Genomic Exploration of the Hemiascomycetous Yeasts: 5. *Saccharomyces bayanus* var. *uvarum*

Elisabeth Bon[a,b], Cécile Neuvéglise[b], Serge Casaregola[b], François Artiguenave[c], Patrick Wincker[c], Michel Aigle[a], Pascal Durrens[a,*]

[a] *Laboratoire de Biologie Cellulaire de la Levure, UMR 5095, 1 rue Camille Saint-Saëns, F-33077 Bordeaux Cedex, France*
[b] *Collection des Levures d'Intérêt Biotechnologique, Laboratoire de Génétique Moléculaire et Cellulaire, INRA UMR 216, CNRS URA 1925, INA-PG, P.O. Box 01, F-78850 Thiverval-Grignon, France*
[c] *Génoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, P.O. Box 191, F-91057 Evry Cedex, France*

Received 3 November 2000; accepted 9 November 2000

First published online 27 November 2000

Edited by Horst Feldmann

Abstract ***Saccharomyces bayanus* var. *uvarum* investigated here is the species closest to *Saccharomyces cerevisiae*. Random sequence tags (RSTs) allowed us to identify homologues to 2789 open reading frames (ORFs) in *S. cerevisiae*, ORFs duplicated in *S. uvarum* but not in *S. cerevisiae*, centromeres, tRNAs, homologues of Ty1/2 and Ty4 retrotransposons, and a complete rDNA repeat. Only 13 RSTs seem to be homologous to sequences in other organisms but not in *S. cerevisiae*. As the synteny between the two species is very high, cases in which synteny is lost suggest special mechanisms of genome evolution. The corresponding RSTs revealed that *S. uvarum* can exist without any *S. cerevisiae* DNA introgression. Accession numbers are from AL397139 to AL402278 in the EMBL databank.**

Key words: Centromere; Ty; Translocation; *Saccharomyces bayanus*

1. Introduction

Among the 13 species studied in this project, *Saccharomyces bayanus* var. *uvarum* was found to be the species closest to *Saccharomyces cerevisiae* in terms of phylogeny. Both of them belong to the *Saccharomyces* sensu stricto group, and a recent classification does not consider *S. uvarum* and *S. bayanus* as different species [1]. However, as measured by DNA–DNA reassociation [2], they have to be considered the most distant species within this group. Like *S. cerevisiae*, the *S. bayanus*/*uvarum* yeast is found in wine fermentation although the process with *S. bayanus*/*uvarum* generally takes place at a lower temperature [3]. As several strains of *S. bayanus* show a variable degree of introgression from the genome of *S. cerevisiae* to that of *S. bayanus*, great care should be taken considering the origin of sequences labelled as *S. bayanus* or *S. uvarum*. Two sub-groups within the *bayanus* species can be distinguished [4–6], though there is genome homogeneity inside the *uvarum* sub-group. The *S. bayanus*/*uvarum* strain CBS 7001, used in this study, is the most distant strain from *S. cerevisiae*, based on the percentage of non-viable spores from *S. cerevisiae*–*S. bayanus*/*uvarum* hybrids [7], and genuine *S. cerevisiae* subtelomeric sequences, found in *S. paradoxus*, are absent from the genome of strain CBS 7001 [8]. The heterogeneity among *S. bayanus* strains is reflected by their variable number of chromosomes, ranging from 14 to 17 [9], whereas the number of chromosomes remains stable at 16 in the *uvarum* sub-group. The chromosome structure of CBS 7001 has been investigated and compared to *S. cerevisiae* chromosomes: several translocations were described [10] between chromosomes II and IV, VIII and XV. We consider here the *uvarum* strain as a variety of the *S. bayanus* species. For the sake of simplicity, we will designate our strain *S. uvarum*. Our DNA sequences from *S. uvarum* CBS 7001 strain have confirmed that this strain is the closest to *S. cerevisiae*, in terms of sequence identity and synteny among the 13 species studied.

