**Figure S1**

**a**, Plot of SNP and gap differences along chromosomes between versions of S288c from SGD (Saccharomyces Genome Database, http://www.yeastgenome.org, downloaded October 2nd 2007) and SGRP (Saccharomyces Genome Resequencing Project, this work). Chromosomes I and II appear to have a higher dissimilarity rate than the other chromosomes and there are cluster of dissimilarities present in subtelomeric regions and multicopy genes; **b**, Example of dissimilarity over *MCM7,* an essential gene on chromosome II, where comparative analysis supported the SGRP sequence; **c**, SGRP quality scores where all other strains agreed with SGD (left graph) and with SGRP (right graph).

**Figure S2**

Distribution of novel hypothetical ORFs (rows) among the sequenced strains (columns). Gray squares indicate the newly-sequenced strains in which the six novel proteins and 38 ORFS were identified.

**Figure S3**

Neighbour-joining (NJ) trees, based on SNP differences between all *S. paradoxus* strains sequenced, with geographic subgroups highlighted in different colours. The *S. cerevisiae* NJ tree is shown to the same scale. Variation in *S. cerevisiae* is comparable to that in a single *S. paradoxus* subpopulation and is substantially less than *S. paradoxus* as a whole.

**Figure S4**

Changing topology of the NJ trees along chromosome VIII. The clean lineages show the same NJ topology across the genome whereas mosaic strains exhibit different topologies for different segments. For example in strain UWOPS83.787.3 (green) the leftmost 80kb of chromosome VIII groups with the Wine/European cluster. In the interval from 240-320 kb the strain groups with the North American strains. For the same intervals the lab strain S288c (violet) is in a long branch followed by the Wine/European cluster.

**Figure S5**

**a**, DNA similarity plots comparing sequences from each clean lineage to a selection of other strains depicted in different colours. Each subplot shows the similarity along the length of a chromosome between the strain printed on the left and each of the strains listed at the bottom. Within each 10kbp block along the chromosome, similarity is defined as N/(D+1), where N is the number of positions in the block for which both strains have a nucleotide of quality 40 or more (maximum 10,000) and D is the number of those positions where the nucleotides differ. Thus, for example, the middle plot shows that DBVPG6765, represented by the yellow line, is more similar to L-1528 across the whole of the chromosome than are any of the other strains listed at the bottom of the diagram. Chromosome X is displayed and similar results were obtained for the rest of the genome. The top four panels are for strains with a mosaic genome, as evidenced by the fact that no single line is consistently highest for any of them. Strain YIIC17-E5 is close to the European/Wine cluster in the phylogenetic tree and most of the chromosomes show high similarity to this lineage (yellow). The bottom five panels are representatives of the five clean lineages showing high similarity to a single query genome right across the chromosome.

**b**, Similar to **a**, but for chromosome II, and with the W303 genome added (purple). This shows large blocks of homology between S288c and W303. Between 150 and 200 kbp W303, but not S288c, shares high homology with the West African lineage (black). Both SK1 and Y55 exhibit high similarity to the West African lineage up to ~600 kbp, when they switch to the sake (red) and Wine/European (yellow) lineages respectively. Y55 also contains a Wine/European segment at around 400kbp.

**Figure S6**

Ribosomal DNA polymorphism associated with genome mosaicism. Variable nucleotide positions (substitutions only) found in the 9.1 kb rDNA repeat are mapped for each *S. cerevisiae* strain. The proportion of sequencing reads showing base substitution at a given variable position is indicated by the colour of the bar. Fully resolved polymorphisms (i.e. SNPs) are excluded. Low frequency polymorphisms, indicated by red bars, predominate, with increasing within-strain variation particularly evident in the non-coding IGS1 and IGS2 regions. Strain names are given on the left, with structured genome strains shown in blue and mosaic strains in red. Ribosomal DNA copy number estimate (in green) and total number of polymorphic positions (in black) are shown for each strain on the right. Strains are sorted top to bottom by increasing number of polymorphic positions. Strains with mosaic genomes possess significantly greater numbers of polymorphisms. The average for mosaic strains was 98 polymorphic sites compared to the structured strain average of 34 sites ($p=1.3 \times 10^{-6}$ under Mann-Whitney U rank test). L-1528 was excluded from this analysis due to a high number of unaligned reads resulting in unreliable data.

**Figure S7**

Abundance of Ty transposable element sequences across *Saccharomyces* strains. The proportion of Ty sequences in ABI shotgun sequencing reads identified by RepeatMasker is shown for each strain of *S. cerevisae* (red) and *S. paradoxus* (blue). Population are defined as in Fig. 1 and abbreviated as follows: WE - Wine/European, CL - Clean Lineage, MO - Mosaic, UK - United Kingdom, RU - Russia, FE - Far East, DK - Denmark, HA - Hawaii, NA - North America, SA - South America. The 6 strains identified as potential clonemates in Table S6 are also excluded from this analysis. Comparison of overall TE abundance in shotgun reads (3.53%) from *S. cerevisiae* strain S288c with the reference genome sequence from the same strain (3.35%) reveals that estimates of overall TE content from reads are similar to finished genome assemblies. Our estimate of Ty abundance for the finished S288c assembly is slightly higher than previous estimates (3.1%)[31] since we used an expanded RepeatMasker library with newly reported *Saccharomyces* Ty variants[32].

**Figure S8**

**a**, Derived allele frequencies in single nucleotide polymorphisms (SNPs) in non-coding regions compared to those in synonymous sites (unfilled symbols). All categories of non-coding SNPs show an excess of low frequency alleles indicating the action of purifying selection, with SNPs in tRNAs (red bars) showing the strongest effect.

**b**, Distribution of McDonald-Kreitman (M-K) ratios for yeast genes. Dotted trace indicates the expectation in the absence of selection or changes in effective population size. Inset is the ration of amino acid changing to synonymous polymorphisms as a function of allele frequency. Because of the excess of amino acid polymorphism at low frequency, only SNPs with minor allele frequency >20% were used for M-K tests.

**Figure S9**

Phenotype variation among sequenced *S. cerevisiae* and *S. paradoxus* strains. Quantification of strain growth phenotypes in 67 environments was performed using high resolution micro-cultivation, automated measurements of population optical density (OD) and calculation of strain doubling times, lags and maximum densities. See Supplementary Materials and Methods for details. Strain (n=2) doubling time, lag and maximal density phenotypes in relation to 20 replicates of the haploid S288c derivative BY4741. Logarithmic strain coefficients, LSC=LN (strain/BY4741) are displayed. Green = poor growth, red = good growth. Hierarchical clustering of strain doubling time, lag and maximum density phenotypes was performed using a centred Pearson correlation metric and average linkage mapping. Blue lines = *S. paradoxus*, pink lines = *S. cerevisiae*, grey line = *S. bayanus* isolate CBS7001.

**Figure S10**

**a**, Phenotypes that distinguish *S. paradoxus* from *S. cerevisiae*. The growth (LSC, n=2) relative the reference strain (n=20) in each environment, averaged over all *S. cerevisiae* and all *S. paradoxus* isolates, is displayed. Error bars indicate standard error. Differences are significant (Student's t-test, Boole-Bonferroni correction; cycloheximide= $p=10^{-17}$, heat= $p=10^{-11}$, paramomycin= $p=10^{-12}$, $CuCl_2$= $p=10^{-10}$).

**b**, Phenotypes that distinguish the *S. cerevisiae* Wine/European & mosaic strains (left cluster group) from other *S. cerevisiae* strains (right cluster group). The growth (LSC, n=2) relative the reference strain (n=20) in each environment, averaged over all strains and all environments is displayed. Error bars indicate standard error. Differences are significant (Student's t-test, rate $p=10^{-55}$, lag $p=10^{-5}$).

## Supplementary Methods

### Strain genome assembly

Assembly was carried out using PALAS, for Parallel ALignment and ASsembly. All *S. cerevisiae* strains were assembled together in one run, and all *S. paradoxus* in another. The PALAS algorithm worked as follows.

The whole-genome sequence for each strain was initialized to the reference sequence for its species, generalized to a (low-complexity) graph by adding edges that allow transposons and one or more copies of tandem repeats to be omitted, and others that allow transposons or long terminal repeats (LTRs) to be inserted at places where an initial alignment of reads to the reference sequence suggests they may be present: i.e. where a read matches a piece of unique sequence and one end of a transposon or LTR. Successive iterations of PALAS then caused each sequence to diverge away from the reference in the direction of the correct sequence for that strain. Each iteration involved the following steps.

1. All still-unplaced Sanger reads for each strain were aligned to the sequence graph for that strain using SsahaSNP (http://portal.acm.org/citation.cfm?id=1099538.1099840&coll=&dl= )with the parameters

```
-diff 2000 -score 20 -seeds 2 -cut 50000
```

to ensure that as many non-trivial matches as possible for any part of each read were found. For this purpose, the sequence graph is represented as a main "backbone" sequence plus smaller sequences extending 1200bp (the maximum length of a read) in each direction from each choice point.

2. Matches for each read, or each read pair when applicable, were assigned to clusters. One match could belong to more than one cluster. Each cluster represented a consistent set, in the sense that all matches for the same read had to have the same orientation and be in the right order. In the case of a pair, the matches for the two reads had to be in opposite orientation and be separated by at most 20kbp, a generous estimate of the upper limit of the insert-size distribution plus an allowance for a large

deletion. Each cluster was assigned a score according to the scores of its component matches and the number of those matches. The highest-scoring clusters consisted of a single match for each of a pair of reads (indicating no large structural variations), consistently oriented, and with each match containing relatively few nucleotide or short-indel differences.

3. Some clusters were discarded on various grounds, including mainly: low absolute score; too much inserted material relative to the matched material, or too many individual insertions; implying a (large) indel at a point across which a read for the same strain had been placed in an earlier iteration, or across which a surviving cluster for another read matches without a break.

4. When the best cluster exceeded that of the next-best cluster (if any) by at least a given threshold, the matches in it were passed forward to the next stage. Otherwise, no matches for that read (pair) were passed forward. However, when two clusters consisted of identical matches to different parts of the genome, both were accepted; in practice, this only occurred for matches to the rDNA region, which is represented as two identical copies in the reference genome.

