

Duplication processes in *Saccharomyces cerevisiae* haploid strains

Joseph Schacherer, Jacky de Montigny*, Anne Welcker, Jean-Luc Souciet and Serge Potier

CNRS/ULP, Laboratoire de Microbiologie et Génétique, Strasbourg, France

Received June 20, 2005; Revised September 21, 2005; Accepted October 17, 2005

ABSTRACT

Duplication is thought to be one of the main processes providing a substrate on which the effects of evolution are visible. The mechanisms underlying this chromosomal rearrangement were investigated here in the yeast *Saccharomyces cerevisiae*. Spontaneous revertants containing a duplication event were selected and analyzed. In addition to the single gene duplication described in a previous study, we demonstrated here that direct tandem duplicated regions ranging from 5 to 90 kb in size can also occur spontaneously. To further investigate the mechanisms in the duplication events, we examined whether homologous recombination contributes to these processes. The results obtained show that the mechanisms involved in segmental duplication are *RAD52*-independent, contrary to those involved in single gene duplication. Moreover, this study shows that the duplication of a given gene can occur in *S.cerevisiae* haploid strains via at least two ways: single gene or segmental duplication.

INTRODUCTION

Susumu Ohno was the first author to present gene duplication as a key to molecular evolution (1). In the light of the data obtained by performing global genome analysis, the importance of gene duplication as an evolutionary process is by now widely recognized, however. One of the most important aspects of molecular evolution is that it leads to the emergence of genes with new functions. Genes encoding novel functions are often derived from preexisting genes by modification of their structure or regulation through a process called neofunctionalization (2). The initial step in this process is a gene duplication step, which leads to redundancies in the genetic material followed by the divergence of one or both copies.

Many of the genomes completely sequenced so far show traces of duplication events. Examples of these duplications

can be found in yeast species. Approximately 40% of the genes in *Saccharomyces cerevisiae* result from duplication events and form families consisting of two to more than twenty paralogous genes (3–5). The presence of redundant sequences of this kind can be partly explained by a whole genome duplication (WGD) that would have been subsequently reduced to the disomic state by deletions and concomitantly rearranged by translocations (6–8). Nevertheless, other molecular mechanisms liable to generate duplications of DNA regions need to be considered to account for the presence of a duplicated copy. Gene duplication can involve either a single gene (9), or segmental (10,11) or single chromosome duplications (12).

To investigate the mechanisms involved in the duplication events at work in *S.cerevisiae*, we used a genetic screening based on the *URA2* gene, which can be used to carry out the positive selection of spontaneous chromosomal rearrangements (13,14) (Figure 1). In the *ura2 15-30-72* mutant strain, the ATCase domain is inactive, resulting in a uracil auxotroph. The functional reactivation of the ATCase domain suffices to generate *Ura*⁺ prototrophs. One type of reactivation event is the duplication of the ATCase coding sequence and its insertion elsewhere in to the genome under the control of a resident promoter (15). Starting with a set of eight haploid revertants carrying a duplication events, it was proposed to determine the chromosomal location, the size and the insertion sites of the duplicate regions with a view to establishing the underlying gene duplication mechanisms. In a previous study, 4 strains among this set of revertants were found to result from a single gene duplication through a retroposition process (9). In the present study, we focussed on other four revertants in which a segmental duplication events leads to the generation of a *Ura*⁺ prototroph. Analysis of the duplicated segments showed that they consist of a variable number of contiguous ORFs which are coduplicated with the ATCase coding region. These segments range from 5 to 90 kb in size.

To further investigate the mechanisms involved in these duplication events, we studied the contribution of homologous recombination to these processes. For this purpose, we examined how the recombination gene *RAD52* affected these mechanisms. Starting with an *ura2 15-30-72 rad52Δ* strain, six revertants carrying a duplication event were selected and

*To whom correspondence should be addressed. Tel: +33 0 3 90 24 20 23; Fax: +33 0 3 90 24 20 28; Email: demontigny@gem.u-strasbg.fr

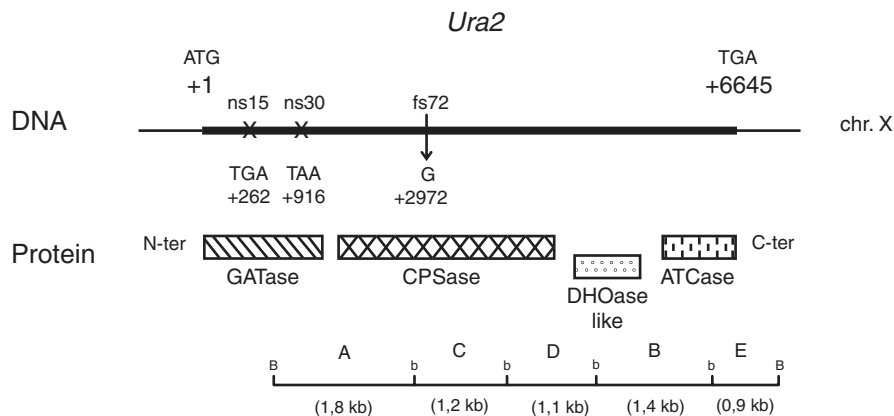


Figure 1. Correspondence between the simplified restriction map of the *URA2* locus and protein domains. The thick line indicates the the *URA2* coding sequence (restriction site: b, BglIII; B, BamHI). A, C, D, B and E denote the BamHI–BglIII subfragments, ns15 and ns30 the positions of the nonsense mutations, and fs72 the position of the frameshift mutation (GATase, glutamine amidotransferase; CPSase, carbamoylphosphate synthetase; DHOase-like, dihydroorotase-like; ATCase, aspartate transcarbamylase).

analyzed. The duplication events corresponded to segments ranging from 5 to 10 kb in size. No single gene duplications were selected in the *rad52Δ* context.

