Physical and Genetic Analysis of the CUP1 Tandem Array in the

*Yeast Saccharomyces cerevisiae*

by

Ying Zhao

Department of Molecular Genetics and Microbiology
Duke University

Date:_______________________

Approved:

___________________________
Thomas D. Petes, Supervisor

___________________________
Sue Jinks-Robertson

___________________________
Fred S. Dietrich

___________________________
John H. McCusker

___________________________
David M. MacAlpine

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Molecular Genetics and Microbiology in the Graduate School
of Duke University

2016
ABSTRACT

Physical and Genetic Analysis of the CUP1 Tandem Array in the Yeast

Saccharomyces cerevisiae

by

Ying Zhao

Department of Molecular Genetics and Microbiology
Duke University

Date:_______________________

Approved:

___________________________
Thomas D. Petes, Supervisor

___________________________
Sue Jinks-Robertson

___________________________
Fred S. Dietrich

___________________________
John H. McCusker

___________________________
David M. MacAlpine

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Molecular Genetics and Microbiology in the Graduate School of Duke University

2016
Abstract

The genomes of many strains of baker’s yeast, *Saccharomyces cerevisiae*, contain multiple repeats of the *CUP1* gene, encoding the copper-binding protein Cup1 (a member of the metallothionein family). The *CUP1* repeats are found in a tandem array on chromosome VIII. In this thesis, I describe studies that characterized these tandem arrays and their mechanism of formation across diverse strains of yeast. I show that *CUP1* arrays are an illuminating model system for observing recombination in eukaryotes, and describe insights derived from these observations.

In Chapter 2, we analyzed 101 natural isolates of *S. cerevisiae* in order to examine the diversity of *CUP1*-containing repeats across different strains. We identified five distinct classes of repeats that contain *CUP1*. We also showed that some strains have only a single copy of *CUP1*. By comparing the sequences of all the strains, we were able to elucidate the likely mechanism of formation of the *CUP1* tandem arrays, which involved unequal non-homologous recombination events starting from a strain that had only a single *CUP1* gene. Our observation of *CUP1* repeat formation allows more general insights about the formation of tandem repeats from single-copy genes in eukaryotes, which is one of the most important mechanisms by which organisms evolve.

In Chapter 3, we delved deeper into our mechanistic investigations by measuring the relative rates of inter-homolog and inter-sister chromatid recombination in *CUP1*
tandem arrays. We used a diploid strain that is heterozygous both for an insertion of a selectable marker (URA3) inside the tandem array, and also for markers at either end of the array. The inter-sister chromatid recombination rate turned out to be more than ten-fold greater than the inter-homolog rate. We also found that the majority of these inter-sister events do not reflect unequal sister-chromatid recombination but likely occur by single-strand annealing, inter-sister gene conversion, or DNA polymerase slippage.

Moreover, loss of the proteins Rad51 and Rad52, which are required for most types of homologous recombination, did not greatly reduce recombination within the CUP1 array. Additionally, we investigated the effects of elevated copper levels on the rate of each type of recombination at the CUP1 locus. Both types of recombination are increased by high concentrations of copper (as is the case for CUP1 transcription). Furthermore, the inter-homolog recombination rate at the CUP1 locus is higher than the average over the genome during mitosis, but is lower than the average during meiosis.

The research described in Chapter 2 was published in 2014.
Dedication

Dedicated to my grandmother, my parents and my husband.
Contents

Abstract ............................................................................................................................................. iv
List of Tables ....................................................................................................................................... xi
List of Figures ...................................................................................................................................... xii
Acknowledgements ........................................................................................................................... xiv
1  Introduction ...................................................................................................................................... 1
  1.1  Mechanisms of homologous recombination and non-homologous end-joining 1
    1.1.1  Homologous recombination pathways for DSB repair ................................................. 2
      1.1.1.1  DSB end resection ........................................................................................................ 2
      1.1.1.2  Synthesis-dependent strand annealing (SDSA) and DSB repair (DSBR) pathways ............................................................ 4
      1.1.1.3  Break-induced replication (BIR) pathway ......................................................... 9
      1.1.1.4  Single-strand annealing (SSA) pathway ......................................................... 11
    1.1.2  Non-homologous end-joining (NHEJ) for DSB repair ........................................... 16
  1.2  Systems to study inter-sister chromatid recombination .............................................. 19
  1.3  CUP1 tandem repeats and copper homeostasis ............................................................... 32
    1.3.1  The unique structure of the CUP1 tandem array .................................................... 32
    1.3.2  Copper homeostasis in the yeast Saccharomyces cerevisiae .................................... 35
2  Structures of naturally-evolved CUP1 tandem arrays indicate that these arrays are generated from single-copy CUP1 genes by unequal non-homologous recombination .. 42
  2.1  Introduction ............................................................................................................................. 42
  2.2  Material and methods ............................................................................................................. 48
2.2.1 Yeast strains .................................................................48
2.2.2 Measurement of the sizes and sequences of CUP1 repeats........49
2.2.3 Analysis of the number of CUP1 copies in tandem arrays.........49
2.2.4 Southern analysis ..................................................................50
2.2.5 Measurements of copper resistance.......................................51
2.2.6 Measurements of the rate of loss of a URA3 gene integrated within the CUP1 repeats in YZ22.................................................................51
2.2.7 DNA sequencing ....................................................................52
2.3 Results .........................................................................................54
2.3.1 Identifying five different CUP1 repeats.................................54
2.3.2 Sequence analysis of CUP1 repeats ......................................59
2.3.3 Variation in the number of CUP1 repeats in different yeast strains........62
2.3.4 Copper resistance of strains with different types and different numbers of CUP1 repeats.................................................................64
2.3.5 Rate of reduction of a tandem CUP1 array to a single copy of CUP1........67
2.4 Discussion......................................................................................69
2.4.1 Previous observations of gene duplications in yeast ..............69
2.4.2 Duplications of CUP1 in yeast...................................................71
2.4.3 Alterations in the number of repeats within the CUP1 tandem array........76
2.5 Summary .......................................................................................77
2.6 Supplementary tables..............................................................77

3 Properties of mitotic and meiotic recombination in the tandemly-repeated CUP1 gene cluster in the yeast Saccharomyces cerevisiae.................................................................175
## 3.1 Introduction

175

## 3.2 Material and methods

179

### 3.2.1 Yeast medium

179

### 3.2.2 Yeast strains and plasmids

179

### 3.2.3 Measurements of recombination rates

182

### 3.2.4 Southern Analysis

184

### 3.2.5 Statistical analysis

185

### 3.2.6 Meiotic analysis

185

## 3.3 Results

186

### 3.3.1 Experimental rationale

186

### 3.3.2 Rates of inter-homolog crossovers within the *CUP1* array and in the *CEN5-hphMX4* interval in YZ103 and YZ104

195

### 3.3.3 Rates of intra-/sister-chromatid events in YZ103 and YZ104

197

### 3.3.4 Rate of inter-homolog gene conversion in YZ104

197

### 3.3.5 Loss of *CUP1* repeats in YZ104 strains that have intra-/sister-chromatid recombination events

200

### 3.3.6 Stimulation of mitotic recombination within the *CUP1* cluster by addition of copper to the growth medium

200

### 3.3.7 Loss of the *URA3* gene from the *CUP1* tandem array in YZ103 is not a consequence of unequal sister-chromatid exchange

203

### 3.3.8 Analysis of the genetic regulation of mitotic recombination within the *CUP1* array

205

### 3.3.9 Meiotic recombination in the *CUP1* array

208

## 3.4 Discussion

211
3.4.1 Relative frequency of inter-homolog and intra-/sister-chromatid events...211
3.4.2 Mechanism of intra-/sister-chromatid events ........................................212
3.4.3 Stimulation of CUP1 recombination by transcription..........................218
3.4.4 Meiotic recombination........................................................................219
3.5 Summary..................................................................................................220
3.6 Supplementary tables.............................................................................221

4 Discussion and future directions................................................................234
4.1 De novo duplications from a single-copy gene ........................................234
4.2 De novo duplications of a single copy of the CUP1 gene .......................239
4.3 Mechanism of intra-/sister-chromatid recombination.............................243

References....................................................................................................245
Biography......................................................................................................271
List of Tables

Table 1.1. Random clone analysis of SCneo recombination events. ......................................................... 31

Table 1.2. SCneo homologous recombination events. .................................................................................. 32

Table 2.1. Analysis of CUP1 gene tandem arrays in 14 yeast strains of Saccharomyces cerevisiae. ................................................................. 66

Table S 2.1 Strain genotypes. ......................................................................................................................... 77

Table S 2.2. Primer names and sequences used in strain constructions and analysis. .................................. 79

Table S 2.3. Sequence analysis of the CUP1 repeats (Type 2, 1.8 kb) of YJM189. .......................... 81

Table S 2.4. Sequence analysis of the CUP1 repeats (Type 2, 1.8 kb) of YJM996. ............................ 92

Table S 2.5. Sequence analysis of the CUP1 repeats (Type 3, 1.2 kb) of YJM789. .......................... 102

Table S 2.6. Sequence analysis of the CUP1 repeats (Type 4, 1.9 kb) of YJM271. .......................... 112

Table S 2.7. Sequence analysis of the CUP1 repeats (Type 4, 1.9 kb) of YJM130. .......................... 123

Table S 2.8. Sequence analysis of the CUP1 repeats (Type 5, 1.6 kb) of YJM456. .......................... 136

Table S 2.9. Sequence analysis of the CUP1 repeats (Type 5, 1.6 kb) of YJM969. .......................... 147

Table S 2.10. Sequence analysis of strain (DTY3) that has a single copy of CUP1. .................. 158

Table 3.1. Inter-homolog and intra-/sister-chromatid mitotic recombination rates in wild-type and mutant yeast strains. .......................... 193

Table 3.2. Rates of inter-homolog and intra-/sister-chromatid in cells grown in medium without added copper, and in medium with 1.4 mM copper. .......................................................... 194

Table S 3.1. Strain genotypes and constructions. .......................................................................................... 221

Table S 3.2. PCR primer names and sequences. ......................................................................................... 227

Table S 3.3. Plasmids used in this study. ................................................................................................. 232
List of Figures

Figure 1.1. Pathways of DSB repair by homologous recombination.......................... 6

Figure 1.2. Single-strand annealing pathway for DSB repair. ................................. 13

Figure 1.3. Sister-chromatid-exchange (SCE) substrate........................................ 21

Figure 1.4. Three types of recombination events generating ADE3.......................... 24

Figure 1.5. Recombination substrate for detecting inter-sister chromatid and inter-
                                                   homolog recombination. Modified from Mozlin et al. (Mozlin et al. 2008).......... 27

Figure 1.6. Recombination substrate Scneo and repair products.............................. 30

Figure 1.7. Structure of the CUP1 locus in S288c................................................ 34

Figure 1.8. Copper homeostasis in yeast.............................................................. 38

Figure 2.1. Mechanisms for altering or generating tandem gene arrays.................... 44

Figure 2.2. Structures of the CUP1 loci in different strains.................................... 56

Figure 2.3. Analysis of PCR fragments representing various types of CUP1 repeats..... 58

Figure 2.4. Sequences at the breakpoints for five different types of CUP1 repeats....... 61

Figure 2.5. Southern analysis of the CUP1 genes in 13 yeast strains.......................... 63

Figure 2.6. Correlations between different assays of CUP1 copy-number and between
                                                   CUP1 copy-number and copper resistance......................................................... 65

Figure 2.7. Loss of a CUP1 tandem array by unequal crossing-over.......................... 68

Figure 2.8. Generation of a de novo duplication of CUP1 by unequal NHEJ............... 73

Figure 2.9. Clustered breakpoints of five different types of CUP1 repeats............... 75

Figure 3.1. Diploid strains (YZ103 and YZ104) used for the detection of inter-homolog
                                                   recombination between CEN8 and CUP1, and within the CUP1 array..................... 188
Figure 3.2. Mitotic recombination events leading to loss of URA3 marker and retention of the flanking hphMX4 and CAN1/kanMX4 markers.........................................................192

Figure 3.3. Loss of the URA3 insertion by single-strand annealing (SSA) or DNA polymerase slippage..........................................................................................................................196

Figure 3.4. Detection of inter-homolog gene conversion by PCR.........................................................199

Figure 3.5. Numbers of remaining CUP1 repeats associated with loss of URA3 on the W303-1A-derived chromosome VIII homolog of YZ104......................201

Figure 4.1. Mechanisms for generating CNVs in yeast. .................................................................236
Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Tom Petes, for welcoming me to his lab and expending so much time and effort training me as a scientist. I feel very lucky to have such a brilliant and successful scientist as my advisor. His pursuit of excellence and passion for science inspire me profoundly and set a quintessential example of the drive I would like to exhibit in all my future professional endeavors. I would also like to thank my co-advisor, Robert Kokoska, for his great guidance and mentorship. Tom and Bob provided invaluable insights and guidance throughout all the stages of my graduate studies. They also devoted substantial efforts to perfecting my writing and presentations. I have matured tremendously both intellectually and professionally under their tutelage.

I would like to thank all the members of my committee: Sue Jinks-Robertson, Fred Dietrich, John McCusker and David MacAlpine. In particular, Sue has provided very valuable advice and support for both of my projects. Fred and John, who collaborated with me during my first project, provided important data and critical insights about yeast genetics that allowed the project to be successful. David has provided highly novel and interesting perspectives based on his expertise. I would also like to thank all the members of the SJR lab for their valuable conversation and comments as well as their generous help and support.
I would like to thank all the members of the Petes Lab. Margaret Dominska kindly spent a lot of time teaching me many important lab techniques. She has also been greatly helpful on several experiments towards the end of my second project. In addition, I am very thankful that Margaret Dominska and Pat Greenwell contributed greatly to our lab’s daily procurement and maintenance as our current and previous lab manager. I want to thank all the other current members of the Petes lab—Anthony Moore, Dennis Kiktev, Aleksandra Petrova, Daoqiong Zheng, Ke Zhang, Lijia Zhang and Halle Bagshaw— for being wonderful colleagues and great collaborators. I also want to thank the past lab members I have worked with—Jordan St. Charles, Wei Song, Sabrina Anderson Cote, Eunice Yim, Felix Kung, Yi Yin and Karen O’Connell—who have been very helpful and supportive. I have learned a great deal from each of them.

I would also like to thank my entire family for being supportive during my time in graduate school. In particular, I want to thank my grandmother, a successful chemist and entrepreneur, who piqued my interest in science since I was little. She passed away during my time in grad school. We had a very special bond, and her passing was a big loss for me. I would also like to thank my parents, who have kept me motivated and passionate about my life and career since I was a child by providing me with their unconditional love and support. My husband, Mark Hallen, has been my role model and cornerstone the whole time. Finally, I would like to thank Mark’s family, who have been very understanding and supportive and helped me feel at home in North Carolina.
1 Introduction

The stability and integrity of DNA, the building block of genomes, is crucial for all life forms. DNA double-strand breaks (DSBs) pose a severe threat to genome stability and cell viability if left unrepaired (Fairhead and Dujon 1993). DSBs can be generated naturally by stalled or collapsed replication forks at regions that are intrinsically susceptible to breakage (fragile sites). In yeast, certain repetitive DNA sequences (Mirkin 2007; Tang et al. 2011; Andersen et al. 2015), and sequences with the ability to form secondary structures (Casper et al. 2009; Gordenin et al. 1993; Lemoine et al. 2005; Lemoine et al. 2008) or to perturb the progression of replication forks (Cha and Kleckner 2002; Paeschke et al. 2011; Song et al. 2014) are hotspots for chromosome breakage. Moreover, programmed genome rearrangements induced by specific enzymes can also lead to DSBs, such as HO-induced DSBs during yeast mating type switching and Spo11-induced DSBs during yeast meiosis. DSBs can also be generated when DNA is exposed to external damaging agents, including agents that induce DNA lesions directly (for example, ionizing radiation (IR)) or agents that lead to replication fork blocks.

1.1 Mechanisms of homologous recombination and non-homologous end-joining

Homologous recombination (HR) is an important mechanism for DSB repair (DSBR) in eukaryotic cells. In S. cerevisiae, the majority of DSBs are repaired by HR (Paques and Haber 1999). In mammalian cells, another major pathway is non-
homologous end-joining (NHEJ). Homologous recombination has been found to account for 30–50% of the events repairing endonuclease-induced DSBs in one of two direct repeats (Liang et al. 1998). Rothkamm et al. examined mammalian cells defective in NHEJ or HR and showed that most DSBs in G1 were repaired by NHEJ, while the majority of replication-associated DSBs were repaired by HR (Rothkamm et al. 2003). This observation is consistent with the notion that the NHEJ pathway requires two broken ends for repair, and many DSBs generated during replication are one-ended DSBs (Cheong et al. 1994).

### 1.1.1 Homologous recombination pathways for DSB repair

Many genes and proteins important or essential to homologous recombination in the yeast *S. cerevisiae* were identified using mutants that are hypersensitive to DNA damaging agents, such as ionizing radiation. Based on studies of these mutants using a variety of recombination assays, at least four main HR pathways are identified: canonical DSB repair (DSBR), synthesis-dependent strand annealing (SDSA), break-induced replication (BIR), and single-strand annealing (SSA) (Paques and Haber 1999) (Figure 1.1).

#### 1.1.1.1 DSB end resection

5’-3’ end resection of the broken ends is an early step for all types of HR. The resection is usually initiated by the MRX complex and Sae2. The MRX complex consists
of three components: Mre11, Rad50 and Xrs2. These three components form hexamers consisting of dimers of each component (van der Linden et al. 2009; Williams et al. 2010). Mre11 and Rad50 are conserved from prokaryotes to higher eukaryotes, whereas Xrs2 and its human counterpart Nbs1 exist only in eukaryotes (Connelly and Leach 2002).

Direct physical evidence, including 3-D structural analysis of components of the MRX complex, indicates that the coiled coil domains of dimerized Rad50 act as the hinge of this complex by directly interacting with Mre11 (de Jager et al. 2001; Hopfner et al. 2002; Dolganov et al. 1996). This observation is consistent with earlier findings suggesting close interactions between Mre11 and Rad50 (Trujillo et al. 1998; Johzuka and Ogawa 1995). In yeast, Xrs2 binds DNA and helps to direct Rad50 and Mre11 to DSB broken ends (Trujillo et al. 2003; Paull and Gellert 1998). The coiled coil also provides a flexible tether for DNA molecules separated by up to 1200 Å (Hopfner et al. 2002). Therefore, the MRX complex both tethers DNA ends and initiates end resection (Mimitou and Symington 2008; Shim et al. 2010; Zhu et al. 2008).

It has been observed that MRX stimulates the 5’-end resection activity of exonucleases Exo1, Dna2 and helicase Sgs1 via cooperative binding of DNA substrates (Chiolo et al. 2005; Zhu et al. 2008; Cejka et al. 2010a; Nicolette et al. 2010). It was also reported that the MRX complex is indispensable in recruiting Exo1, Dna2 and Sgs1 to the DSB sites (Shim et al. 2010).
Extensive resection is thereafter carried out by Exo1 or the combined activities of Sgs1-Top1-Rmi1 (STR) complex and Dna2 (Mimitou and Symington 2009). Studies of Spo11-bound DSB end resection analysis in meiosis indicate that initial end resection removes oligonucleotides of 12-34 nt from the 5’-end (Neale et al. 2005). The average length of end resection in meiotic cells is approximately 850 nt, but is reduced to 270 nt in the absence of Exo1 (Zakharyevich et al. 2010). However, end resection as long as 2-4 kb was observed during mitotic repair between homologous chromosomes (Chung et al. 2010).

The initiation of end resection is a critical step that commits the repair of DSBs to HR since once resection has initiated, the DSB ends become poor substrates for binding by the Ku proteins, a prerequisite for NHEJ (Ira et al. 2004; Zhang et al. 2009b). Resection is regulated during the cell cycle; end resection is most prominent in S and G2 phases, when the sister chromatid becomes available as the template for repair after DNA replication has taken place (Barlow et al. 2008).

1.1.1.2 Synthesis-dependent strand annealing (SDSA) and DSB repair (DSBR) pathways

After extensive 5’-end resection, a long stretch of ssDNA at the 3’-ends of DSBs can be accessed by the single-stranded DNA-binding protein Replication Protein A (RPA), and strand transfer proteins Rad51 (in mitosis) and Rad51 and Dmc1 (in meiosis). These two recombinases are the eukaryotic counterparts of the prokaryotic
recombinase RecA. Dmc1 functions exclusively in meiosis (Bishop et al. 1992), whereas Rad51 plays a major role in mitosis (Paques and Haber 1999).

A dichotomy of stimulatory and inhibitory roles of RPA in facilitating Rad51 binding to the resected 3’-end and the assembly of the Rad51 presynaptic filament has been reported (Sung 1994, 1997; New et al. 1998; Shinohara and Ogawa 1998). On one hand, RPA removes secondary structures of ssDNA (Sugiyama et al. 1997) and sequesters ssDNA generated during HR (Eggler et al. 2002; Van Komen et al. 2002), thus helping Rad51 bind ssDNA and form the presynaptic filament. On the other hand, high concentrations of RPA can saturate ssDNA locally and hinder Rad51 binding (Shinohara and Ogawa 1998; New et al. 1998). As described below, one role of the Rad52 protein is to facilitate the replacement of RPA by Rad51.

Rad52 acts as the recombination mediator facilitating 3’-ssDNA binding and presynaptic filament formation by Rad51 because otherwise, the interactions between RPA, Rad51 and ssDNA would be stochastic and probably unfavorable for Rad51-ssDNA binding. Rad52 forms a homoheptameric ring and plays a pivotal role in recruiting Rad51 to ssDNA substrates. Mutations in the domain of Rad51p that binds Rad52p attenuate the recombination mediator effect of Rad52 (Shinohara and Ogawa 1998; Krejci et al. 2002). In S. cerevisiae, Rad52 also has ssDNA-binding activity (Mortensen et al. 1996). Moreover, ScRad52 has been found to interact with RPA by
Figure 1.1. Pathways of DSB repair by homologous recombination.

In this figure, three pathways of DSB repair by homologous recombination are shown. DNA strands from two homologs are shown in orange and blue; dotted orange and blue lines indicate newly-synthesized DNA. Regions of the duplex containing strands of different colors represent heteroduplexes. This figure is modified from Stafa et al. (Stafa et al. 2014).

A. Synthesis-dependent strand annealing (SDSA) pathway. Following 5’ end resection of the DSB, the 3’ single-stranded DNA of the right end of the DSB invades the template duplex and forms the D-loop. The invading strand primes DNA synthesis using one strand of the blue homolog as the template, becomes unstable, and is then displaced. This 3’ end hybridizes to the left end of the DSB. This pathway gives rise to gene conversion events unassociated with crossovers.
B. Double strand break repair (DSBR) pathway. In this pathway, a double Holliday junction (dHJ) is formed after second-end capture. The dHJ can be resolved by either dissolution or resolution. In the dissolution pathway, Holliday junctions migrate towards each other and decatenate to form noncrossover products. In the resolution pathway dHJ can be cleaved by Holliday junction resolvases, forming crossover or noncrossover products.

C. Break-induced replication (BIR) pathway. One broken end of the DSB invades the homologous chromosome and forms the D-loop. The D-loop migrates from the point of invasion to the telomere and and primes DNA synthesis using the invaded strand as template.

directly binding the largest and middle subunits of RPA (Shinohara et al. 1998; Jackson et al. 2002). All these findings corroborate a model in which the Rad52 heptameric ring binds the Rad51 oligomer and delivers it to the 3’-end of ssDNA, which is usually coated by RPA.

Once the Rad51-ssDNA nucleoprotein filament (the presynaptic filament) is formed, Rad51 functions as the recombinase and searches for a homologous template for repair. Studies of the homology search by RecA indicate that it occurs by random collisions (Bianco et al. 1998). Upon finding the homologous sequence, the presynaptic filament can interact further with the duplex DNA template, either by canonical hydrogen bonding or by formation of a three-stranded intermediate (Gupta et al. 1999). The invading strand and the duplex DNA thereafter form a region of heteroduplex DNA (hDNA) and a displaced, single-stranded D-loop (Figure 1.1, top). As the invading 3’-end primes DNA synthesis using the homologous template, the D-loop enlarges.
The rest of the repair process can continue by either of two mechanisms: synthesis-dependent strand annealing (SDSA) or canonical DSB repair (DSBR). In the SDSA model (Figure 1.1A), when the D-loop collapses, the extended 3’-end is displaced from the duplex DNA and anneals to the 3’-ssDNA on the other end of the DSB (Figure 1.1). This mechanism gives rise exclusively to gene conversion (GC) events unassociated with crossovers (COs), also referred to as noncrossovers (NCOs) (Paques and Haber 1999).

Alternatively, in the canonical DSBR model (Figure 1.1B), the second 3’–end of the broken chromosome, which is not involved in strand invasion, anneals to the D-loop and stabilizes it. This process called “second-end capture” results in the formation of a double Holliday junction (dHJ) (Szostak et al. 1983). Second-end capture is an important step that commits the repair pathway to DSBR. *In vivo* physical analysis has identified an essential role of Rad52 in second end capture, presumably mediated by its affinity for RPA and RPA-coated ssDNA (Nimonkar et al. 2009; Mazloum and Holloman 2009; McIlwraith and West 2008; Lao et al. 2008).

There exist two pathways for resolving dHJs. In the dissolution pathway, Holliday junctions migrate towards each other and decatenate to give rise to noncrossover products. The Sgs1–Top3–Rmi1 (STR) complex was found to play a salient role in dHJ dissolution by this pathway (Cejka et al. 2010b). Alternatively, the dHJ can be
cleaved by Holliday junction resolvases, generating crossover or noncrossover products. In mitotic yeast cells, Mus81/Mms4 complex and Yen1 are the main resolvases for dHJ cleavage (Ho et al. 2010). Mus81/Mms4 complex has a high affinity for nicked HJs (Osman et al. 2003; Heyer 2004). Yen1 is the less important of the two in dHJ resolution, supported by the fact that yen1Δ cells barely showed any phenotypic defects (Schwartz and Heyer 2011). In contrast, Mus81 is indispensable when the dHJ dissolution complex Sgs1–Top3–Rmi1 is unavailable (Mullen et al. 2001). As will be discussed later, mitotic crossovers in diploid cells can lead to loss of heterozygosity (LOH), which can predispose cells to tumorigenesis (Knudson 2001).

1.1.1.3 Break-induced replication (BIR) pathway

A third pathway of HR is BIR. In this mechanism, as shown in Figure 1.1C, one of the broken ends is lost. The other end forms a D-loop that migrates by conservative DNA synthesis to the end of the chromosome (Donnianni and Symington 2013; Saini et al. 2013). BIR serves effectively to repair one-ended DSBs or DSBs in which only one end shares homology with a template.

Bosco and Haber and Malkova et al. studied the repair of a HO-induced DSB located on chromosome III of S. cerevisiae (Bosco and Haber 1998; Malkova et al. 2000). In these experiments, the centromere-distal side of the DSB did not share any sequence homology to the rest of the genome, while the centromere-proximal side of it shared
some 70 bp homology to an ectopic sequence. They found that BIR repair led to a 30 kb non-reciprocal translocation centromere-distal to the DSB. Moreover, repair of this HO-induced DSB was highly dependent on Rad52, but only partially dependent on Rad51; in rad52Δ diploid strains, 99% of cells lost the broken chromosome; while in rad51Δ diploid strains, only 44% cells lost the broken chromosome (Malkova et al. 1996).

However, this conclusion about the relative effects of Rad51 and Rad52 might be dependent on the context of the repair event. In a chromosome fragmentation assay used to measure the efficiency of repair by BIR, it was shown that 95% of BIR events were dependent on Rad51 (Davis and Symington 2004; Donnianni and Symington 2013). Furthermore, Lydeard et al. found that BIR is also highly dependent on POL32 (a sub-unit of DNA polymerase delta), in addition to RAD51 and RAD52 (Lydeard et al. 2007).

Two models were proposed to explain the mechanism of BIR. In the conservative DNA synthesis model, the D-loop created after strand invasion migrates towards the end of the chromosome, displacing a newly-synthesized DNA strand. In the semi-conservative DNA synthesis model, nuclease cleavage of the D-loop recombination intermediate transforms it into a processive replication fork. As the replication fork migrates, newly-synthesized strands segregate into the donor and recipient molecules. In this second model, the recipient molecule is generated by semi-conservative DNA synthesis. Studies by Donnianni and Symington (2013) and Saini et al. (2013)
demonstrated that the first model was correct. In addition, Donnianni and Symington (2013) showed that BIR efficiency decreased as the donor sequence was moved further away from the telomere, requiring more DNA synthesis.

The DNA synthesis associated with BIR is less processive than “normal” DNA synthesis. Smith et al. discovered that BIR sometimes involves by multiple rounds of strand invasion, limited DNA synthesis and dissociation of the invading strand with the template, and re-invasion (Smith et al. 2007). When template switches between dispersed repeats occurred during BIR, non-reciprocal translocations were formed. Such chromosome rearrangements are often deleterious (Bordeianu et al. 2011). In mammalian cells, a microhomology-mediated BIR model involving multiple rounds of template switching was proposed to account for the origin of human gene copy number variations (Hastings et al. 2009). In this model, 3’-ssDNA anneals to sequences based on microhomology, and primes DNA synthesis with a low-processivity DNA polymerase. Subsequently, multiple rounds of template switching occur until processive DNA synthesis is re-established. Complex and nonrecurrent genome rearrangements could be generated by this process.

1.1.1.4 Single-strand annealing (SSA) pathway

When DSBs are flanked by repetitive sequences, a fourth pathway for DSB repair, SSA, can occur (Figure 1.2). In vivo studies in yeast demonstrated that in SSA, 5’ –
3’ resection of DSB ends creates single-stranded 3’ overhangs, which are coated by RPA (Fishman-Lobell et al. 1992; Sugawara and Haber 1992; Lyndaker and Alani 2009).

Single-stranded regions thereafter extend to flanking repetitive sequences so that complementary single-stranded DNAs can anneal to each other. In both yeast and mammalian cells, Rad52 has an important role in SSA in promoting the reannealing of complementary single-stranded DNAs coated by RPA (Mortensen et al. 1996; Sugiyama et al. 1998; Mortensen et al. 2002; Grimme et al. 2010).

It has been reported that another member of the RAD52 epistasis group, RAD59, is also required for Rad51-independent HO-induced deletion formation between direct repeats and inverted repeats by SSA (Bai and Symington 1996; Sugawara et al. 2000). Similar to Rad52, Rad59 binds to both ssDNA and dsDNA, with higher affinity for ssDNA, and promotes annealing of complementary strands (Petukhova et al. 1999; Wu et al. 2006b; Wu et al. 2006a). However, unlike Rad52, annealing of ssDNA by Rad59 is not accelerated by RPA (Wu et al. 2006a). Davis and Symington found that in vivo Rad59 cannot outcompete RPA for binding ssDNA. Therefore, the strand-annealing activity by Rad59 itself is not sufficient for SSA in vivo, although in vitro studies demonstrated an independent strand-annealing function of Rad59 (Davis and Symington 2001).

Overexpression of Rad52p was found to suppress defects in mitotic recombination in a rad59 mutant (Bai and Symington 1996). In addition, certain mutant
alleles of \textit{RAD52} conferred similar recombination defects to a null mutant of \textit{RAD59} (Bai \textit{et al.} 1999). Subsequently, definitive biochemical evidence for the physical interactions

![Figure 1.2. Single-strand annealing pathway for DSB repair.](image)

In this figure, the mechanism of single-strand annealing (SSA) is shown. Each blue horizontal line designates a single DNA strand. Blue vertical lines designate repetitive sequences, with three repeats on each side of the DSB. Following 5' - end resection, ssDNA containing homology is exposed and annealed. Nonhomologous 3' tails are processed by endonucleases. Gaps are filled in by DNA synthesis. This mechanism usually results in deletions of repetitive sequences.
between Rad52 and Rad59 was obtained by the co-immunoprecipitation (Davis and Symington 2001). All of these observations, combined, suggest that Rad59 promotes strand annealing synergistically with Rad52 by formation of a Rad52-Rad59 complex.

Upon strand-annealing between repetitive sequences promoted by Rad52-Rad59, the 3’–ssDNA ends form single-stranded branches on both sides of the annealed region. These branches must be removed before gap-filling DNA synthesis can occur. Genetic evidence has shown that the Rad1-Rad10 excision endonuclease complex (ERCC1-XPF in mammals) is involved in removal of these branches (de Laat et al. 1998). This complex is important in both nucleotide excision repair (NER) (Fagbemi et al. 2011) and SSA (Fishman-Lobell and Haber 1992; Ivanov and Haber 1995; Prado and Aguilera 1995; Bailly et al. 1992). Rad1 and Rad10 have been reported in multiple studies to be required for mitotic recombination between repeated sequences in yeast or between substrates that require removal of heterologous tails (Dowling et al. 1985; Klein 1988; Thomas and Rothstein 1989b; Lyndaker and Alani 2009; Schiestl and Prakash 1990). In addition, Saw1 and Slx4 were discovered to be important in recruiting Rad1-Rad10 complex to the 3’ single-stranded “tails” (Li et al. 2008; Lyndaker and Alani 2009; Flott et al. 2007).

In addition, the Msh2-Msh3 complex, which recognizes DNA base mismatches and is important in DNA mismatch repair, also functions in SSA. Of the six S. cerevisiae homologs of the E. coli MutS protein, the two most important heterodimers for mismatch

Msh2-Msh3 is also important in the Rad1-Rad10-dependent removal of nonhomologous 3' tails (Sugawara et al. 1997). In vitro studies showed that Msh2-Msh3 binds DNA substrates containing 3'–ended ssDNA tails with an affinity comparable to its binding to a mismatched DNA loop of 8 bp (Surtees and Alani 2006). Physical evidence showed that Msh2-Msh3 binds the junctions of double-stranded and single-stranded DNA and opens up the junction slightly, potentially making it more accessible to the Rad1-Rad10 complex (Surtees and Alani 2006). During meiosis in yeast, Msh2-Msh3 and Rad1-Rad10 also function together to remove large loops within heteroduplexes (Jensen et al. 2005; Kirkpatrick and Petes 1997; Kearney et al. 2001).

The efficiency of SSA depends on the length of the repeats that flank the intervening sequence (Sugawara et al. 2000). Although SSA can occur with repeats as short as 29 bp, the efficiency of SSA-associated repair is less than 1%. In contrast, when the length of the repeats is 415 bp, the efficiency of repair is close to 100%.
In addition to its role in SSA, the Rad1-Rad10 complex is important in resolving other types of recombination intermediates that generate branches. For example, the rate of ectopic recombination is substantially reduced in rad1 mutants (Mazon et al. 2012). Gene conversion events in which heterology must be removed to expose homology are also dependent on Rad1-Rad10 and Msh2-Msh3 (Colaiacovo et al. 1999). The degree to which these events are dependent on Rad1-Rad10 and Msh2-Msh3 depends on the length of the “tails” and whether there are one or two non-homologous ends. Although Rad1-Rad10 was found to be required for removing the 3' non-homologous tails of 30 nt or longer (Paques and Haber 1997), removal of only one non-homologous DNA end during conversion occurred in a RAD1- and MSH2-independent manner (Colaiacovo et al. 1999).

1.1.2 Non-homologous end-joining (NHEJ) for DSB repair

Non-homologous end-joining (NHEJ) is a mechanism for DSB repair that does not require sequence homology. It is also a conserved process among prokaryotes and eukaryotes. The frequency and efficiency of NHEJ-dependent repair differs depending on the types of lesions, the phase in the cell cycle, ploidy, and other factors. For instance, NHEJ is more common and efficient in religating complimentary ends of DSBs produced by endonucleases (Milne et al. 1996), compared to its efficiency with blunt or non-compatible ends (Moore and Haber 1996). Moreover, chemically-modified ends
cannot be re-joined by NHEJ unless the modifications have been removed by the action of the MRX-Sae2 complex (Symington and Gautier 2011). Although NHEJ can take place at any time during the cell cycle, it is most efficient during G1, since resection of the ends is inhibited in G1 (Simoneau et al. 2014; Huertas et al. 2008; Huertas and Jackson 2009). In addition, NHEJ is repressed in diploid yeast cells that are heterozygous at the MAT locus because Nej1, a protein required for NHEJ is downregulated by MAT heterozygosity (Frank-Vaillant and Marcand 2001; Kegel et al. 2001; Valencia et al. 2001; Wilson 2002).

The core components of NHEJ in *S. cerevisiae* and mammalian cells have been identified. Ku70-Ku80 heterodimer binding of DSB ends is an early step in the NHEJ pathway. This complex is conserved in all eukaryotes (Doherty et al. 2001). When yeast cells are transformed with a linearized plasmid with complementary ends, NHEJ usually rejoins those ends precisely (Boulton and Jackson 1996). However, strains with mutations in either *KU70* or *KU80* have a reduced efficiency of repair and many of the repair products reflect imprecise joining (Milne et al. 1996). Mutations in the C-terminus of Ku80 impair NHEJ significantly in yeast (Palmbos et al. 2005). This domain interacts with DNA ligase IV (Dnl4) (Daley et al. 2005). Moreover, the forkhead-associated (FHA) domain of Xrs2 interacts with Dnl4/Lif1, helping to recruit this ligase to the DSB region (Palmbos et al. 2008).
In mammalian cells, DNA-dependent protein kinase catalytic subunit (DNA-PKcs) was reported to tether the broken ends during NHEJ (DeFazio et al. 2002). A counterpart of this protein has not been found in yeast, but the MRX complex was reported to perform an equivalent function (Chen et al. 2001). Moreover, the MRX complex can promote DNA end joining by interacting directly with Dnl4 and its associated protein Lif1. MRX was shown to physically bind Dnl4/Lif1 and stimulate the ligation activity of Dnl4/Lif1 (Chen et al. 2001). It is somewhat unclear why the yeast cell needs three different proteins to recruit Dnl4 to the correct position.

*DNL4* encodes DNA ligase IV, a ligase specific for NHEJ (Wilson et al. 1997). It was first shown in mammalian cells that DNA ligase IV binds another NHEJ protein XRCC4 (Sibanda et al. 2001). This observation led to the identification of Lif1, the yeast counterpart of XRCC4, which also binds Dnl4 (Chen et al. 2001; Herrmann et al. 1998). Intriguingly, Lif1 forms a coiled coil domain that resembles the coiled coil of Rad50 (Junop et al. 2000; Sibanda et al. 2001). Dnl4 and Lif1 hence align DSB ends by base pairing the overhangs and ligating the ends. The Lif1 coiled coil domain can provide flexibility in DNA binding during this process, similar to the function performed by Rad50 during HR. Dnl4, when bound to DNA, adopts a configuration that is optimal for recruiting Pol4, a DNA polymerase specialized for DNA synthesis associated with NHEJ (Tseng and Tomkinson 2002). A third NHEJ factor Nej1 was also found to be useful for
efficient NHEJ, but not essential (Wilson 2002). Physically, Nej1 binds strongly to the globular head domain of Lif1, which is linked to its coiled coil domain (Frank-Vaillant and Marcand 2001).

