.

## DISCUSSION

"Mutation at a distance" has been shown to occur in a number of different sequence contexts. For instance, DNA sequences that are prone to double-strand breaks have been indirectly linked to kataegis, a catastrophic mutational event that causes large clusters of mutations dispersed over tens of kilobases in cancer genomes (*51*, *52*). Break-induced replication (BIR) is often initiated in response to double-strand breaks (*53*) and produces long regions of DNA in single-stranded form, which are more susceptible to mutation than double-stranded DNA. Experiments in yeast have shown that the widely distributed clusters of point mutations characteristic of kataegis result from the attack by mutagenic agents upon extended tracts of single-stranded DNA (*53*), whereas studies in yeast (*54*, *55*) and mammalian cells (*52*) have implicated the deregulation of AID/APOBEC deaminases as an alternative cause of kataegis-like mutational events. The elevated mutation rate that we observed in this study is unlikely to be caused by a similar mechanism because *$G_{14}$-ORF* mutagenesis requires *REV1* whereas BIR functions independently of any translesion polymerase (*56*).

Double-strand breaks have also been implicated as the cause of the mutagenicity of long repeats known to form hairpin secondary structures (*25*, *26*). The importance of double-strand breaks as a trigger for induced mutagenesis has been further reinforced by studies that directly induce double-strand breaks using HO (*27*) or I-SCE1 (*28*). In all of these systems, increases in mutation rate are detectable in the surrounding 1 to 2 kb of sequence and dependent on TLS (*25*–*28*, *57*). Conversely, the repeat-induced mutagenesis system studied by Shishkin *et al.* (*58*) and Shah *et al.* (*59*) occurs independently of TLS and can induce mutations both upstream and downstream of the repeat sequence. Mutagenesis in this system only occurs when repeat tracts exceed the length of the Okazaki fragment.

Here, we provide a tentative mechanism for how $G_{14}$ repeats can lead to elevated mutation rates in the surrounding DNA sequence. The 2D gel results suggest that the replication fork is more prone to stall at the *$G_{13}$-URA3* sequence than at the *$G_0$-URA3* sequence. Mutation rates of the *$C_{14}$-ORF*, *ORF-$C_{14}$*, and *ORF-$G_{14}$* shown in Fig. 2C suggest that the mutagenic effect of $G_{14}$ was dependent on the direction of replication. In this experiment, the guanines are seemingly required to be in the leading strand, and the mutagenic effect is only conferred on sequences that are replicated downstream of the guanine repeat sequence. However, this result is also consistent with the direction of transcription playing a role. $G_{14}$ mutagenesis only occurs when transcription proceeds through $G_{14}$ with the *URA3* downstream.

The question of the relative importance of replication or transcription is addressed by the insertion of *$G_{14}$-ORF* constructs in various positions in chromosomes XII and XV (Fig. 4 and table S2). Although all

of these constructs are transcribed through $G_{14}$ to *URA3*, the direction of replication and whether the genes are on the Watson or Crick strand vary by chromosome position (table S2). All constructs had a 4- to 18-fold increase in mutation rate, regardless of the direction of replication or the position of $G_{14}$ on the leading or lagging strand. These results concur with previous work that has shown that $G_{20}$ to $G_{32}$ repeats on plasmids caused high rates of replication fork stalling that depend on transcription (*40*, *41*). In agreement with our results, the equivalent cytosine repeats ($C_{20}$ to $C_{32}$) did not have an effect on replication fork stalling (*40*, *41*). Mutation rates were not measured in these previous experiments, but in our study, even with repressed transcription, the mutation rate remained elevated (Fig. 5D). The tet repressor is known to be extremely effective at repressing expression; thus, it is unlikely that leaky expression can account for the remaining mutagenic effect in the noninduced *tet-*$G_{13}$*-URA3*. However, it has been shown that RNA polymerase II (RNAPII) (sufficient for hindering the DNA replication fork) can bind to the tet promoter even in the absence of induction (*60*). Figure 5 shows that tet repression of transcription causes the $G_0$ strain to accumulate more Y replication fork intermediates. This is consistent with RNAPII binding the repressed tet promoter. However, these accumulated Y fork intermediates do not translate into an increased mutation rate for $G_0$. That $G_0$ does not have an elevated mutation rate despite an increase in replication fork intermediates suggests that the $G_{13+}$ sequence does more than stall the replication fork; it may also interfere with homologous repair, biasing toward TLS repair. The precise role of transcription in $G_{13+}$ mutagenesis requires further work to be fully resolved.