The genetic screening method based on the *URA2* gene provides a powerful tool for selecting duplication events and investigating the mechanisms involved in these events in *S.cerevisiae*. The genomic rearrangements described here constitute a new example of spontaneous segmental duplication events. In addition, the possible role of homologous recombination in the duplication process is discussed in the light of these new molecular data.

MATERIALS AND METHODS

Yeast strains and plasmids

Yeast strain *ura2 15-30-72* is an isogenic derivative of the laboratory strain FL100 (ATCC 28 583). Strain *ura2 15-30-72 rad52Δ* (14) was constructed by effecting a single-step gene replacement of the *ura2 15-30-72* strain with a *kanMX4-rad52* PCR fragment (16).

The integrative pFLB, pFLC and pFLD plasmids were constructed by inserting the corresponding BglIII–BglIII subfragment (Figure 1) in to the unique BamHI site of the pFL35 plasmid (17).

Selection of Ura⁺ revertants from the *ura2 15-30-72* mutant strain

Ura⁺ revertants were isolated as described by Roelants *et al.* (13) from the *ura2 15-30-72* and *ura2 15-30-72 rad52Δ* strains.

Determination of the 5' insertion site by plasmid integration and excision

The 5' junction of the duplicate regions were cloned using a plasmid integration/excision strategy. For this purpose, we first located the beginning of the duplicated fragment, using the BglIII–BamHI restriction map (Figure 1). Depending on the beginning of the duplicated segment, one integrative plasmid (pFLB, pFLC or pFLD) was introduced downstream of the

modified subfragment (B, C or D). If the starting point of the duplicated region was located in the C subfragment, the pFLD plasmid containing the D fragment was integrated (Figure 1). After the transformation step, uracil auxotroph colonies (carrying the plasmid on the duplicated copy) were selected. DNA extraction, HindIII digestion and ligation were carried out in order to recover a plasmid containing the unknown junction, which was sequenced.

Transformation of yeast and bacteria

Yeasts were transformed using the method developed by Becker and Guarante (22). *Escherichia coli* transformation was performed as described by Dower *et al.* (23).

Southern blot analysis

Total DNA from *S.cerevisiae* was prepared as described by Hoffman and Winston (20). Restriction endonuclease digestion steps were carried out as described by the manufacturers. DNA blots were prepared from pulsed-field gel electrophoresis (PFGE) and conventional agarose gels by transferring DNA to Hybond N⁺ membrane (Amersham). DIG-labeled DNA probes were prepared using the DNA labeling and detection kit (Roche).

PCR amplification, DNA sequencing and sequence analysis

Primer sequences used for PCR amplification and sequencing were chosen on the basis of the published genomic sequence of S288C. DNA fragments were obtained by performing PCR amplification using *Taq* DNA polymerase from Q-BIOgene. PCR conditions were those described by the manufacturers. Double strand DNA fragments obtained after PCR amplification were purified using a MicroSpin Column S-400 (Amersham). DNA sequencing was performed on the PCR-purified fragments using the method described by Sanger *et al.* (21). The sequencing chemistry used was *AmpliTaq* FS DNA polymerase and BIGDYE TM terminators (version1). Sequence reactions were analyzed with an Applied Biosystems 373XL sequencer. BLAST analysis was performed after sequencing the PCR product to determine the

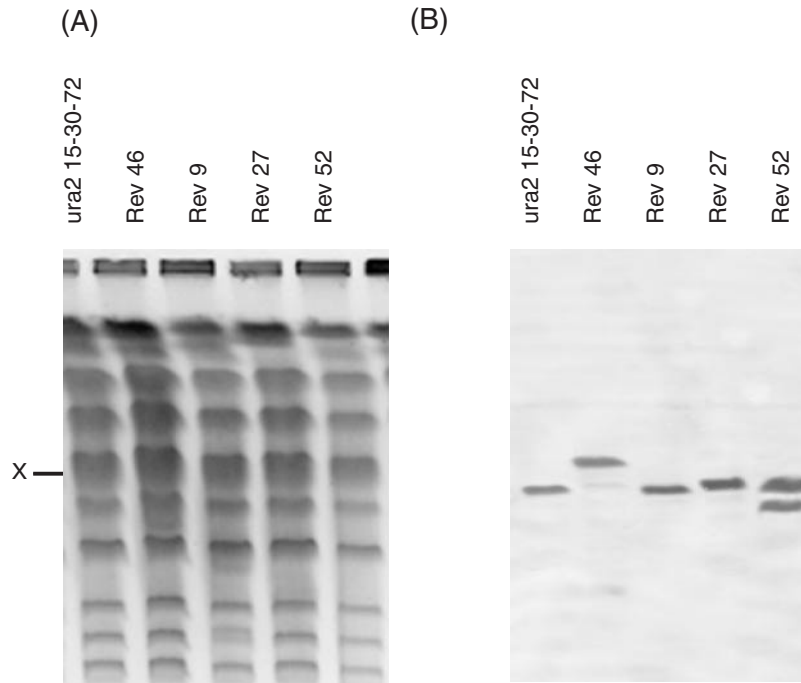


Figure 2. Chromosome location of the duplication events. (A) PFGE of the intact chromosomes from the ancestral strain and each of the revertant strains. (B) Hybridization of the corresponding Southern blot with the ATCase probe.

