DOI: 10.1534/genetics.104.028076

Expansion and Contraction of the *DUP240* Multigene Family in *Saccharomyces cerevisiae* Populations

Véronique Leh-Louis, Bénédicte Wirth, Serge Potier, Jean-Luc Souciet[1] and Laurence Despons

Laboratoire de Microbiologie et de Génétique, FRE 2326 Université Louis Pasteur/Centre National de la Recherche Scientifique, Institut de Botanique, F-67083 Strasbourg, France

Manuscript received February 25, 2004
Accepted for publication May 17, 2004

ABSTRACT

The influence of duplicated sequences on chromosomal stability is poorly understood. To characterize chromosomal rearrangements involving duplicated sequences, we compared the organization of tandem repeats of the *DUP240* gene family in 15 *Saccharomyces cerevisiae* strains of various origins. The *DUP240* gene family consists of 10 members of unknown function in the reference strain S288C. Five *DUP240* paralogs on chromosome I and two on chromosome VII are arranged as tandem repeats that are highly polymorphic in copy number and sequence. We characterized DNA sequences that are likely involved in homologous or nonhomologous recombination events and are responsible for intra- and interchromosomal rearrangements that cause the creation and disappearance of *DUP240* paralogs. The tandemly repeated *DUP240* genes seem to be privileged sites of gene birth and death.

GENE redundancy is apparent in all sequenced genomes. In *Saccharomyces cerevisiae*, ~30% of the genes are present in at least two copies (Dujon 1998). Different mechanisms, acting independently or in combination, have been proposed to explain the origin of the duplicated copies in eukaryotic genomes: duplication of the entire genome (Wolfe and Shields 1997; Lander *et al.* 2001), segmental duplications (Clark 1994; Llorente *et al.* 2000; Emanuel and Shaikh 2001), and single gene duplications (Sankoff 2001). Multigene families are thought to evolve according to two major mechanisms. Studies on tandemly repeated rRNA and U2 snRNA genes have suggested that members of a gene family do not evolve independently of each other but rather evolve in a concerted fashion (Liao 1999). Sequences of family members become homogenized by interlocus recombination events that preserve gene function. In contrast, multigene families associated with the vertebrate immune system comply with the birth-and-death model of evolution (Nei *et al.* 1997), in which repeated gene duplication is followed by functional divergence, gene inactivation, or gene deletion.

The *S. cerevisiae DUP240* family, which consists of 10 genes with a high level of nucleotide identity (from 50 to 98%) in the reference strain S288C, is one of the largest gene families in yeast (Dujon 1998). Members of this family are scattered on four chromosomes and arranged either as tandem repeats or as isolated genes (Figure 1; Feuermann *et al.* 1997). Five open reading frames (ORFs; YAR027w, YAR028w, YAR029w, YAR031w, and YAR033w) are tandemly repeated on chromosome I; YGL051w and YGL053w are directly repeated on chromosome VII. These loci are named tandem I and tandem VII, respectively. *DUP240* orthologs have been identified only in species of the Saccharomyces *sensu stricto* group. Most *DUP240* ORFs encode proteins of ~240 amino acids with two potential transmembrane domains. Simultaneous deletion of the 10 *DUP240* ORFs in strain S288C does not alter cell viability, and the Dup240 proteins are membrane associated (Poirey *et al.* 2002).

The *DUP240* gene family of yeast provides a good system to approach questions of the evolution of duplicated sequences and their influence on chromosomal stability, since its tandem repeats are potential targets for intra- and interchromosomal recombinations that could reshape the chromosome. The presence of a highly conserved sequence unit between the tandem I and tandem VII loci (Figure 1) favors this hypothesis. We supposed that if DNA recombination events occurred at tandemly repeated loci, the analysis of these loci in different strains of the same species should reveal variability in the organization and sequence of the duplicated genes. The polymorphism observed in the solo *DUP240* genes is due mainly to the fixation of point mutations and to allelic recombination (Leh-Louis *et al.* 2004). Here we describe results that suggest that polymorphism within the tandem *DUP240* loci results mainly from variations in the structural organization of the *DUP240* ORFs. The identification of short repetitive

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AJ585103–AJ585108, AJ585190, AJ585524, AJ585525, AJ586490–AJ586508, and AJ586612.

[1]*Corresponding author:* Laboratoire de Microbiologie et de Génétique, FRE 2326 ULP/CNRS, Institut de Botanique, 28 rue Goethe, F-67083 Strasbourg Cedex, France. E-mail: souciet@gem.u-strasbg.fr

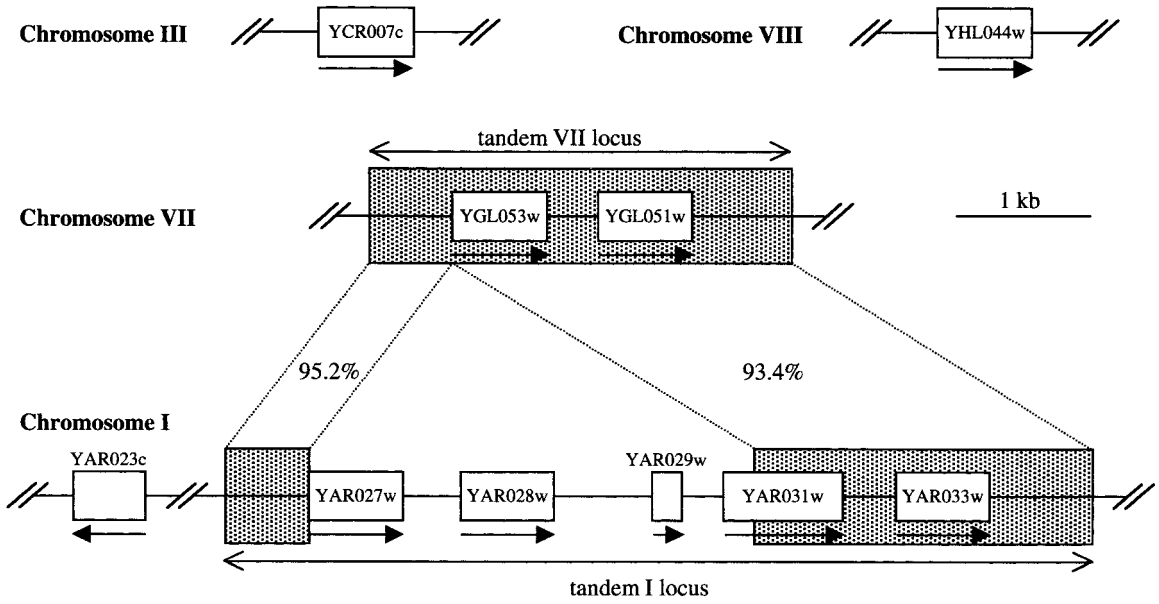

FIGURE 1.—Genetic organization and chromosomal localization of the 10 *DUP240* ORFs of the *S. cerevisiae* strain S288C. ORFs are represented by open boxes; arrows indicate their orientation with respect to the centromere. Percentages represent the level of nucleotide identity between the sequence units indicated by stippled areas.

