Frameshift Mutagenesis: The Roles of Primer–Template Misalignment and the Nonhomologous End-Joining Pathway in *Saccharomyces cerevisiae*

Kevin Lehner,*,[1] Sarah V. Mudrak,*,†,[1] Brenda K. Minesinger,‡,[2] and Sue Jinks-Robertson*,†,[3]

*Department of Molecular Genetics and Microbiology , †University Program in Genetics and Genomics, Duke University Medical Center, Durham, North Carolina 27710, and ‡Biochemistry, Cell and Developmental Biology Program, Graduate Division of Biological and Biomedical Sciences, Emory University, Atlanta, Georgia 30322

ABSTRACT Small insertions or deletions that alter the reading frame of a gene typically occur in simple repeats such as mononucleotide runs and are thought to reflect spontaneous primer–template misalignment during DNA replication. The resulting extrahelical repeat is efficiently recognized by the mismatch repair machinery, which specifically replaces the newly replicated strand to restore the original sequence. Frameshift mutagenesis is most easily studied using reversion assays, and previous studies in *Saccharomyces cerevisiae* suggested that the length threshold for polymerase slippage in mononucleotide runs is 4N. Because the probability of slippage is strongly correlated with run length, however, it was not clear whether shorter runs were unable to support slippage or whether the resulting frameshifts were obscured by the presence of longer runs. To address this issue, we removed all mononucleotide runs >3N from the yeast *lys2ΔBgl* and *lys2ΔA746* frameshift reversion assays, which detect net 1-bp deletions and insertions, respectively. Analyses demonstrate that 2N and 3N runs can support primer–template misalignment, but there is striking run-specific variation in the frequency of slippage, in the accumulation of +1 *vs.* −1 frameshifts and in the apparent efficiency of mismatch repair. We suggest that some of this variation reflects the role of flanking sequence in initiating primer–template misalignment and that some reflects replication-independent frameshifts generated by the nonhomologous end-joining pathway. Finally, we demonstrate that nonhomologous end joining is uniquely required for the *de novo* creation of tandem duplications from noniterated sequence.

THE accumulation of mutations within genomic DNA is precisely regulated; mutations must be kept at a very low level to maintain genome integrity and yet must be frequent enough to support evolutionary change. Most spontaneous mutations are base substitutions or small insertions/deletions (indels) that reflect errors made either when replicating an undamaged DNA template or when synthesizing over a DNA lesion. Indels that are not a multiple of 3 bp are referred to as frameshift mutations because they change the reading frame of a translating ribosome, thereby altering all downstream amino acids and usually resulting in premature termination of translation. Given the very deleterious nature of frameshift mutations, it is critical that the corresponding mutational intermediates be efficiently recognized and removed.

Repetitive sequences such as mononucleotide or dinucleotide repeats are strong hotspots for frameshifts, and most intermediates arise through spontaneous, replication-associated strand slippage (Streisinger *et al.* 1966). As illustrated for a mononucleotide run in Figure 1A, misalignment between the primer and template strands generates an extrahelical repeat on one of the two strands. If not repaired, an extrahelical nucleotide on the primer strand will become a +1 frameshift mutation, while the persistence of an extrahelical nucleotide on the template strand will result in a −1 frameshift mutation. The frequency with which slippage occurs increases as a function of run length *in vitro* (Kunkel 1990) and *in vivo* (Tran *et al.* 1997). Our previous analyses in yeast suggested that only mononucleotide runs >3N accumulate more frameshifts than predicted by chance, indicating a threshold length of 4N for slippage *in vivo* (Greene and Jinks-Robertson

doi: 10.1534/genetics.111.134890
Manuscript received September 21, 2011; accepted for publication November 11, 2011
[1]These authors contributed equally to this work.
[2]Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 01239.
[3]Corresponding author: Department of Molecular Genetics and Microbiology, 228 Jones Building, 213 Research Dr., Duke University Medical Center, Durham, NC 27710. E-mail: sue.robertson@duke.edu

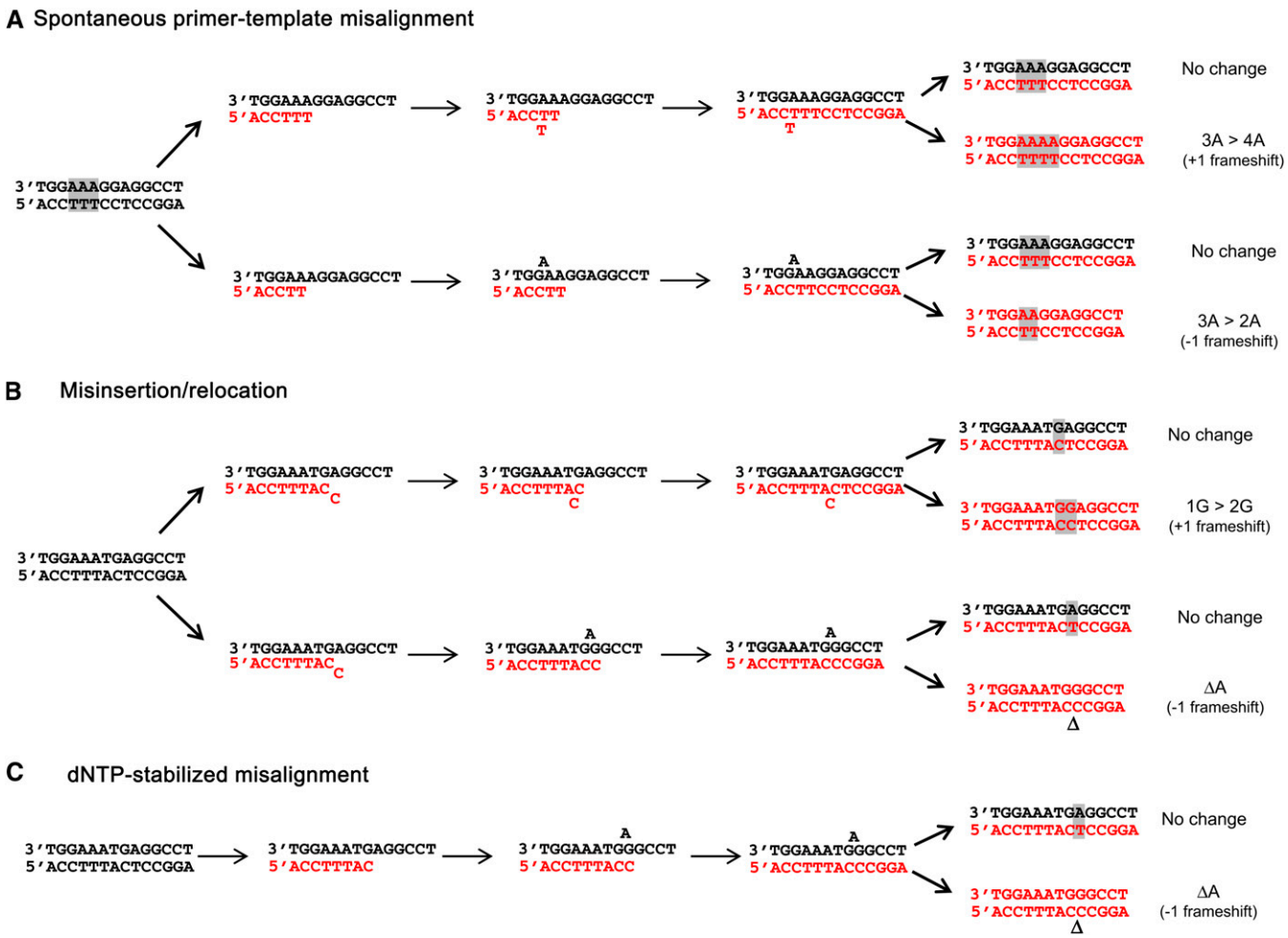

Figure 1 (A–C) Mechanisms of primer–template misalignment. The strands of the original duplex DNA are black and only replication of the top strand is shown. All new strands are red; positions of additions and deletions are highlighted in gray.