One model proposed by Lopes and co-workers (*32*), and later by Huang *et al.* (*33*), postulates that after replication fork stalling, the replication complex decouples from the fork and reinitiates replication downstream. This leaves a patch of single-stranded DNA to be filled using either HR or TLS. We propose that $G_{14}$ repeats confer a higher probability of replication fork stalling than other DNA sequences (Fig. 7). This leads to more TLS activity in the region and, therefore, a higher mutation rate. There is mounting evidence that in early S phase, high-fidelity HR is the synthesis mechanism of choice, whereas in late S phase and early $G_2$ phase, TLS is preferred (*11*, *14*, *33*). Our finding of a strong correlation between $G_{14}$-induced mutation and DNA replication timing adds support to this hypothesis by suggesting that $G_{14}$ repeats in later-replicating DNA are more likely to depend on TLS for repair, leading to a higher average mutation rate in this region.

In one recent study, mismatch repair efficiency was found to vary with time of replication (*15*). We show that most of the difference in mutation rates between $G_{14}$*-URA3* and $G_0$*-URA3* is dependent on having a functional *REV1*. We also show that $G_{14}$ inserts have a stronger correlation between replication timing and mutation rate than $G_0$ inserts. It follows that this stronger correlation is also *REV1*-dependent, and in support of this, a previous comprehensive analysis of $G_0$*-URA3* inserts at 49 sites on chromosome 6 found that the deletion of *REV1* abolished the correlation between replication timing and mutation rate (*11*). These results suggest that mismatch repair plays a background role in the correlation between mutation rate and replication timing that we observed in this study.

The $G_{14}$ repeats that we study do not require an artificial increase in mutation rate or polymerase deletion for the effect to be detectable. Moreover, $G_{14}$ repeats are short, simple, and present in most genomes, suggesting that $G_{14}$ repeats could play an important role in genome evolution. The tight correlation between repeat-induced mutation and replication timing ($R = 0.89$, $P = 0.007$) suggests that the degree to which a repeat sequence will alter mutation rate can be directly combined with knowledge of replication timing. Because such repeats are common throughout all genomes, it is not implausible that these repeats could have implications as locally acting mutator sequences. In addition, short repeat sequences are easy to engineer and could be applied to synthetic biology applications for speeding up evolution of a target gene.

## MATERIALS AND METHODS

### Strain construction

All strains were constructed in a strain isogenic with *W303* (*MAT**a** his3-11,15 leu2-3,112 trp1-1 ura3 ade2-1*). Homopolymeric nucleotide strains were constructed by amplifying *URA3*, with primers containing a homopolymeric nucleotide tract at the position between −4 and −5 of *URA3*, and the resultant polymerase chain reaction (PCR) product was transformed into ura$^-$ yeast cells using the LiAc transformation method. The *URA3* genes of transformants were amplified using PCR, and the sequences were confirmed by Sanger sequencing. Different mutant strains were constructed by amplifying the G418 insertion mutant for each gene of interest from the whole-genome deletion collection. Strains were transformed with PCR products and deletion mutants selected based on their resistance to G418. $G_{14}$*-repeat* was constructed using an alternative *URA3* sequence that carries a point mutation at position 736 (val247ile), which has slightly reduced function compared to the wild-type *URA3* gene. Change in repeat length from $G_{14}$ to $G_{15}$ in the $G_{14}$*-repeat* construct reduced protein translation such that cells containing this mutation were 5-FOA–resistant and detectable using the mutation rate assay.

