Saccharomyces GENOME DATABASE

The modern era of yeast genomics:

Annotation of multiple *Saccharomyces cerevisiae* strains at SGD

Stacia R. Engel

Stanford University

www.yeastgenome.org

SGD Saccharomyces GENOME DATABASE

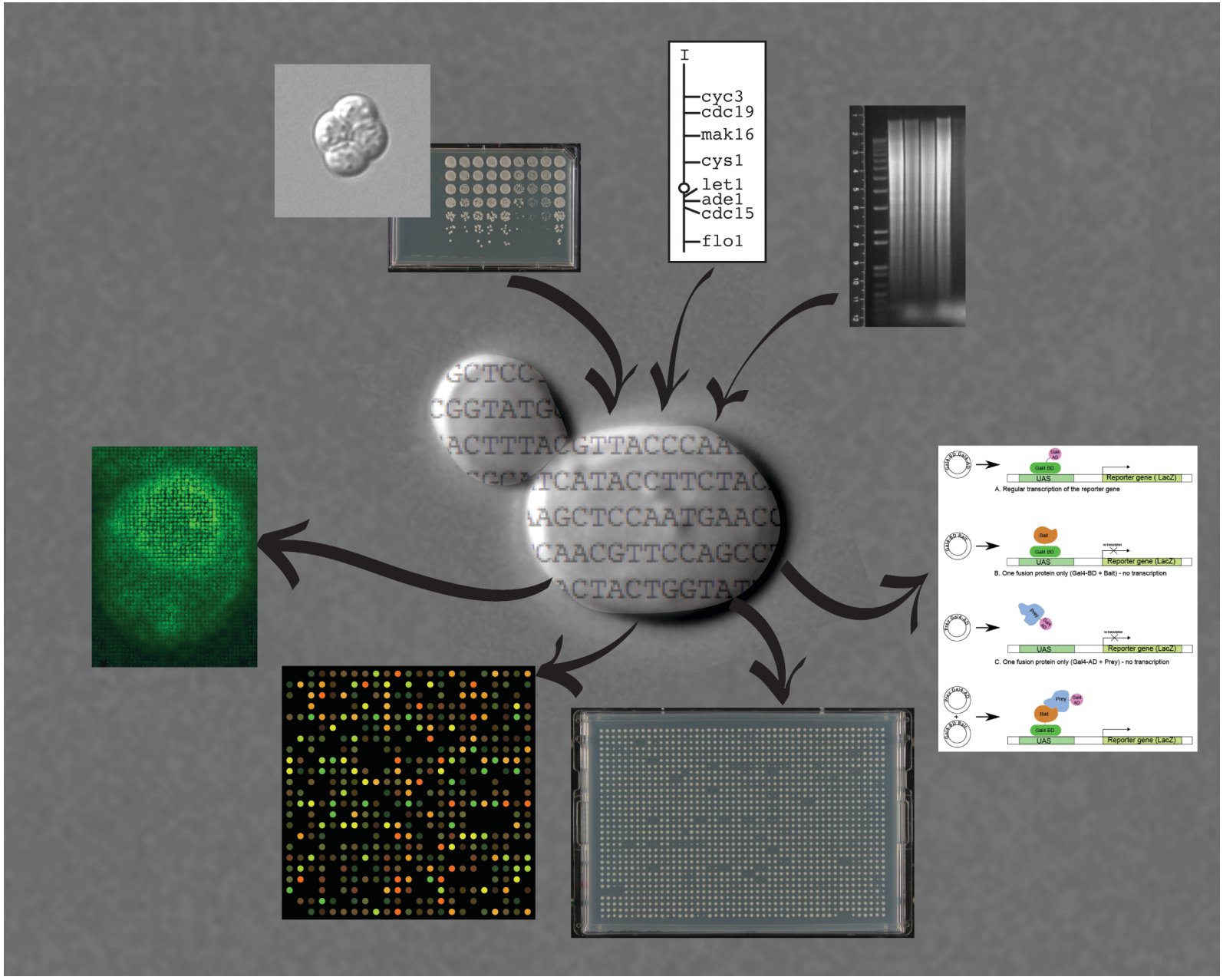
Our mission:

