FEBS Letters 487 (2000) 52–55 FEBS 24372

Genomic Exploration of the Hemiascomycetous Yeasts: 8. *Zygosaccharomyces rouxii*[1]

Jacky de Montigny[a,*], Marie-Laure Straub[a], Serge Potier[a], Fredj Tekaia[b], Bernard Dujon[b], Patrick Wincker[c], François Artiguenave[c], Jean-Luc Souciet[a]

[a] *Laboratoire de Génétique et Microbiologie, UPRES-A 7010 ULP/CNRS, Institut de Botanique, 28 rue Goethe, F-67000 Strasbourg, France*
[b] *Unité de Génétique Moléculaire des Levures, Institut Pasteur/CNRS, Département des Biotechnologies, Institut Pasteur, 25 rue du Docteur Roux, F-75724 Paris Cedex 15, France*
[c] *Génoscope, Centre National de Séquençage, 2, rue Gaston Crémieux, P.O. Box 191, 91057 Evry Cedex, France*

Received 3 November 2000; accepted 9 November 2000

First published online 27 November 2000

Edited by Horst Feldmann

Abstract This paper reports the genomic analysis of strain CBS732 of *Zygosaccharomyces rouxii*, a homothallic diploid yeast. We explored the sequences of 4934 random sequencing tags of about 1 kb in size and compared them to the *Saccharomyces cerevisiae* gene products. Approximately 2250 nuclear genes, 57 tRNAs, the rDNA locus, the endogenous pSR1 plasmid and 15 mitochondrial genes were identified. According to 18S and 25S rRNA cladograms and to synteny analysis, *Z. rouxii* could be placed among the *S. cerevisiae* sensu lato yeasts.

Key words: Comparative genomics; Random sequence tag; Hemiascomycete; *Saccharomyces cerevisiae*

1. Introduction

Zygosaccharomyces rouxii, a homothallic diploid, was first described as *Saccharomyces rouxii* [1] and classified close to *Saccharomyces cerevisiae* on the basis of rDNA sequence comparisons [2].

This hemiascomycete is considered to be a bakery product spoilage yeast since it is halotolerant and osmoresistant. These properties are currently under study. Some of the genes involved in salt tolerance (Zr*PMA1*, Zr*SOD2* [3,4] and Zr*SOD22* [5]) and in osmotic stress response (Zr*HOG1* and Zr*HOG2* [6]) have been cloned and sequenced. In parallel two auxotrophic markers *ADE2* [7] and *HIS3* [8] have been cloned and sequenced.

Z. rouxii contains a plasmid called pSR1 of 6251 bp which shows no homology with the *S. cerevisiae* 2μ plasmid [9]. It has been extensively characterized especially for the recombinases Flp and R [10–14] and its ability to be active in heterologous backgrounds [15,16].

This paper reports the genomic analysis of strain CBS732 which contains seven chromosomes with sizes ranging from 1 to 2.75 Mb corresponding to an estimated total size of 12.8 Mb according to pulsed field gel electrophoresis determination [17]. We explored the sequences of 4934 random sequencing tags (RSTs) of about 1 kb and compared them to the *S. cerevisiae* gene products as described in [18]. Approximately 2250 nuclear genes, 57 tRNAs, the rDNA locus, the endogenous pSR1 plasmid and 15 mitochondrial genes were identified.

2. Materials and methods

2.1. Strain

The yeast *Z. rouxii* strain used was CBS732.

2.2. DNA extraction, library construction, sequencing, RST analysis and annotations

Protocols described in [18–20] were followed with two modifications: the genomic DNA was partially digested (size ranging from 2 to 6 kb) with the endonuclease *Sau*3A and the inserts were cloned in plasmid pRS314 [21].

3. Results and discussion

3.1. Characteristics of the produced sequences

A total of 4934 RSTs which represent 4 367 106 nucleotides have been sequenced by Génoscope with less than 1% ambiguous nucleotides. The deduced GC content is 39.9%. The RSTs have an average length of 885 nucleotides with a standard deviation of 306. The number of inserts sequenced on both ends was 2348 representing 4696 RSTs (95% of all RSTs) with 63 overlapping. The remaining 238 RSTs (5%) were sequenced only at one end.

Using 88 randomly chosen clones, we determined the size of the inserts by digesting the plasmids with *Eco*RI. A total of 69 inserts were 3–5 kb in length, the average size amounting to 3.6 kb with a standard deviation of 1.8 kb. These results are consistent with the expected size of the library inserts prepared to range between 2 and 6 kb.