DNA sequences that are likely involved in recombination processes suggests that the large chromosomal rearrangements observed at the tandem *DUP240* loci are due mainly to nonallelic recombination events. The presence of new *DUP240* paralogs and relics in several yeast strains allows us to conclude that the *DUP240* family evolved by a gene birth-and-death mechanism.

MATERIALS AND METHODS

***S. cerevisiae* strains and media:** All strains used in this study are listed in Table 1. Only the laboratory strains S288C and Σ1278b are heterothallic and haploid; the other strains are homothallic. Cells were cultivated at 30° on YPD medium (1% yeast extract, 2% peptone, 2% glucose, and 2% agar). Asci were obtained after 3 days on sporulation medium (1% potassium acetate and 2% agar) and spores were isolated by tetrad dissection using a Singer MSM micromanipulator.

Molecular biology techniques: Genomic DNA was isolated from yeast cells using the procedure described by HOFFMAN and WINSTON (1987). All PCR amplifications were performed with the Expand long template PCR system (Roche, Indianapolis). Sequences of primers used for PCR amplifications were determined on the basis of the published genomic sequence of the yeast strain S288C (GOFFEAU *et al.* 1997).

DNA sequencing and sequence analyses: DNA fragments obtained after PCR amplification were purified through MicroSpin S-400 HR columns (Amersham Pharmacia Biotech) and sequenced using AmpliTaq FS DNA polymerase and BIGDYE terminators. Sequence reactions were elaborated on an Applied Biosystems (Foster City, CA) 373XL sequencer. BLASTN analyses were performed in the SGD web site (http://www.yeastgenome.org/) to compare the sequence of the PCR products to the genomic sequence of S288C. Nucleic acid or protein sequences were aligned using programs available in the UWGCG package version 8.1 (DEVEREUX *et al.* 1984). A phylogenetic analysis was performed with all *DUP240* DNA sequences identified at the tandem I and tandem VII loci, using the PHYLIP phylogeny inference package version 3.573 (FELSENSTEIN 1989). Phylogenies were estimated with the DNApars program by the parsimony method. The stability of the individual branches was assessed using the bootstrap method (SAITOU and NEI 1987).

Southern blot analysis: PCR amplifications were performed from the genomic DNA of all class A strains with primers specific to YAR027w (5′ ATGCAAACCCCTTCAGAA 3′) and YAR033w (5′ CCGTCTTTTTAAGAAGCG 3′). PCR products were purified, after agarose gel electrophoresis, using the GENECLEAN II system (Q-BIOgene). DNA fragments digested by restriction endonucleases were separated by agarose gel electrophoresis and blotted onto Hybond-N^+ membrane (Amersham Pharmacia Biotech). Digoxygenin (DIG)-labeled DNA probes were prepared using the DIG DNA labeling kit (Roche Diagnostics). PCR products digested with *Bsp*EI were hybridized to the YAR029w and YAR031w gene probes, and those digested with *Pvu*II and *Hae*III were hybridized to the YAR027w, YAR028w, and YAR033w gene probes. Signal detection was performed using an enzyme-catalyzed color reaction (Roche Diagnostics).

RESULTS

Chromosomal organization of the tandem VII locus: ORFs in direct repetition are found in all eukaryotic genomes sequenced so far but are rare in the yeast genome. This particular gene organization is interesting for the purpose of studying genome dynamics. It seems relevant to homologous recombination, which is very efficient in yeast. To demonstrate that the tandemly repeated *DUP240* loci are potential sites for multiple chromosomal rearrangements, we compared the structural organization of these loci in the reference strain S288C with that in 14 other strains of the same species (Table 1). The synteny between the genes flanking this locus, *OLE1* and *ERV14* on one side and YGL050w and

TABLE 1

***S. cerevisiae* strains**

Strain	Isolation	Area and/or country of origin	Reference/source
Laboratory strains			
S288C			
Σ1278b (ATCC 42800)			Sumbu *et al.* (1983)
Natural isolates and industrial strains			
CLIB95 (L1425-4B)[a]	Wine	France	Nguyen and Gaillardin (1997)
CLIB219 (CBS 5287)	Grape berries	Russia	Nguyen and Gaillardin (1997)
CLIB382 (CBS 1782)	Beer	Japan	Ryu *et al.* (1996)
CLIB388 (ATCC 10615)	Beer	Japan	INRA[c], Grignon
CLIB410	Sake	Japan	INRA, Grignon
CLIB413	Fermented rice	China	INRA, Grignon
K1	Wine	France	INRA, Colmar (Oeno-France)
R12	Grape berries	France (Alsace)	INRA, Colmar
R13	Grape berries	France (Alsace)	INRA, Colmar
CLIB556 (TL213[b])	Cheese	France	INRA, Grignon
CLIB630 (TL229[b])	Cheese	France	INRA, Grignon
YIIc12	Wine	France (Sauternes)	Bordeau, France[d]
YIIc17	Wine	France (Sauternes)	Bordeau, France

[a] Collection de Levures d'Intérêt Biotechnologique, INRA Grignon.
[b] Name used in this study to distinguish cheese strains from the others.
[c] Institut National de la Recherche Agronomique, France.
[d] Laboratoire de Biologie Cellulaire de la Levure, UPR CNRS 9026.