exact boundaries of the duplication events. This was done with the SGD database.

PFGE

Chromosomal DNA was prepared as described by Carle and Olson (22). Chromosomes were separated on a 1% agarose gel (Pharmacia) in a 0.5× TBE buffer at 7 V/cm for 22 h with a pulse time of 45 s and an angle of 120°, using a Bio-Rad CHEF-DIII mapper apparatus. The gel was stained with Ethidium bromide to identify the chromosomal pattern specific to each strain.

Comparative genomic hybridization on to microarray

Total genomic DNA of mutants and parent strains was prepared with the Qiagen genomic TIP-100 and hybridized against yeast whole genome arrays (YG-S98) from Affymetrix. Labeling, hybridization and detection steps were performed at the Affymetrix Platform at the 'Génopole Alsace-Lorraine' (IGBMC Illkirch, France). Arrays were analyzed and genomic ratios were calculated using the Affymetrix GeneChip software program.

RESULTS

Characterization and chromosomal location of the duplication events

Starting with an initial *ura2 15-30-72* mutant strain, eight spontaneous revertants carrying a duplication of the ATCase sequence were isolated at a low frequency of $\sim 0.3 \times 10^{-10}$ (13,14). Whereas four revertants were found to result from a single gene duplication, the other four revertants (Rev 9,

Rev 27, Rev 46 and Rev 52) remained uncharacterized and were further analyzed.

The chromosomal location of these duplicated copies was detected by performing PFGE followed by Southern hybridization with an ATCase probe (Figure 2). In the Rev 52 strain, we detected the chromosome X carrying the initial mutated allele, as well as chromosome XI containing the duplicated copy. This strain results from an interchromosomal duplication event in which the chromosome XI acts as a recipient sequence for the duplicated ATCase segment. On the other hand, in Rev 9, Rev 27 and Rev 46, the fact that we detected only chromosome X shows that an intrachromosomal duplication event had occurred. In addition, the hybridization of the karyotypes with the ATCase probe also reflect the occurrence of a 100 kb increase in the size of chromosome X in the Rev 46 strain.

Determination of the size of the duplicated region

The size increase in chromosome X, in the Rev 46 strain, suggests that some genes adjacent to the *URA2* locus may have been coduplicated with the ATCase coding region. In order to test this hypothesis, we estimated the size of the duplicated segment by performing Comparative Genomic Hybridization (CGH) on to microarrays. Total DNA from each revertant was hybridized on to microarrays and compared with the initial *ura2 15-30-72* mutant strain. Genomic ratios were determined for each open reading frame (ORF) between the revertants and the initial *ura2 15-30-72* mutant strain and plotted as a function of their chromosomal location. The results show that in each of the revertants, a region downstream of the *URA2* gene (*YJL130c*) has been duplicated (Figure 3). These regions correspond to segments formed