Independently of pairing RSTs by the inserts, 3200 RSTs could be grouped into 964 contigs consisting of 2–41 RSTs, their sizes ranging from 1309 to 4703 nucleotides. Two thirds of these contigs were composed of two RSTs. Two larger contigs of 250 and 274 RSTs corresponding to the rDNA sequences and the plasmid pSR1 detailed below could be assembled.

3.2. Protein coding genes with orthologues in S. cerevisiae

Open reading frames (ORFs) were identified in all six pos-

*Corresponding author. Fax: (33)-3-88 35 84 84.
E-mail: demontigny@gem.u-strasbg.fr

[1] The sequences have been deposited with EMBL under the accession numbers AL392203–AL397138.

PII: S0014-5793(00)02279-1

Table 1
Distribution of nuclear tRNA genes in *Z. rouxii*

	Anti-codon	Intron length in *Z. rouxii*	Intron length in *S. cerevisiae*	Number of *S. cerevisiae* tRNA genes in the corresponding family	Number of RSTs matching the tRNA gene	Number of contigs matching the tRNA gene	Minimal number of *Z. rouxii* tRNAs deduced from RSTs and contig numbers	Remarks
tRNA-Ala	AGC			11	6	2	2	
tRNA-Cys	GCA			4	3	1	1	
tRNA-Asp	GUC			15	4	2	2	
tRNA-Glu3	UUC			13	5	2	3	duplication on one RST
tRNA-Phe	GAA	18	18	8	3	2	2	
tRNA-Gly2	UCC			3	5	1	1	
tRNA-Gly1	GCC			16	7	3	3	
tRNA-His2	GUG			7	4	3	3	
tRNA-Ile	UAU	63	60	1	1	–	1	
tRNA-Ile	AAU			13	4	2	2	
tRNA-Lys2	UUU	23	23	7	3	1	1	
tRNA-Lys1	CUU			14	3	2	2	
tRNA-Leu1	UAG	22	20	3	2	1	1	
tRNA-Leu4	UAA			7	1	–	2	duplication on one RST
tRNA-Leu3	CAA	39	32	10	3	1	1	
tRNA-Met1	CAU			5	2	1	1	
tRNA-Met3	CAU			5	1	–	1	
tRNA-Asn	GUU			10	2	2	2	
tRNA-Pro	UGG	0	31	9	4	1	2	duplication on one RST
tRNA-Pro2	AGG			2	2	1	1	
tRNA-Gln	UUG			1	3	2	2	
tRNA-Arg2	UCU			11	1	–	1	
tRNA-Arg3	CCG			1	1	–	1	
tRNA-Arg3	ACG			1	2	1	1	
tRNA-Ser2	GCU	18	19	4	2	1	1	
tRNA-Ser2	CGA				2	2	2	
tRNA-Ser2	AGA			11	3	2	2	
tRNA-Thr	UGU			1	2	1	1	
tRNA-Thr	CGU			1	2	1	1	
tRNA-Thr	AGU			11	6	2	2	
tRNA-Val	CAC			2	2	2	2	
tRNA-Val1	AAC			13	2	2	2	
tRNA-Trp	CCA	29	34	6	4	1	1	
tRNA-Tyr	GUA	14	14	8	5	3	3	
Total							56	

sible frames, translated into putative proteins, compared to the *S. cerevisiae* database as described [19], and validated after manual examination of the position of the sequence on RSTs and evaluation of the identity/similarity results. BLASTX results of the 4934 RSTs showed that 1821 ORFs translated from these (36.9%) had only one valid match with a *S. cerevisiae* protein, 577 (11.7%) had at least two valid 'o' matches [19], 310 (6.3%) had an ambiguous 'oo' match and 65 (1.3%) had composite matches of 'o' and 'oo'. The analysis of 2023 RSTs (41%) showed matches with no significant score values to be attributed to homologues in *S. cerevisiae*. Finally 140 comparisons (2.8%) did not show any match.

The non-ambiguous 'o' matches made it possible to identify the *Z. rouxii* homologues to 1872 *S. cerevisiae* ORFs. The analysis of ambiguous 'oo' RSTs yielded a set of 554 additional sequences belonging to 293 gene families. Amino acid identity/similarity was found to be in the range of 57–72%.

The positions of *S. cerevisiae* ORFs identified as non-ambiguous 'o' and ambiguous 'oo' orthologues of *Z. rouxii* were plotted on the complete *S. cerevisiae* map [22]. The matches are equally distributed on each of the chromosomes except for the right arm of chromosome I, for which no match was found. The random distribution of ORFs along the chromosomes indicates that the library used in this work as well as the number of random RSTs examined are representative of the *Z. rouxii* genome. The absence of matches with the right arm of chromosome I could be interpreted to indicate an absence of this region in the *Z. rouxii* genome.