5. Surviving matches were then pruned. Where two matches overlapped and either they were for the same strain, or one (or both) originated in a cluster involving a large indel, the lower-scoring one was rejected. This cautious approach increased the number of PALAS iterations required, but had the advantage of making sequence changes implied by one read available for matching other reads in subsequent iterations, thereby ensuring consistency between overlapping accepted matches. All matches not rejected on these grounds were firmly accepted and incorporated into an overall whole-genome, multi-strain, multi-read alignment, containing the matches accepted in the current and all previous iterations, and quality scores as well as nucleotides.

6. At each position in this alignment, a consensus nucleotide value (or gap symbol) and quality score was inferred for each strain that had one or more reads matching at that point. The consensus value was the one whose sum of quality scores was highest. The initial quality score $q_1$ was this sum, minus the sum for the next-best value if any. $q_1$ was then adjusted to $q_2$ using the formula $q_2 = K\ (1\text{-}e^{-0.029q_1})$, where K was 38.5 if any match specified a gap symbol and 55 otherwise. This formula was determined from statistics on the frequency of mismatches resulting by considering only forward-matching reads

and only reverse-matching reads, as a function of the $q_1$ values from each of those sources. Its main effect is to flatten off $q_2$ values so that they never exceed 55.

7. Because of the low coverage and presence of sequencing errors, the sequence for each strain is neither as complete nor as correct as it could be. This is remedied using an imputation process, in which a sample of ancestral recombination graphs (ARGs) are calculated using the Margarita system[33] (in order to give a sample of possible trees relating the strains at each point in the genome. Felsenstein's algorithm[34] is then applied to infer a value at every strain. The effect of this is to correct sequencing errors that have low quality scores and to fill in missing values.

8. A sequence graph for each strain is then read off from the alignment, preserving the uncertainty that arises when strain S1 has no or very little material placed in a region R; strain S2 has a deletion across R; and strain S3 has some sequence in it. When this happens, the sequence graph for S1 will allow region R to be either included or skipped in subsequent matching.

9. A decision is then made on whether to continue iterating or to finish. Another iteration is started if, on average, more than one genome position in 10,000 has been newly filled in for each strain on the current iteration.

10. Otherwise, a final iteration is run, which differs from the earlier ones in four respects. In step 4, only clusters for paired reads involving one comprehensive match for each read in the pair are created. In step 5, overlapping reads for the same strain are allowed. In step 7, 25 ARGs are sampled, rather than 3 as in earlier iterations, to increase the accuracy of imputation. In step 8, a firm decision is made on which subsequences to include, so as to yield a sequence rather than a graph.

11. Once a sequence has been generated for each strain, Illumina (Solexa) reads are aligned using Maq[35]. Alignment is to the PALAS-derived (imputed) sequence for the strain if one was created, or otherwise to the species reference sequence. A final consensus sequence per strain is then derived by combining Sanger and Illumina nucleotide values, taking the highest-scoring value when there is a clash.

### *S. paradoxus* assembly

First, we built an *S. paradoxus* reference sequence using the Phusion assembler[36] from reads from the following sources.

1) Ten strains collected in London, totalling about 11x coverage. We tried including further European strains, but found that doing so produced less good results, presumably because the greater divergence confused the assembly process more than the extra data helped it.

2) All reads from the original *S. paradoxus* sequencing project[37], again totalling about 11x coverage.

3) Artificial reads constructed by shredding the 832 contigs created by Kellis *et al*[37] into paired 1,000 bp sequences, 5kbp insert size, with reads overlapping by 500bp, giving 2x coverage.

Running Phusion on these reads created 608 contigs organized into 471 supercontigs, for a total length of 12,265,206 bp. The N50 contig size was 172,493, compared to 11,872,617 bp total and N50 49,124 for the Broad assembly.

Next, we aligned the contigs against the Sc reference using Ssaha2 with default settings. The best alignments for the contigs were in good agreement with the supercontigs that Phusion had independently created: we were able to place 52 supercontigs (roughly the largest 52, covering 11,639,177 bp, or 94.9% of our assembled sequence). All but two of these placements were completely syntenic, once allowance had been made for subtelomeric rearrangements. We broke the other two supercontigs: one at a point between two contigs, and the other in the middle of a contig which appeared to be chimeric owing to sequence similarity between two regions of approximately 2,500 bp located at roughly 1,243,250 to 1,245,753 on ChrIV and 575,099 to 577,600 on ChrXV.

We then formed most of the Sp assembly by juxtaposing all the placed contigs, separating contigs within the same supercontig with 50 ``N" symbols and separating adjacent supercontigs by 100 ``N" symbols. However, coverage of the Sc sequence by the Sp contigs was poor in the mitochondrion (only about 20\%, all short contigs) and the rDNA region on chromosome XII because of its multiple repeats. We therefore did not attempt to create a mitochondrial sequence. We carried out a separate Phusion run on the reads aligning to the Sc rDNA region; these assembled into a sequence of 8,393 bp for the each of the rDNA sequences proper (cf 8,375 for Sc) separated by spacers of 721 bp (cf 762 for Sc), and ending with a partial copy of the region containing only the first 1,200 bp.

Second we performed deep 80X coverage using Illumina GA on our version of CBS432 to create an independent assembly. Essentially, we obtained the same genome architecture albeit with strain specific SNPs and indels. This complete reference did not result in any substantial changes in the analysis.

**Multiple-species whole genome alignments**

We created an initial whole-genome alignment between *S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzevii* and *S. bayanus*, using the SGD reference sequence for *S. cerevisiae*, our *S. paradoxus* assembly (see above), and assemblies for the other three species downloaded from

```
    ftp://genome-ftp.stanford.edu/pub/yeast/sequence/fungal_genomes
```

We used the MIT assembly for *S. mikatae* and the Washington University assemblies of the last two species. We aligned each species to *S. cerevisiae*, using ssaha2[38] in each case except for S paradoxus, where we used BLASTN (with parameters `W=9 M=1 N=-1 Q=3 R=2 wordmask=seg`) because ssaha2 was not able to handle chromosome-length query sequences. We found that where we were able to use ssaha2, it gave better results (fewer separate matches, encompassing more sequence) than Blastn. We then formed the multi-species alignment by placing the two-species alignments beside each other, using the *S. cerevisiae* sequence as backbone.

**Inference of ancestral states and allele frequencies**

To infer ancestral states for SNPs in non-coding regions the whole genome alignments described above were used. Where the outgroup matched one of the segregating alleles, this was inferred to be ancestral. Where the outgroup matched neither of the segregating alleles, no ancestral state was inferred.

To infer ancestral states in coding regions, alignments based on proteins rather than DNA are expected to be more accurate.  Therefore, we obtained the one-to-one orthologue assignments of protein coding sequences (CDS) from SGD, translated these and aligned them using T-coffee[39]. Genes with introns were excluded from these analyses. The gaps in these protein alignments were then re-introduced into the sequences of the CDSs. These alignments were used to infer the ancestral states of SNPs, as well as to calculate the number of fixed differences in coding regions for analysis of synonymous and non-synonymous sites. Codons in which there were more than one SNP or substitution, or where the outgroup matched neither segregating allele were not considered.

Insertion and deletion mutations were identified by searching for the largest possible continuous region of gap characters in the alignments, where at least one other strain had sequence observed. Indels and stop codons in essential genes were identified using essential genes as indicated in http://chemogenomics.stanford.edu/supplements/01yfh/files/orfgenedata.txt[40].To compute allele frequencies, residues with quality less than 40 were excluded, as were positions with more than two alleles. The observed counts of each residue were divided by the total number of residues at each position. For indels, data from strains were excluded if any of the two residues flanking the indel on either side had quality less than 40, or if there was another indels within 15 basepairs. Such indels were considered likely to have alignment difficulties, particularly in the case of homopolymeric repeats, where placement of the indel is often arbitrary. Only markers with exactly two alleles and at least eight total observations were included in the analyses. For minor allele frequency analysis, aligning sequence in the outgroup was required, but it was not required to match either the reference or variant codon in the strains. All this analysis was performed on the observed sequence data, rather than imputed data.

**Linkage Disequilibrium**

Calculations of θ, ρ and linkage disequilibrium based on raw (not imputed) data with q>40 with gaps treated as missing data. $\theta_\pi$, $\theta_S$ and Tajima's D were calculated using Variscan (version 2.0.1; option numnuc = 4). Linkage disequilibrium was measured by $r^2$ and calculated using custom Perl scripts.

**Analysis of CNVs.**

A gene coverage table (available at ftp site `ftp://ftp.sanger.ac.uk/pub/dmc/yeast/latest`) was produced with the average aligned depth for each gene across each strain. This number was then divided by the mean coverage of the strain, in order to get an approximation of the copy number for each gene in each strain, giving a total of 36 x 6577 estimates altogether. These estimates were rounded to the nearest integer. For estimates up to $n=4$ the distribution fits a Poisson with mean 1. For $n\geq5$, the observed distribution of estimates is much higher than the Poisson would predict. These discrepancies are evidence of copy number >1, with an estimated 4120 (out of 36*6577) gene-strain combinations showing evidence of five or more copies. After Bonferroni correction only estimates of $n\geq10$ remain significant. Excluding rDNA genes, known to be high-copy, nearly all of the remaining genes with increased copy numbers are from three strains. The counts are:

226 genes in YJM981

202 genes in NCYC361

27 genes in YS2

4 genes in DBVPG1373

1 gene in 273614N

1 gene in BC187

1 gene in UWOPS05_217_3

The karyotypes show that YJM981 and possibly also NCYC361 have an unusual chromosome structure (results not shown). The high-copy genes in these strains do not seem to cluster in specific chromosomes.

The seven (4+1+1+1) CNVs in the strains that have four or fewer CNVs are:

| | |
|---|---|
| YBL071C-B | UWOPS05.217.3 |
| YFR012W | 273614N |
| YGR201C | BC187 |

| | |
|---|---|
| YHR052W-A | DBVPG1373 |
| CUP1-1 | DBVPG1373 |
| YHR054W-A | DBVPG1373 |
| CUP1-2 | DBVPG1373 |

Apart from *CUP1*, the other genes are all either uncharacterised or dubious. The *CUP1* story we have presented may be the only CNV example that we can detect in this way. There is clearly something interesting going on with the three strains with high counts of genes with increased copy number and this will merit further investigation. Because of the large number of gene-strain combinations and the consequent need for a Bonferroni correction, we cannot use read coverage alone to spot copy numbers >1 but <10. For more sensitivity, we would need to look at assembly clashes (multiple apparent within-strain SNPs and/or structural inconsistencies) in alignments from the same strain to the same gene, which is beyond the scope of this paper.