TIF4632 on the other side (Figure 2), is perfectly conserved for all the tested strains. In contrast, we observed that the sizes of the PCR fragments obtained with the primers 1 and 2 flanking the tandem VII region (Figure 2) are subject to numerous variations (data not shown). We determined the DNA sequence of this region of the genome in the 15 strains and found that there are three classes (Figure 2). Class 1 consists of the two laboratory strains S288C and Σ1278b in which YGL051w and YGL053w are arranged as tandem repeats. Class 2 strains contain a single *DUP240* ORF at their tandem VII locus. A phylogenetic analysis was performed with the DNA sequences of all ORFs identified at both tandem VII and tandem I loci. As shown in Figure 3, the ORFs identified in class 2 strains constitute two new members of the *DUP240* multigene family. These new paralogs were named *DUP X* (CLIB413, CLIB410, YIIc12, and YIIc17 strains) and *DUP Y* (CLIB219 strain) according to their phylogenetic relationships. Finally, class 3 is composed of 10 strains in which the tandem VII locus is composed of only Ty and solo LTR elements. Among the natural diploid strains we studied, three cases of heterozygosity are observed in the organization of the tandem VII locus (strains R12, YIIc12, and YIIc17; Figure 2).

Chromosomal organization of the tandem I locus: The gene synteny is also conserved upstream (between *CDC15* and YAR023c) and downstream (between *YAT1* and *SWH1*) of the tandem I locus. However, the tandem I locus varies extensively among the strains, in both ORF copy number and nucleotide sequence. Analysis of the size of the PCR products generated using primers 3 and 4 (Figure 4) and of the corresponding fragments obtained by their digestion with *Hae*III and *Nhe*I allowed a ranking of the strains among classes A, B, and C (data not shown).

For the class C strains TL213 and TL229, the fragment between YAR023c and *YAT1* amplified by the PCR is 2.5 kb long, compared with a size of >10 kb for all the other strains. Both strains are missing the *DUP240* ORFs (revealed from the DNA sequence of this DNA fragment from strain TL229, which contained two Ty LTR elements, and confirmed by a dot blot for strain TL213; Figure 4). The eight strains of class A (YIIc17, YIIc12, K1, R12, R13, Σ1278b, CLIB95, and CLIB388) show an identical fragment profile following digestion of a DNA fragment of this locus with *Hae*III and *Nhe*I, and thus only the tandem I region of strain YIIc17 was sequenced. The five tandemly repeated *DUP240* ORFs present on chromosome I of strain S288C are also present in strain YIIc17 in the same order and orientation (Figure 4). All eight class A strains shared the same Southern blot hybridization pattern using the gene probes YAR027w, YAR028w, YAR029w, YAR031w, and YAR033w (data not shown), so we conclude that the five *DUP240* ORFs, present at the tandem I locus of all class A strains, are organized in the same way as in strain YIIc17 (Figure 4).

DNA fragments of the tandem I locus from each of the four strains of class B (CLIB219, CLIB382, CLIB410, and CLIB413) show a different fragment profile after digestion with *Hae*III and *Nhe*I, indicating that their *DUP240* ORFs are organized differently. The DNA se-