1997, 2001). Frameshifts also occur, however, at low levels in smaller repeats and in noniterated sequence (Greene and Jinks-Robertson 2001).

In addition to the spontaneous strand slippage described above, *in vitro* studies have suggested two additional mechanisms of primer–template misalignment (reviewed in Garcia-Diaz and Kunkel 2006). First, frameshift mutagenesis can be initiated by the insertion of an incorrect nucleotide, which creates a mispaired primer terminus that is difficult for DNA polymerase to extend. Subsequent primer–template misalignment can restore proper base pairing, thereby promoting efficient primer extension (Bebenek and Kunkel 1990). If the misinserted nucleotide is complementary to the next base of the template strand, relocation of the terminus will yield a −1 frameshift intermediate; if complementary to the previous base, realignment will produce a +1 frameshift intermediate (Figure 1B). Second, as an alternative to misinsertion/relocation, *in vitro* studies suggest that there can be dNTP-stabilized misalignment at the active site of polymerase, with the incoming dNTP base pairing with the next base in the template strand (Figure 1C) (Efrati *et al.* 1997). This mechanism generates only −1 frameshift intermediates and might be particularly relevant during the bypass of DNA lesions that lack base-pairing potential.

The first defense against polymerization errors derives from the inherent 3′–5′ exonuclease activity of replicative DNA polymerases, which "proofreads" mistakes as they are made (reviewed in Garcia-Diaz and Kunkel 2006). Mutation intermediates that escape proofreading become targets for the postreplicative mismatch repair (MMR) system, which recognizes distortions in the DNA helix (reviewed by Kunkel and Erie 2005). In the context of replication, the MMR system specifically removes a distortion-containing segment of the newly synthesized strand, providing another opportunity for error-free DNA synthesis using the original template. The role of MMR in removing frameshift intermediates is especially important in long runs, which support very high levels of spontaneous primer–template misalignment and are poor substrates for proofreading. In humans, hereditary nonpolyposis colorectal cancer (HNPCC) is associated with MMR defects, the diagnostic feature of which is highly elevated microsatellite instability (Shah *et al.* 2010).

Because of their association with human disease, most studies of frameshift mutagenesis in yeast have focused on highly repetitive sequences; little attention has been given to events that occur within short repeats or noniterated sequence. In the present study, we have focused on the latter events by removing mononucleotide runs >3N from model frameshift reversion assays used in our earlier analyses (Greene and Jinks-Robertson 1997; Harfe and Jinks-Robertson 1999). Analyses in wild-type (WT) and MMR-defective backgrounds demonstrate that runs of 2N or 3N can promote primer–template misalignment, but do so in a highly sequence-context–dependent manner. Significantly, we find that the nonhomologous end-joining (NHEJ) pathway contributes to frameshift mutations in both iterated and noniterated sequence and is uniquely required to generate *de novo* tandem duplications of noniterated sequence.

Materials and Methods

Strain constructions

The WT *lys2ΔBgl* and *lysΔA746* strains (SJR357 and SJR922, respectively) were derived from SJR195 (*MATα ade2-101*$_{oc}$ *his3Δ200 ura3ΔNco*) and were previously described (Greene and Jinks-Robertson 1997; Harfe and Jinks-Robertson 1999). The *lys2ΔBgl,NR* (*NR*, no run) and *lys2ΔA746,NR* alleles were created by site-directed mutagenesis of integrating plasmids pSR699 (*lys2ΔBgl*) (Steele and Jinks-Robertson 1992) and pSR585 (*lys2ΔA746*) (Harfe and Jinks-Robertson 1999), respectively. These plasmids contain *HIS3* as a selectable marker and a 1.2-kb fragment of *LYS2* that spans the reversion window monitored. The primers used to disrupt the 6A run within the reversion window were 5′-GCTAGCTGAA**TC**AATTCAAAG and 5′-CTTTGAATT**GA**TTCAGCTAGC (underlined bases reflect the extent of the original run, with mutated bases indicated in boldface type); the 5T and 4A runs were removed using primers 5′-CGTTTGGCCT**GTC**TGGA**TAT**CCAAGATTTC and 5′-GAAATCTTGG**ATAT**CCA**GAC**AGGCCAAACG; and the 4C run was removed using primers 5′-GGAAAGGA**GG**CCT-CAGTTG and 5′-CAACTGAGG**CC**TCCTTTCC. The resulting *lys2ΔBgl,NR* (pSR701) and *lys2ΔA726,NR* (pSR700) alleles were introduced at the *LYS2* locus of strain SJR195 by two-step allele replacement (Rothstein 1991), generating strains SJR1468 and SJR1467, respectively. The *MSH2* gene of SJR1467 and SJR1468 was disrupted using a *hisG-URA3-hisG* cassette (Greene and Jinks-Robertson 1997), yielding strains SJR1472 and SJR1473, respectively. The *DNL4* gene of SJR1468 was disrupted using a *dnl4Δ::loxP-URA3Kl-loxP* cassette amplified from pUG72 (Gueldener *et al.* 2002), yielding strain SJR3134. SJR3232 and SJR3296 were constructed by transforming SJR1467 with a *dnl4Δ::kan* and *tdp1Δ::kan* cassette, respectively, amplified from pFA6-kanMX6 (Longtine *et al.* 1998).

Mutation rates and spectra

Mutation rates were determined using at least 20 cultures from each of two independent isolates of each strain. Cultures were grown to saturation at 30° in nonselective YEPGE medium (1% yeast extract, 2% Bacto-peptone, 2% glycerol, 2% ethanol, and 250 mg/liter adenine). Appropriate dilutions were plated onto YEPD medium (YEP plus 2% dextrose) to determine total cell number and onto lysine-deficient synthetic glucose medium to select Lys$^+$ revertants. Mutation rates and 95% confidence intervals were determined by maximum likelihood using Salvador 2.0 software (Zheng 2005). Mutation rates for specific mutation types were calculated by multiplying the proportion of that event in the corresponding spectrum by the total Lys$^+$ rate.