1.2 Systems to study inter-sister chromatid recombination

Despite the significance of HR in the repair of spontaneous and induced DNA damage, there are many details concerning mitotic recombination that are still unclear. One issue is the timing of the DNA lesions that induce spontaneous mitotic recombination events. Using genetic systems that monitor inter-homolog recombination, we (Lee et al. 2009; St Charles et al. 2012) and others (Esposito 1978) concluded that most recombination events are initiated in G1 of the cell cycle. Specifically, St. Charles et al. used single-nucleotide polymorphism (SNP) microarrays to map about 140 spontaneous mitotic crossovers (St Charles et al. 2012). They observed that about two-thirds of the observed conversion events involved the repair of two sister chromatids broken at approximately the same position and they inferred that these events were consequences of G1-induced DSBs on one of the homologs, followed by replication of the broken chromosome and subsequent repair in G2.

This conclusion was unexpected since Rad52p foci (indicative of DNA damage) are much more common in the S-period and G2 than in G1 (Lisby et al. 2001). Also, the RPA proteins, which bind ssDNA intermediates during HR, form larger foci in G2 than
in G1 phase (Barlow et al. 2008). In addition, repair of DSBs by HR is inefficient in G1 (Aylon et al. 2004). A simple model consistent with all of these observations is that most DSBs occur during the S-period but these DSBs are repaired by sister-chromatid recombination rather than inter-homolog recombination (Kadyk and Hartwell 1992; Lee et al. 2009). DSBs that occur in G1 are likely not repaired in G1, but the broken chromosome is replicated to produce two sister chromatids that are broken at the same position. Since the DSBs are at the same position in the two sisters, the breaks cannot be repaired by sister-chromatid recombination and are repaired using the intact homolog as a template (Lee et al. 2009).

Despite the abundance of indirect evidence that S/G2 phase DSBs prefer to recombine with sister chromatids for repair, direct evidence is needed to unequivocally verify this hypothesis. Since HR using the sister chromatid is almost always genetically silent, it is a challenge to detect inter-sister recombination.

In 1992, Kadyk and Hartwell developed a system in which both inter-sister and inter-homolog recombination can be monitored (Kadyk and Hartwell 1992). This system monitors the inter-sister chromatid recombination rate at the ADE3 locus, which they integrated into the right arm of chromosome III. Inter-homolog recombination was examined by determining the rate of heteroallelic recombination at the LEU2 locus. 5' and 3' deletions of the ADE3 gene that fail to complement were integrated into a strain
Figure 1.3. Sister-chromatid-exchange (SCE) substrate.

This figure shows the substrate designed by Kadyk and Hartwell to monitor sister-chromatid exchange (Kadyk and Hartwell 1992).

A. The original $ADE3$ fragment is 5.5 kb long and contains the 2.8 kb $ADE3$ ORF, which is shown as a blue box. The 5’ – truncated $ADE3$ (4.5 kb) removes 507 bp and the 3’ – truncated $ADE3$ (1.3 kb) removes 2026 bp. They share 305 bp sequence homology, which is shown as black boxes. The arrowhead indicates the 3’ end and the tail indicates the 5’ end of the $ADE3$-containing fragment.

B. Structure and integration of the SCE substrate. The SCE substrate consists of a direct repeat of the 5’ – truncated $ADE3$ and 3’ – truncated $ADE3$, separated by the selectable marker $URA3$, which is shown as a purple box. This recombination substrate was referred to as $SCR::URA3$. The $SCR::URA3$ fragment was cloned into a 2.5 kb fragment on the left arm of chromosome III, 1.7 kb away from the $LEU2$ locus, which is shown as a green box.
from which the native ADE3 gene has been deleted. There was a 305 bp region of shared sequence homology between the two truncated ADE3 fragments (Figure 1.3).

In this strain, only recombination events that occur within one chromatid or unequal events between sister chromatids will generate a wild-type ADE3 gene (Figure 1.4). The frequency of generating an ADE3 gene was measured indirectly by a colony color assay. The diploid used in these experiments was homozygous for an ade2 mutation. A strain with both an ade3 and ade2 mutation forms white colonies, but a strain that is ADE3 ade2 forms red colonies. Thus, the frequency of red colonies reflects the frequency of sister-chromatid recombination involving the truncated ade3 gene fragments. As illustrated in Figure 1.4, a wild-type ADE3 gene can be generated by an intra-chromatid “pop-out” (Figure 1.4A), an unequal crossover between sister chromatids (Figure 1.4B), or a gene conversion between sister chromatids (Figure 1.4C). The intra-chromatid pop-out mechanism would be expected to result in a circular plasmid that contains the ADE3 and URA3 genes. Since this plasmid lacks a replication origin, it would not be stably maintained. Recombination between homologs was monitored by measuring the frequency of Leu+ derivatives generated from the diploid that containing leu1-1 and leu1-12 heteroalleles.

In their experiments, Kadyk and Hartwell synchronized the diploids in G1 or G2 phase, and treated the synchronized cells with X-rays. Interhomolog recombination was
induced much more efficiently (about 15-fold more) in G₁-irradiated cells than in G₂-irradiated cells. In contrast, sister-chromatid exchange was induced more efficiently in G₂-irradiated cells than in G₁-irradiated cells. The simplest interpretation of this experiment is that DSBs formed in G₂ are preferentially repaired using the sister chromatid as a template rather than the homolog. The degree of preference is about 15-fold. Although this study is convincing, there are two potential caveats. First, sister-chromatid recombination was measured at a different locus than interhomolog exchange. Second, in these experiments, DNA damage was induced by X-rays. The properties of spontaneous events might be different. In our studies described in Chapter 3, we examined spontaneous sister-chromatid recombination events and interhomolog events at the same locus.

Several other systems have also been developed to study inter-sister chromatid recombination. Arbel et al. used haploid yeast strains, in which only inter-sister recombination exists, to study inter-sister chromatid recombination (Arbel et al. 1999). They employed the power of a phenomenon called return-to-growth (RTG) to induce meiotic level of DSBs. In sir3 yeast strains, haploid cells incubated in sporulation medium can initiate meiotic DSB formation (Gilbertson and Stahl 1994). If these cells are returned to growth medium, these DSBs are repaired. The repair of DSBs in haploid strains by sister chromatid recombination can be monitored physically using CHEF gels.
Figure 1.4. Three types of recombination events generating \textit{ADE3}

This figure shows three recombination events in the \textit{SCR::URA3} substrate that generate the \textit{ADE3} gene. The arrowhead indicates the 3’ end (shown as 5’∆ in this figure) and the tail indicates the 5’ end (shown as 3’∆ in this figure) of the \textit{ADE3}-containing fragment. Black boxes indicate the shared homology between the 5’ end of the \textit{ADE3}-containing fragment and the 3’ end of the \textit{ADE3}-containing fragment.
Purple boxes indicate the URA3 gene. Small blue circles represent centromeres. Thin black lines indicate locations of crossover or gene conversion. This figure is modified from Kadyk and Hartwell (1992).

A. An intramolecular event (pop-out) can occur before or after DNA replication. This recombination mechanism generates an extrachromosomal circle containing ADE3 and URA3. However, this circle does not contain a replication origin, and is expected to be lost at high frequency.

B. After replication, an unequal recombination event can occur between sister chromatids, generating a chromatid containing ADE and URA3 and another chromatid lacking ADE3 and URA3.

C. After replication, a gene conversion event between sister chromatids gives rise to an ADE3 URA3 chromatid and an ade3 URA3 chromatid.

The mechanisms described in panels B and C can be distinguished by examining both sectors of red/white sectored colonies as described in Kadyk and Hartwell (1992).

(Zenvirth et al. 1997). Using this system, Arbel et al. showed that RAD54, which is also involved in strand invasion, was important for inter-sister chromatid recombination (Arbel et al. 1999). The RAD54 homolog TID1, and meiotic recombinase DMC1, however, were not necessary for inter-sister chromatid recombination (Arbel et al. 1999). Although this system is useful in identifying proteins required for sister-chromatid recombination, it does not address the issue of the preference for the sister chromatid over the homolog for mitotic exchange.
Gonzalez-Barrera et al. and Cortes-Ledesma et al. designed a physical assay that involved measurements of the sizes of repaired plasmids to distinguish BIR, equal sister chromatid recombination and unequal sister chromatid recombination (Gonzalez-Barrera et al. 2003; Cortes-Ledesma and Aguilera 2006; Cortes-Ledesma et al. 2007). They concluded that the repair of an HO-induced DSB in this plasmid occurred most frequently by equal sister chromatid recombination (Gonzalez-Barrera et al. 2003). In addition, they showed that sister chromatid cohesion was important for sister-chromatid recombination. Temperature-sensitive mutants of cohesion proteins Smc3, Scc1 and Scc2 significantly decreased the frequency of sister-chromatid recombination, and was associated with an elevation in intra-chromatid recombination (Cortes-Ledesma and Aguilera 2006). In addition, Cortes-Ledesma et al. found that repair of a DSB using the sister chromatid as a template was highly dependent on Rad51 and Rad54, both of which play important roles in strand invasion (Petukhova et al. 2000), and on Rad52 and Rad59, both of which are important in strand annealing (Cortes-Ledesma et al. 2007).
Figure 1.5. Recombination substrate for detecting inter-sister chromatid and inter-homolog recombination. Modified from Mozlin et al. (Mozlin et al. 2008).

A. Recombination substrate for detecting inter-sister chromatid recombination. A direct repeat of ade2 heteroalleles is separated by a TRP1 gene. This substrate is integrated at the ADE2 locus in the genome. Either unequal sister chromatid recombination or inter-sister/intra-chromatid gene conversion can give rise to ADE2 products.

B. Recombination substrate for detecting inter-homolog recombination in a diploid strain. Two ade2 heteroalleles are located at the ADE2 loci on two homologs. Inter-homolog recombination (gene conversion or crossover) can generate ADE2 products. Inter-sister chromatid recombination cannot be detected with this system.
Mozlin et al. designed a recombination substrate in which a direct repeat of ade2 heteroalleles was separated by a TRP1 gene (Figure 1.5A) (Mozlin et al. 2008). This substrate was integrated at the ADE2 locus in the genome. Either unequal sister chromatid gene conversion or intra-chromatid recombination between two ade2 heteroalleles can regenerate the ADE2 gene, which was selectable. This system is similar to that used by Kadyk and Hartwell (1992). In addition, a diploid strain containing the two ade2 heteroalleles at the ADE2 loci on two homologs was constructed to monitor inter-homolog recombination by either GC or CO (Figure 1.5B). Inter-homolog and inter-sister chromatid recombination was decreased by about 1000-fold by the rad51 mutation. In this system, partner choice during DSB repair cannot be examined since inter-sister recombination is measured in a different strain than inter-homolog recombination.

In mammalian cells, Johnson et al. designed a recombination reporter substrate (Figure 1.6A), SCneo, that also is similar to the substrate designed by Kadyk and Harwell (Johnson et al. 1999). Two nonfunctional copies of the neomycin phosphotransferase (neo) gene were oriented as direct repeats, separated by a hygromycin resistance gene (hygR) (Johnson et al. 1999). This recombination substrate was subsequently integrated into the mammalian genome. One copy of the neo gene was missing the 5’ end of the gene (designated as 3’ neo in Figure 1.6). The other copy was
mutated at an *Nco*I site by a 4 bp deletion, and had an insertion of the I-SceI endonuclease cutting site. This second mutated copy is designated as *S2neo* in Figure 1.6. Either intra-sister chromatid gene conversion (GC) or unequal sister chromatid exchange (SCE) could regenerate a functional *neo* gene, while maintaining the *hyg*<sup>R</sup> marker. The *neo<sup>r</sup>* gene confers resistance to G418 in mammalian cells. Utilizing this SCneo assay, researchers were able to determine that the spontaneous recombination rate within the SCneo substrate is between 10<sup>-6</sup>/generation/cell and 10<sup>-5</sup>/generation/cell. Expression of the I-SceI endonuclease elevated the recombination rate to > 10<sup>-3</sup>/generation/cell. They also discovered that XRCC2, one of five mammalian *RAD51* paralogs, is essential for HR in mammalian cells when DSBs are induced by I-SceI endonuclease (Johnson *et al.* 1999).

Moreover, Johnson *et al.* delineated the relative frequency of each recombination mechanism. Using Southern analysis, they examined a total of 300 clones that had undergone I-Sce-induced recombination events. Their analysis indicated that 42 out of 300 clones had lost the I-SceI cleavage site. Of these, 24 (57%) were generated by NHEJ and 18 (43%) were generated by HR (Table 1.1) (Johnson and Jasin 2000).
This figure shows the structure of the recombination reporter substrate SCneo and recombination products. Shaded boxes represent the 0.7 kb neo gene and the clear boxes containing an arrowhead represent the promoter of the S2neo gene. This figure is modified from Johnson and Jasin (Johnson and Jasin 2000).

A. 5’ – truncated neo (shown as 3’ neo) and a copy of mutated nonfunctional neo (shown as S2neo) are separated by hyg. There is an 18 bp I-SceI site inserted in the mutated neo gene, shown as a cleft within S2neo.

B. Predicted neo+ and neo− products of repairing DSBs induced by I-SceI expression. Short tract gene conversion (STGC), long tract gene conversion (LTGC) and unequal sister chromatid exchange (SCE) can generate neo+ products. NHEJ, homologous deletion (HD) and SCE can generate neo− products. STGC and NHEJ products are 4.0 kb long, and the LTGC products are 7.3 kb long. The 0.7 kb products (lower right in Panel B) result from either HD or SCE. HD events are likely a consequence of SSA. One unequal SCE event gives rise to both a neo+ (lower left in Panel B) and a neo− (lower right in Panel B) product.
Table 1.1. Random clone analysis of SCneo recombination events.

<table>
<thead>
<tr>
<th></th>
<th>V79 cell line</th>
<th>AA8 cell line</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clones analyzed</td>
<td>200</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>Clones losing I-SceI site</td>
<td>27</td>
<td>15</td>
<td>42</td>
</tr>
<tr>
<td>Repair products</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STGC</td>
<td>5</td>
<td>3</td>
<td>8 (19%)</td>
</tr>
<tr>
<td>LTGC/SCE</td>
<td>4</td>
<td>2</td>
<td>6 (14%)</td>
</tr>
<tr>
<td>HD/SCE</td>
<td>1</td>
<td>3</td>
<td>4 (10%)</td>
</tr>
<tr>
<td>NHEJ</td>
<td>17</td>
<td>7</td>
<td>24 (57%)</td>
</tr>
</tbody>
</table>

In another set of experiments, 119 clones were found to be G418\(^{R}\) out of 2267 colonies in which I-SceI expression was induced. Out of 119 G418\(^{R}\) clones, 90 were generated by GC events that gave rise to shortened SCneo substrates, referred to as short tract gene conversion (STGC). The other 29 contained expanded SCneo substrates, which could be generated by either unequal SCE or GC; the latter are long tract gene conversion (LTGC). In these 29 G418\(^{R}\) clones with expanded SCneo products, LTGC products could be further distinguished from unequal SCE products by Southern analysis. A detailed breakdown of the data is summarized in Table 1.2 (Johnson and Jasin 2000).
Table 1.2. SCneo homologous recombination events.

<table>
<thead>
<tr>
<th></th>
<th>V79 cell line</th>
<th>AA8 cell line</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clones analyzed</td>
<td>1457</td>
<td>810</td>
<td>2267</td>
</tr>
<tr>
<td>Clones with G418(^r) cells</td>
<td>75</td>
<td>44</td>
<td>119</td>
</tr>
<tr>
<td>neo(^r) repair products</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STGC</td>
<td>55 (74%)</td>
<td>35 (80%)</td>
<td>90 (76%)</td>
</tr>
<tr>
<td>LTGC</td>
<td>19 (25%)</td>
<td>7 (16%)</td>
<td>26 (22%)</td>
</tr>
<tr>
<td>SCE</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>other</td>
<td>1 (1%)</td>
<td>2 (4%)</td>
<td>3 (2%)</td>
</tr>
</tbody>
</table>

The aforementioned systems provided important insights for investigating the template and mechanism used in DSB repair in yeast or mammalian systems. However, none of these experiments allowed the monitoring of spontaneous inter-sister chromatid and inter-homolog recombination at the same locus. As part of this thesis (Chapter 3), a genetic system has been developed that overcomes the limitations of the published assays just described.

1.3 CUP1 tandem repeats and copper homeostasis

1.3.1 The unique structure of the CUP1 tandem array

In the S. cerevisiae genome, there are two arrays that generally have large numbers of repeats, the ribosomal DNA and the CUP1 array. The ribosomal DNA includes about 100 to 150 copies of a 9 kb repeat on chromosome XII. The CUP1 array is
more variable in size, containing between 1 and 20 tandem repeats (varying between about 1 and 2 kb per repeat) located on chromosome VIII. The degree of resistance to copper in the medium correlates with the number of CUP1 repeats (Fogel et al. 1983). Most yeast strains commonly used in laboratories have more than five copies of CUP1 in their arrays.

In the S. cerevisiae strain S288c, which was the first strain sequenced (Saccharomyces Genome Database), the CUP1 locus is depicted as having two repeats (shown in red brackets in Figure 1.7). The repeat has a complex structure that begins about 150 bp upstream of the RUF5-1 gene, which encodes a transcript of unknown function (McCutcheon and Eddy 2003). The CUP1 gene is embedded within the RUF5 gene, but is transcribed in the opposite orientation (http://www.yeastgenome.org/cgi-bin/seqTools). There is a second intergenic region of about 70 bp separating the 3' end of the RUF5-1 gene from the 3' end of YHR054C. YHR054C is identical to the terminal 1 kb of RSC30. There is also an ARS element within each repeat that overlaps with the 3' end of RUF5, and extends to the 3' end of RSC30.

The repeat length in S288c is 1998 bp, of which the CUP1 gene is a small part, encoding a protein of only 61 amino acids (Karin et al. 1984). Although only two repeats are shown in the genomic sequence of S288c in SGD, the closely related strain X2180-1A has about 15 repeats. Below, we find that S288c also has about 15 repeats. A copper-
sensitive strain, BZ31-1-7Ba, had a restriction enzyme digest pattern consistent with the presence of a single \textit{CUP1} gene (Fogel and Welch 1982). Fogel \textit{et al.}, also investigated the tandem repeat patterns of the \textit{CUP1} locus in both laboratory and industrial yeast strains (Welch \textit{et al.} 1983). Researchers found that the structures and sizes of \textit{CUP1} tandem repeats are strain dependent. The repeat unit of about 2.0 kb is common in laboratory strains, possibly because many of these strains were derived from S288c. In industrial yeast strains, \textit{CUP1} repeats of about 1.5 and 1.7 kb were also observed (Welch \textit{et al.} 1983). In addition, Welch \textit{et al.} identified a copper-sensitive strain that was thought to contain a single \textit{CUP1} gene. However, neither this putative single-copy \textit{CUP1} repeat nor any of the variant repeats were sequenced in these early studies.

Figure 1.7. Structure of the \textit{CUP1} locus in S288c.

As described in the text, the \textit{CUP1} gene (blue) is located within a non-coding RNA gene (\textit{RUF5}, pink) between the genes \textit{CIC1} (green) and \textit{RSC30} (orange) on chromosome VIII. We show only two copies of each repeat (bracketed in red), although most strains have more than two copies. Thin horizontal arrows show the direction of transcription. The gray lines are 1 kb apart. \textit{ARS} elements are indicated by purple lines. \textit{YHR054C} is an open reading frame derived from the 3’ end of \textit{RSC30}. Each repeat is 2.0 kb long.
1.3.2 Copper homeostasis in the yeast *Saccharomyces cerevisiae*

Copper homeostasis is essential for cell survival, because copper has both essential biological roles at moderate concentrations and toxic effects at high concentrations. In fungi as well as mammalian cells, copper is an essential component of many biochemically important proteins involved in various vital cellular processes. For instance, the copper ion is required as a cofactor for the activities of galactose oxidase (Whittaker and Whittaker 1988; Wang *et al.* 1998), nitrite reductase (Tocheva *et al.* 2004), nitrous-oxide reductase (Rasmussen *et al.* 2000), and superoxide dismutase (Rae *et al.* 1999).

The specificity of the role of the copper ion in catalysis is derived from its redox property: copper ion can exist in either a reduced state (Cu⁺) or oxidized state (Cu²⁺). Cu²⁺ shows a high affinity for binding negatively-charged oxygen groups (such as the carboxylate oxygens in aspartate and glutamate) and imidazole nitrogen groups (such as the imidazole nitrogens in histidine). However, Cu⁺ prefers to bind divalent sulfur compounds, such as thiols and thioethers (Festa and Thiele 2011). Therefore, the redox metal copper binds a wide spectrum of ligands and provides a diverse chemical environment for many pivotal biological processes.

This redox property of copper, however, also turns out to have negative consequences. In the conversion processes between Cu⁺ and Cu²⁺, free hydroxyl radicals
(•OH) are readily formed. Hydroxyl radicals are toxic and highly reactive, causing damage to DNA and proteins (Halliwell and Gutteridge 1985). Moreover, copper can interfere with iron-sulfur (Fe-S) clusters, which are cofactors essential to many fundamental cellular processes, such as the regulation of gene expression, electron transport, and DNA repair (Johnson et al. 2005).

Copper regulation and homeostasis has been extensively investigated in *S. cerevisiae*. As shown in Figure 1.8, in *S. cerevisiae*, the sensors for copper levels are Mac1 and Ace1/Cup2. Thiele (Thiele 1988) and Welch *et al.* (Welch *et al.* 1989) independently identified the transcription factor activating the transcription of *CUP1*, the main copper metallothionein in yeast, as Ace1 and Cup2, respectively. I therefore refer to this transcription factor as Ace1/Cup2. Both Mac1 and Ace1/Cup2 contain cysteine-rich domains that bind copper with high affinity. Each cysteine-rich domain chelates 4 Cu$^+$ ions and forms a tetra-copper cluster (Brown *et al.* 2002; Dameron *et al.* 1991). Mac1 activates the expression of copper transporter genes until copper ions bind its cysteine-rich domains and inactivate it. Ace1/Cup2 remains inactive when the copper level is low and no copper ions bind it. When the copper level is elevated, the assembly of tetra-copper clusters upon copper binding in both Ace1/Cup2 and Mac1 induces conformational changes, activating Ace1/Cup2 and inactivating Mac1 (Graden and Winge 1997; Dobi *et al.* 1995). When activated, both Ace1/Cup2 and Mac1 bind copper
responsive elements (CuREs), TTTGC(T/G)C(A/G), in the promoter regions of genes that they regulate (Jamison McDaniels et al. 1999).

In response to low cellular copper levels, Mac1 activates the expression of high-affinity plasma membrane copper transporters CTR1 and CTR3 (Labbe et al. 1997), and potentially the low-affinity vacuolar membrane copper transporter CTR2 (Yamaguchi-Iwai et al. 1997). In addition, activated Mac1 induces the expression of the cell-surface ion reductases Fre1 and Fre2 (Georgatsou et al. 1997). 90% to 98% of metal reductase activity at the cell surface is performed by Fre1 and Fre2 (Anderson et al. 1992). Fre1 and Fre2 are important in copper homeostasis in that they convert Cu^{2+} to Cu^{+}, and Cu^{+} is the only form of copper ion that copper transporters Ctr1 and Ctr3 bind with high affinity (Hassett and Kosman 1995).
This figure shows the regulation of copper for maintenance of homeostasis. First, for the purpose of efficient transportation and utilization, extracellular Cu$^{2+}$ needs to be reduced to Cu$^{+}$, mostly by the activity of the cell-surface metal reductase Fre1. In response to low levels of cellular Cu$^{+}$ ions, transcription factor Mac1 activates the expression of high affinity copper transporters Ctr1 and Ctr3, as well as ion...
reductase Fre1. Enhanced Ctr1, Ctr3 and Fre1 transcription elevates copper transport through the membrane, increasing the cellular copper level. In response to elevated cellular copper levels, Cu^{+} ions bind the Mac1 molecule, form tetra-copper clusters and inactivate the TF activity of Mac1 by inducing conformational changes upon copper binding. Cu^{+} ions also bind Ace1/Cup2, inducing conformational changes that convert the inactive apo form of metal-free Ace1/Cup2 into the active copper-binding form. Ace1/Cup2 strongly activates the transcription of yeast metallothionein Cup1 and Crs5, as well as the superoxide dismutase Sod1. Cup1 and Crs5 detoxify excess copper by directly chelating Cu^{+} ions. Sod1 detoxifies radical oxygen species, buffering copper toxicity indirectly. This figure was derived from Festa and Thiele (Festa and Thiele 2011).

In response to elevated cellular copper levels, Ace1/Cup2 activates the copper metallothioneins Cup1 and Crs5, as well as the cytosolic copper-zinc superoxide dismutase Sod1. Metallothioneins refer to a group of cysteine-rich proteins that bind a wide range of metals via the thiol groups of cysteines. They exist in many eukaryotes and are reported to be localized to the Golgi membrane and secreted (Ercolani et al. 1990; Moltedo et al. 2000). In the yeast genome, the CUP1 locus consists of tandemly-repeated CUP1 genes, whereas the CRS5 gene exists in single copy. CUP1 encodes a small 61 AA peptide with an 8 AA lead sequence that is removed in the mature form of the peptide (Winge et al. 1985).

A comprehensive study comparing the Cu^{+} binding efficiency of Cup1 and Crs5 has established that Cup1 is the dominant copper-chelating molecule that confers most of the copper tolerance (Jensen et al. 1996). It was found that a single copy of CUP1 was
much more efficient in conferring copper resistance than the single copper CRS5 gene. The Cup1 molecule binds 8 Cu$^{+}$ ions via 12 cysteines (Winge et al. 1985), whereas one copper-binding stoichiometry experiment for Crs5 showed that each Crs5 molecule binds 11 Cu$^{+}$ ions (Jensen et al. 1996). Intriguingly, the CUP1 promoter region contains four copper responsive elements (CuREs), whereas the CRS5 promoter region contains only one CuRE (Culotta et al. 1994; Furst et al. 1988; Thiele and Hamer 1986). This difference in their respective promoters is relevant to their ability to affect copper resistance because when a single copy of CRS5 was under the control of the CUP1 promoter, its copper-buffering capability was significantly enhanced (Jensen et al. 1996).

Lastly, Ace1/Cup2 also activates the transcription of the superoxide dismutase Sod1, which detoxifies radical oxygen species and therefore buffers copper toxicity (Gralla et al. 1991).

In Chapter 2, I will discuss the structure of the CUP1 locus in more detail. In addition to the 2.0 kb repeats that have been well characterized, I will report four novel types of CUP1 repeats. These experiments were done in collaboration with the McCusker and Dietrich labs, and utilized information derived from their 100-genome sequencing project (Strope et al. 2015). The sequences and structures analyzed in Chapter 2 shed light on the mechanism of de novo duplications from a single copy of the CUP1 gene, which bears important evolutionary significance. In Chapter 3, I will introduce a
system based on the CUP1 array for monitoring inter-sister chromatid versus inter-homolog recombination. I will also discuss some results regarding the genetic regulation of these types of recombination, and the stimulatory effect of high levels of copper on them. In Chapter 4, I will discuss future experiments to investigate de novo duplication mechanisms and intra-/sister-chromatid recombination mechanisms.
2 Structures of naturally-evolved *CUP1* tandem arrays indicate that these arrays are generated from single-copy *CUP1* genes by unequal non-homologous recombination

An important issue in genome evolution is the mechanism by which tandem duplications are generated from single-copy genes. In the yeast *Saccharomyces cerevisiae*, most strains contain tandemly-duplicated copies of *CUP1*, a gene that encodes a copper-binding metallothionein. By screening 101 natural isolates of *S. cerevisiae*, we identified five different types of *CUP1*-containing repeats, as well as strains that had only one copy of *CUP1*. A comparison of the DNA sequences of these strains indicates that the *CUP1* tandem arrays were generated by unequal non-homologous recombination events in a strain that had one *CUP1* gene.

2.1 Introduction

Repetitive sequences exist in many eukaryotic genomes. About 8% of the human genome consists of repetitive DNA sequences, and a substantial fraction of these repeats are tandemly arranged. With the increased availability of whole-genome sequences, it has become increasingly clear that alterations in the number of repeated genes per genome is a common source of genetic diversity. Many of these alterations are larger than 50 base deletions/duplications, but are too small to be detected by analysis of chromosome karyotypes. Such intermediate-sized variations are called copy-number variation (CNV). The term is often used to describe a variant genome that contains
deletions or duplications of sequences (100 bp to 1 Mb) relative to a standard genome. The rate of CNV formation in human cells is about 2 x 10^{-6} to 10^{-4} per generation per locus (Zhang et al. 2009a). For the sake of comparison, the rate of point mutations in human cells is about 2 x 10^{-8}/bp per generation (Zhang et al. 2009a).

In a pre-existing tandem array, one mechanism that can alter the number of repeats is unequal crossovers (Figure 2.1A), a mechanism first demonstrated in studies of Bar eye in Drosophila almost 100 years ago (Sturtevant and Morgan 1923). As shown in Figure 2.1A, homologous recombination between misaligned tandem arrays can result in both deletions and duplications of repeats. The unequal crossover could occur between sister chromatids (as shown) or between homologs. Deletions can also be formed non-reciprocally by several other mechanisms including “pop-outs” (in which an intrachromatid crossover produces a circular DNA molecule containing one or more repeats and a shorter tandem array), or by single-strand annealing (in which processing of a break within the tandem array followed by annealing of the broken ends deletes one or more repeats). In yeast, unequal crossover events within the tandemly repeated ribosomal RNA genes are very frequent, occurring at a mitotic frequency of >10^{-2} per mitotic division (Szostak and Wu 1980) and >10^{-1} per meiotic division (Petes 1980).
Figure 2.1. Mechanisms for altering or generating tandem gene arrays.

Events are depicted as occurring between sister chromatids, and centromeres are shown as ovals or circles.
A. Unequal crossovers between misaligned sister chromatids can generate arrays with more or fewer repeats per array. Repeats are indicated by small rectangles.

B. Unequal crossovers between flanking repeats (triangles) can produce a duplication of single-copy sequences (dashed line).

C. In this model, a broken end invades a sister chromatid utilizing microhomology. Break-induced replication of the invaded chromosome results in a duplication of the sequence (shown as a triangle) initially present in one copy.

Homologous recombination between non-contiguous direct repeats can also result in duplication or deletion of single-copy sequences located between the repeats (Figure 2.1B). Such events have been detected in mammalian/human genomes (reviewed by Zhang et al., 2009) and in yeast (Koszul et al. 2004; Gresham et al. 2008; Zhang et al. 2013; Finn and Li 2013). Although most of these studies were done in wild-type strains, Finn and Li showed that re-replication resulting from premature initiation of a second round of DNA synthesis substantially elevated the frequency of this class of non-allelic homologous recombination.

In addition to duplication events that reflect homologous recombination, two other mechanisms that lead to duplications of single-copy genes have been observed. The first mechanism involves a specialized type of break-induced replication (BIR). As shown in Figure 2.1C, a broken chromosome end invades a non-allelic region of a sister
chromatid, copying sequences from the chromatid. This initial invasion involves little or no sequence homology. In yeast, which has small chromosomes (< 2 Mb), synthesis could continue to the end of the chromosome. In yeast, this type of event (termed “microhomology-mediated break-induced replication” or MMBIR) was demonstrated to be the causal mechanism for some duplications of the ribosomal protein gene RPL20B (Payen et al. 2008). A related mechanism that sometimes involves multiple invasions and template switches was proposed to account for some classes of CNVs in mammalian cells (Lee et al. 2007). In the second mechanism, processing of an inverted repeat (hairpin) structure can generate a multi-copy palindromic plasmid containing a gene that was originally present in a single copy (Narayanan et al. 2006).

Since different selective procedures were used to study gene copy number variations in yeast experiments, it is difficult to compare the rates of duplications by the pathways described above. In experiments in which duplications of a reporter gene were located between two retrotransposons located about 50 kb apart, duplications were observed at an approximate frequency of $10^{-6}$-$10^{-7}$/generation/cell (Zhang et al. 2013); all of the detected duplications in a haploid strain resulted from unequal crossovers between the flanking Ty1 elements (Zhang et al. 2013). This result argues that, at least in this chromosomal context, non-allelic homologous recombination events generate duplications more frequently than MMBIR or related mechanisms. In contrast,
duplications of the *RPL20B* gene were generated by non-allelic homologous recombination and MMBIR with approximately equal frequencies (about $5 \times 10^8$; Payen *et al.*, 2008).

In the studies described below, we characterize the *CUP1* locus in a collection of yeast strains isolated from the wild or from clinical specimens. Whole genome sequencing and assembly has recently been performed for 100 diverse strains of *Saccharomyces cerevisiae* (Strope *et al.* 2015); this collection has been termed the “100-genomes strains.” From these sequences and additional sequencing efforts, we define the structures of five different classes of *CUP1* repeats as well as the sequence of a strain (DTY3) with only a single copy of *CUP1*. The comparison of these sequences suggests a simple mechanism by which a single *CUP1* gene is duplicated to form a tandem array by unequal non-homologous recombination.

We also compared the sizes of the *CUP1* arrays in 13 different strains that had at least two copies of *CUP1*, and found a size range of between two and eighteen repeats per array. The size of the array roughly correlated with the ability of the strain to grow in high levels of copper. Finally, we showed that a strain with 14 *CUP1* repeats could revert to having only one *CUP1* gene at a rate of about $5 \times 10^7$/division.
2.2 Material and methods

2.2.1 Yeast strains

The genotypes of all strains used in our study are shown in Table S 2.1 and primers used in strain construction, PCR analysis, or DNA sequencing are in Table S 2.2. The genotypes of the strains S288c (Engel et al. 2014), W303-1A (Thomas and Rothstein 1989a), and YJM789 (Wei et al. 2007) have been previously described. The strains YJM189, YJM271, YJM456, YJM693, YJM969, YJM972, YJM978, YJM996, YJM1307, and YJM1549 were described by Strope et al. (Strope et al. 2015). The strain YJM799 is isogenic with YJM789 except for changes introduced by transformation, and was obtained from J. McCusker (Duke University). JSC19-1 is also isogenic with YJM789 except for changes introduced by transformation (St Charles and Petes 2013).

The strain JSC10-1 (St Charles and Petes 2013) is isogenic with W303-1A except for changes introduced by transformation and has the genotype: MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::NAT RAD5. We generated a PCR fragment containing a wild-type URA3 gene by amplifying genomic DNA of the strain S288c with the primers VIII212898::URA3 F and VIII212898::URA3 R. This fragment was transformed into JSC10-1 to generate a derivative (YZ22) that had the URA3 gene integrated within the CUP1 cluster at position 212898 in the RUF5 gene of one of the repeats; this allele is called VIII212898::URA3.
2.2.2 Measurement of the sizes and sequences of CUP1 repeats

Our first estimate of the sizes of the CUP1 repeats in different strains was initially based on a PCR analysis. For this analysis, we used primers F1 and R1' located within CUP1 but oriented in different directions. In order to generate a PCR product, the array needs to contain at least two tandem CUP1 genes. The resulting PCR fragments were analyzed by gel electrophoresis. Some of the strains did not generate a PCR fragment with this procedure. We confirmed that these strains had only a single copy of CUP1 by PCR analysis using the primers VIII211849 and VIII216603. With these primers, strains that have a single copy of CUP1 produce a PCR fragment of 2.7 kb.

Sequence analysis of various classes of repeats and their flanking sequences were determined by sequencing PCR fragments with the primers described in supplementary tables in Section 2.5. These sequences are displayed in Tables S 2.3-S 2.10.

2.2.3 Analysis of the number of CUP1 copies in tandem arrays

Genomic DNA isolated from different strains was isolated in plugs of low-melt agarose as described previously (McCulley and Petes 2010). The samples were treated overnight at 37°C with EcoRI. The resulting DNA fragments were separated by CHEF (contour-clamped homogeneous electric field) gel electrophoresis (McCulley and Petes 2010), followed by transfer of the separated fragments to Hybond nylon membranes.
Southern analysis was conducted using *[**CUP1**](#) specific probes. Sizes of the tandem arrays were estimated relative to DNA size standard (Bioline DNA Hyperladders I and VI).

### 2.2.4 Southern analysis

The size of the *[**CUP1**](#) tandem arrays was estimated by gel electrophoresis of *[EcoRI](#)* fragments derived from the various yeast strains. Following electrophoresis, the samples were transferred to Nylon membranes (Roche, Product # 11209272001). The 1 kb hybridization probe included sequences between coordinates 212534 and 213538, containing *[**CUP1**](#) and flanking sequences within the repeat. The probe was synthesized to contain digoxigenin (DIG)-dUTP. Using the primers *[CUP1](#) amp5-2* and *[CUP1](#) amp3* (Table S 2.2), we amplified genomic DNA of strain S288c using the PCR DIG probe synthesis kit of Roche (Product No. 11536090910). We used a concentration of DIG-dUTP of 21 µM in addition to the genomic DNA (10 ng), primers, and other constituents of the reaction. DIG-labeled probes for the DNA ladders (Bioline DNA Hyperladders I and VI (discontinued) were generated with DIG-High Prime DNA Labeling and Detection Starter Kit II (Roche, Product No. 11585614910) using the random-priming labeling technique.

Hybridization was performed according to the procedures described in the Roche DIG Application Manual for Filter Hybridization (http://lifescience.roche.com/wcsstore/RASCatalogAssetStore/Articles/05353149001_08
We used 45 µl of the DIG-labeled CUP1 probe mixture in 25 ml of hybridization buffer (Roche DIG Easy Hyb Granules, Product No. 11796895001). Hybridization was conducted at 420 C. overnight, followed by two high-stringency washes done for 15 minutes at 650C. Chemiluminescent detection of the probes on the membranes was done using Anti-Digoxigenin-AP, Fab fragments (Roche, Product No. 11093274910), and the DIG Wash and Block Buffer Set (Roche, Product No. 11585762001). The washing buffer, maleic acid buffer, and detection buffer were prepared as described in the DIG DNA Labeling and Detection Kit Version 19 protocol (https://cssportal.roche.com/LFR_PublicDocs/ras/11093657910_en_19.pdf).

2.2.5 Measurements of copper resistance

For each strain, about 1000 cells were inoculated into 5 ml of SD complete medium (Guthrie and Fink 1991) containing levels of copper sulfate that varied between 0 and 2.4 mM, changing in 0.1 mM increments. After two days of growth at 30 °C, we measured the OD_{660} of each culture. If the OD_{660} of the undiluted culture was less than 0.1, we scored the concentration as inhibitory to growth of that strain. In the uninhibited cultures, the OD_{660} was between 1.8 and 3, the equivalent of about 2-4 x 10^7 cells/ml.

2.2.6 Measurements of the rate of loss of a URA3 gene integrated within the CUP1 repeats in YZ22

As described above, the haploid strain YZ22 (isogenic with W303-1A) contains a CUP1 array of about 14 repeats with a URA3 gene integrated into the array. As
explained in the text, an unequal crossover or various intra-chromatid recombination events can result in loss of the insertion. Strains that lose the *URA3* insertion result in derivatives that are Ura− and selectable on medium containing 5-fluoro-orotate (Boeke *et al*. 1987). We measured the frequency of 5-FOA R derivatives and the total number of cells in 18 independent cultures of YZ22. These frequencies were converted to rates using the method of the median (Lea and Coulson 1949). We screened 55 independent 5-FOA R derivatives by PCR analysis with primers VII211849 F and VIII216603 to identify strains that contained only one *CUP1* gene. Two of the 55 strains had the 2.7 kb fragment expected for single *CUP1* genes.

