Systematic screen for human disease genes in yeast

Lars M. Steinmetz[1,3*], Curt Scharfe[2,3*], Adam M. Deutschbauer[1], Dejana Mokranjac[4], Zelek S. Herman[3], Ted Jones[3], Angela M. Chu[2], Guri Giaever[3], Holger Prokisch[4], Peter J. Oefner[2,3] & Ronald W. Davis[1–3]

**These authors contributed equally to this work.*

Published online: 22 July 2002, doi:10.1038/ng929

nature genetics • volume 31 • august 2002

High similarity between yeast and human mitochondria allows functional genomic study of *Saccharomyces cerevisiae* to be used to identify human genes involved in disease[1]. So far, 102 heritable disorders have been attributed to defects in a quarter of the known nuclear-encoded mitochondrial proteins in humans[2]. Many mitochondrial diseases remain unexplained, however, in part because only 40–60% of the presumed 700–1,000 proteins involved in mitochondrial function and biogenesis have been identified[3]. Here we apply a systematic functional screen using the pre-existing whole-genome pool of yeast deletion mutants[4–6] to identify mitochondrial proteins. Three million measurements of strain fitness identified 466 genes whose deletions impaired mitochondrial respiration, of which 265 were new. Our approach gave higher selection than other systematic approaches, including fivefold greater selection than gene expression analysis. To apply these advantages to human disorders involving mitochondria, human orthologs were identified and linked to heritable diseases using genomic map positions.

La levure, une aide pour décrypter les maladies mitochondriales humaines ?

Françoise Foury

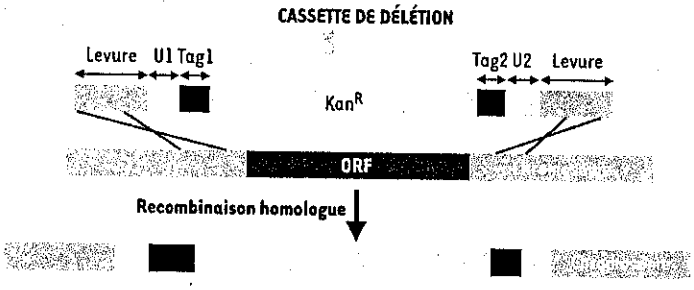


***Figure 1. Cassette utilisée pour la délétion systématique des gènes de* Saccharomyces cerevisiae.** La cassette contient le gène *Kan*^{R} sous le contrôle d'un promoteur reconnu par *S. cerevisiae*, deux séquences « code-barres » Tag1 et Tag2 spécifiques de chaque délétion, les séquences U1 et U2 communes à toutes les cassettes qui sont utilisées pour l'amplification simultanée de ces dernières, et 40 bases homologues au génome de levure, de part et d'autre du gène à déléter. Après transformation de la levure, la cassette se substitue au gène par recombinaison homologue.

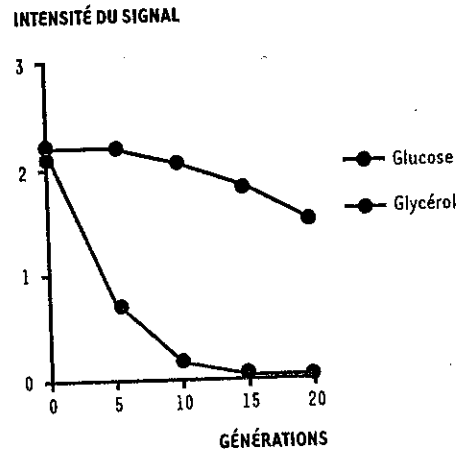


Figure 2. Profil de croissance d'une souche délétée de la classe III. L'intensité du signal détecté sur les puces à ADN diminue plus lentement en présence de glucose (conditions non respiratoires ou fermentescibles) qu'en présence de glycérol (conditions respiratoires obligatoires ou non fermentescibles), situation au cours de laquelle la croissance de la souche s'arrête très rapidement.

Figure

To return to the article, close this browser window. To toggle between the article and the figure, click on the browser window containing the article.

Systematic screen for human disease genes in yeast

Lars M. Steinmetz *et al.*

Nature Genetics 31, 400 – 404 (2002)