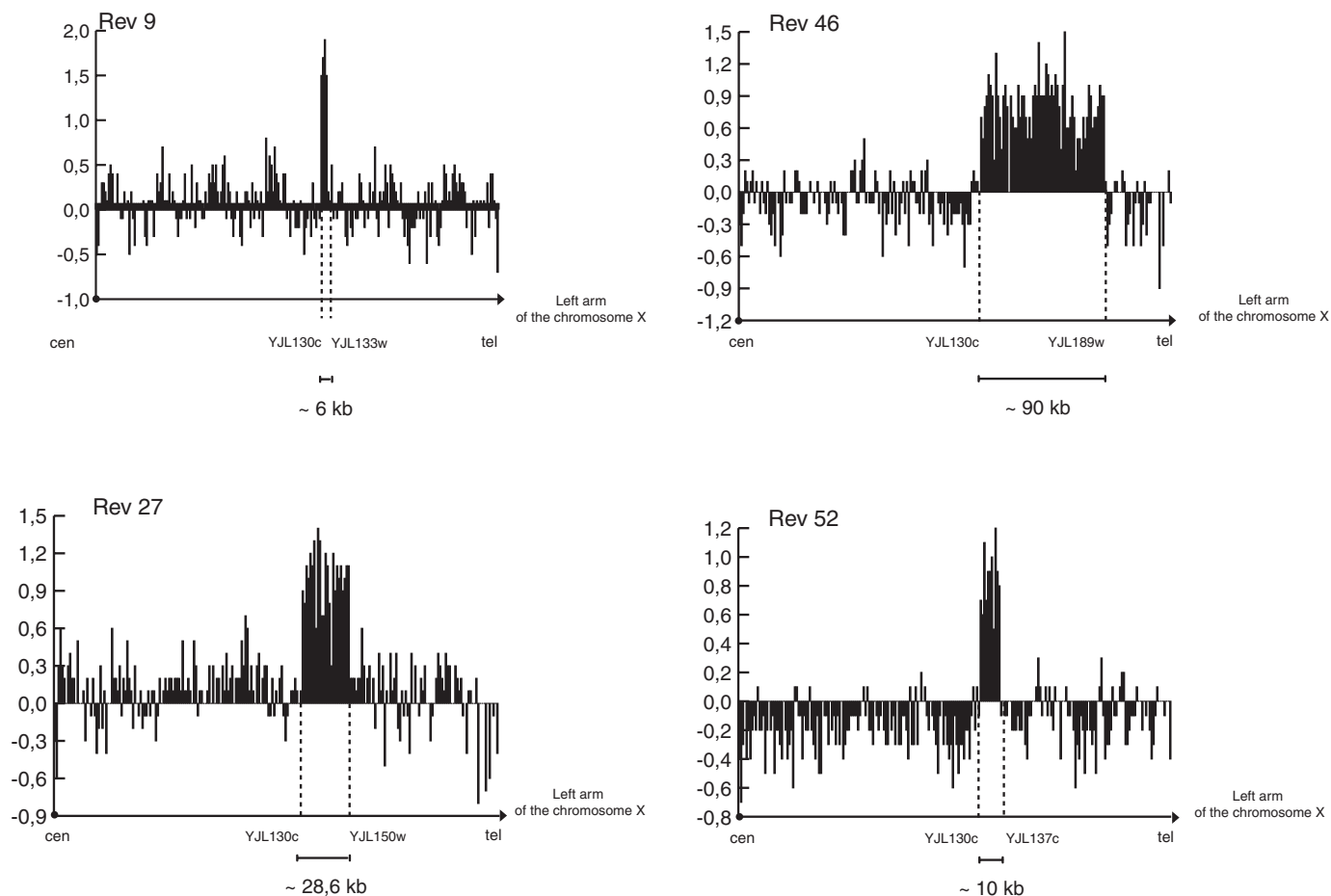


Figure 3. Size of the duplicated region. Diagrams give the profiles obtained upon performing CGH on to microarray. The x-axis correspond to ORFs from the left arm of the chromosome X running from YJL001w to YJL204c. The y-axis correspond to the genomic ratio calculated between the revertant and *ura2 15-30-72* strain.

by a variable number of contiguous ORFs which are coduplicated along with the ATCase coding region. These segments ranging from 5 to 90 kb in size cover a region situated between ORFs YJL131c and YJL190c. These results were confirmed by conventional Southern blot analysis (data not shown).

Sequence analysis of the insertion sites of the duplicated regions

The exact location and orientation of the duplicated segments and their insertion sites were determined by sequencing the 5' and 3' flanking regions.

The 5' junctions of the duplicated regions were cloned using a plasmid insertion and excision strategy. The first step consisted in integrating a plasmid downstream of the starting point of the duplicated region. After this integration step, HindIII digestion and ligation of the digestion product were carried out in order to recover a plasmid containing the unknown junction via an *E. coli* transformation and plasmid extraction procedure. The unknown junction was then sequenced and used to perform BLAST queries against the *Saccharomyces* Genome Database (<http://seq.yeastgenome.org/cgi-bin/SGD/nph-blast2sgd>).

Analysis of the junction indicated that the duplicated copy begins, in each case, in a region located downstream of the

three mutations. The duplicated part of the *URA2* gene is either inserted in to a gene or in to an intergenic region (Figure 4). In the Rev 27 and the Rev 46 strains, the duplicated segment is inserted in to a gene. The 5' junctions consist in an in frame fusion with the non-essential gene *SNA3* (YJL151c) and *RPS22A* (YJL190c), respectively. Since the fusion preserves the coding frame, chimeric genes composed of the 5' end of ORFs (*SNA3* or *RPS22A*) and the 3' duplicated part of the *URA2* gene are formed. In the other two revertants (Rev 9 and Rev 52), the ATCase coding region is fused with an intergenic region which is located downstream the gene *MRS3* (YJL133w) and *GLG1* (YKR058w), respectively. In addition, the junction in Rev 46 shows the presence of a short region of identity (7 bp) between the *RPS22A* gene and the 5' duplicated part of the *URA2* gene (Figure 5). By contrast, in the three other strains (Rev 9, Rev 52 and Rev 27), the junctions do not show any such short regions of homology.

To exactly determine the 3' insertion sites, it was assumed that the integration of the duplicated segments would move the second part of the insertion region directly downstream of the duplication. Based on the 5' chromosomal insertion sites and the 3' ends of the duplicated regions determined by a CGH, we designed a set of oligonucleotides to perform PCR amplification as a means of testing this hypothesis.

