High-resolution mapping of two types of spontaneous mitotic gene conversion events in

Saccharomyces cerevisiae

Eunice Yim*, Karen E. O’Connell*, Jordan St. Charles*¹, and Thomas D. Petes*

*Department of Molecular Genetics and Microbiology, Duke University School of Medicine,
Durham, NC 27710

¹Current address: Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

Copyright 2014.
Running title: Gene conversion in yeast

Key phrases: yeast, mitotic recombination, gene conversion, DNA damage repair, loss of heterozygosity

Corresponding author:

Thomas D. Petes

Department of Molecular Genetics and Microbiology

Duke University School of Medicine

213 Research Drive

Durham, NC 27710

Phone: 919 684-4986

Email: tom.petes@duke.edu
Gene conversions and crossovers are related products of the repair of double-stranded DNA breaks by homologous recombination. Most previous studies of mitotic gene conversion events have been restricted to measuring conversion tracts that are less than 5 kb. Using a genetic assay in which the lengths of very long gene conversion tracts can be measured, we detected two types of conversions, those that have a median size of about 6 kb and those with a median size of more than 50 kb. The unusually long tracts are initiated at a naturally-occurring recombination hotspot formed by two inverted Ty elements. We suggest that these long gene conversion events may be generated by a different mechanism (break-induced replication or repair of a double-stranded DNA gap) than the short conversion tracts that likely reflect heteroduplex formation followed by DNA mismatch repair. Both the short and long mitotic conversion tracts are considerably longer than those observed in meiosis. Since mitotic crossovers in a diploid can result in a heterozygous recessive deleterious mutation becoming homozygous, it has been suggested that the repair of DNA breaks by mitotic recombination involves gene conversion events that are unassociated with crossing over. In contrast to this prediction, we found that about 40% of the conversion tracts are associated with crossovers. Spontaneous mitotic crossover events in yeast are frequent enough to be an important factor in genome evolution.
INTRODUCTION

In the yeast *Saccharomyces cerevisiae*, double-stranded DNA breaks (DSBs) in mitotically-dividing cells are usually repaired by homologous recombination using an unbroken DNA molecule as a template (Symington, 2002). In G2 diploid cells, sister chromatids are the favored substrate for repair rather than the homolog (Kadyk and Hartwell, 1992). Associated with the repair of most DSBs, there is a region of DNA transferred non-reciprocally from the unbroken chromosome to the broken chromosome; when this process alters the sequence of the recipient chromosome, it is called “gene conversion” (Petes et al., 1991; Paques and Haber, 1999). In the current models of recombination, conversion events are produced through several different pathways (Fig. 1). In one pathway (synthesis-dependent strand annealing, SDSA), one broken end invades the homologous template and primes DNA synthesis. The invading end is then ejected from the template and reanneals to the other broken end (Fig. 1A). Mismatches within the resulting heteroduplex (boxed region in figure) can be repaired to generate the conversion event. This pathway produces conversions unassociated with crossovers (NCO conversions). An NCO conversion can also be generated by dissolution of a double Holliday junction (dHJ) intermediate. Cleavage of the dHJ can produce either a conversion event associated with a crossover (CO conversion) or an NCO conversion (Fig. 1B). The resolution of the dHJ is biased toward the crossover, rather than the non-crossover pathway, in mitosis (Mitchel et al., 2010). Lastly, a DSB can be repaired by a non-reciprocal process in which one broken end invades and copies the homolog to the end of the chromosome, and the other chromosome end is lost. This pathway is break-induced replication (BIR, Fig. 1C) (Llorente et al., 2008). In *S. cerevisiae*, in both meiosis (Allers and Lichten, 2001) and mitosis (Mitchel et al., 2010), most NCO gene conversion events reflect the SDSA pathway rather than processing of a dHJ.

Spontaneous mitotic recombination events are about $10^4$-fold less frequent than meiotic events (Esposito and Wagstaff, 1981). Consequently, analysis of spontaneous mitotic gene conversion events requires a selective system or induction of events by DNA damage or site-specific meganucleases such as HO or I-SceI (Paques and Haber, 1999). The most common method of detecting spontaneous
conversions is to use auxotrophic heteroalleles. For example, in one study in which one homolog had a
leu2-K allele and the other a leu2-R allele, Leu⁺ conversion events were detected at a rate of about 1 x
10⁻⁷/division (Lichten and Haber, 1989). The rates of conversion in most heteroallelic studies on
homologs vary between 10⁻⁷ to 10⁻⁶/division (Jinks-Robertson and Petes, 1986; Steele et al., 1991;
Nickoloff et al., 1999; Lettier et al., 2006). In one study, conversion rates involving heteroallelic genes
located close together were higher, 4 x 10⁻⁶/division (Aguilera and Klein, 1989).

Previous studies of the lengths of mitotic gene conversion tracts have yielded disparate results. In a
study of mitotic gene conversion between homologs, Judd and Petes (1986) found that about 20% of
conversion tracts were less than 2 kb, but at least 40% were greater than 4 kb. In this study, conversion
tracts >4 kb could not be measured because of the lack of flanking markers. Similarly, in a study of HO-
induced recombination, Nickoloff et al. (1999) found that about 40% of the conversion tracts were 2 kb
or less. The other conversion tracts were at least 3.3 kb in length, but the lack of flanking markers
prevented an accurate determination of tract length. In a plasmid-chromosome recombination assay
(Mitchel et al., 2010), NCO conversions were usually smaller than 400 bp and CO conversions often
extended to the limits of the homology, a distance of >500 bp from the initiating DNA lesion. In two
other studies in which conversion events were classified as either short (< 1 kb) or long (>1 kb), long
tracts were more frequently associated with crossovers (Aguilera and Klein, 1989; Ho et al., 2010).