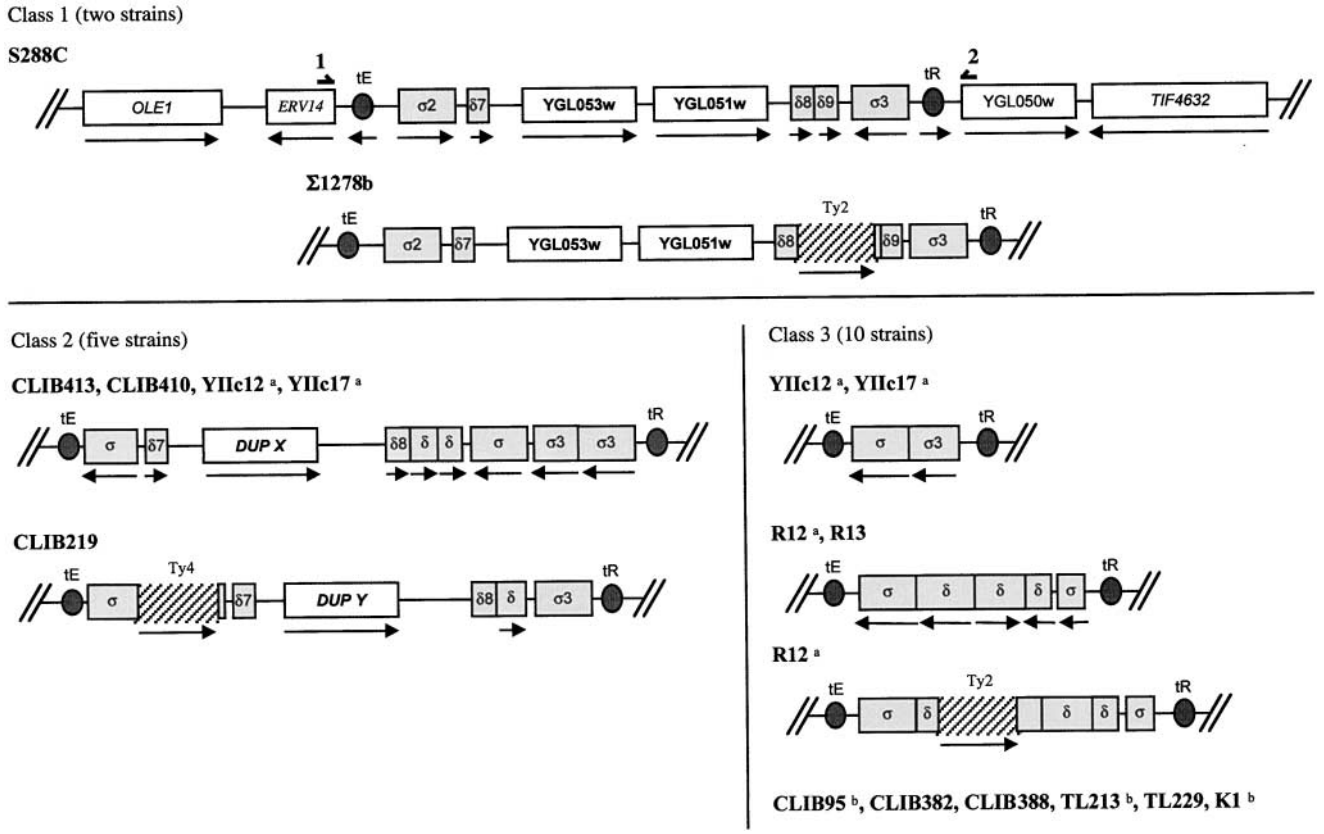

FIGURE 2.—Genetic organization of the tandem VII locus in the 15 studied yeast strains (not to scale). Genes and ORFs are represented by open boxes. Names of *DUP240* ORFs are indicated in boldface type. Genes that surround the tandem VII locus are specified only for the reference strain S288C. Solo long terminal repeats and tRNA genes are denoted by shaded boxes and circles, respectively. Complete Ty sequences appear as hatched areas. The orientation of genetic sequences is indicated by arrows. The tandem VII locus was sequenced in the studied strains after PCR amplification with primers 1 and 2 (arrows above *ERV14* and YGL050w). Strains of classes 1, 2, and 3 contain two, one, and no *DUP240* ORFs, respectively, on chromosome VII between *ERV14* and YGL050w. Superscript a indicates heterozygous strains. Superscript b indicates strains for which the tandem VII locus is not sequenced; presence or absence of *DUP240* ORFs and Ty elements was checked by dot-blot analyses.

quence of each locus was determined (Figure 4). Strain CLIB413 is heterozygous with regard to the structural organization of the tandem I locus, so we examined both E1 and D1 meiotic products. The E1 tandem I locus is composed of seven *DUP240* copies arranged in tandem. Three of them are new paralogs named *DUP A*, *DUP B*, and *DUP C*. They are located upstream of the block YAR028w-YAR029w-YAR031w-YAR033w previously identified in S288C and have several distinctive features. First, *DUP A* is phylogenetically related to *DUP X*, the paralog present at the tandem VII locus of the same strain (Figure 3). Second, *DUP B* is closely related to YAR029w (Figure 3). Whereas YAR029w represents the shortest coding sequence for a member of the *DUP240* family, *DUP B* encodes a 191-amino-acid-long protein, a size closer to the standard size for a member of this family (240 amino acids). Finally, *DUP C* is similar to YAR031w in its first 262 nucleotides, which are 96% identical to the corresponding sequence of YAR031w. The D1 tandem I locus of CLIB413, in contrast, is composed of three *DUP240* copies: YAR028w, YAR029w, and a new paralog designated as YAR031-033w that probably results from an in-frame deletion-fusion event between YAR031w and YAR033w.

The genetic organization of the CLIB410 tandem I locus (Figure 4) is quite similar to that depicted for CLIB413 E1 but with two variations: (i) *DUP B-C* results probably from an in-frame deletion-fusion event between *DUP B* and *DUP C* and (ii) the coding sequence of the YAR033w ORF is 153 bp shorter than its counterpart in CLIB413 E1, because of the presence of a missense mutation in the position of the expected initiation codon.

Strain CLIB382 features seven *DUP240* paralogs (Figure 4). A perfect duplication (100% nucleotide identity) of a large part of *DUP C* is present upstream from *DUP A*. This duplicated copy of *DUP C* has been named *DUP X-C* since the beginning of its coding sequence (the first 66 nucleotides) shares the highest degree of similarity with the *DUP X* ORF located at the CLIB413 tandem