### Fluctuation assays

Strains to be assayed were grown overnight in 3-ml complete supplement mixture (CSM)–uracil medium, diluted $10^{-4}$, and then inoculated into 100-μl cultures so that there were approximately 1000 cells per culture. At least 24 independent cultures were used per assay, and each assay was repeated at least three times. Cultures were left overnight at 30°C until the cultures were assessed to have reached a suitable density, and then the entire culture, except for 5 μl, was plated onto predried 5-FOA plates to detect *ura3* mutants that were 5-FOA–resistant. The remaining culture was pooled and diluted, and then the cell count was assayed using a Scepter cell counter. Mutation rates were calculated using the maximum likelihood method (*61*). To measure the background mutation rate at a site distal from the *URA3* locus, the wild-type *CAN1* locus was restored in $G_0$*-URA3* and $G_{14}$*-ORF* strains. Mutation rate assays were carried out the same as above except that mutations in the *CAN1* gene were detected by plating on CSM-arginine plates supplemented with canavanine (60 μg/ml).

### 2D gel analysis

Strains for the 2D gel analysis were constructed by amplifying *tet-*$G_N$*-URA3* fragments from a plasmid for insertion by HR into the intergenic region between YCL052C and YCL054W at position 34,028 on chromosome III. After we introduced the *tet-*$G_N$*-URA3* fragments into *S. cerevisiae*, we plated the cells on selection plates lacking uracil. Colonies were picked and transformation was confirmed by PCR and Sanger sequencing. For the 2D gel experiments, log-phase cells [$OD_{600}$ (optical density at 600 nm) ≈ 0.5] were synchronized at $G_1$ by α factor (10 μg/ml) [with or without doxycycline (10 μg/ml)] for 2.5 hours and

then released in YPD (yeast extract, peptone, and dextrose) medium (with or without doxycycline) at 25°C. Cells were harvested at 15, 30, 45, and 60 min and terminated immediately using sodium azide, and then in vivo psoralen cross-linking was undertaken as previously described (*62*, *63*). Genomic DNA was digested using Hind III. First-dimension gels were made of 0.35% agarose and second-dimension gels were made of 1% agarose. A DNA fragment correlated to the yeast genome positions 34,028 to 34,999 was used to make the probe. Replication intermediate signals were quantified as previously described (*39*).

### DNA synthesis stop assay

To determine whether $G_{11}$ to $G_{14}$ sequences could form G-quadruplex structures, we conducted experiments comparing a known G-quadruplex–forming sequence from *Tetrahymena* (GGGTTGGGTTGGGTTGGGTT) (*50*) to $G_{11}$, $G_{12}$, $G_{13}$, and $G_{14}$ sequences. We designed oligonucleotides composed of either homopolymeric runs of 11 to 14 G's in a row or the G-quadruplex sequence, integrated into the same sequence context as the genetic constructs used to measure mutation rate in this study. Following Han and co-workers (*64*), a radiolabeled primer ($\Gamma$-$^{32}$P), shown below, was annealed with template DNA (10 nM) in buffer containing 5 mM KCl. To initiate the sequencing reactions, MgCl (3 μM), Taq polymerase (2.5 U per reaction), and deoxynucleotide triphosphates (final concentration of 100 μM) were added and the mix was incubated at either 37° or 55°C. The reactions were stopped and then run on 12% polyacrylamide gel. If the template forms a G-quadruplex, then DNA synthesis will not be completed, and no band can be visualized on the polyacrylamide gel.

Primer: [CTGCACAGAACAAAAACCTGCAGGAAACG]. Templates: G-quadruplex control, GCTTTCGACATGATT(GGGTTGGGTTGGGTTGGGTT)TATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAG; $G_{11}$, d[GCTTTCGACATGATTGGGGGGGGGGGTATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAG; $G_{12}$, GCTTTCGACATGATTGGGGGGGGGGGTATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAG; $G_{13}$, GCTTTCGACATGATTGGGGGGGGGGGGGGTATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAG; and $G_{14}$, GCTTTCGACATGATTGGGGGGGGGGGGGGGTATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAG.