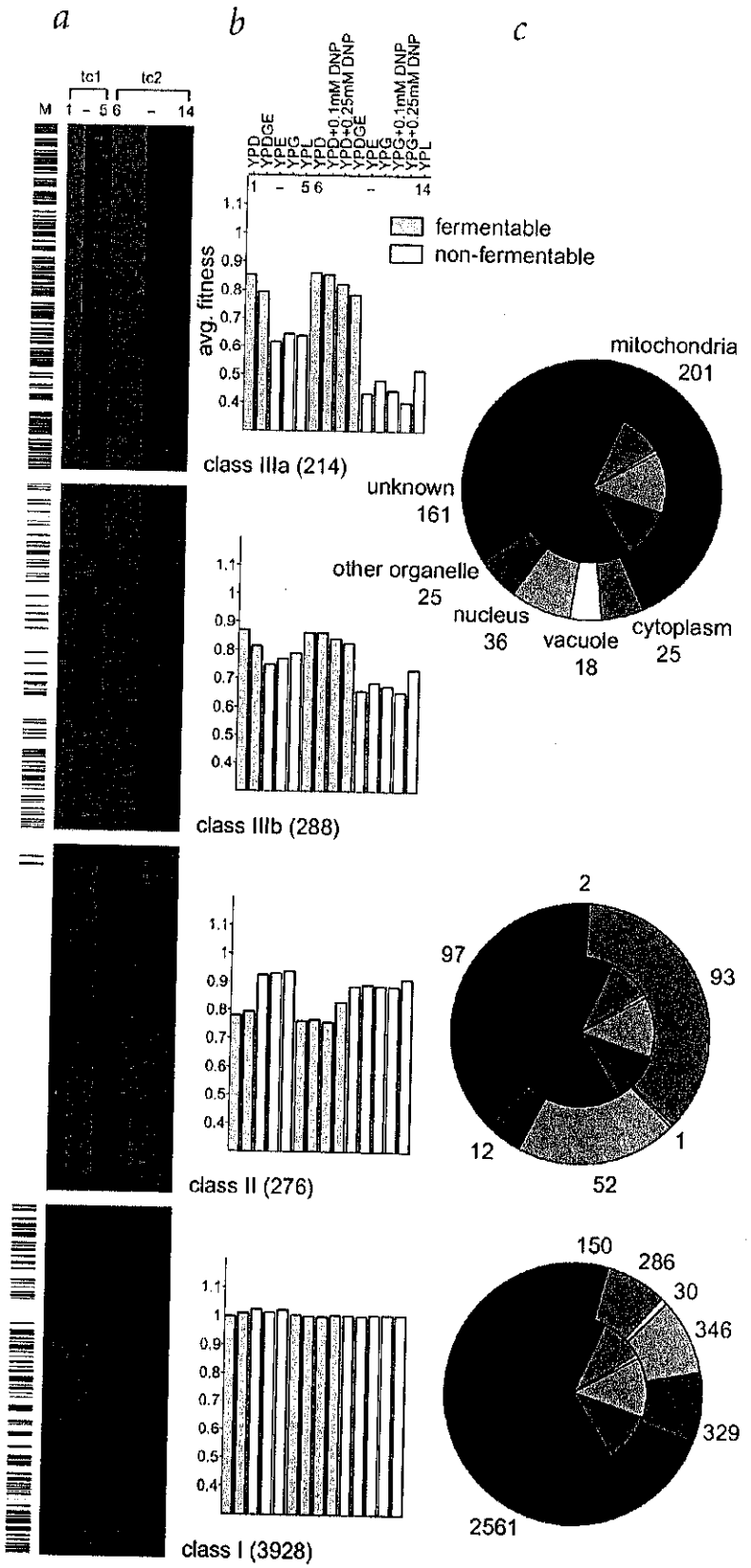


Figure 1: Categorization of the whole genome according to phenotypes associated with gene deletions.

a, Clustergram showing four fitness patterns found among the 4,706 homozygous diploid deletion strains in the pool. Each row represents a strain with a deletion of a different gene and each column a different medium condition. The number of strains in each cluster is indicated in parentheses next to the cluster name. For each strain, fitness values are indicated with a color scale ranging from blue to red, with blue representing levels below, and red levels above, the strain's median. Conditions 1–5 represent measurements from the first experimental time course (tc1) and columns 6–14 those from a repeat experiment (tc2). Marked in black at left are 353 strains with deletions of genes previously known to localize to or function in mitochondria (M). ***b***, The bar graphs show the average fitness profiles for each cluster. For each condition, the height of the bar represents the growth rate of strains in the cluster relative to the average growth rate of the pool. Values of 1.0 indicate no difference, those less than 1.0 strains that grow more slowly than, and those greater than 1.0 strains that grow more quickly than, the pool average. The medium conditions, indicated above the graphs, are in the same order as the columns in the clustergram. ***c***, The outer pie chart shows the composition of genes represented in each cluster according to MIPS localization categories[22], after the removal of all spurious ORFs. The inner pie charts represent the distribution over the genome. Because of the similarity in pattern, the class III clusters were combined.

To return to the article, close this browser window. To toggle between the article and the figure, click on the browser window containing the article.

Figure

To return to the article, close this browser window. To toggle between the article and the figure, click on the browser window containing the article.

Systematic screen for human disease genes in yeast

Lars M. Steinmetz *et al.*

Nature Genetics 31, 400 – 404 (2002)

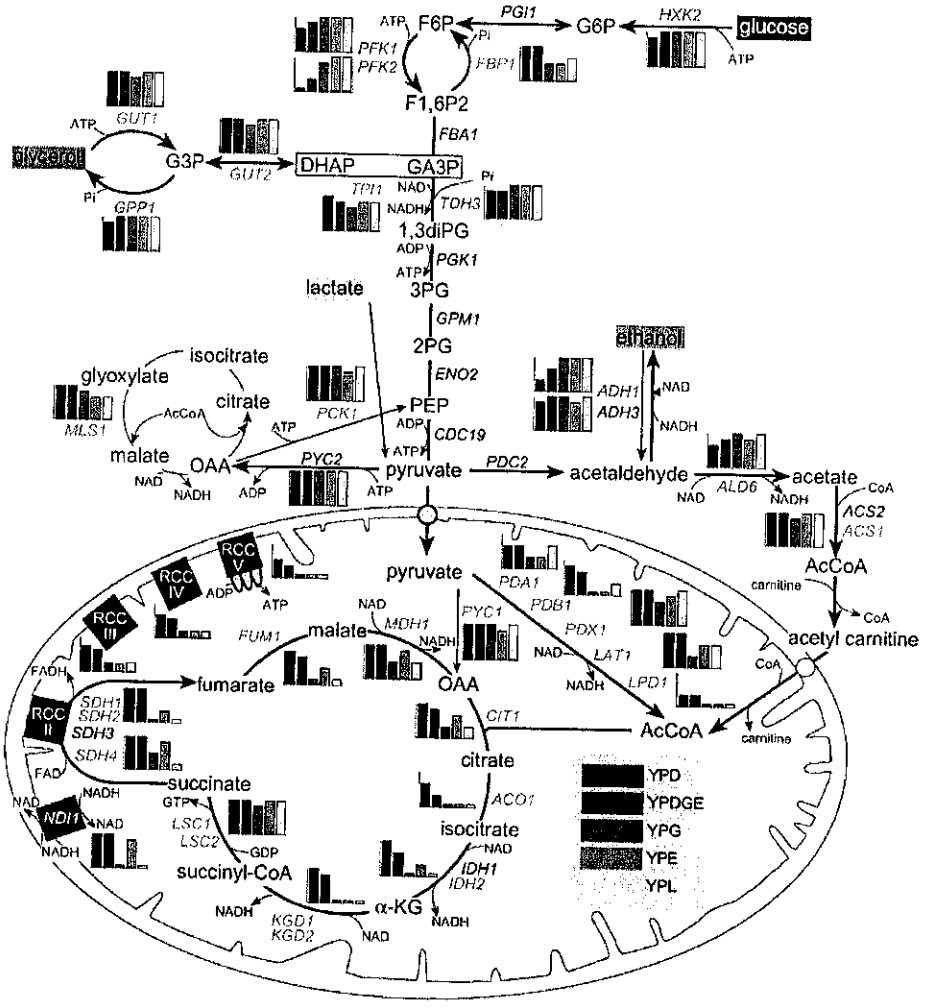


Figure 3: Distinction between mitochondrial and cytoplasmic pathway branches: glycolysis above and TCA cycle and mitochondrial respiratory chain below.