Educate students

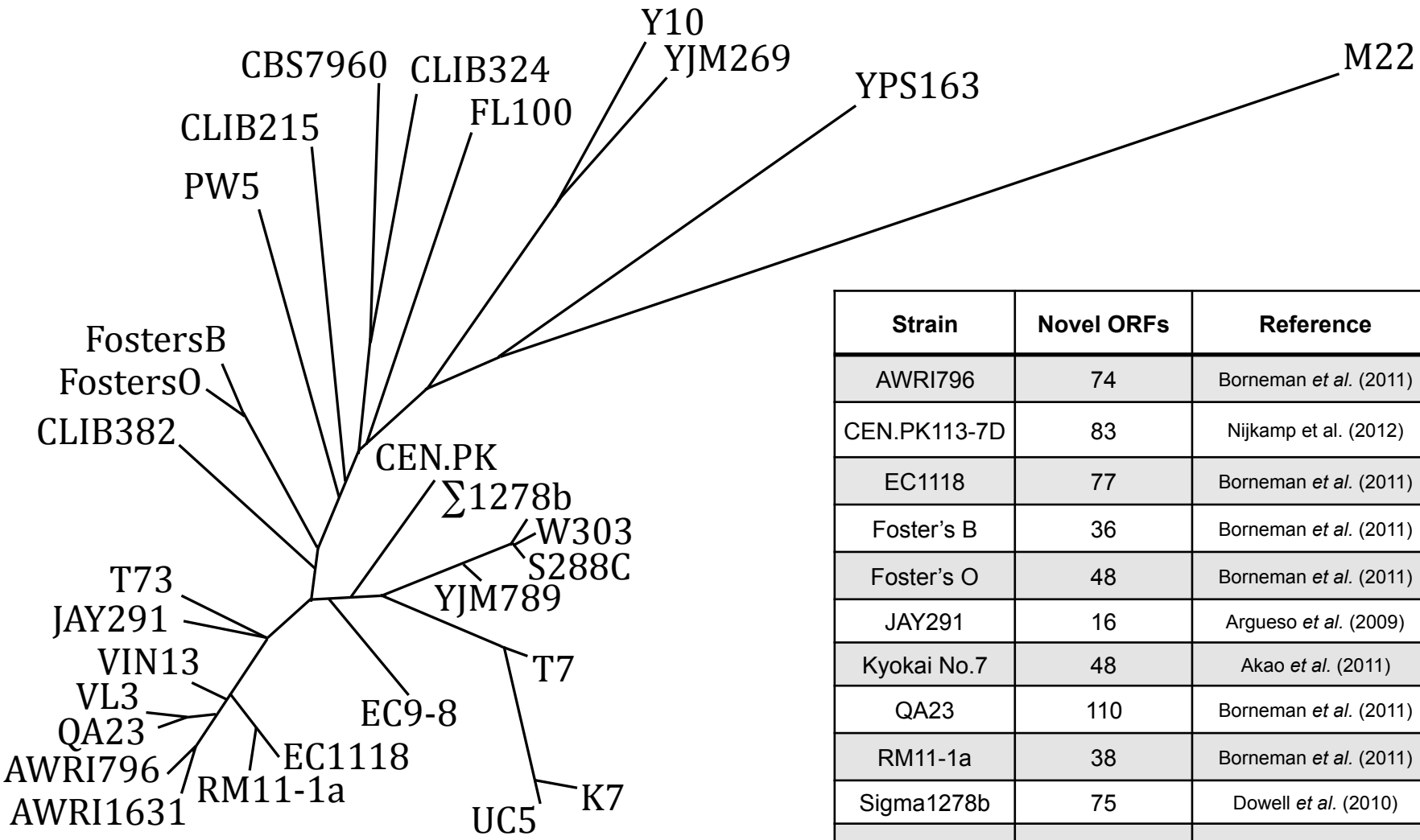
Enable researchers

Facilitate scientific discovery

Saccharomyces GENOME DATABASE



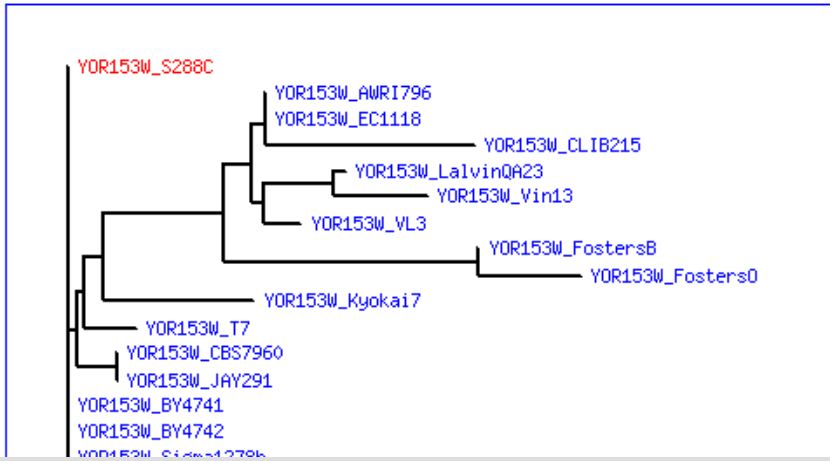
Saccharomyces GENOME DATABASE



Strain	Novel ORFs	Reference
AWRI796	74	Borneman *et al.* (2011)
CEN.PK113-7D	83	Nijkamp et al. (2012)
EC1118	77	Borneman *et al.* (2011)
Foster's B	36	Borneman *et al.* (2011)
Foster's O	48	Borneman *et al.* (2011)
JAY291	16	Argueso *et al.* (2009)
Kyokai No.7	48	Akao *et al.* (2011)
QA23	110	Borneman *et al.* (2011)
RM11-1a	38	Borneman *et al.* (2011)
Sigma1278b	75	Dowell *et al.* (2010)
VIN13	45	Borneman *et al.* (2011)
VL3	54	Borneman *et al.* (2011)
YJM789	34	Borneman *et al.* (2011)

Saccharomyces GENOME DATABASE

PDR5/YOR153W S. cerevisiae Strain Sequence Alignment



ClustalW Protein **Sequence Alignment**

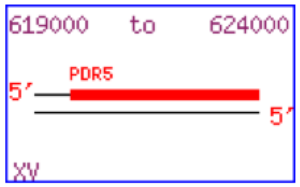
Color Keys: 100% identical | 90-99% identical | 75-89% identical | < 75% identical

```
YOR153W_S288C     1  MPEAKLNNNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_AWRI796   1  MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_BY4741    1  MPEAKLNNNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_BY4742    1  MPEAKLNNNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_CBS7960      --------------------------------------------------
YOR153W_CLIB215   1  MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_EC1118    1  MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_FostersB  1  MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFGEHTEARIQKLARTL  50
YOR153W_FostersO  1  MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFGEHTEARIQKLARTL  50
YOR153W_JAY291       --------------------------------------------------
YOR153W_Kyokai7   1  MPEAKLNNNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_LalvinQA23 1 MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_Sigma1278b 1 MPEAKLNNNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_T7        1  MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_VL3       1  MPEAKLNDNVNDVXSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_Vin13     1  MPEAKLNDNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_W303      1  MPEAKLNNNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50