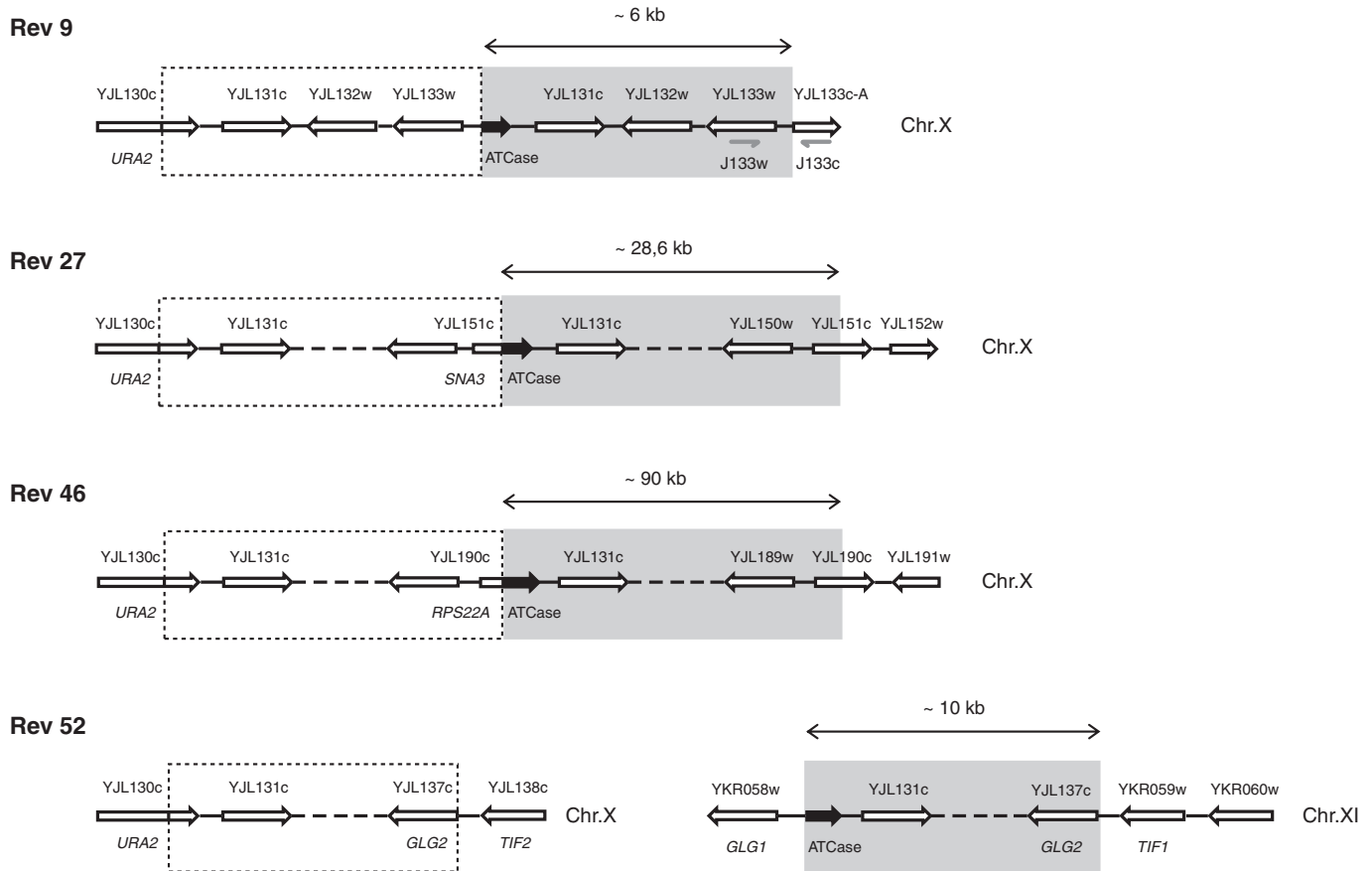


Figure 4. Chromosomal insertion of the segmental duplicated regions. The grey boxes correspond to the duplicated elements whereas the dotted boxes correspond to the original region. The black arrow indicates the duplicated part of the *URA2* gene. The numbers placed above the duplicated structure correspond to the size of the duplicated segment.

As an example, in the strain Rev 9, the duplicated segment is inserted in to the intergenic region downstream of ORF YJL133w (*MRS3*) (Figure 4). Moreover, we established that this segment is 5 kb long and stretches from the duplicated part of the *URA2* gene (YJL130c) to ORF YJL133w (Figure 3). The following two oligonucleotides were designed: J133w specific to ORF YJL133w (the end of the duplicated fragment) and J133c specific to ORF YJL133c-A (the contiguous ORF at the 5' insertion site) in order to amplify the junction. The same strategy was used with the other 3 strains (Rev 27, Rev 46 and Rev 52).

In each case, the oligonucleotides made it possible to amplify and sequence a DNA fragment. The results show that the regions located downstream of the duplicated segment correspond to the adjacent chromosomal region of the 5' insertion site in the *ura2 15-30-72* strain.

In revertants Rev 9, Rev 27 and Rev 46, the intrachromosomal duplications correspond to the chromosomal region located between the *URA2* gene and the insertion site (Figure 4). In addition, they conserve the same orientation as the original copy. All these data show that the selected events corresponded to the duplicated segments repeat as a direct tandem (Figure 4).