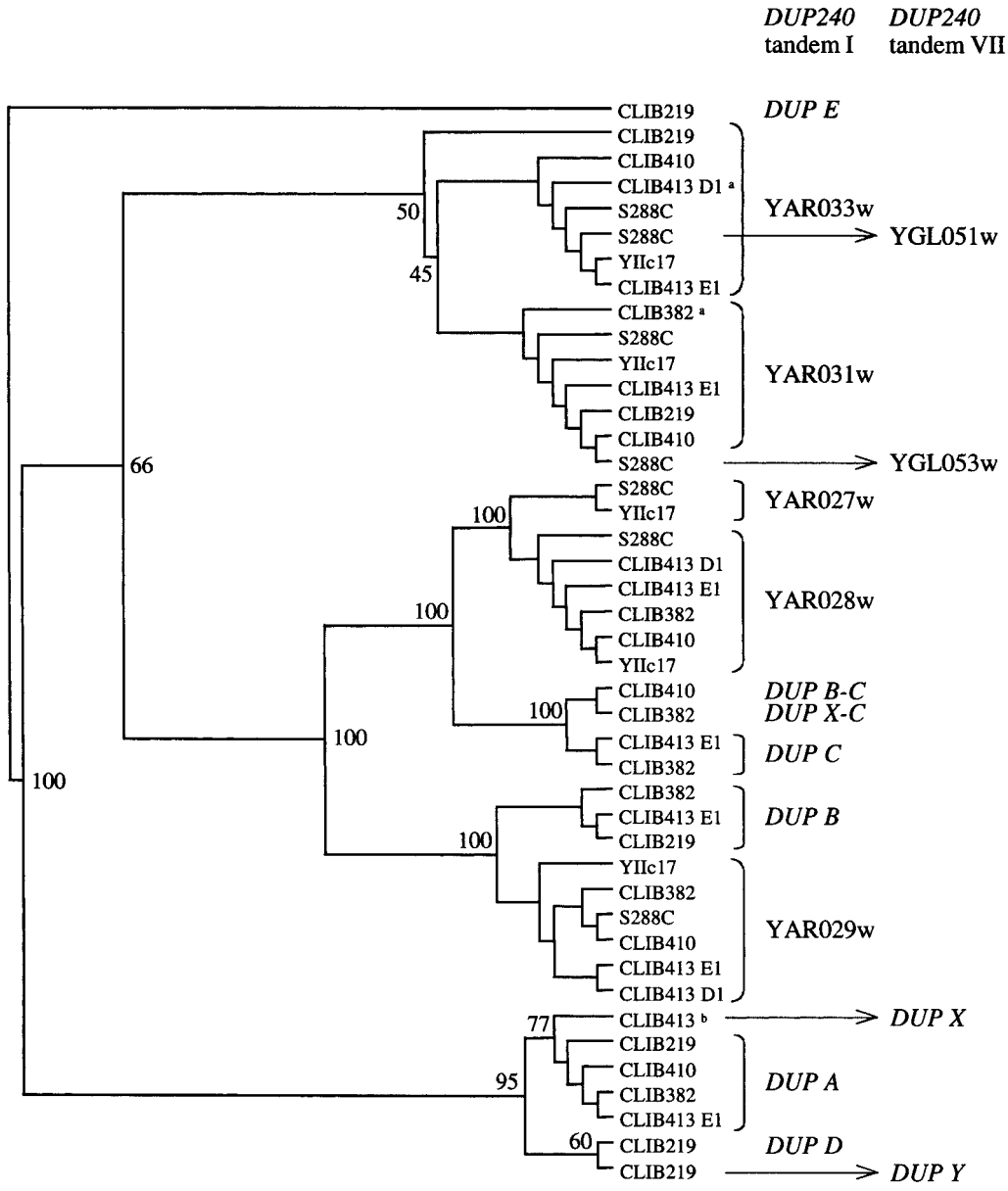

FIGURE 3.—Phylogenetic relationships between all *DUP240* ORFs identified at the tandem I and tandem VII loci in the 15 tested yeast strains. DNApars program of the PHYLIP phylogeny inference package was used. Bootstrap values for 100 replicates are indicated. Superscript a indicates chimeric ORFs YAR031-033w. Superscript b indicates that the *DUP X* ORF was identified in strains CLIB413, CLIB410, YIIc12, and YIIc17; only that of strain CLIB413 is shown.

VII locus. In addition, we found a chimeric YAR031-033w ORF truncated as a result of multiple point mutations.

Another type of polymorphism is apparent with the strain CLIB219 since six paralogs and one relic are detectable (Figure 4). Among the six *DUP240* ORFs, *DUP D* and *DUP E* are new paralogs not yet found in other genomes. *DUP D* is phylogenetically related to *DUP Y*, the paralog recovered at the tandem VII locus of the same strain (Figure 3). In contrast, *DUP E* clearly appears as an outgroup in the phylogenetic tree shown in Figure 3, so no hypothesis about its origin can be deduced. Notably, two missense mutations shorten the YAR033w 5′-coding sequence to 309 bp in length. Finally, analysis of the unusually long intergenic area between *DUP A* and *DUP E* (1512 bp instead of 425 bp on average) suggests that an ORF, closely related to YAR028w of CLIB413 and CLIB382, was previously located here, but the accumulation of numerous mutations led to the erosion of its coding sequence, forming a relic (FISCHER *et al.* 2001).

Identification of DNA sequences potentially implicated in intra- and interchromosomal rearrangements: With the aim of identifying the basis of the gene organization polymorphisms observed at the tandem I and tandem VII loci, we searched for direct repeats of DNA sequences that could play a role in recombination. The DNA sequences of *DUP240* tandem loci were systematically screened for such DNA repeats, but we report only data obtained for strain CLIB413 (Figure 5).

The presence at the CLIB413 tandem I locus of DNA motifs potentially involved in recombination suggests

