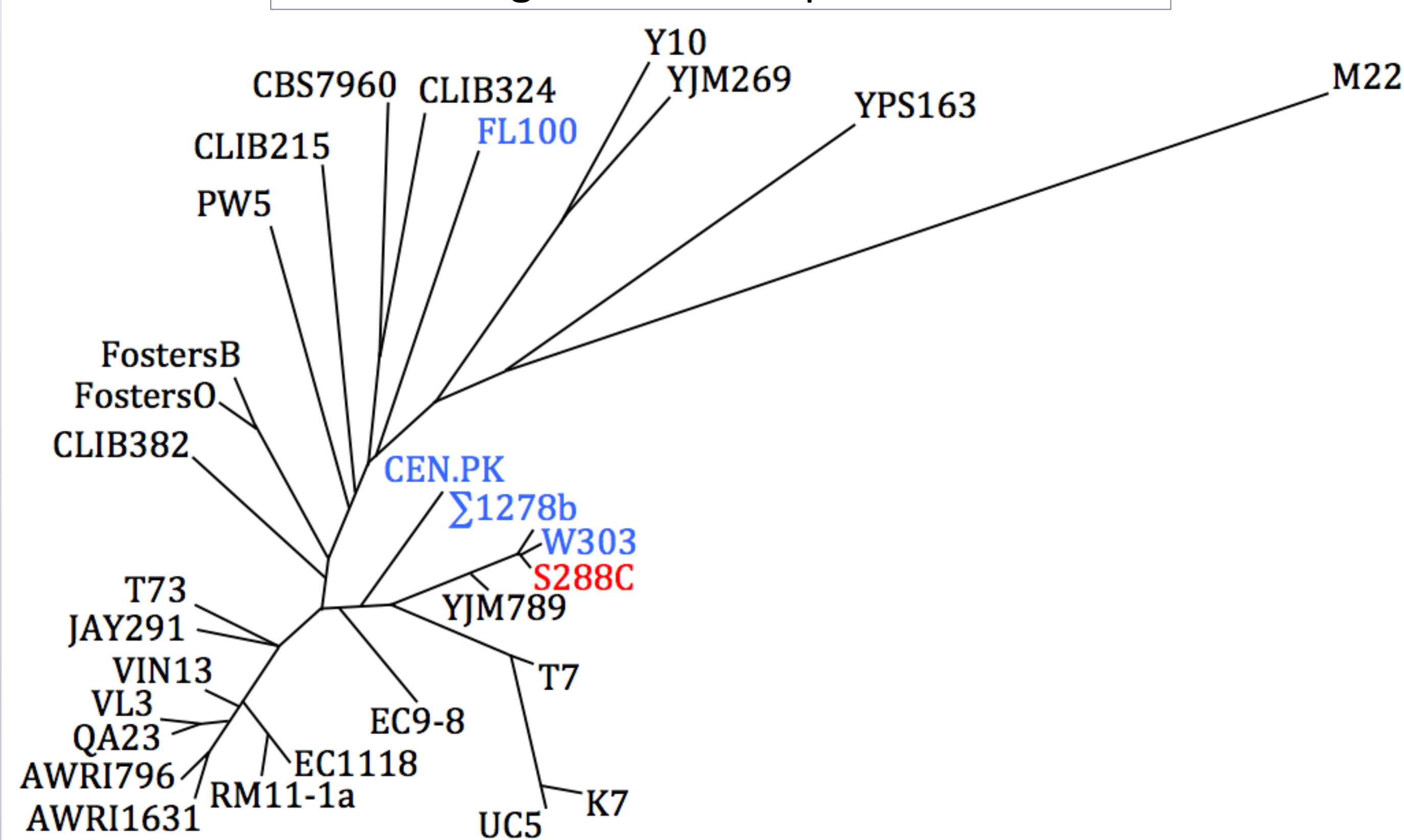


annotation of multiple *Saccharomyces cerevisiae* strains at SGD

Stacia R. Engel, J. Michael Cherry, and The SGD Project
Department of Genetics, Stanford University, Stanford, CA 94305, USA

The first completed eukaryotic genome sequence was that of the yeast *Saccharomyces cerevisiae*, and the *Saccharomyces* Genome Database (SGD; <http://www.yeastgenome.org/>) is the original model organism database. SGD remains the authoritative community resource for the *S. cerevisiae* reference genome sequence and its annotation, and continues to provide comprehensive biological information correlated with *S. cerevisiae* genes and their products. A diverse set of yeast strains have been sequenced to explore commercial and laboratory applications, and a brief history of those strains is provided. The publication of these new genomes has motivated the creation of new tools, and SGD will annotate and provide comparative analyses of these sequences, correlating changes with variations in strain phenotypes and protein function. We are entering a new era at SGD, as we incorporate these new sequences and make them accessible to the scientific community, all in an effort to continue in our mission of educating students, enabling bench researchers and facilitating scientific discovery.