### 2.2.7 DNA sequencing

For all strains containing repeated *CUP1* genes, we determined the sequence of the repeats, as well as the sequences that connected the repeats to single-copy sequences on the centromere-distal and centromere-proximal sides of the tandem array. The information about the flanking sequences was necessary to determine whether the tandem arrays were in the same chromosomal context on chromosome VIII in all strains. The results of this analysis are presented in Tables S 2.3-S 2.10. For our analysis, we sequenced the *CUP1* genes in two strains with 1.8 kb repeats (Type 2, Tables S 2.3 and 2.4), one strain with a 1.2 kb repeat (Type 3, Table S2.5), two strains with 1.9 kb repeats (Type 4, Tables S 2.6 and S 2.7), and two strains with 1.6 kb repeats (Type 5, Tables S 2.8
and 2.9). In addition, we sequenced the comparable genomic region of a strain lacking CUP1 repeats (Table S 2.10). Sequences of all primers are in Table S 2.2.

For the analysis of repeats, we first determined the sequences of the PCR fragments generated using the primers F1 and R1’. Since the fragment generated using these primers does not contain about 30 bp that separate the primer binding sites, we also determined the sequences of PCR fragments generated using the primers VIII212300 and VIII213031 that contains the sequences separating F1 and R1’. The specific primers used to sequence the repeats differed for different types of repeats, and are shown in boldface in Tables S 2.3-S 2.9.

The primers used to produce PCR fragments containing the centromere-proximal junctions (CIC1-CUP1 region) were produced with the following primer pairs: 1) Type 2 (VIII211528 F and R1’), 2) Type 3 (VIII211528 F and R1’), 3) Type 4 (VIII211528 F and R1’), and 4) Type 5 (VIII211528 F and VIII212063 R).

The primers used to produce PCR fragments containing the centromere-distal junctions (CUP1-RSC30 region) were produced with the following primer pairs: 1) Type 2 (VIII213200 F and VIII216603), 2) Type 3 (F1 and VIII213537 R), 3) Type 4 (VIII213601 F and VIII216603 R), and 4) Type 5 (F1 and VIII214195).

We also sequenced the CUP1 region in a copper-sensitive strain DTY3 that contains a single copy of CUP1. Three overlapping PCR fragments were sequenced. One
fragment (generated using primers VIII210632 F and VIII212063 R) contained the CIC1 coding sequence and a portion of the CIC1-CUP1 intergenic sequence. The second fragment (generated using primers VIII211849 F and VIII216603 R) extended from the 3' region of CIC1 to the 3' region of RSC30. The third fragment (generated with primers VIII216314 F and VIII218008) contained the 5' region of RSC30. The composite sequence and the primers used to produce the sequence are in Table S 2.10.

2.3 Results

2.3.1 Identifying five different CUP1 repeats

Until recently, the complete assembled sequence of the CUP1 array was only available for S288c (Johnston et al. 1994). In examining genomic sequences derived from YJM789 (Wei et al. 2007), we noticed that the partial sequence of the CUP1 repeat in contig 18 was different from that of S288c. We decided, therefore, to examine the sequence of the CUP1 repeats and flanking sequences in YJM789 and other yeast strains that had repeats different from S288c. In addition, we sequenced the CUP1 gene in DTY3, a strain that has only one CUP1 gene.

The initial characterization of the repeats was done by PCR analysis using primers located within the CUP1 gene (F1 and R1', Table S 2.2). The location and orientation of the primers are shown in Figure 2.2A. In order to produce a PCR product, there must be at least two tandem CUP1 copies. Since these primers are located about 40
bp apart, the PCR product is about 40 bp smaller than the size of the repeats. In Figure 2.3, we show an analysis of repeat size in 5 different yeast strains: S288c (lanes 2), YJM789 (lane 6) and three strains derived from the 100-genomes sequencing project (lanes 3-5). Although we show the results of only three of the strains from the 100-genomes sequencing project, all these strains were examined by PCR.

These results and those obtained by deep sequencing (Strope et al. 2015) indicate that there are at least five types of CUP1 repeats extant in S. cerevisiae. The approximate sizes of these repeats in kb are: Type 1 (2.0), Type 2 (1.8), Type 3 (1.2), Type 4 (1.9), and Type 5 (1.6). S288c and W303-1A have Type 1 repeats, whereas YJM789 has Type 3 repeats.

From the deep-sequencing analysis, of the 66 strains with at least two CUP1 genes, 57 have only one type of repeat per array, and 9 have more than one type. Of the “pure” arrays, the number of strains with each type are: Type 1 (8), Type 2 (14), Type 3 (18), Type 4 (4), and Type 5 (13).

About one-third of the 100-genome strains examined by PCR failed to produce a product with primers F1 and R1’. Genome DNA from these strains was re-examined using primers VIII211849 and VIII216603 (shown as 849 and 603 in Figure 2.2, indicated by navy short fat arrows) that are located in the genes flanking the CUP1 sequences (Figure 2.2A). In strains that have only one copy of CUP1, we expect to see a PCR fragment of about 2.7 kb. A fragment of this size was observed in the 100-genomes
Figure 2.2. Structures of the CUP1 loci in different strains.

This figure shows the structures of the CUP1 loci for five types of CUP1 repeats and a single-copy CUP1 locus. The CUP1 gene (blue) is located within a noncoding RNA gene (RUF5, pink) between the genes CIC1 (green), and RSC30 (orange) on chromosome VIII. We show only two copies of each repeat (bracketed in red), although most strains have more than two copies. Thin horizontal arrows show the direction of transcription, and short fat arrows indicate the location of primers used in the analysis. The vertical gray lines are 1 kb apart, and the blue arrows show the position of EcoRI sites relevant to Southern analysis. ARS elements are indicated by purple lines. The lines labeled J1-J5 represent the junction fragments between the
end of one repeat and the beginning of the next. Lines 1–1 to 5–1 represent the
leftmost end of the tandem arrays, and lines labeled 1–2 to 5–2 represent the rightmost
end of the array.

A. Type 1 repeats (2.0 kb) found in strains S288c and W303-1A. \textit{YHR054C} is an
open reading frame derived from the 3′ end of \textit{RSC30}.

B. Type 2 repeats (1.8 kb) found in strains YJM189 and YJM972.

C. Type 3 repeats (1.2 kb) found in strains YJM693, YJM789, and YJM1549.

D. Type 4 repeats (1.9 kb) found in strains YJM271, and YJM1307.

E. Type 5 repeats (1.6 kb) found in strains YJM456, YJM969, and YJM978.

F. Single-copy \textit{CUP1} locus found in the copper-sensitive strain DTY3.

strains that failed to generate a PCR product with primers F1 and R1′. We also observed
the 2.7 kb fragment in DNA derived from the copper-sensitive strain DTY3. Deep-
sequencing analysis demonstrated that 30 of the 100-genome strains had only one \textit{CUP1}
gene.
Figure 2.3. Analysis of PCR fragments representing various types of CUP1 repeats.

Using primers F1 and R1’, we PCR-amplified genomic DNA samples from five different yeast strains. Because of the locations of F1 and R1’ within the CUP1 gene, the resulting fragments are about 40 bp shorter than the repeat length determined by DNA sequencing. Lanes 1 and 7 contain sizing ladders. The strain names, and the types and sizes of the repeats in the other lanes are: lane 2 (S288c, type 1, 2.0 kb); lane 3 (YJM271, type 4, 1.9 kb); lane 4 (YJM189, type 2, 1.8 kb); lane 5 (YJM456, type 5, 1.6 kb); lane 6 (YJM789, type 3, 1.2 kb).
2.3.2 Sequence analysis of CUP1 repeats

The structure of the Type 1 repeats of S288c is shown in Figure 2.2. The repeat can be defined by the sequence of the junction (labeled “J1” in Figure 2.2A) that has centromere-proximal sequences derived from RSC30 (named YHR054C) fused to centromere-distal sequences derived from the region located 5’ to RUF5. As shown in Figure 2.4A, the junction sequence can be matched nearly perfectly to sequences derived from within the RSC30 gene (line above the junction) on the left side and to sequences from the CIC1-RUF5 intergenic region (line below the junction) on the right side. The J1 junction has only two sequence differences that are not predicted from a simple fusion of RSC30 and the intergenic region of CIC1-RUF5: an A at the fusion breakpoint and seven T residues inserted into the poly T tract in the region upstream of RUF5. The length of the Type 1 repeat is 1998 bp (Johnston et al. 1994). It should be mentioned that it can be confusing to use the SGD coordinates to describe the repeats, since the sequences within the repeats match to two sets of coordinates, and the coordinates of the flanking sequences are displaced by about two kb (the length of the Type 1 repeat).

We sequenced the CUP1 repeats, as well as about 300 bp of sequences flanking the repeats, in eight strains including one Type 2 (YJM189), two Type 3 (YJM789, YJM969), two Type 4 (YJM271, YJM307), two Type 5 (YJM456, YJM996) strains, and one strain (DTY3) with a single copy of CUP1. The strain S288c, which has a Type 1 repeat,
was used for sequence comparisons by BLAST analysis (http://www.yeastgenome.org/cgi-bin/seqTools). For those strains that were also examined by deep sequencing, the sequences were identical (Strope et al. 2015). The details of the sequence analysis are described in Section 2.2.7, and the resulting sequences are described in Section 2.5.

Based on the sequence analysis, the repeats of each type are depicted in Figure 2.2. In Figure 2.4, we show the junction sequences (middle line) compared to sequences flanking the repeat near RSC30 (top line) and CIC1 (bottom line). As observed for the Type 1 repeat, the Type 2-5 junctions show little or no sequence homology at the breakpoints. As will be explained below, this observation argues strongly that the mechanism that generated the repeats likely involves non-homologous end-joining.

We also compared the sequences of all eight strains described above to the sequence of S288c. There were very few differences, and these differences are summarized in Tables S 2.3-S 2.10. No differences were observed for the CUP1 coding sequence except the termination codon for the repeats in strains YJM789, YJM456, YJM969, and DTY3 was UAA instead of UGA, the termination codon for CUP1 in S288c and the other strains.
Figure 2.4. Sequences at the breakpoints for five different types of CUP1 repeats.

Breakpoint sequences of five different types of CUP1 repeats are shown in comparison to the sequences flanking the repeats. The breakpoints of each repeat are labeled “Junctions 1–5,” and are the middle lines of each comparison. The top lines of each comparison (highlighted in blue) represent sequences from the centromere-distal flanking region, usually including a portion of the RSC30 gene, and the bottom lines are from the centromere-proximal flanking region (highlighted in pink) usually containing a portion of the CIC1-RUF5 intergenic region. For each comparison, the flanking sequences are from the same strain that contained the repeats. The locations of junctions J1-J5, as well as the locations of the flanking sequences, are shown in Figure 2.2. Sequence matches of the junction sequences to the centromere-distal and centromere-proximal flanking sequences are highlighted in blue and pink,
respectively. Bases that are shared homologies of the flanking sequences are shown in boxes. Bases that do not match either flanking sequence or that match both flanking sequences are not highlighted.

A. J1 junction sequences of Type 1 repeats (S288c). As indicated, the T-tract is seven bp longer in the repeat than in the flanking sequences. There is an A/T base pair at the junction that is not derived from either of the flanking sequences.

B. J2 junction sequences (YJM189 and YJM996).

C. J3 junction sequences (YJM789). There is one SNP distinguishing the repeat sequence and that of the centromere-distal flanking sequence (base that is not highlighted within the highlighted region).

D. J4 junction sequences (YJM271, YJM1307).

E. J5 junction sequences (YJM456). The strain YJM769 had the identical sequence as YJM456 except for the presence of one SNP (Table S 2.8 and Table S 2.9).

2.3.3 Variation in the number of CUP1 repeats in different yeast strains

Variation in the number of CUP1 repeats in tandem arrays of different yeast strains has been observed previously by Southern analysis (Welch et al. 1983). In addition, in the analysis of the 100-genome strains, estimates of the numbers of repeats per strain were made based on the number of times CUP1 sequences were present relative to single-copy sequences (Strope et al. 2015). We used Southern analysis to
Figure 2.5. Southern analysis of the CUP1 genes in 13 yeast strains.

Genomic DNA from each strain was treated with EcoRI. There are no recognition sites for EcoRI within the CUP1 repeats (Figure 2.2A). The fragments were separated by gel electrophoresis, transferred to a membrane, and hybridized to a probe containing the CUP1 sequences. The lanes labeled S1 and S2 on the left side of the figure are ethidium bromide-stained fragments representing size standards (Hyperladders VI and I from Bioline). S2 in the gel on the right side of the figure is the same ladder hybridized to a ladder-specific probe. The samples in lanes 1–13 are: 1 (YJM189), 2 (YJM271), 3 (YJM456), 4 (YJM693), 5 (YJM969), 6 (YJM972), 7 (YJM978), 8
(YJM996), 9 (YJM1549), 10 (YJM1307), 11 (S288c), 12 (W303-1A), and 13 (YJM789). The fragment sizes in lanes 1–13 are in Table 2.1.

determine the number of repeats in 10 strains that were part of the 100-genome strain analysis, as well as three other yeast strains (S288c, W303-1A, and YJM789). For this analysis, genomic DNA was treated with EcoRI which does not have a recognition sequence within the repeats (Figure 2.2A). Southern analysis of the strains is depicted in Figure 2.5.

Based on the size of observed fragment, the sizes of the repeats, and the location of the flanking EcoRI sites, we calculated the number of repeats in each strain (Table 2.1). This number was based on three independent measurements. In the strains examined, the number of repeats varied between two copies (YJM978) and 18 copies (YJM1549). There is a strong correlation ($R^2 = 0.76$) between the copy numbers as determined by Southern analysis and the copy numbers as estimated by coverage in the deep-sequencing experiments (Table 2.1 and Figure 2.6A).

**2.3.4 Copper resistance of strains with different types and different numbers of CUP1 repeats**

To determine whether there is a significant positive correlation between the number of CUP1 genes in the tandem array and the copper resistance of different strains, we measured the minimal concentration of copper concentration that prevented
growth (details in Section 2.2.5). The analysis of copper resistance was done in the same strains that were used for the Southern analysis, and the W303-1A, YJM789, and DTY3 strains. These data are summarized in Table 2.1. As shown in Figure 2.6B, there is a strong ($R^2=0.59$), but not perfect correlation between copper resistance and $CUP1$ copy number. Qualitatively similar results were observed previously (Fogel et al. 1983; Welch et al. 1983; Strope et al. 2015).

![Figure 2.6](image)

**Figure 2.6.** Correlations between different assays of $CUP1$ copy-number and between $CUP1$ copy-number and copper resistance.

**A.** Correlation between different assays of $CUP1$ copy-number. From the information in Table 1, we found a strong correlation ($R^2 = 0.76$) between copy-number as estimated by coverage in deep-sequencing experiments (Y-axis) vs. copy-number estimated by Southern analysis (X-axis).

**B.** Correlation between $CUP1$ copy-number (Southern analysis) and the minimal concentration of copper that inhibits growth for diploid strains.
Table 2.1. Analysis of *CUP1* gene tandem arrays in 14 yeast strains of *Saccharomyces cerevisiae*.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Type of repeat (size in kb)</th>
<th>Size of EcoRI fragment (kb)(^1)</th>
<th><em>CUP1</em> copy # (Southern)</th>
<th><em>CUP1</em> copy # (deep sequencing)(^2)</th>
<th>([\text{Cu}^{2+}]) inhibitory concentration (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S288c</td>
<td>1 (2.0)</td>
<td>30.8</td>
<td>14</td>
<td>ND</td>
<td>2</td>
</tr>
<tr>
<td>W303-1A</td>
<td>1 (2.0)</td>
<td>31.3</td>
<td>14</td>
<td>ND</td>
<td>1.8</td>
</tr>
<tr>
<td>YJM189</td>
<td>2 (1.8)</td>
<td>23.3</td>
<td>11</td>
<td>6</td>
<td>0.8</td>
</tr>
<tr>
<td>YJM972</td>
<td>2 (1.8)</td>
<td>18.7</td>
<td>8</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>YJM996</td>
<td>2 (1.8)</td>
<td>25.4</td>
<td>12</td>
<td>12</td>
<td>1.4</td>
</tr>
<tr>
<td>YJM789</td>
<td>3 (1.2)</td>
<td>11.6</td>
<td>7</td>
<td>ND</td>
<td>0.4</td>
</tr>
<tr>
<td>YJM693</td>
<td>3 (1.2)</td>
<td>19.3</td>
<td>13</td>
<td>15</td>
<td>1.8</td>
</tr>
<tr>
<td>YJM1549</td>
<td>3 (1.2)</td>
<td>24.4</td>
<td>18</td>
<td>18</td>
<td>1.0</td>
</tr>
<tr>
<td>YJM271</td>
<td>4 (1.9)</td>
<td>20.7</td>
<td>9</td>
<td>8</td>
<td>0.3</td>
</tr>
<tr>
<td>YJM1307</td>
<td>4 (1.9)</td>
<td>15.5</td>
<td>6</td>
<td>4</td>
<td>0.3</td>
</tr>
<tr>
<td>YJM456</td>
<td>5 (1.6)</td>
<td>19.8</td>
<td>10</td>
<td>5</td>
<td>1.2</td>
</tr>
<tr>
<td>YJM969(^3)</td>
<td>5 (1.6)</td>
<td>8.5</td>
<td>3</td>
<td>4</td>
<td>0.2</td>
</tr>
<tr>
<td>YJM978</td>
<td>5 (1.6)</td>
<td>6.5</td>
<td>2</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>DTY3</td>
<td>No repeat</td>
<td>5.2(^4)</td>
<td>1</td>
<td>NR</td>
<td>&lt;0.1</td>
</tr>
</tbody>
</table>

\(^1\)Average size of *CUP1*-containing *EcoRI* fragment based on two to three independent experiments.
Estimate of \textit{CUP1} copy number based on coverage in deep-sequencing analysis (Strope \textit{et al.} 2015).

This strain had a mixture of Type 3 and Type 5 repeats by deep sequencing (Strope \textit{et al.} 2015), but only Type 5 repeats by PCR analysis.

This size is based on our sequencing results and on Southern analysis of a single-copy \textit{CUP1} strain (Fogel \textit{et al.} 1983).

ND, not determined. NR, not relevant.

2.3.5 Rate of reduction of a tandem \textit{CUP1} array to a single copy of \textit{CUP1}

Although we are primarily concerned with the mechanisms by which single-copy genes become duplicated, a related issue is the mechanism by which a tandem array is reduced to a single-copy gene. We constructed a haploid YZ22 in which a tandem array of about 14 \textit{CUP1} repeats has an integrated \textit{URA3} gene. Unequal crossing (Figure 2.7) or several types of homologous recombination events not depicted (intra-chromatid “pop-out” recombination, and single-strand annealing) can result in a single-copy \textit{CUP1} locus. Derivatives of YZ22 that lose the \textit{URA3} insertion can be identified by selection on solid medium containing 5-fluoro-orotate (details in Section 2.2.6). From measurements of the frequency of 5-FOA\textsuperscript{R} derivatives in multiple independent cultures, we calculated the rate of appearance of 5-FOA\textsuperscript{R} derivatives to be $1.4 \times 10^{-5}$/division.
We depict the \textit{CUP1} array of the haploid strain YZ22 that contains an inserted \textit{URA3} gene within the array. An unequal crossover between the terminal \textit{CUP1} repeat of one array with opposite terminus of another array located on a sister chromatid results in one cell with an elongated array with two \textit{URA3} insertions and another cell with a single \textit{CUP1} repeat and no \textit{URA3} insertion.

To determine what fraction of these derivatives had reduced the array to a single copy, we examined 55 independent 5-FOA\textsuperscript{R} derivatives by PCR using primers from within the flanking \textit{CIC1} and \textit{RSC30} genes (primers VIII211849 F and VIII216603 R, respectively). In strains with a single copy of \textit{CUP1}, a PCR fragment of 2.7 kb was
observed. Two of the 55 strains had a single copy of CUP1. In summary, the approximate rate for reduction of the tandem array of CUP1 genes to a single copy is $2/55 \times 1.4 \times 10^{-5}$ or $5 \times 10^{-7}$/division.

### 2.4 Discussion

Tandemly-repeated genes are a common feature of eukaryotic genomes. These genes often encode proteins or RNA species (such as ribosomal RNA) that are required in large amounts. As discussed in the Introduction, the homologous recombination mechanisms that alter the number of repeats in a pre-existing tandem array are well characterized, and include unequal crossing-over, intra-chromatid crossovers, gene conversion events, and single-strand annealing. The events that generate a tandem duplication of a single-copy gene are less understood. From our analysis, we argue that this duplication event at the CUP1 locus likely reflects the joining of two broken ends by a non-homologous end-joining event.

#### 2.4.1 Previous observations of gene duplications in yeast

In yeast, the number of repeats in the ribosomal RNA gene cluster alters at a high rate, $10^{-1}$/meiotic division (Petes 1980) and $10^{-2}$/mitotic division (Szostak and Wu 1980). In general, the rates of generating de novo duplications are much less, although the rates are variable depending on the context of the reporter gene and the details of the experimental system. If the reporter gene is flanked by repeated elements, the
duplication often occurs by unequal crossovers between these elements. For example, Zhang et al. (Zhang et al. 2013) showed that duplication of a reporter construct located between two Ty elements on chromosome V in a haploid strain was a consequence of unequal crossovers, occurring at a frequency of about $10^{-6}-10^{-7}$. No events in which the duplication was generated by non-homologous end-joining were observed. In strains in which re-replication of an origin could be induced, gene duplications of sequences within the re-replicated region were efficiently generated (Green et al. 2010; Finn and Li 2013). These events usually involved non-allelic homologous recombination between Ty elements flanking the duplicated region.

Haploid yeast strains with a deletion of the ribosomal protein gene RPL20A grow slowly, allowing for selection of fast-growing derivatives that duplicate the related gene RPL20B (Koszul et al. 2004). These duplications involved both flanking repeated genes (delta and Ty elements) and microhomologies. The events involving recombination between large (> 300 bp) regions of homology were Rad52p-dependent, and all classes of duplications had a requirement for Pol32p (Payen et al. 2008). Based on the genetic analysis, Payen et al. (Payen et al. 2008) argued that many of the Rad52p-independent duplications were generated by microhomology-mediated BIR. The rates of duplications of RPL20B were estimated at between $10^{-7}-10^{-10}$, depending on what correction factor was used to determine the relative growth rates of strains with and without the duplication.
(Payen et al. 2008; Koszul et al. 2004). The de novo RPL20B repeats were large, varying between 41 and 655 kb (Koszul et al. 2004). In experiments selecting for duplications of the ADH4 gene, only one large (>100 kb) chromosomal duplication was observed, yielding a rate of about $10^{10}$/cell division (Dorsey et al. 1992).

Although it is difficult to reach a definitive conclusion based on these data, most of the yeast observations suggest that duplications that arise as a consequence of homologous recombination between flanking repeated sequences occur more frequently than those generated by non-homologous end-joining events. In contrast, many duplications in mammalian cells reflect NHEJ or related mechanisms (Zhang et al. 2009a). These differences are consistent with the relative importance of homologous recombination versus non-homologous end-joining in yeast compared to mammals (Jasin and Rothstein 2013).

### 2.4.2 Duplications of CUP1 in yeast

The CUP1 repeats are located on the right arm of chromosome VIII, and there are no flanking Ty elements on this arm. In strains with tandemly-repeated CUP1 genes, Fogel et al. (Fogel et al. 1983) were readily able to select derivatives with longer arrays by growing the strains in high concentrations of copper. They were not, however, able to isolate strains a de novo duplication of CUP1, beginning with a strain that had a single copy. Their results argue that such duplications are infrequent when CUP1 is in its
“normal” location. Our analysis of different types of \textit{CUP1} repeats, however, suggest that such duplications occur during evolution.

All five classes of repeats can be explained by the duplication mechanism shown in Figure 2.8. We suggest that, during replication of a chromosome with only one copy of \textit{CUP1}, two breaks occur, one centromere-distal and one centromere-proximal to the \textit{CUP1} locus. The joining of these broken ends by non-homologous end-joining results in two products, one with a \textit{CUP1} deletion and one with a tandem duplication. An alternative possibility is that the repeats were generated by a single break in the centromere-distal location, followed by a BIR event in which the centromere-proximal site was invaded. Since most BIR events involve either extensive sequence homology (Paques and Haber 1999) or microhomology (Payen \textit{et al}. 2008), and most of the observed breakpoints in our study have very little or no homology, we favor the first alternative.

The differences in the location of the breaks generate repeats that vary in size between about 1.16 kb and 2 kb. The centromere-proximal breaks occur in two locations (Figure 2.9). Most occur in the intergenic region between \textit{CIC1} and \textit{RUF5}, but one occurs within the \textit{CIC1} gene. The centromere-distal breaks occur either between \textit{RUF5} and \textit{RSC30}, or within the \textit{RSC30} gene.
Figure 2.8. Generation of a de novo duplication of CUP1 by unequal NHEJ.

A mechanism for the generation of a two-copy tandem array of CUP1 from a single CUP1 gene (similar to the duplication observed in S288c) is shown in this figure. We suggest that, during DNA replication, one fork is broken in the intergenic region between CIC1 and RUF5/CUP1/ARS810. The second fork is broken within RSC30. The non-homologous end-joining of one broken end to the other (indicated by the dashed line) would produce the duplication. If the other broken ends are also joined, a deletion of RUF5/CUP1/ARS810 would be produced in the sister chromatid.

The locations of the presumptive DSBs required to generate the CUP1 duplications could reflect chromosomal sequences that are intrinsically susceptible to
breakage (fragile sites). In yeast, DNA sequences with the ability to form secondary structures (Voineagu et al. 2008) or to perturb the progression of replication forks (Cha and Kleckner 2002; Paeschke et al. 2011; Song et al. 2014) are hotspots for chromosome breakage.

The minimal size of the duplication is constrained by the sequences required for optimal CUP1 expression. Aside from the CUP1 coding sequence of about 200 bp, optimal expression of the gene and copper-induced transcription requires about 300 bp of upstream sequences (Thiele and Hamer 1986). Since the CUP1 transcript is about 500 bp in size (Karin et al. 1984), the minimal size of a functional CUP1 repeat is expected to be less than 600 bp. However, all of the repeats share sequences from the 3' end of RUF5 to the beginning of the RSC30 gene, a region of about 920 bp. It is possible that these additional sequences are involved in regulating CUP1 gene expression in some environments or enhancing the gene amplification process subsequent to the duplication. It is interesting that none of the duplications include an intact CIC1 or RSC30 gene. Koszul et al. (Koszul et al. 2004) found that duplications of the RPL20B gene often included many flanking genes. One obvious possibility is that yeast cells are intolerant of extra doses of either CIC1 (encoding an essential protein associated with the proteasome; Jäger et al. (Jager et al. 2001) or RSC30 (Angus-Hill et al. 2001).
Figure 2.9. Clustered breakpoints of five different types of \textit{CUP1} repeats.

A. Location of junction breakpoints in the \textit{CIC1} gene and intergenic region for five types of \textit{CUP1} repeats.

B. Sequence in the region of three tightly-cluster breakpoints in the \textit{CIC1-RUF5/CUP1/ARS} intergenic region.

C. Location of junction breakpoints in the \textit{RSC30} gene for five types of repeats.

D. Sequence in the region of two tightly-cluster breakpoints within \textit{RSC30}. 
2.4.3 Alterations in the number of repeats within the *CUP1* tandem array

Several lines of evidence demonstrate that the number of repeats per tandem array alters at high frequency. First, as described above, Fogel *et al.* (Fogel *et al.* 1983) showed that strains with longer arrays of *CUP1* could be isolated by growing the cells in high levels of copper. Second, isolates of yeast obtained from the wild have different numbers of *CUP1* genes in their arrays (Welch *et al.* 1983). In the strains examined in our study, the number of *CUP1* genes varied from one to eighteen (Table 2.1). Third, in several different mutant backgrounds, the rate of alterations in the number of repeats is very high, greater than $10^{-3}$/division (McCulley and Petes 2010; Song *et al.* 2014). Since loss or duplication of *CUP1* repeats occur in integral numbers of units, these events are presumably the result of homologous recombination (unequal crossovers, single-strand annealing, or gene conversion).

About 30% of the 100-genome strains had only one copy of *CUP1*. Such strains could represent the progenitor isolates from which the tandem arrays were derived. Alternatively, these isolates could represent strains in which all copies except one were lost as a consequence of homologous recombination. The second alternative is more plausible, since we found that a strain with a tandem array of 14 *CUP1* repeats could give rise to a derivative with a single *CUP1* repeat at a rate of about $10^{-7}$/division. It is important to stress that the process by which a single *CUP1* repeat becomes duplicated
is inherently different than the process by which a tandem array becomes deleted to a single copy. If the \textit{CUP1} gene is not associated with repeated flanking elements, the duplication process likely occurs by some type of non-homologous end-joining or a BIR event involving microhomologies. The latter type of event can occur by homologous recombination between the terminal repeats of the array.

2.5 \textbf{Summary}

Based on our study and a number of previous studies, we argue that \textit{CUP1} tandem arrays arise as a consequence of two mechanisms: 1) a duplication of a single-copy \textit{CUP1} gene by an infrequent non-homologous recombination event and 2) increases in copy number from the duplication to multiple copies by a frequent homologous recombination event. In environments with high levels of copper, strains with larger numbers of \textit{CUP1} repeats will be selected. Our observation that a tandem array of \textit{CUP1} repeats can be reduced to a single \textit{CUP1} copy emphasizes the necessity of a mechanism for generating \textit{de novo} duplications.

2.6 \textbf{Supplementary tables}

Table S 2.1 Strain genotypes.

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Genotype (reference)\textsuperscript{1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>S288c</td>
<td>\textit{MATa gal2} (Engel et al. 2014)</td>
</tr>
<tr>
<td>W303-1A</td>
<td>\textit{MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100\textDelta::NAT}</td>
</tr>
<tr>
<td></td>
<td>Genetic Information</td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
</tr>
<tr>
<td>YZ22</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::NAT VIII212898::URA3 RAD5</td>
</tr>
<tr>
<td>JSC10-1</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::NAT ho::hisG RAD5 (St Charles and Petes 2013)</td>
</tr>
<tr>
<td>YJM789</td>
<td>MATα lys2 gal2 ho::hisG (Wei et al. 2007)</td>
</tr>
<tr>
<td>YJM799</td>
<td>MATα ura3 gal2 ho::hisG (provided by J. McCusker, Duke Univ.)</td>
</tr>
<tr>
<td>JSC19-1</td>
<td>MATα ade2-1 ura3 gal2 ho::hisG CAN1Δ::NAT (St Charles and Petes 2013)</td>
</tr>
<tr>
<td>DTY3</td>
<td>MATα trp1-1 leu2-3,112 gal1 ura3-50 his cup1s (Hamer et al. 1985)</td>
</tr>
<tr>
<td>YJM189</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM271</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM456</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM693</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM969</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM972</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM978</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM996</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
<tr>
<td>YJM1307</td>
<td>Wild-type diploid from 100-genome strains (Strope et al. 2015)</td>
</tr>
</tbody>
</table>
YJM1549 | Wild-type diploid from 100-genome strains (Strope et al. 2015)

*Strains from the 100-genome collection were provided by Strope, et al. (Strope et al., 2015)*

### Table S 2.2. Primer names and sequences used in strain constructions and analysis.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>CATTGGCACTCATGACCTTCA</td>
</tr>
<tr>
<td>R1</td>
<td>AATAAAGTATCTCCATATGTCGCC</td>
</tr>
<tr>
<td>R1'</td>
<td>AATCATGTAGCTGCCCCAACGG</td>
</tr>
<tr>
<td>cup1 amp5-2</td>
<td>CGAGATGAAATGAATAGCAACGG</td>
</tr>
<tr>
<td>cup1 amp5-3</td>
<td>CTCCTTGTCTTGTATCAATTGCAT</td>
</tr>
<tr>
<td>cup1 amp3</td>
<td>TTCATTTCCCCAGAGCAAGCATGAC</td>
</tr>
<tr>
<td>VIII212898::URA3 F</td>
<td>TACAAGACAAAGGAGTTATTGCTTCTTTATATGAT</td>
</tr>
<tr>
<td></td>
<td>TCTGACAATCCATATTGCGTTGGTATCCTTTTAATGTG</td>
</tr>
<tr>
<td></td>
<td>GCTGTGTTTTCAGG</td>
</tr>
<tr>
<td>VIII212898::URA3 R</td>
<td>ATTCTTTTGTGGCATTTCCTTTCTAGAAGCAAAAAAGAGC</td>
</tr>
<tr>
<td></td>
<td>GATGCGTCTTCCCGCTGAACCCTTCCAGCAAAGATT</td>
</tr>
<tr>
<td></td>
<td>CCCGGGTAATAACTG</td>
</tr>
<tr>
<td>Accession</td>
<td>Sequence</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>VIII211849 F</td>
<td>GGAAATCGCCAATCCTCCGAATT</td>
</tr>
<tr>
<td>VIII216603 R</td>
<td>TATCTCAACTGACAAAACTAGGCG</td>
</tr>
<tr>
<td>VIII212300 F</td>
<td>GAGTTGTAAGTTAGGCAAACCTAGA</td>
</tr>
<tr>
<td>VIII213200 F</td>
<td>ATCATAGAAATCGTGGAAGTTTTCG</td>
</tr>
<tr>
<td>VIII213031 R</td>
<td>CTGATATCTTAGCTTGTACTAG</td>
</tr>
<tr>
<td>VIII212063 R</td>
<td>TACATATGCACCCGACTCTATG</td>
</tr>
<tr>
<td>VIII210632 F</td>
<td>TGCTTCACCGTTGCCTCACAA</td>
</tr>
<tr>
<td>VIII211849 F</td>
<td>GGAAATCGCCAATCCTCCGAATT</td>
</tr>
<tr>
<td>VIII216603 R</td>
<td>TATCTCAACTGACAAAACTAGGCG</td>
</tr>
<tr>
<td>VIII213601 F</td>
<td>GCAATTGAACATTAATCTCCTCAT</td>
</tr>
</tbody>
</table>
Table S 2.3. Sequence analysis of the CUP1 repeats (Type 2, 1.8 kb) of YJM189.

In this table, we show genomic sequences of YJM189 in three regions: 1) the sequences that flank the CUP1 repeats adjacent to CIC1, 2) the sequence of the CUP1 repeat, and 3) the sequences that flank the CUP1 tandem array adjacent to RCS30. The sequences of YJM189 (denoted “Query” below) were compared in a BLAST search with sequences of S288c (denoted “Sbjct”). SNPs that distinguish YJM189 and S288c sequences are summarized at the end of the table. The CUP1 coding sequences are shown in red. The names of the primers used in the sequence analysis are shown in boldface. Additional details about the sequencing are in Section 2.2.7.

