Gene copy number variation (CNV) in haploid and diploid strains of the yeast

*Saccharomyces cerevisiae*

Hengshan Zhang§1, Ane F. B. Zeidler‡, Wei Song§, Christopher M. Puccia†, Ewa Malc†,
Patricia W. Greenwell§, Piotr A. Mieczkowski†, Thomas D. Petes§, and Juan Lucas Argueso‡*

‡ Department of Environmental and Radiological Health Sciences, Colorado State University,
Fort Collins, CO 80523.
§ Department of Molecular Genetics and Microbiology, Duke University Medical Center,
Durham, NC 27710.
† Lineberger Comprehensive Cancer Center, Carolina Center for Genome Science,
Department of Genetics, University of North Carolina, Chapel Hill, NC 27599.

1. Present address: The First Affiliated Hospital of Fujian Medical University, Fuzhou, P. R.
China.

* Corresponding author
Running title: yeast CNV detection assay

Keywords:
Copy Number Variation
Non-Allelic Homologous Recombination
Gene Amplification
Chromosomal Rearrangements

Corresponding author:
Juan Lucas Argueso, PhD
493 MRB Building, 1618 campus delivery
Colorado State University, Fort Collins, CO 80523.
Phone: 1-970-491-3681
Fax: 1-970-491-0623
email: lucas.argueso@colostate.edu
ABSTRACT

The increasing ability to sequence and compare multiple individual genomes within a species has highlighted the fact that copy number variation (CNV) is a substantial and under-appreciated source of genetic diversity. Chromosome-scale mutations occur at rates orders of magnitude higher than base substitutions, yet our understanding of the mechanisms leading to CNVs has been lagging. We examined CNV in a region of chromosome V (Chr5) in haploid and diploid strains of *Saccharomyces cerevisiae*. We optimized a CNV detection assay based on a reporter cassette containing the *SFA1* and *CUP1* genes that confer gene dosage-dependent tolerance to formaldehyde and copper, respectively. This optimized reporter allowed the selection of low order gene amplification events, going from one copy to two copies in haploids, and from two to three copies in diploids. In haploid strains, most events involved tandem segmental duplications mediated by non-allelic homologous recombination between flanking direct repeats, primarily Ty1 elements. In diploids, most events involved the formation of a recurrent non-reciprocal translocation between a Chr5 Ty1 element and another Ty1 repeat on Chr13. In addition to amplification events, a subset of clones displaying elevated resistance to formaldehyde had point mutations within the *SFA1* coding sequence. These mutations were all dominant, and are proposed to result in hyper-active forms of the formaldehyde dehydrogenase enzyme.
INTRODUCTION

As a consequence of studies that utilize genomic microarrays and next-generation DNA sequencing, it has become clear that much of the natural genetic variation that exists between individuals is due to alterations in the number of copies of genes rather than differences in the nucleotide sequence (GIRIRAJAN et al. 2011; VELTMAN and BRUNNER 2012). It has been calculated that 8-25 kb of DNA are deleted or duplicated per generation in humans compared to about 100 bp of point mutations (ITSARA et al. 2010). Although deletions and duplications vary in size from a few base pairs (for example, changes in the lengths of microsatellites) to many Mb (for example, changes in ploidy), Girirajan et al. (2011) define copy-number variants (CNVs) in humans as changes that vary in size between 50 bp and 1 Mb. While most CNV events have no obvious effect, a significant number are associated with human diseases including Charcot-Marie-Tooth syndrome, autosomal dominant leukodystrophy, Williams syndrome, several cancer predispositions, and autism-related and other neuro-developmental disorders (ABRAHAMS and GESCHWIND 2008; GIRIRAJAN et al. 2011; KREPISCHI et al. 2012; MALHOTRA and SEBAT 2012; SULLIVAN et al. 2012). There is also strong evidence suggesting that CNVs played a significant role in human evolution (ISKOW et al. 2012). Finally, multiple somatic CNV events are frequently observed in the altered karyotypes of cancer cells (STRATTON et al. 2009).

CNVs have also been widely observed in natural populations and have been studied in detail in model organisms. In the discussion below, we analyze genomic deletions and duplications in the yeast Saccharomyces cerevisiae. We will define a CNV as a deletion or duplication that includes at least one gene (average of about 2 kb) and no more than one chromosome arm (average of about 400 kb). The most frequently observed mechanisms for the generation of deletions and duplications are illustrated in Figure 1.
Perhaps the most notable early example of CNV in yeast was the discovery of a large deletion on the right arm of chromosome III (Chr3) that spanned the mating type locus and distal genes linked to it (HAWTHORNE 1963). Later studies demonstrated that this ~150 kb deletion was mediated by homologous recombination between two directly oriented non-tandem repeats at the \textit{MAT} and \textit{HMR} loci (HERSKOWITZ 1988). The Hawthorne deletion was unusual in that it was originally isolated in a diploid strain. In contrast, most early deletion studies were done in haploids. As a result, the identified deletions typically involved small segments spanning only a few loci, since larger deletions often include essential genes. For the well-characterized \textit{HIS4} and \textit{CYC1} genes, only a small fraction (less than 1\%) of the mutations were deletions (FINK and STYLES 1974; SHERMAN \textit{et al.} 1974), which occurred at these loci at a frequency lower than $10^{-8}$/cell division.

Strain-specific hotspots for deletions were identified at two genetic locations. Although deletions of \textit{CYC1} are observed at very low frequency in most wild-type strains, Liebman and colleagues (LIEBMAN \textit{et al.} 1979) found a strain in which deletions of three linked genes (one of which was \textit{CYC1}) occurred at a relatively high rate ($10^{-5}$ to $10^{-6}$ events/division). Subsequently, it was shown that in strains that had a high deletion rate, the \textit{CYC1} gene was flanked by two 6 kb Ty retrotransposons in direct orientation, and that the deletion events were a consequence of homologous recombination between the Ty elements (LIEBMAN \textit{et al.} 1981). Similarly, recombination between flanking delta LTR elements (the long terminal repeat sequences associated with Ty1 and Ty2 elements) was implicated in the high frequency of deletions at the \textit{SUP4} locus (ROTHSTEIN \textit{et al.} 1987). These deletions could reflect single-strand annealing between the direct flanking repeats (Figure 1A) or represent one of the products of unequal crossing-over (Figure 1B).

Assay systems have been developed to study the environmental and genetic control of deletions in yeast. One such assay was developed to study deletions mediated by
homologous recombination between direct flanking repeats (DEL assay) (Schiestl 1989). In this system, recombination between two directly repeated mutant his3 genes, one with a deletion at the 5’ end and one with a deletion at the 3’ end, produced a functional copy of HIS3 at a very high frequency (10^{-4}/division). In contrast, a different deletion selection system that relied on recombination between micro- or non-homologous sequences to reactivate a mutant ura2 gene yielded recombinants at a much lower frequency (10^{-10}/division) (Tourrette et al. 2007).

Chen and Kolodner (1999) developed an assay for detecting large deletions involving simultaneous loss of two counter-selectable markers (URA3 and CAN1) located about 10 kb apart (Gross Chromosomal Rearrangements [GCR] assay). These markers were located near the left end of Chr5, about 10 kb distal to PCM1, the first essential gene on the chromosome. The types of genetic events detected by this assay include interstitial deletions, terminal deletions followed by telomere addition, and translocations between Chr5 and other chromosomes (Chen and Kolodner 1999). For both the translocations and interstitial deletions, the rate of events in a wild-type strain is very low (<10^{-9}/division); the breakpoints of the events occur between CAN1 and PCM1, and typically involved micro- or no sequence homology. Most of these rearrangements reflect non-homologous end-joining (NHEJ) events, although homologous recombination (HR) events involving very small homologies are also observed (Putnam et al. 2005). More recent versions of the GCR assay showed that the introduction of homology either from a dispersed repeated gene family member (HXT13) or from a Ty element inserted between CAN1 and PCM1 elevated the rate of URA3-CAN1 co-deletions by two to three orders of magnitude. Most of these deletions were a consequence of non-reciprocal translocations between the repeat on Chr5 and a homologous sequence elsewhere in the genome, and were associated with amplification of the terminal segment of the donor chromosome (Figure 1D) (Putnam et al. 2009; Chan and Kolodner 2011; Chan
and KOLODNER 2012). In summary, the frequency of deletions in yeast is highly dependent on the context of the reporter gene. In regions in which the reporter is flanked by directly repeated sequences, deletions are frequent and occur through homologous recombination; in regions without repeats, deletions are rare and often involve NHEJ or other microhomology pathways.