3.3. Comparison with other S. cerevisiae sequences

3.3.1. tRNA genes. All RSTs matching a tRNA sequence were annotated as 'oo' since all of them belong to defined families. The comparison of the RSTs with the 42 tRNA gene families of *S. cerevisiae* [23] identified 34 families in *Z. rouxii*. Considering the number of different contigs matching the same tRNA gene, a minimal number of 56 tRNA gene sequences was identified (Table 1). Since the coverage of the genome in our library is approximately 30%, we could expect a total number of 190 tRNA genes, which is lower than the number of 274 tRNA genes found in *S. cerevisiae*, provided that no bias was introduced in *Z. rouxii* by the cloning procedure.

For all the tRNA genes identified here, the anticodon was found to be conserved between the two species. Regarding

Table 2
Putative mitochondrial genes found in *Z. rouxii*

Gene name	Gene product
15S-rDNA	15S ribosomal RNA
21S-rDNA	21S ribosomal RNA
tRNA-Pro	proline specific tRNA
AI3/I-SceIII	mobile group I intron of mitochondrial COX1
AI5α	intron of mitochondrial COX1
AI5β	intron of mitochondrial COX1
ATP6/OLI4	ATP synthase subunit 6
ATP9/OLI1	F_o-ATP synthase subunit 9
BI3	mRNA maturase bI3
BI4	mRNA maturase bI4
COB	cytochrome *b*
COX1	cytochrome *c* oxidase subunit I
COX2	cytochrome *c* oxidase subunit II
COX3	cytochrome *c* oxidase subunit III
VAR1	structural protein of ribosome

tRNA introns, one tRNA-Pro contains an intron in *S. cerevisiae* which is absent in the homologue of *Z. rouxii*. In two cases, a tRNA-Glu and a tRNA-Leu gene sequence were found to be duplicated on the same RST.

3.3.2. rDNA. A total of 371 RSTs matched significantly with genomic rDNA sequences of *S. cerevisiae*. These RSTs mainly belong to two contigs of 2278 and 8198 nucleotides which could be assembled in a unique sequence of 9674 nucleotides encoding 25S, 18S, 5.8S and 5S rRNAs which respectively exhibit 95%, 98%, 98% and 87% identity with their *S. cerevisiae* counterparts. These genes have the same organization as those of *S. cerevisiae*. No repetitive sequences were detected at both sides of the 9674 nucleotides contig suggesting that it probably does not cover a complete unit and that the rDNA repeat of *Z. rouxii* is longer than that of *S. cerevisiae*, which is 9.1 kb in length.

3.3.3. Transposons. Transposable elements are ubiquitous in eukaryotes. In *S. cerevisiae*, 52 full-length retrotransposable elements were found as well as 268 solo long terminal repeats [23], which have been attributed to five classes, Ty1, 2, 4 and 5 belonging to the *Drosophila copia*-like family and Ty3 representing a member of the *gypsy*-like family [24].

All RSTs and contigs were compared with the different Ty element sequences identified in the *S. cerevisiae* genome but no significant matches were obtained at the nucleotide level. Comparisons of ORFs from the RSTs with the protein sequences TyA and TyB of the various Ty elements revealed matches of five of these with Ty3B. The scores of the matches were too low to be considered significant. Nonetheless, we cannot exclude that Ty3-like elements are present in *Z. rouxii* but with a sequence and/or an organization different from the *S. cerevisiae* Ty3. Since in *S. cerevisiae* the majority of the Ty elements are located upstream of tRNA genes [25], the RSTs containing tRNA genes were more carefully checked for retrotransposons, though with no success. Another explanation for the absence of significant matches with Ty elements is that similar elements in *Z. rouxii* are present in such a low copy number that we were not able to detect them. The most compelling assumption would be that *Z. rouxii* is devoid of Ty-like retrotransposon sequences, thus constituting a notable exception among the yeasts.