1. CIC1-CUP1 (VIII211739-212339)

VIII211528 F

Query: 182   CTTGATGAACT 192

Sbjct: 211739 CTTGATGAACT 211749

Query: 193   TGAAGCTAAAAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCACCAT 246

Sbjct: 211750 TGAAGCTAAAAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCACCAT 211803
**Sbjct:** 212155 ATCTGTTGTACTATCCGCTTCAAATAAATAGATCATGAAAGTGACGGGGATAACAGCAT 212214

**Query:** 393 TTTACCTTTAAAGACGTTCTCATAATAGATTTTAGGATTAATACATATGCTTTTTTTTT 334

**Sbjct:** 212215 TTTACCTTTAAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCTTTTTTTTT 212274

**Query:** 333 TATTCGAAATCTGGGGATTCTATACAGAGTTGTAAGTTAGGCAAACTAGAATTTGGTAAT 274

**Sbjct:** 212275 -ATTCGAAATCTGGGGATTCTATACAGAGTTGTAAGTTAGGCAAACTAGAATTTGGTAAT 212333

**Query:** 273 AATATT 268

**Sbjct:** 212334 AATATT 212339

---

**2. CUP1 Repeat (VIII212039-213867)**

**R1’**

**Query:** 574 TTTCATAGAGTGGGTTGCATATGTATATATCTATATATGTTTGAAGTGTATATTTAAAAAT 515

**Sbjct:** 212039 TTTCATAGAGTGGGTTGCATATGTATATATCTATATATGTTTGAAGTGTATATTTAAAAAT 212098

**Query:** 514 AAAGTCATTATTTGAAATATTGGGTTTCTCGGTCTAAAGACAATTTAAAATGACTGATCT 455

**Sbjct:** 212099 AAAGTCATTATTTGAAATATTGGGTTTCTCGGTCTAAAGACAATTTAAAATGACTGATCT 212158
<table>
<thead>
<tr>
<th>Query</th>
<th>Sbjct</th>
</tr>
</thead>
<tbody>
<tr>
<td>187</td>
<td>ATATATTAAAGACTATTCTTTATTTTCCAGAGGCAGCATGACTTTCTGTTTCCTTTCAGA 246</td>
</tr>
<tr>
<td>212514</td>
<td>ATATATTAAAGACTATTCTTTATTTTCCAGAGGCAGCATGACTTTCTGTTTCCTTTCAGA 212573</td>
</tr>
<tr>
<td>247</td>
<td>CTTGTTACCAGGGGATTTTGCTGCTGTTACACCCCCGTGGGAGCTACATGATT 306</td>
</tr>
<tr>
<td>212574</td>
<td>CTTGTTACCAGGGGATTTTGCTGCTGTTACACCCCCGTGGGAGCTACATGATT 212633</td>
</tr>
<tr>
<td>307</td>
<td>TTGGCATTGTTCATTATTTTTGCAGCTACCACATTGGCATTGGCAGCTACATGACCTTCATT 366</td>
</tr>
<tr>
<td>212634</td>
<td>TTGGCATTGTTCATTATTTTTGCAGCTACCACATTGGCATTGGCAGCTACATGACCTTCATT 212693</td>
</tr>
<tr>
<td>367</td>
<td>TTGGAAAGTTAAATTTAATTTCGCTGAAACATTATTATATGTGATGATTGATTTGATTGATTG---TACAGT 422</td>
</tr>
<tr>
<td>212694</td>
<td>TTGGAAAGTTAAATTTAATTTCGCTGAAACATTATTATATGTGATGATTGATTTGATTTGATTTGATTTGACGTACAGT 212753</td>
</tr>
<tr>
<td>423</td>
<td>TTGTTTTCTTAATATCTATTTTCATGACTTTCTATATGATATTGCACATAAAGAGATA 482</td>
</tr>
<tr>
<td>212754</td>
<td>TTGTTTTCTTAATATCTATTTTCATGACTTTCTATATGATATTGCACATAAAGAGATA 212813</td>
</tr>
<tr>
<td>483</td>
<td>TTATAATGCAATTGATACAAGACAAGGAGTTATTTGCTTCTCTTTTATATGATTCTGACA 542</td>
</tr>
<tr>
<td>212814</td>
<td>TTATAATGCAATTGATACAAGACAAGGAGTTATTTGCTTCTCTTTTATATGATTCTGACA 212873</td>
</tr>
<tr>
<td>543</td>
<td>ATCCCATATTTCGGGTGGTAGCTTTTTTTGCTGGAACCGGTTACCCGGAAGAGACGCATCGCT 602</td>
</tr>
<tr>
<td>212874</td>
<td>ATCCCATATTTCGGGTGGTAGCTTTTTTTGCTGGAACCGGTTACCCGGAAGAGACGCATCGCT 212933</td>
</tr>
</tbody>
</table>
Query:  603  CTTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAA  662
Sbjct:  212934  CTTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAA  212993

Query:  663  ATGTCTT  669
Sbjct:  212994  ATGTCTT  213000

F1

Query:  297  TTTCTAACTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAAGTAATAAT  356
Sbjct:  213001  TTTCTAACTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAAGTAATAAT  213060

Query:  357  TAAATCATCCGTATAACCTATACATATATGAGGAAAAATAATACAAAAAGTGTGTTTAA  416
Sbjct:  213061  TAAATCATCCGTATAACCTATACATATATGAGGAAAAATAATACAAAAAGTGTGTTTAA  213120

Query:  417  ATACAGATACATACTACATAGATACGTATAGCGTCTTAAATGTCGGTAATGGGATCGG  476
Sbjct:  213121  ATACAGATACATACTACATAGATACGTATAGCGTCTTAAATGTCGGTAATGGGATCGG  213180

Query:  477  CTTACTAATTATAAAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAATACCCAGAT  536

Sbjct: 213181  CTTACTAATTATAAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAATACCCAGAT 213240

Query: 537   TATCAGATTCCAAATCCTTGTCAATAATTATACTCCTTTGGAAAACTTCTCTTTCCATTA 596

Sbjct: 213241  TATCAGATTCCAAATCCTTGTCAATAATTATACTCCTTTGGACAACTTCTCTTTCCATTA 213300

Query: 597   AAAAATCTGAAATCTCCTTAAATTTTAAATAGATTCTGTTCAGTTCACTAACGGGGAATT 656

Sbjct: 213301  AAAAATCTGAAATCTCCTTAAATTTTAAATAGATTCTGTTCAGTTCACTAACGGGGAATT 213360

VIII213200 F

Query: 116   TCAAGAGAACAT 127

Sbjct: 213361  TCAAGAGAACAT 213372

Query: 128   TTTTGGTCTTCCGCGACTGACTATAATCTGTAACATTATTGTTATCAGAGTTTCTCGCAA 187

Sbjct: 213373  TTTTGGTCTTCCGCGACTGACTATAATCTGTAACATTATTGTTATCAGAGTTTCTCGCAA 213432

Query: 188   AATTTGGTTTTTTTCTTGCTAAATCTCAGCATATATTTAATCAGATTCAAAACCTTGTTGA 247

Sbjct: 213433  AATTTGGTTTTTTTCTTGCTAAATCTCAGCATATATTTAATCAGATTCAAAACCTTGTTGA 213492

Query: 248   AACCTTTAATAGATTGAAATTTCGTTGCTATTTCATCTCGTAAAAGGATACGA 307

Sbjct: 213493  AACCTTTAATAGATTGAAATTTCGTTGCTATTTCATCTCGTAAAAGGATACGA 213552
Query: 308    TAATTTCTATTTTTTTAAAATTTCCAAAATCTTGTACATGAAATCAATAGCAATTGAACAT 367
Sbjct: 213553 TAATTTCTATTTTTTTAAAATTTCCAAAATCTTGTACATGAAATCAATAGCAATTGAACAT 213612

Query: 368    TAATCTCCTCATTTGGAAAGATTTTTGTAAAATTCGTCATATAATATTACTTCACAACGTT 427
Sbjct: 213613 TAATCTCCTCATTTGGAAAGATTTTTGTAAAATTCGTCATATAATATTACTTCACAACGTT 213672

Query: 428    GGAAAATAGCAAATGTGATTGCTATAAAATTCTGTAAGATTTCAATAAAATGATTTGCGA 487
Sbjct: 213673 GGAAAATAGCAAATGTGATTGCTATAAAATTCTGTAAGATTTCAATAAAATGATTTGCGA 213732

Query: 488    ATAAAAATTCTTTACCA
Sbjct: 213733 ATAAAAATTCTTTACCA 213792

R1'

Query: 625    ATGTTATTGAGTAAAATAATGTGCATATTAGAAATAATTTTCATCAGATCCT
Sbjct: 213817 ATGTTATTGAGTAAAATAATGTGCATATTAGAAATAATTTTCATCAGATCCT 213867
3. **CUP1-RSC30 (VIII213567-214167)**

**VIII213234 F**

Query: 304  
TTTAAAATTTCCAAATCTTGTCATGATCAATAGCAATTGAACATTAATCTCCTCATTT  363
Sbjct: 213567  
TTTAAAATTTCCAAATCTTGTCATGATCAATAGCAATTGAACATTAATCTCCTCATTT  213626

Query: 364  
GAAAGATTTTTGTAAAATTCGTCATATAATATTACTTCAAACGTTGGAATAATGCAAAT  423
Sbjct: 213627  
GAAAGATTTTTGTAAAATTCGTCATATAATATTACTTCAAACGTTGGAATAATGCAAAT  213686

Query: 424  
GTGATTGCTATAAAATCTGTTGATTTCCAAATATAATATTACTTTCCCAGTTGGAATAATGCAAAT  483
Sbjct: 213687  
GTGATTGCTATAAAATCTGTTGATTTCCAAATATAATATTACTTTCCCAGTTGGAATAATGCAAAT  213746

Query: 484  
CCATTAGAATGAAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAG  543
Sbjct: 213747  
CCATTAGAATGAAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAG  213806

Query: 544  
ATAAATTTAAATTTACTGACATAAAAATGTGACATTTATGATTTCAATTTTCAATTTTATGCTCAGATCC  603
Sbjct: 213807  
ATAAATTTAAATTTACTGACATAAAAATGTGACATTTATGATTTCAATTTTCAATTTTATGCTCAGATCC  213866

Query: 604  
TTTGCACATCTTTCAAGGCCAGTTGCTTTATTGTTAGTTGGAATAATGCTGCATCATCAGATCC  663
Sbjct: 213867  
TTTGCACATCTTTCAAGGCCAGTTGCTTTATTGTTAGTTGGAATAATGCTGCATCATCAGATCC  213926
Sbjct: 213867 TTTGCACATCTTTCAAGTGTTTGCAGTTTATTTGTGTTAGAAGAATGTGTACTGCCATG

Query: 664 GACAAAGAGGATTCGTTTTGAAACAAAAAGGAAAAAATTTGTATAAAACATGGTATTGATA

Sbjct: 213927 GACAAAGAGGATTCGTTTTGAAACAAAAAGGAAAAAATTTGTATAAAACATGGTATTGATA

VIII216603 R

Query: 588 AAAT 585

Sbjct: 213987 AAAT

Query: 584 TTTAAGGTGCTTTTCATTTTCTGACTTTGTTGTCATGAAAATATAAGTCTACTGTAT

Sbjct: 213991 TTTAAGGTGCTTTTCATTTTCTGACTTTGTTGTCATGAAAATATAAGTCTACTGTAT

Query: 524 TACTCACGCCCATAGTCAAGGTTTTCTAACAGACTTTTCAATTTTTGTTAATTTTACTGGCA

Sbjct: 214051 TACTCACGCCCATAGTCAAGGTTTTCTAACAGACTTTTCAATTTTTGTTAATTTTACTGGCA

Query: 464 AGTAGAAAGGAACATCTTTGCAGAAATTTATTAATTTTGGCTTTGCTTTCCAGTAATT

Sbjct: 214111 AGTAGAAAGGAACACCTTTGCAGAAATTTATTAATTTTGGCTTTGCTTTCCAGTAATT

SNPs between YJM189 and S288c
<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in YJM189</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CIC1-CUP1 VIII211739-212339</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>211825</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>212243</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 poly(T)</td>
<td>9 poly(T)</td>
</tr>
<tr>
<td><strong>CUP1 repeat VIII212039-213867</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>212243</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 poly(T)</td>
<td>9 poly(T)</td>
</tr>
<tr>
<td></td>
<td>212744-212747</td>
<td>4 bp deletion</td>
<td>ATTG</td>
</tr>
<tr>
<td></td>
<td>213157</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213283</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213513</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td><strong>CUP1-RSC30 VIII213567-214167</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>213823</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>214125</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>
Table S 2.4. Sequence analysis of the CUP1 repeats (Type 2, 1.8 kb) of YJM996.

In this table, we show genomic sequences of YJM96 in three regions: 1) the sequences that flank the CUP1 repeats adjacent to CIC1, 2) the sequence of the CUP1 repeat, and 3) the sequences that flank the CUP1 tandem array adjacent to RCS30. The sequences of YJM189 (denoted “Query” below) were compared in a BLAST search with sequences of S288c (denoted “Sbjct”). SNPs that distinguish YJM189 and S288c sequences are summarized at the end of the table. The CUP1 coding sequences are shown in red. The names of the primers used in the sequence analysis are shown in boldface. Additional details about the sequencing are in Section 2.2.7.

1. **CIC1-CUP1 (VIII211739-212339)**

R1′

<table>
<thead>
<tr>
<th>Query: 863</th>
<th>CTTGATGAACCTGAAGCTAAA 843</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sbjct: 211739</td>
<td>CTTGATGAACCTGAAGCTAAA 211759</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query: 842</th>
<th>AAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCACCATTGATGGTGTACAAGTT 783</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sbjct: 211760</td>
<td>AAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCACCATTGATGGTGTACAAGTT 211819</td>
</tr>
</tbody>
</table>
Query: 782  CATTATCTACCTTCAACAAGGTTTTGATGGAAATCGCCAATCCTCCGAATTTGGTTCA 723
\[\text{Sbjct: 211820 CATTTGTCTACCTTCAACAAGGGTTTGATGGAAATCGCCAATCCTCCGAATTTGGTTCA 211879}\]

Query: 722  ATTTTCTCTAAACAAATTAACAATGCAAAAAAGAGATCTTCTAGCGAGCTTGAAAAAGAA 663
\[\text{Sbjct: 211880 ATTTTCTCTAAACAAATTAACAATGCAAAAAAGAGATCTTCTAGCGAGCTTGAAAAAGAA 211939}\]

Query: 662  TCTAGCGAGTCAGAAGCTGTCAAGAAGGCTAAAAGTTAATTTGTTTCCTCCTTATCTATC 603
\[\text{Sbjct: 211940 TCTAGCGAGTCAGAAGCTGTCAAGAAGGCTAAAAGTTAATTTGTTTCCTCCTTATCTATC 211999}\]

Query: 602  TTTTCTCTCTATTTTTTTCTTGTGAAGAAAAAAATTTGAATTTCATAGAGTGCGGTGCATA 543
\[\text{Sbjct: 212000 TTTTCTCTCTATTTTTTTCTTGTGAAGAAAAAAATTTGAATTTCATAGAGTGCGGTGCATA 212059}\]

Query: 542  TGTATATATCTATATATGTTTGAAGTGTATATTAAAAATAAAGTCATTATTTGAATATTG 483
\[\text{Sbjct: 212060 TGTATATATCTATATATGTTTGAAGTGTATATTAAAAATAAAGTCATTATTTGAATATTG 212119}\]

Query: 482  GTTTCTCGGTCTAAGAGCTTATACGTTTTAGACTGATCTGTTGTACTATCCGCTTCAAAT 423
\[\text{Sbjct: 212120 GTTTCTCGGTCTAAGAGCTTATACGTTTTAGACTGATCTGTTGTACTATCCGCTTCAAAT 212179}\]

Query: 422  AAATAGATCATTTGAAAGTGACGGGGATAACAGCATTTTACCTTTAAAAAGAGCTTCTCATA 363
\[\text{Sbjct: 212180 AAATAGATCATTTGAAAGTGACGGGGATAACAGCATTTTACCTTTAAAAAGAGCTTCTCATA 212239}\]
2. **CUP1 repeat (VIII212039-213867)**

R1’

Query: 572    TTTCATAGAGTGCGGTGCATATGTATATCTATATATGTTTGAAGTGTATATTAAAAAT 513
Sbjct: 212039 TTTCATAGAGTGCGGTGCATATGTATATCTATATATGTTTGAAGTGTATATTAAAAAT 212098

Query: 512    AAAAGTCATTATTTGAATATTGGTTTCTCGGTCTAAGAGCTTATACGTTTTAGACTGATCT 453
Sbjct: 212099 AAAAGTCATTATTTGAATATTGGTTTCTCGGTCTAAGAGCTTATACGTTTTAGACTGATCT 212158

Query: 452    GTTGTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTTA 393
Sbjct: 212159 GTTGTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTTA 212218

Query: 392    CTTTTAAAGACGTTCTCATAATAGATTATTTAGGATTAATACATATGCTTTTTTTTTATT 333
Sbjct: 212218 CTTTTAAAGACGTTCTCATAATAGATTATTTAGGATTAATACATATGCTTTTTTTTTATT 212278
Sbjct: 212219  CCTTTAAAAAGACGTTCTCATAATACATTAGGATTAATAACATATGCTTTTTTTTT-ATT 212277

Query: 332  CGAAATCTGGGAGATTCTAAGAGTCTAGAGATTTGTAATAATA 273
Sbjct: 212278  CGAAATCTGGGAGATTCTAAGAGTCTAGAGATTTGTAATAATA 212337

Query: 272  TTTTATTCTTTGCGACATATGGGAGATTTTATCTTTCTTTTTTATTATTACGTAT 213
Sbjct: 212338  TTTTATTCTTTGCGACATATGGGAGATTTTATCTTTCTTTTTTATTATTACGTAT 212397

Query: 212  ACCTATAAAATTAACAAAGTATCTAAACAAAATACATAAGTGTACTCAAACTGAGTAGAAT 153
Sbjct: 212398  ACCTATAAAATTAACAAAGTATCTAAACAAAATACATAAGTGTACTCAAACTGAGTAGAAT 212457

### VIII 212300 F

Query: 131  CGTGGATTAAACTTCCTTCTCCTTAAAAATTAAAAACAGCAAATAGTTAGATGA 186
Sbjct: 212458  CGTGGATTAAACTTCCTTCTCCTTAAAAATTAAAAACAGCAAATAGTTAGATGA 212513

Query: 187  ATATATTAAGACTATTCTGAATCTTCCCCAGAGCAGCATATGACTTTGTGTTCTTCAGA 246
Sbjct: 212514  ATATATTAAGACTATTCTGAATCTTCCCCAGAGCAGCATATGACTTTGTGTTCTTCAGA 212573

Query: 247  CTTGTTACCAGGAGTCATTTCTGCCTGTGGTACACCCCCGTGGGCACTACATGGATTT 306
Sbjct: 212574  CTTGTTACCAGGAGTCATTTCTGCCTGTGGTACACCCCCGTGGGCACTACATGGATTT 212633
Sbjct: 212948 AAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAAATGTCTTTTTCTAA 213007
Query: 305  CTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAGTAATAATTAAATCA 364
Sbjct: 213008 CTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAGTAATAATTAAATCA 213067
Query: 365  TCCGTATAAAACCTATACACATATATGAGGAAAAATAATACAAAAAGTGGTTTTAAATACAGA 424
Sbjct: 213068 TCCGTATAAAACCTATACACATATATGAGGAAAAATAATACAAAAAGTGGTTTTAAATACAGA 213127
Query: 425  TACATACATGAACATATGCACGTATAGCGTCCAAATGTCGGTAATGGGATCGGCTTACTA 484
Sbjct: 213128 TACATACATGAACATATGCACGTATAGCGTCCAAATGTCGGTAATGGGATCGGCTTACTA 213187
Query: 485  ATTATAAAATGCATCATHAAATCGTTGAAGTTGGCCGTAGTAATACCCAGATTACAGA 544
Sbjct: 213188 ATTATAAAATGCATCATHAAATCGTTGAAGTTGGCCGTAGTAATACCCAGATTACAGA 213247
Query: 545  TTCCAAATCCTTGTCAATAATTATACCTCTCCTTTTGGAAACTTTCTCTTTCCATTTAAAAATC 604
Sbjct: 213248 TTCCAAATCCTTGTCAATAATTATACCTCTCCTTTTGGAAACTTTCTCTTTCCATTTAAAAATC 213307

VIII213200 F
Query: 84  TGAAATCTCCTAAATTTTTAAATAGATTTCTGTTCAGTTCACTAAC 128
Sbjct: 213308 TGAAATCTCCTAAATTTTTAAATAGATTTCTGTTCAGTTCACTAAC 213352
3. **CUP1-RSC30 (VIII213567-214167)**

**VIII213200 F**

Query: 342  TTTAAAATTTCAGAAAATCTGT  364  
Sbjct: 213567 TTTAAAATTTCAGAAAATCTGT  213588

Query: 365  CATGAATCAATAGCAATTTGAACATTTACCACCAATTTTGAAGATTTTTTGTAAAATTCGT  424  
Sbjct: 213589 CATGAATCAATAGCAATTTGAACATTTACCACCAATTTTGAAGATTTTTTGTAAAATTCGT  213648

Query: 425  CATATAATTTACTTCCAAACGCATTGGAAATAGCAAATGTGATTGCTATAAAATTCTGTA  484  
Sbjct: 213649 CATATAATTTACTTCCAAACGCATTGGAAATAGCAAATGTGATTGCTATAAAATTCTGTA  213708
Query: 485  AGATTTCAATAAAATGATTTGCGAATAAAAATTCTTTACCATTAGAATGAAAGCGATTAT  544
Sbjct: 213709 AGATTTCAATAAAATGATTTGCGAATAAAAATTCTTTACCATTAGAATGAAAGCGATTAT  213768

Query: 545  TGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAAAATTAAATGTTATTGAGT  604
Sbjct: 213769 TGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAAAATTAAATGTTATTGAGT  213828

Query: 605  AAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTTTGCACATCTTTCAGAGTTCG  664
Sbjct: 213829 AAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTTTGCACATCTTTCAGAGTTCG  213888

Query: 665  AGGTCTTATTGTTGTTAGAAGAATGTTGAACTGCCATGGACAAAGAGGATTCGTTTTGAA  724
Sbjct: 213889 AGGTCTTATTGTTGTTAGAAGAATGTTGAACTGCCATGGACAAAGAGGATTCGTTTTGAA  213948

VIII216603 R

Query: 630  CAAAAAGGA  622
Sbjct: 213949 CAAAAAGGA  213957

Query: 621  AAAAATTTGTATAAAACATGGTATTGATAAATTTAAAAGTGTCTTTCCATTCTTTCTGA  562
Sbjct: 213958 AAAAATTTGTATAAAACATGGTATTGATAAATTTAAAAGTGTCTTTCCATTCTTTCTGA  214017
SNPs between YJM996 and S288c

<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in YJM996</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIC1-CUP1 VIII211739-212339</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>211825</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>212243</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 poly(T)</td>
<td>9 poly(T)</td>
</tr>
<tr>
<td>CUP1 repeat</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S 2.5. Sequence analysis of the CUP1 repeats (Type 3, 1.2 kb) of YJM789.

In this table, we show genomic sequences of YJM789 in three regions: 1) the sequences that flank the CUP1 repeats adjacent to CIC1, 2) the sequence of the CUP1 repeat, and 3) the sequences that flank the CUP1 tandem array adjacent to RCS30. The sequences of YJM789 (denoted “Query” below) were compared in a BLAST search with sequences of S288c.
SNPs that distinguish YJM789 and S288c sequences are summarized at the end of the table. The CUP1 coding sequences are shown in red. The names of the primers used in the sequence analysis are shown in boldface. Additional details about the sequencing are in Section 2.2.7.

1. CIC1-CUP1 (VIII211728-212328)

VIII211528 F

Query: 165  ATCAGGACGTTCCTTGAT 181

Query: 182  GAACCTGAAGCTAAAAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTTCACCATT 241

Query: 242  GATGGTGTACAAAGTTGGAATTTTCTCTACTTCAACAAGGGTTTGATGGAAATCGCAATCCTCCTCT 301

Query: 302  TCCGAATTGGGTTCAATTTTTCTCTCTAAACAAATTAACAATGCAAAAAAGAGATCTTCTAGC 361
Query: 362   GAGCTTGAAAAAGAATCTAGCGAGTCAGAAGCTGTCAAGAAGGCTAAAAGTTAATTTGTT 421
                    |                                                  |
Sbjct: 211925 GAGCTTGAAAAAGAATCTAGCGAGTCAGAAGCTGTCAAGAAGGCTAAAAGTTAATTTGTT 211984

Query: 422   TCCTCCTTATCTATCTTTTCTCTCATTTTTTTCTTGTGAAGAAAAAAATTTGAATTTCAT 481
                    |                                                  |
Sbjct: 211985 TCCTCCTTATCTATCTTTTCTCTCATTTTTTTCTTGTGAAGAAAAAAATTTGAATTTCAT 212044

Query: 482   AGAGTGCGGTGCATATGTATATATCTATATATGTTTGAAGTGTATATTAAAAATAAAGTC 541
                    |                                                  |
Sbjct: 212045 AGAGTGCGGTGCATATGTATATATCTATATATGTTTGAAGTGTATATTAAAAATAAAGTC 212104

Query: 542   ATTATTTGAATATTGGTTTCTCGGTCTAAGAGCTTATACGTTTTAGACTGATCTGTTGTA 601
                    |                                                  |
Sbjct: 212105 ATTATTTGAATATTGGTTTCTCGGTCTAAGAGCTTATACGTTTTAGACTGATCTGTTGTA 212164

Query: 602   CTATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTTACCTTTA 661
                    |                                                  |
Sbjct: 212165 CTATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTTACCTTTA 212224

Query: 378   AAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCTTTTTTTT 328
                    |                                                  |
Sbjct: 212225 AAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCTTTTTTTT- 212274

Query: 327   ATTGCAAATCTGGGGATTTTTATACAGAGTCAAGTTAGGCAAACTAGAATTTG 274
2. **CUP1 repeat (VIII212028-213190)**

**F1**

Query: 487  ATAAATTTGAATTTCATAGAGTGCGGTGCATATGTATATATCTATATATGTTTGAAGTGT 546

Sbjct: 212028  AAAAATTTGAATTTCCATAGAGTGCGGTGCATATGTATATATCTATATATGTTTGAAGTGT 212087

Query: 547  ATATTAAAAATAAAGTCATTATTTGAATATTGGTTTCTCGGTCTAAGAGCTTATACGTTT 606

Sbjct: 212088  ATATTAAAAATAAAGTCATTATTTGAATATTGGTTTCTCGGTCTAAGAGCTTATACGTTT 212147

**R1’**

Query: 456  TAGACTGATCTGTTGTACTATCCGCTTCAAATAAATAGATCATTTGGAAGTACGGGGATA 397

Sbjct: 212148  TAGACTGATCTGTTGTACTATCCGCTTCAAATAAATAGATCATTTGGAAGTACGGGGATA 212207

Query: 396  ACAGCATTTTACCTTTAAAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCTT 337

Sbjct: 212208  ACAGCATTTTACCTTTAAAAGACGTTCTCATAATACATTTTAGGATTAATACATATGCTT 212267

Query: 336  TTTTTTTTATTCGAAATCTGGGGATTTTATACAGAGTTGTAAGTTAGGCAAACCTAGAATT 277

Sbjct: 212268  TTTTTTTTATTCGAAATCTGGGGATTTTATACAGAGTTGTAAGTTAGGCAAACCTAGAATT 212326
Query: 276    TGGTAATAATATTTTATTTTCTCCTTTCTTTCTTAAT 217
Sbjct: 212327 TGGTAATAATATTTTATCTTGGGGCGACATATGGAGATACTTTATTTCCTTTTCTTAAT 212386

Query: 216    TATTAACGTATACCTATAAATTAACAAAGTATCTAAACAAAATACATAAGTGTACTCAAA 157
Sbjct: 212387 TATTAACGTATACCTATAAATTAACAAAGTATCTAAACAAAATACATAAGTGTACTCAAA 212446

Query: 156    CTGAGTAGAATCGTCGATTAAACTTCCTTCTCCTTTTAAAAATTAAAAACAGTAAATAGT 97
Sbjct: 212447 CTGAGTAGAATCGTCGATTAAACTTCCTTCTCCTTTTAAAAATTAAAAACAGTAAATAGT 212506

**VIII212300 F**

Query: 180    TAAATGAA 187
Sbjct: 212507 TAGATGAA 212514

Query: 188    TATATTAAGACTATTCGTTTTATTTTCCAGAGCAGCATGATTTTCTTTCAGAC 247
Sbjct: 212515 TATATTAAGACTATTCGTTTTATTTTCCAGAGCAGCATGATTTTCTTTCAGAC 212574

Query: 248    TTGTTACCAGGGGACATTATTGTCGTCGTGTTACACCACCCGGTGGGCAGCTACATGATTTT 307
Sbjct: 212575 TTGTTACCAGGGGACATTATTGTCGTCGTGTTACACCACCCGGTGGGCAGCTACATGATTTT 212634
Query: 290    TGTCTTTTT 298
Sbjct: 212995 TGTCTTTTT 213003

Query: 299    CTAACATGAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAGTAATAATTAA 358
Sbjct: 213004 CTAACATGAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAGTAATAATTAA 213063

Query: 359    ATCATCCGTATAAACCTACATCATATGCACGTGCTCAATGTCGGTAATGGGATCGGCT 478
Sbjct: 213064 ATCATCCGTATAAACCTACATCATATGCACGTGCTCAATGTCGGTAATGGGATCGGCT 213122

Query: 419    ACAGATACATACATGAACATATGCACGTATAGCGTCACAATGTCGGTAATGGGATCGGCT 478
Sbjct: 213123 ACAGATACATACATGAACATATGCACGTATAGCGTCACAATGTCGGTAATGGGATCGGCT 213182

Query: 479    TACTAATT 520
Sbjct: 213183 TACTAATT 213190

3. CUP1-RSC30 (VIII212890-213490)

F1

Query: 186    TAGTCTTTTTTGGCTAGGTTGCTAGGTAACCGGAAAGACGCACTCGCTTTTTGCTT 239
SNPs between YJM789 and S288c

<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in YJM789</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>CIC1-CUP1</em> VIII211728-212328</td>
<td>211729</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 T’s</td>
<td>9 T’s</td>
</tr>
<tr>
<td>CUP1 repeat</td>
<td>212293</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>VIII212028-213190</td>
<td>211729</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 T’s</td>
<td>9 T’s</td>
</tr>
<tr>
<td></td>
<td>212293</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212499</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212509</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>212536</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212556</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212744-212747</td>
<td>4 bp deletion</td>
<td>ATTG</td>
</tr>
<tr>
<td></td>
<td>212751</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>212762</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212828</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>213097-213101</td>
<td>6 A’s</td>
<td>5 A’s</td>
</tr>
<tr>
<td></td>
<td>213157</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>CUP1-RSC30</td>
<td>213097-213101</td>
<td>6 A’s</td>
<td>5 A’s</td>
</tr>
<tr>
<td>VIII212890-213490</td>
<td>213157</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213174</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>213283</td>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>
Table S 2.6. Sequence analysis of the CUP1 repeats (Type 4, 1.9 kb) of YJM271.

In this table, we show genomic sequences of YJM271 in three regions: 1) the sequences that flank the CUP1 repeats adjacent to CIC1, 2) the sequence of the CUP1 repeat, and 3) the sequences that flank the CUP1 tandem array adjacent to RCS30. The sequences of YJM271 (denoted “Query” below) were compared in a BLAST search with sequences of S288c (denoted “Sbject”). SNPs that distinguish YJM271 and S288c sequences are summarized at the end of the table. The CUP1 coding sequences are shown in red. The names of the primers used in the sequence analysis are shown in boldface. Additional details about the sequencing are in Section 2.2.7.

1. **CIC1-CUP1 (VIII211758-212358)**

VIII211528 F

Query: 202   AAAAGGACAAATCGAAGAACCACGAAGATGACATGGTCCCAT 247

Sbjct: 211758   AAAAGGACAAATCGAAGAACCACGAAGATGACATGGTCCCAT 211803

Query: 248   TGATGGTGTTACAGCTTTCATTTATCTACCTCAACCEPTGATGGGATCGCCAATCC 307
2. **CUP1 repeat (VII\(\text{212058-213988}\))**

**R1’**

Query: 555  
TATGTATATATCTATATATGTTT  533

Sbjct: 212058  
TATGTATATATCTATATATGTTT  212080

Query: 532  
GAAGTGTATATTAAATAAGTCATTATTGAATATTTGTCTCCTGAATAGCTTA  473

Sbjct: 212081  
GAAGTGTATATTAAATAAGTCATTATTGAATATTTGTCTCCTGAATAGCTTA  212140

Query: 472  
TACGTTTTTAGACTGATCTGTTGTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGAC  413

Sbjct: 212081  
TACGTTTTTAGACTGATCTGTTGTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGAC  212140
Sbjct: 212141 TACGTTTTAGACTGATCTGTTGTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGAC 212200
Query: 412 GGGGATAACAGCATTTTACCTTTAAAAGACGTTCTCATAATAGATTTTAGGATTAATACA 353
Sbjct: 212201 GGGATACAGCATTTTACCTTTAAAAGACGTTCTCATAATACATTTTAGGATTAATACA 212260
Query: 352 TATGCTTTTTTTTTATTCGAAATCTGGGGATTCTATACAGAGTTGTAAGTTAGGCAAAC 293
Sbjct: 212261 TATGCTTTTTTTTTCTCGAAATCTGGGGATTCTATACAGAGTTGTAAGTTAGGCAAAC 212319
Query: 292 TAGAATTTTGTAATTATATATTTATCTTTGCGACATATGGAGATACTTTATTTCCTTT 233
Sbjct: 212320 TAGAATTTTGTAATTATATATTTATCTTTGCGACATATGGAGATACTTTATTTCCTTT 212379
Query: 232 TCTTAATTATTAACGTTACCTATAAATTTAACAAGATATCTAAACCAAAATACATAAGTGT 173
Sbjct: 212380 TCTTAATTATTAACGTTACCTATAAATTTAACAAGATATCTAAACCAAAATACATAAGTGT 212439
Query: 172 ACTCAAACTGAGTAGAGGTCGATCTCTCTTCTCTTTTTAAATATAAAAACAGC 113
Sbjct: 212440 ACTCAAACTGAGTAGAGGTCGATCTCTCTTCTCTTTTTAAATATAAAAACAGC 212499

VIII212300 F
Query: 173 AAATAGTTAGTGA 187
Sbjct: 212875 TCCATATTGCTTGGTAGTCTTTTTTGCTGGAACGGTTCAGCGGAAAAGACGCATCGCTC 212934

Query: 604 TTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAAA 663

Sbjct: 212935 TTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAAA 212994

F1

Query: 292 TGTCTT 297

Sbjct: 212995 TGTCTT 213000

Query: 298 TTTCTAACTAGTAAACAAGGCTAAGATATCGCTGAATAAAAGGGTGGTGAAGTAATAAT 357

Sbjct: 213001 TTTCTAACTAGTAAACAAGGCTAAGATATCGCTGAATAAAAGGGTGGTGAAGTAATAAT 213060

Query: 358 TAAATCATCCGTATAAACCCTATACATACATATATAGGAAAAATAATACAAAAGGTGTTTTAA 417

Sbjct: 213061 TAAATCATCCGTATAAACCCTATACATACATATATAGGAAAAATAATACAAAAGGTGTTTTAA 213120

Query: 418 ATACAGATAACATACATGAACATATGCACGTATAGCGTCCCAAATGTCGGTAATGGGATCGG 477

Sbjct: 213121 ATACAGATAACATACATGAACATATGCACGTATAGCGTCCCAAATGTCGGTAATGGGATCGG 213180

Query: 478 CTTACTAATTATAAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAATACCAGAT 537

Sbjct: 213181 CTTACTAATTATAAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAATACCAGAT 213240
Sbjct: 213181 CTTACTAATTATAAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAATACCCAGAT 213240

Query: 538   TATCAGATTCCAAATCTCTTGTCAATAATTACTCTTCTTGGAAAACTTCTCTTCCATTA 597

Sbjct: 213241 TATCAGATTCCAAATCTCTTGTCAATAATTACTCTTCTTGGAAAACTTCTCTTCCATTA 213300

Query: 598   AAAAATCTGGAATCTCCTTTAAAATTTAAATAGATTCTGTTCAGTCTCACTAAGGGGAATT 657

Sbjct: 213301 AAAAATCTGGAATCTCCTTTAAAATTTAAATAGATTCTGTTCAGTCTCACTAAGGGGAATT 213360

VIII213200 F

Query: 135   TCAAGAGAACATTTTTGTTCTTCGCCGACTGAGTATAATCTGTAACATTATT 186

Sbjct: 213361 TCAAGAGAACATTTTTGTTCTTCGCCGACTGACTATAATCTGTAACATTATT 213412

Query: 187   ATTATCAGAGTTTCTCGCAAAATTTTGTTTTTTCTTGCTAAATCTCAGCATATATTTAAT 246

Sbjct: 213413 GTTATCAGAGTTTCTCGCAAAATTTTGTTTTTTCTTGCTAAATCTCAGCATATATTTAAT 213472

Query: 247   CAGATTCAAAACCTTTGGTAAACCTTTAATAGATTGAAATTTCCGTTGCTATTTCATTC 306

Sbjct: 213473 CAGATTCAAAACCTTTGGTAAACCTTTAATAGATTGAAATTTCCGTTGCTATTTCATTC 213532

Query: 307   ATCTCGTAAAAGGATACGATAATTTCTATTTTTTTAAAATTTCCAAATCTTGTACATG 366

Sbjct: 213533 ATCTCGTAAAAGGATACGATAATTTCTATTTTTTTAAAATTTCCAAATCTTGTACATG 213592
Query: 367    AATCAATAGCAATTGAACATTAATCTCCTCATTTGAAAGATTTTTGTAAAATTCGTCATA 426
Sbjct: 213593 AATCAATAGCAATTGAACATTAATCTCCTCATTTGAAAGATTTTTGTAAAATTCGTCATA 213652

Query: 427    TAATATTACTTCACAACGTTGGAAAATAGCAAATGTGATTGCTATAAAATTCTGTAAGAT 486
Sbjct: 213653 TAATATTACTTCACAACGTTGGAAAATAGCAAATGTGATTGCTATAAAATTCTGTAAGAT 213712

Query: 487    TTCAATAAAATGATTTGCGAATAAAAATTCTTTACCATTAGAATGAAAGCGATTATTGCC 546
Sbjct: 213713 TTCAATAAAATGATTTGCGAATAAAAATTCTTTACCATTAGAATGAAAGCGATTATTGCC 213772

Query: 547    GCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAAAATTAAATGTTACTGAGTAAAA 606
Sbjct: 213773 GCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAAAATTAAATGTTACTGAGTAAAA 213832

Query: 607    AATGTGCATATTAGAAATAATTTTCATCAGATCCTTTGCACATCTTTCAGAGTTCGAGGT 666
Sbjct: 213833 AATGTGCATATTAGAAATAATTTTCATCAGATCCTTTGCACATCTTTCAGAGTTCGAGGT 213892

Query: 667    CTTATTGTTGTTAGAAGAATGTTGAACTGCCATGGACAAAGAGGATTCGTTTTGAACAAA 726
Sbjct: 213893 CTTATTGTTGTTAGAAGAATGTTGAACTGCCATGGACAAAGAGGATTCGTTTTGAACAAA 213952

R1'}
3. **CUP1-RSC30 (VIII213688-214288)**

**VIII213601 F**

Query: 595  
AAGGAAAAAAAAATTTGTATAAAC 575

Sbjct: 213953 AAGGAAAAAAAAATTTGTATAAAC 213973

Query: 574  
AATGGTATTGATAAA 560

Sbjct: 213974 AATGGTATTGATAAA 213988

Query: 61  
TGA 63

Sbjct: 213688 TGA 213690

Query: 64  
TTGCTATAAAATCTGTAAGATTATTTCAATTTAATGATTGTGCAATAAAAATTTCTTTACC 123

Sbjct: 213691 TTGCTATAAAATCTGTAAGATTATTTCAATTTAATGATTGTGCAATAAAAATTTCTTTACC 213750

Query: 124  
TAGAAATGAAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAA 183

Sbjct: 213751 TAGAAATGAAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAA 213810

Query: 184  
AATTAAATGTACTGAGTAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTTTG 243

Sbjct: 213811 AATTAAATGTACTGAGTAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTTTG 213870
SNPs between YJM271 and S288c

<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in YJM271</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CIC1-CUP1 VIII211758-212358</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>211825</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>212243</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 poly(T)</td>
<td>9 poly(T)</td>
</tr>
</tbody>
</table>
Table S 2.7. Sequence analysis of the CUP1 repeats (Type 4, 1.9 kb) of YJM130.

In this table, we show genomic sequences of YJM1307 in three regions: 1) the sequences that flank the CUP1 repeats adjacent to CIC1, 2) the sequence of the CUP1 repeat, and 3) the sequences that flank the CUP1 tandem array adjacent to RCS30. The sequences of YJM1307 (denoted “Query” below) were compared in a BLAST search with

<table>
<thead>
<tr>
<th>CUP1 repeat VIII212058-213988</th>
<th>212243</th>
<th>G</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>212266-212274 10 poly(T) 9 poly(T)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>212744-212747 4 bp deletion ATTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>212751</td>
<td>G</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>213157</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>213283</td>
<td>A</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>213413</td>
<td>A</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>213513</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>213823</td>
<td>C</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>

| CUP1-RSC30 VIII213688-214288 | 213823 | C  | T  |
sequences of S288C (denoted “Sbjct”). SNPs that distinguish YJM1307 and S288C sequences are summarized at the end of the table. The CUP1 coding sequences are shown in red. The names of the primers used in the sequence analysis are shown in boldface. Additional details about the sequencing are in Section 2.2.7.