The use of heteroalleles to monitor the frequency of conversion and the length of conversion tracts
has a substantial limitation. If conversion is a consequence of heteroduplex formation followed by
mismatch repair and if mismatches in a heteroduplex are usually repaired using the same template
strand, a conversion event that includes both mismatches will not generate a wild-type allele. Thus,
estimates of the frequency of conversion based on heteroalleles will usually underestimate the
frequency of heteroduplex formation, and will underestimate tract size. We have recently used
microarrays or related methods to measure the conversion tract lengths associated with recombination
between homologs using a system that does not involve heteroalleles (Lee et al., 2009; St. Charles et
al., 2012; St. Charles and Petes, 2013; Yin and Petes, 2013). Most of these studies examined CO
conversion tracts. The median size of spontaneous CO conversion tracts was about 10 kb; NCO conversion tracts were not examined in these studies. In a study of UV-induced recombination events, NCO tracts were significantly shorter than CO tracts with median sizes of 4.9 kb and 7.6 kb, respectively (Yin and Petes, 2013).

One controversial issue concerning mitotic gene conversion in yeast is the relative frequencies of CO and NCO conversion events. Mitotic crossing over has the negative effects of allowing harmful recessive mutations to become homozygous, and ectopic crossovers cause changes in chromosome structure. It has been suggested, therefore, that most mitotic repair events should be associated with NCO conversions rather than CO conversions. In yeast, however, the percentage of CO conversion events varies greatly in different studies: 10% (Jinks-Robertson and Petes, 1986), 12% (Inbar and Kupiec, 1999), 17% (Chua and Jinks-Robertson, 1991), 20% (Nickoloff et al., 1999), 25% (Malkova et al., 1996), 25% (Haber and Hearn, 1985), 33% (Yin and Petes, 2013), 50% (Aguilera and Klein, 1989), and 50% (Welz-Voegele and Jinks-Robertson, 2008). Most of these studies involved heteroallelic recombination, DNA damage-induced events, and/or ectopic recombination.

Below, we describe experiments to measure the frequency of spontaneous gene conversion between homologous chromosomes, the lengths of gene conversion tracts, and the fraction of CO and NCO conversions. We utilize a system that allows detection of very long (>100 kb) conversion tracts and that does not have the biases associated with heteroallelic conversion.
MATERIALS AND METHODS

Strain construction and genetic methods: Experiments were conducted using the diploid strain EY7. The construction of EY7 is described in Table S1, and primers used for strain construction and sequence analysis are listed in Table S2. EY7 was generated by mating two haploids isogenic with W303-1A and YJM789. The resulting diploid is heterozygous for about 55,000 single-nucleotide polymorphisms (Lee et al., 2009). These polymorphisms allow the mapping of gene conversion events. Aside from these markers, the relevant genotype of EY7 is: MATa/MATα::natMX4 leu2-3,112/LEU2 his3-11,15/HIS3 ade2-1/ade2-1 ura3-1/ura3-p trp1-1/TRP1 can1-100/CAN1 GAL2/gal2 ho/ho::hisG IV957578::hphMX4/IV957578 IV1013217::URA3/IV1013217::ura3-e IV1510386::SUP4-o/IV1510386. The mutations in ura3-1, ura3-p, and ura3-e are: G to A at position 701 of URA3, G to A change at position 608, and T to A alteration at position 170, respectively.

Transformation, mating, media preparation, and tetrad dissection were performed according to standard methods (Guthrie and Fink, 1991). Cells were grown at 30°C. As explained in the Results section, a complete analysis of the conversion events in EY7 required meiotic analysis. Although EY7 has a disruption of the MATα locus, which prevents sporulation under normal conditions, sporulation can be induced in sporulation medium with 5mM nicotinamide (St Charles and Petes, 2013). The sporulation medium also contained 25 µg/ml uracil.

Ultraviolet radiation (UV) was used in one experiment (described elsewhere) to determine coupling relationships of markers on the two homologs. For these experiments, cells from overnight cultures were placed on solid medium (modified SD-Arg, Barbera and Petes, 2006) at a density of about 1000 cells/plate. The cells were then treated with UV (15 J/m² using a TL-2000 Translinker) and allowed to form colonies. About 20 red/white sectored colonies were obtained per plate. Their analysis is described in the Results section.

Identification of strains with gene conversion events at the URA3 gene on chromosome IV: A gene conversion event in which information is transferred between ura3-e and the wild-type URA3 gene on chromosome IV in EY7 results in derivatives that are resistant to 5-fluoro-orotate (5-FOA<sup>R</sup>),
hygromycin-resistant (Hyg\textsuperscript{R}), tryptophan prototrophs (Trp\textsuperscript{+}), and form pink colonies. To select 5-FOA\textsuperscript{R} derivatives of EY7, we made patches of individual colonies of EY7 grown on rich medium (YPD) on solid medium containing 5-FOA (Boeke \textit{et al.}, 1987). We purified the resulting 5-FOA\textsuperscript{R} colonies non-selectively, and determined their phenotypes by replica-plating to media lacking uracil or tryptophan, or containing hygromycin. Only one 5-FOA\textsuperscript{R} colony per patch was examined to ensure that the observed events were independent. The color of the colony was determined by allowing the colonies to grow for three days at 30°C, followed by one day-incubation at 4°C. We also measured the rate of 5-FOA\textsuperscript{R} derivatives by measuring the frequency of 5-FOA\textsuperscript{R} in thirty-nine independent cultures, and then calculating the rate by using the method of the median (Lea and Coulson, 1949).