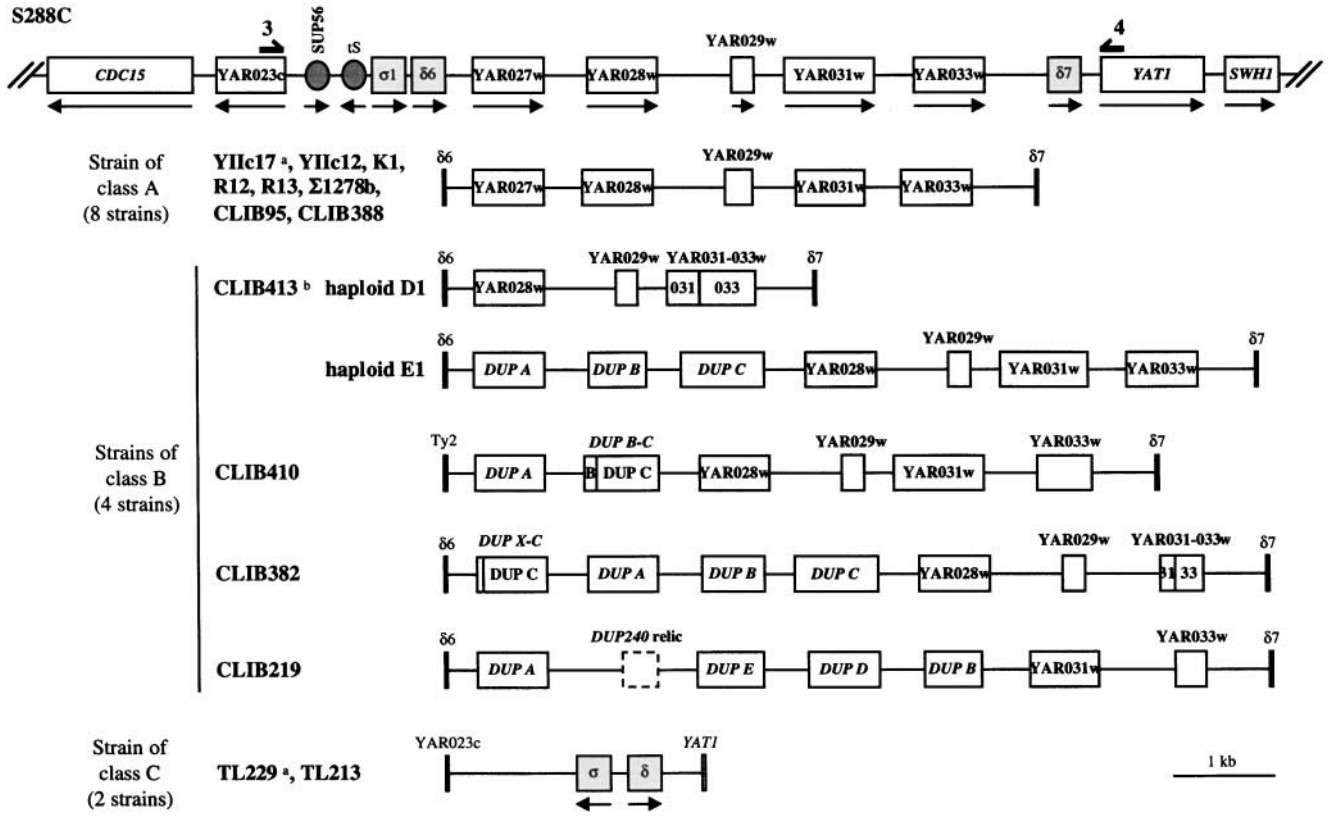

FIGURE 4.—Genetic organization of the tandem I locus in the 15 studied yeast strains (to scale except for the genetic elements upstream from δ6 and downstream from δ7 in the S288C map). Names of genes and ORFs are indicated inside or above open boxes. Genes that surround the tandem I locus are mentioned only on the map of the reference strain S288C. tRNA genes and long terminal repeats are represented by circles and shaded boxes, respectively. Arrows specify the orientation of genetic sequences. Primers 3 and 4 were used to amplify by PCR the genomic region located between YAR023c and *YAT1* in all the studied strains. The characteristics of the obtained PCR products allowed grouping of the strains into class A, B, or C. Superscript a indicates that among strains of classes A and C, only the tandem I loci of YIIc17 and TL229 were sequenced. Superscript b indicates that for the heterozygous strain CLIB413, the tandem I locus was sequenced from two haploid strains (named D1 and E1) derived from the sporulation of the diploid strain.

that the strain was previously homozygous at this locus, with the ORF structure of the E1 allele (Figure 5). Two intrachromosomal recombination events inside one tandem I locus could be the cause of the observed D1 allele ORF organization. A first event of nonhomologous recombination [nonhomologous end joining (NHEJ)], involving a 9-bp microhomology (CATGCAAAC), between *DUP A* and YAR028w could explain the loss of sequence upstream of the D1 YAR028w ORF; a second recombination event between YAR031w and YAR033w probably allowed the creation of the new paralog YAR031-033w in D1. In this latter case, the in-frame deletion-fusion event could have been achieved either with the TACAC CAG direct repeats by an NHEJ mechanism or through homologous recombination between a 78-bp DNA sequence repeat (S78). The high level of nucleotide identity is not restricted to the *DUP240* coding sequences but is also observed for some intergenic regions. For example, *DUP B-DUP C* block is 91.8% identical to the YAR029w-YAR031w block (Figure 5). This suggests that the duplication of a sequence unit encompassing at least two adjacent *DUP240* ORFs and their corresponding intergenic region have occurred at the tandem I locus of CLIB413 E1.

We considered the possible occurrence of interchromosomal events between the tandem I and tandem VII loci of CLIB413. Sequence units between these two ectopic loci share a high level of nucleotide identity (90.8 and 98.6%; Figure 5). This suggests that, during genome evolution, exchange(s) of genetic material between the tandem I and tandem VII loci led to the presence of the same gene organization in these two nonallelic loci, probably the one recovered in the tandem I region. Such interchromosomal rearrangements may have been followed by intrachromosomal deletion event(s) at the tandem VII locus to yield the present genetic map comprising only *DUP X*. This hypothesis is supported by the presence of the microhomology ATATCGATGCGC at the potential deletion junctions (in *DUP A* and YAR033w and also in *DUP X*; Figure 5).

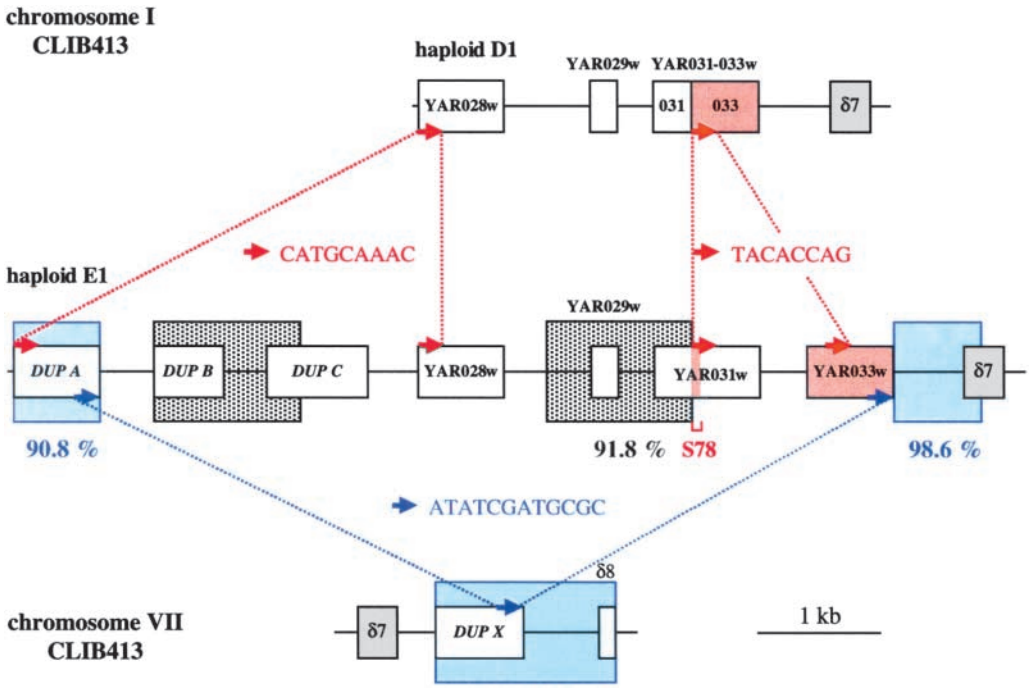

FIGURE 5.—Schematic of the DNA sequences potentially implicated in chromosomal rearrangements at the tandem I and tandem VII loci of the heterozygous CLIB413 strain. Intra- and interchromosomal duplicated sequence units are represented by stippled and blue boxes, respectively. Percentages show the level of nucleotide identity between these duplicated units. Red and blue arrows indicate the positions of very short repeated DNA sequences that could be involved in nonhomologous recombinations inducing deletion events. The in-frame deletion-fusion event probably at the origin of the YAR031-033w ORF could also have been achieved by homologous recombination, since a 78-bp sequence (S78) in YAR031w is 97% identical to the corresponding sequence in YAR033w. Red boxes correspond to YAR033w sequences.