**Analysis of novel genes**

One of the advantages of direct re-sequencing over the microarray approach is the detection of novel genes not present in a reference genome. Putative novel genes in the newly sequenced *S. cerevisiae* isolates were identified from the sequence reads. Here we describe the approach used to identify potential new ORFs.

1. To start our search we have used the reads that had not been aligned by PALAS to the reference genome sequence S288c. These reads have at least 400 bp of good sequence and are therefore expected to align.

2. The sequences of the unaligned reads were compared to the published genome sequence using BLAST[41] and matches discarded. These sequences included the coding sequences of genes with multiple copies in the newly sequenced isolates but only one copy in the published sequence.

3. We predicted translations in all six frames in the 2506 reads for which no match was found in the reference genome and found 1118 open reading frames (ORFs) of 100 amino acids or more in 923 different reads.

4. Of these 923 reads 278 found a match only in the same strain when compared to all 923 sequences using BLAST search. Analyses of these reads suggested that they are likely to be contamination during the DNA extraction or sequencing process.

5. Among the remaining predicted peptides, we determined that 38 were likely to be real genes because they contained matches to protein domains (Pfam) or matched other sequences in the NCBI database.

6. In order to identify which isolates and lineages encode these putative novel genes, the six previously known proteins and the 38 newly identified ORFs were compared to the complete set of SGRP reads using BLAST. A match was accepted (Figure S2) if the ends of the match were determined either by the ends of the ORF or by the ends of the read and at least 70% of the residues were identical. Two previous non-reference proteins and five hypothetical novel ORFs were identified in our version of S288c.The seven proteins found in our version of S288c (BIO6, RTM1, SGRP_9, SGRP_10, SGRP_16, SGRP_24 and SGRP_28) were searched for in the published genome. BIO6, RTM1, SGRP_9, SGRP_10, SGRP_24 and SGRP_28 have matches with the published reference S288c genome, however they are quite divergent and the matched sequence seems to be the same as the matched S288c reads. The level of divergence resulted in rejection by PALAS. SGRP_16 finds no match in published genome, but a good match in our S288c (97% id., 99% positives) with first 92 residues of the ORF.

---

**Identifier:** SGRP_Hypothetical protein_1

**Sequencing reads:** SK1-7c04.p1k and SK1-10i11.p1k

**Predicted protein sequences:**

```
MEPISSGSSLQSNMRETVQSQNAEILPQMSNENKNKPSCEHFESVGEYIMVGGRLFKKSDTNTYLEDLG
PGTPVEQPGQVGFANPLPLGLASFSFMCLTLGLVNARVRGVTNLYLLNASFIFGGAVVLLSGLLSFCVG
DTFCMTVFGSFGGFWISWGCLNLEQFGVTKAYADDPQALQNVLGFYLAGWTVFNFLVMVCSMKSTWGIF
LLLLFLDLTFLMLCIGSFTQSTNVSMAGGYFGILSSCCGWYSLYCSIANKDSSYVPLVAYPMPGSQIV
```

**Top BLAST hit:** hypothetical protein SCRG_02225, *S. cerevisiae* RM11-1a

**Putative functional domain:** pfam01184, Grp1_Fun34_YaaH, GPR1/FUN34/yaaH family.

---

**Identifier:** SGRP_Hypothetical protein_2

**Sequencing reads:** UWOPS05_227_2-15k24.p1k

**Predicted protein sequences:**

```
MFAIITPVALTPAILVMAYLEHQANKTGEIPVGSDPLAKKKVEVTESHISGFKQYLELLKASLIEIDAF
GLILLGFAFSLILLPCSLYSYAEGGWNNPSMIAMEVVGGIFLITYVVFEVFFAPFPLLPKRVLMNRTFI
CCVIIDFIYQMAGYFSLLFFTSYTFVVLNLSYRDWVYLSNTTTMGLCFFGVVWGALFRCFHRYKIFQVV
GIAIKLIGMGLYVACSKQGW
```

**Top BLAST hit:** GENE ID: 2894316, KLLA0E24849g, hypothetical protein, *Kluyveromyces lactis*

**Putative functional domain:** Fungal trichothecene efflux pump (TRI12). This family consists of several fungal specific trichothecene efflux pump proteins. Many of the genes involved in trichothecene toxin biosynthesis in Fusarium sporotrichioides are present within a gene cluster.It has been suggested that TRI12 may play a role in F. sporotrichioides self-protection against trichothecenes.

---

**Identifier:** SGRP_Hypothetical protein_3

**Sequencing reads:** UWOPS05_217_3-3b07.p1k

**Predicted protein sequences:**

```
MGLNILLEAVKARAGIDVYHISLFCGVASLFDRITCSENGAEKQYNHPGRIVLMEITCIFIFVRVCCLV
YRQYKKVSKEELMAILTDFDHTTARFCNETLDIRSDLFQHIIRDKNKSDYHRDIIHGIEKVLGREIITT
IESCEREVLSSDEYRQAQYMGNVATKDLDYLRYYLNLDIFPELNTDDELWDDLKEIDKYYCSSV
```

**Top BLAST hit:** hypothetical protein SCY_1426, *S. cerevisiae* YJM789

**Putative functional domain:** Not detected

---

**Identifier:** SGRP_Hypothetical protein_4

**Sequencing reads:** L_1374-12c04.q1k, L_1374-15n06.q1k

**Predicted protein sequences:**

```
MSDKKSKKACEVCKRRKKRCSGGRPCDYCIKIDKQLACTYRVKVSSKTVKVTEKYLVNLKSKIKDLELQ
LATRSNCHPNDVSTNDNPLVSSEDDEEDRDGMDDPSEGNNYYRLGNSACGKFLLRIKDSLGKSCQLRGD
VRPSVIETISLETSPNMALIEQIVRENCPSPSEAKNWILAASNVIGADYMYIEPDYEKSVLDELIWTSD
SHNADFVKYATEVTRFFTYLALGCLFNKDRSPEK
```

**Top BLAST hit:** KLLA0C18953p, *Kluyveromyces lactis*

**Putative functional domain:** Not detected

---

**Identifier:** SGRP_Hypothetical protein_5

**Sequencing reads:** DBVPG6044-29h08.q1k, DBVPG6044-24h21.p1k

**Predicted protein sequences:**

```
MSAYLDNSINAANYQKNRITYPKSLYETVLQHHLGERNLAVDVGCGTGIGSFPLLDYFEKVVGCDPSEK
MLQTARMIADTIPESSKRNVEFKETGGETLGKYFKEDSVDLIIAGESLQYTKFEQFFEQAHKILKPNGT
LAYWFYCDPIFIDYPKANEIFKYFVYEDERFFKAFWPPEMDYVRHLGSTIEIPKNSFTDVYSEKYIPLK
SEKAGKFLISEMISL
```

**Top BLAST hit:** hypothetical protein CaO19.2468, *Candida albicans*.

**Putative functional domain:** pfam08241, Methyltransferase domain. Members of this family are SAM dependent methyltransferases.

**Notes:** This hypothetical protein is specific to the West African lineage.

---

**Identifier:** SGRP_Hypothetical protein_6

**Sequencing reads:** NCYC110-12f22.p1k

**Predicted protein sequences:**

```
MLLHYFLLVASFMKFTIGNQTIWLDDLEHNLANGAPISIYDDIPDLNKTAMEIYSENTLDINWNGMNDF
LDHLSETDNATLAKREDLLGVIINAIDPSKATNSDKLAKREESCEQGYTRSLYSRLTNWRTIRSLSSNI
REFYYTYATAMIDNASGLASLGYSVAVDLKNRSNKQSCNGGSDWFDVYNKDGSKHTYLVAIAPWTTGKN
CDTTATSGMLKEMTTWILDKAQEEHLSAWCSRYDNGGSWHADVRS
```

**Top BLAST hit:** KLLA0D00660g, hypothetical protein *Kluyveromyces lactis*

**Putative functional domain:**

**Note:** This hypothetical protein is specific to the West African lineage and is also present in *S. mikatae* and *S. bayanus*

---

**Identifier:** SGRP_Hypothetical protein_7

**Sequencing reads:** UWOPS03_461_4-5k23.p1k

**Predicted protein sequences:**

```
MLRPLWCLLSCTVTKEQPLESYCIATKDLYGIDGVKAGESSRAYYNIYDGANATCPSVQRLIDMSAVIV
GTLKLTHFENRETPTADYVDYHTPFNLRGDGYQSHFSSSCASGATEAAYD
```

**Top BLAST hit:** *S. cerevisiae* (unspecified strain)

**Putative functional domain:** PRK05962, Amidase superfamily

**Note:** ORF previously cloned nearby MEL1

---

**Identifier:** SGRP_Hypothetical protein_8

**Sequencing reads:** DBVPG6044-18g13.q1k, DBVPG6044-26f04.q1k, DBVPG6044-36o12.q1k

**Predicted protein sequences:**

```
DPLDFQGLSIISCKTCRQRKIKCGRQFPKCQNCIKRSCECTYPRTFRKTSTKLTRKRRDVNARFYGFSS
VNRSLFEVGMPFSNVDFELEGENAANQMKRFSSSPVFKKYIGDTKLILAAIQSVRSSISCSFFDETVDL
NLLEQKIFSKQGADYQTLLLSYAVIIVSERFYETPPDVREVVSELDILLNECSDCSEKVSSLILLSEYY
HYNFKIETAWKCIFLAASIGYALGLHTTSSKVWTMLVLQDSLLCSVLGRPTSISCVNSKLVSDQCDGWG
EIAILLREGNDMLLNLKSETCVEKAISLDFKIDDVIERTKKNMSSSEKSDSSVNLLVGYLKVCILSASR
IKLLFPFFTKHRSIKAQLDENCSSLAGCLCGLFQLLNASNLTSGDKKFPLRPHFFPAYCSVFQGFLLQF
LYTSNELFKNFDETTNGANSIFLPKDLGRTGLFLPSLDVTSVLMEDYDLITGKVKFCSFMTDLFASFRS
LLNQKKS
```

**Top BLAST hit:**

**Putative functional domain:** cd00067, GAL4, GAL4-like Zn2Cys6 binuclear cluster DNA-binding domain, found in transcription regulators like GAL4. Domain consists of two helices organized around a Zn(2)Cys(6 )motif; Binds to sequences containing 2 DNA half sites comprised of 3-5 C/G combinations.