To generate mutation spectra, DNA was extracted from purified Lys$^+$ colonies isolated from independent cultures (http://jinks-robertsonlab.duhs.duke.edu/protocols/yeast_prep.html). An appropriate portion of the *LYS2* gene was amplified by PCR and sequenced by the Duke University DNA Analysis Facility (Durham, NC), using primer 5′-GTAA CCGGTGACGATGAT. The proportions of mutations in different spectra were compared by Fisher's exact test (http://faculty.vassar.edu/lowry/VassarStats.html). A P-value <0.05 was considered statistically significant.

Results

The *lys2ΔBgl* allele was constructed by filling in *Bgl*II-generated, 4-nt overhangs, which yields a direct duplication of the sequence GATC and creates the equivalent of a +1 frameshift mutation (Steele and Jinks-Robertson 1992). The *lys2ΔA746* allele was constructed by deleting an adenine nucleotide located at position 746 (relative to the upstream *Xba*I site) of *LYS2* and hence contains a −1 frameshift mutation (Harfe and Jinks-Robertson 1999). The *lys2ΔBgl* and *lys2ΔA746* alleles have largely coincident, ~150-bp reversion windows that fall within a nonessential region of the Lys2 protein, allowing the detection of any compensatory frameshift mutation that restores the correct reading frame. Use of these two alleles thus allows a comparison of the relative locations, types, and rates of net +1 and −1 frameshift mutations that occur within a common region of DNA. The longest, naturally occurring mononucleotide run in this region is composed of six adenines (6A run), with an additional 5T, 4A, and 4C run.

We previously reported that most compensatory frameshifts in the *lys2ΔBgl* and *lys2ΔA746* assays were deletions and insertions, respectively, of a single nucleotide within the mononucleotide runs noted above (Greene and Jinks-Robertson 1997; Harfe and Jinks-Robertson 1999). In a repair-proficient background, such mutations comprised 57% and 74% of the reversion spectra, respectively (see Figures 2A and 3A). Because these percentages greatly exceeded the proportion of the window occupied by these runs, and events at smaller runs or noniterated positions were

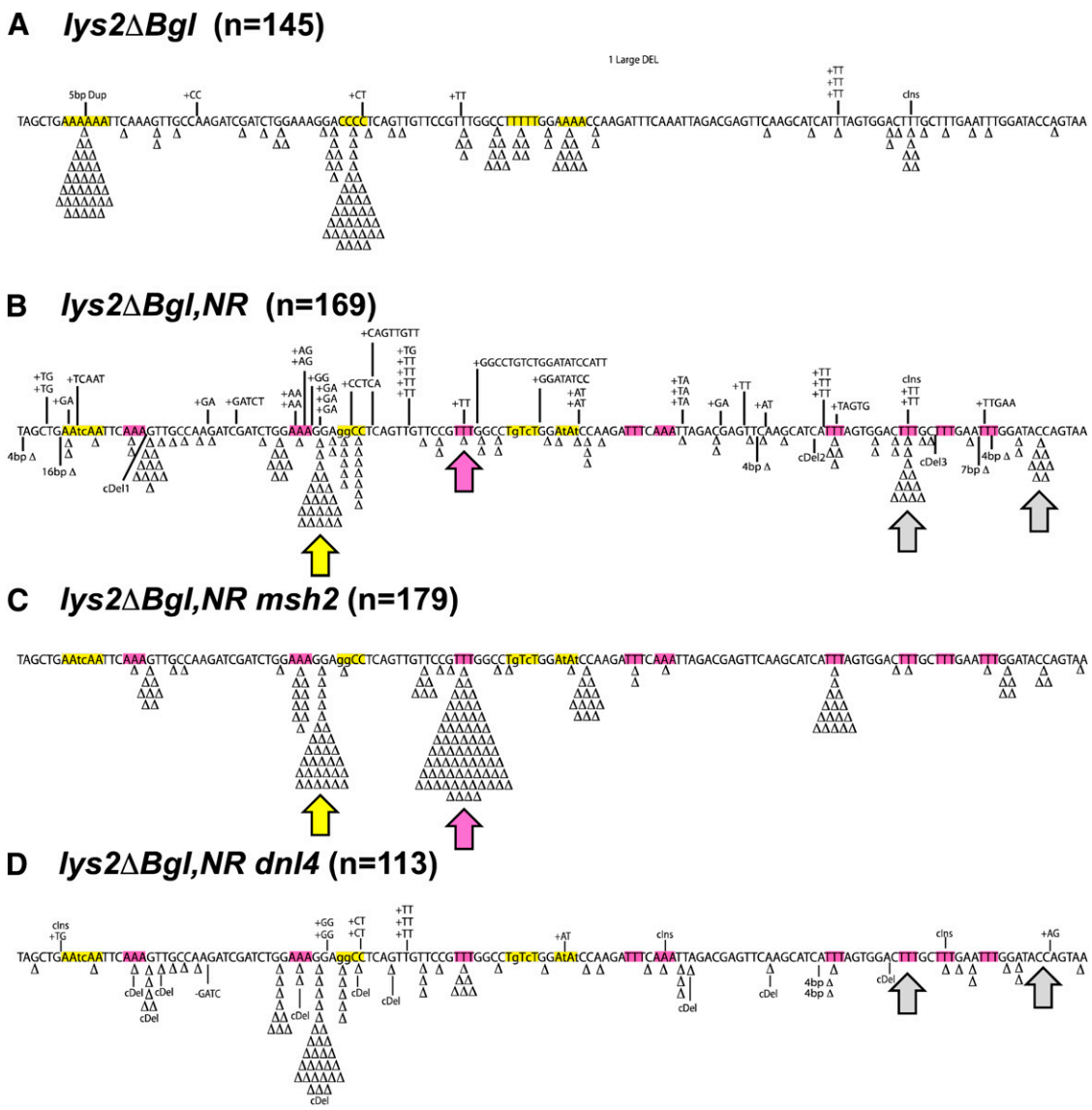

Figure 2 (A–D) *lys2ΔBgl,NR* reversion spectra. The theoretical reversion window on the coding strand is shown, with runs >3N (or the original positions of these runs) highlighted yellow and 3N runs highlighted pink. All deletions are below the sequence, with each "Δ" signifying loss of a single base pair. All insertions are above the sequence. Vertical arrows indicate specific hotspots that are described in the text. *n*, number of independent Lys$^+$ colonies sequenced; cins, complex 2-bp insertion with associate base substitution; cdel, complex 1-bp deletion; DEL, deletion. The WT spectrum was published previously (Greene and Jinks-Robertson 1997).

underrepresented, we concluded that the size threshold for spontaneous primer–template misalignment is 4N. In an *msh2Δ* background, which completely lacks the ability to recognize replication-generated mismatches (Kunkel and Erie 2005), the reversion rate of each allele was elevated several hundredfold and there was further skewing of events toward the longer runs. More than 98% of reversion events were in these runs, which completely obscured events that might be occurring at 3N runs, 2N runs, and noniterated sequence. To specifically examine these latter types of events, site-directed mutagenesis was used to remove the four mononucleotide runs >3N from the *lys2ΔBgl* and *lys2ΔA746* reversion windows (highlighted in yellow in Figures 2 and 3). We refer to the resulting alleles as *NR* alleles, although there remain multiple 3N and 2N runs within the region monitored. As in analyses with the original *lys2ΔBgl* and *lys2ΔA746* alleles, the *lys2ΔBgl,NR* and *lys2ΔA746,NR* alleles were located at the endogenous *LYS2* locus on chromosome II in all analyses reported here.