### Circular dichroism

Following the work by Dexheimer *et al.* (*50*), we incubated cuvettes containing 5 μM of oligomeric DNA dissolved in tris-HCl (50 mM, pH 7.6) containing either 100 mM KCl or 100 mM NaCl for 5 min at 90°C and then let them slowly cool to 25°C. Circular dichroism spectra were measured on a spectropolarimeter (J-815, JASCO) using a 1-cm path length quartz cuvette, over a range of 200 to 320 nm, with a response time of 1 s and a scanning speed of 100 nm min$^{-1}$. Three replicate measurements were taken and measured at 25°C.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/5/e1501033/DC1

fig. S1. Expansion of the polyguanine repeat ($G_{14}$ to $G_{15}$) reduces the Ura3 protein abundance but not the mRNA level.

fig. S2. $G_{11}$ to $G_{14}$ sequences do not stop DNA polymerase from synthesizing DNA, whereas G-quadruplex does.

table S1. Summary table of 318 sequenced ura3 mutants from $G_0$, $G_{14}$-*ORF*, and $G_{14}$-*repeat* strains.

table S2. Chromosome insertion position and replication timing for engineered $G_{14}$-*URA* inserts.

References (*66*, *67*)

## REFERENCES AND NOTES


1. P. D. Keightley, M. Lynch, Toward a realistic model of mutations affecting fitness. *Evolution* **57**, 683–685 (2003).
2. R. Sanjuán, M. R. Nebot, N. Chirico, L. M. Mansky, R. Belshaw, Viral mutation rates. *J. Virol.* **84**, 9733–9748 (2010).
3. J. W. Drake, The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Ann. N. Y. Acad. Sci.* **870**, 100–107 (1999).
4. A. Giraud, M. Radman, I. Matic, F. Taddei, The rise and fall of mutator bacteria. *Curr. Opin. Microbiol.* **4**, 582–585 (2001).
5. M. J. McDonald, Y.-Y. Hsieh, Y.-H. Yu, S.-L. Chang, J.-Y. Leu, The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. *Curr. Biol.* **22**, 1235–1240 (2012).
6. S. Wielgoss, J. E. Barrick, O. Tenaillon, M. J. Wiser, W. J. Dittmar, S. Cruveiller, B. Chane-Woon-Ming, C. Médigue, R. E. Lenski, D. Schneider, Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 222–227 (2013).
7. A. Datta, S. Jinks-Robertson, Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**, 1616–1619 (1995).
8. B. Schuster-Böckler, B. Lehner, Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).
9. Y. Yang, J. Sterling, F. Storici, M. A. Resnick, M. A. Gordenin, Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. *PLOS Genet.* **4**, e1000264 (2008).
10. K. Chan, J. F. Sterling, S. A. Roberts, A. S. Bhagwat, M. A. Resnick, D. A. Gordenin, Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. *PLOS Genet.* **8**, e1003149 (2012).
11. G. I. Lang, A. W. Murray, Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol. Evol.* **3**, 799–811 (2011).
12. J. A. Stamatoyannopoulos, I. Adzhubei, R. E. Thurman, G. V. Kryukov, S. M. Mirkin, S. R. Sunyaev, Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393–395 (2009).