**Note:** well conserved in *Saccharomyces sensu stricto* complex

**Identifier:** SGRP_Hypothetical protein_9

**Sequencing reads:** DBVPG6044-22p13.q1k

**Predicted protein sequences:**

```
DPLGAVDKNKYSGQLYTLPMLQAFNTLDSLGPGMFVTAQSVAISDGFSGNKTVSEIQFPVLFDSGTTYS
TLPTEIADSIGKIFDGKYSSDDQGYIADCSKMNNTLLSIDFGGFDISANISNFVTRTKDHCLLNIEATD
SGFVLGDAFLVDAYVVYDLEDYEVSIAQASFNSQGEDIEIISNSIPGAIPAPGYSSTWVYTPDSPIGTG
DFLNMSWTSYSDYSEYQTLLSTATASSSSSSSSSDQTTTXKRNSGDRIQQSFFSFSLIPLLSYILL
```

**Top BLAST hit:** Aspartic proteinase yapsin-6 precursor *S. cerevisiae* RM11-1a

**Putative functional domain:** pfam00026, Eukaryotic aspartyl protease.

**Identifier:** SGRP_Hypothetical protein_10

**Sequencing reads:** 273614N-42n13.q1k

**Predicted protein sequences:**

```
SIFHQWTSVIACFFSWVHTVCFLYQAFREGGTDGMQYQWKSQLIWRTGVPPLLFLTLLWLFSLLFIRKY
IYELFLQFHWILAIGFYVSLFYHVYPELNTHMYLVGTIVIWFAQLLYRLVSKGYLRPGKTFMTSSIATI
TLKGIGCIELIVKDIDMDYSPGQHILLRTIDKDVVENHPFSIFPSSHSPGALKILMRAQNGFTKSLYLS
QCTAKRILVDGPYGGIERDVRSFTNLYLICSGSGISSCLPFLIRYGPLLQETNLRFIRLEWIIRYEEDI
SWVSDELRYLTTILKRSFLEGRIVIKIYICSSGDP
```

**Top BLAST hit:** EDN63726.1, *S. cerevisiae* YJM789

**Putative functional domain:** pfam08030, NAD_binding_6, Ferric reductase NAD binding domain.

**Identifier:** SGRP_Hypothetical protein_11

**Sequencing reads:** 273614N-41n14.q1k

**Predicted protein sequences:**

```
PEHMVFAWGGATIFCEVMAEMKRPMDFWKGMLCAQSLILVVYLFYGLFVYAYNGQFSYVTANMAIGSIG
LQNAGNVLTIITGIIAMVLYGNIGIKVIYQGFLVTDFNFPSLTSRKGTFAWAGFVVVYWAIAYILGTAI
PSISALVAIVGAFCILNFSYTFPFLFGFCLLCRQDAALADNFDAKTLTVEKADSYRSWSRWKRALGYGG
TYRILIKVSLFLLFLASLATCGLCSYSAISGAIAVYQTNPAQPFTCTSPVA
```

**Top BLAST hit:** GENE ID: 2900766, DEHA0C00484g, hypothetical protein*, Debaryomyces hansenii*

**Putative functional domain:** pfam01490, Transmembrane amino acid transporter protein.

---

**Identifier:** SGRP_Hypothetical protein_12

**Sequencing reads:** 322134S-11b12.p1k

**Predicted protein sequences:**

```
DPIVTFGGAGGQHAVAVAESLGINEILAHRYSSILSAYGIFLADVVEEKQEPCFLNLNDPDDAKSARKR
LDQLVKTCSESLIIQGFSETQILHEKYLNLRYEGTETSLMILEQNENWEFEKWFAEAHKREFGFAFSEK
CVIVDDVRVRATAKSCVRDEEPVDEQLKRYKPRSVFAAKEASFFKNVYFDNGWLKTPVFKIDDMTYGSV
VKGPAILADGTQTNIIPENSEAIVLKSHIFVKILRKSEENVSDEQKVPVDPVMLXIFSHRFMDIAXAMG
TXLKKTSVSTXCEGRXDFSCALFXPRWXXVANAPHVPVHFGSMSTCIAA
```

**Top BLAST hit:** hypothetical protein SCRG_03375 *S. cerevisiae* RM11-1a

**Putative functional domain:** pfam02538, Hydantoinase B/oxoprolinase. This family includes N-methylhydaintoinase B which converts hydantoin to N-carbamyl-amino acids, and 5-oxoprolinase EC:3.5.2.9 which catalyses the formation of L-glutamate from 5-oxo-L-proline. These enzymes are part of the oxoprolinase family and are related to pfam01968.

---

**Identifier:** SGRP_Hypothetical protein_13

**Sequencing reads:** UWOPS05_217_3-1n08.q1k

**Predicted protein sequences:**

```
MMPASFARVYLRENADRCIDTITEGSKNNISQERVNHKRISLYERNGCKLSFDSSTKDLVNVPLFPWNT
EDMLVIMGSTTFMDYSLASHIVIEYGLYLRAIADLRIS
```

**Top BLAST hit:** Low similarity with *S. cerevisiae RDS1*, Zinc cluster protein involved in conferring resistance to cycloheximide

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_14

**Sequencing reads:** UWOPS05_227_2-15g20.q1k

**Predicted protein sequences:**

```
MFTERSXVXARNDAGEKDLVSVNDGLXQVPTSISSCNGEKIIKTRGVTRIEVVRERMSTKVTWILGLSI
FLTSWVAALDATTTYNYQPYATSSFNRHSMLSTLTIANSVIGAVCKPFIAKISDLSSRPVTYFVVLVLY
VIGFVITACSPTIAAYVIGSVFIAIGQSGISLMNMVIIADTTTLKWRSFFTSLLSVPYLVTTWISGYIV
EDIINSNWRWGYGMFAIITPVALTPAILVMAYLEHQANKTGEIPVGSDPLAKKKVEVTESHISGFKQYL
ELLKASGDP
```

**Top BLAST hit:** GENE ID: 2894316 KLLA0E24849g, hypothetical protein *Kluyveromyces lactis*

**Putative functional domain:** pfam00083, Sugar (and other) transporter

---

**Identifier:** SGRP_Hypothetical protein_15

**Sequencing reads:** UWOPS83_787_3-1g13.p1k

**Predicted protein sequences:**

```
FVWNCTYAKIVRGIRCVGIDTSKYFEKVQLLIPVDVLFMKNFIPIVLAFALNNACLFSCLYIIDQYSQY
AERDSPLLAGVKLVPLIICMVIGNALCAFESTKLKPRVGVALGFFLALAGSVILIQLHLVKEDVFWKIF
FSSQALVGFGVAIFYPYALQIAVGGAPDQSKGIASGVAQTFGQLGIEITFSVMASVLGNINEMKGRTDA
VQKFRTGFQNCSYFTVAVGALGFLVTAICIRDIHPPNDDNSDLESSIHRTKIEIDQEKSGSEGEA
```

**Top BLAST hit:** GENE ID: 2892277, KLLA0C18931g, hypothetical protein, *Kluyveromyces lactis*.

**Putative functional domain:** pfam07690, Major Facilitator Superfamily

---

**Identifier:** SGRP_Hypothetical protein_16

**Sequencing reads:** SK1-5f12.p1k

**Predicted protein sequences:**

```
MGVEEFSIHVSENELEDLKRRLSSARIPKNVERKNWNFGTNAEYLAEVINYWKNSYDWRNIEQKLNGFH
HFQTTISNIRIHYIHEKGKSANSIPIILTHGWPDSFLRYTKLIPLLTDPEKFGVSSGISFDVVIPSLPG
FGFSDYPAGGSINNDTISDIWLELMKNKLGYDRFLAAGGDIGSGVTRYLGFKYPQNLIGIHLTDVGIIR
DLLNQSQLQSFSSEEQEYCKIASDWLDKEAGYMKIQSTKPQTLAFGLTDSPVGLAAWILESFIPGETYR
QFTPR
```

**Top BLAST hit:** *gb|ABE98164.1*, epoxide hydrolase *S. paradoxus*

**Putative functional domain:** pfam06441, EHN, Epoxide hydrolase N terminus

---

**Identifier:** SGRP_Hypothetical protein_17

**Sequencing reads:** SK1-31g11.p1k

**Predicted protein sequences:**

GMGLYVACSKRDGSPGIGLVVAALVVTNFGDAANVMGTQVAAQAAVPHQDMAATISVLSLYSSIGAAIG
TAITSAVWTDKLPGALSKYVPDKEKAAAFFESLTSIWEEPWGSVNREGAINAYQKVNYTLFCMGLGVSS
IMFIVALFQTNYYLGDQQNCVEGEQKEDYHHNANGSKKTLLNRAFDFWK

**Top BLAST hit:** GENE ID: 2894305, KLLA0E24871g, hypothetical protein, *Kluyveromyces lactis*

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_18

**Sequencing reads:** DBVPG6765-30e14.q1k

**Predicted protein sequences:**

EDPLNVDYMDLDVPDNNISNEDGFDPKILIANTKLALHIGSVMKKVYVTSGDNDFITNIVDSLKQLELF
RDSLDPELRVLPDIIESNRSIAVLTLRYHQIIIVTGRPLYLSLLKSTIRVTDELQDAKVKCVLAAVTNI
CILNNLWNSGWFCIFGFLEAQCCFSSILMLVMETLNGDAFPELQIALSLNASMCKAGNITALDNYSRLK
ELDSILFEAEKERQNREESSMKSLQSNRTTITEDGENSLVENESLHAKKISTLQDNAPLSEELRQDELN
ANLPSLIELKSSGAGFELFSPETFRNLSKKLERWDASLDLPTGNSL

**Top BLAST hit:** GENE ID: 4840364 FST10, putative zinc-finger protein*, Pichia stipitis*

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_19

**Sequencing reads:** YIIc17_E5-3o17.q1k

**Predicted protein sequences:**

HPFTGYLGSLMISKNVEKNFGLIFSPSIMFVMSAAFLIRILNNAHKYKLQSRHGVQENNIFFLSWGEFS
AINSHHSPSHPNQEPELFFEPKTAANKTRTLLEALLGVAQFSIHAIRCLRINSPICSFHLIGAFISCAL
WLPILIVIFCRITVYNQSMRWLSFNLFDLWVVCIVCYTILFSISIVAFRSVSVGHVDDEAEAFYIKWQF
FTNLTLFFLTFTGKIEKSGSTTKTSA