Although duplication events in yeast have been less investigated than deletions, there appear to be more diverse types of mechanisms involved. One distinction between duplication events is whether they involve an increase in copy number of a pre-existing duplication, or duplication of a sequence that is initially present in single copy. Alterations in copy number of the tandemly-repeated ribosomal RNA (rRNA) genes or *CUP1* genes represent the first class of events. For these types of genes, alterations in copy number occur as a consequence of homologous recombination (HR) of a variety of types including unequal crossovers (PETES 1980; SZOSTAK and Wu 1980; WELCH et al. 1990), gene conversion (WELCH et al. 1990; GANGLOFF et al. 1996), and single-strand annealing (SSA) (OZENBERGER and ROEDER 1991). Some of these events are remarkably frequent. For example, unequal crossovers within the rRNA gene array occur at a frequency of at least $10^{-2}$/mitotic division (SZOSTAK and Wu 1980) and a frequency of about $10^{-1}$/meiotic division (PETES 1980).

In contrast, the frequency of duplicating sequences that are initially present in single copy is much lower, although very dependent on chromosome context and ploidy. A number of reporter systems have been used to detect duplications; differences in the types of events recovered are likely to reflect the chromosome context of the reporter gene rather than the reporter gene *per se*. Duplications of the *ADH2* or *ADH4* genes can be selected on medium containing antimycin A (DORSEY et al. 1992). Duplications of *RPL20B* can be selected as normal-growing derivatives that display a rescue of the slow-growth phenotype of their
The types of duplications commonly observed can be classified into six groups: 1- interstitial segmental duplications in which the breakpoints are in repeated sequences in direct orientation (KOSZUL et al. 2004; LIBUDA and WINSTON 2006; ARGUESO et al. 2008;
PAYEN et al. 2008; McCULLEY and PETES 2010); 2- segmental duplications unassociated with repeats at the duplication breakpoints (KOSZUL et al. 2004; SCHACHERER et al. 2005; SCHACHERER et al. 2007; PAYEN et al. 2008); 3- interstitial duplications in which the duplicated segments are in an inverted orientation (MOORE et al. 2000; RATTRAY et al. 2005; WATANABE and HORIZUCHI 2005; NARAYANAN et al. 2006); 4- terminal duplications of part of one chromosome arm associated with a terminal deletion of another chromosome (DUNHAM et al. 2002; UMEZU et al. 2002; ARGUESO et al. 2008; KIM et al. 2008; VERNON et al. 2008; HOANG et al. 2010; McCULLEY and PETES 2010; CHAN and KOLODNER 2011; CHAN and KOLODNER 2012); 5- linear extrachromosomal plasmids with the amplified copies in inverted orientation (DORSEY et al. 1992; NARAYANAN et al. 2006); and, 6- Interstitial triplications of a genomic segment containing an origin of replication flanked by short inverted repeats (BREWER et al. 2011).

Class 1 duplications can be generated by unequal crossovers (Figure 1B) or by break-induced replication (BIR) in which a DSB in one repeat is repaired using a non-allelic repeat on a sister chromatid or a homolog. In one system (PAYEN et al. 2008), duplications of this type were dependent on Pol32p, arguing that these events occurred by BIR. Class 2 duplications are likely to reflect BIR events in which the invading end utilizes very small (<10bp) regions of homology (Figure 1C) (SCHACHERER et al. 2005; SCHACHERER et al. 2007; PAYEN et al. 2008); such events have been termed “microhomology-induced replication” (HASTINGS et al. 2009). Class 3 duplications are associated with the processing or replication of inverted repeats located near the reporter gene. In one study, the initiating event is likely to be processing of a cruciform (NARAYANAN et al. 2006). Most Class 4 events are a consequence of a DSB within or near a repeat that is repaired by a BIR event involving a repeat located elsewhere in the genome (resulting in a translocation as in Figure 1D). Since the net result of the translocation event is a large terminal deletion of one chromosome arm
and a large duplication of another chromosome arm, Class 4 events are more often observed in diploid strains (Dunham et al. 2002; Umez u et al. 2002; Argueso et al. 2008; Kim et al. 2008; Vernon et al. 2008; Hoang et al. 2010; McCulley and Petes 2010). In haploids, only repeats very close to the end of chromosomes can support this type of event (Putnam et al. 2009; Chan and Kolodner 2011; Chan and Kolodner 2012). In the two Class 5 events analyzed, the reporter genes were located near the telomeres (Dorsey et al. 1992; Narayanan et al. 2006). The formation of the palindromic plasmid in one system reflected the processing of a small palindromic repeat located centromere-proximal to the reporter gene (Narayanan et al. 2006). Finally, class 6 inverted triplication events observed at the SUL1 locus have been proposed to occur through the extrusion of a circular DNA intermediate formed during replication, followed by re-integration into the genome via homologous recombination (Brewer et al. 2011).

In summary, there are a variety of mechanisms that produce and select for CNVs in S. cerevisiae. The relative importance of these mechanisms is dependent on the chromosome context (in particular, whether the reporter gene is flanked by repeats) and ploidy. In our study, we optimized a system to select for low order amplification events (one extra copy) that more closely resemble the duplications associated with human disease. We used this system to examine gene duplications in both haploids and diploids, and in wild-type and mismatch-repair defective strains. In the chromosome context that we examined (reporter gene in the middle of the right arm of Chr5), most duplications reflected homologous recombination between flanking Ty repeats in haploids, whereas diploids displayed mostly non-reciprocal translocations between Ty elements on Chr5 and Ty elements on other chromosomes, particularly Chr13.
MATERIALS AND METHODS

Yeast strains and plasmids

All strains used in the CNV assay were isogenic with strain MS71, except for noted locus-specific changes introduced by transformation (LEMOINE et al. 2005). Allele combinations were obtained by mating isogenic haploid strains of opposite mating types and selecting segregant spores with the desired genotypes. MS71 is essentially isogenic to the CG379 strain background (MORRISON et al. 1991; ARGUESO et al. 2008). The specific strains and oligonucleotides used in this study are listed in Supporting Information, Tables S1 and S2, respectively. A detailed description of the constructions of specific strains and plasmids is presented in Supporting Information.

Culture media and CNV selection conditions

Unless otherwise noted, yeast cells were grown in YPD rich media or in SC drop-out media (ROSE et al. 1990). The selection of clones carrying amplification events was done by first streaking cells to single colonies in YPD plates (2 days, 30⁰), then inoculating these colonies in 5 ml liquid YPD (18 hr, 30⁰). Dilutions containing ~2x10⁶ cells from these cultures were plated on SC supplemented with a complete drop-out mix and 150 μM CuSO₄ and 1 mM Formadehyde. Formadehyde is not stable in aqueous solution, therefore 1M dilutions (813 μl of 37% methanol stabilized stock [Fischer Scientific BP-531] in 9.187 ml of sterile water) were prepared fresh for each batch of media and mixed in immediately before pouring the plates. Plates were always used within 24 hr of pouring. Cells were incubated in this media at 30⁰ and FA/Cu resistant colonies were observed after 3 to 5 days. Only one colony per culture was examined in downstream experiments to ensure independence. Candidate CNV clones (CFRs) were streaked to single colonies non-selectively in YPD without FA or Cu
for 2 days, and then the purified clones were patched in YPD, grown overnight, then frozen. The FA/Cu resistance phenotype of each clone was retested by growing overnight cultures in 5 ml liquid YPD, then serial dilutions from the cultures were spotted on to SC-complete Cu 150 µM / FA 1mM alongside their corresponding parental strain. Plates were incubated at 30° for 4 days before the differential growth phenotype was assessed.

The nature of the chromosomal rearrangements present in the selected CFR clones was determined using a molecular karyotype analysis, including PFGE and array-CGH, as described previously (ARGUESO et al. 2008). The detailed procedures are presented in Supporting Information. The full set of copy number variation events detected by array-CGH in all clones examined is shown in Supporting Information, Figure S1.