2. Materials and methods

2.1. Strain

The strain (kindly given by E. Louis) is derived from MCYC 623 (NRRL Y-11845, CBS 7001). It is homozygous as regards the HO homothallism allele. The monosporic culture 623-6C, issued from the MCYC 623, yields fertile clones in *S. bayanus* crosses [7]. The 623-6C culture was sporulated, the spores were submitted to a low dose of UV radiation, and plated on complete medium. Diploid colonies growing after mating type switch were replica-plated on 5-FOA plates. The resulting diploid strain used in this study, called 623-6C *ura3-1*, has normal karyotype and meiosis, is homozygote (E. Louis, personal communication) and is equivalent to a haploid as regards genome sequencing.

2.2. DNA library construction

Cells were spheroplasted with Glucanex (Novo Nordisk A/S) (25 mg/ml in 0.9 M Sorbitol, 0.1 M Na_2EDTA, pH 7.5) and lysed in 50 mM Tris–HCl buffer pH 7.4, 20 mM EDTA, 1% (w/v) sodium dodecyl sulfate for 30 min at 65°C. After RNase treatment and isopropanol precipitation, DNA fibers were picked with a glass rod and dissolved in 2 ml of TE buffer. DNA was partially digested with endonuclease *Cvi*JI (Chimerix, Madison, WI, USA) to obtain DNA fragments in a size range of 2–5 kb. *Sma*I linearized and dephosphorylated pBAM3 vector [11] (8 ng) was ligated to 20 ng of yeast DNA. Randomly, 3250 *Escherichia coli* DH10B clones were picked and stored in triplicates at −80°C. The average size of the inserts was estimated at 3.5 kb (±1.3 kb).

2.3. Sequences

Sequence processing was performed as described [11,12] except (i) a minmatch = 30 bp parameter in PHRAP; (ii) the BLOSUM62 substitution matrix in comparison with *S. cerevisiae* proteome; (iii) only alignments spanning a minimum of 30 residues were considered.

*Corresponding author.
E-mail: pascal.durrens@ibgc.u-bordeaux2.fr

0014-5793/00/$20.00
PII: S0014-5793(00)02276-6

Table 1
ORFs present in multiple copies in *S. uvarum*

A. Non-ambiguous matches		B. Ambiguous matches	
ORF	Number	ORF	Number
SuYBR112c	2	*SuYAL068c*	5
SuYBR148w	2	*SuYBR301w*	5
SuYBR164c	2	*SuYEL049w*	5
SuYCL029c	2	*SuYGL261c*	5
SuYCL031c	2	*SuYGR294w*	5
SuYDL042c	2	*SuYHL046c*	5
SuYDR028c	2	*SuYIL176c*	5
SuYDR030c	2	*SuYJL223c*	5
SuYDR421w	3	*SuYLL064c*	5
SuYDR503c	2	*SuYLR461w*	5
SuYDR505c	2	*SuYNR076w*	5
SuYDR533c	2	*SuYOL161c*	5
SuYER054c	2	*SuYOR394w*	5
SuYFL036w	2	*SuYPL282c*	5
SuYGL092w	2		
SuYGL250w	2	*SuYDR342c*	2
SuYGR065c	2	*SuYDR343c*	2
SuYHR029c	2		
SuYIR038c	2	*SuYFL062w*	3
SuYIR039c	2	*SuYGR295c*	3
SuYJL217w	2	*SuYNL336w*	3
SuYKL164c	2		
SuYKR097w	2	*SuYHR210c*	2
SuYLL057c	2	*SuYNR071c*	2
SuYLR081w	2		
SuYLR098c	2	*SuYOR389w*	2
SuYMR240c	2	*SuYPL277c*	2
SuYNL001w	2		
SuYNL130c	2		
SuYOL130w	2		
SuYOL155c	2		
SuYOR051c	2		
SuYOR100c	2		
SuYOR187w	2		
SuYOR244w	2		
SuYOR304w	2		
SuYOR323c	2		
SuYOR330c	2		
SuYPL273w	3		
SuYPL274w	2		
SuYPR018w	2		

ORF names are indicated with the minimal number of copies necessary to account for the alignments. A: Non-ambiguous matches corresponding to a unique match in *S. cerevisiae*, suggesting duplication in *S. uvarum* compared to *S. cerevisiae*. B: Ambiguous matches corresponding to multiple paralogues in *S. uvarum* and *S. cerevisiae*.