**Top BLAST hit:** gb|EDO14338.1, hypothetical protein Kpol_175p1, Vanderwaltozyma polyspora

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_20

**Sequencing reads:** SK1-62i15.p1k

**Predicted protein sequences:**

```
MAFNKVTLLLKWFGDIRNKCQKSRYQLNVNVGLIFLGCTFDLLNVASMISLIDDLAKTYNISYTTASWS
LTSYAVTFAGFIACMGRLGDIVGNSVLFTISCSLFAILSLLCAVMPNFPAFAVFRAIQGICAAGLVPCA
YALIPILAPKEKVQTYFSIVSCGFSSTIGLGLIIGGAFAATKIGYRGIFYLTFAVMSVISLVALFFHL
```

**Top BLAST hit:** GENE ID: 2892277, KLLA0C18931g, hypothetical protein*, Kluyveromyces lactis*

**Putative functional domain:** COG2814, Arabinose efflux permease. Carbohydrate transport and metabolism.

---

**Identifier:** SGRP_Hypothetical protein_21

**Sequencing reads:** 322134S-20m09.q1k

**Predicted protein sequences:**

```
MKGKIASVTGASGGIGYEIAIGFAQAGADVAMWYNTNPSIEKEVEGLSMKYGVKVKAYQCALTDGAKVA
QTIKKIENDFGKIDIMIANAGIAWASGPLVDFAEXDSESCDAEWMKIMDVDVNSVYYVSKSIGSIFQKQ
GYGSLVITASMSGHIANIPQYKLLITQLGQL
```

**Top BLAST hit:** GENE ID: 2906628, YALI0B16192g, hypothetical protein*, Yarrowia lipolytica*

**Putative functional domain:** PRK12825, fabG, 3-ketoacyl-(acyl-carrier-protein) reductase

---

**Identifier:** SGRP_Hypothetical protein_22

**Sequencing reads:** DBVPG6040-20m04.q1k

**Predicted protein sequences:**

```
GSLRALYSKVLRQEGVEAAARVNADFDRQTVRIDGRLVSFEYVREMKRDLARDFQVQMGLLGEWLEFGQ
VADVYTTVMANKGVKVVDNISGRGSVFNVLDASSVAGYCMRFRNQEEAGEEHKRCRILKVISSITKLLM
LSVWFNPGLPLRFPELSILSFGGSQRNLYFDAGDRVFIIRSRYNKNTKYDTRLLFLDAGVSAQLFWLIY
VLWPF
```

**Top BLAST hit:** hypothetical protein SCY_5432*, S. cerevisiae* YJM789

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_23

**Sequencing reads:** BC187-29g19.q1k

**Predicted protein sequences:**

```
EDPHYEKSVLDELIWTSDSHNADFVKYATEVTRFFTYLALGCLFNKDRSPEKTGSKFPGLQYFETALRL
QSELLKVYDMMANTSLVQSFLYVAYYALSLDKPEFAYLTIGSAIRMVFTLNYHKKTTTFTENRVFWLCF
VYDRLVSVRFGFPLMINELDIDVSLLNEPTISSQETLIDSCHFNWQVKLANIITQTLRKIYTRNSFSFI
HNCYTVLKELKYWLDGLPDDLKIDMDNFSTTQPRSTINLHINYNYTIIITTRPVFLYVFNXVADSEQKA
EELFPKKLLHTITTLLESSGQAAQIQSFIFTK
```

**Top BLAST hit:** GENE ID: 2892177, KLLA0C18953g, hypothetical protein*, Kluyveromyces lactis*

**Putative functional domain:** pfam04082, Fungal specific transcription factor domain.

---

**Identifier:** SGRP_Hypothetical protein_24

**Sequencing reads:** 322134S-17d07.q1k

**Predicted protein sequences:**

```
MLQSLKQSGKVDKLWLFTNAYKNHGVRCVKLLGVADLFDGITYCDYAQRDTLICKPDVRAFERAKLQSG
LGDYHNAWFVDDSGNNIDQGIALRMRKCIHLVEKEVDENLGKTPAGSHVIQEIIHLPKALPELF
```

**Top BLAST hit:** EDV10424.1, protein SSM1, *S. cerevisiae*, RM11-1a

**Putative functional domain:** COG1011, Predicted hydrolase (HAD superfamily)

---

**Identifier:** SGRP_Hypothetical protein_25

**Sequencing reads:** 322134S-18n10.q1k

**Predicted protein sequences:**

```
MNAEERTSDSTLATIMMLAAFDIFFSDKRRKWRAHVYGAGRLIMERLCDSGSNMLTISDEGESNDLFFI
TRWFSYVDIIGSLSSTSKVITSEKLRAIKYKFEKMTDQENWSRRRINLKDIEAGTGLEAKVLSYLADVS
WLIREREQRQDANGGEITQKLLSQVLELDYEINAHLNESERERDEIFKAYYSQGRPEIHKGYRILRATN
LIFGLTGPYD
```

**Top BLAST hit:** Match with hypothetical protein SCRG_03376, *S. cerevisiae* RM11-1a

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_26

**Sequencing reads:** 322134S-1g05.q1k

**Predicted protein sequences:**

```
MEKIELLNKQIDNSFTKDQTITLTSNLEKTTLRYHLKRSHVSDYYIDLIIKESISGKYLDWHFITENCY
FEELSPCVSFFSCGNGKLEATKMLEKYPSLKDLLLSFGVSEATMDMLSTKESPITCTRQHDTLCEISPT
ESMMHALSLYNNDNIYFLKYFICFILDRKVYESMDCDMIWCTKIYNTLSKEHFTKVYMSLVDKQDYLMH
YR
```

**Top BLAST hit:** KRE29, *S. cerevisiae*, S288c

**Putative functional domain:** pfam08691, DNA repair protein Nse5. Nse5 is a non essential nuclear protein that is critical for chromosome segregation in fission yeast. Nse5 forms a dimer with Nse6 and facilitates DNA repair as part of the Smc5-Smc6 holocomplex.

---

**Identifier:** SGRP_Hypothetical protein_27

**Sequencing reads:** 322134S-7n18.q1k

**Predicted protein sequences:**

```
MSKARKRSKVACSRCSCRKVRCSGDRPSCRACIISSNGNSCTYPLKSRKISVLDTDIKKMEEKMNALEA
ECCRLRSIQRGHDCSSHPGKHMENLIDSNHFFPTMKPSADSSNCTTDELMLGSLEIKPKVPGNSSCQRF
VGALRWHLIRNASGHDSAGGVAECLKYNVNEGRESLTKRLYLDDQERGDNQKAILPERAYASELIHRVY
QFFAKEYGLFSITEFHVRLEETYRDVQSQDPSWLAYLMVTFAVGEQYTNDAAGVGI
```

**Top BLAST hit:** ref|XP_001483600.1|  hypothetical protein PGUG_04329, *Pichia guilliermondii*

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_28

**Sequencing reads:** DBVPG6040-15p11.q1k

**Predicted protein sequences:**

MTKTDQKTAVALNNQAIQDWNIPDIEVPTYPRKAVKEGIVHLGVGGFHRSHLALYMNRLMQNHGTKDWS
ICGVGLMRFDAPMRDALQSQDCLYTLMERGIEKTTTQVIGSITSYMGI

**Top BLAST hit:** hypothetical protein SCRG_03070, *S. cerevisiae RM11-1a*

**Putative functional domain:** pfam01232, Mannitol dehydrogenase Rossmann domain

---

**Identifier:** SGRP_Hypothetical protein_29

**Sequencing reads:** DBVPG6040-1p12.q1k

**Predicted protein sequences:**

MIKQITNVTSEELVAILDIWLQANIDAHHFIPKEYWERNYEFVRSTLPKATLFTYCVGNEIVGFLGLMG
SYIAGIFVKKQWRSCGIGRKLINTVKAEKMRLSLSVYDKNERAISFYLSEGFTLKEKKIESETNEIESI
LFWASNR

**Top BLAST hit:** YJM-GNAT, *S. cerevisiae YJM789*

**Putative functional domain:** PRK10562, predicted acyltransferase with acyl-CoA N-acyltransferase domain

---

**Identifier:** SGRP_Hypothetical protein_30

**Sequencing reads:** DBVPG6040-2b02.q1k

**Predicted protein sequences:**

LHXXVCLQILLLVPRSAGMTIVSALLYLCALRVTYNFEKMQDAMTXSGLTLGFYGGIAIVIGIPYQLLC
MPETKNRTLEEIDDIFEKPTRQIIRENLSHLRKGRISY

**Top BLAST hit:** emb|CAC08232.1, fructose symporter, *S. pastorianus*

**Putative functional domain:** pfam00083, Sugar (and other) transporter.

---

**Identifier:** SGRP_Hypothetical protein_31

**Sequencing reads:** DBVPG6040-16j12.q1k

**Predicted protein sequences:**

MNRIACLGCRESKRQCDSEQPICSRCIKTGRTCRYELFSKRKPATNRYVQSLKNRIRSLEGVLKVSHEV
EFENYKSKDVNVYPNLRELNFGRDVLEHPKKIR

**Top BLAST hit:** XP_001482887.1, hypothetical protein PGUG_04842*, Pichia guilliermondii*

**Putative functional domain:** pfam00172, Fungal Zn(2)-Cys(6) binuclear cluster domain

---

**Identifier:** SGRP_Hypothetical protein_32

**Sequencing reads:** DBVPG6044-25i22.p1k

**Predicted protein sequences:**

MTLLVSSALAIDSNDLGFYNITLSWIDRNVTDSPENLVHIPPGANISDYITMYANTSTILKDLSTNIVY
TDLLIVPPANLTDGQALNARSVKSYNYDGWQQSWKQVQTLRSGQWWSPWYPASHCFWNGKNAGGSVSVP
IEMGYQYTWSMDWSAGLSSNVLSATVGMSVSSAVSRSMQVDCTXQRETMGQVW