Additional genomes incorporated into SGD

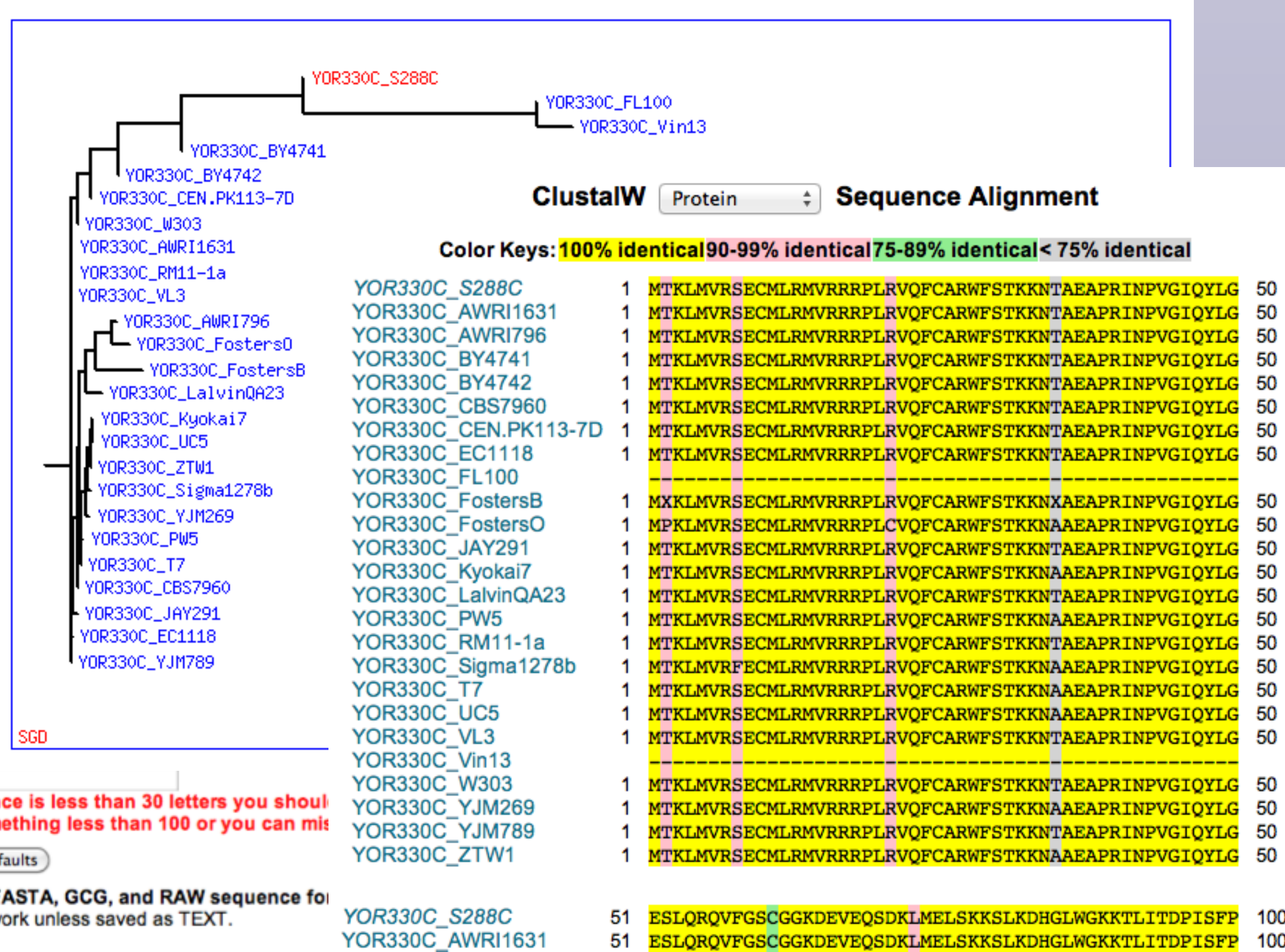


S. cerevisiae genomes contain ORFs not in S288C reference

Strain	ORFs not in S288C	Reference
AWRI796	74	Borneman <i>et al.</i> (2011)
CEN.PK113-7D	83	Nijkamp <i>et al.</i> (2012)
EC1118	77	Borneman <i>et al.</i> (2011)
Foster's B	36	Borneman <i>et al.</i> (2011)
Foster's O	48	Borneman <i>et al.</i> (2011)
JAY291	16	Argueso <i>et al.</i> (2009)
Kyokai No.7	48	Akao <i>et al.</i> (2011)
QA23	110	Borneman <i>et al.</i> (2011)
RM11-1a	38	Borneman <i>et al.</i> (2011)
Sigma1278b	75	Dowell <i>et al.</i> (2010)
VIN13	45	Borneman <i>et al.</i> (2011)
VL3	54	Borneman <i>et al.</i> (2011)
YJM789	34	Borneman <i>et al.</i> (2011)

Akao, T., Yashiro, I., Hosoyama, A. *et al.* (2011) Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no 7. *DNA Res.*, 18, 423-434.
 Argueso, J.L., Carrozzolle, M.F., Mieczkowski, P.A. *et al.* (2009) Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.*, 19, 2258-2270.
 Borneman, A.R., Desany, B.A., Riches, D. *et al.* (2011) Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.*, 7, e1001287.
 Dowell, R.D., Ryan, O., Jansen, A. *et al.* (2010) Genotype to phenotype: a complex problem. *Science*, 328, 469.
 Nijkamp, J.F., van den Broek, M., Datema, E. *et al.* (2012) De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.*, 11, 36.