Reversion of the lys2ΔBgl,NR allele in a WT background

The reversion rate of the *lys2ΔBgl,NR* allele was approximately twofold lower than that of the original *lys2ΔBgl* allele (Table 1), consistent with the elimination of events in runs >3N. Similar to the *lys2ΔBgl* spectrum, the *lys2ΔBgl, NR* spectrum was dominated by simple, 1-bp deletions (121/169 = 72%), but a greater variety of additional

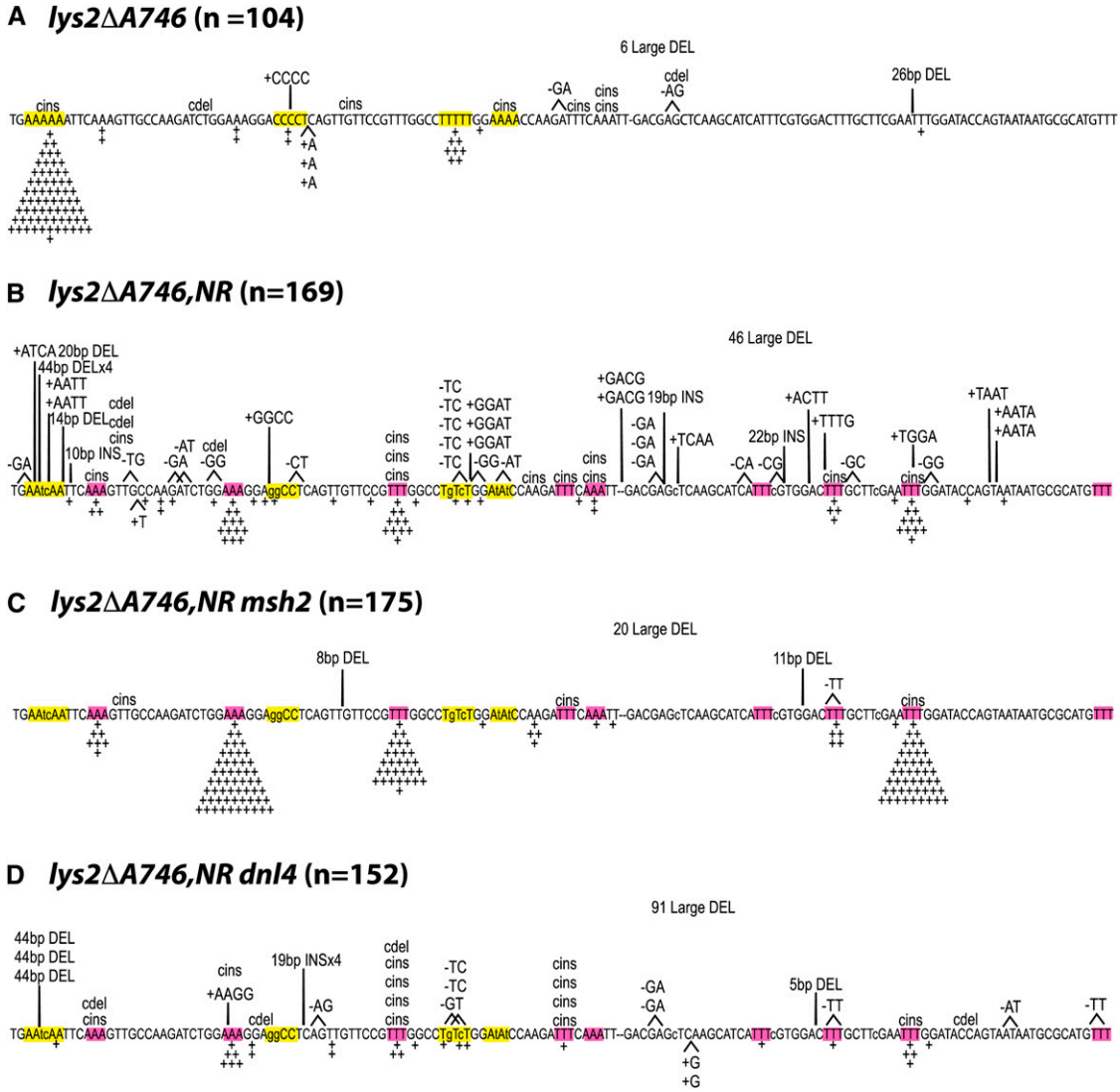

Figure 3 (A–D) *lys2ΔA746,NR* reversion spectra. The theoretical reversion window is shown, with runs >3N (or the original positions of these runs) highlighted yellow and 3N runs highlighted pink. All simple, 1-bp insertions are indicated by "+" and are below the sequence; all other mutation types are above the sequence. *n*, number of independent Lys$^+$ colonies sequenced; cins, complex 1-bp insertion; DEL, deletion. The WT spectrum was published previously (Harfe and Jinks-Robertson 1999).

mutation types and positions was evident (Figure 2, A and B). We expected that most 1-bp deletions in the *lys2ΔBgl,NR* spectrum would shift to the 3N runs (highlighted in pink), but only one of the nine 3N runs (indicated with the gray arrow) within the reversion window accumulated more −1 events than predicted by chance ($P < 0.0001$; expected number was based on proportion of reversion window occupied by the run). Although the overall number of events in the 3N runs (reflecting primarily events in a single 3T run) did not exceed that based on a random distribution of events ($P = 0.57$), there were many more 1-bp deletions in 2N runs ($P < 0.0001$) and many fewer events in noniterated sequence ($P = 0.003$) than expected. Almost 20% (22/121) of the 1-bp deletions occurred in a single 2G run (indicated by the yellow arrow), a run where only one event was observed in the *lys2ΔBgl* spectrum. We note that this 2G run is only 1 nt removed from the 4C run that was eliminated when constructing the *lys2ΔBgl,NR* allele (GGA*CCCC* changed to GGA*ggCC*), suggesting that local sequence context likely drives 2G hotspot activity. Even if one discounts the 2G hotspot, there was still an excess of 1-bp deletions within the remaining 2N runs ($P = 0.0005$).

Whereas sequence duplications were rare in the reversion spectrum of the *lys2ΔBgl* allele (7/145 = 5%), duplications of 2–20 bp accounted for 23% (39/169) of the *lys2ΔBgl,NR* spectrum. Significantly, more than half (23/39) of these duplications corresponded to the *de novo* creation of a repeat rather than the expansion of a preexisting repeat. Finally, there were a small number of events (9/169) within the *lys2ΔBgl,NR* spectrum that did not fall within either the duplication or the 1-bp deletion class, but these were too few in number to analyze in detail.