**Top BLAST hit:** GENE ID: 2886474 CAGL0A02255g | hypothetical protein*, Candida glabrata*

**Putative functional domain:** COG4143, ABC-type thiamine transport system, periplasmic component, coenzyme metabolism.

---

**Identifier:** SGRP_Hypothetical protein_33

**Sequencing reads:** L_1374-11i16.q1k

**Predicted protein sequences:**

IPSALLFERAAGCESNEGSNGGWGSGWGGWKGGWGSGWKGGWGTGSKGGSKGGSKGGSIYNVAPHCPNL
DFGWRADECDRMNFSMDLLAVDWIESETYGVTVKVQGAESIDWKYLSSLKLTGIDGPQSCVEVTKSNKD
CSINSATEFTVSFPVYAQKIENEPCQVLMPSFQIEYEFLKGEAQQYSEGWKWGETCFNLESGCQHQSSS
SKANCDFPLWHWNCGHIPGCPSSSSSTSTASGSSTTWTPITSETPCTETP

**Top BLAST hit:** gb|EDV12484.1| hypothetical protein SCRG_03373, *S. cerevisiae* RM11-1a

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_34

**Sequencing reads:** L_1528-23b02.q1k

**Predicted protein sequences:**

MDDLATNNATILKRDSSDVSCVNETCQYVDYHVDDEGVITIDISTYRIPVEWDNGSAGNASYGVSKRDT
KYETFCKKKICGINVSGFCNAYDFAVPAFDFGGNVYNLVSGITDRIKEATKRDKTECLGYELDHVGD

**Top BLAST hit:** gb|EDN61464.1| killer toxin [Saccharomyces cerevisiae YJM789]

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_35

**Sequencing reads:** NCYC361-14b18.q1k

**Predicted protein sequences:**

LSSSKSTRQTTLVTVTSCESGICSETASPAIISTATTTINDAVTVYTTWCSLTANDKGDTVEMNTSVGS
TPAILEGSQTETVTKTNEKYSISESETHLPTATQNIIHSNGASSAPDTSKVEAIASTYLTTISQQPRRG
S

**Top BLAST hit:** BAG49462.1, flocculin*, S. pastorianus*

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_36

**Sequencing reads:** NCYC361-33h22.p1k

**Predicted protein sequences:**

SCISFKRANVKYYMSFLVFSWSIITLSSGFVRSHKSLLALRVLLGTFEGGFFPAMTLIISIVYKPQEQA
KRIAFFFGSAALSGAFGGLIATGLSSVKNGGGLEGWRWLYIIEGLISVCASVWLFFGLPAKFEVLPFLN
ERECHLMSIRSKQRTQYMGSTEKFSWSCVKDAFLDFKTYLSFTIQFCQDTIMYGFSTFLTAILKMGLGF
SSMQAQYLSVPVYILAGVVFLISAFLSDRFKMRGPIFFCYNLLGIVGYILLLSVGNDAVKYFACYLITF
SLYTGTXLNISWXTNNMAPHYKRXTAF

**Top BLAST hit:** KLLA0C19019g, hypothetical protein, *Kluyveromyces lactis*

**Putative functional domain:** COG2271, UhpC, Sugar phosphate permease, Carbohydrate transport and metabolism

---

**Identifier:** SGRP_Hypothetical protein_37

**Sequencing reads:** UWOPS83_787_3-2i14.p1k

**Predicted protein sequences:**

```
DLLIIAQMFFVLNVFNPFQGSFSYLAVCFPNYYASVMAGNGLFRAVFACAFPLFGRAMYKNLGTTKYPV
AWGSSLIGFFGVGLAAIPFIMYKNGSKLRGRSRYAAA
```

**Top BLAST hit:** EDN64624.1, major facilitator transporter, *S. cerevisiae* YJM789

**Putative functional domain:**

---

**Identifier:** SGRP_Hypothetical protein_38

**Sequencing reads:** 273614N-8k04.p1k

**Predicted protein sequences:**

```
MSKLVPIASPNLSMMTSRAYDEAKKEKGSQGTIDHEKGEYVDISGREIISSESEKHGRQLPLQAYVHYA
EIQREFERERDEDGLYAVQREQMYTDPHSNVSPSKLDSQRMLRIAKAYSVFFLITTDILGPSNAPYAVA
QMGWVPGVILYVIFGVAAAFGGWLLNFCFCKVDSNNYPIRTFSDLAARVVAPWFRYPFGLLQFIQMILN
CGLLLLSTAQSVSQMLV
```

**Top BLAST hit:** XP_505107.1, hypothetical protein*, Yarrowia lipolytica*

**Putative functional domain:** pfam01490, Transmembrane amino acid transporter protein

---

**rDNA analysis**

Raw sequencing reads were processed in three stages. In stage one, using the rDNA consensus sequence derived from the *S. cerevisiae* reference strain S288c[42,43], a series of 100 bp (rDNA) query sequences were selected at 20bp sliding intervals. These sequences were used for gapped BLAST queries against the complete *S. cerevisiae* whole-genome shotgun sequencing (WGSS) database to produce a fresh database of reads definitely containing rDNA sequence. In the second stage, less stringent BLAST searches were performed to produce alignments that might be quite divergent from the S288c consensus sequence. False positives due to sequencing error were accepted at this stage in order to ensure comprehensive sampling of variability. In the third stage, multiple alignments were generated using MUSCLE[44] with the default parameters. At this stage, the original BLAST window query was included with each stack of read fragments to be aligned to ensure the resulting output alignments could be compared to the S288c rDNA consensus sequence. In order to distinguish authentic sequence variation from variation due to sequencing error, PHRED quality (q) scores, published along with the WGSS database, were extracted and examined. To minimise the expected number of false positives due to sequencing error to less than one, candidate polymorphisms (base substitutions only) were accepted only if they appeared either in two or more covering reads with q scores >38, or in three or more reads if all q scores were >25. Using this strategy, rDNA polymorphisms were identified, their frequencies calculated, and their positions mapped (in relation to the S288c rDNA consensus sequence) for each individual *S. cerevisiae* strain.

**Analaysis of Ty abundance**

Ty element abundance was estimated from the ABI shotgun sequencing reads using a custom RepeatMasker library for both *S. cerevisiae* and *S. paradoxus* constructed as follows: (1) all "Saccharomyces" repeats (including TEs, simple repeats and RNA genes) were exported from the RepeatMasker 20061006 library; (2) the TY element was removed since it is redundant with Ty1; (3) the Ty4 entry was replaced with two entries: one for the internal region (633-6116) and one LTR entry (262-632) from GenBank accession X67284; (4) the LTR (AY198187.1) and internal region (AY198186.1) for Ty3-1p[32] was added; (5) all names were standardized to a common format. RepeatMasker open-3.1.6 was run on each strain using the parameters "`-s -nolow -xsmall -gff -no_is`" and Ty element content for each strain was estimated from the LTR component of the RepeatMasker ".tbl" file.

**Global phenotyping of yeast isolates**

Sequenced strain isolates as well as 10 individual replicates of the haploid reference strain BY4741[45] were subjected to precise phenotyping in 67 experimental conditions using a high-resolution micro-cultivation approach. Two consecutive rounds of 48-hour pre-cultivation in SC media (0.14% yeast nitrogen base, 0.1% monosodium glutamic acid, 1% succinic acid, 2% (w/v) glucose and 0.077% Complete Supplement Mixture (ForMedia), pH 5.8) were followed by a 72-hour cultivation in stress media (see Figure S9) using micro-cultivation instruments Bioscreen C (Growth curve Oy, Finland).

Readings of optical density were taken every 20 minutes. Strains were tested as duplicates (N=2). Growth curves were calibrated and the growth variables growth lag (adaptation time, h), growth rate (doubling time, h) and growth efficiency (change in cell density, OD units) calculated as earlier described[46,47]. Growth variables were normalized to the behaviour of the 20 BY4741 replicates forming the ratio LN (BY4741/isolate). This ratio is referred to as LSC, Logarithmic Strain Coefficient, and reflects strain isolate sensitivity relative the reference strain BY4741.

$$\mathrm{LSC}_{\mathrm{ij}} = \tfrac{1}{2}\sum_{r=1}^{2}\left[\tfrac{1}{20}\sum_{k=1}^{20}\log(wt_{kj}^{r}) - \log(x_{kj}^{r})\right]$$

$wt_{kj}$ is the growth variable of the k^th^ measurement of the wild-type in environment j, $x_{kj}$ is the growth variable of strain i in environment j and r indicates the run. In total, 60,000 growth curves, corresponding to 12 million measurements of optical density, were evaluated.

To test for statistical significance of environments where *S. paradoxus* phenotypes are clearly distinct from *S. cerevisiae* phenotypes a two sample Student's t-test and Boole-Bonferroni correction was carried out. The phenotypes most clearly ($p<10^{-9}$) separating the two species were strong *S. paradoxus* resistance to cycloheximide and sensitivity to paramomycin, heat and copper. Similarly, using a two sample Student's t-test it was found that the overall characteristic separating the two main groups of *S. cerevisiae* strains, one containing most of the Wine/European lineage and most of the long-branch recombinants, the other mainly consisting of the North American, Malaysian and African lineages, was rapid growth (short lag and steep slope in rate, $p<10^{-4}$) for the Wine/European and mosaics. To test for differences in degree of phenotypic variation among *S. paradoxus* strains (excluding the Hawaiian isolate) as compared to among *S. cerevisiae* strains a two sample Student's t-test was carried out comparing the overall phenotypic variance averaged over all environments (equal weights) within each species. It was found that *S. paradoxus* show significantly (p=0.002) lower phenotypic variation. A similar test also revealed that there was no significant (p=0.78) deviation in phenotypic variation between the two main groups of *S. cerevisiae* strains.

| | Source or location[a] | Strains | [b]ABI | [c]IGA |
|---|---|---|---|---|
| *S. cerevisiae* | | 36 | 43.3 | 37.3 |
| | Fermentation | 13 | 11.4 | |
| | Clinical | 6 | 5.6 | |
| | Wild | 9 | 10.0 | |
| | Laboratory | 4 | 10.6 | 37.3 |
| | Baking | 3 | 2.6 | |
| | Unknown | 1 | 3.1 | |
| *S. paradoxus* | | 35 | 38.5 | 151.8 |
| | England | 18 | 10.7 | 71.8 |
| | Continental Europe/Siberia | 6 | 12.6 | 80.0 |
| | Far East Russia/Japan | 4 | 6.8 | |
| | North & South America | 6 | 6.8 | |
| | Hawaii | 1 | 1.6 | |

**Table S1.** Origin of strains and sequence coverage

[a]Geographic origin of *Saccharomyces* strains and more detailed information are given in Supplementary table S2. Number of [b]ABI and [c]Illumina GA (Solexa) nucleotides successfully aligned and divided by the genome size estimate to give total coverage depth for each category of strains.