3. Results and discussion

3.1. Random sequence tags (RSTs), inserts and contigs

Among the 5140 RSTs, a contig of 201 RSTs corresponded to the rDNA and nine contigs, containing 74 RSTs altogether, represented parts of the mitochondrial genome. After removal of these 201+74 RSTs, 4865 RSTs constituted the database used in the subsequent comparisons. Apart from mitochondrial or ribosomal DNA, the (G+C) content of *S. uvarum* DNA amounted to 39.3%, identical to that of *S. cerevisiae* [13]. As the sequences of *S. uvarum* and *S. cerevisiae* are highly similar, we were able to measure an average of 0.57 frameshift/kb and of 0.12 gap/kb as sequencing artifacts.

3.2. Comparison with S. cerevisiae sequences

S. uvarum homologues to 2789 different open reading frames (ORFs) from *S. cerevisiae* were found in non-ambiguous matches, and another 25 in ambiguous matches, out of a total of 4326 RSTs. However, a closer examination of non-ambiguous matches coordinates within the hits and from contig information, allowed us to find ORFs in several copies in *S. uvarum*, which were represented by only singletons in *S. cerevisiae*. The minimal number of gene copies necessary to account for the alignments is reported in Table 1A. Thus 43 ORFs have to be added to the sum of identified ORFs. The same type of analysis on ambiguous matches showed that these correspond to paralogues in *S. cerevisiae* (Table 1B).

Inserts containing centromeres were detected on the basis of the ORFs flanking a centromere in *S. cerevisiae* (Table 2A). *Su*CEN11 and *Su*CEN15 were compared with their homologous centromeres in *S. cerevisiae* (Table 2B).

All sequences which had no match to an ORF in *S. cerevisiae* were compared to *S. cerevisiae* tRNA sequences [12], which defined 14 families out of the 42 found in *S. cerevisiae*. A minimum of 34 different genes were identified, indicating that each tRNA gene exists in several copies as in *S. cerevisiae*. In all cases, the respective anticodon is conserved as well as an intron if present. In some cases the *S. uvarum* introns were found to be longer by one or two nucleotides than their homologues in *S. cerevisiae*. With a genome coverage of 45%, about 120 tRNA genes were expected. Assuming that our set of 34 sequences is not biased, this discrepancy probably reflects a smaller complement of tRNA genes in *S. uvarum*.

Comparison of the RSTs to *S. cerevisiae* Ty nucleotide sequences or to Ty proteins yielded (i) matches to Ty4 or Ty1/Ty2 proteins (27 inserts), (ii) matches to non-coding DNA of Ty (11 inserts) with alignments against Ty1/Ty2 giving the best scores, probably corresponding to solo LTRs. As these alignments did not discriminate between Ty1 and Ty2 sequences, possibly only one type of Ty1/Ty2 homologue exists in *S. uvarum*. Further, there was no evidence that retroelements similar to Ty3 or Ty5 occur in *S. uvarum*. Thus it appears that *S. uvarum* contains fewer types of retroelements than *S. cerevisiae* [14], because otherwise some 100 matches were expected.

The rDNA repeat length was found to be 8226 nucleotides, whereby the 18S, 5.8S and 26S sequences are highly conserved [15]. About 500 RSTs were expected instead of the 201 we observed. If there is no sequence bias, this together with the low number of tRNA genes might reflect that *S. uvarum* has a lower rate of protein synthesis metabolism than *S. cerevisiae*, in accordance with the fact that fermentation in *S. uvarum* generally takes place at a lower temperature than in *S. cerevisiae*.