**Analysis of Loss-of-Heterozygosity (LOH)***

Independent clones derived from MS71xYJM789 hybrid background diploids carrying allelic mitotic recombination products resulting in LOH on the right arm of Chr13 were selected in SC-complete plates with 5-FOA. The LOH breakpoints were analyzed from two sets of clones. The first set of 5-FOA resistant clones was derived from strain JAY408, and was analyzed with custom SNP genotyping microarrays and data analysis methods described previously (ST CHARLES et al. 2012). We also isolated a second set of LOH clones derived from strain JAY800. This strain was the one used to measure the rate of LOH on the right arm of Chr13, by counting colonies growing in 5-FOA (non-permissive) and in SC-complete (permissive), and using the values in calculations through the method of the median (LEA and COULSON 1949). LOH breakpoints from this set were analyzed by PCR and restriction fragment length polymorphisms at EcoRI (primers JAO1031/JAO1032) and BglII (primers JAO1029/JAO1030) as described in RESULTS.
RESULTS

A new selection system for copy number variation (CNV)

*De novo* CNV is a pervasive source of genetic diversity that has been increasingly associated with phenotypic consequences in humans, including disease. We set out to develop a new assay system to detect chromosomal rearrangements involving *de novo* gene amplification in yeast, especially those events resulting in only one additional copy of the affected genomic segments. To do this, we took advantage of a previous system that used the *SFA1* and *CUP1* genes to detect amplification through their gene dosage-dependent phenotypes. *SFA1* encodes formaldehyde dehydrogenase, which detoxifies formaldehyde (FA), and *CUP1* encodes methallothionein, which sequesters copper ions (Cu) from solution. Cells harboring amplification of these genes display increased tolerance to formaldehyde and copper, respectively, and these phenotypes have been used to select for gene amplification (Narayanan et al. 2006). A limitation of this earlier system, however, was that it primarily identified clones carrying multiple copies of the reporter genes (>10 fold increase in some cases) because the difference in phenotype between amplification levels was subtle. We therefore decided to optimize the selection conditions with the goal of identifying clones with genome rearrangements that more closely resembled those seen in human *de novo* CNV (i.e. only one extra copy of the amplified genes).

We first deleted the endogenous copy of *SFA1* and its regulatory regions, and we also removed the entire cluster of tandem *CUP1* repeats. Next, we constructed a CNV reporter cassette containing one copy of the *SFA1* and *CUP1* genes, as well as a selective marker for integration (*Hph*, resulting in hygromycin B resistance). This cassette was integrated on Chr5 (Figure 2A) between the *DDI1* and *UBP5* genes. This region is also flanked by the *YERCTy1*-1 and *YERCTy1*-2 repetitive elements that we had previously shown to be the most active for the formation of CNV-associated non-allelic homologous recombination (NAHR)
chromosomal rearrangements following induction of DNA double strand breaks (DSBs) by ionizing radiation (ARGUESO et al. 2008). This region, and the YERCTy1-1 element in particular, had also been shown by other studies to frequently participate in NAHR under various conditions, including spontaneous rearrangements (NARAYANAN et al. 2006; McCULLEY and PETES 2010; CHAN and KOLODNER 2011; CHAN and KOLODNER 2012; CHENG et al. 2012).

Previous studies that used SFA1 and/or CUP1 as amplification reporters used either single or sequential selection for resistance to FA or Cu. Given that these genetic resistance mechanisms are completely independent of each other, we reasoned that simultaneous selection for FA and Cu might create a synergistic effect that would provide a more discrete phenotypic differential between strains carrying one, two, or three copies of the reporter. To test this hypothesis, we built both haploid and diploid strains carrying known numbers of reporter insertions either at the Chr5 site described above, or on Chr4, at the native SFA1 locus. These strains were then tested in media containing incremental concentrations of FA and Cu at various combinations in order to identify optimal conditions that would inhibit the growth of haploid cells carrying one copy of the reporter, but allow normal growth of cells carrying two or more copies. Likewise, we looked for a combination of FA and Cu that inhibited the growth of diploid cells carrying two homozygous copies of the reporter, but supported growth of cells carrying three or more copies. Representative FA/Cu concentration optimization assays showing the gene dosage-dependent resistance phenotype are presented in Supporting Information, Figures S2 and S3. We found that haploid cells with one copy of the reporter had a similar level of FA/Cu resistance as diploids carrying two homozygous copies, and that minimal media containing 1.0 mM FA and 150 µM CuSO₄ prevented colony formation by these strains. However, this inhibition could only be reliably
maintained below a critical plating density of roughly 200 cells per mm², or ~10⁶ cells per standard Petri dish. Above this cell concentration, we observed sporadic growth of small background colonies, possibly due to a filtering effect by the excess stagnant cells.

Once the optimal conditions for simultaneous FA/Cu selection had been established, we set out to isolate clones carrying spontaneous Chr5 amplification events. We grew cultures of the haploid and diploid test strains, plated appropriate dilutions on FA/Cu, and incubated them for 4 days at 30⁰ to select for spontaneous copy number variants. To ensure independence, we selected only one FA/Cu resistant colony from each culture, and isolated single-colony derivatives of these strains using rich media without FA/Cu. These purified candidate clones were named CFR (for Copper Formaldehyde Resistant) and were re-tested in media containing FA/Cu comparing their phenotype to that of the respective parent strain (Figures 2B and 3A). The majority of the clones re-tested positive (102 resistant out of 115 CFRs in the full strain set; Table 1), confirming that the selection conditions worked as intended. Even though the sporadic growth of background colonies on densely plated cultures prevented us from calculating high confidence CNV mutation rates, we estimated that gene amplification on Chr5 occurred spontaneously at a rate ranging from 10⁻⁶ to 10⁻⁷ events/cell division. We focused the remainder of our study on the characterization of the spectra of chromosomal rearrangements associated with the various experimental conditions described below.

Characterization of gene amplification chromosomal rearrangements

After confirming the resistance phenotype, we analyzed all CFR clones for possible chromosome size changes using pulse field gel electrophoresis (PFGE). Among the haploid CFRs, we observed loss of the parental-sized Chr5 coupled with the appearance of larger new chromosomal bands (Figure 2C). In addition, we also analyzed several CFRs by array-
CGH to directly test whether the FA/Cu resistance phenotype was due to CNV (Figure 2D). As expected, most CFRs had additional copies of the Chr5 region where the CNV reporter had been inserted. The parallel analysis of the PFGE and array-CGH data showed that in haploids the CNV events were all segmental duplications, and that the rearrangement breakpoints were two flanking Ty1 elements or flanking delta elements (the long-terminal repeats associated with Ty1) (Figures 2A and 2D; Supporting Information, Figure S1). For example, clone CFR19 had breakpoints at the Watson-oriented LTR elements YERCWdelta20B and YERWdelta22, while CFR20 had breakpoints at the Crick-oriented full-length Ty1 elements YERCTy1-1 and YERCTy1-2. The size increases in the chromosomal molecules carrying these two segmental duplications (~40 kb and ~50 kb, for CFR19 and CFR20 respectively), and the log2 Cy5/Cy3 amplification signals (+1.07 and +1.09, for CFR19 and CFR20 respectively) were consistent with a doubling of the amplified Chr5 region mediated by NAHR between the direct repeats.

Another class of rearrangements observed was exemplified by the pattern seen in CFR17. In this case, the breakpoints were exactly the same as in CFR20, but the rearranged chromosome was about 100 kb longer than the parental Chr5, and the array-CGH amplification signal was also higher (Log2 ratio of +1.57). This result confirmed that CFR17 contained a segmental triplication, mediated by the two Ty1 elements in the region. Accordingly, as shown in Figure 2B, the FA/Cu resistance phenotype of CFR17, with its three copies of the CNV reporter, was noticeably stronger than that of CFR20 with two copies. It is unclear whether the triplication events occurred prior to the initial FA/Cu selection plating, or if they arose initially in cells carrying a duplication that then underwent a secondary NAHR event during the growth of the colony to expand to the 3x level. Since 3x cells are hyper-resistant to FA/Cu, they would be expected to rapidly outnumber the 2x parent cells in the
colony. It is also possible that such secondary rearrangements could be further stimulated by the growth in media containing FA, a known genotoxic agent (KUMARI et al. 2012).

The phenotypic difference between the 2x and 3x reporter dosage levels in haploid strains, as well as our optimization trials, encouraged us to attempt the selection of spontaneous amplification rearrangements in diploid strains homozygous for the Chr5 reporter insertion. We isolated several independent WT diploid CFR clones in FA/Cu media, re-purified them under non-selective conditions, and found that most re-tested positive for FA/Cu resistance (Figure 3A; Table 1). In analyzing the karyotype changes in these diploid clones, we found that 5 of 25 CFRs had one Chr5 homolog containing Ty1-mediated segmental duplications that were essentially identical to those seen in haploids, in addition to one parental copy of Chr5 (e.g. CFR73 in Figure 3C; CFR85 in Supporting Information, Figure S6).