MIP1/YOR330C *S. cerevisiae* Strain Sequence Alignment



In the years since the publication of the S288C genome, dozens of complete yeast genome sequences have been published

Strain	Year	Description	NCBI BioProject	Contig N50 ^a	Scaffold N50 ^a
S288C	1996	Laboratory strain	PRJNA128	N/A ^b	N/A
RM11-1a	2005	Haploid derivative of California vineyard isolate	PRJNA13674	263 288	795 018
YJM789	2007	Haploid derivative of opportunistic human pathogen	PRJNA13304	429 709	N/A
M22	2008	Italian vineyard isolate	PRJNA28815	2207	N/A
YPS163	2008	Pennsylvania woodland isolate	PRJNA28813	2901	N/A
AWRI1631	2008	Haploid derivative of South African commercial wine strain N96	PRJNA30553	7704	N/A
JAY291	2009	Haploid derivative of Brazilian industrial bioethanol strain PE-2	PRJNA32809	64 336	N/A
EC1118	2009	Commercial wine strain	PRJEA37863	776 014	N/A
Sigma1278b	2009	Laboratory strain	PRJNA39317	365 700	N/A
Foster's O	2010	Commercial ale strain	PRJNA48567	195 316	N/A
Foster's B	2010	Commercial ale strain	PRJNA48569	204 208	626 897
VIN13	2010	South African white wine strain	PRJNA48563	308 189	700 638
AWRI796	2010	South African red wine strain	PRJNA48559	403 341	565 854
CLIB215	2010	New Zealand bakery isolate	PRJNA60143	16 813	47 217
CBS7960	2011	Brazilian bioethanol factory isolate	PRJNA60391	18 761	65 099
CLIB324	2011	Vietnamese bakery isolate	PRJNA60415	4260	24 472
CLIB382	2011	Irish beer isolate	PRJNA60145	840	2711
EC9-8	2011	Haploid derivative of Israeli canyon isolate	PRJNA73985	15 539	541 605
FL100	2011	Laboratory strain	PRJNA60147	4244	26 506
Kyokai No.7	2011	Japanese sake yeast	PRJNA45827	120 978	902 266
QA23	2011	Portuguese Vinho Verde white wine strain	PRJNA48561	182 942	182 942
PW5	2011	Nigerian Raphia palm wine isolate	PRJNA60181	14 234	393 105
T7	2011	Missouri oak tree exudate isolate	PRJNA60387	147 205	476 142
T73	2011	Spanish red wine strain	PRJNA60195	2945	36 287
UC5	2011	Japanese sake yeast	PRJNA60197	17 142	356 094
VL3	2011	French white wine strain	PRJNA48565	293 399	656 188
W303	2011	Laboratory strain	PRJNA167645	149 943	367 966
Y10	2011	Philippine coconut isolate	PRJNA60201	2730	22 204
YJM269	2011	Austrian Blauer Portugieser wine grapes	PRJNA60389	23 452	58 353
BY4741	2012	S288C-derivative laboratory strain	N/A	N/A	N/A
BY4742	2012	S288C-derivative laboratory strain	N/A	N/A	N/A
CEN.PK 113-7D	2012	Laboratory strain	PRJNA52955	48 196	918 791
ZTW1	2012	Chinese corn mash bioethanol isolate	PRJNA174065	556 921	N/A

S. cerevisiae WU-BLAST2 Search

Query: YNR066C fused in RM11-1a, Sigma1278b, W303

Results: All hits shown. p=0. s=16755 ORF Uncharacterized [Sc strain BY4741] Protein of unknown function; pri

Split ORFs in BY4741, BY4742

Download strain sequences: <http://www.yeastgenome.org/download-data/sequence>
 Search sequence patterns: <http://www.yeastgenome.org/cgi-bin/PATMATCH/nph-patmatch>
 BLAST multiple strains: <http://www.yeastgenome.org/cgi-bin/blast-sgd.pl>
 Align multiple strains: <http://www.yeastgenome.org/cgi-bin/FUNGI/alignment.pl>

SGD tools provide access to strain genomes →

Coming soon to SGD! Select your reference strain...SEY6210, Cen.PK2, JK9-3d, SK1, W303, Sigma1278b, FL100...