13. M. S. Lawrence, P. Stojanov, P. Polak, G. V. Kryukov, K. Cibulskis, A. Sivachenko, S. L. Carter, C. Stewart, C. H. Mermel, S. A. Roberts, A. Kiezun, P. S. Hammerman, A. McKenna, Y. Drier, L. Zou, A. H. Ramos, T. J. Pugh, N. Stransky, E. Helman, J. Kim, C. Sougnez, L. Ambrogio, E. Nickerson, E. Shefler, M. L. Cortés, D. Auclair, G. Saksena, D. Voet, M. Noble, D. DiCara, P. Lin, L. Lichtenstein, D. I. Heiman, T. Fennell, M. Imielinski, B. Hernandez, E. Hodis, S. Baca, A. M. Dulak, J. Lohr, D.-A. Landau, C. J. Wu, J. Melendez-Zajgla, A. Hidalgo-Miranda, A. Koren, S. A. McCarroll, J. Mora, R. S. Lee, B. Crompton, R. Onofrio, M. Parkin, W. Winckler, K. Ardlie, S. B. Gabriel, C. W. M. Roberts, J. A. Biegel, K. Stegmaier, A. J. Bass, L. A. Garraway, M. Meyerson, T. R. Golub, D. A. Gordenin, S. Sunyaev, E. S. Lander, G. Getz, Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
14. L. S. Waters, G. C. Walker, 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 (2006).
15. S. A. Lujan, A. R. Clausen, A. B. Clark, H. K. MacAlpine, D. M. MacAlpine, E. P. Malc, P. A. Mieczkowski, A. B. Burkholder, D. C. Fargo, D. A. Gordenin, T. A. Kunkel, Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.* **24**, 1751–1764 (2014).
16. G. I. Lang, L. Parsons, A. E. Gammie, Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3* **3**, 1453–1465 (2013).
17. G. Levinson, G. A. Gutman, Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221 (1987).
18. K. Hammerschmidt, C. J. Rose, B. Kerr, P. B. Rainey, Life cycles, fitness decoupling and the evolution of multicellularity. *Nature* **515**, 75–79 (2014).
19. S. M. Mirkin, Expandable DNA repeats and human disease. *Nature* **447**, 932–940 (2007).
20. R. Moxon, C. Bayliss, D. Hood, Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **40**, 307–333 (2006).
21. K. Usdin, E. Grabczyk, DNA repeat expansions and human disease. *Cell. Mol. Life Sci.* **57**, 914–931 (2000).
22. D. Tautz, Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**, 6463–6471 (1989).
23. E. Zlotorynski, A. Rahat, J. Skaug, N. Ben-Porat, E. Ozeri, R. Hershberg, A. Levi, S. W. Scherer, H. Margalit, B. Kerem, Molecular basis for expression of common and rare fragile sites. *Mol. Cell Biol.* **23**, 7143–7151 (2003).
24. Y. F. Chan, M. E. Marks, F. C. Jones, G. Villarreal Jr., M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, D. M. Kingsley, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
25. W. Tang, M. Dominska, M. Gawel, P. W. Greenwell, T. D. Petes, Genomic deletions and point mutations induced in *Saccharomyces cerevisiae* by the trinucleotide repeats (GAA.TTC) associated with Friedreich's ataxia. *DNA Repair* **12**, 10–17 (2013).