YOR153W_YJM269    1  MPEAKLNNNVNDVTSYSSASSSTENAADLHNYNGFDEHTEARIQKLARTL  50

YOR153W_S288C    51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_AWRI796  51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_BY4741   51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_BY4742   51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_CBS7960      --------------------------------------------------
YOR153W_CLIB215  51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_EC1118   51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_FostersB 51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_FostersO 51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYXPKLDPNSE  100
YOR153W_JAY291       --------------------------------------------------
YOR153W_Kyokai7  51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_LalvinQA23 51 TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_Sigma1278b 51 TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_T7       51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_VL3      51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_Vin13    51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_W303     51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
YOR153W_YJM269   51  TAQSMQNSTQSAPNKSDAQSIFSSGVEGVNPIFSDPEAPGYDPKLDPNSE  100
```

SEQUENCE INFORMATION

ChrXV:619840 to 624375 | ORF Map | GBrowse



Genetic position: 85 cM

Last Update Coordinates: 2011-02-03 | Sequence: 1996-07-31

Subfeature details

	Relative Coordinates	Chromosomal Coordinates	Most Recent Updates Coordinates	Most Recent Updates Sequence
CDS	1..4536	619840..624375	*2011-02-03*	*1996-07-31*

Retrieve sequences S.c. Strain W303 | DNA Coding | Get Sequence

ANALYZE SEQUENCE

S288C only BLASTP | ATCC/WashU Clones | BLASTN | Design Primers | Restriction Fragment Map | Restr
Six-Frame Translation