1. CIC1-CUP1 (VIII211758-212358)

VIII211528 F

Query: 202  AAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCAC 244

Sbjct: 211758 AAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCAC 211800

Query: 245  CATTGATGGTGTCAAGTTCATTTGTCTACCTTCAACAAGGGTTTGATGGAAATCGCCAA 304

Sbjct: 211801 CATTGATGGTGTCAAGTTCATTTGTCTACCTTCAACAAGGGTTTGATGGAAATCGCCAA 211860

Query: 305  TCCTTCCGAATTTGCTTCAATTCTCTCTAAACAAATTAACAATGCAAAAAAGAGATCTTC 364

Sbjct: 211861 TCCTTCCGAATTTGCTTCAATTCTCTCTAAACAAATTAACAATGCAAAAAAGAGATCTTC 211920

Query: 365  TAGCGAGCTTGAAAAAGAATCTAGCGAGTCAGAAGCTGTCAAGAAGGCTAAAAGTTAATT 424

Sbjct: 211921 TAGCGAGCTTGAAAAAGAATCTAGCGAGTCAGAAGCTGTCAAGAAGGCTAAAAGTTAATT 211980
2. **CUP1 repeat (VIII212058-213988)**

R1’

Query: 630    TATGTATATATCTATATATGTTT 537
---
Sbjct: 212058 TATGTATATATCTATATATGTTT 212080

Query: 536    GAAGTGTATATTAAAAATAAAGTCATTATTTGAATATTGGTTTCTCGTCTAAGAGCTTA 477
---
Sbjct: 212081 GAAGTGTATATTAAAAATAAAGTCATTATTTGAATATTGGTTTCTCGTCTAAGAGCTTA 212140

Query: 476    TACGTTTTAGACTGATCTGTTGTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGAC 417
---
Sbjct: 212141 TACGTTTTAGACTGATCTGTTGTACTATCCGCTTCAAATAAATAGATCATTGAAAGTGAC 212200

Query: 416    GGGGATAACAGCATTTTTACCTTTAAAGACGTTTCCTATAATAGATTATTTAGGATTAATACA 357
---
Sbjct: 212201 GGGGATAACAGCATTTTTACCTTTAAAGACGTTTCCTATAATAGATTATTTAGGATTAATACA 212260

Query: 356    TATGCTTTTTTTTTTTATTCGAAATCTGGGGATTCTATACAGAGTTGTAAGTTAGGCAAAC 297
---
Sbjct: 212201 GGGGATAACAGCATTTTTACCTTTAAAGACGTTTCCTATAATAGATTATTTAGGATTAATACA 212260
VIII212300 F

Query: 174    AAATAGTTAGATGA 187
Sbjct: 212500 AAATAGTTAGATGA 212513

Query: 188    ATATATTAAGACTATTCGTTTCATTCCCAGAGCAGCATGACTTTCTTGTTTCTTCAGA 247
Sbjct: 212514 ATATATTAAGACTATTCGTTTCATTCCCAGAGCAGCATGACTTTCTTGTTTCTTCAGA 212573

Query: 248    CTTGTTACCCGAGGGCATTTTGTCGCTGTTACACCCCGTTGGGCAGCTACATGATTT 307
Sbjct: 213302 AAAATCTGAAATCTCCTTAAATTTTAAATAGATTCTGTTCAGTTCACTAACGGGAATTT 213361

VIII213200 F

Query: 134 CAAGAGAACATTTTTGTTTCTCCTCGCCAATTTTGGTTTCTGCTAAATCTCAGCTATATATTATATATT 183
Sbjct: 213362 CAAGAGAACATTTTTGTTTCTCCTCGCCAATTTTGGTTTCTGCTAAATCTCAGCTATATATTATATATT 213411

Query: 184 TATTATCAGAGTTTCTCGCCAATTTTGGTTTCTGCTAAATCTCAGCTATATATTATATATT 213412
Sbjct: 213412 TATTATCAGAGTTTCTCGCCAATTTTGGTTTCTGCTAAATCTCAGCTATATATTATATATT 213471

Query: 244 TCAGATTCAAAACCTTGGAAACCTTTAATAGATTTGAAATTTCCGTTGCTATTCATTT 303
Sbjct: 213472 TCAGATTCAAAACCTTGGAAACCTTTAATAGATTTGAAATTTCCGTTGCTATTCATTT 213531

Query: 304 CATCTCGTTAAAAAGGATACGATAATTTCTATTTTTTTAAAATTTCCAAAATCTGTCAT 363
Sbjct: 213532 CATCTCGTTAAAAAGGATACGATAATTTCTATTTTTTTAAAATTTCCAAAATCTGTCAT 213591

Query: 364 GAATCAATAGCAATTGAACATTAATCTCCTCATTTGAAAGATTTTTGTAAAATTCGTCAT 423
Sbjct: 213592 GAATCAATAGCAATTGAACATTAATCTCCTCATTTGAAAGATTTTTGTAAAATTCGTCAT 213651

Query: 424 ATAATATTACTTCAACAGTTGGAAAAATAGCAAATGTGATTGCTATAAAATTCTGTAAGA 483
Sbjct: 213652 ATAATATTACTTCAACAGTTGGAAAAATAGCAAATGTGATTGCTATAAAATTCTGTAAGA 213711
3. **CUP1-RSC30 (VIII213688-214288)**

**VIII213601 F**

Query: 63   GATGCTATA    73

Sbjct: 213688 GATGCTATA 213698

Query: 74   AAATTCTGTAAGATTCCATAAAAATGATTTGCGAATAAAATCTTTACCATTAGAATG    133

Sbjct: 213699 AAATTCTGTAAGATTCCATAAAAATGATTTGCGAATAAAATCTTTACCATTAGAATG 213757

Query: 134   AAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAAAATTAAA    193

Sbjct: 213758 AAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGATAAAATTAAA 213817

Query: 194   TGTTACTGAGTAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTTTGCACATCT    253

Sbjct: 213818 TGTTATTGAGTAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTTTGCACATCT 213877

Query: 254   TTCAGAGTTCGAGGTCTTATTGTTGTTAGAAGAATGTTGAACTGCCATGGACAAAGAGGA    313

Sbjct: 213878 TTCAGAGTTCGAGGTCTTATTGTTGTTAGAAGAATGTTGAACTGCCATGGACAAAGAGGA 213937

Query: 314   TTCGTTTTGAAACAAAAAGGAAAAAATTGTATAAAACAAATGGTATTGATAAAAAATTTAAGT    373

Sbjct: 213938 TTCGTTTTGAAACAAAAAGGAAAAAATTGTATAAAACAAATGGTATTGATAAAAAATTTAAGT 213997
Query: 374  GTCTTTCCATTCGACTCGTTGTCATGAAAATATAAGTCTACTGTATTACTCAC 433
Sbjct: 213998 GTCTTTCCATTCGACTCGTTGTCATGAAAATATAAGTCTACTGTATTACTCAC 214057

Query: 434  GCCCATAGTCAAGGTTTCTAACAGACTTTCAATT
TTGGTTAAATTTACTGGCAAGTAGAA 493
Sbjct: 214058 GCCCATAGTCAAGGTTTCTAACAGACTTTCAATTGTTAAATTTACTGGCAAGTAGAA 214117

Query: 494  AGGAACATCTTGCAGAATATTTATCAATTTTGCTTGCGTTTCCAGTAATTTTAAATCGTT 553
Sbjct: 214118 AGGAACATCTTGCAGAATATTTATCAATTTTGCTTGCGTTTCCAGTAATTTTAAATCGTT 214177

Query: 554  AGCAATTAAAGGAATGTCGTTCGTATCAATAGAGGCAGGTATCGGAGATAGGTTTTCAGC 613
Sbjct: 214178 AGCAATTAAAGGAATGTCGTTCGTATCAATAGAGGCAGGTATCGGAGATAGGTTTTCAGC 214237

Query: 614  AGCGGGTACCATGAAT 629
Sbjct: 214238 AGCGGGTACCATGAAT 214253

VIII216603 R

Query: 325  GAAGACTGAC 316
Sbjct: 216252 GAAGACTGAC 216261

Query: 315  CTAGAAGCGAATGTCTTGAGTAATA 291
Sbjct: 216252 GAAGACTGAC 216261
Sbjct: 216262 CTAGAAGCGAATGTCTTGAGTAATA 216286
### SNPs between YJM1307 and S288c

<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in YJM1307</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CIC1-CUP1 VIII211758-212358</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>212243</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 poly(T)</td>
<td>9 poly(T)</td>
</tr>
<tr>
<td><strong>CUP1 repeat</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VIII212058-213988</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>212243</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 poly(T)</td>
<td>9 poly(T)</td>
</tr>
<tr>
<td></td>
<td>212744-212747</td>
<td>4 bp deletion</td>
<td>ATTG</td>
</tr>
<tr>
<td></td>
<td>212751</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>213157</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213283</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213393</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213413</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>213513</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213823</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><strong>CUP1-RSC30</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VIII213688-214288</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S 2.8. Sequence analysis of the *CUP1* repeats (Type 5, 1.6 kb) of YJM456.

In this table, we show genomic sequences of YJM456 in three regions: 1) the sequences that flank the *CUP1* repeats adjacent to *CIC1*, 2) the sequence of the *CUP1* repeat, and 3) the sequences that flank the *CUP1* tandem array adjacent to *RCS30*. The sequences of YJM456 (denoted “Query” below) were compared in a BLAST search with sequences of S288c (denoted “Sbject”). SNPs that distinguish YJM456 and S288c sequences are summarized at the end of the table. The *CUP1* coding sequences are shown in red. The names of the primers used in the sequence analysis are shown in boldface. Additional details about the sequencing are in Section 2.2.7.

<table>
<thead>
<tr>
<th></th>
<th>Query: 62</th>
<th>Sbject: 211275 AAT 211277</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>213823</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>214125</td>
<td>T</td>
</tr>
</tbody>
</table>

**1. CIC1-CUP1 (VIII211275-211875)**

VIII211185 F

Query: 62  | AAT 64
---------|--------
||
Sbject: 211275 AAT 211277
Query: 65  TAGATTCAGAAGGAATCAAGGTTGATGAAATCATTTGCGGGAAAGACTTAAAGACCGTTT 124
Sbjct: 211278 TAGATTCAGAAGGAATCAAGGTTGATGAAATCATTTGCGGGAAAGACTTAAAGACCGTTT 211337

Query: 125  ACAAGGCATATGAGACTAGAAACGCTTTTATATCTCAGTTTTCTTTGATTTTGGCTGACG 184
Sbjct: 211338 ACAAGGCATATGAGGCTAGAAACGCTTTTATATCTCAGTTTTCTTTGATTTTGGCTGACG 211397

Query: 185  ACAGTATAGTTACATCTTTGCCAAAACTTATGGGAGGCAAAGCCTACAACAAAGTAGAAA 244
Sbjct: 211398 ACAGTATAGTTACATCTTTGCCAAAACTTATGGGAGGCAAAGCCTACAACAAAGTAGAAA 211457

Query: 245  CTACCCCTATATCAATTAGAACACATGCAAATAAGGAATTTTCCTTGACCACTTTGACGA 304
Sbjct: 211458 CTACTCCTATATCAATTAGAACACATGCAAATAAGGAATTTTCCTTGACCACTTTGACGA 211517

Query: 305  ACAATATCAAAAAGGTATACATGAATCAGTTGCCCGTTAAACTTCCAAGAGGTACCACGT 364
Sbjct: 211518 ACAATATCAAAAAGGTATACATGAATCAGTTGCCCGTTAAACTTCCAAGAGGTACCACGT 211577

Query: 365  TGAATGTCCATTTGGGTAATTTAGAATGGTTAAGGCCAGAAGAGTTTGTAGATAACGTTG 424
Sbjct: 211578 TGAATGTCCATTTGGGTAATTTAGAATGGTTAAGGCCAGAAGAGTTTGTAGATAACGTTG 211637
Query: 425  AATTAATTCTGAACAGGTAATCAAAAGCATACCAAATCAGATCCATTTTTATCAAAACCA 484
Sbjct: 211638 AATTAATTCTGAACAGGTAATCAAAAGCATACCAAATCAGATCCATTTTTATCAAAACCA 211697

Query: 485  ATAGTCACCCGTTATTGCATTATACTATAACCAGAGCTTTGATGAAAATTGAAGCTA 544
Sbjct: 211698 ATAGTCACCCGTTATTGCATTATACTATAACCAGAGCTTTGATGAAAATTGAAGCTA 211757

Query: 545  AAAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCACCATTGATGGTGTACAAG 604
Sbjct: 211758 AAAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGTCACCATTGATGGTGTACAAG 211817

Query: 605  TTCATTTATCTACCTTCAACAAGGGTTTGATGGAAATCGCCAATCCTTCCGAATTGGG 662
Sbjct: 211818 TTCATTTGCTCTACCTTCAACAAGGGTTTGATGGAAATCGCCAATCCTTCCGAATTGGG 211875

2. **CUP1 repeat (VIII211575-213213)**

**VIII212063 R**

Query: 455  CGTTGAATGTCCATTTGGGTAAT 433
Sbjct: 211575 CGTTGAATGTCCATTTGGGTAAT 211597

Query: 432  TTGAATGGTTAAGCCAGAGATTTGAGTAACGTTAATATTTCTGAACAGGTA 373
Sbjct: 211598 TTGAATGGTTAAGCCAGAGATTTGAGTAACGTTAATATTTCTGAACAGGTA 211657
Query: 372    ATCAAAGCATACCAAATCAGATCCATTTTTATCAAAACCAATAAGTCACCCGTATTGCCA 313
Sbjct: 211658 ATCAAAGCATACCAAATCAGATCCATTTTTATCAAGACCAATAGGTCGCCCGTATTGCCA 211717

Query: 312    TTATACTATAACCAGGACGTTCTTGATGAACTTGAAGCTAAAAAGGACAAAATCGAAGAA 253
Sbjct: 211718 TTATACTATAACCAGGACGTTCTTGATGAACTTGAAGCTAAAAAGGACAAAATCGAAGAA 211777

Query: 252    ACCCACGAAGATGACATGGTCACCATTGATGGTGTACAAGTTCATTTATCTACCTTCAAC 193
Sbjct: 211778 ACCCACGAAGATGACATGGTCACCATTGATGGTGTACAAGTTCATTTGTCTACCTTCAAC 211837

Query: 192    AAGGGTTTGATGGAAATCGCCAATCCTTCCGAATTGGGTTCAATTTTCTCTAAACAAATT 133
Sbjct: 211838 AAGGGTTTGATGGAAATCGCCAATCCTTCCGAATTGGGTTCAATTTTCTCTAAACAAATT 211897

R1

Query: 450    ACAATGCAAAAAAGAGATCTTCTAGCGAGCTTGAAAAAGAATCTAGC 403
Sbjct: 211898 ACAATGCAAAAAAGAGATCTTCTAGCGAGCTTGAAAAAGAATCTAGC 211945

Query: 402    GAGTCAGAAGCTGTAAGAAGGCTAAAAGTTAATTTGTTTCCTCCTTATCTATTTTCT 343
Sbjct: 211946 GAGTCAGAAGCTGTAAGAAGGCTAAAAGTTAATTTGTTTCCTCCTTATCTATTTTCT 212005

Query: 342    CTCATTTTTTTCTTGTAAGAAAAAAATTTGAATTTCATAGAGTGCGGTGCATATGTATA 283
Sbjct: 212634  TTGGCATATTGTCATTATTTTTTGCAGCTACCACATTTGCGATTGGCACTCATGACCTTCTCATT 212693

Query: 368  TTGGAGTTAATTAATTCGCTGAACATTTTTATGTGATGATTTGATTGATATTG----TACGGT 423  
Sbjct: 212694  TTGGAGTTAATTAATTCGCTGAACATTTTTATGTGATGATTTGATTGATATTGACCATG 212753

Query: 424  TTGTTTTCTTAATATCCTATATCCGATGACTCTCATTATGACATTGCACTAAACAAGAAGATA 483
Sbjct: 212754  TTGTTTTCTTAATATCCTATATCCGATGACTCTCATTATGACATTGCACTAAACAAGAAGATA 212813

Query: 484  TTATAATGCAATTGATACAAGACAAGGAGTTATTTGCTTCTCTTTTATATGATTCTGACA 543
Sbjct: 212814  TTATAATGCAATTGATACAAGACAAGGAGTTATTTGCTTCTCTTTTATATGATTCTGACA 212873

Query: 544  ATCCATATTGGCGTTGGTAGCTTTTTTGCTGGAACGGTTCAGCGGAAAAGACGCATCGCT 603
Sbjct: 212874  ATCCATATTGGCGTTGGTAGCTTTTTTGCTGGAACGGTTCAGCGGAAAAGACGCATCGCT 212933

Query: 604  CTTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAA 663
Sbjct: 212934  CTTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAA 212993

F1

Query: 291  ATGTCTTTTT 300
Sbjct: 212994  ATGTCTTTTT 213003
3. **CUP1-RSC30 (VIII212913-213513)**

**F1**

Query: 214 CAGCGAAAAAGACGCATCGCTCTTTTTGCTTCTTAGAAGAA 253

Sbjct: 212913 CAGCGAAAAAGACGCATCGCTCTTTTTGCTTCTTAGAAGAA 212952

Query: 254 ATGCCAGCAAAAATCGACTCTCTTTTGTAGACAGCAAGGAAAAATGCTTTTTACTAGGT 313
Sbjct: 212953 ATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAAAATGTCTTTTTCTAACTAGT 213012
Query: 314    AACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAGTAATAATTAAATCATCCGT 373
Sbjct: 213013 AACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAGTAATAATTAAATCATCCGT 213072
Query: 374    ATAAACCTATACATATATGAGGAAAAATAATACAAAAGTGTTTTAAATACAGATACAT 433
Sbjct: 213073 ATAAACCTATACATATATGAGGAAAAATAATACAAAAGTGTTTTAAATACAGATACAT 213132
Query: 434    ACATGAACATATGCACGTATAGCGTCCAAATGTCGGTAATGGGATCGGCTTACTAATTAT 493
Sbjct: 213133 ACATGAACATATGCACGTATAGCGTCCAAATGTCGGTAATGGGATCGGCTTACTAATTAT 213192
Query: 494    AAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAATACCCAGATTATCAGATTCCA 553
Sbjct: 213193 AAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAATACCCAGATTATCAGATTCCA 213252
Query: 554    AATCCTTGTCAATAATTATATCTACTGCCTTTTGGGAATTCTTCTTCTTTCCATTAATAATCTGAA 613
Sbjct: 213253 AATCCTTGTCAATAATTATATCTACTGCCTTTTGGGAATTCTTCTTCTTTCCATTAATAATCTGAA 213312
Query: 614    TCTCCTAAAAATTATAGATTCTGTTCCAGTTCAACTACAAGGGAAATTCTGAAACAT 673
Sbjct: 213313 TCTCCTAAAAATTATAGATTCTGTTCCAGTTCAACTACAAGGGAAATTCTGAAACAT 213372
**VIII213200 F**

Query: 146    TTTGTCTTTCGCCGACTGACTATAATCTGTAACATTA 183

Sbjct: 213373 TTTGTCTTTCGCCGACTGACTATAATCTGTAACATTA 213410

Query: 184    TTGTTATCAGGTTCGCAAAATTTTCTTCTTGCTAAATCTCAGATATTTA 243

Sbjct: 213411 TTGTTATCAGGTTCGCAAAATTTTCTTCTTGCTAAATCTCAGATATTTA 213470

Query: 244    ATCAGATTCAAACCTTTGTTGAAACCTTTAATAGATTTGAAAT 286

Sbjct: 213471 ATCAGATTCAAACCTTTGTTGAAACCTTTAATAGATTTGAAAC 213513

**SNPs between YJM456 and S288c**

<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in YJM456</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>CIC1-CUP1 VIII211275-211875</em></td>
<td>211352</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211462</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>211534</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>211655</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>211693</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>211701</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211705</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211825</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td><strong>CUP1 repeat</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VIII211575-213213</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>211655</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>211693</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211701</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211705</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>212243</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212266-212274</td>
<td>10 T's</td>
<td>9 T's</td>
</tr>
<tr>
<td></td>
<td>212536</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212744-212747</td>
<td>4 bp deletion</td>
<td>ATTG</td>
</tr>
<tr>
<td></td>
<td>212751</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>213157</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td><strong>CUP1-RSC30</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VIII212913-213513</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>213157</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>
Table S 2.9. Sequence analysis of the CUP1 repeats (Type 5, 1.6 kb) of YJM969.

In this table, we show genomic sequences of YJM969 in three regions: 1) the sequences that flank the CUP1 repeats adjacent to CIC1, 2) the sequence of the CUP1 repeat, and 3) the sequences that flank the CUP1 tandem array adjacent to RCS30. The sequences of YJM969 (denoted “Query” below) were compared in a BLAST search with sequences of S288c (denoted “Subject”). SNPs that distinguish YJM969 and S288c sequences are summarized at the end of the table. The CUP1 coding sequences are shown in red. The names of the primers used in the sequence analysis are shown in boldface. Additional details about the sequencing are in Section 2.2.7.

1. **CIC1-CUP1 (VII211275-211875)**

<table>
<thead>
<tr>
<th></th>
<th>213283</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
</table>

VII21185 F

Query: 1  AATTAGATTCAAGGAAATCAAGGTTGATGAATCATTTCGGAAGACTAAAGACCG  60

Sbjct: 211275  AATTAGATTCAAGGAAATCAAGGTTGATGAATCATTTCGGAAGACTAAAGACCG  211334
Query:  61    TTTACAAGGCATATGAGGCTAGAAACGCTTTTATATCTCAGTTTTCTTTGATTTTGGCTG  120
Sbjct: 211335 TTTACAAGGCATATGAGGCTAGAAACGCTTTTATATCTCAGTTTTCTTTGATTTTGGCTG 211394

Query:  121    ACGACAGTATAGTTACATCTTTGCCAAAACTTATG
Sbjct: 211395 ACGACAGTATAGTTACATCTTTGCCAAAACTTATGGGAGGCAAAGCCTACAACAAAGTAG 211454

Query:  181    AAACTACCCCTATATCAATTAGAACACATGCAAATAAGGAATTTTCCTTGAC
Sbjct: 211455 AAACTACCCCTATATCAATTAGAACACATGCAAATAAGGAATTTTCCTTGACCACTTTGA 211514

Query:  241    CGAACAATATCAAAAAGGTATACATGAATCAGTTGCCCGTTAAACTTCCAAGAGGTACCA
Sbjct: 211515 CGAACAATATCAAAAAGGTATACATGAATCAGTTGCCCGTTAAACTTCCAAGAGGTACCA 211574

Query:  301    CGTGTATGCATTTGGGTATTTTTGAAATGGCCGAGAAGGTTGTAGATAACG
Sbjct: 211575 CGTGTATGCATTTGGGTATTTTTGAAATGGCCGAGAAGGTTGTAGATAACG 211634

Query:  361    TGGGTTTAATTTTCTCAGACATACGGTAAATCAAGCTACAAATCCGATATCCTCAGATTTTATCAA
Sbjct: 211635 TGGGTTTAATTTTCTCAGACATACGGTAAATCAAGCTACAAATCCGATATCCTCAGATTTTATCAA 211694

Query:  421    CCAATAAGTATGACCGCTATTTGCATTATACTGATAACGGGACTTTTACTTGAAGAATG
Sbjct: 211695 CCAATAAGTATGACCGCTATTTGCATTATACTGATAACGGGACTTTTACTTGAAGAATG 211754
Query: 481    CTAAAAAGGACAAAAATCGAAGAAACCCACGAAGATGACATGGTCACCATTGATGGTGTAC 540

Sbjct: 211755 CTAAAAAGGACAAAAATCGAAGAAACCCACGAAGATGACATGGTCACCATTGATGGTGTAC 211814

Query: 541    AAGTTCATTTATCTACCTTCAACAAGGGTTTGATGGAAATCGCCAATCCTTCCGAATTGG 600

Sbjct: 211815 AAGTTCATTTGTCTACCTTCAACAAGGGTTTGATGGAAATCGCCAATCCTTCCGAATTGG 211874

Query: 601    G

Sbjct: 211875 G

2. **CUP1 Repeat (VIII211575-213213)**

**F1**

Query: 514    CGTTGAATGTCCATTTGGGTAAT 536

Sbjct: 211575 CGTTGAATGTCCATTTGGGTAAT 211597

Query: 537    TTAGAATGGTTAAGGCCAGAAGGTTGTAGATAACGTTGAATTATTTCTGAACAGGTA 596

Sbjct: 211598 TTAGAATGGTTAAGGCCAGAAGGTTGTAGATAACGTTGAATTATTTCTGAACAGGTA 211657

Query: 597    ATCAAAGCATACCAAATCAGATTTTTATCCAATCTCAAACCAGTAGGCGCTTGTTGCA 656
Sbjct: 211658 ATCAAGCATACAAATCAGATCCATTTTTATCAAGACCAATAGGTCGCCCGTATTGCCA 211717

Query: 657    TTATACTATAACCAGGACGTTCTTGATGAACTTGAAGCTAAAAAGGACAAAATCGAAGAA 716

Sbjct: 211718 TTATACTATAACCAGGACGTTCTTGATGAACTTGAAGCTAAAAAGGACAAAATCGAAGAA 211777

Query: 717    ACCCA 721

Sbjct: 211778 ACCCA 211782

VIII212063 R

Query: 249    CGAAGATGACATGGTGCAACCATTGATGCTACAAGTTCTATTTATCTACCTTCAACAAGGG 190

Sbjct: 211783 CGAAGATGACATGGTGCAACCATTGATGCTACAAGTTCTATTTATCTACCTTCAACAAGGG 211842

Query: 189    TTTGATGGAAATCGCCAATCCTTCCGAATTGGGTTCAATTTTCTCTAAACAAATTAACAA 130

Sbjct: 211843 TTTGATGGAAATCGCCAATCCTTCCGAATTGGGTTCAATTTTCTCTAAACAAATTAACAA 211902

Query: 129    TGCAAAAAAGAGATCTTCTAGCGAGCTTGAAAAAGAATCTAGCGAGTCAGAAGCTGTCAA 70

Sbjct: 211903 TGCAAAAAAGAGATCTTCTAGCGAGCTTGAAAAAGAATCTAGCGAGTCAGAAGCTGTCAA 211962

R1'

Query: 665    GAAGGCTAAAAGTTAATTTGTTTCCTTATCTATTTTCTCTCAT 594
VIII212300 F

Query: 42   TATTTCCCTTTCTTAAATTATTAAC 65
Sbjct: 212370 TATTTCCCTTTCTTAAATTATTAAC 212393

Query: 66   GTATACCTATAAAATTAACAAAGTATCTAAACATAAGTTACTCAAATCTGAGTA 125
Sbjct: 212394 GTATACCTATAAAATTAACAAAGTATCTAAACATAAGTTACTCAAATCTGAGTA 212453

Query: 126  GAATCGTCGATTAACTTCCTTCTTTTAAAAATTAAAAACAGCAAATAGTTAGATGA 185
Sbjct: 212454 GAATCGTCGATTAACTTCCTTCTTTTAAAAATTAAAAACAGCAAATAGTTAGATGA 212513

Query: 186  ATATATTAAGACTATTCGTTTTATTTCCAGAGCAGCATGACTTCTTGGTTTCTTCAGA 245
Sbjct: 212514 ATATATTAAGACTATTCGTTTTATTTCCAGAGCAGCATGACTTCTTGGTTTCTTCAGA 212573

Query: 246  CTTGGCATTGTTCATTATTATTTTTGCAGCTACCACATTGGCATTGGCACTCATGACCTTCATT 305
Sbjct: 212574 CTTGGCATTGTTCATTATTATTTTTGCAGCTACCACATTGGCATTGGCACTCATGACCTTCATT 212633

Query: 306  TTGGCATTGGTTCATTATTATTTTTGCAGCTACCACATTGGCATTGGCACTCATGACCTTCATT 365
Sbjct: 212634 TTGGCATTGGTTCATTATTATTTTTGCAGCTACCACATTGGCATTGGCACTCATGACCTTCATT 212693
Query: 366  TTGGAAGTTAATTTAATTCCGTGAACATTTTATATGTGATGATTGATTGATTG---TACGGT 421
Sbjct: 212694  TTGGAAGTTAATTTAATTCCGTGAACATTTTATATGTGATGATTGATTGATTGATTGTCACGT 212753

Query: 422  TTGTTTTTCTTAATATCTATTTCGATGACTTCTATATGATATTGCACTAACAAGAAGATA 481
Sbjct: 212754  TTGTTTTTCTTAATATCTATTTCGATGACTTCTATATGATATTGCACTAACAAGAAGATA 212813

Query: 482  TTATAATGCATAATGACAAGGAGTTATTTGGCTTCTCTCTTATATGATTCTGACA 541
Sbjct: 212814  TTATAATGCATAATGACAAGGAGTTATTTGGCTTCTCTCTTATATGATTCTGACA 212873

Query: 542  ATCCATATTGCCGTTGGTGATCTTTTTTGGCCTGGGAACGGTTCAGCGGAAAAGACGCATCGCT 601
Sbjct: 212874  ATCCATATTGCCGTTGGTGATCTTTTTTGGCCTGGGAACGGTTCAGCGGAAAAGACGCATCGCT 212933

F1
Query: 234  CTTTTTGCTCTCTA 246
Sbjct: 212934  CTTTTTGCTCTCTA 212946

Query: 247  GAAGAAATGCCAGCAAAGAAAATCTCTCGACAGTGACTGACAGCAAAAATGTCTTTTTCTA 306
Sbjct: 212947  GAAGAAATGCCAGCAAAGAAAATCTCTCGACAGTGACTGACAGCAAAAATGTCTTTTTCTA 213006

Query: 307  ACTAGTAACAAAGCCTAAGATATCACGGCTGAATGAAATAAAGGTGGTGAAGTAAATTAAATCTC 366

3. **CUP1-RSC30 (VIII212913-213513)**

**VIII212300 F**

Query: 581  CAGCGGAAAAGACGCATCGCT 601

Sbjct: 212913 CAGCGGAAAAGACGCATCGCT 212933

**F1**

Query: 234  CTTTTTGTCTTCTA 246

Sbjct: 212934 CTTTTTGTCTTCTA 212946
SNPs between YJM969 and S288c

<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in YJM969</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>CIC1-CUP1</em> VIII211275-211875</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>211462</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>211534</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>211550</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>211655</td>
<td>G</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>--------</td>
<td>---</td>
</tr>
<tr>
<td><strong>CUP1 repeat</strong></td>
<td><strong>VIII211575-213213</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>211655</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>212243</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>212266-212274</td>
<td>4 bp deletion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>212744-212747</td>
<td>4 bp deletion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>212751</td>
<td>G</td>
</tr>
<tr>
<td><strong>CUP1-RSC30</strong></td>
<td><strong>VIII212913-213513</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>212973</td>
<td>C</td>
</tr>
</tbody>
</table>
Table S 2.10. Sequence analysis of strain (DTY3) that has a single copy of \textit{CUP1}.

To determine the sequence of the single-copy \textit{CUP1} gene and its flanking sequences, we generated three PCR fragments that had overlapping sequences as described in Supporting Data File S1. These fragments were sequenced using the primers shown in boldface below. The region between the 5' end of \textit{CIC1} and the 5' end of \textit{RSC30}, including the \textit{CUP1} gene, is shown as two continuous sequences. One sequence matches with SGD coordinates 210848 (5' end of \textit{CIC1}) to 213868 (in the 3' region of \textit{RSC30}). The other sequence matches between 215867 (3' region of \textit{RSC30}) to 217834 (the 5' end of \textit{RSC30}). The discontinuity in the comparison between the DTY3 sequences and those of S288c arises because the S288c sequence contains two copies of the 2.0 kb Type 1 \textit{CUP1} repeat instead of the single \textit{CUP1} gene present in DTY3.

1. \texttt{VIII210848-VIII213868}

\texttt{VIII210632 F}

Query: 189  \texttt{ATGGCTAAAGAGATCTCAAAAGAAATCTACGCCCTGTAAGTACACCAAGCAAAG 243}

Sbjct: 210848 \texttt{ATGGCTAAAAGAGATCTCAAAAGAAATCTACGCCCTGTAAGTACACCAAGCAAAG 210902}

Query: 244  \texttt{AAAAGAAAGGTTATTGAAAGAAATCTTTCCACAGCCATTCTAGGGAAAGAGTTATTA 303}
Query: 664  ACTTAAAGACCGTTTACAAGGCATATGAGGCTAGAAACGCTTTTATATCTCAGTT 718
Sbjct: 211323 ACTTAAAGACCGTTTACAAGGCATATGAGGCTAGAAACGCTTTTATATCTCAGTT 211377

VIII212063 R

Query: 658  TTCTTTGATTTTGGCTGACGACAGTATAGTTACATCTTTGCCAAAACTTATGGGAGGCAA 599
Sbjct: 211378 TTCTTTGATTTTGGCTGACGACAGTATAGTTACATCTTTGCCAAAACTTATGGGAGGCAA 211437

Query: 598  AGCCTACAACAAAGTAGAAACTACTCCTATATCAATTAGAACACATGCAAATAAGGAATT 539
Sbjct: 211438 AGCCTACAACAAAGTAGAAACTACTCCTATATCAATTAGAACACATGCAAATAAGGAATT 211497

Query: 538  TTCCTTGACCACTTTGACGAACAATATCAAAAAGGTTTACATGAATCAGTTGCCCGTTAA 479
Sbjct: 211498 TTCCTTGACCACTTTGACGAACAATATCAAAAAGGTTTACATGAATCAGTTGCCCGTTAA 211557

Query: 478  ACTTCCAAGAGGTACCACGTTGAATGTCCATTTGGGTAATTTAGAATGGTTAAGGCCAGA 419
Sbjct: 211558 ACTTCCAAGAGGTACCACGTTGAATGTCCATTTGGGTAATTTAGAATGGTTAAGGCCAGA 211617

Query: 418  AGAGTTTTGTAGATAAACGTTGAATTTTCTGAACAGTTAATCAAAGCATACCAAATCAG 359
Sbjct: 211618 AGAGTTTTGTAGATAAACGTTGAATTTTCTGAACAGTTAATCAAAGCATACCAAATCAG 211677

Query: 358  ATCCATTTTATCAAACCAATAAGTCGCCCGTATTGCCATTATACTATAACCAGGACGT 299
Sbjct: 211678 ATCCATTTTTATCAAGACCAATAGGTCGCCCGTATTGCACTATAACCAGGACGT 211737

Query: 298  TCTTGATGAACTTGAAGCTAAAAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGT 239
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 211738 TCTTGATGAACTTGAAGCTAAAAAGGACAAAATCGAAGAAACCCACGAAGATGACATGGT 211797

Query: 238  CACCATGATGGTGACTGGGATAAGTCATTTACCTCAACATACAGGTGGTGATGGAAATCGC 179
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 211798 CACCATGATGGTGACTGGGATAAGTCATTTACCTCAACATACAGGTGGTGATGGAAATCGC 211857

Query: 178  CAATCCTTCCGAATTGGGTTCAATTTTCTCTAAACAAATTAACAATGCAAAAAA 112
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 211858 CAATCCTTCCGAATTGGGTTCAATTTTCTCTAAACAAATTAACAATGCAAAAAA 211910

VIII211849 F

Query: 33   GAGATCTTCTAGGAGTTGAAAAAGAATCTA 64
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 211911 GAGATCTTCTAGGAGTTGAAAAAGAATCTA 211943

Query: 65   GCGAGTCAGAAGCTGTCAAGAAGACTAAAAGTTAATTTGTGTCCTCCTTATCTATCTTTT 124
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 211944 GCGAGTCAGAAGCTGTCAAGAAGCTAAAAGTTAATTTGTGTCCTCCTTATCTATCTTTT 212003

Query: 125  CTCTCATTTTTTTCTTGTGAAGAAAAATTTGAATTTCATAGAGTGCGGTGCATATGTA 184
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 212004 CTCTCATTTTTTTCTTGTGAAGAAAAATTTGAATTTCATAGAGTGCGGTGCATATGTA 212063
Query: 185
TATATCTATATATGTTTGAAGTGTATATTAAAAATAAGTCATTATTTGAATATTGGTTT

Sbjct: 212064
TATATCTATATATGTTTGAAGTGTATATTAAAAATAAGTCATTATTTGAATATTGGTTT

Query: 245
CTCGGTCTAAGAGCTTATACGTTTTAGACTGTCTCTCCCTCAAAAATAAAT

Sbjct: 212124
CTCGGTCTAAGAGCTTATACGTTTTAGACTGTCTCTCCCTCAAAAATAAAT

Query: 305
AGATCATTGAAAGTGACGGGGATAACAGCATTTTACCTTTAAAAGACGTTC

Sbjct: 212184
AGATCATTGAAAGTGACGGGGATAACAGCATTTTACCTTTAAAAGACGTTC

Query: 365
ATTTTAGGATTAATACATATGCTTTTTTTTTATTCAAAATCTGGGGATTTTATACAGAGT

Sbjct: 212244
ATTTTAGGATTAATACATATGCTTTTTTTTTATTCAAAATCTGGGGATTTTATACAGAGT

Query: 425
TGTAAGTTAGGCAAACTAGAATTTGGTAATAATATTATTCTTGGGGCGACATATGGAG

Sbjct: 212304
TGTAAGTTAGGCAAACTAGAATTTGGTAATAATATTATTCTTGGGGCGACATATGGAG

Query: 485
ATACCTTTATTTTCTTTTCTTTAATTATTACGTATACCTATAAAATTTAACAAAGTGATCTAAA

Sbjct: 212364
ATACCTTTATTTTCTTTTCTTTAATTATTACGTATACCTATAAAATTTAACAAAGTGATCTAAA

Query: 545
CAAAATACATAAGTGAACGTCGATTTAAACCTTCCTCTCTCTTCTTT

Sbjct: 212424
CAAAATACATAAGTGAACGTCGATTTAAACCTTCCTCTCTCTTCTTT
VIII212300 F

Query: 157    AAAAAATAAAAACAGCAAATTAAATGG 185
Sbjct: 212484 AAAAAATTAAACAGCAAATAGTCTTGGTCTTCAG 212512

Query: 186    AATATATTAAGCTATTCCGTGTGTTTTTCTTGGTTTTCTTCAAG 245
Sbjct: 212513 AATATATTAAGCTATTCCGTGTGTTTTTCTTGGTTTTCTTCAAG 212572

Query: 246    ACTTGTTACCGCAGGGGACTTTGTGCTGCTTACACCCCGTGGGCAGCTACATGATT 305
Sbjct: 212573 ACTTGTTACCGCAGGGGACTTTGTGCTGCTTACACCCCGTGGGCAGCTACATGATT 212632

Query: 306    TTTGGCATTGTCTTATTTTGCAGCTACCACATTGGCATTGGCACTCATGACCTTCAT 365
Sbjct: 212633 TTTGGCATTGTCTTATTTTGCAGCTACCACATTGGCATTGGCACTCATGACCTTCAT 212692

Query: 366    TTTGGGACTTTAATTCCGCTGAAACTTTATGTGATGATTGATTGATTG---TACGG 421
Sbjct: 212693 TTTGGGACTTTAATTCCGCTGAAACTTTATGTGATGATTGATTGATTGATTGATGCAG 212752

Query: 422    TTTGGGAGTTATAATTCTCCTCGATGACTTTATATGATATTGCACTAACAAGAGAT 481
Sbjct: 212753 TTTGGGAGTTATAATTCTCCTCGATGACTTTATATGATATTGCACTAACAAGAGAT 212812
Query: 482  ATTATAATGCAATTGGTACAAGACAAGGAGTTATTTGCTTCTCTTTTATATGATTCTGAC 541
Sbjct: 212813 ATTATAATGCAATTGATACAAGACAAGGAGTTATTTGCTTCTCTTTTATATGATTCTGAC 212872

Query: 542  AATCCATATTGCGTTGGTAGTCTTTTTTGCTGGAACGGTTCAGCGGAAAAGACGCATCGC 601
Sbjct: 212873 AATCCATATTGCGTTGGTAGTCTTTTTTGCTGGAACGGTTCAGCGGAAAAGACGCATCGC 212932

F1
Query: 242  TCTTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAA 301
Sbjct: 212933 TCTTTTTGCTTCTAGAAGAAATGCCAGCAAAAGAATCTCTTGACAGTGACTGACAGCAAA 212992

Query: 302  AATGTCTTTTTCTAACTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAA 361
Sbjct: 212993 AATGTCTTTTTCTAACTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAA 213052

Query: 362  GTAATAATTAAATCATCCGTATAAACCTATACACATATATGAGGAAAAAATAATACAAAA 421
Sbjct: 213053 GTAATAATTAAATCATCCGTATAAACCTATACACATATATGAGGAAAAAATAATACAAAA 213111

Query: 422  GTGTTTTAATACAGATACATACATGAACATATGCACGTATAGCGCCCAAATGTCGGTAA 481
Sbjct: 213112 GTGTTTTAATACAGATACATACATGAACATATGCACGTATAGCGCCCAAATGTCGGTAA 213171

Query: 482  TGGGATCGGCTTACTAATTATAAAATGCATCATAGAAATCGTTGAAGTTTGCCGTAGTAA 541
Sbjct: 213172 TGGGATCGGCTTACTAATTATAAAATGCATCATA
GAAATCGTTGAAGTTTGCCGTAGTAA 213231

Query: 542 TACCCAGATTATCAGATTCCAAATCCTTGTCAATAATTATACTCCTTTGGAAAACTTCTC 601

Sbjct: 213232 TACCCAGATTATCAGATTCCAAATCCTTGTCAATAATTATACTCCTTTGGA
CAACTTCTC 213291

Query: 602 TTTCCATTAAAAAATCTGAAATCTCCTTAAATTAAAAATAGATTCTGTTCAGTTCACTAA 661

Sbjct: 213292 TTTCCATTAAAAAATCTGAAATCTCCTTAAATTAAAAATAGATTCTGTTCAGTTCACTAA 213351

VIII213234 F

Query: 87 CGGGGAATTCTCAAGAGAACATTATTATTCTTGCTCAGCGGA 123

Sbjct: 213352 CGGGGAATTCTCAAGAGAACATTATTATTCTTGCTCAGCGGA 213388

Query: 124 CTGACTATAATCTGTAACATTATTATTATCAGAGTTTCTCGCAAAATTTTGTTTTTTCTT 183

Sbjct: 213389 CTGACTATAATCTGTAACATTATTATTATCAGAGTTTCTCGCAAAATTTTGTTTTTTCTT 213448

Query: 184 GCTAAATCTCAGCATATATTTATATTTAATCAGATTCTAAAAATCTGTTGAAACCTTTAATAGATT 243

Sbjct: 213449 GCTAAATCTCAGCATATATTTATATTTAATCAGATTCTAAAAATCTGTTGAAACCTTTAATAGATT 213508

Query: 244 GAAATTTCCGTTGCTATTTCAATTTCATCCCGTAAAAAGGATACGATAATTCTATTTTTTT 303

Sbjct: 213509 GAAATTTCCGTTGCTATTTCAATTTCATCCCGTAAAAAGGATACGATAATTCTATTTTTTT 213568
Query: 304    TAAAATTTCCAAAATCTTGTCATGAATCAATAGCAATTGAACAT
Sbjct: 213569 TAAAATTTCCAAAATCTTGTCATGAATCAATAGCAATTGAACAT

Query: 364    AAGATTTTTGTAAAAATTCGTCATATAATATTACTTCACAACGTTGGAAAATAGCAAATGT
Sbjct: 213629 AAGATTTTTGTAAAAATTCGTCATATAATATTACTTCACAACGTTGGAAAATAGCAAATGT

Query: 424    GATTGCTATAAAATTCTGTAAGATTTCAATAAAATGATTTGCGAATAAAAATTCTTTACC
Sbjct: 213689 GATTGCTATAAAATTCTGTAAGATTTCAATAAAATGATTTGCGAATAAAAATTCTTTACC

Query: 484    ATTAGAATGAAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGAT
Sbjct: 213749 ATTAGAATGAAAGCGATTATTGCCGCTTGAAAATGACTTTATCGACTTTATGGGGAAGAT

Query: 544    AAAATTAAATGTTACTGAGTAAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTT
Sbjct: 213809 AAAATTAAATGTTACTGAGTAAAAAATGTGCATATTAGAAATAATTTTCATCAGATCCTT

2. VIII215871-VIII217834

VIII213601 F
Query: 240  TGCA 243
Sbjct: 215867 TGCA 215870

Query: 244  CATCTTTCCAGGCTGCTTATTGTTGAGAATGCTGAACTGCCATGGACAAA 303
Sbjct: 215871 CATCTTTCCAGGCTGCTTATTGTTGAGAATGCTGAACTGCCATGGACAAA 215930

Query: 304  GAGGATCGTTTGAACAAAAAGGAAAAATTGTATAAAACATGGTATTGATAAAAATTT 363
Sbjct: 215931 GAGGATCGTTTGAACAAAAAGGAAAAATTGTATAAAACATGGTATTGATAAAAATTT 215990

Query: 364  AAAAGTGCTTTTCCAGATTCTTTCTGACTTCGTTGTCATGAAAATATAAGTCTACTGTATTA 423
Sbjct: 215991 AAAAGTGCTTTTCCAGATTCTTTCTGACTTCGTTGTCATGAAAATATAAGTCTACTGTATTA 216050

Query: 424  CTCACGCCCATAGTCAAGGTTTCTAACAGACTTTCAATTTTGGTTAAATTTACTGGCAAG 483
Sbjct: 216051 CTCACGCCCATAGTCAAGGTTTCTAACAGACTTTCAATTTTGGTTAAATTTACTGGCAAG 216110

Query: 484  TAGAAAGGAACATCTTTGACAAATATTTATCAATTTTGCTTGGTTATCCAGTAAATTTAAA 543
Sbjct: 216111 TAGAAAGGAACATCTTTGACAAATATTTATCAATTTTGCTTGGTTATCCAGTAAATTTAAA 216170

Query: 544  TCGTTAGCAATTAAAGGAATGTCGTTTGTATCAATTAGAGGCTGATACGAGATAGGTTT 603
Sbjct: 216171 TCGTTAGCAATTAAAGGAATGTCGTTTGTATCAATTAGAGGCTGATACGAGATAGGTTT 216230
Query: 604  TCAGTAGCGGTACCATGAATGAAAGACTGACCTAGAAGCGAATGTCTTGAGTAATACATT 663
Sbjct: 216231 TCAGCAGCGGTACCATGAATGAAAGACTGACCTAGAAGCGAATGTCTTGAGTAATACATT 216290

Query: 664  TTTTTCAAAAAATGGAATAACTTGAAATCTTTTATCTGGAAAGCTTAA 710
Sbjct: 216291 TTTTTCAAAAAATGGAATAACTTGAAATCTTTTATCTGGAAAGCTTAA 216337

VIII216859 R

Query: 489  GAGACAGCTCAAATCTTCG 471
Sbjct: 216338 CAGACAGCTCAAATCTTCG 216356

Query: 470  TCTAAACTTGATTTATGATCATCATGCATGTACAGAGATTCATAAAATTTTGTGATGGCA 411
Sbjct: 216357 TCTAAACTTGATTTATGATCATCATGCATGTACAGAGATTCATAAAATTTTGTGATGGCA 216416

Query: 410  ATAAATTTGATTGTCTCTTGGTCATAATATTCCAGGTTTATCAATGTGATTTGGCTTCGA 351
Sbjct: 216417 ATAAATTTGATTGTCTCTTGGTCATAATATTCCAGGTTTATCAATGTGATTTGGCTTCGA 216476

Query: 350  ATCAACCTCAAAATTATTTATGCGAAACATGGTTATTAATAGCCTGCTTATTGAAT 291
Sbjct: 216477 ATCAACCTCAAAATTATTTATGCGAAACATGGTTATTAATAGCCTGCTTATTGAAT 216536

Query: 290  AGGGTCATTTGAAATCGTTTACAAACAACAGCAGAAAGCACAGTTAATTCGCCTAGTGGTCAGT 231
Sbjct: 216537 AGGGTCATTGAATCGTTTAACAACAGTAGAAGCACAGTTAATTCGCCTAGTTTGGTCAGT 216596
Query: 230 TGAGATAAAGGAAGATC GACCGTTGAGAATGAAGGACAGGTATCACCATTTGGAAAAGAG 171
Sbjct: 216597 TGAGATAAAAGGAAGATCGACCGTTGAGAATGAAGGACAGGTATCACCATTTGGAAAAGAG 216656
Query: 170 TTATGAACAAATTCTAGTAAGAGGGTTTTGTCAAAGATTGGCAAAAATTCTCAATATTC 111
Sbjct: 216657 TTATGAACAAATTCTAGTAAGAGGGTTTTGTCAAAGATTGGCAAAAATTCTCAATATTC 216716
Query: 110 AGATTGTTAAGGGATTCAGAGTTTAGTGTCTCAATAATTATTGCTTTTGGGGAAGGTA 51
Sbjct: 216717 AGATTGTTAAGGGATTCAGAGTTTAGTGTCTCAATAATTATTGCTTTTGGGGAAGGTA 216776

VIII216894 R
Query: 89 ACTTTGCAAAACTCCATATATTGGGTATTTATCGTTATGGAGGACT 40
Sbjct: 216777 ACTTTGCAAAACTCCATATATTGGGTATTTATCGTTATGGAGGACT 216826

VIII216763 F
Query: 38 TACCTTTGAACTGATTAATATGGC 60
Sbjct: 216827 TACCTTTGAACTGATTAATATGGC 216849
Query: 481   TACCAACATTAAAGAGTATGGGTTATCTGTATAGTCAAAACTTTGAGCGTTAACCAGTT  540
                  |
Sbjct: 217270 TACCAACATTAAAGAGTATGGGTTATCTGTATAGTCAAAACTTTGAGCGTTAACCAGTT 217329

Query: 541   GTAATCGTGTGTTATATTCTCTAATCTGATCCATGGAGGGCACATCTGGACTATCGCTTC  600
                  |
Sbjct: 217330 GTAATCGTGTGTTATATTCTCTAATCTGATCCATGGAGGGCACATCTGGACTATCGCTTC 217389

Query: 601   TCACAATAGTGTGGTTCATGAACATTACTACTAACTAGTACGAGGAGCAG  660
                  |
Sbjct: 217390 TCACAATAGTGTGGTTCATGAACATTACTACTAACTAGTACGAGGAGCAG 217449

VII1218008 R

Query: 551   TATTATT  545
          |
Sbjct: 217450 TATTATT 217456

Query: 544   ATGGGTGTATTATCTGTCATGACTGGGCTTTGGGCTCCACTGCATGCAGTATCCACAGA  485
                  |
Sbjct: 217457 ATGGGTGTATTATCTGTCATGACTGGGCTTTGGGCTCCACTGCATGCAGTATCCACAGA 217516

Query: 484   GGGGTGAAAGAGAAGTTGGATGCCTCTATCGACGTTTGCATAATCGGATACTGCGTTTG  425
                  |
Sbjct: 217517 GGGGTGAAAGAGAAGTTGGATGCCTCTATCGACGTTTGCATAATCGGATACTGCGTTTG 217576
SNPs between DTY3 and S288c
<table>
<thead>
<tr>
<th>Sequenced interval</th>
<th>Coordinate(s)</th>
<th>SNP in DTY3</th>
<th>SNP in S288c</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIII210848-VIII213868</td>
<td>211693</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211701</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211825</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211967</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>211984</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>212279</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>212293</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212509</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>212536</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212556</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212744-212747</td>
<td>4 bp deletion</td>
<td>ATTG</td>
</tr>
<tr>
<td></td>
<td>212751</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>212762</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>212828</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>213097-213101</td>
<td>6 A’s</td>
<td>5 A’s</td>
</tr>
<tr>
<td></td>
<td>213157</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213283</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213413</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>213513</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>213536</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>213823</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><strong>VIII215871-VIII217834</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>215192</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>216123</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>216235</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>216563</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>216707</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>216814</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>217026</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>217047</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>217626</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>217536</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>217604</td>
<td>A</td>
<td>G</td>
</tr>
</tbody>
</table>
3 Properties of mitotic and meiotic recombination in the tandemly-repeated \textit{CUP1} gene cluster in the yeast \textit{Saccharomyces cerevisiae}

In the yeast \textit{Saccharomyces cerevisiae}, the genes encoding the metallothionein protein Cup1 are located in a tandem array on chromosome VIII. Using a diploid strain that is heterozygous for an insertion of selectable marker (\textit{URA3}) within this tandem array, and heterozygous for markers flanking the array, we measured inter-homolog recombination and intra-/sister-chromatid exchange in the \textit{CUP1} locus. The rate of intra-/sister chromatid recombination exceeded the rate of inter-homolog recombination by more than ten-fold. Loss of the Rad51 protein, required for most inter-homolog recombination, led to a relatively small reduction of recombination in the \textit{CUP1} array. Although inter-homolog mitotic recombination in the \textit{CUP1} locus is elevated relative to the average genomic region, we found that inter-homolog meiotic recombination in the array is reduced compared to most regions. Lastly, we showed that high levels of copper (previously shown to elevate \textit{CUP1} transcription) lead to a substantial elevation in rate of both inter-homolog and intra-/sister-chromatid recombination in the \textit{CUP1} array.

3.1 Introduction

Homologous recombination (HR) is an important mechanism for the repair of double-stranded DNA breaks (DSBs) in yeast (Symington \textit{et al.} 2014) and in higher
eukaryotes (Liang et al. 1998). *Saccharomyces cerevisiae* strains that lack HR are very sensitive to DNA damaging agents such as X-rays (Resnick and Martin 1976), and have a high rate of spontaneous chromosome loss (Song and Petes 2012). HR is essential for survival of mammalian cells (Helleday 2003).

Despite the importance of HR in the repair of spontaneous and induced DNA damage, there are many details concerning mitotic recombination that are still unclear. One issue is the timing of the DNA lesions that induce spontaneous mitotic recombination events. Using genetic systems that monitor inter-homolog recombination, we (Lee et al. 2009; St Charles et al. 2012) and others (Esposito 1978) concluded that most recombination events are initiated in G₁ of the cell cycle. This conclusion was unexpected since Rad52p foci (indicative of DNA damage) are much more common in the S-period and G₂ than in G₁ (Lisby et al. 2001). In addition, repair of DSBs by HR is inefficient in G₁ (Aylon et al. 2004). A simple model consistent with all of these observations is that most DSBs occur during the S-period but these DSBs are repaired by sister-chromatid recombination rather than inter-homolog recombination (Kadyk and Hartwell 1992; Lee et al. 2009). DSBs that occur in G₁ are likely not repaired in G₁, but the broken chromosome is replicated to produce two sister chromatids that are broken at the same position. Since the DSBs are at the same position, the breaks are repaired using the intact homolog as a template (Lee et al. 2009).
One set of experiments with data supporting the model described above was performed by Kadyk and Hartwell (Kadyk and Hartwell 1992). Since sister-chromatids have identical sequences, it has been challenging to determine the frequency of sister-chromatid recombination by genetic approaches. Kadyk and Hartwell (Kadyk and Hartwell 1992) developed a diploid yeast strain in which both inter-homolog and inter-sister recombination could be monitored. Inter-homolog recombination was monitored by measuring the frequency of gene conversion between leu1 heteroalleles. Sister-chromatid exchange was examined using two incomplete copies of the ADE3 gene separated by a wild-type URA3 gene, a method similar to that developed by Fasullo and Davis (Fasullo and Davis 1987). A detailed description of this system was reviewed in Section 1.2.

A variety of other systems for the analysis of sister-chromatid mitotic recombination have been developed (Jackson and Fink 1981; Szostak and Wu 1980; Arbel et al. 1999; Gonzalez-Barrera et al. 2003; Mozlin et al. 2008). Most of these assays have been done in haploid strains or utilized plasmids, preventing a direct comparison of inter-sister and inter-homolog events. Below, we describe a diploid yeast strain in which both inter-homolog and inter-sister chromatid recombination within the CUP1 locus can be monitored.
Most strains of *S. cerevisiae* have between two and twenty tandem copies of
*CUP1*, a gene encoding a copper-binding metallothionein, located on chromosome VIII
(Zhao *et al.* 2014; Strope *et al.* 2015). The degree of resistance to copper is proportional to
the number of *CUP1* genes in the tandem array (Fogel and Welch 1982). As will be
discussed in detail below, unequal sister-chromatid recombination events within
tandem arrays can be monitored using single-copy genes inserted within the array
(Petes 1980; Szostak and Wu 1980). In addition, inter-homolog recombination between
*CUP1* arrays can be measured by using selectable drug resistance markers flanking the
array. Thus, a direct comparison of inter-sister and inter-homolog events can be made.

We found that inter-sister recombination events are >10-fold more frequent than
inter-homolog events. Most of these inter-sister events do not reflect unequal sister-
chromatid recombination but likely occur by single-strand annealing, inter-sister gene
conversion, or DNA polymerase slippage. We also found that, although Rad51 is
required for most types of homologous recombination, absence of this protein has a
relatively small effect on recombination within the *CUP1* array.

Transcription of the *CUP1* genes is induced by high levels of copper (Karin *et al.*
1984). We demonstrated that mitotic recombination within the *CUP1* array is strongly
induced by transcription. We were also able to compare the properties of mitotic and
meiotic recombination within the *CUP1* locus. In meiosis, unlike mitosis, recombination
occurs less frequently than for an average genomic interval, and meiotic recombination is not elevated by high levels of copper.

3.2 Material and methods

3.2.1 Yeast medium

Standard media were used (Guthrie and Fink 1991) unless noted below. Copper-containing medium was made by adding an aqueous solution of copper sulfate to SD-complete medium. HYG+CAN medium contained 120 mg/liter L-canavanine sulfate and 300 mg/liter hygromycin B in SD-arg omission medium. Before adding hygromycin B, we adjusted the liquid medium to a pH of 5.5 by addition of sodium hydroxide. The medium was then filter-sterilized, and mixed with autoclave-sterilized agar. HYG and GEN medium was made by supplementing YPD medium with either 300 mg/liter hygromycin B or 200 mg/liter geneticin. CAN medium was made by adding 120 mg/liter L-canavanine sulfate to SD-arg omission medium. 5-fluoro-orotate (5-FOA)-containing medium was SD-complete with the addition of 40 milligrams of uracil and 1 gram of 5-FOA per liter.

3.2.2 Yeast strains and plasmids

The genotypes and additional details of the constructions of all strains used in this study are shown in Table S 3.1. All haploid strains are isogenic derivatives of strains S1/S288c (Engel et al. 2014), W303-1A (Thomas and Rothstein 1989a), or YJM789 (Wei et
A complete list of primers used in strain construction or PCR analysis is provided in Table S 3.2. Plasmids used in our work are listed in Table S 3.3. Haploid strains with inserted genetic markers and various gene deletions were constructed by transformation. All insertions and deletions were confirmed by PCR. In addition, the rad52Δ::natMX4 strains YZ43, YZ44, and YZ113, and the rad1Δ::natMX4 strains YZ49, YZ50, and YZ116 were confirmed by their sensitivity to 50 J/m² of UV radiation. The cup2Δ::natMX4 strains YZ51, YZ52 and YZ117 were confirmed by their sensitivity to growth in medium containing 0.2 mM copper sulfate.

The strain YZ18 was constructed by integrating the URA3 marker into one of the CUP1 repeats in the strain YZ15. Genomic DNA from the wild-type strain S1 was amplified using primers VIII214177::URA3 F and VIII214177::URA3 R. Genomic DNA of independent YZ18 isolates was isolated and digested with HindIII, which cuts once within the URA3 marker, and in the regions flanking the CUP1 cluster. In YZ18-10, the strain used in subsequent constructions, the URA3 insertion was in the middle of the 18-repeat CUP1 array.

Previously, we found that the CUP1 clusters in strains YJM789 and JSC19-1 contained seven copies of the 1.2 kb CUP1 repeats (Zhao et al. 2014). The strain YZ17 was derived by selecting JSC19-1 derivatives with enhanced copper resistance; a concentration of 0.4 mM copper in Synthetic dextrose complete (SDC) medium was used.
for this selection. Southern analysis verified that YZ17 had an expanded CUP1 cluster with about 22 repeats.

The 2-micron-based plasmid p425-GPD-CUP1 was constructed by amplifying genomic DNA of the strain S1 using the primers HindIII-CUP1 stop ligate and CUP1 start-BamHI ligate. The resulting fragment was treated with HindIII and BamHI, and inserted into HindIII-BamHI-treated p426-GPD (ATCC 87359). In this plasmid, the CUP1 gene is regulated by the strong constitutive GPD promoter (alternative name of the TDH3 promoter) (Mumberg et al. 1995). This plasmid also contains a LEU2 marker for the maintenance and selection of the plasmid.

The haploid strain YZ26 was constructed by switching the mating type of the strain JSC10-1 (St Charles and Petes 2013), a W303-1A derivative with the genotype \( \text{MAT}^a \text{leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100\Delta::natMX RAD5} \). The plasmid pSR109 (pGAL-HO) (Herskowitz and Jensen 1991) was transformed into JSC10-1 in omission medium lacking uracil and containing 2% raffinose and 0.1% glucose. HO expression was transiently induced by incubating cells for 5 hours in medium containing 2% galactose and 0.1% glucose. After this treatment, cells were plated on rich growth medium (YPD) and allowed to form colonies. The mating type of each colony was scored using “lawns” of tester strains (YJM789 and YJM790, Table S 3.1). Colonies were also tested on omission medium lacking uracil to confirm the loss of pSR109.
3.2.3 Measurements of recombination rates

The diploid strains YZ103 and YZ104, and related derivatives are sensitive to 5-FOA and canavanine (Can), and resistant to the drugs hygromycin (Hyg) and Geneticin (Gen). As described in the Results, we measured inter-homolog recombination rates in the *CEN8-hphMX4*, *hphMX4-URA3*, *URA3-CAN1/kanMX4* genetic intervals by measuring the rate of derivatives that are 5-FOA<sup>R</sup>Hyg<sup>S</sup>Gen<sup>S</sup>Can<sup>R</sup>, 5-FOA<sup>R</sup>Hyg<sup>R</sup>Gen<sup>S</sup>Can<sup>R</sup>, and 5-FOA<sup>S</sup>Hyg<sup>R</sup>Gen<sup>R</sup>Can<sup>R</sup>, respectively. Derivatives resulting from intra-/sister-chromatid events have the phenotype 5-FOA<sup>R</sup>Hyg<sup>R</sup>Gen<sup>R</sup>Can<sup>S</sup>. To determine the rates of these events shown in Table 3.1, for each strain, we allowed cells to form single colonies at 30° C. on rich growth medium (YPD). Each colony was suspended in water, and various dilutions were plated on SD complete, 5-FOA-containing medium, and HYG+CAN medium. Colonies formed on 5-FOA plates were replica-plated to medium containing hygromycin, geneticin, and canavanine, and the numbers of colonies on each plate with the phenotypes described above were counted. For each experiment, we calculated these numbers for about 20 independent colonies, and we performed experiments using two derivatives of each strain. For each strain, these values were converted to rate estimates using the method of the median (Lea and Coulson 1949).

In experiments to examine the effect of copper on mitotic recombination in strains YZ118 and YZ103 (Table 2), experiments were done somewhat differently. YZ118
contains a plasmid with a CUP1 gene regulated by the GPD promoter as well as a LEU2 gene. Colonies were grown on plates containing SD-complete medium with 1.4 mM copper sulfate. These colonies were suspended in water and plated to three types of medium (all containing 0.2 mM copper sulfate): SD-complete, 5-FOA, and HYG + CAN-containing medium. Rate estimates were performed as described above. Copper was included in the diagnostic media to force retention of the plasmid. For reasons that are not clear, the YZ118 strain fails to form colonies on medium lacking leucine and containing copper. For the control experiments in the absence of added copper, cells of YZ118 were grown on SD omission medium lacking leucine; subsequent analysis was done using the diagnostic plates with 0.2 mM copper sulfate. For the YZ103 strain, cells were allowed to form colonies on YPD medium or SD-complete medium with 1.4 mM copper sulfate. The media for diagnosis were the same as those used in the analysis of YZ118.

Although we expect that derivatives of YZ103 and YZ104 that have the phenotype 5-FOA<sup>R</sup> Hym<sup>S</sup> Gen<sup>S</sup> Can<sup>S</sup> primarily reflect crossovers between CEN8 and hphMX4, loss of the homolog that carries the hphMX4, URA3, and CAN1/kanMX markers could result in a strain with the same phenotype. In the diploid YZ104, which has 55,000 heterozygous single-nucleotide polymorphisms (St Charles et al. 2012), loss of one homolog would result in loss of heterozygosity (LOH) for all polymorphisms on
chromosome VIII. Consequently, in 20 independent 5-FOA<sup>R</sup>Hyg<sup>S</sup>Gen<sup>S</sup>Can<sup>R</sup> derivatives of YZ104, we checked for LOH for a polymorphism located within 3 kb of CEN8. Using primers VIII 108799-F and VIII 109229-R (Table S 3.2), we amplified genomic DNA from the 20 derivatives. The resulting 430 bp fragment from the YJM789-derived homolog has an Asel site that the W303-1A-derived fragment lacks. Thus, when the PCR products from diploids heterozygous for this polymorphism were treated with Asel, we found three fragments of sizes 430, 260, and 170 bp. In strains in which the W303-1A-derived homolog was lost, we found only the 260 and 170 bp fragments. Of the 20 isolates, 18 retained heterozygosity for the centromere-linked polymorphism, demonstrating that most 5-FOA<sup>R</sup>Hyg<sup>S</sup>Gen<sup>S</sup>Can<sup>R</sup> derivatives are a consequence of crossovers rather than chromosome loss.

### 3.2.4 Southern Analysis

Genomic DNA from different strains was isolated in plugs of low-melt agarose as described previously (McCulley and Petes 2010). The samples were treated overnight at 37° C with a diagnostic restriction enzyme (HindIII or EcoRI). The resulting DNA fragments were separated by CHEF (contour-clamped homogeneous electric field) gel electrophoresis (McCulley and Petes 2010), followed by transfer of the separated fragments to nylon membranes. The hybridization probes were prepared using digoxygenin (DIG)-dUTP labeling (Roche); details of the hybridization conditions were
described previously (Zhao et al. 2014). DIG-labeled CUP1 probes that hybridize to both 2 kb and 1.2 kb CUP1 repeats were synthesized by DIG-PCR labeling using primers CUP1 amp5-2 and CUP1 amp3 (Table S 3.2). Probes that hybridize specifically to the 2 kb CUP1 repeats were DIG-labeled using primers CUP1 W303 spec amp5 and CUP1 W303 spec amp3 (Table S 3.2). The sizes of the tandem arrays were estimated relative to DNA size standard (Bioline DNA Hyperladders I and VI).

3.2.5 Statistical analysis

We performed comparisons using chi-square, Fisher exact, or the Mann-Whitney non-parametric tests, and the VassarStat website (http://vassarstats.net). 95% confidence limits on rate estimates were calculated as described in Yin and Petes (Yin and Petes 2014). For comparisons of rate measurements in different strains or strains grown in different conditions, we calculated rates for each culture using the method of the median (Lea and Coulson 1949). We then compared these rates using the Mann-Whitney non-parametric test.

3.2.6 Meiotic analysis

For meiotic studies, we constructed a derivative of YZ103 that was heterozygous for a trp1 mutation (details in Table S 3.1). This diploid (MD692) was sporulated at room temperature using three types of medium: standard sporulation medium (Guthrie and
Fink 1991), sporulation medium with 1.4 mM copper sulfate, and sporulation medium with 5 mM nicotinamide. Tetrads were dissected and analyzed using standard methods.

3.3 Results

3.3.1 Experimental rationale

One goal of this project was to develop diploid yeast strains that would allow a comparison of mitotic recombination events between homologs with events that involve intra-chromatid or sister-chromatid interactions. Two diploids (YZ103 and YZ104) with similar arrangements of markers on chromosome VIII were constructed (Figure 3.1A). In both diploids, one homolog (derived from the haploid strain W303-1A) contained the \textit{hphMX4} gene (resulting in hygromycin resistance, Hyg\textsuperscript{R}) at the centromere-proximal end of the \textit{CUP1} tandem array, and a cassette containing the \textit{CAN1-kanMX4} genes (resulting in sensitivity to canavanine, and resistance to geneticin, Can\textsuperscript{S} Gen\textsuperscript{R}) at the centromere-distal end of the cluster. In addition, this homolog has a copy of \textit{URA3} integrated within the \textit{CUP1} tandem array. In this homolog, there are eighteen copies of a 2 kb \textit{CUP1} repeat, and the \textit{URA3} gene is integrated in the middle of the cluster (shown by Southern analysis described in Materials and Methods). In the diploid YZ103, the other homolog is also derived from W303-1A but lacks the three insertions. In the diploid YZ104, the other homolog is derived from the haploid YJM789 (Wei et al. 2007). The \textit{CUP1} locus on this chromosome has 22 copies of a 1.2 kb repeat (Zhao et al. 2014).
Both YZ103 and YZ104 are homozygous for can1 and ura3 mutations at their normal loci on chromosomes V.

Inter-homolog exchanges on chromosome VIII were monitored in three intervals as shown in Figure 3.1B-D. Crossovers in the CEN8-hphMX4 interval (Figure 3.1B) can produce isolates that are Ura\(^-\) (and, therefore, resistant to 5-fluoro-orotate, 5-FOA\(^R\)) Hyg\(^S\) Gen\(^S\) Can\(^R\). A crossover in the hphMX4-URA3 interval (Figure 3.1C) results in 5-FOA\(^R\) Hyg\(^R\) Gen\(^S\) Can\(^R\) isolates, and a crossover in the URA3-CAN1/kanMX4 interval (Figure 3.1D) results in 5-FOA\(^S\) Hyg\(^R\) Gen\(^S\) Can\(^R\) isolates. We determined the frequencies of each of these phenotypic classes in multiple (>20) independent cultures, and converted these frequencies to rates using the method of the median as described in Section 3.2.3.

Although we expected that most of the derivatives with the phenotype 5-FOA\(^R\) Hyg\(^S\) Gen\(^S\) Can\(^R\) would represent crossovers (Figure 3.1B), loss of the homolog with the hphMX4, URA3, and CAN1/kanMX4 markers would produce the same phenotype. The diploid YZ104 is heterozygous for about 55,000 SNPs distributed throughout the genome. To determine what fraction of the 5-FOA\(^R\) Hyg\(^S\) Gen\(^S\) Can\(^R\) derivatives were crossovers, we examined 20 independent 5-FOA\(^R\) Hyg\(^S\) Can\(^R\) Gen\(^S\) derivatives for loss of heterozygosity (LOH) for a single-nucleotide polymorphism (SNP) located near CEN8.
Figure 3.1. Diploid strains (YZ103 and YZ104) used for the detection of inter-homolog recombination between *CEN8* and *CUP1*, and within the *CUP1* array.
A. Arrangement of markers on chromosome VIII in YZ103 and YZ104. The centromeres are shown as circles. In both diploids, one homolog has the marker hphMX4 inserted in single-copy sequences at the centromere-proximal end of the CUP1 array, an insertion of URA3 in the middle of the CUP1 cluster, and a cassette with CAN1/kanMX4 markers at the centromere-distal end of the array. Both homologs in YZ103 are derived from the haploid W303-1A and have 2.0 kb CUP1 repeats (Zhao et al., 2014); although there are 18 repeats in the CUP1 cluster of W303-1A, only ten are shown, each in brackets. In YZ104, one homolog (shown in black) is derived from W303-1A, and the other (shown in blue) is derived from YJM789. Only 16 of the approximately 22 1.2 kb CUP1 repeats of this homolog are shown. Both strains are phenotypically Hyg\(^R\) FOA\(^S\) Can\(^R\) Gen\(^S\).

B. Detection of a crossover between CEN8 and hphMX4. After a crossover in this genetic interval (shown as an X), followed by co-segregation of chromatids 1 and 3 into one daughter cell and 2 and 4 in the other, one daughter cell (boxed in thin lines) would have the same phenotype as the parental strain, whereas the other daughter (boxed in thick lines) would be phenotypically distinct (Hyg\(^S\) FOA\(^R\) Can\(^R\) Gen\(^S\)).

C. Detection of a crossover between hphMX4 and URA3. A crossover in this interval would produce one daughter cell with the same phenotype as the starting diploid and a second daughter with the unique phenotype Hyg\(^R\) FOA\(^R\) Can\(^R\) Gen\(^S\).

D. Detection of a crossover between URA3 and CAN1/kanMX4. One daughter cell would have the same phenotype as the starting strain, and the second daughter would have the phenotype Hyg\(^R\) FOA\(^S\) Can\(^R\) Gen\(^S\).

This diagnosis was done using primers flanking the polymorphism, followed by digestion of the product using an enzyme that cuts one allele but not the other; the details of this method are described in Section 3.2. In the wild-type strain YZ104, 18 of 20
of the isolates retained heterozygosity for the centromere-linked marker, indicating that these derivatives reflected crossovers rather than chromosome loss.

In addition to those recombination events involving the two homologs, we observed a high frequency of events in which the *URA3* gene was lost and the flanking *hphMX4* and *CAN1/kanMX4* markers were retained. Strains of this genotype have the phenotype 5-FOA<sup>R</sup> Hyg<sup>R</sup> Gen<sup>R</sup> Can<sup>S</sup>. A variety of different recombination mechanisms can generate this genotype including intrachromatid “pop-out” recombination (Figure 3.2A), unequal sister-chromatid recombination (Figure 3.2B), inter-sister chromatid gene conversion (Figure 3.2C), and inter-homolog gene conversion (Figure 3.2D). In addition, the phenotype could reflect single-strand annealing (Figure 3.3A and 3.3B) or DNA polymerase realignment/slippage (Figure 3.3C). By the experiments described below, we eliminated some of these possibilities. In the discussion below, all of the mechanisms except inter-homolog gene conversion will be described as intra-/inter-sister chromatid events.
Figure 3.2. Mitotic recombination events leading to loss of *URA3* marker and retention of the flanking *hphMX4* and *CAN1/kanMX4* markers.

The chromosomes are shown following DNA replication, and the two homologs are shown in different colors; in this depiction, one homolog has 2 kb *CUP1* repeats and the other 1.2 kb repeats. *CUP1* repeats are indicated by brackets. The chromosomes of 5-FOA<sup>R</sup> daughter cells are outlined by thick lines.

A. Intra-chromatid "pop-out" recombination. A crossover occurs within a chromatid, producing a shorter *CUP1* array and a plasmid with the *URA3* gene and three *CUP1* repeats. The cell with the chromosomes outlined with thick lines would be 5-FOA<sup>R</sup>. Since each *CUP1* repeat has an *ARS* element, the *URA3*-containing plasmid would be capable of autonomous replication. It is shown segregated into the daughter cell that also contains an integrated *URA3* gene.

B. Unequal sister-chromatid crossover. As a consequence of this event, the 5-FOA<sup>R</sup> daughter cell would contain a shorter *CUP1* array, and the Ura<sup>+</sup> daughter cell would contain a longer array with two *URA3* insertions.

C. Inter-sister chromatid gene conversion. A DSB (shown as a red arrow) occurs near the *URA3* insertion in one chromatid, and is repaired using the sister chromatid as a template. The net result of this event would be a loss of *URA3* and one or more *CUP1* repeats in one daughter cell with no alteration in the second daughter.

D. Inter-homolog gene conversion. As in Figure 3.2C, this event is initiated with a DSB near or within the *URA3* insertion. The repair template, however, is a chromatid of the other homolog instead of the sister chromatid. Associated with the loss of *URA3* and the loss of some of the 2.0 kb *CUP1* repeats, one would expect insertion of one or more 1.2 kb *CUP1* repeats derived from the other homolog.
Table 3.1. Inter-homolog and intra-/sister-chromatid mitotic recombination rates in wild-type and mutant yeast strains.¹

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Relevant genotype (background)</th>
<th>CEN8-hphMX4 inter-homolog</th>
<th>hphMX4-URA3 inter-homolog</th>
<th>URA3-kanMX4/CAN1 inter-homolog</th>
<th>hphMX4-kanMX4/CAN1 inter-homolog²</th>
<th>Intra-/sister chromatid</th>
</tr>
</thead>
<tbody>
<tr>
<td>YZ103</td>
<td>Wild-type (W303-1A x W303-1A)</td>
<td>8.2 x 10⁻⁶(6.4-14) [1]</td>
<td>3.3 x 10⁻⁶(3.0-3.8) [1]</td>
<td>5.1 x 10⁻⁶(3.9-5.8) [1]</td>
<td>8.4 x 10⁻⁶ [1]</td>
<td>8.3 x 10⁻⁵ (6.5-12) [1]</td>
</tr>
<tr>
<td>YZ104</td>
<td>Wild-type (W303-1A x YJM789)</td>
<td>4.6 x 10⁻⁶(3.0-6.3) [0.56]</td>
<td>2.8 x 10⁻⁶(2.3-3.1) [0.85]</td>
<td>3.2 x 10⁻⁶(3.0-3.5) [0.62]</td>
<td>6 x 10⁻⁶ [0.71]</td>
<td>1.5 x 10⁻⁴ (1-1.9) [1.8]</td>
</tr>
<tr>
<td>YZ113</td>
<td>rad52/rad52 (W303-1A x W303-1A)</td>
<td>9.7 x 10⁻⁷ [0.12]</td>
<td>4 x 10⁻⁷ [0.12]</td>
<td>4.8 x 10⁻⁷ [0.094]</td>
<td>8.8 x 10⁻⁷ [0.10]</td>
<td>5.2 x 10⁻⁸ [0.62]</td>
</tr>
<tr>
<td>YZ114</td>
<td>rad51/rad51 (W303-1A x W303-1A)</td>
<td>4.5 x 10⁻⁸(3.4-5.2) [0.0055]</td>
<td>4.5 x 10⁻⁷(3.2-6.4) [0.14]</td>
<td>1.0 x 10⁻⁶(0.9-1.4) [0.20]</td>
<td>1.5 x 10⁻⁶ [0.18]</td>
<td>1.9 x 10⁻⁵ (1.7-2.2) [0.23]</td>
</tr>
<tr>
<td>YZ115</td>
<td>mre11/mre11 (W303-1A x W303-1A)</td>
<td>6.2 x 10⁻⁵(5.7-7.4) [7.6]</td>
<td>2.4 x 10⁻⁵(2.0-3.2) [7.3]</td>
<td>1.5 x 10⁻⁵(1.2-1.9) [2.9]</td>
<td>3.9 x 10⁻⁵ [4.6]</td>
<td>1.1 x 10⁻⁴ (0.9-1.4) [1.3]</td>
</tr>
<tr>
<td>YZ116</td>
<td>rad1/rad1 (W303-1A x W303-1A)</td>
<td>1.8 x 10⁻⁵(1.3-2.3) [2.2]</td>
<td>3.8 x 10⁻⁴(3.4-4.1) [1.2]</td>
<td>4.0 x 10⁻⁴(3.7-5.1) [0.78]</td>
<td>7.8 x 10⁻⁵ [0.93]</td>
<td>1.5 x 10⁻⁴ (1.3-1.8) [1.8]</td>
</tr>
</tbody>
</table>

¹The rates represent the number of events per cell division as determined by fluctuation analysis as described in the text. Numbers in parentheses represent 95% confidence limits, and bold-faced numbers in brackets are the rates divided by the rate of the wild-type strain YZ103.