3.3.4. Mitochondrial DNA. The BLASTX comparison of the 4934 RSTs to the *S. cerevisiae* mitochondrial database revealed 40 RSTs with significant matches. The alignments allowed different proteins and mobile introns to be identified (Table 2). Concerning the RNA genes, one RST matched mitochondrial tRNA-Pro, three RSTs 21S rRNA, one 15S RNA and none 9S RNA. These RSTs belong to 14 different contigs which also contain the sequences of 31 RSTs showing no match with the mitochondrial sequences. If only one RST of an insert or a contig including the 40 identified RSTs matched a mitochondrial sequence and the second one showed no significant match, we supposed that both of them belonged to the *Z. rouxii* mitochondrial genome. By this proposition we identified another 45 putative mitochondrial RSTs. The total of 85 RSTs were included in 18 contigs representing 25.3 kb of mitochondrial DNA. Using this by excess size evaluation, the size of the entire mitochondrial

Table 3
Homologues of *Z. rouxii* in organisms other than *S. cerevisiae*

Bacteria		
Burkholderia cepacia	P16932	2,2-dialkylglycine decarboxylase (EC 4.1.1.64)
Campylobacter jejuni	HP0006	pantoate β-alanine ligase
Archaea		
Pyrococcus horikoshii	F71085	transcriptional regulatory protein
Ascomycetes		
Aspergillus niger	Q12556	copper amine oxidase 1 (EC 1.4.3.6)
Candida albicans	CAB65618	sorbitol utilization protein
Emericella nidulans	P08158	acetamidase (EC 3.5.1.4)
Emericella nidulans	P48777	purine permease
Emericella nidulans	Q07307	uric acid xanthine permease
Kluyveromyces lactis	P49374	high affinity glucose transporter
Kluyveromyces lactis	AJ243800	ubiquitin-like protein
Schizosaccharomyces pombe	P32747	dihydroorotate dehydrogenase
Schizosaccharomyces pombe	P78771	fission yeast protein P78771
Schizosaccharomyces pombe	Q09329	*Mlo2* protein involved in mitosis
Schizosaccharomyces pombe	Q10088	putative agmatinase precursor (EC 3.5.3.11)
Schizosaccharomyces pombe	CAA17034	hypothetical protein of the major facilitator superfamily
Schizosaccharomyces pombe	CAA19062	major facilitator superfamily
Schizosaccharomyces pombe	Q10082	hypothetical protein
Other eukaryotes		
Brassica juncea	Q39287	ω6 fatty acid desaturase
Drosophila melanogaster	P56538	eIF-6 eukaryotic initiation translation factor
Rhizobium	P55441	hypothetical monooxygenase

Potential functions encoded by *Z. rouxii* RSTs having no validated homologues in the genome of *S. cerevisiae*.

genome would be around 83 kb since one third of the genome was covered by our library. Clearly, this is close to the 86 kb of the *S. cerevisiae* mitochondrial DNA [26].

3.3.5. Plasmid. Strain CBS732 of *Z. rouxii* contains a plasmid called pSR1, 6251 bp in length. The comparison of its sequence with a large contig formed by 274 RSTs containing the entire sequence of the plasmid revealed less than 0.5% discrepancies at the nucleotide level.

3.3.6. Proteins with no orthologues in S. cerevisiae. The occurrence of yeast specific genes not found in *S. cerevisiae* has already been described for several yeast species [27,28]. Therefore, we performed a general comparison of the RSTs not attributable to genes in *S. cerevisiae* with several entirely sequenced genomes and the SwissProt database deprived of *S. cerevisiae* and homologous sequences [19]. We were thus able to identify 20 putative *Z. rouxii* genes (Table 3). Two particularities should be pointed out. First, in *Z. rouxii* a purine permease and a uric acid xanthine permease were identified, while in *S. cerevisiae* these products are transported by a unique pleiotropic permease [29]. Second, the dihydroorotate dehydrogenase was only identified by similarity with the *S. pombe* gene product. In contrast to *S. cerevisiae*, the latter has been localized in mitochondria, so that one has to argue that the *Z. rouxii* gene product is also a mitochondrial enzyme and not a cytoplasmic one as in *S. cerevisiae* [30].

In our analysis we also found the specific proteins of the pSR1 plasmid (see above), namely the *trans*-acting factors B and C encoded by the *REP1* and *REP2* genes and the recombinase FLP encoded by the *FLP* gene. The numbers of RSTs showing significant matches with the corresponding genes were 78, 66 and 98%, respectively.

3.4. Phylogenetic classification

The cladogram established from 18S and 25S rRNA sequence comparisons [22] and measurement of synteny [31] allowed us to unequivocally place *Z. rouxii* in the *S. cerevisiae* sensu lato group of yeasts.

Acknowledgements: We thank Catherine Spehner and Yves Tourrette for technical assistance and Cécile Neuvéglise for helpful discussions. This work was supported by Grant 11-0926-99 from BRG (Bureau des Ressources Génétiques). The yeast genetic department of the UPRES-A 7010 is a member of the 'Génopole Alsace-Lorraine'. B.D. is a member of Institut Universitaire de France.