The majority of the diploid CFR clones, however, had a class of rearrangements not observed in haploids: non-reciprocal translocations, exemplified by CFR82 (Figure 3B and 3D). In this case, the two parental copies of Chr5 were unchanged, but one of the copies of Chr13 was rearranged by the loss of a 176 kb segment between \textit{YMRCTy1-5} and the right telomere (\textit{TEL13R}), and the addition of a 128 kb Chr5 segment from \textit{YERCTy1-1} to \textit{TEL05R}, which included the CNV reporter insertion. \textit{YMRCTy1-5} is a Crick-oriented full Ty1 retrotransposon insertion that is present in our strain background near the \textit{tR(UCU)M1} tRNA gene (at SGD coordinate 748,219), but that is absent in the S288c \textit{S. cerevisiae} reference genomic sequence. This Chr13/Chr5 translocation results in an \textsim 875 kb chromosome visible in PFGE, with an overall reduction in size of \textsim 48 kb relative to the parental Chr13. The presence of one remaining parental copy of Chr13 made the loss of the large Chr13 segment viable in the diploids, while the amplified segment of Chr5 was responsible for the third copy.
of the CNV reporter that conferred the FA/Cu resistance phenotype. This type of event is also shown schematically in Figure 1D.

Surprisingly, 20 of 25 independent WT diploid CFRs had the same 875 kb translocation band as CFR82. We used array-CGH to analyze genomic DNA from three other of these clones, and found the same Chr13 deletion and Chr5 amplification pattern in all three. These results indicated that the 875 kb Chr13/Chr5 translocation was a recurrent NAHR event. The fact that no other reporter-associated translocations were observed in the WT diploids suggested that 875 kb event must occur spontaneously at a substantially higher rate than other possible ectopic recombination interactions.

In addition to the chromosomal rearrangements that were associated with amplification of the CNV reporter, in a few cases, we also observed unselected karyotype changes in other regions of the genome. These unselected events were only observed in diploids, and are shown in Supporting Information, Figure S1. They included three trisomies (Chr1, Chr3, Chr16), and five chromosomal rearrangements: four segmental duplications (Chr3, Chr4, Chr12, Chr13), and one non-reciprocal translocation between the right arm of Chr3 and the left arm of Chr5 in clone CFR100. The breakpoints in all five unselected rearrangements occurred at full length Ty or LTR repeats, therefore they were similar in nature to the selected FA/Cu resistant Chr5 amplifications.

One or more extra copies of the SFA1-CUP1 gene dosage reporter were detected in all but two haploid and one diploid CFR clones (CFR16, 42, and 77, respectively). These three exceptions re-tested positive in the selective plates, but did not show any detectable karyotype changes in PFGE. In addition, we examined CFR16 by array-CGH and also did not detect any gene dosage changes. These results suggested the existence of a secondary mechanism of FA/Cu resistance. We later found that these clones had acquired dominant
Investigation of mechanisms for the recurrent Chr13/Chr5 translocation

We considered two hypotheses to explain the recurrence of the 875 kb translocation. The first was that \textit{YMRCTy1-5} might be more similar in sequence to \textit{YERCTy1-1} than any other Ty element in the genome. Since the \textit{YMRCTy1-5} sequence was not available in the reference genome, we PCR-amplified and fully sequenced this element from our strain. We then constructed a sequence alignment containing \textit{YERCTy1-1}, \textit{YMRCTy1-5}, and the other 30 existing full-length Ty1 element sequences in the S288c reference genome (Supporting Information, Figure S4). The alignments showed that \textit{YMRCTy1-5} was indeed among the most similar Ty1 elements to \textit{YERCTy1-1} (99.6% identity; with a 2,981 bp segment of perfect homology). Nonetheless, we did find three other Ty1 elements (\textit{YMLWTy1-2}, \textit{YLRCTy1-1}, and \textit{YPLWTy1-1}) that had even higher identity, and had the same orientation with respect to the centromere as \textit{YERCTy1-1}. These elements should in principle be competent to participate in NAHR with \textit{YERCTy1-1} to generate other stable chromosomal translocations.

One function of the DNA mismatch repair system is to reduce the frequency of recombination between repeated genes that have sequence differences (Harfe and Jinks-Robertson 2000; George and Alani 2012). If near perfect sequence identity was important in affecting the recombination partners of \textit{YERCTy1-1}, then by removing the DNA mismatch repair system we might be able to relax this constraint, and observe translocations involving diverged Ty1 sequences from other chromosomes. We therefore tested haploid and diploid derivatives with a deletion of the \textit{MSH2} gene, which is required for all functions of the mismatch repair system, including anti-recombination activity (Surtees et al. 2004). We

point mutations in the \textit{SFA1} gene. These mutations are presented at the end of the RESULTS section.
recovered 44 msh2Δ and msh2Δ/msh2Δ CFRs, and examined their karyotypes by PFGE and array-CGH. Only two classes of rearrangements were found in these mutants, and they were the same seen in the WT strains: Chr5 segmental duplications and Chr13/Chr5 translocations. This result, coupled with the existence of other suitable Ty1 elements in the genome (>99% identity to YERCTy1-1), ruled out the recurrence of the Chr13/Chr5 translocation due to a sequence identity mechanism. However, one of the msh2Δ/msh2Δ clones (CFR99) had a Chr13/Chr5 translocation chromosome that was slightly larger than the others in its class (~895 kb vs. 875 kb), and had a Chr13 array-CGH deletion breakpoint at YMRWdelta19, a solo LTR element positioned 20.6 kb distal to YMRCTy1-5 (Supporting Information, Figure S5). The amplified region from Chr5 was the same as in the other translocations, but the recombination event likely involved YERWdelta20B, which is in the compatible orientation to generate a stable monosomic rearrangement.

Even though there were no major qualitative differences in the chromosomal rearrangements found in the mismatch repair-deficient CFRs, we did find a substantial shift in their abundance relative to the number of clones without any detectable PFGE or array-CGH changes (e.g. CFR93, Supporting Information, Figure S6). Clones with normal karyotypes accounted for ~65% of all msh2Δ and msh2Δ/msh2Δ CFRs, suggesting the prevalence of a nucleotide mutation FA/Cu resistance mechanism, rather than CNV (Table 1 and discussed below).

The second possibility we considered to explain the recurrence of Chr13/Chr5 875 kb translocation was the presence a putative fragile site on the right arm Chr13 located near the right telomere. By this explanation, YMRCTy1-5 would not be actively responsible for this translocation, but instead, it would simply be the predominant site of repair of the precursor DSB associated with this fragile site. This scenario is plausible because YMRCTy1-5 is the
most distal full-length Ty1 element on the right arm of Chr13, and therefore the first large
NAHR substrate available to repair a DSB being resected from the right end of the
chromosome. This explanation was also consistent with the observation of the CFR99 clone
(Supporting Information, Figure S5), which, possibly due to a relaxed sequence identity
requirement in msh2Δ/msh2Δ, was repaired at an LTR repeat located distal (SGD coordinate
768,548) to the larger homology at YMRCTy1-5.

To test the fragile site hypothesis, we generated isogenic diploids with homozygous
deletions of YMRCTy1-5. In this strain, DNA lesions formed at distal positions on the right
arm and resected toward the centromere would no longer find a suitable recombination
substrate, preventing the formation of the 875 kb Chr13/Chr5 translocation seen in WT cells.
We analyzed the karyotypes of ten independent CFRs derived from this strain and observed
that the pattern of genome rearrangements had changed (Table 1). Remarkably, three of
these CFRs still displayed deletions on the right arm of Chr13, but the breakpoint in all three
was at YMRCTy1-4 (e.g. CFR502, Figure 4C) located near SGD coordinate 375,000. In the
absence of YMRCTy1-5, YMRCTy1-4 became the first available full-length Ty1 NAHR
substrate for repair of a lesion being resected from the right. Assuming that such lesion
originated at a putative fragile site distal to YMRCTy1-5, this resection would have to span at
least an additional 370 kb to reach YMRCTy1-4. Chromosomal rearrangements that involve
long range processing of a distally-located DSB up to a Ty element have also been observed
previously (HOANG et al. 2010).

Another interesting observation in the strain with the deletion of YMRCTy1-5 was the
appearance of a different recurrent rearrangement. Five of ten CFRs had the same 925 kb
Chr7/Chr5 translocation that co-migrated with Chr16 and Chr13 in PFGE (e.g. CFR501;
Figure 4A). This translocation was composed of a deletion of Chr7 right arm sequences distal
to YGRCTy1-3, and amplification of Chr5 sequences from YERCTy1-1 to the right telomere (Figure 4B). In addition, we observed one strain that was trisomic for Chr5, and one strain with a Chr5/Chr5 translocation (iso-chromosome) involving YERCTy1-1 and the Watson-oriented Ty1 element inserted at the *ura3-52* allele on the left arm.