Use of microarrays to measure conversion tract length: Mitotic gene conversion events in EY7 result in an interstitial region of loss of heterozygosity (LOH) flanked by heterozygous markers. To determine the extent of LOH, we used DNA microarrays capable of distinguishing whether the strain was homozygous for a single-nucleotide polymorphism (SNP) specific for the W303-1A-derived homolog, homozygous for a SNP specific for YJM789-derived homolog, or heterozygous. The design of a microarray to detect LOH for SNPs between W303-1A and YJM789 was described in detail by St Charles \textit{et al.}, 2012. In brief, four 25-base oligonucleotides centered on each SNP of interest were designed, two representing Watson and Crick sequences for one homolog and two representing Watson and Crick sequences for the other. Under the appropriate hybridization conditions, the detected hybridization of the genomic DNA to the perfectly-matched oligonucleotides is stronger than the detected hybridization to oligonucleotides that have a mismatch, allowing detection of patterns of LOH. Additional details about the microarray analysis are provided in File S1 of the Supporting Information.

Of the 59 samples isolated from 5-FOA\textsuperscript{R} Hyg\textsuperscript{R} Trp\textsuperscript{+} pink colonies, most had LOH events involving at least five adjacent SNPs. For those samples with no LOH event detectable by microarrays, we sequenced the \textit{ura3} gene on the W303-1A-derived chromosome IV. For this analysis, we sporulated the diploid, and isolated DNA from spore cultures. We identified which spores contained the W303-1A-
derived homolog (as described in the Results section). We amplified the *ura3* gene on chromosome IV using the primer pairs URA3-EY F and URA3-EY R (Table S2), and sequenced the resulting product.

The microarrays used for mapping do not contain all of the SNPs that distinguish the two homologs. For conversion events that had a breakpoint in a region sparsely represented by oligonucleotides on the microarray, we refined the mapping using a different method (Lee et al., 2009). This method is described in File S1 and Table S3 of Supporting Information.
RESULTS

The purposes of these experiments are: 1) to measure the frequency of mitotic gene conversion in a system that is not limited by the length of sequence homology, 2) to determine what fraction of mitotic gene conversion events are associated with crossovers, and 3) to analyze the length distribution of mitotic gene conversion tracts. The gene conversion events in the current study fall into two categories, those with a median length of about 6 kb and those with a median length of > 50 kb. Even the short class of conversions is longer than most of the tracts described by previous studies. As discussed in the Introduction, the short tracts observed in previous studies likely reflect the use of heteroalleles or systems involving ectopic recombination between repeats with limited homology.

Experimental rationale: The strain used for this analysis is shown in Figure 2. The diploid EY7 was constructed by crossing two sequenced-diverged haploids (W303-1A and YJM789) resulting in a diploid that is heterozygous for about 55,000 single-nucleotide polymorphisms (SNPs). Similar strains have been used previously to map spontaneous and UV-induced mitotic crossovers (Lee et al., 2009; St. Charles et al., 2012; St. Charles and Petes, 2013; Yin and Petes, 2013). In our experiments, we monitor conversion and crossover events on chromosome IV. In our description of the approximate location of diagnostic markers or the boundaries of recombination breakpoints, we use Saccharomyces Genome Database [SGD] coordinates. The first base on the left end of chromosome IV is assigned the coordinate 1 bp and the last base on the right end is coordinate 1531933 bp. The diploid is heterozygous or hemizygous for the following markers (SGD coordinate of the mid-point of the gene or position of insertion rounded to a kb value): TRP1 (462 kb), hphMX4 which encodes an antibiotic that confers resistance to hygromycin (958 kb), URA3 (1013 kb), and SUP4-o (1510 kb); the centromere is located at SGD coordinate 450 kb. The diploid is also homozygous for the ade2-1 ochre mutation. Diploids with the ade2-1 mutation form white, pink, and red colonies in the presence of two, one or no copies of the SUP4-o ochre suppressor gene, respectively (Barbera and Petes, 2006).

We selected for conversion events in which information is transferred from the mutant ura3-e allele to the wild-type allele using solid medium containing 5-fluoro-orotate that selects against cells with wild-
type URA3 activity (Boeke et al., 1987). In order to distinguish conversion events from other events that would result in a 5-FOA^R derivative, we also examined phenotypes of the other markers on chromosome IV. As shown in Fig. 2A, a gene conversion event unassociated with a crossover (NCO conversion) would be expected to result in a 5-FOA^R colony that is Trp^+ Hyg^R, and pink (Class IA). A conversion associated with a crossover (CO conversion) could produce colonies of two different phenotypes (Fig. 2B). If the recombinant products co-segregated (left side of Fig. 2B, Class IB), then the 5-FOA^R colony would have the same phenotype as the conversion unassociated with the crossover (Trp^+ Hyg^R pink). If the recombinant products do not co-segregate (right side of Fig. 2B, Class II), then the 5-FOA^R product would be Trp^+ Hyg^R and red. A 5-FOA^R derivative could be generated by a crossover between the URA3 genes and centromere without a conversion involving URA3 (Fig. 2C). For example, a crossover between the TRP1 marker (closely linked to the centromere) and the heterozygous hphMX4 insertion would generate a 5-FOA^S Trp^+ red colony (Fig. 2C, Class III). Since the length of a mitotic gene conversion tract is expected to be small relative to the distance between the TRP1 marker and URA3 marker (about 500 kb), we expect the event depicted in Fig. 2C to be much more common than the events shown in Figs. 2A and 2B. In addition, a Class III strain could result from a BIR event in which a break occurring between trp1 and hphMX4 on the blue chromosome is repaired using the red chromosome as a template. Crossovers and BIR events are indistinguishable without recovery of both daughter cells derived from the cell in which the recombination event occurred (Barbera and Petes, 2006).

Other than the mitotic recombination events shown in Fig. 2, several other genetic changes could result in a 5-FOA^R strain. Chromosome loss would generate a 5-FOA^R Trp^+ Hyg^S red colony. A point mutation or deletion of the wild-type URA3 gene would result in a 5-FOA^R Trp^+ Hyg^R pink colony.