| OS | STRAIN | Geographic, Isolated by, Year and references | Source | Provided by | Notes, genotype |
|---|---|---|---|---|---|
| 96 | S288c | Merced, California, USA, Mrak E, 1938[48] | Rotting fig | Haber JE | Laboratory strain used in the genome project. Mat α |
| 17/A | SK1 | USA, Kane S, pre-1974[49] | Soil | Haber JE | Laboratory strain used in meiotic studies. |
| 281 | W303 | Created by Rothstein R by multiple crossing[50,51] | NA | EUROFAN | Laboratory strain,64/A *Mat a, ura3-1, trp1-d2, leu2-3, his3-11, ade2-1, can1-100* |
| 97/A | Y55 | France, Winge Ö, between 1930-60[52] | Grape | Haber JE | Laboratory strain |
| 284/A | 322134S | Royal Victoria Infirmary, Newcastle UK, Galloway A | Clinical isolate (Throat-sputum) | Mackenzie D | *Mat a* |
| 287/A | 378604X | Royal Victoria Infirmary, Newcastle UK, Galloway A | Clinical isolate (Sputum) | Mackenzie D | *Mat α* |
| 288/A | 273614N | Royal Victoria Infirmary, Newcastle UK, Galloway A | Clinical isolate (Fecal) | Mackenzie D | |
| 258/A/A | YS2 | Australia[53] | Baker strain | Bell P | |
| 259/A/A | YS4 | Netherlands, 1975, Barnett J[53] | Baker strain | Bell P | NCYC817 |
| 262/A/A | YS9 | Singapore[53] | Baker strain | Bell P | Le Saffre yeast, commercial |
| 270/A | UWOPS83-787.3 | Great Inagua Island, Bahamas, 1983, Lachance M | Fruit, *Opuntia stricta* | Lachance M | |
| 271/A | UWOPS87-2421 | Puhelu Road, Maui, Hawaii, Lachance M, 1987 | Cladode, *Opuntia megacantha* | Lachance M | |
| 220/A | L-1374 | Cauquenes, Chile, Ganga A, 1999 | Fermentation from must País | Martinez C | |
| 221/A | L-1528 | Cauquenes, Chile, Ganga A, 1999 | Fermentation from must Cabernet | Martinez C | |
| 181 | BC187 | Napa Valley, Bisson L, USA[54] | Barrel fermentation | Gerke J | Spore derivative of UCD2120 |
| 150/A | DBVPG1106 | Australia, 1947, Fornachon J | Grapes | Vaughan A | |
| 91/A | DBVPG1373 | Netherlands, Capriotti A, 1952[55,56] | Soil | Vaughan A | |
| 3/A | DBVPG6765 | Unknown[55,56] | Unknown | Vaughan A | Previously regarded as *S. boulardii* |
| 174 | YIIc17_E5 | Sauternes, France | Wine | Souciet JL | |
| 155/A | DBVPG6040 | Netherlands, 1947[57] | Fermenting fruit juice | Vaughan A | Previously regarded as *S. fructuum*. |
| 248/A/A | NCYC361 | Ireland, Gilliland R, 1952[58] | Beer spoilage strain from wort | NCYC | Previously regarded as *S. diastaticus*. |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
| 84/A | DBVPG1788 | Turku, Finland, Capriotti A, 1957[55,56] | Soil | Vaughan A |  |
| 92/A | DBVPG1853 | Ethiopia, Rossi J, 1959[55,56] | White Teff | Vaughan A |  |
| 303/A | YJM978 | Ospedali Riuniti di Bergamo, Italy, 1994-6[59] | Isolated from vagina of patient suffering from vaginitis | McCusker J |  |
| 304/A | YJM981 | Ospedali Riuniti di Bergamo, Italy, 1994-6[59] | Isolated from vagina of patient suffering from vaginitis | McCusker J |  |
| 308/A | YJM975 | Ospedali Riuniti di Bergamo, Italy, 1994-6[59] | Isolated from vagina of patient suffering from vaginitis | McCusker J |  |
| 278/A | UWOPS03-461.4 | Telok Senangin, Malaysia, Wiens F, 2003[60] | Nectar, Bertram palm | Lachance M |  |
| 279/A | UWOPS05-217.3 | Telok Senangin, Malaysia, Lachance M, 2005 | Nectar, Bertram palm | Lachance M |  |
| 280/A | UWOPS05-227.2 | Telok Senangin, Malaysia, Lachance M, 2005 | *Trigona* spp (Stingless bee) collected near Bertam palm flower | Lachance M |  |
| 251/A/A | K11 | Japan, 1981[61] | Shochu sake strain | Fay J | Awamori-1 |
| 252/A | Y9 | Indonesia, pre-1962[61] | Ragi (similar to sake wine) | Fay J |  |
| 253/A/A | Y12 | Ivory Coast, pre-1981[61] | Palm wine strain | Fay J |  |
| 182 | YPS606 | Pennsylvania, USA, Sniegowski P, 1999[62] | Bark of *Q. rubra* | Gerke J | Woodland isolate. Spore derivative of YPS142 |
| 104/A | YPS128 | Pennsylvania, USA, Sniegowski P, 1999[62] | Soil beneath *Q. alba* | Sniegowski P | Woodland isolate |
| 247/A | NCYC110 | West Africa, Guilliermond A, pre-1914[63] | Ginger beer from *Z. officinale* | NCYC | Previously regarded as *S. chavalieri*. |
| 60/A | DBVPG6044 | West Africa, Guilliermond A, 1925[55,56] | Bili wine, from *Osbeckia grandiflora* | Vaughan A | Previously regarded as *S. manginii* |

Table S2 A

| OS | STRAIN | Geographic, Isolated by, Year | Source | Provided by | Notes and genotype |
|---|---|---|---|---|---|
| 298/A | Q31.4 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 167/A | Q32.3 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 168/A | Q59.1 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 169/A | Q62.5 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 296/A | Q69.8 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 294/A | Q74.4 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 170/A | Q89.8 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 171/A | Q95.3 | Windsor Great Park, UK Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 172/A | S36.7 | Silwood Park, UK, Koufopanou V, 1997[64] | Bark of *Quercus* spp | Koufopanou V | |
| 40/A | T21.4 | Silwood Park, UK, Koufopanou V, 1998[64] | Bark of *Quercus* spp | Koufopanou V | |
| 297/A | W7 | Silwood Park, UK, Koufopanou V, 1996[64] | Bark of *Quercus* spp | Koufopanou V | |
| 165/A | Y6.5 | Silwood Park, UK, Koufopanou V, 2003[65] | Bark of *Quercus* spp | Koufopanou V | |
| 164/A | Y7 | Silwood Park, UK, Koufopanou V, 2003[65] | Bark of *Quercus* spp | Koufopanou V | |
| 302/A | Y8.1 | Silwood Park, UK, Koufopanou V, 2003[65] | Bark of *Quercus* spp | Koufopanou V | |
| 299/A | Y8.5 | Silwood Park, UK, Koufopanou V, 2003[65] | Bark of *Quercus* spp | Koufopanou V | |
| 293/A | Y9.6 | Silwood Park, UK, Koufopanou V, 2003[65] | Bark of *Quercus* spp | Koufopanou V | |
| 301/A | Z1 | Silwood Park, UK, Koufopanou V, 2003[65] | Bark of *Quercus* spp | Koufopanou V | |
| 173/A | Z1.1 | Silwood Park, UK, Koufopanou V, 2003[65] | Bark of *Quercus* spp | Koufopanou V | |
| 26/A | N-17 | Tartastan, Russia[66] | Exudate of *Q. robur* | Naumov G | |
| 142 | CBS432 | Moscow area, Russia, pre-1931[66] | Bark of *Quercus* spp | Naumov G | Neotype strain of *S. paradoxus* |
| 98/A | CBS5829 | Denmark, Jensen V, pre-1967[66] | Mor soil, pH3.6 | Naumov G | |

| | | | | | |
|---|---|---|---|---|---|
| 28/A | DBVPG4650 | Marche, Italy, Bartolini, pre-1992[55,56] | Fossilized guano in a cavern | Vaughan A | |
| 254/A | KPN3828 | Novosibirsk, Siberia, Russia, Yurkow A, 2003[67] | Bark of *Q. robur* | Iurkow A | |
| 255/A | KPN3829 | Novosibirsk, Siberia, Russia, Yurkow A, 2003[67] | Bark of *Q. robur* | Iurkow A | |
| 76/A | N-43 | Vladivostok, Russia, Naumov G, 1987[68] | Exudate of *Q. mongolica* | Naumov G | |
| 77/A | N-44 | Ternei, Russia, Naumov G, 1987[68] | Exudate of *Q. mongolica* | Naumov G | |
| 78/A | N-45 | Ternei, Russia, Naumov G, 1987[68] | Exudate of *Q. mongolica* | Naumov G | |
| 137/A | IFO1804 | Japan[68] | Bark of *Quercus* spp | Pérez-Ortín J | |
| 115/A | YPS138 | Pennsylvania, USA, Sniegowski P, 1999[62] | Soil beneath *Q. velutina* | Sniegowski P | |
| 32/A | DBVPG6304 | Yosemite, California, USA, Phaff H, 1951[69] | *Drosophila pseudoobscura* | Vaughan A | |
| 186/A | A4 | Mont St-Hilaire, Quebec, Canada, Bell G and Replansky T, 2003[65] | Bark of *Quercus rubra* | Koufopanou V | |
| 187/A | A12 | Mont St-Hilaire, Quebec, Canada, Bell G and Replansky T, 2003[65] | Soil beneath *Q. rubra* | Koufopanou V | |
| 20/A | UFRJ50791 | Catalao point, Rio de Janeiro, Brazil, pre-1992[70] | *Drosophila spp* | Naumov G | Previously regarded as *S. cariocanus* |
| 21/A | UFRJ50816 | Tijuca Forest, Rio de Janeiro, Brazil, pre-1992[70] | *Drosophila spp* | Naumov G | Previously regarded as *S. cariocanus* |
| 273/A | UWOPS91-917.1 | Saddle Road, Island of Hawaii, 1991, Lachance M | Flux of *Myoporum sandwichense* | Lachance M | |

Table S2 B

**Table S2** provides additional information on *S. cerevisiae* (A) and *S. paradoxus* (B) strains sequenced. OS column is the accession number in the internal collection at the University of Nottingham. OSNNN/A means a single spore was isolated from the original diploid and NNN/A/A indicates this process was repeated. The absence of any /A indicates that either the strain was haploid, as for S288c and W303, or a monosporic culture was provided.