One explanation for the changes observed in the CNV spectrum in the *ymrcty1-5Δ/ymrcty1-5Δ* diploids was that the extensive resection required for ectopic repair of the distal DNA lesion lowered the recovery of CFRs with Chr13 deletions. This lengthier and more time-consuming resection would provide additional opportunities for allelic repair of the DSB using the intact sister chromatid or homolog, an outcome that is not detectable in the CNV assay. As a consequence, other less abundant NAHR events became more prevalent among the FA/Cu resistant clones. The results obtained with the *ymrcty1-5Δ/ymrcty1-5Δ* diploids also uncovered the existence of a second recurrent translocation event, involving the right arm of Chr7.

**Supporting evidence and initial mapping of a candidate Chr13 right arm fragile site**

The results of the CNV assays described above suggested the existence of a frequent DNA lesion on the right end of Chr13, responsible for triggering the recurrent translocations observed in the diploid CFRs. To obtain independent support for this possibility, we generated a hybrid diploid strain marked to allow the detection of LOH events on Chr13. Allelic recombination repair is a much more frequent event than NAHR, however, because our diploid CNV strain is isogenic, it was not suitable to detect LOH. We therefore crossed our haploid strain to a haploid isogenic with YJM789, a highly diverged strain background (Wei *et al.* 2007). The resulting hybrid diploid has a large number of heterozygous single-nucleotide polymorphisms (SNPs) that can be followed to analyze recombination on a
genome-wide basis (Lee et al. 2009). In addition to the SNPs, we inserted a **URA3-Kan** cassette on the CG379-derived homolog of Chr13, downstream of the **ADH6** gene about 10 kb from the right telomere (Figure 5A). A DSB formed in the CG379 homolog and repaired through allelic recombination with the YJM789 homolog may result in LOH of the markers between the site of recombination repair and the right telomere, including the loss of the **URA3** marker. The presumption in this experiment was that a discrete fragile site in this region would produce a bias in the distribution of sites of LOH recombination, causing an excess of breakpoints to be detected in its vicinity (TANG et al. 2011).

Independent spontaneous LOH events were selected in plates containing 5-FOA, and the rate of these events was determined to be $2.4 \times 10^{-5}$ LOH events/cell division (95% confidence interval: 1.7-3.9 x $10^{-5}$). We then analyzed twelve FOA$^R$ clones using custom allele-specific microarrays to determine their genome-wide SNP genotype (CG379, YJM789, or heterozygous) (St Charles et al. 2012). As expected from the 5-FOA selection, all twelve clones analyzed displayed terminal LOH on the right arm of Chr13. 50% of the detected LOH breakpoints (inverted triangles in Figure 5B) were found in the 41 kb region immediately proximal to the **URA3-Kan** insertion site, which corresponded to only 6.3% of the possible recombination window (distance **CEN13 to URA3-Kan** = 643 kb). The other half of the 5-FOA$^R$ clones had breakpoints scattered through the rest of the right arm.

Since the pattern observed in the SNP microarray experiment was suggestive of a bias in the distribution of LOH breakpoints, we examined a larger number of FOA$^R$ clones to assess the significance of this find. We genotyped these additional clones for a SNP marker 40 kb proximal to **URA3-Kan**. This SNP was located at the approximate boundary of the breakpoint cluster detected by microarrays, and corresponded to an **EcoRI** restriction site (SGD Chr13 coordinate 870,864) in the CG379 homolog that was absent in the YJM789 homolog. Using primers that flanked the polymorphism, we generated a PCR fragment for
130 independent 5-FOA<sup>R</sup> diploids, treated the fragments with EcoRI, and examined the products by gel electrophoresis. Of the 130 new LOH clones: 109 were homozygous for the YJM789 SNP, 20 were heterozygous, and 1 was homozygous for the CG379 SNP (possibly within a mitotic gene conversion tract associated with LOH). Therefore, 21 of 130 clones had an LOH breakpoint distal to the EcoRI site. Based on the size of this interval relative to the distance between CEN13 and the URA3-Kan insertion, we expect only eight events. By chi-square analysis, this difference is very significant (\( p = 2.8 \times 10^{-6} \)). This deviation was even more pronounced when we considered the PCR-RFLP and microarray data together (\( p = 2.5 \times 10^{-9} \)). We interpret these results as evidence for the presence of a frequent recombination initiating lesion near the right end of Chr13, perhaps a DSB formed at a chromosomal fragile site. Alternatively, it is possible that the right telomere of Chr13 is poorly maintained and frequently becomes “uncapped”, allowing resection beginning at the telomere itself and extending into telomere-proximal chromosome sequences.

**Characterization of new SFA1 dominant alleles**

As discussed above, we detected CFR clones that re-tested positive for the FA/Cu resistance phenotype but did not have any detectable chromosomal rearrangements. This class was particularly abundant in the mismatch-repair deficient strains suggesting a nucleotide mutation-based mechanism of resistance. In addition, because we identified several such isolates from diploid strains, these point mutations would have to be dominant / gain-of-function alterations in order to display the resistance phenotype in the presence of the second wild-type allele.

To characterize this possible mutation, we first crossed two msh2Δ haploid CFR isolates from this class (CFR49 and CFR50, MATα) to a wild type strain (JAY377, MATα). All
three haploid strains had only one copy of the CNV reporter, but CFR49 and CFR50 had the Kan G418 resistance marker linked to the reporter, whereas JAY377 had the Hph hygromycin B resistance marker at the same position (Figure 2A). We tested the corresponding diploids for FA/Cu resistance. Both diploids derived were FA/Cu resistant, confirming that the mutations in the two CFR clones were dominant (Figure 6A). Next, we sporulated and dissected tetrads from these diploids to examine the segregation pattern of the resistance phenotype. We observed two resistant and two sensitive spores, indicating that the mutant phenotype was controlled by only one locus. In addition, we also scored the spores for the Kan and Hph markers associated with CNV reporter. As expected, both markers segregated 2:2, and spores were either G418R HygS or G418S HygR. However, we noticed that FA/Cu resistance always co-segregated with G418 resistance. Since both CFR49 and CFR50 had CNV reporters marked with Kan, this result indicated that their FA/Cu resistance mutations were tightly linked to the CNV reporter, presumably in the SFA1 or CUP1 genes.

We then PCR amplified and sequenced the CNV reporter from thirteen independent CNV-less CFRs to find the nucleotide mutations. This analysis showed that all sequenced clones carried mutations in the SFA1 gene. Since the mutations were in SFA1, we repeated the resistance assays in plates containing FA or Cu separately. This analysis showed that the SFA1 mutations did not alter resistance to Cu (data not shown), and that their effect on resistance to FA was quite pronounced (Figure 6A). In fact, a single copy of the mutant SFA1 alleles conferred a level of FA resistance comparable to that seen in a strain carrying three copies of the wild type SFA1 gene (data not shown).

A total of five different dominant SFA1 alleles were identified. Two of the alleles arose independently multiple times: SFA1-Val208Ile (nucleotide mutation G622A) occurred in nine cases, and SFA1-Ala303Thr (nucleotide mutation G907A) occurred three times. Alleles
SFA1-Met283Ile, SFA1-Ile309Met, and SFA1-Pro313Ala occurred once each. All alleles sequenced were isolated from msh2Δ haploids, except for SFA1-Met283Ile that was isolated in a MSH2 haploid (CFR16).

To gain insight into the potential effect of these mutations on the enzymatic activity of Sfa1p, we analyzed them in the context of the human glutathione-dependent formaldehyde dehydrogenase enzyme (hFDH) for which a crystal structure has been solved (SANGHANI et al. 2002a; SANGHANI et al. 2002b). All five dominant alleles corresponded to residues that are conserved between the human and yeast enzymes (Figure 6B). Interestingly, the two alleles most frequently isolated (SFA1-Val208Ile and SFA1-Ala303Thr) are located in the active site of hFDH, in direct contact with the NAD(P)+ co-enzyme. This structural arrangement raises the possibility that mutation at these specific residues may destabilize the binding of the co-enzyme, facilitating its release from the active site after the reaction with the S-hydromethylglutathione. This would accelerate the rate at which FA is detoxified, therefore enhancing FA resistance without increasing the number of copies of the SFA1 gene.