In summary, we expect three common phenotypic classes of 5-FOA^R derivatives of EY7: Class I (5-FOA^R Trp^+ Hyg^R pink), Class II (5-FOA^R Trp^+ Hyg^R red), and Class III (5-FOA^R Trp^+ Hyg^S red). Of a total of 625 independent 5-FOA^R derivatives, the numbers of derivatives observed for each class were: 59 Class I, 46 Class II, and 520 Class III. Our subsequent analysis concerns only Class I strains. There
are four expected sub-classes of Class I: conversion in which information is transferred from *ura3-e* to *URA3* unassociated with a crossover (Class IA), conversion between *ura3-e* and *URA3* associated with a crossover (Class IB), a point mutation inactivating the wild-type *URA3* allele (Class IC), and deletion of the wild-type *URA3* allele (Class ID).

**Measurements of gene conversion tract lengths:** For both Class IA and IB events, the minimum conversion tract length would be one base. We showed previously, however, that spontaneous and damage-induced mitotic gene conversion events have a median tract length of 5-10 kb (St. Charles and Petes, 2013; Yin and Petes, 2013). Thus, we expect that most of the conversion events detected as Class I events will result in loss of heterozygosity (LOH) for multiple SNPs located near the mutant *ura3-e* allele. In our previous studies, we used oligonucleotide-containing microarrays to identify regions of LOH (St. Charles *et al.*, 2012; St. Charles and Petes, 2013; Yin and Petes, 2013). The details of this method are described in Materials and Methods and in File S1.

Not all SNPs on chromosome IV near the *URA3* insertion are represented on the microarray. To refine the mapping of the breakpoints, for some events, we used a different method of determining LOH termed “SPA” (Single-nucleotide-polymorphism PCR Analysis; St. Charles *et al.*, 2010). In brief (details in File S1 and Table S3), we identified polymorphisms in the genomic region of interest in which one homolog had a restriction site that the other lacked. By PCR, we generated a small (< 700 bp) fragment from genomic DNA of the strain of interest, and treated the fragment with the diagnostic restriction enzyme. By examining the resulting products on an agarose gel, we could distinguish which allele was present. Gene conversion tracts were primarily examined by microarrays, but a few were also examined by SPA.

Figure 3 shows a microarray analysis of one Class I strain (EY7-40). In this strain, the conversion event is flanked by heterozygous sites at SGD coordinates 1008881 and 1013909, and the homozygous sites closest to the transitions are located at SGD coordinates 1009116 and 1013370. The lengths of conversion tracts in such strains were estimated by averaging the distance between the
heterozygous sites and the homozygous sites and adding the size of the inserted \textit{URA3} gene (about 1.1 kb). The tract length of the event shown in Fig. 3 is 5.8 kb.

Of the Class I strains examined by microarray, most had undergone LOH for one or more SNPs flanking the \textit{URA3} gene on chromosome IV. The coordinates for the transitions between heterozygous and homozygous SNPs for all conversion events are in Table S4. Most of the conversion events had only two transitions between heterozygous and homozygous regions. In Table S4, the centromere-proximal and centromere-distal transitions are called “a” and “b”, respectively. There were two conversion events with additional transitions (EY7-63 and EY7-69). These events are described in File S1.

In five of the Class I strains, no flanking SNPs had undergone LOH. There are three mechanisms that could produce such strains: 1) short allelic conversion events that failed to include flanking markers, 2) new mutations within the \textit{URA3} gene on the W303-1A-derived homolog, and 3) ectopic gene conversion between the \textit{URA3} gene on chromosome IV and one of the mutant \textit{ura3} genes (\textit{ura3-1} or \textit{ura3-p}) located on chromosome V. To distinguish among these possibilities, we sporulated these five diploid strains and identified spores that had the mutant \textit{ura3} gene on the W303-1A-derived homolog. We PCR-amplified the \textit{ura3} gene on chromosome IV from these haploid strains, and sequenced the resulting product. In one of these strains (EY7-8), the mutant allele generated by conversion had the same \textit{ura3-e} allele that was originally heterozygous in EY7; this strain, therefore, had a short gene conversion tract (< 1 kb). The other four strains (EY7-19, -24, -32, -68) that did not have co-conversion of flanking markers had one or more new mutations in \textit{URA3}. The mutational positions in \textit{URA3} (1 being the first base of the initiating ATG) and types of alterations in these strains were: 175, A to G (EY7-19); 62, T to C (EY7-24); 345, G to A (EY7-32); 84, G to A; 90, A to T; 91, A to T (EY7-68).

Lastly, the one strain (EY7-3) had a deletion of the wild-type \textit{URA3} allele on the W303-1A-derived homolog, resulting in hemizygosity for the \textit{ura3-e} allele. The heterozygous/homozygous transitions for the deletion were at SGD coordinates 980838 and 993113 (centromere-proximal) and at coordinates 1079615 and 1089446 (centromere-distal); the resulting deletion is about 100 kb.
In summary, of the 59 Class I strains, 54 were gene conversions (Classes IA and IB), 4 were new ura3 point mutations (Class IC), and one was a large deletion (Class ID). The locations of the 54 conversion tracts are in Fig. 4, and the tract sizes are shown in a histogram in Fig. 5. It is apparent that about half of the tracts were very long (>25 kb) and the remainder had a median length of less than 10 kb. Considering all of the data, the median tract length was 16.6 kb. The tracts shorter or longer than 25 kb had median lengths of about 6.4 kb (4.4-10 kb, 95% confidence limits) and 54 kb (45-69 kb), respectively. Table S4 lists all conversion tract lengths.