In the Rev 52 strain, the interchromosomal duplicated segment is inserted in to one of the two regions of duplicated

block 41, which is a trace of the ancestral WGD process (6). Block 41 is composed of two distinct regions (<http://acer.gen.tcd.ie/~khwolfe/yeast/>). The first of these is located on chromosome X between ORFs YJL078c (*YUR1*) and YJL139c (*PRY3*), encompassing the *URA2* gene (YJL130c). The second one, which acts as the recipient for the duplicated segment, is located on chromosome IX between ORFs YKR061w (*KTR2*) and YKR013w (*PRY2*) inserted between the *GLG1* and *TIF1* genes. This region does not contain a paralog of the *URA2* gene but contains the *GLG1* (YKR058w) and *TIF1* (YKR059w) genes which are paralogs of the *GLG2* (YJL137c) and *TIF2* (YJL138c) genes, respectively (Figure 4).

Deletion of the *RAD52* gene results in the suppression of the selection of the single gene duplication events

Molecular analysis of the revertants isolated using the *URA2* genetic screening system made it possible to define two types of duplication events: single gene duplications and segmental duplications. In the case of single gene duplication, the results described in Schacherer *et al.* (9) strongly suggest that single ATCase duplication results from retroposition, which means that the corresponding mRNA is retrotranscribed in to complementary DNA (cDNA) and then inserted in to the genome. Since the duplicated ATCase is embedded in Ty sequences,

Rev 9	168890	<i>URA2</i>
	AAAACCATGCAGGTAGAAT	
	AATTTATAATGTAAGTATT	
<i>MRS3</i>	159787	
Rev 27	168898	<i>URA2</i>
	TATCCAGTTTTGGTACGTC	
	GACAACTTGTATGGGAATG	
<i>SNA3</i>	136557	
Rev 46	169127	<i>URA2</i>
	GAATTTGA TTGGTGTGCTGTTAC	
	CACGGTTACATTGGT GAATTTGA	
<i>RPS22A</i>	75150	
Rev 52	166807	<i>URA2</i>
	GGCGCCAAAGGAAAATGCA	
	ATATACAAGATGTCTGAAG	
<i>GLG1</i>	554638	

Figure 5. Sequence of the 5' junction. The numbers placed above and below the sequence correspond to the chromosomal coordinates. The bold face type indicates a short region of identity.

the chromosomal insertion process could either have involved homologous recombination with endogenous Ty sequences or have been mediated by Ty integrase. In the case of segmental duplication, the mechanism generally suggested to account for the formation of duplicated regions is an unequal crossing over (UCO) mechanism (23). To understand how homologous recombination is involved in the integration of the cDNA and the segmental duplication events, we examined how these mechanisms are affected by the *RAD52* gene, which is essential to the homologous recombination pathway, by investigating the *RAD52* dependence of the two types of duplication events observed in our genetic system.

Starting with a *ura2 15-30-72 rad52Δ* strain, six revertants carrying a duplication events were isolated at a low frequency of $\sim 0.7 \times 10^{-10}$ and characterized (14). In each of these six revertants, the BamHI restriction pattern detected upon performing Southern blot hybridization using an ATCase coding sequence probe consisted of two bands: one corresponding to the *ura2 15-30-72* resident copy and the other to a duplicated copy inserted in to a novel chromosomal region. The chromosomal location of each duplicated copy was detected by performing PFGE followed by Southern hybridization with an ATCase probe. In the six revertants, the fact that only one chromosome was detected suggests that an intrachromosomal duplication event occurred on chromosome X. To determine the type of events occurring in the six revertants, we estimated the size of the duplicated segment by performing CGH on to microarrays. The segments are formed by a variable number of contiguous ORFs, which are coduplicated with the ATCase coding region (Table 1). These segments ranging from 5 to

Table 1. Duplication events observed in revertants selected with the *ura2 15-30-72 rad 52Δ* strain

Revertants	Chromosomal location	Left border ^a	Right border ^a	Segment size
173	X	YJL130c	YJL137c	10 kb
185	X	YJL130c	YJL133c	5 kb
193	X	YJL130c	YJL135c	7.5 kb
195	X	YJL130c	YJL135c	7.5 kb
196	X	YJL130c	YJL135c	7.5 kb
198	X	YJL130c	YJL133c	5 kb

^aORFs located at the right and left borders were determined by a CGH.

10 kb in size cover a region situated between ORFs YJL131c and YJL137c. In addition, the 5' junctions do not show any short regions of homology. The results obtained show that the inactivation of the homologous pathway resulting from the deletion of the *RAD52* gene does not affect the segmental duplication selection process. By contrast, no single gene duplications were selected in this context.

DISCUSSION

Gene duplication plays an important role in Evolution because it constitutes a source of evolutionary novelty. The aim of the present study was to identify the ways in which a given DNA sequence may be duplicated in *S.cerevisiae* haploid strains and to elucidate the mechanisms involved. Using a genetic system based on the *URA2* gene, strains carrying a duplication events were selected and analyzed. In a previous study, we established that half of the selected events resulted from a single gene duplication process (9). In the present study, we report that in the second half of the revertants, the selected events resulted from a segmental duplication process. All these data provide direct experimental evidence that a cell is able to duplicate part of its genome via at least 2 different mechanisms, resulting in either segmental duplication or single gene duplication events.