The bar graphs indicate the relative fitness of a homozygous deletion mutant of a gene under different medium conditions, color-coded in the legend. Genes without bar graphs were not detected. For the respiratory-chain complexes (RCC) III, IV and V, an average profile is shown. Deletions of genes shown in red lettering result in deficiency in growth on fermentable substrates, and deletion of those in green in a deficiency in growth on non-fermentable substrates. Ac, acetyl; CoA, coenzyme A; DHAP, dihydroxyacetone phosphate; 1,3diPG, 1,3-bisphosphoglycerate; F6P, fructose-6-phosphate; F1,6P2, fructose-1,6-bisphosphate; GA3P, glyceraldehyde-3-phosphate; α-KG, α-ketoglutarate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; Pi, phosphate.

To return to the article, close this browser window. To toggle between the article and the figure, click on the browser window containing the article.

Figure

To return to the article, close this browser window. To toggle between the article and the figure, click on the browser window containing the article.

Systematic screen for human disease genes in yeast

Lars M. Steinmetz *et al.*

Nature Genetics **31**, 400 – 404 (2002)

QDP	Yeast gene	Human gene	OMIM	Pathophysiological defect
	*SDH1*	*SDHA*	600857	RCC II subunit
	*SDH2*	*SDHB*	185470	RCC II subunit
	*BCS1*	*BCS1L*	603647	RCC III assembly
	*SHY1*	*SURF1*	185620	RCC IV assembly
	*SCO1*	*SCO1*	603644	RCC IV assembly
	*SCO1*	*SCO2*	604272	RCC IV assembly
	*COX10*	*COX10*	602125	RCC IV assembly
	*LAT1*	*PDX1*	245349	pyruvate DH
	*PDA1*	*PDHA1*	312170	pyruvate DH
	*PDA1*	*BCKDHA*	248600	AA catabolism (DH)
	*PDB1*	*BCKDHB*	248611	AA catabolism (DH)
	*KGD2*	*DBT*	248610	AA catabolism (DH)
	*LPD1*	*DLD*	246900	pyruvate/AA/α-KGDH
	*MIS1*	*MTHFD1*	172460	AA metabolism
	*GCV3*	*GCSH*	238330	AA metabolism
	*YHM1*	*SLC25A15*	603861	small-molecule transport
	*PET8*	*SLC25A15*	603861	small-molecule transport
	*FUM1*	*FH*	136850	TCA-cycle enzyme
	*MIP1*	*POLG*	174763	maintenance of mtDNA
	*HEM14*	*PPOX*	600923	heme biosynthesis
	*YTA12*	*SPG7*	602783	ATP-dependent protease
	*CCC2*	*ATP7B*	277900	copper-transport ATPase

Figure 4: Human mitochondrial-related genes that give rise to disease and for which there is an associated quantitative deletion phenotype in yeast.

Bar graphs represent the relative fitness of homozygous yeast deletion mutants grown in different carbon sources, as defined in the legend of Fig. 3. AA, amino acid; DH, dehydrogenase; KG, ketoglutarate; QDP, quantitative deletion phenotype; RCC, respiratory-chain complex.

To return to the article, close this browser window. To toggle between the article and the figure, click on the browser window containing the article.

Genome evolution in yeasts

Bernard Dujon[1], David Sherman[5,6], Gilles Fischer[1], Pascal Durrens[6,7], Serge Casaregola[8], Ingrid Lafontaine[1], Jacky de Montigny[9], Christian Marck[10], Cécile Neuvéglise[8], Emmanuel Talla[1], Nicolas Goffard[6], Lionel Frangeul[2], Michel Aigle[7], Véronique Anthouard[11], Anna Babour[8], Valérie Barbe[11], Stéphanie Barnay[8], Sylvie Blanchin[8], Jean-Marie Beckerich[8], Emmanuelle Beyne[5,6], Claudine Bleykasten[9], Anita Boisramé[8], Jeanne Boyer[1], Laurence Cattolico[11], Fabrice Confanioleri[12], Antoine de Daruvar[6], Laurence Despons[9], Emmanuelle Fabre[1], Cécile Fairhead[1], Hélène Ferry-Dumazet[6], Alexis Groppi[6], Florence Hantraye[3], Christophe Hennequin[1], Nicolas Jauniaux[9], Philippe Joyet[8], Rym Kachouri[13], Alix Kerrest[1], Romain Koszul[1], Marc Lemaire[14], Isabelle Lesur[5], Laurence Ma[2], Hélo'ise Muller[1], Jean-Marc Nicaud[8], Macha Nikolski[5], Sophie Oztas[11], Odile Ozier-Kalogeropoulos[1], Stefan Pellenz[1], Serge Potier[9], Guy-Franck Richard[1], Marie-Laure Straub[9], Audrey Suleau[8], Dominique Swennen[8], Fredj Tekaia[1], Micheline Wésolowski-Louvel[14], Eric Westhof[13], Bénédicte Wirth[9], Maria Zeniou-Meyer[9], Ivan Zivanovic[12], Monique Bolotin-Fukuhara[12], Agnès Thierry[1], Christiane Bouchier[2], Bernard Caudron[4], Claude Scarpelli[11], Claude Gaillardin[8], Jean Weissenbach[11], Patrick Wincker[11] & Jean-Luc Souciet[9]

Identifying the mechanisms of eukaryotic genome evolution by comparative genomics is often complicated by the multiplicity of events that have taken place throughout the history of individual lineages, leaving only distorted and superimposed traces in the genome of each living organism. The hemiascomycete yeasts, with their compact genomes, similar lifestyle and distinct sexual and physiological properties, provide a unique opportunity to explore such mechanisms. We present here the complete, assembled genome sequences of four yeast species, selected to represent a broad evolutionary range within a single eukaryotic phylum, that after analysis proved to be molecularly as diverse as the entire phylum of chordates. A total of approximately 24,200 novel genes were identified, the translation products of which were classified together with *Saccharomyces cerevisiae* proteins into about 4,700 families, forming the basis for interspecific comparisons. Analysis of chromosome maps and genome redundancies reveal that the different yeast lineages have evolved through a marked interplay between several distinct molecular mechanisms, including tandem gene repeat formation, segmental duplication, a massive genome duplication and extensive gene loss.