DISCUSSION

A new assay for Copy Number Variation

In this study, we took advantage of the gene dosage-dependent phenotypes of the SFA1 and CUP1 genes to develop an optimized assay for the detection of gene amplifications. While several analogous systems have been previously reported in yeast, few of them have the sensitivity necessary to select low-order amplification events. In contrast, our system is capable of identifying spontaneous chromosomal rearrangements resulting in a simple doubling of the region of interest, going from one to two copies; and also subtler (50%) increases in gene dosage, going from two to three copies. This sensitivity makes the system
equally suited for the analysis of amplifications in haploid and diploid cells, whereas most previous systems function exclusively in haploids. We believe our system is an attractive and more germane model for the somatic CNVs commonly observed in humans (FROHLING and DOHNER 2008; GIRIRAJAN et al. 2011), and represents a valuable new tool for the investigation of structural genomic variation in a biomedically relevant context.

Another distinctive quality of this assay is that it allows the detection of chromosomal rearrangements initiated by rare spontaneous DNA lesions, regardless of where in the genome they occur. These lesions may include not only DSBs, but also uncapped telomeres, or collapsed replication intermediates that switch DNA template, or any other recombination initiating event. For example, in the case of the translocations detected in diploids, the precursor lesions likely occurred on Chr13 or on Chr7, but the CNV reporter was inserted on Chr5. Conventional assays that are based on deletions such as DEL or GCR (SCHIESTL 1989; CHEN and KOLODNER 1999), or amplification assays using haploids (PAYEN et al. 2008), are thought to detect rearrangements initiated predominantly by lesions in the immediate vicinity of their respective reporter sequences. The same issue is even more pronounced in assays that rely on the induction of site-specific DSBs (e.g. HO, I-SceI endonucleases) to raise the frequency of recombination up to a high level required for detection. The spontaneous and genome-wide nature the FA/Cu resistance assay is a highly desirable feature because, in principle, it should provide experimental results that more broadly represent global genomic instability processes, rather than the local properties of specific loci that may not necessarily apply to other regions.

The versatility of this detection system allowed us to conduct parallel genome stability experiments in the same chromosomal context in haploids and diploids. We observed that the classes of chromosomal rearrangements associated with each ploidy state were quite specific and distinct. In haploids we identified only segmental duplications, and, as expected,
no associated large deletions. Segmental duplications were generally also observed in most other previously reported haploid genome rearrangement assays, with the notable exception of the translocations detected in the recent iterations of the GCR assay, in which a deletion of the relatively small and non-essential terminal region of Chr5 left arm is associated with the duplication of a terminal segment from a different chromosome (Putnam et al. 2009; Chan and Kolodner 2011; Chan and Kolodner 2012). The diploid CFR clones displayed two main classes of rearrangements: about 15% were segmental duplications similar to those found in the haploids, but most others were non-reciprocal translocations resulting in amplification of the right arm of Chr5 (including the CNV reporter) coupled with a deletion of a terminal segment of a different chromosome. We also detected a clone (CFR512) containing a trisomy of Chr5, and incidentally, this same clone also carried two unselected trisomies of Chr3 and Chr16. These and other unselected karyotype changes were only observed in diploid clones. In summary, these data confirmed that diploid cells can tolerate a more diverse repertory of rearrangements than haploids, and therefore provide a richer platform to study genome stability.

Another key observation was that all chromosomal rearrangement breakpoints detected in this study contained dispersed repetitive DNA sequences, either full length Ty or solo LTR elements. No breakpoints were detected at single copy or microhomology sequences. This pattern was consistent with several other studies that implicated Ty sequences as hotspots for chromosomal rearrangements (Mieczkowski et al. 2006; Scheifele et al. 2009), and confirmed that NAHR is the primary DSB repair pathway responsible for such events in yeast. Interestingly, the spontaneous rearrangements selected in the CFR diploid clones were reminiscent of those we found in isogenic, G2-synchronized, diploids that survived exposure to 800 Gy of ionizing radiation resulting in ~250 DSBs per cell (Argueso et al. 2008). The NAHR-mediated segmental duplications and translocations
identified in that high dose γ-ray experiment were qualitatively indistinguishable from the ones selected spontaneously in this study. The only meaningful difference was the massive drop in the overall frequency of occurrence: 2-3 rearrangements per 800 Gy surviving cell, compared to one rearrangement per approximately \( 10^6 \) viable cells at the zero Gy dose. This suggests that the FA/Cu resistance assay may be successfully used to study CNV formation following low doses of radiation, as well as other clastogenic environmental exposures.

**Dominant formaldehyde hyper-resistant alleles of SFA1**

During the course of our study, we unexpectedly identified dominant mutations in the *SFA1* gene that increased the resistance of cells to formaldehyde, presumably through increased activity of the formaldehyde dehydrogenase (FDH) enzyme. This was intriguing and unusual, since mutations that result in the substitution of conserved amino acid residues at the core an enzyme’s active site are often associated with reduction or complete loss of activity and are typically recessive. This result also suggests that the natural selection forces that have acted on the sequence of *SFA1* stopped short of the maximum possible biochemical activity, therefore implying that a super-active FDH may somehow be associated with negative fitness consequences. Regardless of the their role in the evolution of *SFA1*, the fact that the mutations have significantly increased activity, and consequently a larger FA resistance phenotypic differential between one and two mutant copies, provided an opportunity for the improvement of this gene as a reporter for gene dosage. Recent work in our laboratories using the *SFA1* mutant alleles in the CNV reporter showed a pronounced reduction in background colony growth when cultures were plated at high cell densities (to be presented elsewhere), in effect enabling future versions of the CNV assay that will be fully quantitative.
Detection of recurrent chromosomal rearrangements

Our analysis of spontaneous amplifications selected from diploid clones revealed a remarkable recurrence of a specific chromosomal rearrangement between the right arms of Chr13 and Chr5. We took two experimental approaches to address this result. First, we created a derivative strain deleted for a central component of the DNA mismatch repair system \((msh2\Delta/msh2\Delta)\) that should relax sequence identity constraints imposed on recombination partner choice (GEORGE and ALANI 2012). This strain produced the very same rearrangements as the wild type, indicating that sequence similarity is not an important factor driving the recurrence of this translocation.

The second factor we investigated was the possible presence of a chromosomal fragile site on the right arm of Chr13. Previous studies from our group and others have shown that frequent breaks are associated with certain Ty sequences under conditions of replication stress, and that these breaks can be potent inducers of chromosomal rearrangements (CHA and KLECKNER 2002; LEMOINE et al. 2005). Replication stress is a well characterized activator of fragile site expression in humans (ARLT et al. 2012). However, it is important to note that the fragile nature may not be entirely dependent on replication stress. For example, the same yeast fragile sites have also been detected as hotspots for NAHR even under normal growth conditions (HOANG et al. 2010; CHAN and KOLODNER 2012).

We tested the candidate Chr13R fragile site in two ways. We deleted the \(YMRCTy1-5\) element and observed that in some cases the rearrangements recovered involved a proximal Ty element on the right arm of Chr13. We then examined spontaneous allelic recombination clones containing LOH on the right arm of Chr13 and found a statistically significant clustering of breakpoints consistent with the presence of a candidate fragile site in the region.
Taken together, these results suggested the possibility of a frequent DSB distal to the
YMRCTy1-5 element, which could be the trigger for the recurrent Chr13/Chr5 translocations.
Alternatively, a telomere with a partial defect in capping might also account for the high
incidence of Chr13R rearrangements.

In either the fragile site or weak telomere scenarios, a region of Chr13 at least ~150 kb
would have to be resected to expose the homology at YMRCTy1-5 and initiate NAHR.
Chromosomal rearrangements involving similarly long resection tracts (Break-Distal
Recombination; BDR) have been recently reported (HOANG et al. 2010; TAN et al. 2012).
While the DSB at Chr13R explanation is attractive, it is unlike to be the only factor involved.
Several other Ty repeats exist in the genome that should in principle be competent to
participate in translocations, yet none of those other sites were detected with the wild type
diploid strain. This suggests that there may be other significant contributing cellular
mechanisms.

We considered three other factors that may be favoring the detection of this specific
NAHR interaction in the FA/Cu resistance assay. The first is not related to a recombination
mechanism, but rather may be associated with the viability of clones. This would be possible
if cells carrying the Chr13/Chr5 translocation somehow had higher FA/Cu resistance, and
would therefore form colonies more often than cells with other rearrangements. We do not
think this is the case for two reasons. We compared the relative viability and growth of clones
carrying the Chr13/Chr5 rearrangements with clones carrying the Chr7/Chr5 translocation or
the Chr5 SDs in media containing FA/Cu (data not shown). These three classes of clones
appeared to be just as viable between them. In addition, even if Chr13/Chr5 translocation had
a more robust growth, this would not prevent the recovery of other, less robust
rearrangements that should also be viable in diploid cells.
Another factor that could create the observed detection bias is the relative proximity between potential recombination partner sequences in the tridimensional structure of the genome. This mechanism has been invoked as a contributor to the formation of recurrent chromosomal rearrangements in cancer cells (Wijchers and de Laat 2011). Thus, we analyzed the available structure of the \textit{S. cerevisiae} genome (Duan et al. 2010), and asked whether proximity between the right arms of Chr5 and Chr13 might explain the observed rearrangements. This analysis showed that the reverse is true: these two chromosomes are relatively farther apart from each other than from other chromosomes. This observation suggests that the relative position of a sequence in the static model of the genome may not be as important for recombination as its ability to move around the nucleus during DSB repair.