DISCUSSION

Extensive polymorphism for the *DUP240* tandem loci: We documented the high degree of intraspecies polymorphism for the structural organization and sequence of the tandemly repeated *DUP240* ORFs in *S. cerevisiae*. This polymorphism is restricted to the *DUP240* loci since the synteny is always conserved in the flanking regions. Interestingly, we also observed some cases of heterozygosity in the gene organization of the tandem loci, a situation rarely reported since most of the genetic studies in yeast are performed on haploid or homozygous diploid cells. Another example is given by BROWN *et al.* (1998) who analyzed one population of diploid yeast that underwent 450 generations of glucose-limited growth. They found that the predominant cell type at the end of the assay shared a single copy of each of the target genes *HXT6* and *HXT7* (parental genotype) on one chromosome and multiple duplicated *HXT7/HXT6* copies (recombinant genotype) on the homologous chromosome.

Expansion and contraction of the *DUP240* family: The *DUP240* tandem loci are a breeding ground of new *DUP240* paralogs. Indeed, the analysis of duplicated blocks (such as *DUP B-DUP C* and YAR029w-YAR031w in CLIB413 E1) strongly suggests that new *DUP240* ORFs can be generated via the duplication of a sequence unit composed of ORFs and their corresponding intergenic regions. Such blocks have also been observed by BROWN *et al.* (1998) with the *HXT* genes. These authors proposed that the increase in gene copy number results from unequal crossover between sister chromatids, a hypothesis that can also be envisaged in the case of the tandemly repeated *DUP240* ORFs. We also identified, at the tandem I locus of some class B strains, different chimeric ORFs (like YAR031-033w), which probably originated from two distinct preexistent *DUP240* ORFs by deletion-fusion events. In this latter case, the creation of a new paralog at the *DUP240* tandem loci is accompanied by the loss of other *DUP240* ORFs. The disappearance of ORFs could also be the result of other evolutionary processes. A deletion event (without fusion) could explain, for example, the situation observed for CLIB413 D1 where three paralogs were apparently lost (Figure 5). Furthermore, we identified one relic of a *DUP240* ORF in an unusually long intergenic region (tandem I locus in CLIB219). Such an occurrence of genetic drift by point mutations has also led to the loss of parts of the coding sequence of other *DUP240* ORFs. For instance, the YAR029w ORF is a derivative of the *DUP B* paralog through the fixation of several mutations (missense, nonsense, and frameshift) affecting the 5′-terminal part of the coding sequence.

The *DUP240* tandem loci probably act as hot spots for ectopic recombination: We identified DNA motifs

that may represent the signature of the molecular events at the origin of chromosomal rearrangements in the *DUP240* tandem regions. The absence of a *DUP240* ORF at the tandem I and/or tandem VII loci (strains of classes 3 and C) could be explained by ectopic recombination between two homologous solo LTR sequences that surround each *DUP240* tandem locus (Figures 2 and 4). Indeed, Ty elements or solo LTRs are well known to induce chromosomal deletion, duplication, translocation, and inversion events by allelic or ectopic recombination in yeast (Roeder *et al.* 1984; Kupiec and Petes 1988; Rachidi *et al.* 1999). It is likely that ectopic recombination also plays a major role in the creation of the newly identified *DUP240* paralogs with a chimeric structure. The high level of nucleotide identity among most members of the *DUP240* family further suggests that a homologous recombination mechanism is involved in the in-frame deletion-fusion events at the origin of these chimeric ORFs. Nevertheless, we cannot exclude a nonhomologous recombination (NHEJ) mechanism, possibly involving microhomology stretches (2–20 bp; Mezard *et al.* 1992; Kramer *et al.* 1994; Mezard and Nicolas 1994). Involvement of very short DNA tandem repeats in deletions leading to new genes encoding a fusion protein was experimentally demonstrated previously in yeast by Welcker *et al.* (2000). Thus, *DUP240* paralogs could represent another example of multigene families involved in a recombination hot spot, such as the previously described murine immunoglobulin (Raynard *et al.* 2002) and human β-globin genes (Schneider *et al.* 2002).

Birth-and-death model of evolution for the *DUP240* family: We have shown previously that the three isolated *DUP240* ORFs (YAR023c, YCR007c, and YHL044w) evolve by nucleotide substitutions and allelic recombination events (Leh-Louis *et al.* 2004). The accumulation of multiple point mutations in coding sequences also takes an important part in the evolution of tandemly repeated *DUP240* ORFs, but in this case, nonallelic recombination events are the predominant forces at the origin of the extensive structural polymorphism that is observed. For both tandem and solo *DUP240* loci, we do not observe homogenization of ORF sequences, but rather, a propensity toward the accumulation of nucleotide substitutions and toward recombination. Therefore, the *DUP240* paralogs evolve by a birth-and-death process, which allows an increase in genetic diversity.

We do not believe the *DUP240* gene family represents a new form of selfish DNA. The major argument favoring this idea is that the protein encoded by YCR007c, the solo *DUP240* ORF present in all tested yeast strains, is under strong selective pressure (Leh-Louis *et al.* 2004). Poirey *et al.* (2002) suggested that the Dup240 proteins have different specialized functions since they have different subcellular localizations and nonredundant interacting partners. Thus, the *DUP240* family is another intriguing example of an *S. cerevisiae* family of nonessential genes with a high level of nucleotide identity but nevertheless distinct functions, like the hexose transporter gene family (Kruckeberg 1996; Wieczorke *et al.* 1999). Considering that *DUP240* orthologs have been identified only in species of the Saccharomyces *sensu stricto* group, a working hypothesis is that the birth of a new *DUP240* paralog could provide a new function probably highly specialized and specific to these species. The *DUP240* gene family can be considered as a genetic marker of evolution and is a model to study genome plasticity.