The map locating all matches to *S. cerevisiae* along its respective chromosomes indicated that equivalents to *S. cerevisiae* chromosomes were evenly represented in *S. uvarum*, as expected from the linear relationship between the number of identified ORFs and chromosome length [16].

The average similarity was found to be 84.5% with a standard deviation of 11.4%. The average length of an alignment was 440 nucleotides (±290 nucleotides), measured from the extreme coordinates of the alignment, ignoring frameshifts or gaps. Only 46 alignments longer than 30 amino acids showed complete identity, whereby only 16 of them had an identity greater or equal to 90% at the nucleotide level, none of them reaching 100% identity, confirming that there was no *S. cerevisiae* sequence in our inserts.

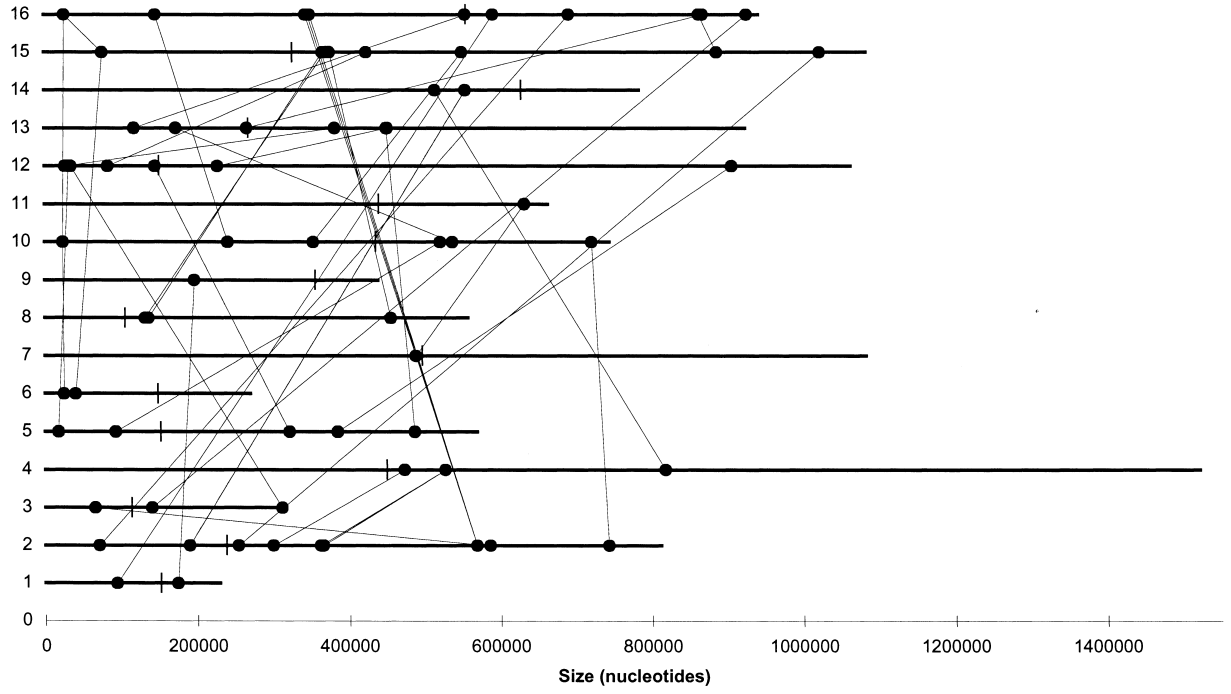

Fig. 1. Localization of *S. uvarum* non-synteny pairs on *S. cerevisiae* chromosomes. Non-synteny pairs, corresponding to a single insert in *S. uvarum*, are represented as dots joined by a thin line. The two hits are on different chromosomes in *S. cerevisiae*. Only non-redundant pairs are shown (i.e. *SuYAL028w–SuYPR015c, SuYAR019c–SuYIL089w, SuYBL017c–SuYNL040w, SuYBL082c–SuYPR072w, SuYBR008c–SuYOR363c, SuYBR030w–SuYDR012w, SuYBR060c–SuYDR037w, SuYBR061c–SuYDR037w, SuYBR164c–SuYPL108w, SuYBR164c–SuYPL109c, SuYBR164c–SuYPL110c, SuYBR177c–SuYCL031c, SuYBR269c–SuYJR152w, SuYCR014c–SuYPR194c, SuYCR106w–SuYLL054c, SuYDR177w–SuYNL059c, SuYEL031w–SuYJR045c, SuYEL071w–SuYLL055w, SuYER080w–SuYLL003w, SuYER111c–SuYLR392c, SuYER157w–SuYMR090w, SuYFL048c–SuYOL130w, SuYFL053w–SuYJL217w, SuYGL006w–SuYKR097w, SuYHR013c–SuYOR019w, SuYHR015w–SuYOR018w, SuYHR176w–SuYOR022c, SuYJL046w–SuYOR119c, SuYJL099w–SuYPL216w, SuYJL217w–SuYPL273w, SuYJR054w–SuYML051w, SuYLL029w–SuYOR049c, SuYLL057c–SuYMR053c, SuYLR039c–SuYMR091c, SuYML002w–SuYPR162c, SuYML075c–SuYPL002c, SuYOL130w–SuYPL273w, SuYOR304w–SuYPR160w*). Vertical