Table 1 Reversion of the *lys2ΔBgl,NR* allele in WT, *msh2Δ*, and *dnl4Δ* backgrounds

lys2 allele	Relevant genotype	Lys$^+$ rate $\times 10^{-9}$ (95% C.I.)	Lys$^+$ rate relative to *ΔBgl,NR* WT strain	Rate of individual mutation type relative to *ΔBgl,NR* WT strain					
				1-bp deletions				Small dups	Other events
				3N runs	2N runs	No run	Total		
ΔBgl	WT	2.94 (2.18–3.85)	1.6	1.3	0.47	1.5	0.84	0.24	0.43
ΔBgl,NR	WT	1.79 (1.33–2.36)	1.0	1.0	1.0	1.0	1.0	1.0	1.0
ΔBgl,NR	*msh2Δ*	32.2 (27.5–37.0)	18	86	15	7.9	24	ND	ND
ΔBgl,NR	*dnl4Δ*	0.97 (0.69–1.31)	0.54	0.44	0.72	0.32	0.58	0.21	1.44

C.I., confidence interval; dups, duplications; ND, none detected.

Removal of 1-bp deletion intermediates by the MMR machinery

In our previous analysis, elimination of Msh2 elevated the reversion rate of the *lys2ΔBgl* allele almost 200-fold, and all but one of 50 revertants analyzed contained a 1-bp deletion within the runs >3N (Greene and Jinks-Robertson 1997). While this demonstrated very efficient repair of −1 frameshift intermediates that arise in these runs, it was not clear whether other types of events seen in the WT background were simply repaired less efficiently or escaped MMR altogether. This was addressed by examining reversion of the *lys2ΔBgl,NR* allele in an *msh2Δ* background. Loss of Msh2 was associated with an 18-fold increase in reversion rate of the no-run allele (Table 1), a 10-fold smaller increase than observed with the original *lys2ΔBgl* allele.

In contrast to the diversity of mutation types observed in the WT background, all of the 179 *lys2ΔBgl,NR* revertants sequenced from the *msh2Δ* background contained a simple, 1-bp deletion event (Figure 2C). The 1-bp deletions localized to discrete hotspots, some of which were prominent both in the WT and in the MMR-defective backgrounds (*e.g.*, the 2G hotspot indicated with the yellow arrow in Figure 2, B and C) and some of which were evident only in the absence of MMR. For example, 70 events occurred at a single 3T run in the *msh2Δ* background (indicated by the pink arrow in Figure 2, B and C), whereas only one event was seen at this location in WT. The reverse pattern was also evident; the 3T run that was hottest in the WT background (gray arrow in Figure 2, B and D) contained only a single event in the *msh2Δ* background. Of the three 3A runs, one contained 10 events and the other two each contained only 1 event; of the six 3T runs, one contained 70 events, one contained 16 events, and the remaining four contained at most 2 events. Because the mutations that are elevated in an *msh2Δ* background presumably reflect replication errors, the data indicate that the probability of persistent primer–template misalignment varies dramatically between runs of the same size and composition.

Nonhomologous end joining produces small duplications in the lys2ΔBgl,NR assay

Given the large reversion-rate increase in the *msh2Δ* background, the absence of the small duplication class from the corresponding spectrum would be consistent either with dependence on functional MMR or with no change in rate. With regard to the former possibility, we previously reported that suppression of recombination by the MMR system promotes Polζ-dependent mutagenesis via the alternative translesion synthesis pathway, making such mutations dependent on functional MMR (Lehner and Jinks-Robertson 2009). We thus examined whether small duplications depend on the presence of Polζ. Deletion of the *REV3* gene, which encodes the catalytic subunit of Polζ (Nelson *et al.* 1996), neither affected the rate of *lys2ΔBgl,NR* reversion nor reduced the proportion of small duplications in the corresponding spectrum (data not shown).

The lack of an effect of Msh2 or Polζ loss on small duplications suggests that most are generated outside the context of DNA replication. Because tandem duplications (as well as deletions) can arise when double-strand breaks (DSBs) are repaired via the NHEJ pathway (Daley *et al.* 2005), we examined the effect of deleting the *DNL4* gene, which encodes the ligase required for NHEJ (Teo and Jackson 1997), on reversion of the *lys2ΔBgl,NR* allele. Relative to the WT background, the rate of *lys2ΔBgl,NR* reversion was reduced almost twofold in the *dnl4Δ* background (Table 1), and there were two notable changes in the reversion spectrum (Figure 2D). First, there was a significant reduction in duplications—from 39/169 mutations in WT to 10/113 in the *dnl4Δ* strain ($P = 0.001$). Second, there was a loss of simple deletions at two specific positions (indicated by gray arrows in Figure 2, B and D): the 3T hotspot noted previously in the WT background (16/179 *vs.* 1/113 events; $P = 0.002$), as well as a 2C run (8/169 *vs.* 0/113 events; $P = 0.016$). These data demonstrate that simple deletions within mononucleotide runs can result from error-prone end joining as well as from classical primer–template misalignment.

Reversion of the lys2ΔA746,NR allele in a WT background

The reversion rate of the *lys2ΔA746,NR* allele was approximately threefold lower than that of the original *lys2ΔA746* allele (Table 2), a decrease consistent with the loss of simple 1-bp insertions within the runs >3N (Figure 3A). Simple 1-bp insertions comprised 38% (64/169) of the *lys2ΔA746,NR* reversion spectrum and were primarily clustered in a subset of the 3N runs (Figure 3B; 3N runs are highlighted in pink). In addition to +1 events, the spectrum contained a large number of 2-bp deletions and 4-bp duplications (20 and

Table 2 Reversion of the *lys2ΔA746,NR* allele in WT, *msh2Δ*, *dnl4Δ* and *tdp1Δ* backgrounds

				Rate of individual mutation type relative to *ΔA746,NR* WT strain							
				1-bp insertions							
lys2 allele	Repair genotype	Lys$^+$ rate $\times 10^{-9}$ (95% C.I.)	Lys$^+$ rate relative to *ΔA746,NR* WT strain	3N runs	2N runs	No run	Total	4-bp dups	2-bp dels	Large dels	Other events
ΔA746	WT	5.74 (4.66–6.95)	2.9	0.53	0.33	ND	0.52	0.29	0.46	0.61	2.3
ΔA746,NR	WT	2.00 (1.72–2.31)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
ΔA746,NR	*msh2Δ*	13.3 (12.1–14.7)	6.7	21	3.2	1.3	15	ND	0.32	2.8	1.3
ΔA746,NR	*dnl4Δ*	1.93 (1.51–2.43)	0.97	0.39	0.53	0.64	0.44	0.07	0.49	2.1	1.1
ΔA746,NR	*tdp1Δ*	2.12 (1.49–2.90)	1.1	0.55	1.3	0.32	0.69	0.40	1.2	1.8	0.94

C.I., confidence interval; dels, deletions; dups, duplications; ND, none detected.

16 events, respectively), neither of which was associated with repetitive sequence elements. Finally, large (144 bp) deletions accounted for a much larger proportion of the *lys2ΔA746,NR* than of the original *lys2ΔA746* spectrum (46/169 and 6/104, respectively), which is consistent with the Lys$^+$ rate differences. These large deletions have endpoints in 10-bp direct repeats and are affected by the direction of DNA replication (Abdulovic *et al.* 2007), suggesting that most reflect repeat-mediated realignment of a blocked 3′ end during replication.