**Rate of gene conversion:** By measuring the frequency of 5-FOA$^R$ derivatives in multiple independent cultures of EY7 and applying the method of the median (Lea and Coulson, 1949), we calculated a rate of 5-FOA$^R$ of $1.5 \times 10^{-5}$/division (95% confidence limits of 0.9-2.0 x $10^{-5}$). From the analysis described above, of 625 independent 5-FOA$^R$ derivatives of EY7, 54 resulted from gene conversion. The rate of allelic conversion in which information is transferred from the ura3-e allele to the wild-type URA3 allele on the other homolog is about $1.3 \times 10^{-6}$/division (54/625 x [1.5 x $10^{-5}$]).

**Association of crossovers with conversions:** As shown in Fig. 2A and 2B, conversion events at the URA3 locus could be unassociated with crossovers (NCO) or crossover-associated (CO). By microarray analysis, NCO and CO conversions are indistinguishable. The distinction can be made by examining the coupling of heterozygous markers flanking the conversion tract. If the conversion event is NCO, then heterozygous markers flanking the tract have the same coupling arrangement as in the original EY7. In contrast, if the conversion event is associated with a crossover, the coupling of markers flanking the tract is reversed. We determined the coupling arrangements by meiotic analysis of all strains with a conversion event (Fig. 6). For this analysis, we used the hphMX4 insertion as the heterozygous marker centromere-proximal to the tract. The centromere-distal markers were SNPs that were close to the conversion tract, and could be diagnosed by SPA analysis as described above. The SNP that was used for the meiotic analysis for each conversion strain is given in Table S5. In Fig. 6, we show the heterozygous SNP used in the analysis as SNP$^W$ for the W303-1A-derived SNP and SNP$^Y$ for the allelic YJM789-derived SNP.
If two markers in yeast are closely-linked, most of the tetrads will be of two classes, parental ditype (all four spores with markers in the original parental coupling relationship, PD) or tetratype (two spores with markers in the parental configuration and two markers with the recombinant arrangement, TT). Non-parental ditype tetrads (all four spores with markers in recombinant arrangements, NPD) for linked markers are rare since such tetrads require four-strand double crossovers (Petes et al., 1991). For most of the conversion events, the distance between the \textit{hphMX4} marker and the centromere-distal end of the conversion tract is 80 kb or less. On chromosome IV, the average association between genetic and physical distance is about 0.3 cM/kb (Saccharomyces Genome Database). Therefore, the genetic distance between the \textit{hphMX4} marker and the centromere-distal heterozygous SNP is less than 25 cM. For a 25 cM distance, PD and TT tetrads are expected to be about equally frequent and there would be very few NPD tetrads. As shown in Fig. 6, the genotypes of spores that constitute PD and NPD segregation patterns are reversed for NCO and CO conversion events. We will refer to the PD and NPD patterns of the strains with an NCO conversion event as PD-1 and NPD-1, and the patterns of the strains with CO conversions as PD-2 and NPD-2.

For each strain with a conversion event, we analyzed all spores derived from two to eight tetrads with four viable spores for the segregation of the \textit{hphMX4} marker and the centromere-distal SNP. If at least two of the tetrads had a PD-1 segregation pattern and no tetrads had the PD-2 segregation pattern, we concluded that the conversion was an NCO conversion. Similarly, if at least two of the tetrads had a PD-2 segregation pattern and no tetrads had the PD-1 segregation pattern, we concluded that the event was a CO conversion. By this criterion, we showed that 39 of 54 conversion events were NCO, and 15 were CO (summarized in Table S4 and S5).

In strains with very long conversion tracts, the procedure for determining the coupling of markers by tetrad dissection was tedious, since most of the tetrads had tetratype segregation. We developed a second method of examining coupling relationships by mitotic recombination using the 5-FOA\textsuperscript{R} derivative EY7-33 that had a very long (121 kb) conversion tract. Both Class IA (NCO) and IB (CO)
strains are heterozygous for the *hphMX4* and *SUP4-o* markers (Fig. S1). We treated EY7-33 with ultraviolet light (details in Materials and Methods) to stimulate an additional mitotic crossover event on chromosome IV. As shown in Figure S1A, a crossover between the *hphMX4* marker and *CEN4* in strains with the coupling relationship shown in Class IA strains could generate a red/white sectored colony in which the cells in the white sector would be Hyg<sup>R</sup> and the cells in the red sector would be Hyg<sup>S</sup>. In contrast, if the strain has the markers in the reverse coupling relationship (Fig. S1B), the cells in the white and red sectors would be Hyg<sup>S</sup> and Hyg<sup>R</sup>, respectively. Of 76 red/white sectored colonies examined, 29 showed co-sectoring of the hygromycin-resistance phenotype. In all 29 of these colonies, the white sector was Hyg<sup>R</sup> and the red sector was Hyg<sup>S</sup>, as expected if the EY7-33 conversion event was unassociated with a crossover.
DISCUSSION

In this study, we examined spontaneous mitotic gene conversion in a diploid strain in which conversion events could be measured without selecting against long conversion tracts. Our analysis showed two distinct size classes of events likely generated by two different mechanisms. About 40% of the conversions were associated with crossovers, demonstrating that there is not a strong bias against the generation of crossovers in mitosis.

Frequency of gene conversion

The rate of allelic gene conversions in this study was about $1.3 \times 10^{-6}$/division, similar to rates of allelic conversions in other studies that vary between $10^{-7}$ to $10^{-6}$ (Petes et al., 1991; Paques and Haber, 1999). Although the rate estimates obtained in heteroallelic experiments are usually lower than observed in our study (presumably because of the more restrictive system for detecting events), rates of conversion as high or higher than ours were observed in some heteroallelic studies (Golin and Esposito, 1984). These high rates may be the result of a marker being located near a mitotic recombination hotspot. It should also be pointed out that, although the rate of mitotic recombination events is much less than the rate of meiotic recombination events, these events are still likely to be important in generating novel combinations of alleles. It has been estimated that there are about 30,000 clonal (mitotic) generations to one sexual (meiotic) cycle for yeast (Magwene et al., 2011).