References

[1] Boutroux, L. (1884) Ann. Sci. Nat. 17, 144–209.
[2] Wilmotte, A., Van de Peer, Y., Goris, A., Chapelle, S., De Baere, R., Nelissen, B., Neefs, J.M., Hennebert, G.L. and De Wachter, R. (1993) Syst. Appl. Microbiol. 16, 436–444.
[3] Watanabe, Y., Shiramizu, M. and Tamai, Y. (1991) J. Biochem. (Tokyo) 110, 237–240.
[4] Watanabe, Y., Miwa, S. and Tamai, Y. (1995) Yeast 11, 829–838.
[5] Iwaki, T., Higashida, Y., Tsuji, H., Tamai, Y. and Watanabe, Y. (1998) Yeast 14, 1167–1174.
[6] Iwaki, T., Tamai, Y. and Watanabe, Y. (1999) Microbiology 145, 241–248.
[7] Sychrova, H., Braun, V. and Souciet, J.L. (1999) Yeast 15, 1399–1402.
[8] Sychrova, H., Braun, V. and Souciet, J.L. (2000) Yeast 16, 581–587.
[9] Araki, H., Jearnpipatkul, A., Tatsumi, H., Sakurai, T., Ushio, K., Muta, T. and Oshima, Y. (1985) J. Mol. Biol. 182, 191–203.
[10] Matsuzaki, H., Araki, H. and Oshima, Y. (1988) Mol. Cell. Biol. 8, 955–962.
[11] Evans, B.R., Chen, J.W., Parsons, R.L., Bauer, T.K., Teplow, D.B. and Jayaram, M. (1990) J. Biol. Chem. 265, 18504–18510.
[12] Chen, J.W., Evans, B.R., Yang, S.H., Teplow, D.B. and Jayaram, M. (1991) Proc. Natl. Acad. Sci. USA 88, 5944–5948.
[13] Araki, H., Nakanishi, N., Evans, B.R., Matsuzaki, H., Jayaram, M. and Oshima, Y. (1992) J. Mol. Biol. 225, 25–37.
[14] Lee, J., Serre, M.C., Yang, S.H., Whang, I., Araki, H., Oshima, Y. and Jayaram, M. (1992) J. Mol. Biol. 228, 1091–1103.
[15] Onouchi, H., Yokoi, K., Machida, C., Matsuzaki, H., Oshima, Y., Matsuoka, K., Nakamura, K. and Machida, Y. (1991) Nucleic Acids Res. 19, 6373–6378.
[16] Onouchi, H., Nishihama, R., Kudo, M., Machida, Y. and Machida, C. (1995) Mol. Gen. Genet. 247, 653–660.
[17] Sychrova, H., Braun, V., Potier, S. and Souciet, J.L. (2000) Yeast (in press).
[18] Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B. (2000) FEBS Lett. 487, 31–36 (this issue).
[19] Tekaia, F., Blandin, G., Malpertuy, A., Llorente, B., Durrens, P. et al. (2000) FEBS Lett. 487, 17–30 (this issue).
[20] Artiguenave, F., Wincker, P., Brottier, P., Duprat, S., Jovelin, F. et al. (2000) FEBS Lett. 487, 13–16 (this issue).
[21] Sikorski, R.S. and Hieter, P. (1989) Genetics 122, 19–27.
[22] Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G. Bolotin-Fukuhara, M. et al. (2000) FEBS Lett. 487, 3–12 (this issue).
[23] Hani, J. and Feldmann, H. (1998) Nucleic Acids Res. 26, 689–696.
[24] Jordan, I.K. and McDonald, J.F. (1999) Genetics 151, 1341–1345.
[25] Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A. and Voytas, D.F. (1998) Genome Res. 8, 464–478.
[26] Foury, F., Roganti, T., Lecrenier, N. and Purnelle, B. (1998) FEBS Lett. 440, 325–331.
[27] Ozier-Kalogeropoulos, O., Malpertuy, A., Boyer, J., Tekaia, F. and Dujon, B. (1998) Nucleic Acids Res. 23, 5511–5524.
[28] Barth, G. and Gaillardin, C. (1997) FEMS Microbiol. Rev. 19, 219–237.
[29] Chevallier, M.R., Jund, R. and Lacroute, F. (1975) J. Bacteriol. 122, 629–641.
[30] Nagy, M., Lacroute, F. and Thomas, D. (1992) Proc. Natl. Acad. Sci. USA 89, 8966–8970.
[31] Llorente, B., Malpertuy, A., Neuvéglise, C., de Montigny, J., Aigle, M. et al. (2000) FEBS Lett. 487, 71–75 (this issue).