Table 1 **Genome assemblies of the four yeast species**

Species	Strain	Number of chromosomes	Total reads	Coverage (sequence)	Coverage (clones)	N50 contigs (kb)	N50 scaffolds (kb)	Total gaps	Assembly size (without rDNA) (kb)
C. glabrata	CBS138	13	188,853	×8	×30	1,000	1,025	6	12,280
K. lactis	CLIB210	6	152,071	×11.4	×56	1,670	1,670	0	10,631
D. hansenii	CBS767	7	150,570	×9.7	×36	102	2,038	207	12,221
Y. lipolytica	CLIB99	6	247,279	×10	×59	704	3,453	11	20,503

Sequencing and assembly were performed as described in Methods. Sequences of *C. glabrata*, *K. lactis* and *Y. lipolytica* are finished (no gap) or contain very few gaps. The sequence of *D. hansenii* is in the form of a high-quality draft. In all cases, each chromosome of each yeast is either complete (single contig) or represented by a single super-contig (scaffold). Most remaining gaps are small or contain repeated sequences. Some subtelomeric regions are missing from the assembly because they are too similar to one another to be assigned to a specific chromosome. rDNA repeats are assembled separately. N50, median values.

Table 2 **General characteristics of the yeast genomes and predicted proteomes**

Species	Genome size (Mb)	Average G+C content (%)	Total CDS	Total tRNA genes	Average gene density (%)	Average G+C in CDS (%)	Average CDS size (codons)	Median CDS size (codons)	Maximum CDS size (codons)
S. cerevisiae	12.1	38.3	5,807	274	70.3	39.6	485	398	4,911
C. glabrata	12.3	38.8	5,283	207	65.0	41.0	493	409	4,881
K. lactis	10.6	38.7	5,329	162	71.8	40.1	461	381	4,916
D. hansenii	12.2	36.3	6,906	205	79.2	37.5	389	307	4,190
Y. lipolytica	20.5	49.0	6,703	510	46.3	52.9	476	399	6,539

Figures are calculated from final chromosome sequences or scaffolds, after annotation. Genome sizes do not include rDNA. Average gene density represents the fraction of each genome occupied by the protein-coding genes (other genetic elements are not considered). Figures for *D. hansenii* are only tentative; figures for *S. cerevisiae* were recently recomputed from http://mips.gsf.de/genre/proj/yeast.

Table 3 **Classification of yeast proteins in families**

Family class	Number of families	Number of CDS					
		SACE	CAGL	KLLA	DEHA	YALI	Total
Robust (identical + reconciled)	3,410	4,094	3,651	3,504	3,832	3,296	18,377
Consensus	1,311	1,287	1,201	1,176	1,831	1,894	7,389
Non-assigned	–	426	431	649	1,243	1,513	4,262
Total	4,721	5,807	5,283	5,329	6,906	6,703	30,028

The table shows the total number of protein families in each class and the corresponding numbers of CDS in each yeast species. Families were classified as explained in Methods. See Fig. 1 for species abbreviations.

Open Access

Method

Large-scale exploration of growth inhibition caused by overexpression of genomic fragments in *Saccharomyces cerevisiae*

Jeanne Boyer*, Gwenaël Badis*†, Cécile Fairhead*, Emmanuel Talla*‡, Florence Hantraye†, Emmanuelle Fabre*, Gilles Fischer*, Christophe Hennequin*§, Romain Koszul*, Ingrid Lafontaine*, Odile Ozier-Kalogeropoulos*, Miria Ricchetti*¶, Guy-Franck Richard*, Agnès Thierry* and Bernard Dujon*

Abstract

We have screened the genome of *Saccharomyces cerevisiae* for fragments that confer a growth-retardation phenotype when overexpressed in a multicopy plasmid with a tetracycline-regulatable (Tet-off) promoter. We selected 714 such fragments with a mean size of 700 base-pairs out of around 84,000 clones tested. These include 493 in-frame open reading frame fragments corresponding to 454 distinct genes (of which 91 are of unknown function), and 162 out-of-frame, antisense and intergenic genomic fragments, representing the largest collection of toxic inserts published so far in yeast.

refereed research

Table 1

Distribution of the toxic inserts between the different genetic objects

Genetic objects represented	Number of toxic inserts	Percentage of total	Mean size ± SD (nucleotides) (minimum-maximum)	Phenotypes				Inserts encoding artificial peptides
				3/0, 3/1	3/2	2/0, 2/1	1/0	
In-frame ORF fragments	493	68.7	743 ± 311 (220-2,120)	375	87	23	8	–
Antiparallel ORF fragments	68	9.6	532 ± 247 (140-1,220)	37	11	12	8	53
Out-of-frame ORF fragments	53	7.5	733 ± 306 (170-1,620)	12	11	22	8	12
Intergenic regions	41	6.0	625 ± 358 (170-1,820)	13	4	16	8	27
LTRs	2	0.3	595 (320-1,120)	1	0	0	1	1
Ty elements	15 (10)	2.1	633 ± 265 (320-870)	7	4	2	2	–
Y' elements	9 (3)	1.2	678 ± 370 (320-1,320)	9	0	0	0	6
RNA genes	4	0.5	662 ± 246 (470-1,020)	3	0	1	0	3
2 μm plasmid	17 (10)	2.4	564 ± 288 (170-1,220)	13	3	1	0	5
Mitochondrial DNA	12	1.7	483 ± 201 (200-920)	9	3	0	0	10
Total	714	100	703 ± 313 (140-2,120)	479	123	77	35	117

The first column indicates nature of sequence in toxic inserts. Second and third columns contain, respectively, actual number of inserts of each type and corresponding percentages. For Tys, Y' and 2 μm plasmid, numbers in brackets represent numbers of in-frame fragments of natural ORFs. The fourth column shows the mean size of insert in nucleotides ± standard deviation (SD) with minimum and maximum sizes in brackets. Scoring of each type of phenotype is shown in the next four columns. The last column shows the number of inserts in which artificial ORFs of more than 24 codons were detected.