S288C vs. other species Fungal Alignment | BLASTN vs. fungi | BLASTP at NCBI | BLASTP vs. fungi | Synteny Viewer

S288C vs. other strains Strain Alignment

Saccharomyces GENOME DATABASE

Home | Analyze | Sequence | Function | Literature | Community

Genome Browser
BLAST
Fungal BLAST
Gene / Sequence Resources
Reference Genome
Strains and species
Homology
Resources

Align Strain Sequences
Synteny Viewer

S BLAST2 Search

Datasets updated: September 6, 2011
This form allows BLAST searches of *S. cerevisiae* s rch multiple fungal sequences, go to the Fungal BLAST search form.

Query Comment (optional, will be added to output
YNR066C
NOTE: If the input sequence is less than 30 lette
Cutoff Score value to something less than 100 c

Run WU-BLAST or Select Defaults

Upload Local TEXT File: FASTA, GCG, and RAW okay
WORD Documents do not work unless saved as TEXT.
Choose File no file selected

Type or Paste a Query Sequence : (FASTA or RAW format, or No Comments, Numbers are okay)
ATGATATTACTCCAAGTCATATGCACGATTTGGACATGTCTCTTTATTCCGTTACTCAATGCAGAGGAATTCGTC
CCCAAAGTAACGGAGACTCTTTCAGAATATTCATTTAGTCTAGAGAGCTTTGACGATTCCAACAGTTTAATCAG
ATTAGATAATCAAGTCGTGTGGATAAGTTCCGATTCTGGAGAAAATTGGGAAGCGGTCAAAGAAATTGAAGGGC
ATATTCTCGAATTAATTGTTGATCCTTTGCATGGACAGGACAGGGCTTTTGTTTCGATACATTTATCACCCAAAT
TTTACGTCACCGATGATCGTGGAAAATCATGGAGGGCTC

Choose the Appropriate BLAST Program:
BLASTN – nucleotide query to nucleotide db

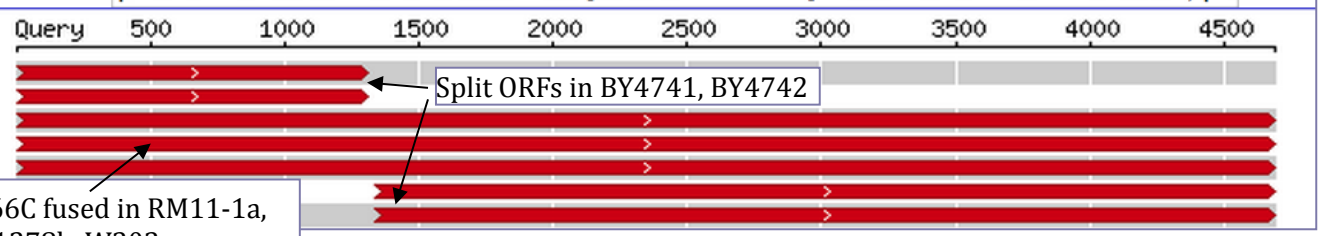
Choose one or more Sequence Datasets:
Select or unselect multiple datasets by pressin
while clicking. Selecting a category label select

REFERENCE (S288C) GENOMIC SEQUENCE
. . . Nuclear chromosomes (DNA)
. . . Mitochondrial chromosome (DNA)
. . . 2-micron plasmid (DNA)
GENES: PROTEIN ENCODING (S288C)
. . . Open Reading Frames (DNA or Protein)
. . . Genomic (Coding and Introns) Sequences of define
. . . Open Reading Frames + 1000 bp Up & Downstrear
GENES: ncRNA (S288C)
. . . RNA Coding (DNA)
. . . RNA Genomic (coding and introns) (DN
. . . RNA Genomic + 1000 bp Up & Downstr
NON-GENIC SEQUENCES (S288C)
. . . Intergenic Genomic DNA between ORFs
OTHER PUBLIC SEQUENCES
. . . Yeast public sequences from GenBank and UniProt (DNA or Protein)
. . . Yeast Cloning Vector Sequences from VectorDB (DNA)
COMPLETE S.c STRAIN GENOMES
. . . S. cerevisiae strain AWRI1631
. . . S. cerevisiae strain AWRI796
. . . S. cerevisiae strain BY4741
. . . S. cerevisiae strain BY4742
. . . S. cerevisiae strain CBS7960
. . . S. cerevisiae strain CLIB215
. . . S. cerevisiae strain CLIB324