We thank especially Marc Sultan for technical assistance and Stéphane Vuilleumier and Mikael Dubow for careful reading of the manuscript. We are grateful to Michel Aigle, Claude Gaillardin, and Huu-Vang Nguyen for providing *S. cerevisiae* strains and Philippe Hammann and Malek Alioua for automated DNA sequencing in the Strasbourg Centre National de la Recherche Scientifique (CNRS)/Institut de Biologie Moléculaire des Plantes department facilities. This work was supported in part by an E.U. Comprehensive Yeast Genome Database grant (QLRI CT 1999 01333) and by the Génolevures-2 sequencing consortium GDR CNRS 2354. B.W. is supported by a grant from the French Ministère de l'Education Nationale, de la Recherche et de la Technologie.

LITERATURE CITED

Brown, C. J., K. M. Todd and R. F. Rosenzweig, 1998 Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol. Biol. Evol. **15:** 931–942.

Clark, A. G., 1994 Invasion and maintenance of a gene duplication. Proc. Natl. Acad. Sci. USA **91:** 2950–2954.

Devereux, J., P. Haeberli and O. Smithies, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

Dujon, B., 1998 European Functional Analysis Network (EUROFAN) and the functional analysis of the *Saccharomyces cerevisiae* genome. Electrophoresis **19:** 617–624.

Emanuel, B. S., and T. H. Shaikh, 2001 Segmental duplications: an 'expanding' role in genomic instability and disease. Nat. Rev. Genet. **2:** 791–800.

Felsenstein, J., 1989 PHYLIP—Phylogeny inference package (version 3.2). Cladistics **5:** 164–166.

Feuermann, M., J. de Montigny, S. Potier and J. L. Souciet, 1997 The characterization of two new clusters of duplicated genes suggests a 'Lego' organization of the yeast *Saccharomyces cerevisiae* chromosomes. Yeast **13:** 861–869.

Fischer, G., C. Neuveglise, P. Durrens, C. Gaillardin and B. Dujon, 2001 Evolution of gene order in the genomes of two related yeast species. Genome Res. **11:** 2009–2019.

Goffeau, A., R. Aert, M. L. Agostini-Carbone, A. Ahmed, M. Aigle *et al.*, 1997 The yeast genome directory. Nature **387** (Suppl.): 5–105.

Hoffman, C. S., and F. Winston, 1987 A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene **57:** 267–272.

Kramer, K. M., J. A. Brock, K. Bloom, J. K. Moore and J. E. Haber, 1994 Two different types of double-strand breaks in *Saccharomyces cerevisiae* are repaired by similar *RAD52*-independent, nonhomologous recombination events. Mol. Cell. Biol. **14:** 1293–1301.

Kruckeberg, A. L., 1996 The hexose transporter family of *Saccharomyces cerevisiae*. Arch. Microbiol. **166:** 283–292.

Kupiec, M., and T. D. Petes, 1988 Allelic and ectopic recombination between Ty elements in yeast. Genetics **119:** 549–559.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature **409:** 860–921.

Leh-Louis, V., B. Wirth, L. Despons, S. Wain-Hobson, S. Potier *et*

al., 2004 Differential evolution of the *Saccharomyces cerevisiae DUP240* paralogs and implication of recombination in phylogeny. Nucleic Acids Res. **32:** 2069–2078.

Liao, D., 1999 Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. **64:** 24–30.

Llorente, B., A. Malpertuy, C. Neuveglise, J. de Montigny, M. Aigle *et al.*, 2000 Genomic exploration of the hemiascomycetous yeasts. 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. FEBS Lett. **487:** 101–112.

Mezard, C., and A. Nicolas, 1994 Homologous, homeologous, and illegitimate repair of double-strand breaks during transformation of a wild-type strain and a *rad52* mutant strain of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **14:** 1278–1292.

Mezard, C., D. Pompon and A. Nicolas, 1992 Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell **70:** 659–670.

Nei, M., X. Gu and T. Sitnikova, 1997 Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. USA **94:** 7799–7806.

Nguyen, H. V., and C. Gaillardin, 1997 Two subgroups within the *Saccharomyces bayanus* species evidenced by PCR amplification and restriction polymorphism of the non-transcribed spacer 2 in the ribosomal DNA unit. Syst. Appl. Microbiol. **20:** 286–294.

Poirey, R., L. Despons, V. Leh, M. J. Lafuente, S. Potier *et al.*, 2002 Functional analysis of the *Saccharomyces cerevisiae DUP240* multigene family reveals membrane-associated proteins that are not essential for cell viability. Microbiology **148:** 2111–2123.

Rachidi, N., P. Barre and B. Blondin, 1999 Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. Mol. Gen. Genet. **261:** 841–850.

Raynard, S. J., L. R. Read and M. D. Baker, 2002 Evidence for the murine IgH mu locus acting as a hot spot for intrachromosomal homologous recombination. J. Immunol. **168:** 2332–2339.

Roeder, G. S., M. Smith and E. J. Lambie, 1984 Intrachromosomal movement of genetically marked *Saccharomyces cerevisiae* transposons by gene conversion. Mol. Cell. Biol. **4:** 703–711.

Ryu, S. L., Y. Murooka and Y. Kaneko, 1996 Genomic reorganization between two sibling yeast species, *Saccharomyces bayanus* and *Saccharomyces cerevisiae*. Yeast **12:** 757–764.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sankoff, D., 2001 Gene and genome duplication. Curr. Opin. Genet. Dev. **11:** 681–684.

Schneider, J. A., T. E. Peto, R. A. Boone, A. J. Boyce and J. B. Clegg, 2002 Direct measurement of the male recombination fraction in the human beta-globin hot spot. Hum. Mol. Genet. **11:** 207–215.

Sumbu, Z. L., P. Thonart and J. Bechet, 1983 Action of patulin on a yeast. Appl. Environ. Microbiol. **45:** 110–115.

Welcker, A. J., J. de Montigny, S. Potier and J. L. Souciet, 2000 Involvement of very short DNA tandem repeats and the influence of the *RAD52* gene on the occurrence of deletions in *Saccharomyces cerevisiae*. Genetics **156:** 549–557.

Wieczorke, R., S. Krampe, T. Weierstall, K. Freidel, C. P. Hollenberg *et al.*, 1999 Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. FEBS Lett. **464:** 123–128.

Wolfe, K. H., and D. C. Shields, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387:** 708–713.

Communicating editor: M. Johnston