bars represent centromeres. The horizontal bars represent chromosomes.

3.3. Comparisons with other genomes

In all, only 13 new genes, such as α-galactosidase, were detected in *S. uvarum* (Table 3). The *S. cerevisiae* S288C reference strain is Mel$^-$ but *S. uvarum* and brewer's yeast strains are Mel$^+$ as they contain the *MEL* genes [17]. Other matches include two putative MSF pumps, two proteins probably implicated in oxidoreduction, and two metabolic enzymes. Perhaps these proteins belong to at least one metabolic pathway

Table 2
Centromeres of *S. uvarum*

A

Inserts	Corresponding centromere
SuYHL002w–SuYHR001w	CEN8
SuYIL001w–SuYIR001c	CEN9
SuYJL002c–SuYJR001w	CEN10
SuYKL003c–SuYKR001c	CEN11
SuYML001w–SuYMR001c	CEN13
SuYML002w–SuYMR001c	
SuYNL001w–SuYNR001c	CEN14
SuYOL001w–SuYOR001w	CEN15
SuYOL002c–SuYOR001w	

B

Centromere	Centromere elements		
	CDE I	CDE II	CDE III
*Sc*CEN11	GTCACATG	85 bp (95% AT)	TGTTCATGATTTCCGAACGTATAAA
*Su*CEN11	– ––––––	86 bp (87% AT)	– ATT– –A– ––––A–A–T
*Sc*CEN15	ATCACGTG	86 bp (91% AT)	TGTATATGACTTCCGAAAAATATAT
*Su*CEN15	– ––––––	86 bp (87% AT)	– –––––––T– –––––C–A–A–

A: Inserts probably encompassing a centromere. Each line corresponds to an insert. Matches found in the inserts are indicated side by side. The centromeres are those of *S. cerevisiae*. B: Sequence comparisons between centromere elements [21] of *S. uvarum* and *S. cerevisiae*. Dashes correspond to identical nucleotide residues. For CDE II, length and percentage of (A+T) content are indicated.