Figure 3

Construction d'une banque génomique de S. *cerevisiae*.

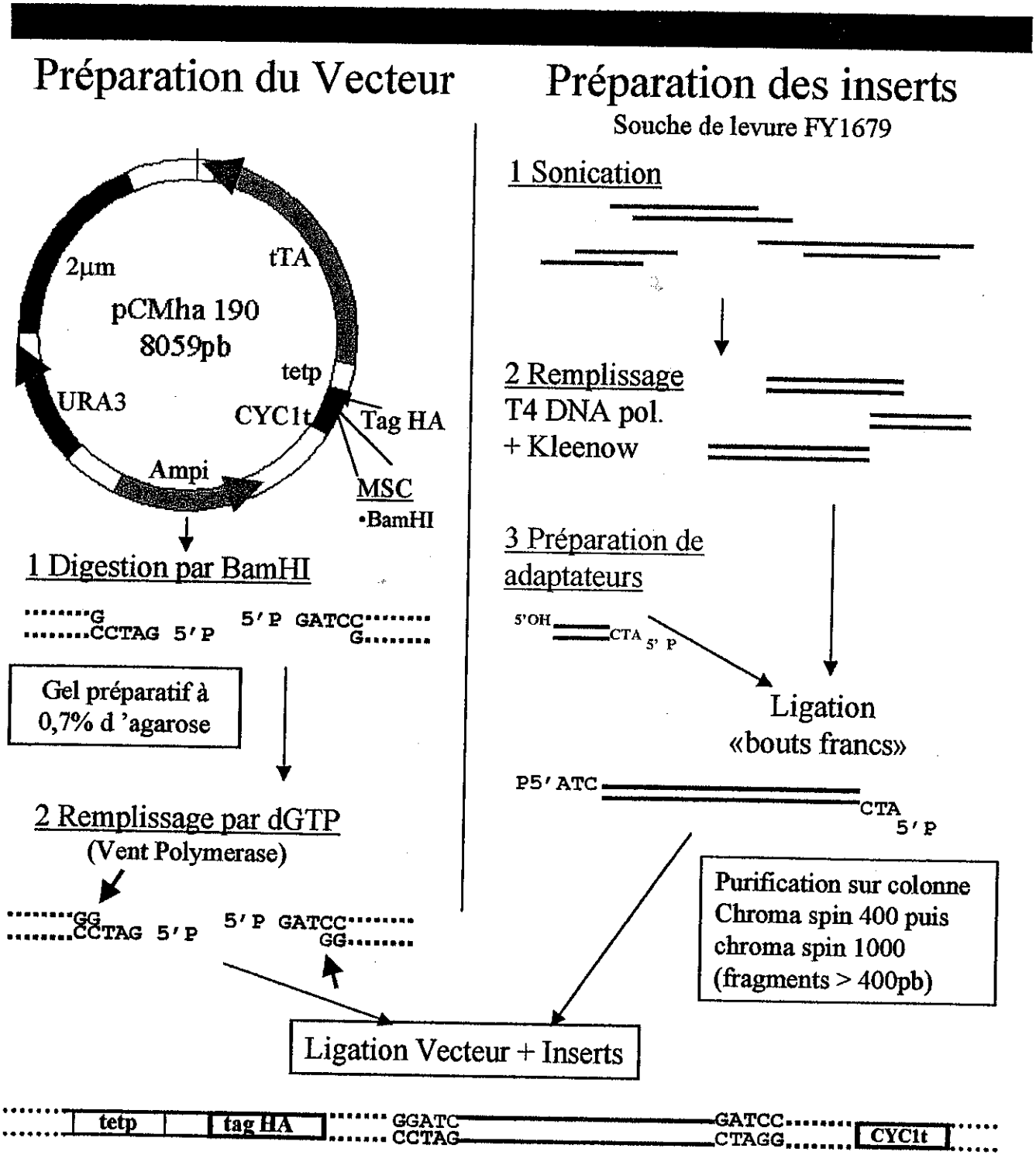
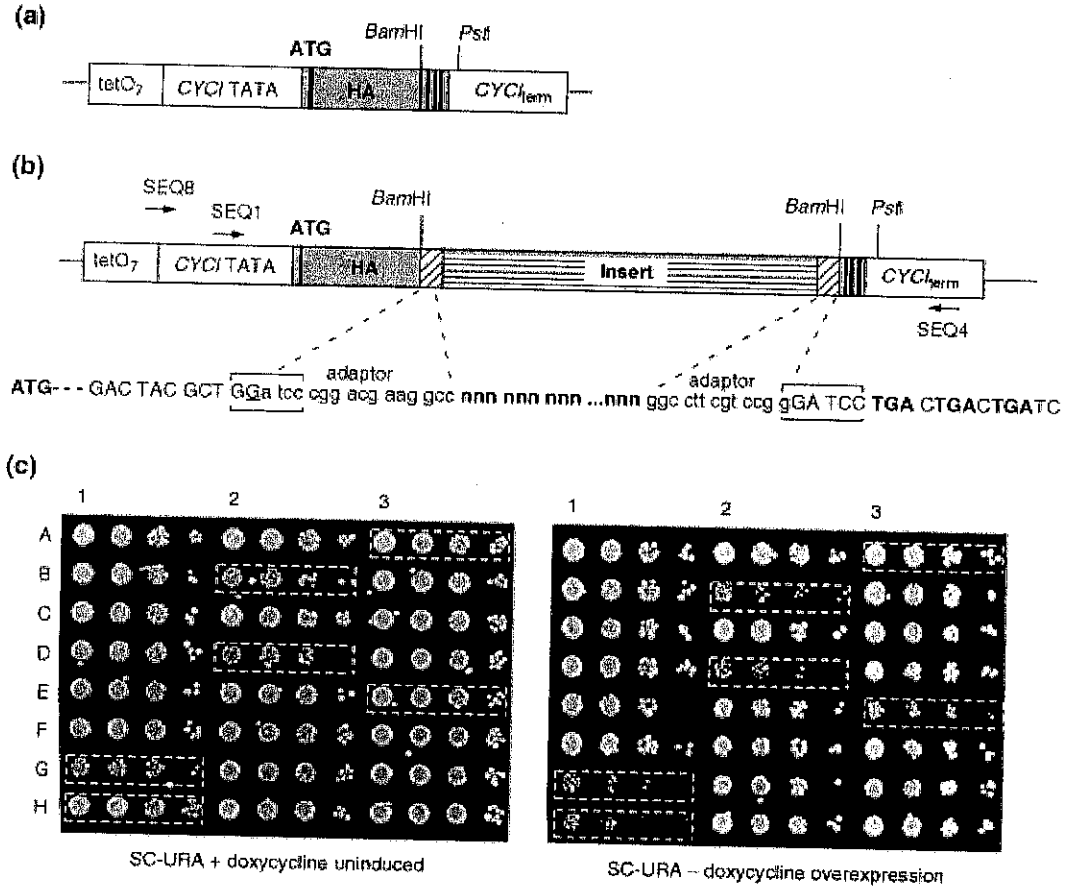