26. N. Saini, Y. Zhang, Y. Nishida, Z. Sheng, S. Choudhury, P. Mieczkowski, K. S. Lobachev, Fragile DNA motifs trigger mutagenesis at distant chromosomal loci in *saccharomyces cerevisiae*. *PLOS Genet.* **9**, e1003551 (2013).
27. A. J. Rattray, B. K. Shafer, C. B. McGill, J. N. Strathern, The roles of *REV3* and *RAD57* in double-strand-break-repair-induced mutagenesis of *Saccharomyces cerevisiae*. *Genetics* **162**, 1063–1077 (2002).
28. C. Shee, J. L. Gibson, S. M. Rosenberg, Two mechanisms produce mutation hotspots at DNA breaks in *Escherichia coli*. *Cell Rep.* **2**, 714–721 (2012).
29. M. J. McDonald, W. C. Wang, H.-D. Huang, J.-Y. Leu, Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLOS Biol.* **9**, e1000622 (2011).
30. C. A. Nieduszynski, Y. Knox, A. D. Donaldson, Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev.* **20**, 1874–1879 (2006).
31. E. V. Mirkin, S. M. Mirkin, Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.* **71**, 13–35 (2007).
32. M. Lopes, M. Foiani, J. M. Sogo, Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol. Cell* **21**, 15–27 (2006).
33. D. Huang, B. D. Piening, A. G. Paulovich, The preference for error-free or error-prone post-replication repair in *Saccharomyces cerevisiae* exposed to low-dose methyl methanesulfonate is cell cycle dependent. *Mol. Cell Biol.* **33**, 1515–1527 (2013).
34. C. Yu, H. Gan, J. Han, Z.-X. Zhou, S. Jia, A. Chabes, G. Farrugia, T. Ordog, Z. Zhang, Strand-specific analysis shows protein binding at replication forks and PCNA unloading from lagging strands when forks stall. *Mol. Cell* **56**, 551–563 (2014).
35. P. V. Shcherbakova, I. J. Fijalkowska, Translesion synthesis DNA polymerases and control of genome stability. *Front. Biosci.* **11**, 2496–2517 (2006).
36. X. Zhong, P. Garg, C. M. Stith, S. A. Nick McElhinny, G. E. Kissling, P. M. J. Burgers, T. A. Kunkel, The fidelity of DNA synthesis by yeast DNA polymerase zeta alone and with accessory proteins. *Nucleic Acids Res.* **34**, 4731–4742 (2006).
37. S.-K. Quah, R. C. von Borstel, P. J. Hastings, The origin of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* **96**, 819–839 (1980).
38. C. Cassier, R. Chanet, J. A. P. Henriques, E. Moustacchi, The effects of three *PSO* genes on induced mutagenesis: A novel class of mutationally defective yeast. *Genetics* **96**, 841–857 (1980).
39. M. Fumasoni, K. Zwicky, F. Vanoli, M. Lopes, D. Branzei, Error-free DNA damage tolerance and sister chromatid proximity during DNA replication rely on the Polα/Primase/Ctf4 complex. *Mol. Cell* **57**, 812–823 (2015).
40. M. M. Krasilnikova, G. M. Samadashwily, A. S. Krasilnikov, S. M. Mirkin, Transcription through a simple DNA repeat blocks replication elongation. *EMBO J.* **17**, 5095–5102 (1998).
41. B. P. Belotserkovskii, R. Liu, S. Tornaletti, M. M. Krasilnikova, S. M. Mirkin, P. C. Hanawalt, Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12816–12821 (2010).
42. S. Prakash, R. E. Johnson, L. Prakash, Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annu. Rev. Biochem.* **74**, 317–353 (2005).
43. J. F. Lemontt, Mutants of yeast defective in mutation induced by ultraviolet light. *Genetics* **68**, 21–33 (1971).
44. K. A. Schürer, C. Rudolph, H. D. Ulrich, W. Kramer, Yeast *MPH1* gene functions in an error-free DNA damage bypass pathway that requires genes from Homologous recombination, but not from postreplicative repair. *Genetics* **166**, 1673–1686 (2004).
45. U. H. Mortensen, C. Bendixen, I. Sunjevaric, R. Rothstein, DNA strand annealing is promoted by the yeast Rad52 protein. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10729–10734 (1996).
46. P. L. Foster, J. M. Trimarchi, Adaptive reversion of a frameshift mutation in Escherichia coli by simple base deletions in homopolymeric runs. *Science* **265**, 407–409 (1994).
47. R. S. Harris, S. Longerich, S. M. Rosenberg, Recombination in adaptive mutation. *Science* **264**, 258–260 (1994).
48. S. M. Rosenberg, S. Longerich, P. Gee, R. S. Harris, Adaptive mutation by deletions in small mononucleotide repeats. *Science* **265**, 405–407 (1994).
49. J. Lopes, A. Piazza, R. Bermejo, B. Kriegsman, A. Colosio, M.-P. Teulade-Fichou, M. Foiani, A. Nicolas, G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **30**, 4033–4046 (2011).
50. T. S. Dexheimer, D. Sun, L. H. Hurley, Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the *bcl-2* P1 promoter. *J. Am. Chem. Soc.* **128**, 5404–5415 (2006).