Table 3
Potential functions encoded by *S. uvarum* RSTs having no validated homologues in the genome of *S. cerevisiae*

Organism	Accession number	Putative function and ORF name
Bacteria		
Bacillus subtilis	P49852	flavohemoglobin (HMPA_BACSU)
Coxiella burnetii	P51837	ribonuclease III (RNC_COXBU)
Mycobacterium tuberculosis	CAA17396.1	oxidoreductase (Rv0439c)
Rhodococcus rhodochrous	Q03217	aliphatic nitrilase (NRL2_RHORH)
Ascomycetes		
Kluyveromyces lactis	Q00970	mitochondrial genome maintenance protein (M101)
K. lactis	P09806	protein from killer plasmid (RF4)
Schizosaccharomyces pombe	S62589	unknown (SPAC21E11.04)
S. pombe	Q92341	drug efflux protein MSF (SPBC4F6.09)
S. pombe	CAA22200.1	putative MSF transporter (SPBC1271.10c)
S. pombe	Q09833	nuclease (YAD7)
Other eukaryotes		
Caenorhabditis elegans	T15878	unknown (D1009.4)
Coffea arabica	Q42656	α-galactosidase (A-GAL-COF)
Homo sapiens	O43488	aflatoxin aldehyde reductase (AFLA-HUM)

Putative functions and accession numbers are those indicated in SwissProt, PIR, WormPD, or ENTREZ files.

absent in *S. cerevisiae*, which could account for the ecological differences between the two species.

3.4. Synteny

2166 pairs of neighboring ORFs were obtained from non-ambiguous annotations of the inserts. In 2121 pairs, the *S. uvarum* homologues were located on the same chromosome as in *S. cerevisiae*, whereas in 45 pairs the homologues were located on different *S. cerevisiae* chromosomes. These latter pairs indicate that 38 distinct chromosomal rearrangements must have occurred (Fig. 1) during speciation of the two genomes. This figure when extrapolated to the whole genome would correspond to 84 chromosomal rearrangements. Recently, four chromosomal translocations have been characterized in *S. uvarum* CBS 7001 strain [10] used in this study. Two of these translocations are represented among our non-synteny pairs (Fig. 1): a translocation II–IV comprising the pair *SuYDR012w–SuYBR030w*, and a translocation VIII–XV comprising the pairs *SuYOR018w–SuYHR015w* and *SuYOR019w–SuYHR013w*. Translocation VI–X has not been included in Fig. 1, because in the pair *SuYJL046w–SuYOR119c*, the breakpoint on chromosome X is linked to an ORF from the homologue of *S. cerevisiae* chromosome XV right arm (translocated to chromosome VIII in *S. uvarum*). Apart from a chimeric insert, several hypotheses can be proposed: (i) the *SuYJL046w* ORF is duplicated in the homologue of chromosome XV right arm (and can even be absent from chromosome X), (ii) *SuYOR119c* has been inserted in chromosome X, near *SuYJL046w* and the translocation breakpoint. Altogether, the non-synteny pairs indicate far more rearrangements (about 80) to have occurred than the translocations described in [10] might imply (see [18]). Other events than translocations, such as in situ inversions (three occurrences) or deletions (one occurrence), seem to be rare.

4. Conclusions

This study differentiates more precisely *S. bayanus* var. *uvarum* from *S. cerevisiae*. (i) The average similarity between the orthologous protein sequences was about 85%, much lower than the intraspecific allelic variations in *S. cerevisiae* [19], and very few genes of *S. uvarum* have no orthologue in *S. cerevisiae*; (ii) in spite of a similar karyotype, huge remodelling occurred in chromosome organization between the two species (about 80 events). This may account, at least in part, for the high sterility of the hybrids [7,10]. Facts (i) and (ii) allow to assess the relative speed of mutational evolution of the genes (15% at the protein level) and synteny disruption events (circa 80) since *S. uvarum* and *S. cerevisiae* diverged from their common ancestor. (iii) The paucity of rDNA and tRNA genes in *S. uvarum* could be related to differences in ecology of the two species, *S. cerevisiae* growing preferentially at a higher temperature than *S. uvarum*. Alternatively, ecological differences may originate from the averaged 15% of protein sequence variation, though it is surprising that such sequence divergence has not generated more obvious physiological differences (see [20]). (iv) Absence of bona fide *S. cerevisiae* sequences in our samples is in accordance with previous work [8]. Although this does not prove that *S. bayanus* var. *uvarum* is totally devoid of such sequences, it strongly suggests that 'pure' *S. bayanus* var. *uvarum*, without *S. cerevisiae* DNA introgression, can exist in nature.