S. cerevisiae WU-BLAST2 Search

All hits shown.

p=0. s=16755 ORF Uncharacterized [Sc strain BY4741] Protein of unknown function; pr



Split ORFs in BY4741, BY4742

YNR066C fused in RM11-1a, Sigma1278b, W303

Saccharomyces GENOME DATABASE

ALK2/YBL009W Phen

Summary | Locus History | Literature | Gene Ontology | Phenotype | Interactions | Ex

This page lists all curated single mutant phenotypes associated with ALK2. Click on a that term. Browse phenotype terms

12 Single Mutant Phenotype(s) for					
Experiment type	**Mutant Information**	**Strain background**	**Phenotype**	**Chemical**	
classical genetics	Description: overexpression	S288C	cell cycle progression in G2 phase: abnormal		Deta S8 fc Deta 2C D
classical genetics	Description: overexpression	W303	cell cycle progression in M phase: arrested		Deta accu DNA short

Strain	Curated phenotypes
CEN.PK	165
D273-10B	270
FL100	57
JK9-3d	104
RM11-1a	8
S288C	62145
SEY6210	355
SK1	888
Sigma1278b	1302
W303	2793
X2180-1A	285
Y55	22

S288C Reference Genome

Current and archived versions of data files containing nucleic acid and protein sequences of *Saccharomyces cerevisiae* strain S288C.

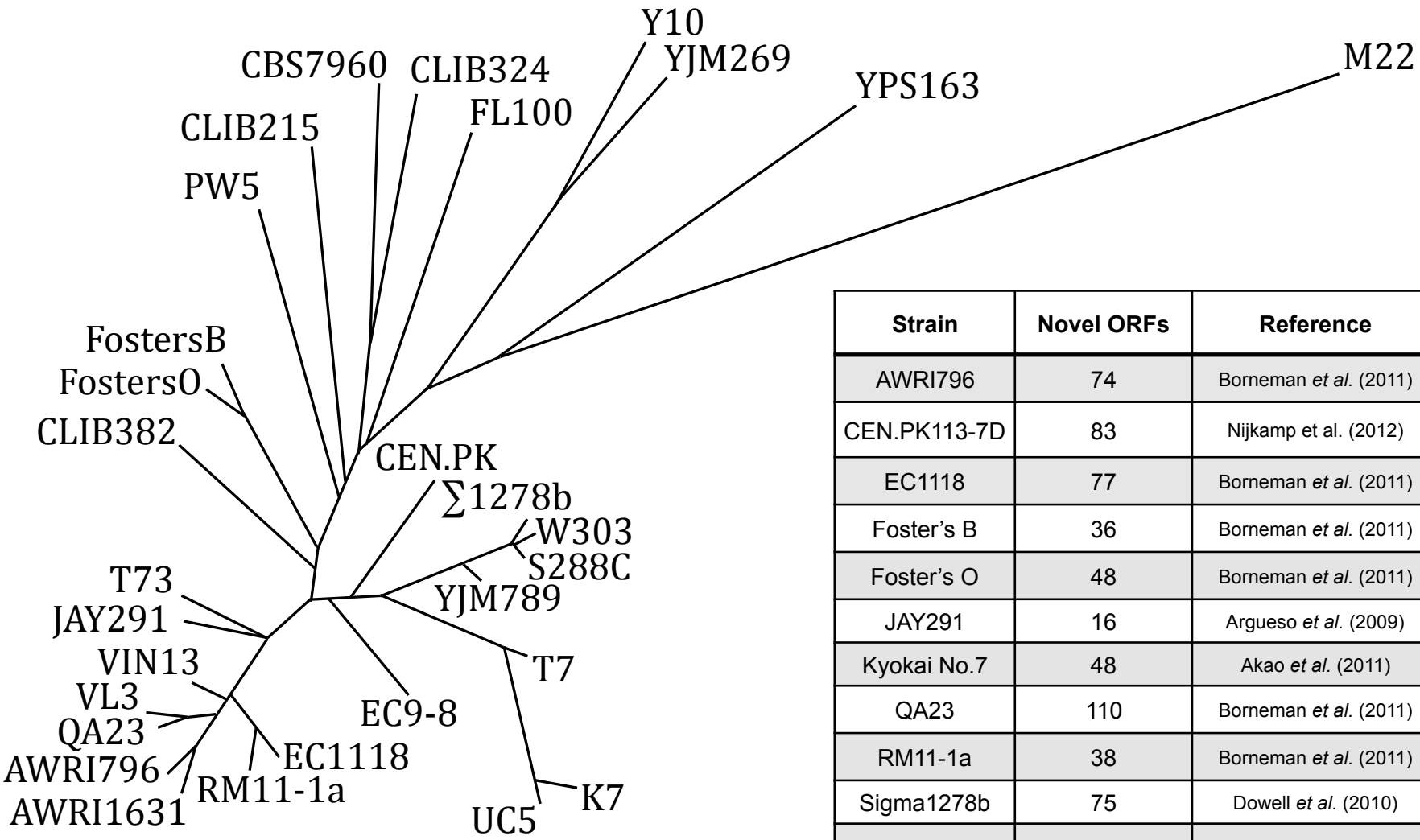
Name	Description	README
chromosomes/	FASTA sequence files of the 16 nuclear chromosomes and mitochondrial genome.	
genome_releases/	S288C reference genome sequence and annotation.	
intergenic/	Intergenic DNA sequences that are not contained within any annotated feature.	
NCBI_genome_source/	Sequence files used by NCBI to create the *S. cerevisiae* reference genome sequence.	
orf_dna/	ORF genomic and coding DNA sequences in FASTA format.	
orf_protein/	Protein translations of ORFs and pseudogenes in FASTA format.	
other_features/	Sequences of features other than ORFs in FASTA format.	
primer_sequences/	Primers used to amplify ORFs and intergenic regions of S288C.	
rna/	RNA gene sequences in FASTA format.	
dates_of_genome_releases.tab	Dates of various genome releases.	