²The rates in this column are the sum of the rates in the fourth and fifth columns. These rates reflect the frequency of inter-homolog recombination within the entire CUP1 array.
Table 3.2. Rates of inter-homolog and intra-/sister-chromatid in cells grown in medium without added copper, and in medium with 1.4 mM copper.¹

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant genotype</th>
<th>Condition</th>
<th>CEN8-hphMX4 inter-homolog</th>
<th>hphMX4-URA3 inter-homolog</th>
<th>URA3-kanMX4/CAN1 inter-homolog</th>
<th>hphMX4-kanMX4/CAN1 inter-homolog</th>
<th>Intra-/sister-chromatid</th>
</tr>
</thead>
<tbody>
<tr>
<td>YZ103</td>
<td>Wild-type</td>
<td>No added copper</td>
<td>$1.1 \times 10^{-5}$ (0.9-1.7) [1]</td>
<td>$5.1 \times 10^{-6}$ (4.4-7.0) [1]</td>
<td>$6.0 \times 10^{-6}$ (4.4-8.1) [1]</td>
<td>$1.1 \times 10^{-5}$ [1]</td>
<td>$1.0 \times 10^{-4}$ (0.7-1.2) [1]</td>
</tr>
<tr>
<td>YZ103</td>
<td>Wild-type</td>
<td>1.4 mM copper sulfate</td>
<td>$1.3 \times 10^{-5}$ (1.1-1.6) [1.2]</td>
<td>$8.6 \times 10^{-5}$ (7.4-11) [17]</td>
<td>$8.8 \times 10^{-5}$ (7.5-11) [15]</td>
<td>$1.7 \times 10^{-4}$ [15]</td>
<td>$2.3 \times 10^{-3}$ (2.0-2.7) [23]</td>
</tr>
<tr>
<td>YZ117</td>
<td>cup2/cup2</td>
<td>No added copper</td>
<td>$8.8 \times 10^{-6}$ (6.2-13) [0.8]</td>
<td>$3.6 \times 10^{-6}$ (2.3-4.8) [0.71]</td>
<td>$4.2 \times 10^{-6}$ (3.1-5.7) [0.7]</td>
<td>$7.8 \times 10^{-6}$ [0.71]</td>
<td>$1.1 \times 10^{-4}$ (0.7-1.4) [1.1]</td>
</tr>
<tr>
<td>YZ118</td>
<td>cup2/cup2 + p425-GPD-CUP1</td>
<td>No added copper</td>
<td>$6.8 \times 10^{-6}$ (4.6-10) [0.62]</td>
<td>$4.3 \times 10^{-6}$ (2.7-4.9) [0.84]</td>
<td>$4.1 \times 10^{-6}$ (3.1-5.2) [0.68]</td>
<td>$8.4 \times 10^{-6}$ [0.76]</td>
<td>$6.7 \times 10^{-5}$ (5.0-7.7) [0.67]</td>
</tr>
<tr>
<td>YZ118</td>
<td>cup2/cup2 + p425-GPD-CUP1</td>
<td>1.4 mM copper sulfate</td>
<td>$1.3 \times 10^{-5}$ (0.9-2.5) [1.2]</td>
<td>$1.2 \times 10^{-5}$ (1.1-1.6) [2.4]</td>
<td>$9.9 \times 10^{-6}$ (6.9-11) [1.7]</td>
<td>$2.2 \times 10^{-5}$ [2]</td>
<td>$8.4 \times 10^{-5}$ (7.3-10) [0.84]</td>
</tr>
</tbody>
</table>

¹Numbers in parentheses after the rate estimates are the 95% confidence limits. Numbers in brackets in bold are the rates divided by the rate observed in YZ103 grown in medium without added copper. The number of cells per culture for all strains except YZ117 was determined using SD-complete medium containing 0.2 mM copper; the number of cells per culture for YZ117 was determined in the absence of copper similar to the analysis of the strains in Table 3.1.
By fluctuation analysis (described in Section 3.2), we measured the rates of mitotic recombination in the CEN8-hphMX4, the hphMX4-URA3, and the URA3-kanMX4/CAN1 intervals as $8.2 \times 10^{-6}$, $3.3 \times 10^{-6}$, and $5.1 \times 10^{-6}$ per cell division, respectively (Table 3.1). Since the relative sizes of the CEN8-hphMX4 and CUP1 arrays are about 107 kb and 30 kb, respectively, we expect the recombination rate in the CEN8-hphMX4 interval to be about four-fold higher than the rate for the CUP1 cluster (hphMX4-URA3 added to URA3-kanMX4/CAN1). The observed rate for the CEN8-hphMX4 interval was about the same as the rate within the cluster, indicating that the cluster is about four-fold “hotter” for inter-homolog exchange than the CEN8-CUP1 interval. A similar comparison of the frequency of inter-homolog recombination within the CUP1 locus to the frequency of recombination on the 1 Mb right arm of chromosome IV (St. Charles and Petes, 2013) indicates that the CUP1 cluster is about eight-fold elevated for inter-homolog mitotic recombination compared to the genomic average.

We also examined inter-homolog crossovers in YZ104 in which the CUP1 clusters on the two homolog have different repeat lengths (Zhao et al., 2014). The inter-homolog recombination rates in YZ103 and YZ104 are similar for all intervals, although the rates are slightly lower (25-50% lower) in YZ104. This difference may reflect the sequence heterogeneity between the interacting homologs or small differences in the enzymes that catalyze mitotic recombination in the hybrid background.
Figure 3.3. Loss of the *URA3* insertion by single-strand annealing (SSA) or DNA polymerase slippage.

In this diagram, we show the chromatid as double-stranded DNA molecules with the 3’ ends marked by arrows. *CUP1* repeats are shown as thick arrows and the *URA3* insertion is indicated by a rectangle.

A. SSA resulting in a single-stranded DNA loop. Following a DSB (shown with a thick red arrow), the broken ends are processed by 5’ to 3’ degradation. The single-stranded *CUP1* repeat at the terminus at the right broken end anneals with repeats on the top strand of the left broken end. The resulting intermediate has a single-stranded loop containing the *URA3* gene. This loop could be removed by cellular endonucleases (shown with thin red arrows) or the resulting DNA molecule could be replicated without removing the loop. The latter event would result in one daughter molecule that retains the *URA3* insertion but has fewer *CUP1* repeats, and a second molecule that loses the insertion and several *CUP1* repeats.

B. SSA resulting in a single-stranded flap. This mechanism is very similar to that of Fig. 3A except for the amount of processing and the pattern of annealing. The resulting intermediate has a single-stranded flap that contains both the *URA3* gene and one repeat. Removal of the flap would result in a shorter array that lacks the *URA3* insertion.
C. DNA polymerase slippage. During DNA replication, the primer strand (the top strand) dissociates from the template and reassociates beyond the position of the *URA3* insertion. The resulting intermediate has a single-stranded loop with the *URA3* marker and several *CUP1* repeats. Removal of the loop or replication of the intermediate would result in a daughter molecule that lacks the *URA3* insertion and several *CUP1* repeats.

3.3.3 Rates of intra-/sister-chromatid events in YZ103 and YZ104

Most of the events that produce 5-FOA<sup>R</sup> Hyg<sup>R</sup> Gen<sup>R</sup> Can<sup>S</sup> derivatives are likely to involve intra- or inter-sister chromatid recombination (Figures 3.2 and 3.3). The rates with which the 5-FOA<sup>R</sup> Hyg<sup>R</sup> Gen<sup>R</sup> Can<sup>S</sup> genotype is produced are $8.3 \times 10^5$/division in YZ103 and $1.5 \times 10^4$/division in YZ104. These rates are 10-fold (YZ103) and 25-fold (YZ104) higher than the rates of inter-homolog recombination within the *CUP1* cluster for each strain. These observations are consistent with the conclusion of Kadyk and Hartwell (Kadyk and hartwell 1992) that sister-chromatids are preferred over homologs as the substrate for the repair of DNA lesions.

3.3.4 Rate of inter-homolog gene conversion in YZ104

As described above, derivatives of YZ103 or YZ104 with the 5-FOA<sup>R</sup> Hyg<sup>R</sup> Gen<sup>R</sup> Can<sup>S</sup> could reflect various types of intra-/inter-sister chromatid events or inter-homolog gene conversion unassociated with crossing-over (Figure 3.2D). Such events can be detected in the YZ104 diploid. Loss of the *URA3* gene, which is inserted in the 2 kb *CUP1* repeats characteristic of the W303-1A-derived homolog, by gene conversion will
result in insertion of one or more 1.2 kb \textit{CUP1} repeats from the YJM789-derived homolog. To restrict our analysis to the W303-1A-derived homolog, we sporulated strains with the 5-FOA\textsuperscript{R} Hyg\textsuperscript{R} Gen\textsuperscript{R} Can\textsuperscript{S} phenotype, and identified spores that were Hyg\textsuperscript{R} Gen\textsuperscript{R} Can\textsuperscript{S}.

We did two types of PCR analysis to determine if there were 1.2 kb \textit{CUP1} repeats within the cluster of 2.0 kb repeats (Figure 3.4). The primers F1 and R1' hybridize to both the 2 kb and 1.2 kb, and PCR amplification with these primers leads to 2 and 1.2 kb products, respectively. Among 42 independent haploid isolates examined, none contained a mixture of both types of repeats. To confirm this result, we used primers 4038F (specific for the 2 kb repeats) and 2172R (specific for the 1.2 kb repeats). Arrays with a mixture of these two types of repeats should produce a 1.2 kb PCR product. None of the 42 isolates had such a product. We conclude that inter-homolog gene conversion is not responsible for a significant fraction of the 5-FOA\textsuperscript{R} Hyg\textsuperscript{R} Gen\textsuperscript{R} Can\textsuperscript{S} strains.
Figure 3.4. Detection of inter-homolog gene conversion by PCR.

*CUP1* repeats are outlined by thick black lines. Sequences in common between the 2.0 and 1.2 kb *CUP1* repeats are in blue, sequences unique to the 2.0 kb repeat are in white, and those unique to the 1.2 kb repeat are in red. Primers F1 and R1' amplify both the 2.0 and 1.2 kb repeats whereas the 4038F and 2172R primers are specific for the 2.0 and 1.2 kb repeats, respectively.

A. 2.0 kb-repeat array of W303-1A. Only six repeats of the 18-repeat array are shown.

B. 1.2 kb-repeat array of YJM789. Only 11 repeats of the 22-repeat array are shown.

C. An array containing both 2.0 and 1.2 kb repeats. As an example of an inter-homolog conversion, we show a 2.0 kb array with an insertion of two 1.2 kb repeats. A PCR reaction with primers 4038F and 2172R would produce a unique 1.2 kb band.
3.3.5 Loss of CUP1 repeats in YZ104 strains that have intra-/sister-chromatid recombination events

In most of the models shown in Figures 3.2 and 3.3, one would expect loss of one or more CUP1 repeats in the 5-FOA<sup>+</sup> Hyg<sup>R</sup> Gen<sup>R</sup> Can<sup>S</sup> strains that have lost the URA3 gene. To determine the number of CUP1 repeats that were lost, we examined the sizes of the CUP1 clusters on the W303-1A-derived homolog in 33 independent 5-FOA<sup>+</sup> Hyg<sup>R</sup> Gen<sup>R</sup> Can<sup>S</sup> derivatives of YZ104. Genomic DNA was digested with EcoRI (which does not cut the repeats), and Southern analysis was conducted with a probe that hybridized specifically to the 2 kb CUP1 repeats (details in Materials and Methods). These data are summarized in Figure 3.5. The starting CUP1 copy number is 18, and copy numbers in the resulting 5-FOA<sup>+</sup> Hyg<sup>R</sup> Gen<sup>R</sup> Can<sup>S</sup> daughter cells range from 1 to 16. The median number of retained repeats was eight.

3.3.6 Stimulation of mitotic recombination within the CUP1 cluster by addition of copper to the growth medium

The transcription of the CUP1 genes is greatly elevated (about 20-fold) by addition of 1 mM CuSO<sub>4</sub> to the medium (Keller et al. 2005). To determine if copper also leads to an elevated mitotic recombination rate, we grew the YZ103 strain in medium containing 1.4 mM CuSO<sub>4</sub>. The amount of CuSO<sub>4</sub> in synthetic medium unsupplemented with extra copper is about 0.25 µM

The starting strain had 18 repeats on this homolog. We isolated 33 independent spontaneous 5-FOA<sup>®</sup>Hyg<sup>®</sup>Gen<sup>®</sup>Can<sup>®</sup>derivatives. CUP1 copy numbers were determined by measuring the size of CUP1 cluster by Southern analysis of EcoRI-digested genomic DNA. A hybridization probe specific for the W303-1A type of CUP1 repeat was used.

Rates of recombination in various genetic intervals were measured as described above. The crossover rate in the CEN8-hphMX4 interval was similar in the absence or presence of added copper (Table 3.2). In contrast, the rate of inter-homolog crossovers in the CUP1 array was elevated about 15-fold in high-copper medium. Similarly, the rate of intra-/inter-sister chromatid events was elevated 23-fold by growth in high-copper medium.
Since the stimulation of mitotic recombination by copper primarily elevates recombination in the \textit{CUP1} locus and since transcription of the \textit{CUP1} repeats is induced by copper, we tested whether the recombinogenic effect of copper was dependent on \textit{CUP1} transcription. We constructed a strain that was isogenic with YZ103 (YZ117) but was also homozygous for a deletion of \textit{CUP2}. The Cup2p binds upstream of \textit{CUP1} to elevate its transcription in response to high levels of copper (Buchman \textit{et al.} 1989; Huibregtse \textit{et al.} 1989; Szczypka and Thiele 1989; Welch \textit{et al.} 1989). The \textit{cup2} diploid YZ117 was unable to grow in concentrations of copper of 0.2 mM whereas the isogenic YZ103 strain could grow in concentrations ten-fold higher. In the absence of copper, YZ117 has similar crossover rates and intra-/inter-sister chromatid events to YZ103 (Table 3.2).

Since YZ117 is very sensitive to copper, in order to determine whether determine whether the recombinogenic effects of copper required copper-stimulated transcription of the \textit{CUP1} array, we constructed a derivative of YZ117 (YZ118) that contained a high-copy number plasmid (p426-GPD-CUP1) in which transcription of the plasmid-borne \textit{CUP1} gene was regulated by the GPD (glceraldehyde-3-phosphate dehydrogenase) promoter; details of this construction are described in Section 3.2. The plasmid also contained a wild-type \textit{LEU2} gene. The YZ118 strain grew in medium with high levels of copper sulfate (1.4 mM) but failed to grow in medium lacking leucine with high levels of copper. Therefore, in all experiments involving YZ118, to force retention of the plasmid,
all diagnostic media contained 0.2 mM copper sulfate. In the experiments summarized in Table 3.2, the control strain YZ103 was grown and analyzed under the same conditions as YZ118.

As shown in Table 3.2, in YZ118, unlike YZ103, the presence of high levels of copper has little effect on inter-homolog or on intra-/sister-chromatid recombination in the CUP1 array. This result strongly argues that the elevated rate of CUP1 recombination induced by copper in YZ103 is a consequence of elevated levels of CUP1 transcription acting in cis.

3.3.7 Loss of the URA3 gene from the CUP1 tandem array in YZ103 is not a consequence of unequal sister-chromatid exchange.

As shown in Figure 3.2, although most of the events resulting in loss of URA3 in 5-FOA\(^8\) Hyg\(^8\) Gen\(^5\) Can\(^5\) derivatives also lead to loss of CUP1 repeats in one daughter cell, the models differ in predictions about repeats in the other daughter cell. More specifically, if URA3 loss is a consequence of unequal sister-chromatid crossovers, the Ura\(^+\) daughter cells would be expected to have additional CUP1 repeats and two copies of URA3 within one array (Figure 3.2B). In addition, “pop-out” recombination might produce a Ura\(^+\) daughter cell with one URA3 gene within the CUP1 array and one plasmid-borne URA3 gene (Figure 3.2A). Such plasmids would be capable of autonomous replication because the CUP1 repeats of W303-1A contain ARS elements (Zhao et al. 2014). For most of the other models shown in Figures 3.2 and 3.3, in the Ura\(^+\)
sector, we would expect only one copy of the *URA3* gene in the array and no change in the number of *CUP1* repeats per array.

To determine copy number for the *URA3* and *CUP1* genes in the Ura⁺ daughter cell that was produced in the same cell division as the Ura⁻ cell, we used the haploid strain YZ18-10, one of the parental strains of YZ103. This strain has *URA3* integrated into the middle of the *CUP1* array, and has the *hphMX4* and *kanMX4/CAN1* genes flanking the array. The strain was grown overnight in medium lacking copper, and then grown for six hours in medium with 1.4 mM copper sulfate in order to induce recombination. The cells were then plated on rich growth medium, and allowed to form colonies. The colonies were replica-plated to medium lacking uracil. About 1% of the colonies had approximately equal-sized Ura⁺/Ura⁻ sectors; in such colonies, the Ura⁺ sector should represent the daughter cell produced in the same event as the Ura⁻ daughter.

Genomic DNA derived from the Ura⁺ sector was isolated from 17 independent sectored colonies. The DNA was treated with *Hind*III, a restriction enzyme that cuts within the *URA3* gene and sequences that flank the array, but does not cut within the *CUP1* repeat. The resulting fragments were examined by Southern analysis using *CUP1* sequences as the hybridization probe. If the tandem array in the Ura⁺ sector had two non-tandem *URA3* genes, as expected for an unequal crossover event, we would expect to detect three hybridizing bands. All samples had only two. This observation indicates
that loss of URA3 is not a consequence of unequal sister-strand crossing over. In
addition, since a plasmid containing URA3 and CUP1 would also produce a third
hybridizing fragment, our results suggest that loss of the URA3 insertion is not likely the
consequence of “pop-out” recombination.

3.3.8 Analysis of the genetic regulation of mitotic recombination
within the CUP1 array

There are a large number of proteins required for wild-type levels of mitotic
recombination in yeast (Symington et al. 2014). We constructed diploids that were
isogenic with YZ103 except for homozygous mutations affecting various genes involved
in recombination, and determined the effect of these mutations on inter-homolog and
intra-/sister-chromatid exchange.

The diploid YZ114 lacks Rad51p, a RecA-related protein required to form a
filament on single-stranded DNA required for strand invasion. Rad51p is required for
inter-homolog recombination in most assays (Symington et al. 2014; Paques and Haber
1999). As expected, the rad51 mutation greatly (>100-fold) reduced the frequency of
crossovers in the CEN8- hphMX4 interval (Table 3.1). Inter-homolog recombination
within the CUP1 cluster and intra-/sister-chromatid events were much less affected with
reductions of four- to eight-fold compared to the wild-type strain.

The strain YZ113 was homozygous for the rad52 mutation. The Rad52 protein,
which aids in the loading of Rad51p onto single-stranded DNA coated with RPA, is
required for most types of HR involving non-repeated sequences (Symington et al. 2014).
Loss of Rad52p reduced the rate of inter-homolog events at the CUP1 locus about ten-fold, and had no significant effect on the rate of intra-/sister-chromatid events.

Surprisingly, the rate of inter-homolog crossovers in the CEN8-hphMX4 interval was also decreased only ten-fold, considerably less than the reduction observed for the rad51/rad51 strain. This observation is likely to be an artifact. As described previously, there are two ways of producing 5-FOA\(^R\) Hyg\(^S\) Can\(^R\) Gen\(^S\) derivatives of YZ113 and related strains: a crossover in the CEN8-hphMX4 interval, and loss of the homolog containing the hphMX4, URA3, and CAN1-kanMX4 markers. Since rad52/rad52 diploids have a very high rate of chromosome loss (Song and Petes 2012), it is likely that the many of the 5-FOA\(^R\) Hyg\(^S\) Can\(^R\) Gen\(^S\) derivatives of YZ113 are a consequence of chromosome loss rather than mitotic recombination. Thus, we conclude that the rad52 mutation reduces crossovers in the CEN8-hphMX4 interval at least ten-fold.

Because of our uncertainty in measurements of the rates of recombination in the rad52/rad52 diploid, we are currently repeating these experiments. A preliminary analysis of these data suggests that the rates of recombination in the CEN8-hphMX4 interval and the hphMX4-CAN1-kanMX4 interval are 5- to 10-fold lower than those shown in Table 3.1. In addition, the rate of intra-/sister-chromatid recombination in these newer experiments is about 10-fold lower than observed in our first experiments. These discrepancies will be resolved before submitting our manuscript describing this research.
We also examined recombination in the homozygous \textit{rad1/rad1} strain YZ116. The Rad1 endonuclease, in addition to its role in nucleotide excision repair (Tomkinson \textit{et al.} 1993), is involved in processing the single-stranded branches that are intermediates in the single-stranded annealing pathway of HR (Ivanov \textit{et al.} 1996). Loss of Rad1 had no effect on interhomolog recombination within the \textit{CUP1} array, and resulted in a small (two-fold) increase in recombination between \textit{CEN8} and the \textit{hphMX4} marker. Intra-/sister-chromatid recombination was also relatively unaffected (two-fold increase).

Lastly, we measured recombination rates in YZ115, an \textit{mre11/mre11} strain. Mre11p, acting in a complex with Rad50p and Xrs2p (the MRX complex), has multiple cellular roles. The complex is involved in telomere length regulation (Kironmai and Muniyappa 1997), processing of broken DNA molecules (Mimitou and Symington 2009), and promoting post-replicative cohesion assembly (Unal \textit{et al.} 2004). In assays of inter-homolog mitotic recombination, mutants lacking any of the MRX proteins are hyper-Rec, and it has been suggested that the absence of damage-induced cohesin might reduce the efficiency of sister-chromatid recombination, and elevate the frequency of inter-homolog recombination (Symington \textit{et al.} 2014). We found an approximately five-fold elevation in inter-homolog events, both in the \textit{CEN8-hphMX4} interval and within the \textit{CUP1} array (Table 3.1). In contrast, the rate of intra-/sister-chromatid events was elevated only slightly (1.3-fold).
In summary, both the inter-homolog and intra-/sister-chromatid recombination events within the CUP1 cluster are unusual relative to most HR events involving non-repeated genes. More specifically, although some reduction in recombination rates is observed in the rad51 strain, the degree of this reduction is small. In contrast, inter-homolog recombination in the CEN8-hphMX4 interval is strongly reduced as expected from previous studies. We will discuss these results as well as our interpretation of the results obtained in rad1 and mre11 strains in detail below.

3.3.9 Meiotic recombination in the CUP1 array

We examined meiotic recombination within the CUP1 array in the diploid MD692 (isogenic with YZ103 except for being heterozygous for the centromere-linked trp1 marker). In the hphMX4-CAN1/kanMX4 interval (the markers that span the CUP1 locus), we observed 355 parental ditype (PD), 1 non-parental ditype (NPD), and 44 tetratype (T) tetrads. Applying the standard mapping equation (Perkins, 1949) to these data, we calculate that the CUP1 locus is about 6 cM. Based on the physical length of the yeast genome (excluding ribosomal DNA) and the number of crossovers per meiosis (Mancera et al. 2008), we calculate that the genomic average is about 2.7 kb/cM. Since the CUP1 locus is 30 kb in MD692, the expected map distance for the CUP1 locus is about 11 cM. Thus, there is significant (p<0.001, chi-square test) suppression of meiotic recombination within the CUP1 locus. These results are consistent with the low frequency of Spo11-induced DSBs at the CUP1 locus (Pan et al. 2011), although a direct
comparison is difficult because the strain used for mapping DSBs (SK1) contained only one copy of CUP1.

Because copper induces mitotic recombination, we also sporulated MD692 in medium containing 1.4 mM copper sulfate to determine whether there was a similar stimulation of meiotic exchange. The numbers of PD, NPD, and T tetrads for the hphMX4-CAN1/kanMX4 interval were 214, 0, and 20, respectively. The calculated map distance for the CUP1 array is 4.2 cM, and the numbers of tetrads in each category are not significantly (p>0.05; chi-square test) different from that observed in cells sporulated in the absence of copper.

Meiotic recombination within the tandemly repeated ribosomal RNA gene cluster is strongly suppressed (Petes and Botstein 1977), and this suppression is dependent on the Sir2p histone deacetylase (Gottlieb and Esposito 1989). Nicotinamide, a negative regulator of Sir2p, relieves the meiotic suppression of meiotic recombination in the rDNA (Bitterman et al. 2002). Consequently, we sporulated MD692 in medium containing 5 mM nicotinamide. The numbers of PD, NPD, and T tetrads were 377, 1, and 60, respectively; these numbers are not significantly different from those observed for MD692 sporulated under standard conditions (p>0.05; chi-square test).

In addition to crossovers in the CUP1 array, we found a small number of tetrads in which the URA3 marker showed non-Mendelian segregation. The numbers and types of such tetrads for each sporulation condition were: standard sporulation condition (1,
3:1; 4, 1:3; total of 400 tetrads), high-copper medium (1, 1:3; 234 tetrads), and medium plus nicotinamide (1, 1:3; 438 tetrads). The 1:3 class of tetrads could represent a gene conversion events involving sister-chromatids or homologs, or an unequal sister-chromatid crossover.
### 3.4 Discussion

Our current study reveals several important features of mitotic and meiotic recombination of tandemly-repeated genes including: 1) events resulting in lost of a selectable marker inserted within the \textit{CUP1} array usually occur by mechanisms other than inter-homolog exchange, 2) unequal sister-chromatid recombination does not contribute substantially to marker loss; our results argue that loss of the marker within the \textit{CUP1} locus occurs by single-strand annealing and/or sister-chromatid gene conversion, 3) inter-homolog \textit{CUP1} mitotic recombination is elevated relative to average genomic intervals, and is induced by high levels of \textit{CUP1} transcription, and 4) in contrast to mitotic recombination, meiotic recombination within the \textit{CUP1} array occurs at relatively low levels.

#### 3.4.1 Relative frequency of inter-homolog and intra-/sister-chromatid events

There are three methods that have been used to measure intra-/sister-chromatid recombination events in yeast: recombination between two tandem heteroallelic genes (usually done in haploid strains) (Klein and Petes 1981; Jackson and Fink 1981), loss of a marker located within a tandem array of repeats (Petes 1980; Szostak and Wu 1980), and formation of a dimeric circle from a monomeric circular chromosome or plasmid (Game et al. 1989; Gonzalez-Barrera et al. 2003). The most direct comparison of inter-homolog and intra-/sister-chromatid events was done by Kadyk and Hartwell (Kadyk and Hartwell 1992) in a diploid with heteroalleles on different homologs to monitor inter-
homolog exchange and different tandem heteroalleles to monitor intra-/sister-chromatid recombination. Our system has the advantage of measuring both types of recombination at the same genetic locus. Using X-rays to stimulate recombination, Kadyk and Hartwell concluded that intra-/sister-chromatids were the preferred substrate (relative to the homolog) for DSBs generated in G2 recombination by a factor of about twenty. Our finding that spontaneous mitotic intra-/sister-chromatid events are about 10-20-fold more frequent than inter-homolog events at the CUP1 locus is in good agreement with this conclusion. Since our assay detects only those intra-/sister-chromatid events that are associated with loss of the URA3 gene, our measurement is an underestimate of the true rate of intra-/sister chromatid exchanges.

3.4.2 Mechanism of intra-/sister-chromatid events

The loss of the URA3 gene within the CUP1 array was not usually a consequence of inter-homolog recombination. Figures 3.2 and 3.3 show some of the mechanisms that could result in loss of URA3 by intra-chromatid or sister-chromatid interactions. In studies in which a selectable marker was integrated in the ribosomal RNA genes, loss of the marker was shown to occur by both unequal crossing over (Szostak and Wu 1980; Zamb and Petes 1981) and gene conversion between sister chromatids (Gangloff et al. 1996). Since marker loss from the ribosomal DNA (rDNA) is independent of Rad52 (Zamb and Petes 1981; Ozenberger and Roeder 1991), it was argued that marker loss sometimes reflected single-strand annealing rather than “classic” crossovers or gene
conversions (Ozenberger and Roeder 1991). In addition, since extra-chromosomal plasmids containing ribosomal RNA genes are observed in some yeast strains (Larionov et al. 1980), “pop-out” recombination also occurs.

Similar to the ribosomal RNA genes, marker loss from the CUP1 array is likely to occur through more than one mechanism. From our analysis of the Ura+ sector of Ura-/Ura+ sectored colonies, however, we argue that “pop-out” (Figure 3.2A) and reciprocal unequal crossover events (Figure 3.2B) are not common modes of marker loss, since we failed to detect the expected products of these classes. This observation has the caveat that we examined sectored colonies generated by the hyper-Rec conditions of high copper in the medium. We can also rule out inter-homolog gene conversion unassociated with crossovers (Figure 3.2D) as a common mechanism for marker loss. As described above, none of 42 independent derivatives that lost URA3 had undergone gene conversion with the other homolog.

A model that we cannot exclude, but appears unlikely, is DNA polymerase template switching (Figure 3.3C). In this mechanism, during DNA replication, the replicating strand dissociates from one repeat and reassociates with another. Although such events have been observed in E. coli as RecA-independent interactions between repeats, these exchanges usually involve short (<200 bp) repeats (Lovett et al. 2002); above that threshold, RecA-dependent exchanges predominate. Second, deletions formed by template switching might be expected to involve preferentially adjacent or
close repeats. The median number of repeats associated with loss of the URA3 marker was eight. Lastly, as discussed below, high levels of copper in the medium elevate inter-homolog recombination and intra-/sister-chromatid events to approximately the same extent. If these types of events proceed by fundamentally different mechanisms, this congruence is surprising.

The two mechanisms that fit our observations best are unequal inter-sister chromatid gene conversion (Figure 3.2C) and single-strand annealing (SSA, Figure 3.3A, 3.3B). Both of these mechanisms have been invoked to explain loss of markers within the rDNA (Ozenberger and Roeder 1991; Gangloff et al. 1996). In current models of recombination, gene conversion unassociated with crossovers involves synthesis-dependent strand annealing (SDSA). One of the ends resulting from a DSB invades an intact template, and the invading end primes DNA synthesis. Subsequently, this end disengages from the template and reattaches to the other broken end. If the URA3 marker is within the region of heteroduplex, the resulting single-stranded loop could be removed by cellular endonucleases; heteroduplexes that include large heterologies and the processing of large single-stranded loops have been detected during meiotic recombination in yeast (Kearney et al. 2001). An alternative possibility is that the URA3 marker is removed by double-stranded degradation of broken ends, leading to a gap. Although broken ends are usually processed by 5’ to 3’ processing of only one strand, degradation of both strands has been observed producing a double-stranded gap
(Zierhut and Diffley 2008); such gaps are readily repaired by HR (Orr-Weaver et al. 1981).

In general, recombination events requiring strand invasion would be expected to require the RecA-homolog Rad51p and the Rad51p-mediator/strand-exchange protein Rad52p (Symington et al. 2014). In our experiments, although loss of these proteins substantially reduced the rate of recombination in the interval between CEN8 and the CUP1 locus, intra-/sister-chromatid events were reduced only two- to five-fold in the rad51 strain (Table 3.1). One interpretation of this result is that loss of the URA3 marker occurs by SSA (discussed below) rather than inter-sister conversion. A complication of this interpretation is that inter-homolog recombination within the CUP1 locus is also reduced by only five- to ten-fold in the rad51 strain, and such events are not explicable by SSA. It is possible that the absence of Rad51p greatly reduces the probability of a successful strand invasion, but this reduction is partly balanced by templates that contain many sites at which strand invasion can occur. More specifically, for single-copy sequences, there is only one position on an intact template molecule that has homology to a broken end. In the CUP1 array, however, each repeat in the intact template has homology to the DNA ends generated by DSBs in the CUP1 array.

The second plausible model consistent with our results is SSA (Figure 3.3A, 3.3B). Most studies of SSA are performed in strains in which direct repeats of heteroallelic genes flank an intervening marker. Often, researchers select cells in which
gene conversion produces a wild-type allele from the heteroalleles, and the presence or absence of the intervening marker is scored. In most such experiments, gene conversion events are reduced in rad51 strains, whereas deletions of the intervening marker occur at near wild-type levels (Prado et al. 2003; Symington et al. 2014). The frequency of deletion events involving direct repeats in rad52 strains relative to the wild-type frequency varies from 1% (1 kb repeats) to nearly 100% (>10 kb) (Paques and Haber 1999). In a previous study of HO-induced DSBs within the CUP1 locus, Ozenberger and Roeder (Ozenberger and Roeder 1991) found efficient Rad52p-independent repair. Thus, our observation that marker loss is relatively unaffected by the rad51 mutation is roughly in agreement with previous studies of SSA.

“Classical” SSA usually requires the Rad1p for efficient removal of the single-stranded branches generated when heteroduplexes are formed between two repeats that contain intervening heterology (Sugawara et al. 1997), although exceptions have been observed (Nag et al. 2005; McDonald and Rothstein 1994). In our study, loss of the URA3 marker was not substantially affected by the rad1 mutation (Table 3.1). Depending on the extent of processing of the broken ends and how the broken ends reanneal, SSA at the CUP1 locus may be associated with single-stranded loops (Figure 3.3A) or with single-stranded flaps (Figure 3.3B). Although the role of Rad1p in the removal of flaps has been demonstrated, the removal of loops by Rad1p has not been examined.
Although it is possible that the *URA3* marker is excised from the intermediates shown in Figure 3.3 by cellular endonucleases, two other mechanisms could be responsible for marker loss. First, replication of the intermediates shown in Figure 3.3 without removal of the loops would result in loss of *URA3* and *CUP1* repeats from one of the resulting DNA molecules. Second, as discussed above, double-stranded degradation of the broken molecules could result in loss of *URA3* and *CUP1* repeats unassociated with formation of single-stranded loops or branches.

By either the inter-sister chromatid conversion or SSA models, in order to have deletions involving large numbers of repeats, the broken DNA ends would have to be extensively processed by excising one or both strands of the broken ends. In HO-induced recombination in the *CUP1* array, Ozenberger and Roeder (Ozenberger and Roeder 1991) found a similar distribution of deletions to one observed in our study (Figure 3.5). Previously, we showed that some inter-homolog mitotic gene conversions that include a heterozygous Ty element were very long with a median size of >50 kb (Yim et al. 2014). Thus, the yeast cell has the capacity to generate conversion events and single-stranded annealing events that are large enough to explain our deletions.

The *mre11* mutation elevated inter-homolog recombination and no substantial effect on intra-/sister-chromatid exchanges. Since Mre11p is required for DNA damage-induced sister chromatid cohesion (Unal et al. 2004), it is possible that *mre11* mutants channel the repair of DNA damage from the sister chromatid to the homolog. By this
model, we expect to observe reduced sister chromatid exchange in addition to elevated inter-homolog recombination. Although no such reduction was detectable in our experiments, even a relatively small change in intra-/sister-homolog recombination (which is much more frequent than inter-homolog exchange) could significantly elevate inter-homolog recombination. Alternatively, the *mre11* mutants may have an elevated level of DSBs in addition to altering the ratio of inter-homolog to intra-/sister-chromatid recombination in favor of inter-homolog exchange.

### 3.4.3 Stimulation of *CUP1* recombination by transcription

The phenomenon of transcription-associated recombination (TAR) was reported in yeast about thirty years ago (Voelkel-Meiman *et al.* 1987; Thomas and Rothstein 1989a; Kim *et al.* 2013; Aguilera and Garcia-Muse 2013). TAR likely results from a number of different mechanisms including head-on collisions between the replication fork and the transcriptional machinery, DNA secondary structures (for example, hairpins) formed in single-stranded DNA resulting from transcription, accumulated supercoils in transcribed DNA, R-loop accumulation, damage induced on single-stranded DNA resulting from R-loop formation and/or the transcription “bubble”, and positioning of highly-transcribed genes at the nuclear pores (Aguilera and Garcia-Muse 2013). Since each *CUP1* repeat has a bidirectional origin and a promoter, there will be head-on collisions between the replication and transcription proteins. However, the other models are also plausible.
An important point is that the copper-induced elevations in the levels of recombination are similar for the inter-homolog CUP1 genes (15-fold) and the intra-/sister-chromatid CUP1 events (23-fold) (Table 3.2). The simplest interpretation of this result is that high levels of transcription elevate the frequency of recombinogenic DNA lesions, but do not affect the ratio of inter-homolog to intra-/sister-chromatid events. This observation is more difficult to explain if inter-homolog events are a consequence of HR whereas intra-/sister-chromatid events reflect DNA polymerase slippage.

The high levels of recombination induced by copper may be evolutionarily advantageous. Growth of yeast cells in high copper selects for derivatives with longer CUP1 arrays (Fogel and Welch 1982). Since these derivatives likely arise by unequal crossovers, an elevation in the rate of recombination will result in an increased rate of formation of these copper-resistant strains.

### 3.4.4 Meiotic recombination

In contrast to the hyper-Rec phenotype associated with mitotic recombination in the CUP1 cluster, meiotic recombination is about two-fold lower than the average genomic interval. Unlike mitotic exchange, this level is unaffected by high levels of copper in the medium. Inter-homolog meiotic recombination within the ribosomal RNA genes is much more strongly suppressed (>100-fold; Petes and Botstein 1977), although unequal meiotic recombination occurs frequently, in about 10% of unselected tetrads (Petes 1980). This suppression is dependent on Sir2p (Gottlieb and Esposito 1989). Our
observation that nicotinamide-containing sporulation medium does not elevate inter-homolog mitotic recombination in the CUP1 array suggests a different mechanism for the modest reduction observed in our study. It is possible that the CUP1 sequences are simply not a good substrate for Spo11p-mediated DSB formation.

In our analysis, we observed ten 1:3- and one 3:1-segregation events for the URA3 marker in a total of 1072 tetrads. This frequency of about 1% is consistent with the low level of meiotic crossovers. In contrast, in a study in which meiotic recombination events in the CUP1 array were detected by Southern analysis of the length of the array, Welch et al. (Welch et al. 1990, 1991) found 10-20% of the tetrads contained one or more spores with an altered CUP1 array. In most of these events, only a single spore of the four had an altered array, indicating that the alteration likely reflected SSA or sister-chromatid gene conversion. This result argues that our analysis detects only a small fraction of the meiotic intra-/sister-chromatid events, although our study should detect all of the inter-homolog exchanges. In summary, the CUP1 array has a two-fold reduced rate of meiotic inter-homolog exchange but a higher proportion of sister-strand recombination than observed for most meiotic events for which there is a strong inter-homolog bias (Hunter 2015).

**3.5 Summary**

In this chapter, we described a diploid strain that can be used to monitor inter-homolog and intra-/sister-chromatid recombination events at the CUP1 locus, and
demonstrated that spontaneous intra-/sister-chromatid events are at least ten-fold more frequent than inter-homolog events. The rates of both types of recombination are elevated by high concentrations of copper in the medium, and this effect is dependent on the Cup2 transcription factor. In contrast to mitotic recombination, meiotic recombination is suppressed at the CUP1 locus and is unaffected by high concentrations of copper.