Acknowledgements: We thank E. Louis for kindly providing *S. uvarum* strain 623-6C *ura3-1*. We are grateful to G. Naumov and E. Naumova for helpful discussion. Elisabeth Bon was supported by the Comité Interprofessionnel des Vins de Bordeaux. This work was supported in part by a BRG Grant (Ressources génétiques des microorganismes no. 11-0926-99).

References

[1] Vaughan-Martini, A. and Martini, A. (1998) in: The Yeasts, a Taxonomic Study (Kurtzman, C.P. and Fell, J.W., Eds.), 4th edn., pp. 358–371, Elsevier, Amsterdam.
[2] Vaughan-Martini, A. and Kurtzman, C.P. (1985) Int. J. Syst. Bacteriol. 35, 508–511.
[3] Kishimoto, M., Shinohara, T., Soma, E. and Goto, S. (1993) J. Ferment. Bioeng. 75, 451–453.
[4] Nguyen, H.V. and Gaillardin, C. (1997) Syst. Appl. Microbiol. 20, 286–294.
[5] Naumov, G., Naumova, E. and Gaillardin, C. (1993) Syst. Appl. Microbiol. 16, 274–279.
[6] Rainieri, S., Zambonelli, C., Hallsworth, J.E., Pulvirenti, A. and Giudici, P. (1999) FEMS Microbiol. Lett. 177, 177–185.
[7] Naumov, G.I. (1987) Stud. Mycol. 30, 469–475.
[8] Louis, E.J., Naumova, E.S., Lee, A., Naumov, G. and Haber, J.E. (1994) Genetics 136, 789–802.
[9] Cardinali, G. and Martini, A. (1994) Int. J. Syst. Bacteriol. 44, 791–797.

[10] Fisher, G., James, S.A., Roberts, I.N., Oliver, S.G. and Louis, E.J. (2000) Nature 405, 451–454.
[11] Artiguenave, F., Wincker, P., Brottier, P., Duprat, S., Jovelin, F. et al., FEBS Lett. 487, 13–16 (this issue).
[12] Tekaia, F., Blandin, G., Malpertuy, A., Llorente, B., Durrens, P. et al., FEBS Lett. 487, 17–30 (this issue).
[13] Dujon, B. (1996) Trends Genet. 12, 263–270.
[14] Boeke, J.D. and Sandmeyer, S.B. (1991) in: The Molecular and Cellular Biology of the Yeast *Saccharomyces cerevisiae* (Broach, J.R., Pringle, J.R. and Jones, E.W., Eds.), pp. 193–261, Cold Spring Harbor Laboratory Press, New York.
[15] Llorente, B., Durrens, P., Tekaia, F., Aigle, M., Artiguenave, F. et al., FEBS Lett. 487, 71–75 (this issue).
[16] Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M. et al., FEBS Lett. 487, 3–12 (this issue).
[17] Naumov, G.I., Naumova, E.S. and Louis, E.J. (1995) Yeast 11, 481–483.
[18] Malpertuy, A., Llorente, B., Neuvéglise, C., de Montigny, J., Aigle, M. et al., FEBS Lett. 487, 61–65 (this issue).
[19] Carvajal, E., van den Hazel, H.B., Cybularz-Kolaczkowska, A., Balzi, E. and Goffeau, A. (1997) Mol. Gen. Genet. 256, 406–415.
[20] Gaillardin, C., Casaregola, S., Duchateau-Nguyen, G., Toffano-Nioche, C., Llorente, B. et al., FEBS Lett. 487, 134–149 (this issue).
[21] Hieter, P., Pridmore, D., Hegemann, J.H., Thomas, M., Davis, R.W. and Philippsen, P. (1985) Cell 42, 913–921.