Strains

Complete genomic DNA and protein sequences of *S. cerevisiae* strains other than S288C.

Name	Description	README
archive/	Archived sequences of *S. cerevisiae* strains other than S288C.	
AWRI1631/	Haploid derivative of South African commercial wine strain N96.	
AWRI796/	South African red wine strain.	
BY4741/	S288C-derivative laboratory strain.	
BY4742/	S288C-derivative laboratory strain.	
CBS7960/	Brazilian bioethanol factory isolate.	
CEN.PK/	Laboratory strain.	
CLIB215/	New Zealand bakery isolate.	
CLIB324/	Vietnamese bakery isolate.	
CLIB382/	Irish beer isolate.	

Saccharomyces GENOME DATABASE



Strain	Novel ORFs	Reference
AWRI796	74	Borneman *et al.* (2011)
CEN.PK113-7D	83	Nijkamp et al. (2012)
EC1118	77	Borneman *et al.* (2011)
Foster's B	36	Borneman *et al.* (2011)
Foster's O	48	Borneman *et al.* (2011)
JAY291	16	Argueso *et al.* (2009)
Kyokai No.7	48	Akao *et al.* (2011)
QA23	110	Borneman *et al.* (2011)
RM11-1a	38	Borneman *et al.* (2011)
Sigma1278b	75	Dowell *et al.* (2010)
VIN13	45	Borneman *et al.* (2011)
VL3	54	Borneman *et al.* (2011)
YJM789	34	Borneman *et al.* (2011)

Saccharomyces GENOME DATABASE

Pan-genome

- All genes found within a species

- Core – conserved, essential
  - Actin, histones, replication, translation, etc.
- Frequent – found in some genomes, not others
  - Adaptation to specific environments
- Rare – rapidly evolving, especially mutable
  - High rates of gene birth and death

Saccharomyces GENOME DATABASE

Expanding our scope:

- Annotation and analyses of major strains
  - history of use and experimental results
- "Shift the reference"
  - choose most appropriate genome

SGD Saccharomyces GENOME DATABASE

Awesome power of yeast genetics:

- Investigation of individual gene products
- Fertile environment for understanding biology
- Continued need for 'small science'

SGD Saccharomyces GENOME DATABASE

Our mission:

Educate students

Enable researchers

Facilitate scientific discovery