### 3.6 Supplementary tables

**Table S 3.1. Strain genotypes and constructions.**

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Genotype</th>
<th>Strain construction (reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>MATa gal2</td>
<td>S288c background, Engel et al. 2014</td>
</tr>
<tr>
<td>W303-1A</td>
<td>MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100 RAD5</td>
<td>Thomas and Rothstein 1989</td>
</tr>
<tr>
<td>YJM789</td>
<td>MATα lys2 gal2 ho::hisG</td>
<td>Wei et al. 2007</td>
</tr>
<tr>
<td>YJM790</td>
<td>MATa lys2 gal2</td>
<td>John McCusker, Duke University</td>
</tr>
<tr>
<td>JSC10-1</td>
<td>MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::natMX4 RAD5</td>
<td>W303-1A background, St. Charles and Petes 2013</td>
</tr>
<tr>
<td>JSC19-1</td>
<td>MATα ade2-1 ura3 gal2 ho::hisG can1Δ::natMX4</td>
<td>YJM789 background, St. Charles and Petes 2013</td>
</tr>
<tr>
<td>YZ08</td>
<td>MATa gal2 V31539::kanMX4</td>
<td>S1 transformed with PCR fragment amplified from plasmid pFA6a-kanMX4 using primers V31539::DR F and V31539::DR R; insert kanMX4 next to CAN1</td>
</tr>
<tr>
<td>YZ10</td>
<td>MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::natMX4 RAD5 VIII212140::hphMX4</td>
<td>JSC10-1 transformed with PCR fragment amplified from plasmid pAG32 using primers VIII212140::HYG F and</td>
</tr>
<tr>
<td>ZY15</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::natMX4 RAD5 VIII212140::hphMX4 VIII218090::CAN1-kanMX4</td>
<td>YZ10 transformed with PCR fragment amplified from ZY08 using primers VIII218090::CAN1-KANMX F and VIII218090::CAN1-KANMX R; insert CAN1-kanMX4 at VIII218090</td>
</tr>
<tr>
<td>ZY18-10</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::natMX4 RAD5 VIII212140::hphMX4, VIII214177::URA3, VIII218090::CAN1-kanMX4</td>
<td>YZ15 transformed with PCR fragment amplified from S1 using primers VIII214177::URA3 F and VIII214177::URA3 R; insert URA3 into one of the CUP1 repeats at VIII214177; the CUP1 repeat with URA3 inserted was physically confirmed to be at the center of the cluster</td>
</tr>
<tr>
<td>ZY17</td>
<td>MATα ade2-1 ura3 gal2 ho::hisG can1Δ::natMX4</td>
<td>A derivative of JSC19-1 with enhanced copper resistance; constructed by selecting 0.4 mM copper sulfate-resistant derivatives of JSC19-1 on SDC + 0.4 mM copper sulfate; the CUP1 cluster in this strain is physically measured to be 30 kb</td>
</tr>
<tr>
<td>ZY26</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::natMX4 RAD5</td>
<td>A MATα derivative of JSC10-1; constructed by mating type switching of JSC10-1 using the plasmid pGAL-HO</td>
</tr>
<tr>
<td>ZY41</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 VIII212140::hphMX4, VIII214177::URA3, VIII218090::CAN1-kanMX4</td>
<td>YZ18-10 transformed with PCR fragment amplified from S1 using primers CAN1::HIS3 F and CAN1::HIS3 R; replace can1-100Δ::natMX4 with HIS3</td>
</tr>
<tr>
<td>ZY42</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1</td>
<td>YZ26 transformed with PCR</td>
</tr>
<tr>
<td>Strain</td>
<td>Description</td>
<td>Transformation</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>ade2-1 trp1-1 can1-100Δ::HIS3 RAD5</td>
<td>fragment amplified from S1 using primers CAN1::HIS3 F and CAN1::HIS3 R; replace can1-100Δ::natMX4 with HIS3</td>
<td>YZ43 MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 VIII212140::hphMX4, VIII214177::URA3, VIII218090::CAN1-kanMX4 rad52Δ::natMX4</td>
</tr>
<tr>
<td>YZ44 MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 rad52Δ::natMX4</td>
<td>YZ42 transformed with PCR fragment amplified from plasmid pAG25 using primers RAD52::NAT F and RAD52::NAT R; replace RAD52 with natMX4</td>
<td></td>
</tr>
<tr>
<td>YZ45 MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 VIII212140::hphMX4, VIII214177::URA3, VIII218090::CAN1-kanMX4 rad51Δ::natMX4</td>
<td>YZ41 transformed with PCR fragment amplified from plasmid pAG25 using primers RAD51::NAT F and RAD51::NAT R; replace RAD51 with natMX4</td>
<td></td>
</tr>
<tr>
<td>YZ46 MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 rad51Δ::natMX4</td>
<td>YZ42 transformed with PCR fragment amplified from plasmid pAG25 using primers RAD51::NAT F and RAD51::NAT R; replace RAD51 with natMX4</td>
<td></td>
</tr>
<tr>
<td>YZ47 MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 VIII212140::hphMX4, VIII214177::URA3, VIII218090::CAN1-kanMX4 mre11Δ::natMX4</td>
<td>YZ41 transformed with PCR fragment amplified from plasmid pAG25 using primers MRE11::NAT F and MRE11::NAT R; replace MRE11 with natMX4</td>
<td></td>
</tr>
<tr>
<td>YZ48 MATa leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 mre11Δ::natMX4</td>
<td>YZ42 transformed with PCR fragment amplified from plasmid pAG25 using primers MRE11::NAT</td>
<td></td>
</tr>
<tr>
<td>YZ49</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 VIII212140::hphMX4, VIII214177::URA3, VIII218090::CAN1-kanMX4 rad1Δ::natMX4</td>
<td>F and MRE11::NAT R; replace MRE11 with natMX4</td>
</tr>
<tr>
<td>YZ49</td>
<td>YZ41 transformed with PCR fragment amplified from plasmid pAG25 using primers RAD1::NAT F and RAD1::NAT R; replace RAD1 with natMX4</td>
<td></td>
</tr>
<tr>
<td>YZ50</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 rad1Δ::natMX4</td>
<td>YZ42 transformed with PCR fragment amplified from plasmid pAG25 using primers RAD1::NAT F and RAD1::NAT R; replace RAD1 with natMX4</td>
</tr>
<tr>
<td>YZ51</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 VIII212140::hphMX4, VIII214177::URA3, VIII218090::CAN1-kanMX4 cup2Δ::natMX4</td>
<td>YZ41 transformed with PCR fragment amplified from plasmid pAG25 using primers CUP2::NAT F and CUP2::NAT R; replace CUP2 with natMX4</td>
</tr>
<tr>
<td>YZ52</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 trp1-1 can1-100Δ::HIS3 RAD5 cup2Δ::natMX4</td>
<td>YZ42 transformed with PCR fragment amplified from plasmid pAG25 using primers CUP2::NAT F and CUP2::NAT R; replace CUP2 with natMX4</td>
</tr>
<tr>
<td>YZ60</td>
<td>MATα leu2-3,112 his3-11,15 ura3-1 ade2-1 can1-100Δ::HIS3 RAD5</td>
<td>YZ42 transformed with PCR fragment amplified from S1 using primers IV461633 F and IV462765 R; replace trp1-1 with TRP1</td>
</tr>
<tr>
<td>YZ103</td>
<td>MATa/MATα leu2-3,112/leu2-3,112 his3-11,15/his3-11,15 ura3-1/ura3-1 ade2-1/ade2-1 trp1-1/trp1-1 can1-100Δ::natMX4/can1-100Δ::natMX4 RAD5/RAD5 VIII212140::hphMX4/VIII212140 VIII214177::URA3/VIII214177</td>
<td>YZ18-10 x YZ26</td>
</tr>
<tr>
<td>Strain</td>
<td>Details</td>
<td>Crosses</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>YZ104</td>
<td>MATα MATα HO/ ho::hisG leu2-3,112/LEU2 his3-11,15/HIS3 ura3-1/ura3 ade2-1/ade2-1 trp1-1/TRP1 can1-100Δ::natMX4/can1Δ::natMX4 RAD5/RAD5 VIII212140::hphMX4/VIII212140 VIII214177::URA3/VIII214177 VIII218090::CAN1-kanMX4/VIII218090</td>
<td>YZ18-10 x YZ17</td>
</tr>
<tr>
<td>YZ113</td>
<td>MATα MATα leu2-3,112/leu2-3,112 his3-11,15/his3-11,15 ura3-1/ura3-1 ade2-1/ade2-1 trp1-1/trp1-1 can1-100Δ::HIS3/can1-100Δ::HIS3 RAD5/RAD5 VIII212140::hphMX4/VIII212140 VIII214177::URA3/VIII214177 VIII218090::can1-kanMX4/VIII218090 rad52Δ::natMX4/ rad52Δ::natMX4</td>
<td>YZ43 x YZ44</td>
</tr>
<tr>
<td>YZ114</td>
<td>MATα MATα leu2-3,112/leu2-3,112 his3-11,15/his3-11,15 ura3-1/ura3-1 ade2-1/ade2-1 trp1-1/trp1-1 can1-100Δ::HIS3/can1-100Δ::HIS3 RAD5/RAD5 VIII212140::hphMX4/VIII212140 VIII214177::URA3/VIII214177 VIII218090::CAN1-kanMX4/VIII218090 rad51Δ::natMX4/ rad51Δ::natMX4</td>
<td>YZ45 x YZ46</td>
</tr>
<tr>
<td>YZ115</td>
<td>MATα MATα leu2-3,112/leu2-3,112 his3-11,15/his3-11,15 ura3-1/ura3-1 ade2-1/ade2-1 trp1-1/trp1-1 can1-100Δ::HIS3/can1-100Δ::HIS3 RAD5/RAD5</td>
<td>YZ47 x YZ48</td>
</tr>
<tr>
<td>Strain</td>
<td>Genotype and Transformants</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>YZ116</td>
<td>MATα/MATα leu2-3,112/ura3-1 trp1-1/can1-100Δ::HIS3/can1-100Δ::HIS3 rad1Δ::natMX4/rad1Δ::natMX4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>YZ49 x YZ50</td>
<td></td>
</tr>
<tr>
<td>YZ117</td>
<td>MATα/MATα leu2-3,112/ura3-1 trp1-1/can1-100Δ::HIS3/can1-100Δ::HIS3 rad1Δ::natMX4/rad1Δ::natMX4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>YZ51 x YZ52</td>
<td></td>
</tr>
<tr>
<td>YZ118</td>
<td>MATα/MATα leu2-3,112/ura3-1 trp1-1/can1-100Δ::HIS3/can1-100Δ::HIS3 rad1Δ::natMX4/rad1Δ::natMX4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>YZ117 transformed with plasmid p425-GPD-CUP1</td>
<td></td>
</tr>
</tbody>
</table>
with plasmid p425-GPD-CUP1.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Strain description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD692</td>
<td>MATa/MATα leu2-3,112/leu2-3,112his3-11,15/his3-11,15 ura3-1/ura3-1 ade2-1/ade2-1 trp1-1/TRP1 can1-100Δ::natMX4/can1-100Δ::HIS3 RAD5/RAD5 VIII212140::hphMX4/VIII212140 VIII214177::URA3/VIII214177 VIII218090::CAN1-kanMX4/VIII218090</td>
</tr>
<tr>
<td>YZ18-10 x YZ60</td>
<td></td>
</tr>
</tbody>
</table>

Table S 3.2. PCR primer names and sequences.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ to 3’)</th>
<th>Strain constructed /analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>V31539::DR F</td>
<td>AGGTCTGAAGGAGTTTCAATGCTTTACTCCGTCTGCTTTCTTTTCGGGACGTA CGCTGCAGGTCGAC</td>
<td>YZ08 (insert kanMX4 into S1 at V31539)</td>
</tr>
<tr>
<td>V31539::DR R</td>
<td>CTATAATTCGTATATTATTTTTATC GATGATTCACAAACAATCTTGCTATCG ATGAATTGAGCTCG</td>
<td>YZ08 (insert kanMX4 into S1 at V31539)</td>
</tr>
<tr>
<td>EXT V31539 F</td>
<td>TGGTTCTAGGTCGGTGACG</td>
<td>YZ08 (confirm V31539::kanMX4 insertion)</td>
</tr>
<tr>
<td>EXT V31539 R</td>
<td>CACATTCCACGCCATTTTCG</td>
<td>YZ08 (confirm V31539::kanMX4 insertion)</td>
</tr>
<tr>
<td>VIII212140::H YG F</td>
<td>ATATATGTTGGATGATATATTAATAAAGGCAATTATTTTGGATGATTTCT CGGTCTAGGCTACGTAGCAGGTGCAG GTCGAC</td>
<td>YZ10 (insert hphMX4 into JSC10-1 at VIII212140)</td>
</tr>
<tr>
<td>VIII212140::H YZ F</td>
<td>TGTTATCCCGTCACCAATCGATCTA</td>
<td>YZ10 (insert hphMX4 insertion)</td>
</tr>
<tr>
<td>Gene/Vector</td>
<td>Sequence</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>YG R</td>
<td>TTTATTTCAGGGGATAGTACAACAGA TCAGTCATAAACGTATGAGATTC GAGCTCG</td>
<td>into JSC10-1 at VIII212140</td>
</tr>
<tr>
<td>EXT VIII211913 F</td>
<td>AGATCTTCACGAGCTTG</td>
<td>YZ10 (confirm VIII212140::hphMX4 insertion)</td>
</tr>
<tr>
<td>HYG INT R</td>
<td>CCTCCTACATCGAAGCTG</td>
<td>YZ10 (confirm VIII212140::hphMX4 insertion)</td>
</tr>
<tr>
<td>VIII218090::C AN1-KANMX F</td>
<td>TTATTTCAATAGATATGATTTCGGCATTTTTGAAT CGGATGTAACGTCGTAACGCTGCA GGTCGAC</td>
<td>YZ15 (insert CAN1-kanMX4 into YZ10 at VIII218090)</td>
</tr>
<tr>
<td>VIII218090::C AN1-KANMX R</td>
<td>TAGAGAACGTGGGAACAGTTCCGAC TGTGACCACCGGAACGTACGTCTGGTT AGCGCAGCATA GCGAACTGGAAGAATAACCAAGTC</td>
<td>YZ15 (insert CAN1-kanMX4 into YZ10 at VIII218090)</td>
</tr>
<tr>
<td>EXT VIII217877 F</td>
<td>CAACGACAATGCTCCCTCTCATAA</td>
<td>YZ15 (confirm VIII218090::CAN1-kanMX4 insertion)</td>
</tr>
<tr>
<td>CAN1 INT R</td>
<td>GGTGTTAGCTTGGCTGCCG</td>
<td>YZ15 (confirm VIII218090::CAN1-kanMX4 insertion)</td>
</tr>
<tr>
<td>VIII214177::URA3 F</td>
<td>GCAAGTACAGGGAAACACCTTGCACA ATATTTATCAATTTGCTTGGGCTGAAT CTAATTTATCGTTATGTGGCTGA GGTTTCCAGG</td>
<td>YZ18 (insert URA3 into YZ15 at VIII214177)</td>
</tr>
<tr>
<td>VIII214177::URA3 R</td>
<td>GGTACCAGCTGCAAAAAAATCATCTCC GTAACCTCCCATTTGACGAGCAAT TTCCTTTATTTGCTAGATTCCCAGGT AATAACTG</td>
<td>YZ18 (insert URA3 into YZ15 at VIII214177)</td>
</tr>
<tr>
<td>VIII214038 F</td>
<td>AAGTCTACTGATTTACTCACGCC</td>
<td>YZ18 (confirm VIII214177::URA3 insertion)</td>
</tr>
<tr>
<td><strong>Gene</strong></td>
<td><strong>Primer Name</strong></td>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>URA3</strong> R (V116511 R)</td>
<td>CCATTCTGCTATTCTGTATACCC</td>
<td>YZ18 (confirm VIII214177::URA3 insertion)</td>
</tr>
<tr>
<td><strong>CAN1::HIS3 F</strong></td>
<td>ATTAGTATTAGCGTGAATAATGTACACA TATACTATATATATATATATATATATATTAT ATCTGAGGCTCTGAAGGACTAGCATGT ACCTGAGCC</td>
<td>YZ41/YZ42 (use HIS3 to replace CAN1 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td><strong>CAN1::HIS3 R</strong></td>
<td>CAGCAAAGGCACAGAAGGTATTACCA TGGTAACTTCTGAATATCAAATACACTTAC TGGCAAGTTCGTATATACCACATTGCCA CCTATACCC</td>
<td>YZ41/YZ42 (use HIS3 to replace CAN1 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td><strong>HIS3 INT R</strong></td>
<td>GACTTCCCTGACTAATGCCC</td>
<td>YZ41/YZ42 (confirm can1Δ::HIS3 replacement, use with primer EXT V31539 F)</td>
</tr>
<tr>
<td><strong>RAD52::NAT F</strong></td>
<td>ATATAGGAAGAAATAATATAAAAATC AGAATGGAAAGGTAATATAGGGATT TTGGAGTAATAATAAATAATGATCGTACGC TGCAAGGTCGAC</td>
<td>YZ43/YZ44 (use natMX4 to replace RAD52 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td><strong>RAD52::NAT R</strong></td>
<td>ATGGAAATGAAGATAAAAATGTACGG AACGCAACCTAAGAAAAAGGGAGGA ATGGAAATGTAAACCTGTGTAATCGATG AATTCGAGCTCG</td>
<td>YZ43/YZ44 (use natMX4 to replace RAD52 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td><strong>XIII212091 F</strong> (RAD52)</td>
<td>GTGGGAATGGATAGGGTCCGA</td>
<td>YZ43/YZ44 (use with INT NAT R to confirm rad52Δ::natMX4 replacement)</td>
</tr>
<tr>
<td><strong>INT NAT R</strong> (AMC067)</td>
<td>CCGGTAAGCCGTGTCGTC</td>
<td>YZ43/YZ44/YZ45/YZ46/ YZ47/YZ48/YZ49/YZ50/ YZ51/YZ52 (to confirm the existence of natMX4)</td>
</tr>
<tr>
<td><strong>INT RAD52 R</strong></td>
<td>TGGGAATCAAGTACCGCGT</td>
<td>YZ43/YZ44 (use with XIII212091 F to confirm the deletion of RAD52)</td>
</tr>
<tr>
<td><strong>RAD51::NAT</strong></td>
<td>TCTTCTATATCCGTCAGTTCCATAT</td>
<td>YZ45/YZ46 (use natMX4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>F</td>
<td>ACTAGTAGTTAGTGTAGCGACAAAAG AGCAGACGTAGTTATTTG CGTACGCCTGAGGTACGAC</td>
<td>to replace RAD51 at its endogenous locus in YZ18-10/YZ26</td>
</tr>
<tr>
<td>RAD51::NAT</td>
<td>ATGGAAATGAAGATAAAAAATGTACGG AACGCAACCTAAGAAAGAGAGGA ATTGAAAGTAACCTGTGTAATCGATG AATTCGAGCTCG</td>
<td>YZ45/YZ46 (use natMX4 to replace RAD51 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V349750 F</td>
<td>GACTAGGCCACACTTCGTTAC</td>
<td>YZ45/YZ46 (use with INT NAT R to confirm rad51∆::natMX4 replacement)</td>
</tr>
<tr>
<td>(RAD51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT RAD51 R</td>
<td>CTATGGATACCAATCTTACCG</td>
<td>YZ45/YZ46 (use with V349750 F to confirm the deletion of RAD51)</td>
</tr>
<tr>
<td>MRE11::NAT</td>
<td>TCATGAATTTAGAATGCAAATTTGC TCCTCAAAAATGGCATACTTTGTGTC TCGAAGGCAAGGCCCTTTGCGTACGCT GCAGTACGAC</td>
<td>YZ47/YZ48 (use natMX4 to replace MRE11 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>F</td>
<td>TGGAAAAACACATTAAGAGAATGCAGAA CAAATTGACGCAAGTTGTACCTGTCAG ATCCGATAAAACTCGACTATCGATGA ATTCGAGCTCG</td>
<td>YZ47/YZ48 (use natMX4 to replace MRE11 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>MRE11::NAT</td>
<td>CATATTCCTATCACAGTTAACGC</td>
<td>YZ47/YZ48 (use with INT NAT R to confirm mre11∆::natMX4 replacement)</td>
</tr>
<tr>
<td>R</td>
<td>AGACATGGTTGTACAGTCCCGG</td>
<td>YZ47/YZ48 (use with XIII718186 F to confirm the deletion of MRE11)</td>
</tr>
<tr>
<td>XIII718186 F</td>
<td>GAGAGAGCACAGGTGTACTGGAGGGT TCAGGACGTTGAGGAGGTT TCAGGACGTTGAGGAGGTT TCAGGACGTTGAGGAGGTT AGCCTGCGGTCGAC</td>
<td>YZ49/YZ50 (use natMX4 to replace RAD1 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>(MRE11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT MRE11 R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAD1::NAT</td>
<td>AAGATTCAAAGAGCATGTCTAATTAT AACATATACGAGTGAAGTACCAATC</td>
<td>YZ49/YZ50 (use natMX4 to replace RAD1 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAD1::NAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GATGAATTTCGAGCTCG</td>
<td>endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>AMC052</td>
<td>AGTAAGCTATAGCCACAG</td>
<td>YZ49/YZ50 (use with INT NAT R to confirm rad1Δ::natMX4 replacement)</td>
</tr>
<tr>
<td>INT RAD1 R (AMC053)</td>
<td>TAACACTGCATCGTCATC</td>
<td>YZ49/YZ50 (use with AMC052 to confirm the deletion of RAD1)</td>
</tr>
<tr>
<td>CUP2::NAT F</td>
<td>GTAAGTATTTAAGCTTGAGACGGAAATA GCAATGGCCATGCGATGAATTATT AGACGGCGGGTGGATATAACGTACG TGGAGATCGTGC</td>
<td>YZ51/YZ52 (use natMX4 to replace CUP2 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>CUP2::NAT R</td>
<td>ATCTCGGCCGACTGCGAGGCGGCGAATATATATTGATATGTATATATTATATAGATCGTGGATT TGGAGATCGTGC</td>
<td>YZ51/YZ52 (use natMX4 to replace CUP2 at its endogenous locus in YZ18-10/YZ26)</td>
</tr>
<tr>
<td>VIII190919 F (CUP2)</td>
<td>CCTTCCACCACGCATACGATCG</td>
<td>YZ51/YZ52 (use with INT NAT R to confirm cup2Δ::natMX4 replacement)</td>
</tr>
<tr>
<td>INT CUP2 R</td>
<td>GAGCGACCATTATTAGGCC</td>
<td>YZ51/YZ52 (use with VIII190919 F to confirm the deletion of CUP2)</td>
</tr>
<tr>
<td>CUP1 amp5</td>
<td>CGAGATGAAATGAATAGCAACGG</td>
<td>Synthesize DIG probes that hybridize to both 2.0 kb and 1.2 kb CUP1 repeats</td>
</tr>
<tr>
<td>CUP1 amp3</td>
<td>TTCATTT CCCAGACGCATCGCATGAC</td>
<td>Synthesize DIG probes that hybridize to both 2.0 kb and 1.2 kb CUP1 repeats</td>
</tr>
<tr>
<td>CUP1 W303 spec amp5</td>
<td>AGATTTCTGTCAGTTCACTAACGG</td>
<td>Synthesize DIG probes that hybridize specifically to the 2.0 kb</td>
</tr>
<tr>
<td>Plasmid name</td>
<td>Relevant features</td>
<td>Source</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>pFA6a – kanMX4</td>
<td>kanMX4</td>
<td>Wach et al. 1994</td>
</tr>
</tbody>
</table>

Table S 3.3. Plasmids used in this study.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAG25</td>
<td>natMX4</td>
<td>Goldstein and McCusker 1999</td>
</tr>
<tr>
<td>pAG32</td>
<td>lphMX4</td>
<td>Goldstein and McCusker 1999</td>
</tr>
<tr>
<td>pGAL-HO</td>
<td>Galactose-regulated HO gene</td>
<td>Herskowitz and Jensen 1991</td>
</tr>
<tr>
<td>p425-GPD-CUP1</td>
<td>CUP1 gene regulated by GPD promoter; LEU2</td>
<td>Constructed in this study</td>
</tr>
</tbody>
</table>
4 Discussion and future directions

4.1 De novo duplications from a single-copy gene

With advancements in whole-genome sequencing, it has become clear that CNVs (genetic variations smaller than chromosomal aberrations and bigger than small deletions and insertions) account for a substantial amount of genetic diversity. The CNV mutation rate is estimated to range between $1.7 \times 10^{-6}$ to $1.0 \times 10^{-4}$ per locus per generation in humans (Zhang et al. 2009a), while the mutation rate for single-nucleotide variants (SNVs) is about $1.2 \times 10^{-8}$ per base pair per genome (Campbell et al. 2012). Therefore, it is reasonable to believe that CNVs, instead of SNVs, are the predominant driving forces in genetic and genomic evolution.

As recorded in the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation/), 29.7% of the human genome is subject to copy number variations (Zhang et al. 2009a). Moreover, CNVs are not randomly distributed in the human genome, but highly correlated with other genomic elements, such as segmental duplications (SDs) ($p < 2 \times 10^{-16}$), sequences that occur twice or more times in a haploid genome (Cooper et al. 2007). Certain types of CNVs have been correlated with specific types of cancer (Krepischi et al. 2012). In addition, some CNVs appear to be associated with other types of human disease including susceptibility to HIV infection, autism and schizophrenia (Fanciulli et al. 2007; Abrahams and Geschwind 2008). Thus, an
understanding of the mechanisms involved in CNV formation may have relevance to human health.

In order to better understand how CNVs are formed, it is necessary to define the breakpoints of the duplication with single-base-pair resolution. There are two distinct mechanisms for forming CNVs in yeast. The first type requires substantial sequence homology, such as pre-existing tandem repeats, or flanking dispersed repeats such as Ty and delta elements. Copy-number changes in tandemly-repeated rDNA genes and \textit{CUP1} genes are representative of this type of CNV. In pre-existing tandem repeats, CNVs can be generated by various homologous recombination (HR) events depicted in Figure 4.1. When genes are flanked by homologous sequences, such as Ty elements in the yeast genome, non-allelic homologous recombination (NAHR) can result in the duplication of a single copy gene (Figure 4.1D). NAHR can also generate gross chromosomal rearrangements (GCRs) if the broken molecule uses homologous sequences on a different chromosome for repair (Figure 4.1E).

The second type of CNVs is generated by events that do not require extensive sequence homology. \textit{De novo} duplications from a single-copy gene are representative of this type of CNVs. They occur at a much lower frequency compared to the first type of CNVs, and are dependent on the genetic context of the duplication. Microhomology/Microsatellite-Induced End-joining (MMEJ) events are BIR events in which little homology is present at the point of strand invasion (Figure 4.1F). In Chapter
Figure 4.1. Mechanisms for generating CNVs in yeast.

A. Unequal crossover between pre-existing tandem repeats can generate gene amplifications and deletions.

B. Gene conversions can also generate amplifications and deletions.

D. Crossover between Ty elements flanking a gene can generate the duplication of this gene.

E. Break-induced replication (BIR) mediated by Ty elements on different chromosomes generates gene duplication and gross chromosomal rearrangements.

F. Microhomology-mediated BIR is another mechanism for generating a de novo duplication of a gene.

G. Non-homologous end-joining (NHEJ) can generate a de novo duplication of a gene.

2, we proposed a non-homologous end joining (NHEJ) mechanism in which CNVs are generated from sequences with little or no homology (Figure 4.1G).

Researchers have developed several different systems for detecting or selecting CNVs in yeast. CGH arrays have been used to determine gene-copy numbers in yeast genome (Pinkel et al. 1998). Clamped homogenous electric fields (CHEF) gel electrophoresis can separate yeast chromosomes based on their sizes, and hence enable us to detect GCR events that result from CNVs (Michel et al. 2005). Copy-number changes occur fairly frequently in pre-existing tandem repeats. For instance, the frequency of CNVs within the rRNA tandem repeats was measured to be at least $10^{-2}$/mitotic division and about $10^{-1}$/meiotic division (Petes 1980). In 1990, Welch et al. analyzed meiotic recombination in the CUP1 locus in a diploid, in which both arrays contained 12 copies of CUP1. They showed by Southern analysis that approximately
11% of 202 unselected tetrads had at least one non-parental spore array of CUP1 (Welch et al. 1990).

*De novo* duplications of a single-copy genes occur much less frequently than changes within tandem repeats presumably because the rate of spontaneous NHEJ or MMEJ is much lower than the rate of HR. Dorsey *et al.* developed an assay selecting for *de novo* gene duplication using the ADH system. Cells lacking ADH1 cannot grow on glucose medium with antimycin A, while additional copies of ADH2 or ADH4 restore growth (Dorsey *et al.* 1992). Using this system, Walton *et al.* found a type of ADH4 *de novo* duplication involving formation of a linear, palindromic plasmid; ADH4 is near the end of chromosome VII (Walton *et al.* 1986).

Another system (previously discussed in Chapter 2) was developed by Koszul *et al.* (Koszul *et al.* 2004). Haploid strains with a deletion of RPL20A (encoding a ribosomal protein) grow slowly, but an additional copy of RPL20B can restore growth rate. After 80 generations, 83 of 100 haploid cultures contained revertants with normal growth rates. Several types of duplication breakpoints were identified: GTT/AAC repeats, poly A/T, microhomology regions (2-9 bp), and delta and Ty elements16. Later, Payen *et al.* used the same assay to study gene duplications in replication-compromised strains, and they found two duplication mechanisms: BIR events that require little or no sequence homology (MMBIR, Figure 4.1F), and various Rad 52-dependent recombination pathways (Payen *et al.* 2008) (Figure 4.1A-4.1E).
4.2 *De novo duplications of a single copy of the CUP1 gene*

Despite the abundance of CNVs studied in yeast, *de novo* duplication events of a single gene are very rare and difficult to identify. In Chapter 2, we investigated the structures of five different naturally-occurring CUP1 tandem repeats of various lengths and sequences. We had also mimicked the natural evolutionary process that generates *de novo* duplications of the CUP1 gene, by selection with increased copper concentration. The rationale is that we believe high concentrations of copper have exerted pressure on yeast strains during natural evolution, providing the driving force for generating CUP1 *de novo* duplications and tandem repeats.

The genomic DNA of a haploid (DTY3) that contains only a single copy of CUP1 was also sequenced. Also in Chapter 2, we studied the frequency of a tandem CUP1 array contracting to a single copy CUP1 gene via homologous recombination in the W303-isogenic strain YZ22. The YZ22 – derived strain YZ23 was confirmed contain a single copy CUP1 gene at the CUP1 locus. Since the both DTY3 and YZ23 strains are sensitive to 0.2 mM copper, independent DTY3 derivative colonies were selected on SD complete medium containing 0.2 mM – 0.4 mM copper. These colonies were double checked for their resistance to 0.2 mM – 0.4 mM copper, and resistant derivatives were analyzed by PCR using primers F1 and R1' (Figure 2.2). With these primers, one would expect to get PCR products only if *de novo* duplications had occurred.
400 colonies that were resistant to 0.2 mM copper were analyzed by PCR in this way. Another 600 colonies resistant to 0.3 mM copper were also screened by PCR analysis. However, none of these one thousand copper resistant derivatives contained a de novo duplication of the CUP1 gene. Among them, eight copper-resistant DTY3 derivatives were analyzed by CGH microarrays. One of these derivatives was disomic for chromosome VIII whereas all the other seven derivatives showed no duplications of CUP1. It is possible that the majority of copper-resistant derivatives contain mutations in other genes involved in the copper homeostasis in S. cerevisiae (reviewed in Section 1.3). Specifically, CTR1, CTR2 and CTR3 encode copper transporters on the plasma membrane and on the vacuolar membrane. Mutants of these genes can either block excessive copper outside the plasma membrane or can accumulate high levels of copper in the vacuole, making the cells resistant to high levels of copper in the medium. Also, the copper regulator gene CUP2 encodes a transcriptional factor that activates CUP1 transcription. In addition, Fre1 and Fre2 are the major cell-surface iron reductases, converting Cu$^{2+}$ to usable Cu$^{+}$, the only form that can be transported through plasma membrane. Mutations of these genes could lead to a copper-resistance phenotype in strains with a single CUP1 gene.

To overcome the effects of mutations in other copper-regulating genes, another haploid strain YZ24 that contains a single copy CUP1 gene was constructed by mating type switch of YZ23. A diploid strain YZ102, in which both homologs contain a single
CUP1 gene, was constructed by crossing YZ23 and YZ24. A recessive mutation will not convey phenotypes in diploids, hence reducing the frequency of analyzing copper-resistant strains with undesirable mutations. A few hundred copper-resistant derivatives of the diploid strain YZ102 were analyzed by PCR. However, none of them contained \textit{de novo} duplication of the CUP1 gene.

Although the CUP1 repeats observed in the wild yeast isolates are 2.0 kb or less, it is possible that some of the duplications derived in the lab will be too large to produce PCR products. Therefore, some of the copper-resistant isolates that fail to produce a PCR product could be examined by CGH microarrays that can detect duplications of any size. We have previously used such arrays to examine gene duplications associated with low levels of DNA polymerase (Lemoine \textit{et al}. 2005).

We expect that the frequency of CUP1 duplications will be low; the rate of duplication of the yeast RPL20B gene was about $10^{-7}$/cell division. It is possible that it will be difficult to select for these low frequency events using only the copper resistance selection. An alternative approach is to construct a haploid that has one CUP1 gene adjacent to an SFA1 gene. SFA1, which affects resistance to formaldehyde, is located on chromosome IV. By selecting strains that were simultaneously copper- and formaldehyde-resistant, yeast strains with a duplication of this cassette can be selected (Zhang \textit{et al}. 2013). In the experiments of Zhang \textit{et al}., in which the cassette was located on chromosome V, most of the \textit{de novo} duplications in haploid strains occurred via non-
allelic homologous recombination (NAHR) between Ty elements. In diploids, most de novo duplications were generated by a recurrent nonreciprocal translocation event involving Ty elements (Gresham et al. 2008).

To detect de novo duplications at the “normal” CUP1 locus, we could delete the SFA1 gene from chromosome IV and insert it immediately adjacent to CUP1 on chromosome VIII. Derivative cells containing potential de novo duplications could be selected in medium containing both copper and formaldehyde at concentrations that are inhibitory to the growth of the starting strain. Derivatives with enhanced resistance to both copper and formaldehyde could be analyzed by PCR using primers F1 and R1’ and/or by comparative genome hybridization arrays.

Once the rate of de novo duplications at the CUP1 locus can be measured directly, the structures and sequences of these duplications will be characterized in a similar way as we described in Chapter 2. Genetic regulations of the de novo duplications can also be investigated systematically in mutant strains in which HR and NHEJ pathways are impaired. This study can potentially provide direct evidence to corroborate the de novo duplication mechanism via NHEJ, proposed by us in Chapter 2. Moreover, we can gain important insights into the mechanisms and regulations of de novo duplications. If we find no de novo duplications in the wild-type strain, we could look for such events in strains that have a very high level of chromosome rearrangements such as tel1 mec1 strains.
4.3 Mechanism of intra-/sister-chromatid recombination

In Section 3.3.8, we described the effects of mutations in various genes known to have a role in HR. The absence of Rad51 had a surprisingly small effect in recombination at the CUP1 locus, although this absence greatly reduced the level of exchange between the centromere and the CUP1 array. Three approaches to extend these observations could be done. First, we could examine the effects of other mutations that reduce the levels of HR in at least some contexts. One particularly interesting experiment would be to examine the effects of the *rad57* mutation. In experiments done in haploid strains, Mozlin *et al.* reported the *rad57* mutation reduced sister chromatid recombination more than the *rad51* mutation (Mozlin *et al.* 2008). There are many other such mutations whose effects we could examine in this system.

A second approach would be to examine the effects of genetic alterations that perturb DNA replication intra-/sister-chromatid events. If template switching can sometimes result in loss of CUP1 repeats, such alterations may elevate the frequency of marker loss. For example, one could utilize strains in which the levels of DNA polymerase can be regulated by the amount of galactose in the medium (Lemoine *et al.* 2005).

As a third approach, we could directly screen for new mutations that affect the ratio of interhomolog and intra-/sister chromatid events. Since our assays are performed in a diploid strain, this approach, however, could only detect dominant mutations.
affecting recombination. An alternative strategy is to transform the diploid strain with recombinant plasmids that over-express various yeast proteins, looking for those that elevate or reduce interhomolog or intra-/sister-chromatid events. The characterization of such genes may lead to insights into the mechanisms that control the substrate used in repairing DNA breaks.

Lastly, we have restricted our analysis of recombination in the CUP1 array to spontaneous events. A useful comparison would be to measure interhomolog and intra-/sister-chromatid events that are induced by I-SceI or HO. For this analysis, we would integrate the target site for the endonuclease within the array and include the GAL-inducible endonuclease elsewhere in the genome. We would then analyze the induced events using the methods described in Chapter 3. Another extension of this experiment would be to express the endonuclease at different times in the cell cycle. For this experiment to be feasible, we would need a form of the enzyme that can be rapidly induced and, subsequently, rapidly degraded.

In summary, the approaches described in this thesis are only the first steps to fully understanding the mechanisms that control the formation and genetic stability of tandem repeats in yeast and other eukaryotes.
References


Arbel, A., D. Zenvirth, and G. Simchen, 1999 Sister chromatid-based DNA repair is mediated by RAD54, not by DMC1 or TID1. EMBO J 18 (9):2648-2658.


Ercolani, L., J.L. Stow, J.F. Boyle, E.J. Holtzman, H. Lin *et al.*, 1990 Membrane localization of the pertussis toxin-sensitive G-protein subunits alpha i-2 and alpha i-3 and


Klein, H.L., 1988 Different types of recombination events are controlled by the RAD1 and RAD52 genes of Saccharomyces cerevisiae. Genetics 120 (2):367-377.


Pinkel, D., R. Segraves, D. Sudar, S. Clark, I. Poole et al., 1998 High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20 (2):207-211.

Prado, F., and A. Aguilera, 1995 Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat
recombination in yeast: different requirements for the \textit{RAD1}, \textit{RAD10}, and \textit{RAD52} genes. \textit{Genetics} 139 (1):109-123.


Williams, G.J., S.P. Lees-Miller, and J.A. Tainer, 2010 Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. DNA Repair (Amst) 9 (12):1299-1306.


Biography

Ying Zhao was born on February 11, 1988 in Wuhan, China. She earned a BS from Tsinghua University in 2010, majoring in biology and ranking in the top 12% of her class. While at Tsinghua, she performed research in structural biology, supervised by Zihe Rao, the director of the Chinese Academy of Sciences Institute of Biophysics. This work included solving the crystal structure of the extracellular domain of the human anthrax toxin receptor 1. By the time she graduated college, she was the co-author of two research papers, including one paper on which she was a first co-author (Cai et al. 2011), and another describing a promising treatment for anthrax (Fu et al. 2010). She also spent a summer at the University of Michigan, performing research under the HHMI investigator Tom Kerppola. Ying matriculated at Duke as a Chancellor’s Scholar in 2010, and performed research with Jingdong Tian before joining Tom Petes’ lab to complete this thesis (Zhao et al. 2014).