Lengths of gene conversion tracts

In Fig. 5, there appear to be two discrete size classes of conversion tracts, those with a median size of about 6.4 kb (range 1-21 kb) and those with a median of 54 kb (range 36-121 kb). The smaller size class is similar in length to the conversion events that we have examined in other studies (Lee et al., 2009; St. Charles et al., 2012; St. Charles and Petes, 2013; Yin and Petes, 2013). Our classification of the conversion tracts into two classes is based on three considerations: 1) using R software and the Kolmogorov-Smirnov test, the distribution of tract lengths is significantly different from a uniform
distribution ($p<0.001$) and a normal distribution ($p=0.01$), 2) there are no tracts between 21-36 kb in length, although there are 30 tracts shorter than 21 kb, and 24 tracts greater than 36 kb, and 3) all of the tracts greater than 36 kb overlap the inverted Ty elements located centromere-proximal to ura3 (Fig. 4). The asymmetric distribution of the long conversion tracts relative to the ura3 gene argues strongly that these long conversion tracts are inherently different from the short conversion tracts.

In our previous mapping of unselected crossovers on the right arm of chromosome IV (St. Charles and Petes, 2013), the region containing the inverted Ty elements was associated with a high frequency of crossovers and was termed HS4 (HotSpot4). We previously noted that the crossover-associated conversion events involving HS4 had long tracts with a median size of 48 kb. Our current analysis confirms this estimate that was based on a smaller number of events and, in addition, demonstrates that NCO conversions involving HS4 are also very long (median length of 59 kb). In our previous study, we showed that deletion of one of the Ty elements composing HS4 or expanding the distance between the two Ty elements resulted in loss of hotspot activity (St. Charles and Petes, 2013). We argued, therefore, that the hotspot was a consequence of cleavage of a hairpin structure formed when a single-stranded gap was generated near the center of HS4. Because both the CO and NCO long conversion tracts in the current study overlap with HS4, we assume that they are also initiated as a consequence of processing the hairpin structure associated with the inverted Ty elements.

Several other points are relevant concerning the HS4-associated gene conversion events. First, we previously showed that CO conversions involving HS4 were a consequence of the repair of two sister chromatids broken at approximately the same position, likely reflecting a DSB formed in G1 of the cell cycle (St. Charles and Petes, 2013). In the current study, in which we recover only one of the two daughter cells with recombinant products, we cannot determine whether the HS4-associated events are associated with a G1- or G2-initiated DNA lesion. Second, HS4 is homolog-specific (St. Charles and Petes, 2013). The W303-1A-derived chromosome has the inverted pair of Ty element, whereas the YJM789-derived chromosome has only a portion of one Ty element (Fig. 4). Thus, for the broken ends
of the W303-1A homolog to invade the YJM789 homolog, the ends would have to be resected by about 10 kb to expose flanking homology. This extent of resection is much less than the median length of the long conversion tracts (54 kb). Although in the current study, any conversion tracts initiated at HS4 must extend about 28 kb in order to include the URA3 marker, most of the tracts are considerably longer than 28 kb. Third, the very long tracts associated with HS4 are not typical, but are also not unique. Long conversion events are also associated with a second pair of inverted Ty elements on chromosome IV (HS3, St. Charles and Petes, 2013), and a mitotic recombination hotspot associated with the trinucleotide repeats (GAA) results in conversion tracts with a median length of 20 kb (Tang et al., 2011). In addition, conversion tracts greater than 30 kb have been observed in multiple other studies (Lee et al., 2009; St. Charles and Petes, 2013), although these events are usually less than 10% of the conversion events. One interpretation of the data is that most mitotic recombination events are initiated by random DNA lesions that are processed to produce conversion tracts with a median length of about 6 kb. The other class of recombination events is initiated at sequences capable of forming secondary DNA structures and is processed to produce very long conversion tracts.

An important question is whether the mechanism of generating the short tracts is the same as that which generates the long tracts. Previous studies have demonstrated small mitotic tracts (one kb or less) are a consequence of heteroduplex formation, followed by mismatch repair as predicted by the models in Fig. 1 (Nickoloff et al., 1999; Mitchel et al., 2010). Although it is possible that the very long conversion tracts reflect the formation of very long heteroduplexes and concerted repair of the resulting mismatches, there are two other plausible alternatives. First, these tracts could be a consequence of the repair of a very long double-stranded DNA gap (Lee et al., 2009). Second, these events could represent a BIR event that copies a region of one homolog before switching to the acentric chromosome fragment (Fig. 7, NCO conversion). A long CO conversion could be generated by processing of the BIR structure (shown as the intermediate between Steps 1 and 2 in Fig. 7) to yield a recombinant chromosome and a broken centromere-associated red chromosome. Invasion of this
broken red chromosome into an unbroken blue chromosome, followed by BIR, would produce a long gene conversion tract associated with a crossover. Production of a crossover by two BIR events has been proposed previously (Fig. S4, Lee et al., 2009). Template switching during BIR is common (Smith et al., 2007), and the generation of long conversion tracts by BIR has been suggested previously (Voelkel-Meiman and Roeder, 1990; Llorente et al., 2008; Lee et al., 2009, Yin and Petes, 2013; Chandramouly et al., 2013). It is important to distinguish BIR events that proceed to the end of the chromosome (Voelkel-Meiman and Roeder, 1990) from those in which a long interstitial conversion event is generated; the latter class requires a template switch.