Figure 1

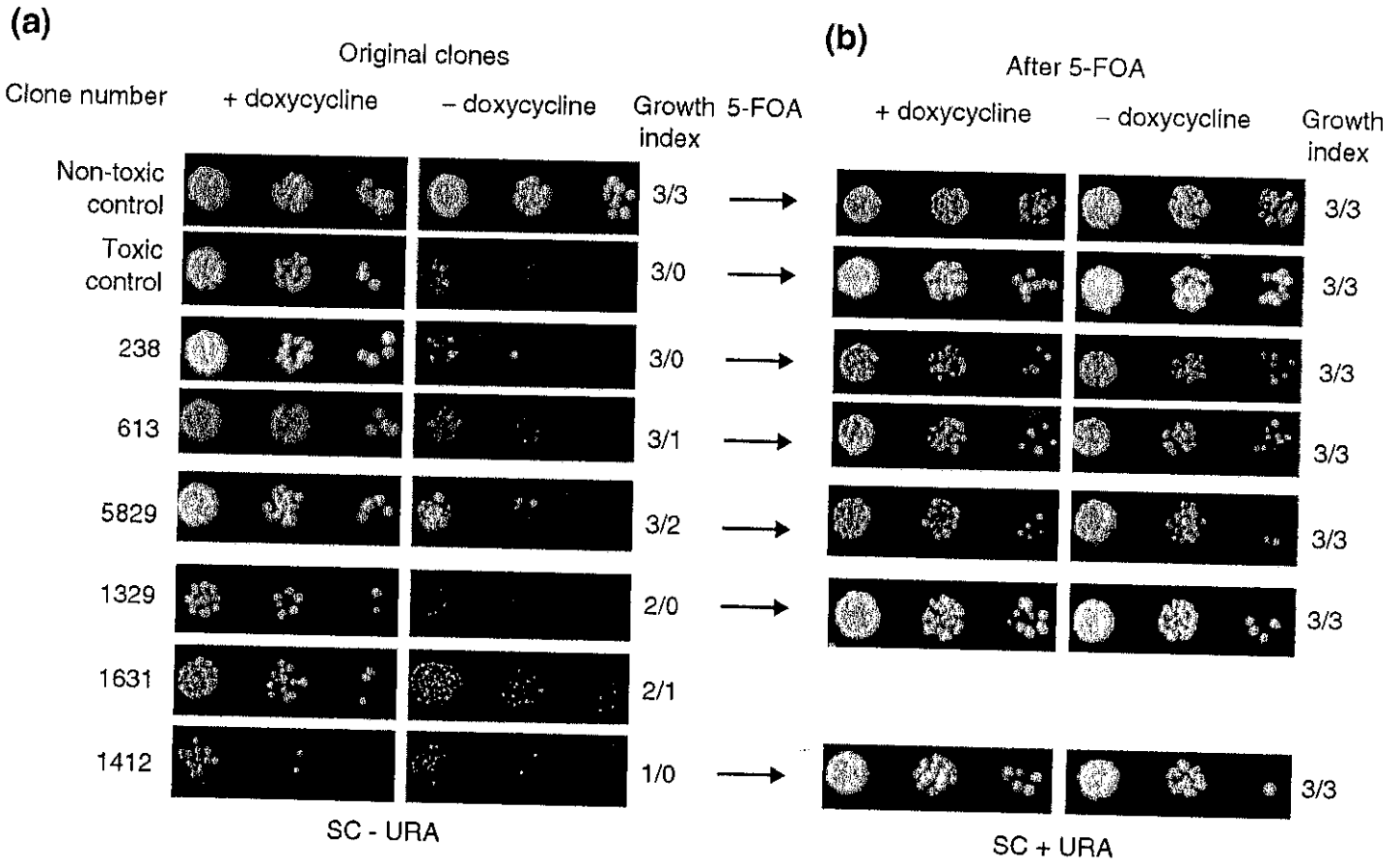
Resolution: **standard** / high



Overexpression library construction and screening. **(a)** Construction of an HA-tagged vector. The pCMha190 vector used here was constructed by insertion of a linker (gray box) in place of the multiple cloning site in vector pCM190 [31]. Features shown include the promoter and TATA box as well as the terminator from the original plasmid (open boxes), and the start codon, HA-tag, *Bam*HI site and stop codons (thick vertical bars) from the introduced linker sequence. The linker was composed from the following annealed oligonucleotides: EXP3: 5'-GATCGTTTAAACCATATGTACCCATACGACGTCCCAGACTACGCTGG ATCCTGACTGACTGATC-3', EXP4: 5'-GGCCGATCAGTCAGTCAGGATCCAGCGT AGTCTGGGACGTCGTATGGGTACATATGGTTTAAAC-3'. **(b)** Library construction in pCMha190 (see Materials and methods for experimental details). The resulting ligation product is schematized, with the insert as a striped box and adaptors as hatched boxes. Sequences shown below are from junctions, with uppercase letters corresponding to vector (the extra nucleotide from filling-in is underlined), lowercase letters to adaptors and bold nnn's to insert. Arrows indicate the different primers used: SEQ8 and SEQ4 are used for PCR amplification of the insert, and SEQ1 for sequencing (see sequences in Additional data file 8). **(c)** First-round screening of toxic phenotypes. The growth of random and control clones on selective medium