Two recent studies have elegantly demonstrated that yeast DNA sequences undergo a rapid transition from a relatively constrained nuclear localization to an increased mobility state immediately after a nearby DSB is induced (Dion et al. 2012; Mine-Hattab and Rothstein 2012). According to this model, a lesion on Chr13 would engage the Chr5 donor sequence not because of static proximity, but rather through a higher-than-normal degree of freedom to explore the nuclear space in search for homology. It would be interesting to investigate if the newly discovered transition from static to mobile status varies between different regions of the genome, and whether some sequences are able to explore larger nuclear volumes than others following breakage.

Finally, we should note that a recent haploid yeast study also identified recurrent non-reciprocal translocations involving dispersed Ty1 elements (Chan and Koledner 2012). In that study, six Ty1 elements (including \textit{YERCTy1-1}) were used as donor sequences repeatedly (~70% of cases) at frequencies much higher than those expected from random choice. As was the case in our analysis, the authors did not find a significant correlation between the relative spatial proximity in nucleus of the recombining elements, nor their
degree of nucleotide sequence similarity. However, their results and our own clearly showed that strong, yet unknown, mechanisms exist in the yeast nucleus that drive specific Ty repeats toward preferential NAHR with others. Understanding these processes could shed some much needed light into the origins of recurrent chromosomal rearrangements in cancer cells.

ACKNOWLEDGMENTS

We would like to thank Thomas Hurley for helpful discussions on the modeling on the SFA1 dominant mutations onto the crystal structure of the hFDH enzyme. The CNV research in the Petes, Argueso, and Mieczkowski laboratories was supported by an ARRA NIH-NIEHS Challenge Grant (5RC1ES018091-02). In addition, T.D.P. was supported by NIH grants GM24110 and GM52319, and J.L.A. was supported by grant ACS IRG #57-001-53 from the American Cancer Society.

FIGURE LEGENDS

Figure 1. General mechanisms for the generation of deletions and duplications.

In this figure, chromosomes are shown as thin horizontal blue or red lines, DNA repeats as black arrows, reporter genes as gray block arrows, and centromeres as circles. Single DNA strands are not represented. Variations of these models are also possible.

(A) Deletion formation by the single-strand annealing (SSA) pathway. A double-stranded DNA break (DSB) between the flanking repeats results in two broken ends that are resected 5’ to 3’. When complementary single-stranded regions of the flanking repeats are exposed, re-annealing occurs, resulting in loss of the sequences between the repeats (Paques and Haber 1999).
(B) Unequal crossovers resulting in deletions and duplications. This homologous recombination event could involve either sister chromatids in haploids or diploids, or homologous chromosomes in diploids.

(C) Microhomology-mediated replication. A free 3’ end associated with centromere-containing DNA fragment invades a sister chromatid or a homologous chromosome using a non-allelic microhomology sequence. The subsequent break-induced replication (BIR) event generates a duplication in which the breakpoints share little sequence homology. It is assumed that the DNA fragment without a centromere is lost.

(D) Coupled terminal duplication and deletion resulting in a non-reciprocal translocation. The different homologues are shown in blue and red. A DSB in or distal to a repeat in the blue chromosome is repaired by a BIR event using a repeat on a non-homologous chromosome (red). This event results in a duplication of sequences from the red chromosome and a deletion of sequences from the blue chromosome.

Figure 2. Identification of Chr5 amplifications in haploids: Segmental duplications.

(A) Schematic representation of the region of Chr5 where the SFA1-CUP1 CNV reporter was inserted (gray arrows). The reporter also contained a drug resistance marker, either Hph (hygromycin B) as shown, or Kan (G418 - geneticin). Terminal boxes correspond to the left and right Chr5 telomeres (5L and 5R, respectively), and the circle represents CEN5. Bottom, expanded view of the region involved in segmental amplifications (440 kb to 502 kb), showing full-length Ty1 elements as black block arrows and solo LTRs as arrowheads according to their orientation.

(B) FA/Cu resistance phenotypic differential between the parental haploid strain (JAY372) containing one copy of the reporter and CFR clones carrying two (CFR20) or three (CFR17)
copies. Serial dilutions of these strains were spotted on plates containing Cu and FA at the concentrations indicated to the left.

(C) PFGE showing the karyotypes of the parental haploid and three CFR clones carrying different segmental amplifications on Chr5. The parental size Chr5 is indicated to the left and the segmental amplifications are indicated to the right (5 S.D.).

(D) Detailed view of Chr5 array-CGH plots showing the Log2 Cy5/Cy3 signal intensity (copy number) and the rearrangement breakpoints for the three CFRs from C. The plots are aligned directly below to the expanded section of A such that the breakpoints correspond to the repeat sequences. Each blue dot corresponds to the signal of a specific probe in the region. The black horizontal lines correspond to the average signal for probes in a region of amplification. Neutral signal corresponds to no copy number change (1x in haploids); Positive signals corresponding to two copies (~1.0) or three copies (~1.5) are shown.

**Figure 3. Identification of Chr5 amplifications in diploids: Chr13/Chr5 translocation.**

Schematic representations are as in Figure 2. Chr5 is shown in red, Chr13 is shown in green.

(A) FA/Cu resistance phenotype in the parental diploid strain JAY350 (2 reporter copies) compared to a derivative resistant clone CFR3 (3 copies).

(B) PFGE showing the karyotype of the parental diploid strain JAY350 and of the derivative resistant clone CFR82 carrying the indicated recurrent 875 kb Chr13/Chr5 translocation. An identical translocation was also present in clone CFR3 shown in A. The parental size Chr5 and Chr13 are indicated to the left; Chr16 co-migrates with Chr13 and is also indicated.

(C) Chr5 array-CGH plot (top) and schematic representation (bottom) of the karyotype in the CFR75 clone showing one parental size homolog and one homolog containing a segmental duplication between the *YERCTy1-1* and *YERCTy1-2* elements (black arrowheads).
(D) array-CGH plots and schematic representation of Chr5 and Chr13 in the CFR82 clone. The breakpoints correspond to YERCTy1-1 in Chr5 and YMRCTy1-5 in Chr13. For position reference, the Chr13 centromere-proximal (left) element YMRCTy1-4 is also represented (see also Figures 4C and 5B).

Figure 4. Alternative translocations in diploids lacking the YMRCTy1-5 repeat element.
Schematic representations are as in Figure 2. Chr5 is shown in red, Chr13 is shown in green, Chr7 is shown in blue.

(A) PFGE showing the karyotypes of parent diploid strain JAY510, and of derivative clones CFR501 (containing a Chr7/Chr5 translocation) and CFR502 (containing a Chr13/Chr5 translocation). The parental bands for Chr5, Chr7, and Chr13 involved in the rearrangements are indicated to the left; also indicated are Chr16 and Chr15, which co-migrate with Chr13 and Chr7, respectively.

(B) array-CGH plots and schematic representation of Chr5 and Chr7 in the CFR501 clone. The breakpoints correspond to YERCTy1-1 in Chr5 and YGRWTy2-2/YGRCTy1-3 in Chr7.

(C) array-CGH plots and schematic representation of Chr5 and Chr13 in the CFR502 clone. The breakpoints correspond to YERCTy1-1 in Chr5 and YMRCTy1-4 in Chr13. Note the deletion of the YMRCTy1-5 element in this strain, represented by the gray “X” over the normal position of this element (distal/right arrowhead).

Figure 5. Initial mapping of a candidate fragile site on the right arm of Chr13.

(A) Experimental rationale. Schematic representation of Chr13 in the hybrid diploid strain (JAY800 and JAY801), formed by mating haploids of the diverged strain backgrounds YJM789 and CG379. The empty circles in the YJM789-derived chromosome represent SNP
positions. The \textit{\textit{KlURA3-Kan}} marker was inserted in the CG379-derived chromosome, downstream of the \textit{ADH6} gene, near the right telomere. Mitotic crossovers initiated by DNA breaks in the CG379-derived chromosome and repaired using the YJM789 homolog as template result in clones that are resistant to 5-FOA and homozygous for the right end of the YJM789-derived Chr13. Genotyping with SNP microarrays can determine the precise site of allelic mitotic recombination (LOH breakpoint, inverted triangle). This site should occur in close proximity to the precursor DNA lesion.