The MMR system efficiently removes +1 frameshift intermediates in 3N runs

Deletion of the *MSH2* gene was associated with a 6.7-fold increase in the reversion rate of the *lys2ΔA746,NR* allele (Table 2). This increase was accompanied by a proportional increase in +1 events in the corresponding spectrum: from 38% in the WT background to 85% (149/175) in the MMR-defective background (Figure 3C). Most of the simple +1 events were within only three of the nine 3N runs, however, again suggesting that the frequency of replication-associated strand misalignment within individual runs is highly variable. As reported previously, the rate of large deletions was also elevated 3- to 4-fold upon loss of MMR (Harfe *et al.* 2000). In contrast to the increases in 1-bp insertion and large-deletion rates upon loss of MMR, the 2-bp deletion and 4-bp duplication classes were almost completely absent in the *msh2Δ* background.

Loss of NHEJ alters the lys2ΔA746,NR reversion spectrum

Given the dependence of 2-bp insertions on NHEJ in the *lys2ΔBgl,NR* assay, we examined the relevance of this pathway to the 2-bp deletion and 4-bp duplication classes detected in the *lys2ΔA746,NR* assay. Deletion of *DNL4* did not change the overall reversion rate of the *lys2ΔA746,NR* allele (Table 2), but it did significantly alter the reversion spectrum in several important ways (Figure 3D). Significant decreases in 2-bp deletions ($P = 0.048$) and especially 4-bp tandem duplications ($P < 0.001$) were associated with Dnl4 loss, indicating that both types of events are predominantly produced via NHEJ. There was also a decrease in the proportion of 1-bp insertions ($P < 0.001$), with reductions being distributed across the spectrum rather than concentrated in specific locations. Finally, there was a twofold proportional increase in the large deletion class ($P < 0.001$), indicating that, in addition to a DNA polymerase-based realignment mechanism, large deletions with endpoints in direct repeats can result from a DSB repair mechanism that is an alternative to NHEJ. We suggest that the single-strand annealing pathway, which specifically generates deletions between direct repeats (Symington 2002), is the most likely NHEJ alternative.

In a plasmid-based NHEJ assay, 4-bp duplications arise at a low frequency following transformation with linear molecules containing complementary, 4-nt 5′ overhangs. Such events are specifically elevated in the absence of Tdp1, a 3′ nucleosidase whose action presumably blocks the filling in of the recessed ends (Bahmed *et al.* 2010). We thus examined whether loss of Tdp1 affects reversion of the *lys2ΔA746,NR* allele. Neither the total rate of Lys$^+$ revertants nor the proportion of 4-bp duplications in the corresponding spectrum was elevated in a *tdp1Δ* background (Table 2).

Discussion

In this study, we have used the complementary *lys2ΔA746, NR* and *lys2ΔBgl,NR* alleles to identify net +1 and −1 frameshift mutations, respectively, within a common, ∼150-bp segment of yeast genomic DNA. A key feature of the region monitored is that it contains no mononucleotide runs >3N, thereby allowing detection of rare indels and other mutation types that are normally masked by frequent, spontaneous slippage in longer runs. In a WT background, the total rates of 1-bp insertions *vs.* 1-bp deletions were similar in the region monitored, but their distributions were very different. This is evident in the compiled spectrum presented in Figure 4A, where events in the eight common 3N runs are highlighted pink to facilitate comparisons. Whereas 70% of 1-bp insertions were in 3N runs, <20% of 1-bp deletions were in these runs. The 1-bp deletions were not randomly distributed, however, but clustered at several 2N hotspots (highlighted in yellow).

Mutations elevated upon loss of MMR reflect errors made by the replicative DNA polymerases that fail to be removed by the associated proofreading activity. In an *msh2Δ* background, 1-bp insertion and deletion rates increased 15- and

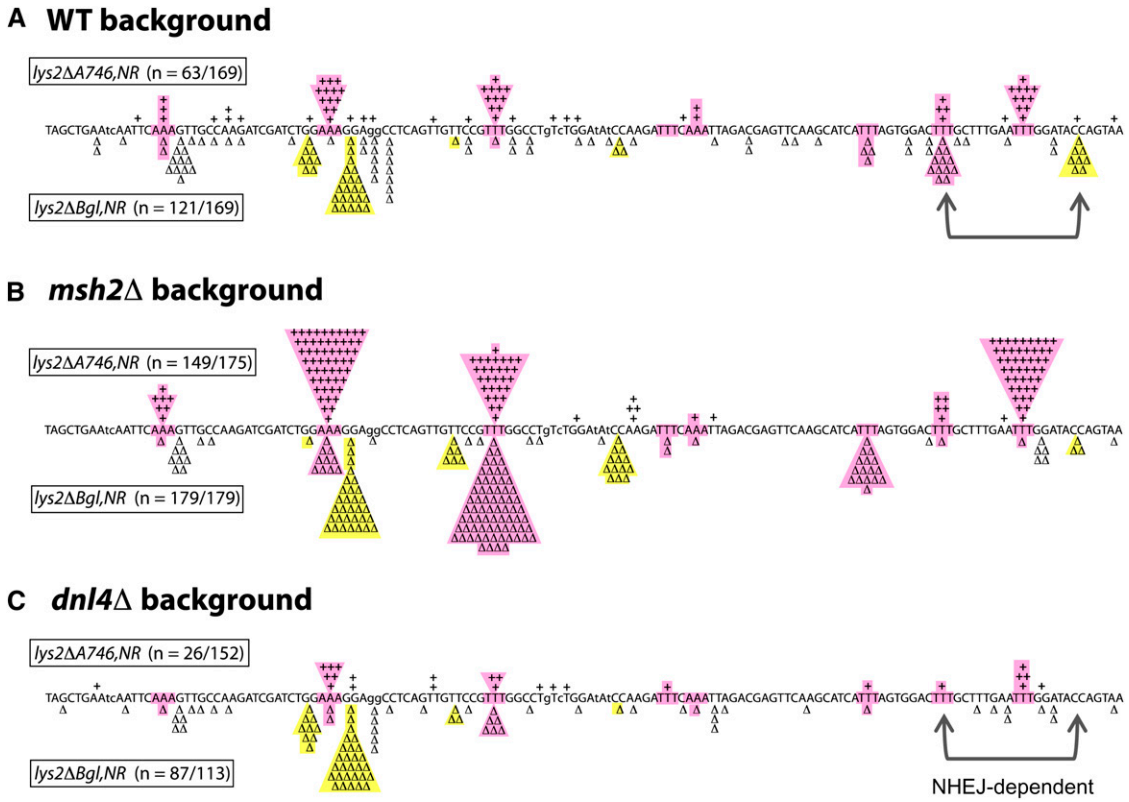

Figure 4 (A–C) Comparison of simple 1-bp indels in the *lys2ΔBgl,NR* and *lys2ΔA746,NR* reversion spectra. The sequence common to both reversion windows is shown. Insertions (+) and deletions (Δ) are above and below the sequence, respectively. 3N runs as well as indels at these positions are highlighted pink; select 2N hotspots are highlighted yellow. *n*, proportion of indels among revertants sequenced.