Among the four revertants analyzed here, three were found to result from a process of intrachromosomal duplication, whereas one was due to interchromosomal duplication. These segments ranging from 5 to 90 kb in size cover a region situated between ORFs YJL130c and YJL190c (Figure 3). In order to be reactivated, the duplicated ATCase has to be under the control of a functional promoter sequence. In two revertants (Rev 27 and Rev 46), this sequence can be provided by the promoter of the non-essential genes *SNA3* and *RPS22A* in to which the duplicated segment is inserted. In these strains, the ATCase sequence is fused in frame with the 5' of these disrupting genes leading to chimeric genes. These experimental data support the idea that the fusion genes result from a process of domain accretion, which make it possible for a new function to emerge (24). In the other two revertants (Rev 9 and Rev 52), the duplicated segment is fused in to an intergenic region which probably acts as a functional promoter. These two duplication events constitute an example of the changes in the regulatory sequences which may provide a key to a molecular evolution (25).

Analysis of the single gene duplication events suggests that this duplication process involved a retroposition mechanism

(9). In the case of the segmental duplications described in this paper, reverse transcription (RT) can be ruled out, since the duplicated region is much longer than the *URA2* mRNA. Several hypotheses can be put forward to account for the occurrence of tandem segmental duplication processes. First, segmental duplication is one possible outcome of UCO which results from homologous recombination between paralogous sequences. In human, Bailey *et al.* (26) established that Alu elements are present at the junctions between the original and duplicated copy sequences. In *S.cerevisiae*, homologous recombination can also occur between two repeated sequences such as two identical LTRs or Ty elements. In all the chromosomal rearrangements described thus far such as translocation, the breakpoint is surrounded by repeated sequences which are mostly Ty or δ elements (27–29). However, a mechanism via UCO does not fit the duplication events described here. The 5' and 3' insertion sites of the duplicated segment determined here show that no long repeated sequences are present in the vicinity. Actually, in the Rev 46 strain, only a short region of homology (7 bp) is present at the junction. In the other three strains (Rev 9, Rev 52 and Rev 27), the junctions do not show any short regions of homology. The possibility that mechanism via UCO might be involved is not at all consistent with the fact that segmental duplications can be selected in a *rad52* Δ context. These findings show that the segmental duplications result from a *RAD52*-independent mechanism, i.e. one in which homologous recombination plays no part.

Because the segmental duplications appear to be *RAD52*-independent, the second hypothesis which can be put forward to account for their formation is a NHEJ (non-homologous end joining) mechanism as it was suggested by Kunes *et al.* (30). After DNA replication during mitosis, the duplications could result from two double strand breaks (one located at the *URA2* locus and the other one at the 5' insertion site of the duplicated region) followed by a simple non-homologous end joining event between two DNA ends. In fact, this repair pathway requires little (2–20 bp) or no sequence homology between the ends of the DSBs. Nevertheless, this mechanism is unlikely because it requires two DSBs.

Finally, the third hypothesis which can be put forward to account for the occurrence of tandem segmental duplication is a replication error mechanism (11). A replication fork pauses and collapses, generating a chromosome breakage. The double strand can then be processed in to a new replication fork. This model is also consistent with the segmental duplication events observed in three revertants (Rev 9, Rev 27 and Rev 46). In all these revertants, the 3' insertion site of the duplicated segment is that of a wild-type sequence without any rearrangement scar. Only the 5' insertion site differs as the result of the duplication event. To investigate this hypothesis, it would be interesting to study the occurrence of duplication events in backgrounds known to affect the replication fork, such as the *rad27* or *clb5* context.

The duplication observed in strain Rev 52 was found to be a non-tandem repetition inserted in to the chromosome XI. However, the insertion occurs in the region corresponding to the ancestral block 41 defined by Wolfe and Shields (6). It therefore seems likely that the interchromosomal duplication observed may have resulted from 2 successive rearrangements. First, a segmental duplication in tandem similar to

those observed in the other three revertants may have occurred in chromosome X. Secondly, this events may have been followed by a homologous recombination between the paralogous regions *GLG1* and *TIF1*, which belong to the ancestral block 41. This finding shows that the ancestral duplicated blocks resulting from a WGD may be responsible for chromosomal rearrangements and may therefore play an evolutionary role.

An interesting point which arises about single gene duplication via retroposition is the question as to how the cDNA is inserted in to the genome. Although only six revertants were selected and analyzed in the *rad52* Δ context, it is surprising that no single gene duplications were observed. In fact, among the 8 spontaneous duplications selected, 4 events (amounting to 50%) were found to result from single gene duplication. All these data strongly suggest that homologous recombination play a role in the genesis of the single gene duplications mediated by retroposition and that the insertion of the cDNA therefore probably implicates a homologous recombination event.

The present data support the idea that the mechanisms whereby genes become duplicated are various. Starting with a haploid *ura2 15-30-72* mutant strain, two distinct type of duplication events were identified in our study: segmental duplication and gene duplication events. However, in studies on haploid cells, only a limited proportion of the rearrangements can be detected because the chromosomal rearrangements that lead to the loss of essential genes or DNA segments would make the cells inviable.