(B) LOH breakpoint positions for twelve independent 5-FOA resistant clones determined by SNP microarray genotyping showing a clustering of breakpoints near the right end of Chr13. Also indicated are the relative positions of the two Ty1 elements involved in the Chr13/Chr5 translocations described in Figures 3 and 4, and the position of the SNP-RFLP markers \textit{BglII} and \textit{EcoRI}. The \textit{EcoRI} marker is described in the RESULTS section. The genotype at \textit{BglII} marker site was also determined and found to be heterozygous in most 5-FOA\textsuperscript{R} clones (117/131; 89.3%).

**Figure 6.** Characterization of dominant mutant alleles of the \textit{SFA1} gene.

(A) Formaldehyde (FA) resistance phenotype in haploid (left) and heterozygous diploid (right) strains carrying three different mutant alleles of \textit{SFA1}. The genotypes are indicated at the top of the panel. Serial dilutions were plated in media containing 0 or 1.5 mM FA as indicated to the left. No \textit{CuSO}_4 was added.

(B) Amino acid residue sequence alignment between the \textit{S. cerevisiae} Sfa1p and the human FDH enzymes. Conserved residues (60.9% identity) are shown in black letters, and diverged residues are shown in gray. Block arrows point to the five residues that were substituted in the dominant \textit{SFA1} mutations.


ERHART, E., and C. P. HOLLenberg, 1983 The presence of a defective LEU2 gene on 2 μ DNA recombinant plasmids of Saccharomyces cerevisiae is responsible for curing and high copy number. Journal of Bacteriology 156: 625-635.


McCULLEY, J. L., and T. D. PETES, 2010 Chromosome rearrangements and aneuploidy in yeast strains lacking both Tel1p and Mec1p reflect deficiencies in two different mechanisms. Proceedings of the National Academy of Sciences of the United States of America 107: 11465-11470.


PUTNAM, C. D., V. PENNANEACH and R. D. KOLODNER, 2005 *Saccharomyces cerevisiae* as a model system to define the chromosomal instability phenotype. Molecular and Cellular Biology **25**: 7226-7238.


ROTHSTEIN, R., C. HELMS and N. ROSENBERG, 1987 Concerted deletions and inversions are caused by mitotic recombination between delta sequences in *Saccharomyces cerevisiae*. Molecular and Cellular Biology **7**: 1198-1207.


**Table 1.** Summary of the karyotype changes associated with the CNV reporter in the CFR clones.

<table>
<thead>
<tr>
<th>Karyotype changes</th>
<th>Haploids</th>
<th></th>
<th>Diploids</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT msh2Δ</td>
<td></td>
<td>WT msh2Δ</td>
<td>ymrcy1-5Δ</td>
</tr>
<tr>
<td>CFR clones isolated</td>
<td>26 22</td>
<td></td>
<td>29 22</td>
<td>16</td>
</tr>
<tr>
<td>Positive CFR upon re-test</td>
<td>24 21</td>
<td></td>
<td>26 21</td>
<td>10</td>
</tr>
</tbody>
</table>

Karyotype changes:

- **Chr5 segmental duplication**
  - a. Longer than normal Chr5 molecules harboring tandem duplications or triplications of a segment of Chr5 containing the CNV reporter flanked by direct repetitive full length Ty1 or solo δ LTR elements.
  - b. CFR511, a clone derived from the *ymrcy1-5Δ* homozygous diploid (JAY510), had a complex rearrangement consisting of a deletion of the right arm ofChr13 from *YMRCTy1-4* to the right telomere, joined by a segment of Chr5 containing a quadruplicaition of the segment containing the CNV reporter between *YERCTy1-1* and *YERCδ26*, in addition to one copy of the terminal segment between *YERCδ26* and *TEL05R*.
  - c. CFR510, a clone derived from the *ymrcy1-5Δ* homozygous diploid (JAY510), had an isochromosome 5 formed by loss of the left arm from *TEL05L* to the Watson Ty1 element inserted at the *ura3-52* allele, joined by the segment from right arm from *YERCTy1-1* to *TEL05R*. This isochromosome was present in addition to the two parental size copies of Chr5, both of which were retained in CFR510.
  - d. CFR512, a clone derived from the *ymrcy1-5Δ* homozygous diploid (JAY510), had three copies of the parental size Chr5, a simple aneuploidy of Chr5.

- **Chr13/Chr5 translocation**
  - 0 0 20 8 3

- **Chr7/Chr5 translocation**
  - 0 0 0 0 5

- **Chr5 isochromosome**
  - 0 0 0 0 1

- **Chr5 aneuploidy**
  - 0 0 0 0 1

- **No karyotype change**
  - 2 16 1 11 0

- CFR clones with CGH data
  - 9 5 9 10 10
Figure 1
Figure 2

A

5L 5R

SFA1-CUP1-Hph

YERCTy1-1 YERW020B YERW022 YERCTy1-2

440 kb 502 kb

B

Parent CFR20 CFR17

Cu 0 µM FA 0 mM

Cu 100 µM FA 1.0 mM

Cu 200 µM FA 1.5 mM

C

Parent CFR19 CFR20 CFR17

Chr5 S.D.
Figure 3

(A) Parent CFR3

Cu 0 µM
FA 0 mM

Cu 200 µM
FA 1.0 mM

(B) Parent CFR82

(C) CFR73-Chr5

D) CFR82-Chr5

Chr16
Chr13
Chr5

~875 kb
13/5

Chr13
Chr5

13L 13 13R

13L 13 5R
Figure 4
Figure 5

A

YJM789

13L - 13 - 13L

CG379

13L - 13 - 13L

5-FOA selection

URA3-Kan
911 kb

LOH breakpoint

B

13L - 13 - 13L

YMRCY1-4
378 kb

BglII
315 kb

YMRCY1-5
747 kb

EcoRI
871 kb

Figure 5
Figure 6

A

FA 0 mM

FA 1.5 mM

haploids
diploids

B

Yeast Sfa1p: MSAATAVGKPI KCI AAVYDA NKPLSV EEE T VADPK AHEVR 40
Human FDH: ANEV ----- I KCI AAVWEA NKPLS EEIE VAPPK AHEVR 35

Yeast Sfa1p: IKKEYTAVCH TDAVTL SGSD PEGLFP CVL VG HEGAGIVESV 80
Human FDH: IKK IATAVCH TDAVTL SGSD PEGLFP CVL VG HEGAGIVESV 75

Yeast Sfa1p: GDIVTVKRG DIVI AYTAE CGC CKFTSG KTNLC GAVRA 120
Human FDH: GECVTLKRG DIVI AYTAE CGC CKFTSNP KTNLC GAVRA 115

Yeast Sfa1p: TQCGI MDPG TRFHNK GE DITYHMG CST FSEYTVADV 160
Human FDH: TQCGI MDPG TRFHNK GE DITYHMG CST FSEYTVADV 154

Yeast Sfa1p: SVAAIDP KAP LDAACL LCGG VTTGE GA ALK TANVQG DT 200
Human FDH: SVAKIDP KAP LDAACL LCGG VTTGE GA ALK TANVQG DT 194

Yeast Sfa1p: AVFGC GVGVL SVIOG KLRG ASKKI AIDIN NKKSQYCSQF 240
Human FDH: AVFGC GVGVL AVFGC KVRG ASKKI AIDIN KKTKRABEK 234

Yeast Sfa1p: GATDFVNPKE DIAKDQIV E KLEMTD QG DFTF O GT 280
Human FDH: GATDFVPN Q Q DSKPQ E KLEMTD QG DFTF O GT 271

Yeast Sfa1p: KMRDALEAC HKGW GOSI E GAAEGEI A TRPFQ LTG 320
Human FDH: KMRDALEAC HKGW GOSI E GAAEGEI A TRPFQ LTG 311

Yeast Sfa1p: TWG5AFGGII KGRSEMCGLI KDOYQKA LK EE FTRRBFF 360
Human FDH: TWGTA FGGI KGRSEMCGLI KDOYQKA LK EE FTRRBFF 351

Yeast Sfa1p: KEINAFEDL HIGDCLRTVL KSDEIK 386
Human FDH: KEINAFELM HIGR SIVTVV K 373

Figure 6