24-fold, respectively; relative to the WT strain, there was an enrichment of each within 3N runs. Although this demonstrates that 3N runs can promote primer–template misalignment during replication, there was dramatic run-to-run variation with respect to the accumulation of +1 and/or −1 events (Figure 3B). Two of the 3N runs were hotspots for insertions and deletions, one accumulated only insertions, and one accumulated only deletions. Because of the strong context effects observed, we suggest that 1-bp indels in these small runs are most likely derived from misinsertion/primer relocation or dNTP-stabilized misalignment rather than from spontaneous primer–template misalignment. Misinsertion/relocation is expected to generate both 1-bp insertions and deletions, while dNTP-stabilized misalignment is predicted to produce only 1-bp deletions. The accumulation of 1-bp insertions and deletions in 3N runs, but only 1-bp deletions in 2N runs, is intriguing and may indicate that 3N is the lower threshold for misinsertion/relocation. An alternative explanation for the highly variable distribution of the 1-bp indels among 3N runs in the *msh2Δ* background is that the efficiency of polymerase-asssociated proofreading is dependent on local sequence context.

Changes in the spectra of spontaneous 1-bp indels upon elimination of MMR are most simply interpreted as site-specific differences in the efficiency of MMR. The efficiency of MMR could be affected, for example, by glycosylase-associated shielding of extrahelical nucleotides (Klapacz *et al.* 2010). An alternative possibility, however, is the existence of additional mutagenic processes that act outside the context of DNA replication and/or do not generate mismatch-containing intermediates. Indeed, data from the *dnl4Δ* background indicate that ∼50% of the 1-bp indels in a WT background are generated via the NHEJ pathway. Although there appeared to be a general deficit of 1-bp insertions at all positions, two examples of NHEJ-dependent, 1-bp deletion hotspots were evident in the *lys2ΔBgl, NR* assay (indicated by the arrows in Figure 4). Such NHEJ-associated deletions presumably reflect the removal of nucleotides from one or both ends of the initiating DSB, which may or may not be associated with inappropriate annealing between overhangs and gap-filling reactions. The possible origins of NHEJ-generated insertions as well as duplications are discussed in more detail below.

In addition to facilitating examination of 1-bp indels in very short mononucleotide runs and noniterated sequence, use of the complementary *lys2ΔBgl,NR* and *lys2ΔA746,NR* alleles allowed the efficient detection of larger insertions and deletions. In the *lys2ΔBgl,NR* spectrum, *de novo* tandem duplications, most of which were 2 bp, were frequent and were clearly NHEJ dependent. In addition to *de novo* duplications, there were a small number of the 2-bp insertions that expanded a preexisting repeat. Similar insertions in

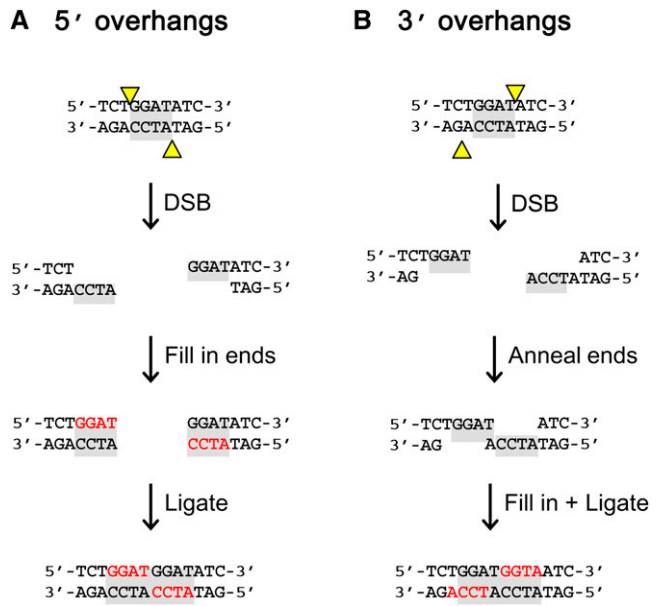

Figure 5 (A and B) Generation of tandem duplications by NHEJ. The original DNA is black and newly synthesized DNA is red. Yellow triangles indicate the positions of nicks that create DSBs; the 4-bp sequence duplicated is highlighted gray.

mononucleotide runs were previously reported among *lys2ΔBgl* revertants isolated in one WT strain background (Heidenreich *et al.* 2003), but this particular class was not observed in at least two other backgrounds (Marsischky *et al.* 1996; Greene and Jinks-Robertson 1997). In the *lys2ΔA746,NR* assay, 2-bp deletions and 4-bp tandem duplications each comprised ~10% of the reversion spectrum, and each class was significantly reduced in the *dnl4Δ* background.

The tandem, 4-bp duplications seen here are of particular interest as they are similar to those recently reported using a plasmid-based NHEJ assay (Bahmed *et al.* 2010, 2011). Because such duplications were observed only following transformation of linear molecules with cohesive 5′ overhangs, it was proposed that they are generated by the precise ligation of filled-in, blunt ends (Figure 5A). In the plasmid-based assay, tandem duplications were elevated upon loss of either Tdp1 (Bahmed *et al.* 2010) or Exo1 (Bahmed *et al.* 2011). It was suggested that the 3′-nucleosidase activity of Tdp1 converts the recessed 3′-OH to a recessed 3′-phosphate, thereby preventing the filling in of the enzyme-generated end (Bahmed *et al.* 2010). In the case of Exo1, either its 5′ > 3′-exonuclease or its 5′-flap endonuclease activity could remove the complementary sequence following the fill-in reaction (Bahmed *et al.* 2011). Although a similar, end-filling mechanism could be generating tandem duplications in the *lys2ΔA746,NR* assay, we saw no increase in these events in either a *tdp1Δ* or an *exo1Δ* background (Table 2 and data not shown). This could reflect a plasmid–chromosome difference in how similar ends are processed (*e.g.*, the ends of spontaneous chromosomal breaks are not accessible to Tdp1 or Exo1), but we think it more likely that the ends are different. Duplications of the sort seen here can be generated, for example, by a misannealing of 3′ (or 5′) overhangs, followed by the filling in of gaps (Figure 5B). This type of mechanism has been proposed to explain the creation of small duplications following the cleavage of yeast genomic DNA with the HO endonuclease, which creates 4-nt, 3′ overhangs (Moore and Haber 1996).

Spontaneous primer–template misalignment requires at least two copies of a repeat unit and so expands only preexisting repeats. Although the alternative misalignment models presented in Figure 1, B and C, are, in principle, capable of creating 2N mononucleotide runs from noniterated sequence, they cannot be used to generate larger repeat units. For repeat units ≥2 bp, NHEJ can provide a mechanism for creating tandem duplications from noniterated sequence. Bioinformatic studies support this type of mechanism for the origin of microsatellites (Zhu *et al.* 2000; Leclercq *et al.* 2010), and data presented here demonstrate that NHEJ-mediated duplications do indeed arise spontaneously in yeast genomic DNA. Finally, we note that NHEJ could provide a mechanism for adding (or deleting) multiple repeat units in a single step. This could, for example, contribute to trinucleotide expansions and may be especially relevant in slow-growing or post-mitotic cells.