in uninduced and overexpression conditions is shown. Drops of serial dilutions (1/100 to 1/100,000) of cultures were grown for 45 h at 30°C. A3, non-toxic control clone transformed by pCMha190; H1, toxic control clone transformed by *MCM1* gene cloned in pCMha190; G1, B2, D2, E3, library transformed clones, exhibiting different levels of toxicity in overexpression conditions (see Figure 2).

Figure 2

Resolution: standard / **high**



Second-round scoring of toxic phenotypes and control. **(a)** Selected clones from the first round were diluted and three drops (1/100, 1/1,000 and 1/10,000) were spotted and grown for 42 h at 30°C, with controls on same plates, for confirmation of toxicity. Growth levels in the presence and absence of doxycycline were scored as described in the text. Each clone was assigned a growth index where the first number represents the growth in uninduced conditions and second number the growth in induced conditions; for example, 3/3 indicates a non-toxic insert; 3/0 indicates a highly toxic insert. Clone numbers are the same as in the tables describing the toxic inserts (see Additional file 1,2,3,4). **(b)** After 5-FOA-induced plasmid loss, growth of surviving clones is scored in the same way as in (a). Wild-type phenotypes in overexpression conditions are indicative of plasmid-borne toxicity.

THE EMBO JOURNAL

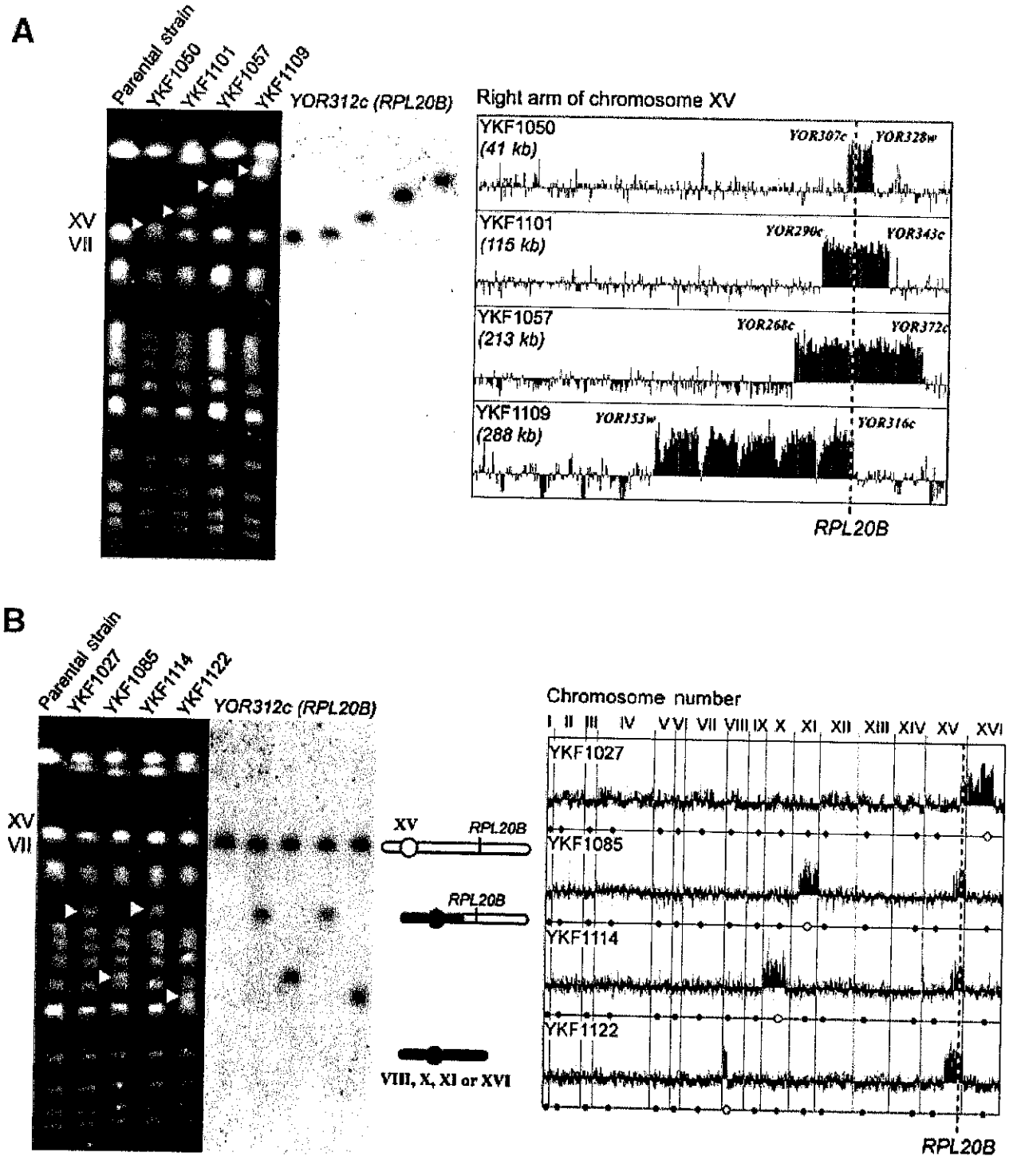
The EMBO Journal (2004) **23,** 234–243, doi: 10.1038/sj.emboj.7600024