51. S. Nik-Zainal, L. B. Alexandrov, D. C. Wedge, P. Van Loo, C. D. Greenman, K. Raine, D. Jones, J. Hinton, J. Marshall, L. A. Stebbings, A. Menzies, S Martin, K. Leung, L. Chen, C. Leroy, M. Ramakrishna, R. Rance, K. W. Lau, L. J. Mudie, I. Varela, D. J. McBride, G. R. Bignell, S. L. Cooke, A. Shlien, J. Gamble, I. Whitmore, M. Maddison, P. S. Tarpey, H. R. Davies, E. Papaemmanuil, P. J. Stephens, S. McLaren, A. P. Butler, J. W. Teague, G. Jönsson, J. E. Garber, D. Silver, P. Miron, A. Fatima, S. Boyault, A. Langerød, A. Tutt, J. W. M. Martens, S. A. J. R. Aparicio, Å. Borg, A. V. Salomon, G. Thomas, A.-L. Børresen-Dale, A. L. Richardson, M. S. Neuberger, P. A. Futreal, P. J. Campbell, M. R. Stratton, Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
52. R. S. Harris, Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res.* **17**, 8 (2015).
53. C. J. Sakofsky, S. A. Roberts, E. Malc, P. A. Mieczkowski, M. A. Resnick, D. A. Gordenin, A. Malkova, Break-induced replication is a source of mutation clusters underlying kataegis. *Cell Rep.* **7**, 1640–1648 (2014).
54. A. G. Lada, E. I. Stepchenkova, I. S. Waisertreiger, V. N. Noskov, A. Dhar, J. D. Eudy, R. J. Boissy, M. Hirano, I. B. Rogozin, Y. I. Pavlov, Genome-wide mutation avalanches induced in diploid yeast cells by a base analog or an APOBEC deaminase. *PLOS Genet.* **9**, e1003736 (2013).
55. B. J. Taylor, S. Nik-Zainal, Y. L. Wu, L. A. Stebbings, K. Raine, P. J. Campbell, C. Rada, M. R. Stratton, M. S. Neuberger, DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* **2**, e00534 (2013).
56. J. R. Lydeard, Z. Lipkin-Moore, Y. J. Sheu, B. Stillman, P. M. Burgers, J. E. Haber, Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes Dev.* **24**, 1133–1144 (2010).
57. S. L. Holbeck, J. N. Strathern, A role for *REV3* in mutagenesis during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **147**, 1017–1024 (1997).
58. A. A. Shishkin, I. Voineagu, R. Matera, N. Cherng, B. T. Chernet, M. M. Krasilnikova, V. Narayanan, K. S. Lobachev, S. M. Mirkin, Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol. Cell* **35**, 82–92 (2009).
59. K. A. Shah, A. A. Shishkin, I. Voineagu, Y. I. Pavlov, P. V. Shcherbakova, S. M. Mirkin, Role of DNA polymerases in repeat-mediated genome instability. *Cell Rep.* **2**, 1088–1095 (2012).
60. T. Uhlmann, S. Boeing, M. Lehmbacher, M. Meisterernst, The VP16 activation domain establishes an active mediator lacking CDK8 in vivo. *J. Biol. Chem.* **282**, 2163–2173 (2007).
61. B. M. Hall, C. X. Ma, P. Liang, K. K. Singh, Fluctuation AnaLysis CalculatOR: A web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564–1565 (2009).
62. K. L. Friedman, B. J. Brewer, Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. *Methods Enzymol.* **262**, 613–627 (1995).
63. G. Liberi, C. Cotta-Ramusino, M. Lopes, J. Sogo, C. Conti, A. Bensimon, M. Foiani, Methods to study replication fork collapse in budding yeast. *Methods Enzymol.* **409**, 442–462 (2006).
64. H. Han, L. H. Hurley, M. Salazar, A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.* **27**, 537–542 (1999).
65. M. Gossen, H. Bujard, Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5547–5551 (1992).
66. Z.-F. Wang, M.-H. Li, S.-T. Hsu, T.-C. Chang, Structural basis of sodium-potassium exchange of a human telomeric DNA quadruplex without topological conversion. *Nucleic Acids Res.* **42**, 4723–4733 (2014).
67. D. Sen, W. Gilbert, A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344**, 410–414 (1990).


**Acknowledgments:** We thank A. Murray, P. Rainey, M. Fumasoni, and three anonymous reviewers for their comments on the manuscript. We thank members of the Leu laboratory for helpful discussions. **Funding:** M.J.M. was funded by an Academia Sinica distinguished postdoctoral fellowship. J.-Y.L. was supported by Academia Sinica of Taiwan (100-CDA-L04) and the Taiwan Ministry of Science and Technology (NSC103-2321-B-001-001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **Author contributions:** M.J.M. and J.-Y.L. designed the study. M.J.M., Y.-H.Y., S.Y.C., C.-F.K., and J.-F.G. performed experiments. M.J.M. and J.-Y.L. analyzed the data and wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 3 August 2015
Accepted 29 April 2016
Published 27 May 2016
10.1126/sciadv.1501033

**Citation:** M. J. McDonald, Y.-H. Yu, J.-F. Guo, S. Y. Chong, C.-F. Kao, J.-Y. Leu, Mutation at a distance caused by homopolymeric guanine repeats in *Saccharomyces cerevisiae*. *Sci. Adv.* **2**, e1501033 (2016).