Analysis of the genetic requirements for the HS4-associated gene conversion tracts could help distinguish among the mechanisms discussed above. Mutations in POL32 and PIF1 reduce the frequency of BIR (Llorente et al., 2008; Chung et al., 2010) and would be expected to reduce the frequency of long conversion tracts if BIR is involved. However, since Pol32p is a general processivity factor for DNA polymerase and since gene conversion by the SDSA pathway requires extensive DNA synthesis primed from the invading chromosome end, it is not clear that the pol32 mutation would affect BIR exclusively. Ho et al. (2010) showed that pol32 strains had shorter conversion tracts in a standard mitotic recombination assay.

**Relationship between conversion events and crossovers**

In the past, it was assumed that CO and NCO conversions reflect alternative patterns of cleavage of the double Holliday junction (Szostak et al., 1983). It is now clear that most NCO conversions in meiosis (Allers and Lichten, 2001) and mitosis (Mitchel et al., 2010) are a consequence of SDSA and/or dHJ dissolution pathways (Fig. 1) rather than cleavage of a dHJ. In our study, 15 of 54 conversions (28%) were crossover-associated and 39 of 54 (72%) were NCO. In our system, we screen for those conversion events in which the recombinant chromosomes co-segregate. Since events in which one recombinant and one non-recombinant chromosome co-segregate are equally frequent (Chua and
Jinks-Robertson, 1991), we conclude that 43% of the conversions were CO ([15x2]/[54+15]) and 57% were NCO. In earlier studies (summarized in the Introduction), NCO conversions often substantially exceeded CO conversions. These results may have been influenced by the heteroallelic systems used to select the conversion or by limited amounts of homology available for ectopic recombination. Our findings argue strongly that the mitotic repair of recombinogenic lesions in yeast is often associated with crossovers. In other genetic systems such as Drosophila, however, NCO conversions are much more frequent than CO conversions (Andersen and Sekelsky, 2010).

In our previous analysis of UV-induced conversion events, we showed that CO tracts were significantly longer than NCO tracts (Yin and Petes, 2013). In our current study, however, the numbers of CO and NCO events for the short tracts (7 and 23, respectively) and long conversions (8 and 16, respectively) were not significantly different (Fisher exact test, p=0.54). Within the short-tract category, the nine shortest tracts (those less than 4 kb) were all NCO events. The comparison with the short tracts larger than 4 kb (4-22 kb) was, however, not statistically significant (p=0.07). The difference between conversion tracts generated by UV and our current results may reflect the difference in the recombinogenic lesions. Alternatively, the ratio of NCO to CO events could be affected by chromosome context. Mancera et al. (2008) reported such differences for meiotic events in yeast.

In conclusion, our analysis of spontaneous mitotic gene conversion events demonstrates the existence of two types of conversions distinguishable by their size. For both types of conversions, CO and NCO events are recovered with approximately equal frequencies.

ACKNOWLEDGMENTS

We thank members of the Petes and Jinks-Robertson labs for useful discussions, Y. Yin, S. Jinks-Robertson, A. Guo, and E. Hum for comments on the manuscript, and Y. Yin for help with the statistical analysis. The research was supported by National Institutes of Health grants
GM24110 and GM52319 to T. P., and K. O. was supported by a GRFP fellowship from NSF (1106401).
LITERATURE CITED


Esposito, M. S., and J. E. Wagstaff, 1981 Mechanisms of mitotic recombination, pp. 341-370 in

Golin, J. E., and M. S. Esposito, 1984 Coincident gene conversion during mitosis in
*Saccharomyces*. Genetics **107**: 355-365.


Haber, J. E., and M. Hearn, 1985 Rad52-independent mitotic gene conversion in *Saccharomyces cerevisiae* frequently results in chromosomal loss. Genetics **111**: 7-22.


Judd, S. R., and T. D. Petes, 1986 Physical lengths of meiotic and mitotic gene conversion tracts in *Saccharomyces cerevisiae*. Genetics **118**: 401-419,


St Charles, J., E. Hazkani-Covo, Y. Yin, S. L. Andersen, F. S. Dietrich et al., 2012 High-resolution genome-wide analysis of irradiated (UV and gamma-rays) diploid yeast cells
reveals a high frequency of genomic loss of heterozygosity (LOH) events. Genetics 190: 1267-1284.


FIGURE LEGENDS

Figure 1. Pathways of repair of DNA breaks by homologous recombination. Two recombining double-stranded DNA molecules are shown as paired red and blue lines. Dotted lines indicate DNA synthesis. Recombination events are initiated by a DSB, followed by 5’ to 3’ processing of the broken ends. All pathways are initiated by invasion of one end of the broken chromosome into the unbroken chromosome, followed by formation of a D-loop caused by DNA synthesis from the invading 3’ strand.

A. Synthesis-dependent strand annealing (SDSA). Following DNA synthesis, the invading strand disassociates from the template and re-associates with the other broken end. The net result is a heteroduplex on one side of the position of the original DSB with flanking markers in the original configuration (non-crossover, NCO). The region of the heteroduplex is boxed. Repair of mismatches within the heteroduplex can result in a gene conversion (two blue strands) or a restoration event (two red strands).

B. Double Holliday junction (dHJ). The non-invading broken end anneals with the D-loop forming two junctions. Depending of the mode of cleavage of the junctions, this structure can be resolved as a non-crossover (left) or a crossover (middle). Alternatively, the structure can be dissolved without junction cleavage (right).

C. Break-induced replication (BIR). The end from the left portion of the broken chromosome sets up a moving D-loop that replicates the intact chromosome by conservative replication. The right portion of the broken chromosome is lost.