While the frequency of primer–template misalignment within mononucleotide runs is strongly correlated with the number of repeat units *in vivo* (Tran *et al.* 1997), whether a lower threshold exists has been unclear. While early studies using the original *lys2ΔBgl* and *lys2ΔA746* frameshift-reversion assays suggested that 4N is the likely threshold in yeast (Greene and Jinks-Robertson 1997; Harfe and Jinks-Robertson 1999), more recent bioinformatic studies have concluded that even 2N is sufficient for slippage in yeast (Pupko and Graur 1999) as well as humans (Leclercq *et al.* 2010). By limiting the current analyses to a region where there are no mononcleotide runs >3N, we have been able to confirm that smaller repeats can be hotspots for indels in yeast, but are not universally so. Importantly, we have shown that the replication-independent mechanism of NHEJ also contributes to 1-bp indels in very short runs and additionally provides a mechanism for the *de novo* creation of tandem duplications of variable size. Given the high conservation of DNA metabolic processes, the results obtained in the yeast system will likely be of relevance to issues of genome stability and evolution in higher eukaryotes.

Acknowledgment

This work was supported by grant GM038464 from the National Institutes of Health (to S.J.-R.).

Literature Cited

Abdulovic, A. L., B. K. Minesinger, and S. Jinks-Robertson, 2007 Identification of a strand-related bias in the PCNA-mediated

bypass of spontaneous lesions by yeast Polη. DNA Repair (Amst.) 6: 1307–1318.
Bahmed, K., K. C. Nitiss, and J. L. Nitiss, 2010 Yeast Tdp1 regulates the fidelity of nonhomologous end joining. Proc. Natl. Acad. Sci. USA 107: 4057–4062.
Bahmed, K., A. Seth, K. C. Nitiss, and J. L. Nitiss, 2011 End-processing during non-homologous end-joining: a role for exonuclease 1. Nucleic Acids Res. 39: 970–978.
Bebenek, K., and T. A. Kunkel, 1990 Frameshift errors initiated by nucleotide misincorporation. Proc. Natl. Acad. Sci. USA 87: 4946–4950.
Daley, J. M., P. L. Palmbos, D. Wu, and T. E. Wilson, 2005 Nonhomologous end joining in yeast. Annu. Rev. Genet. 39: 431–451.
Efrati, E., G. Tocco, R. Eritja, S. H. Wilson, and M. F. Goodman, 1997 Abasic translesion synthesis by DNA polymerase β violates the "A-rule". Novel types of nucleotide incorporation by human DNA polymerase β at an abasic lesion in different sequence contexts. J. Biol. Chem. 272: 2559–2569.
Garcia-Diaz, M., and T. A. Kunkel, 2006 Mechanism of a genetic glissando: structural biology of indel mutations. Trends Biochem. Sci. 31: 206–214.
Greene, C. N., and S. Jinks-Robertson, 1997 Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. Mol. Cell. Biol. 17: 2844–2850.
Greene, C. N., and S. Jinks-Robertson, 2001 Spontaneous frameshift mutations in *Saccharomyces cerevisiae*: accumulation during DNA replication and removal by proofreading and mismatch repair activities. Genetics 159: 65–75.
Gueldener, U., J. Heinisch, G. J. Koehler, D. Voss, and J. H. Hegemann, 2002 A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Res. 30: e23.
Harfe, B. D., and S. Jinks-Robertson, 1999 Removal of frameshift intermediates by mismatch repair proteins in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 19: 4766–4773.
Harfe, B. D., B. K. Minesinger, and S. Jinks-Robertson, 2000 Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast. Curr. Biol. 10: 145–148.
Heidenreich, E., R. Novotny, B. Kneidinger, V. Holzmann, and U. Wintersberger, 2003 Non-homologous end joining as an important mutagenic process in cell cycle-arrested cells. EMBO J. 22: 2274–2283.
Klapacz, J., G. M. Lingaraju, H. H. Guo, D. Shah, A. Moar-Shoshani *et al.*, 2010 Frameshift mutagenesis and microsatellite instability induced by human alkyladenine DNA glycosylase. Mol. Cell 37: 843–853.
Kunkel, T. A., 1990 Misalignment-mediated DNA synthesis errors. Biochemistry 29: 8003–8011.
Kunkel, T. A., and D. A. Erie, 2005 DNA mismatch repair. Annu. Rev. Biochem. 74: 681–710.
Leclercq, S., E. Rivals, and P. Jarne, 2010 DNA slippage occurs at microsatellite loci without minimal threshold length in humans: a comparative genomic approach. Genome Biol. Evol. 2: 325–335.
Lehner, K., and S. Jinks-Robertson, 2009 The mismatch repair system promotes Polζ-dependent translesion synthesis in yeast. Proc. Natl. Acad. Sci. USA 106: 5749–5754.
Longtine, M. S., A. Mckenzie Iii, D. J. Demarini, N. G. Shah, A. Wach *et al.*, 1998 Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. Yeast 14: 953–961.
Marsischky, G. T., N. Filosi, M. F. Kane, and R. Kolodner, 1996 Redundancy of *Saccharomyces cerevisiae MSH3* and *MSH6* in *MSH2*-dependent mismatch repair. Genes Dev. 10: 407–420.
Moore, J. K., and J. E. Haber, 1996 Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 16: 2164–2173.
Nelson, J. R., C. W. Lawrence, and D. C. Hinkle, 1996 Thymine-thymine dimer bypass by yeast DNA polymerase ζ. Science 272: 1646–1649.
Pupko, T., and D. Graur, 1999 Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. J. Mol. Evol. 48: 313–316.
Rothstein, R., 1991 Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. 194: 281–301.
Shah, S. N., S. E. Hile, and K. A. Eckert, 2010 Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes. Cancer Res. 70: 431–435.
Steele, D. F., and S. Jinks-Robertson, 1992 An examination of adaptive reversion in *Saccharomyces cerevisiae*. Genetics 132: 9–21.
Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita *et al.*, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. 31: 77–84.
Symington, L. S., 2002 Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. Microbiol. Mol. Biol. Rev. 66: 630–670.
Teo, S. H., and S. P. Jackson, 1997 Identification of *Saccharomyces cerevisiae* DNA ligase IV: involvement in DNA double-strand break repair. EMBO J. 16: 4788–4795.
Tran, H. T., J. D. Keen, M. Kricker, M. A. Resnick, and D. A. Gordenin, 1997 Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol. Cell. Biol. 17: 2859–2865.
Zheng, Q., 2005 New algorithms for Luria-Delbruck fluctuation analysis. Math. Biosci. 196: 198–214.
Zhu, Y., J. E. Strassmann, and D. C. Queller, 2000 Insertions, substitutions, and the origin of microsatellites. Genet. Res. 76: 227–236.

Communicating editor: E. Alani