Published online 18 December 2003

Figure 1

Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments

Romain Koszul, Sandrine Caburet, Bernard Dujon and Gilles Fischer



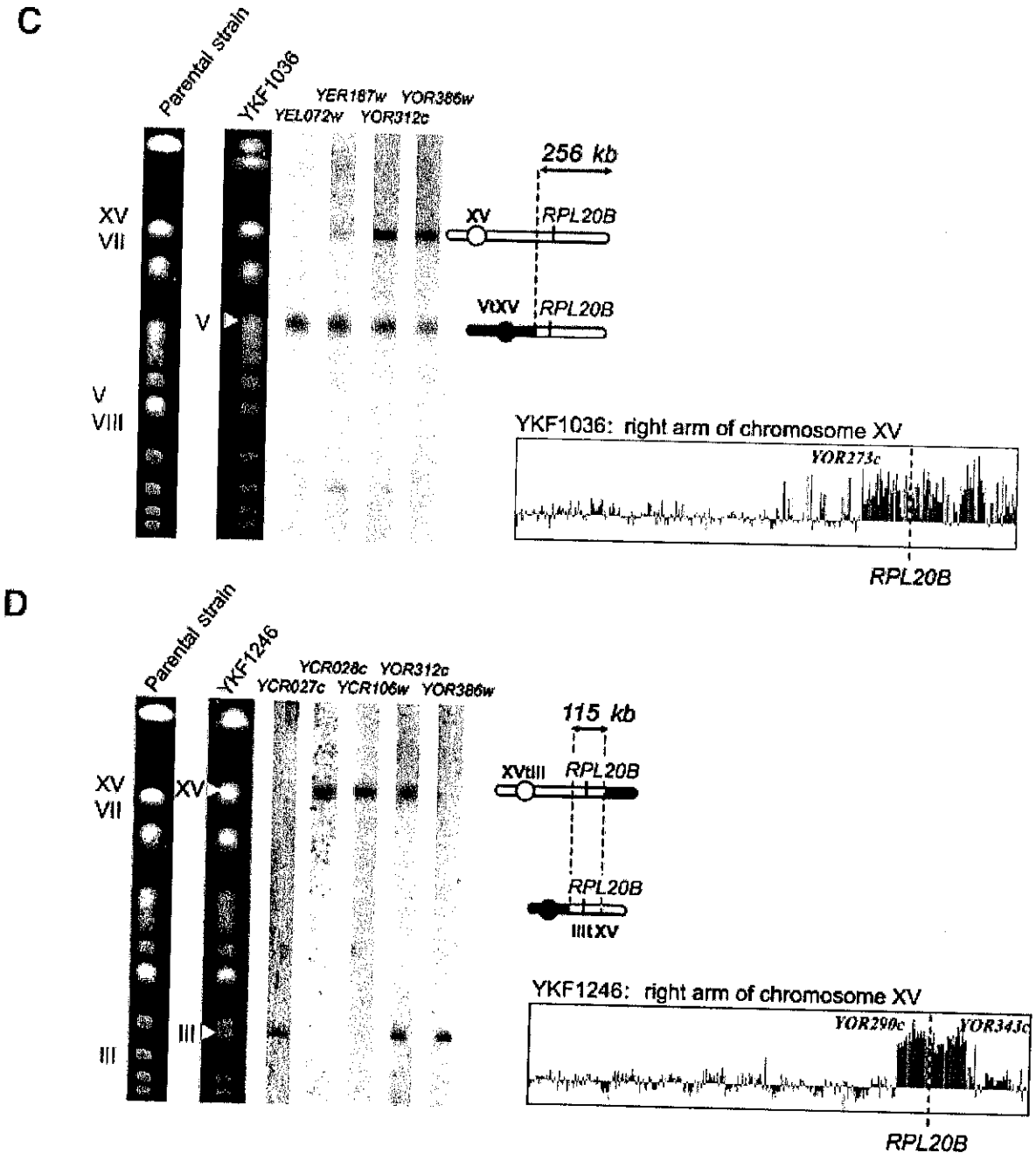


Figure 1
Intra- and interchromosomal duplications of large DNA segments from the right arm of chromosome XV. Left panels show PFGE and hybridizations of the corresponding Southern blots with various ORFs whose names are indicated above each lane. Chromosomes XV and VII comigrate in the parental strain, resulting in a band with a double intensity on the gel. Relevant chromosome numbers are indicated, and the white arrowheads show the position of the modified or additional chromosomes in the karyotypes. Right panels represent CGH-array profiles with the *X*-axes consisting either in the ORFs from the right arm of chromosome XV ordered from *YOR001w* to *YOR394w* (A, C and D) or in all yeast ORFs ordered from the left telomere of chromosome I to the right telomere of chromosome XVI (B). The *Y*-axes correspond to the genomic ratios calculated between the revertant and the parental strains. Scale ranges from -0.5 to +1.5 (A, C and D) and from -1 to +2 (B). The position of *RPL20B* (*YOR312c*) is indicated by the dotted line. The last duplicated ORFs on each side of the blocks are mentioned (A, C and D). In (C) and (D), a schematic representation of the chromosomes involved in the interchromosomal duplication–translocation events is shown. (**A**) Representative revertant strains from class I. The sizes of the duplicated segments are indicated in parentheses and range from 41 to 288 kb. (**B**) Revertant strains from class II. Black and open circles on the *X*-axes of the array profiles symbolize single and duplicated centromeres, respectively. (**C**) Class III strain YKF1036. VtXV means a translocation of the right arm of chromosome XV onto chromosome V. (**D**) Class III strain YKF1246. IIItXV and XVtIII mean a translocation of the right arm of chromosomes XV and III onto chromosomes III and XV, respectively. The ORFs YCR027c and YCR028c flank the translocation breakpoint onto chromosome IIItXV.

◂ previous | next ▸