Figure 2. Genetic system for the detection of gene conversion and crossover events on chromosome IV. Conversion and crossover events are shown as occurring between replicated chromatids. Centromeres are depicted as ovals, and red and blue lines indicate the YJM789- and the W303-1A-derived homologs, respectively. The hphMX4 and the SUP4-o genes are located on only one of the two homologs. Recombination events are selected on plates containing 5-FOA that selects for loss of the wild-type URA3 gene. Diploid cells with 0, 1, and 2 copies of SUP4-o form red, pink, and white colonies, respectively.
A. Gene conversion without an associated crossover (NCO, Class IA). In this event, wild-type URA3 sequences are replaced by mutant sequences as indicated by the arrow. As shown by the pair of arrows, one type of segregation will result in a Trp$^+$ Hyg$^R$ 5-FOA$^R$ pink colony with the flanking markers in the original coupling arrangement.

B. Conversion with an associated crossover (CO, Class IB). Following the crossover, if the recombinant chromatids co-segregate (left side), a Trp$^+$ Hyg$^R$ 5-FOA$^R$ pink colony will be observed as in Fig. 2A. If one recombinant and one non-recombinant chromatid co-segregate (right side), a Trp$^+$ Hyg$^R$ 5-FOA$^R$ red colony will be formed.

C. Crossover centromere-proximal to the *hphMX4* marker. In this event, the 5-FOA$^R$ derivative is generated without a conversion. Following the crossover, if one of the recombinant chromosomes co-segregates with the non-recombinant YJM789-derived homolog, a Trp$^+$ Hyg$^S$ 5-FOA$^R$ red colony will be generated.

**Figure 3. Microarray analysis of the extent of a gene conversion event in EY7-40.**

The strain EY7-40 has the phenotype indicative of a gene conversion event (Class I, 5-FOA$^R$ Trp$^+$ Hyg$^R$ pink). DNA was isolated and hybridized to a SNP-specific microarray, and the ratio of hybridization (EY7-40 versus DNA from a control heterozygous strain) to W303-1A-specific and YJM789-specific SNPs was measured. The red lines or boxes show hybridization to YJM789-specific SNPs, and the blue lines or diamonds indicate hybridization to W303-1A-specific SNPs.

A. Low-resolution analysis. The values on the X-axis are SGD coordinates in bp. The *URA3* insertion is located between bases 1013217 and 1013218. The ratio was calculated in a moving window of 9 SNPs.

B. High-resolution analysis. In this depiction, each square and diamond shows the hybridization signal to a specific oligonucleotide on the microarray.

C. Schematic depiction of the conversion event showing the transitions between heterozygous and homozygous SNPs (green indicating heterozygous SNPs and red showing the region homozygous for
the YJM789-derived SNPs). The “a” transition is located between coordinates 1008881 and 1009116, and the “b” transition is between 1013370 and 1013909 (Table S4).

**Figure 4. Map locations of gene conversion events.** Our mapping of 54 conversion events that include the *ura3-e* mutation is summarized. The blue and red lines indicate the homologs derived from W303-1A and YJM789, respectively. The lengths of independent conversion events are shown by horizontal lines labeled with the number of the EY7 isolate. Green and purple lines indicate NCO conversions and CO conversions, respectively. The black lines at the bottom of the figure show the distribution of SNPs on the microarray, and the yellow lines show the location of SNPs examined by SPA.

**Figure 5. Histogram of gene conversion tract lengths.** Based on the analysis of Table S4, the 54 conversion events appear to have two distinct size distributions. All of the events with tracts lengths greater than 25 kb include HS4.

**Figure 6. Meiotic analysis of the coupling of markers flanking a mitotic gene conversion event.** To distinguish whether the conversion was unassociated (NCO) or associated (CO) with a crossover, we examined Class I strains by tetrad analysis. We analyzed the patterns of segregation for the centromere-proximal *hphMX4* marker and a SNP located centromere-distal to the conversion tract. The red and blue lines signify chromosome regions derived from the YJM789-derived and W303-1A-derived homologs, respectively.

A. Meiotic segregation patterns expected for Class I strains with an NCO conversion. If there is no meiotic crossover (italicized NCO), we expect a parental ditype (PD) tetrad, two Hyg\(^R\) SNP\(^W\) spores: two Hyg\(^S\) SNP\(^Y\) spores. A single crossover (italicized SCO) between the *hphMX4* marker and the diagnostic SNP would produce a tetratype tetrad (TT): one Hyg\(^R\) SNP\(^W\) spore: one Hyg\(^R\) SNP\(^Y\) spore: one Hyg\(^S\) SNP\(^W\) spores: one Hyg\(^S\) SNP\(^Y\) spore. Lastly, a four-stranded double-crossover (DCO) between the *hphMX4* marker and the diagnostic SNP would produce a non-parental ditype tetrad (NPD): two Hyg\(^R\) SNP\(^Y\) spores: two Hyg\(^S\) SNP\(^W\) spores.
B. Meiotic segregation patterns expected for Class I strains with a CO conversion. Similar segregation patterns to those observed in Fig. 6A are expected, but the coupling of the markers in the PD and NPD tetrads would be reversed from those observed in Fig. 6A.

**Figure 7. Generation of a long gene conversion tract by a double BIR event.** We show two double-stranded recombining DNA molecules with the centromere shown as a circle. Dotted lines show DNA synthesis. The conversion tract is formed by two BIR events in the following steps: Step 1, a DSB occurs on the blue chromosome (thick arrow) with 5' to 3' resection of the broken ends. Step 2, the left end invades the red chromosome and initiates DNA synthesis (dotted red line). Step 3, synthesis continues in a moving D-loop mode with second strand synthesis occurring on the displaced strand. Step 4, the chromosome involved in the BIR event disengages from the red chromosome and invades the centromere-distal fragment of the blue chromosome to begin a second BIR event. Step 5, the BIR event continues to the end of the chromosome, and the acentric blue chromosome fragment is lost.
A. NCO conversion

B. CO conversion

Meliotic segregation patterns

NCO

SCO

DCO

TT

PD-1

NPD-1

PD-2

NPD-2