ACKNOWLEDGEMENTS

The authors wish to thank Maitreya Dunham for helpful discussions. We are very grateful to Christelle Thibault from the 'Génopole Alsace-Lorraine' for the Affymetrix array hybridizations. This work was partly supported by the CNRS (GDR 2354 'Génolevure II'). Purchase of the Affymetrix arrays was supported by the French 'Ministère de l'Enseignement Supérieur et de la Recherche'. J.S. is supported by a grant from the French 'Ministère de l'Enseignement Supérieur et de la Recherche'. The Open Access Publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

- Ohno, S. (1970) *Evolution by Gene Duplication*. Springer, Berlin.
- Zhang, J., Rosenberg, H.F. and Nei, M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl Acad. Sci. USA*, **95**, 3708–3713.
- Dujon, B. (1996) Yeast genome project: what did we learn? *Trends Genet.*, **12**, 263–270.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M. *et al.* (1996) Life with 6000 genes. *Science*, **274**, 563–567.
- Llorente, B., Durrens, P., Malpertuy, A., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 20 evolution of gene redundancy compared to *Saccharomyces cerevisiae*. *FEBS Lett.*, **487**, 122–133.
- Wolfe, K.H. and Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, **387**, 708–713.

7. Kellis,M., Birren,B.W. and Lander,E.S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, **428**, 617–614.
8. Dietrich,F.S., Voegeli,S., Brachat,S., Lerch,A., Gates,K., Steiner,S., Mohr,C., Pohlmann,R., Luedi,P., Choi,S. *et al.* (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science*, **304**, 304–307.
9. Schacherer,J., Tourrette,Y., Souciet,J.L., Potier,S. and De Montigny,J. (2004) Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. *Genome Res.*, **14**, 1291–1297.
10. Bailey,J.A., Gu,Z., Clark,R.A., Reinert,K., Samonte,R.V., Schwartz,S., Adams,M.D., Myers,E.W., Li,P.W. and Eichler,E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003–1007.
11. Koszul,R., Caburet,S., Dujon,B. and Fischer,G. (2004) Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.*, **23**, 234–243.
12. Hughes,T.R., Roberts,C.J., Dai,H., Jones,A.R., Meyer,M.R., Slade,D., Burchard,J., Dow,S., Ward,T.R., Kidd,M.J., Friend,S.H. and Marton,M.J. (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature Genet.*, **25**, 333–337.
13. Roelants,F., Potier,S., Souciet,J.L. and de Montigny,J. (1995) Reactivation of the ATCase domain of the URA2 gene complex: a positive selection method for Ty insertions and chromosomal rearrangements in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **246**, 767–773.
14. Welcker,A.J., de Montigny,J., Potier,S. and Souciet,J.L. (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.
15. Bach,M.L., Roelants,F., de Montigny,J., Huang,M., Potier,S. and Souciet,J.L. (1995) Recovery of gene function by gene duplication in *Saccharomyces cerevisiae*. *Yeast*, **11**, 169–177.
16. Wach,A., Brachat,A., Pohlmann,R. and Philippsen,P. (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast*, **10**, 1793–1808.
17. Bonneaud,N., Ozier-Kalogeropoulos,O., Li,G.Y., Labouesse,M., Minvielle-Sebastia,L. and Lacroute,F. (1991) A family of low and high copy replicative, integrative and single-stranded *S.cerevisiae/E.coli* shuttle vectors. *Yeast*, **7**, 609–615.
18. Becker,D.M. and Guarente,L. (1991) High-efficiency transformation of yeast by electroporation. *Methods Enzymol.*, **194**, 182–187.
19. Dower,W.J., Miller,J.F. and Ragsdale,C.W. (1988) High efficiency transformation of *E.coli* by high voltage electroporation. *Nucleic Acids Res.*, **16**, 6127–6145.
20. Hoffman,C.S. and Winston,F. (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene*, **57**, 267–272.
21. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
22. Carle,G.F. and Olson,M.V. (1985) An electrophoretic karyotype for yeast. *Proc. Natl Acad. Sci. USA*, **82**, 3756–3760.
23. Zhang,J. (2003) Evolution by gene duplication: an update. *Trends Ecol. Evol.*, **18**, 292–298.
24. Eichler,E.E. (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.*, **17**, 661–669.
25. Lynch,M. (2002) Gene duplication and evolution. *Science*, **297**, 945–947.
26. Bailey,J.A., Liu,G. and Eichler,E.E. (2003) An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.*, **73**, 823–834.
27. Fischer,G., James,S.A., Roberts,I.N., Oliver,S.G. and Louis,E.J. (2000) Chromosomal evolution in *Saccharomyces*. *Nature*, **405**, 451–454.
28. Dunham,M.J., Badrane,H., Ferea,T., Adams,J., Brown,P.O., Rosenzweig,F. and Botstein,D. (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **99**, 16144–16149.
29. Leh-Louis,V., Wirth,B., Potier,S., Souciet,J.L. and Despons,L. (2004) Expansion and contraction of the DUP240 multigene family in *Saccharomyces cerevisiae* populations. *Genetics*, **167**, 1611–1619.
30. Kunes,S., Botstein,D. and Fox,M.S. (1990) Synapsis-mediated fusion of free DNA ends forms inverted dimer plasmids in yeast. *Genetics*, **124**, 67–80.