| Strain | Bases aligned to ref | Total bases aligned | Reads placed | % unplaced reads | Sequenced (S0) | Sequenced at q≥40 (S40) | Imputed (I0) | Imputed at q>=40 (I40) | SNP rate for S0 | SNP rate for S40 | SNP rate for I0 | SNP rate for I40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *S. cerevisiae* | | | | | | | | | | | | |
| 273614N | 10603108 | 10805160 | 10843 | 2.46 | 55.54 | 34.10 | 95.56 | 56.18 | 12.59 | 4.92 | 4.35 | 3.54 |
| 322134S | 11139085 | 11313055 | 11524 | 3.01 | 55.52 | 34.43 | 95.80 | 54.77 | 13.06 | 4.98 | 4.20 | 3.36 |
| 378604X | 12240646 | 12411973 | 12465 | 2.06 | 60.44 | 38.17 | 95.81 | 59.80 | 12.67 | 4.92 | 4.20 | 3.32 |
| BC187 | 7960499 | 8039840 | 9337 | 2.18 | 45.65 | 29.11 | 95.59 | 55.34 | 7.76 | 4.16 | 4.02 | 3.32 |
| DBVPG1106 | 8305777 | 8393548 | 8760 | 1.34 | 44.80 | 26.04 | 95.50 | 47.98 | 11.48 | 4.50 | 3.87 | 3.24 |
| DBVPG1373 | 15433642 | 15582905 | 17794 | 1.28 | 68.31 | 51.71 | 95.89 | 80.64 | 6.75 | 4.06 | 3.97 | 3.52 |
| DBVPG1788 | 13704629 | 13833361 | 16399 | 1.68 | 64.13 | 46.19 | 96.04 | 79.35 | 7.27 | 4.13 | 3.95 | 3.51 |
| DBVPG1853 | 11023893 | 11147625 | 12930 | 3.59 | 55.26 | 38.67 | 95.61 | 55.66 | 8.18 | 5.17 | 4.64 | 3.45 |
| DBVPG6040 | 9974326 | 10088246 | 10503 | 2.75 | 52.04 | 31.68 | 95.01 | 51.64 | 11.38 | 4.55 | 4.13 | 3.19 |
| DBVPG6044 | 16748247 | 17019143 | 20328 | 2.52 | 69.80 | 53.95 | 95.79 | 90.32 | 9.07 | 7.09 | 7.12 | 6.56 |
| DBVPG6765 | 37690033 | 38082045 | 43970 | 2.27 | 91.54 | 83.25 | 97.02 | 93.20 | 5.04 | 4.11 | 4.17 | 3.78 |
| K11 | 10427849 | 10572577 | 10850 | 1.28 | 55.79 | 35.81 | 95.47 | 57.23 | 12.09 | 5.89 | 5.26 | 4.35 |
| L-1374 | 11737402 | 11841250 | 13685 | 1.51 | 57.73 | 41.29 | 96.02 | 73.46 | 7.24 | 4.06 | 3.98 | 3.48 |
| L-1528 | 12776138 | 12885697 | 15212 | 1.51 | 60.81 | 43.97 | 95.79 | 75.35 | 7.46 | 4.23 | 3.98 | 3.42 |
| NCYC110 | 9984080 | 10152441 | 10537 | 3.38 | 49.60 | 30.75 | 95.88 | 85.34 | 13.24 | 7.44 | 7.06 | 5.78 |
| NCYC361 | 6831285 | 6950628 | 7949 | 4.51 | 31.93 | 20.09 | 95.34 | 31.40 | 12.29 | 5.14 | 4.18 | 3.52 |
| RM11_1A | 27763127 | 28126688 | 27862 | 10.62 | 95.96 | 93.26 | 96.55 | 95.89 | 4.25 | 3.98 | 4.19 | 3.64 |
| S288c | 14587575 | 14646276 | 17059 | 2.30 | 97.63 | 87.44 | 98.45 | 97.80 | 0.18 | 0.06 | 0.06 | 0.05 |
| SK1 | 39677434 | 40422817 | 47667 | 2.80 | 96.75 | 95.56 | 97.18 | 96.61 | 6.90 | 6.75 | 6.83 | 6.59 |
| UWOPS03.461.4 | 11594012 | 11771328 | 12086 | 2.46 | 59.20 | 38.17 | 94.61 | 82.72 | 13.59 | 7.37 | 6.91 | 5.70 |
| UWOPS05.217.3 | 11372857 | 11558874 | 11732 | 2.83 | 51.68 | 34.64 | 95.16 | 77.77 | 13.65 | 7.27 | 6.87 | 5.50 |
| UWOPS05.227.2 | 12250830 | 12443738 | 12906 | 1.93 | 60.33 | 40.29 | 94.70 | 83.67 | 12.25 | 7.31 | 6.93 | 5.80 |
| UWOPS83.787.3 | 10942739 | 11121026 | 11371 | 1.28 | 54.93 | 35.47 | 95.62 | 54.54 | 12.39 | 5.76 | 5.30 | 4.11 |
| UWOPS87.2421 | 11287925 | 11414229 | 11504 | 1.76 | 56.99 | 36.56 | 95.43 | 56.99 | 13.06 | 6.20 | 5.38 | 4.31 |
| W303 | 30587568 | 30825148 | 30491 | 2.08 | 96.20 | 71.05 | 98.03 | 97.19 | 1.77 | 0.88 | 0.89 | 0.72 |
| Y12 | 10286038 | 10405760 | 10500 | 2.03 | 54.42 | 33.32 | 94.92 | 67.96 | 13.10 | 5.98 | 5.35 | 4.59 |
| Y55 | 41917558 | 42532877 | 50329 | 2.82 | 96.66 | 94.10 | 97.29 | 96.59 | 6.16 | 6.00 | 6.08 | 5.84 |
| Y9 | 9593428 | 9711095 | 9705 | 1.50 | 51.55 | 31.84 | 94.97 | 66.44 | 12.94 | 6.06 | 5.32 | 4.61 |
| YIIc17-E5 | 11773495 | 11934408 | 12212 | 1.82 | 59.48 | 38.99 | 95.74 | 60.57 | 10.63 | 4.85 | 4.47 | 3.85 |
| YJM789 | 22743561 | 22902904 | 22961 | 0.65 | 96.60 | 93.36 | 97.27 | 96.47 | 5.13 | 4.78 | 5.10 | 4.18 |
| YJM975 | 12519002 | 12654653 | 12916 | 1.16 | 61.69 | 41.32 | 95.32 | 81.55 | 9.28 | 4.09 | 3.98 | 3.53 |
| YJM978 | 12339657 | 12490681 | 12526 | 1.18 | 60.79 | 40.65 | 95.50 | 81.29 | 10.03 | 4.25 | 4.02 | 3.59 |
| YJM981 | 8505186 | 8712822 | 8140 | 6.25 | 33.17 | 19.54 | 95.48 | 50.76 | 11.06 | 4.60 | 3.92 | 3.37 |
| YPS128 | 15381113 | 15571993 | 17720 | 1.61 | 68.68 | 51.66 | 95.75 | 87.09 | 8.14 | 5.83 | 5.75 | 5.05 |
| YPS606 | 18570601 | 18799911 | 22003 | 1.77 | 73.85 | 57.44 | 95.67 | 87.05 | 8.16 | 5.76 | 5.78 | 5.08 |
| YS2 | 7463314 | 7622329 | 8868 | 3.19 | 38.26 | 24.01 | 95.41 | 38.23 | 11.67 | 4.93 | 4.22 | 3.39 |
| YS4 | 12542187 | 12739908 | 12717 | 1.90 | 60.61 | 40.43 | 95.71 | 59.65 | 10.88 | 4.73 | 4.59 | 3.50 |
| YS9 | 11094911 | 11259544 | 11488 | 6.01 | 56.76 | 35.03 | 95.41 | 55.23 | 12.05 | 4.73 | 4.22 | 3.16 |
| *S. paradoxus* | | | | | | | | | | | | |
| A12 | 13246933 | 13637501 | 16192 | 2.48 | 64.33 | 46.86 | 96.14 | 63.63 | 41.30 | 36.70 | 37.99 | 37.15 |
| A4 | 13705162 | 14093209 | 16964 | 3.39 | 66.56 | 49.51 | 96.18 | 65.92 | 40.70 | 36.98 | 37.96 | 37.35 |
| CBS432 | 49822103 | 50456282 | 53230 | 1.68 | 96.72 | 89.14 | 99.24 | 96.43 | 2.46 | 1.41 | 1.39 | 0.77 |
| CBS5829 | 31072885 | 31360291 | 36052 | 2.15 | 90.22 | 79.16 | 99.15 | 89.91 | 2.56 | 1.32 | 1.24 | 0.68 |
| DBVPG4650 | 17021020 | 17215542 | 20954 | 2.64 | 73.18 | 55.87 | 98.63 | 72.76 | 3.63 | 1.41 | 1.15 | 0.52 |
| DBVPG6304 | 16922088 | 17359001 | 21136 | 2.03 | 72.93 | 57.54 | 96.18 | 72.30 | 40.26 | 37.48 | 38.11 | 37.36 |
| IFO1804 | 10069462 | 10248290 | 11102 | 1.35 | 54.57 | 35.47 | 97.37 | 53.88 | 16.79 | 12.07 | 12.26 | 11.68 |
| KPN3828 | 9906538 | 10017405 | 11094 | 2.51 | 54.62 | 35.99 | 98.12 | 53.71 | 6.38 | 1.75 | 1.24 | 0.68 |
| KPN3829 | 9704300 | 9814838 | 10851 | 1.76 | 52.70 | 34.47 | 98.06 | 51.87 | 5.44 | 1.58 | 1.22 | 0.69 |
| N-17 | 33782511 | 34131107 | 39860 | 2.20 | 91.49 | 81.54 | 98.98 | 91.20 | 2.62 | 1.48 | 1.44 | 0.74 |
| N-43 | 18308249 | 18605967 | 21242 | 1.60 | 75.76 | 59.63 | 97.49 | 75.87 | 14.60 | 12.46 | 12.54 | 12.08 |
| N-44 | 15293652 | 15560627 | 18502 | 1.36 | 69.88 | 52.11 | 97.44 | 69.61 | 14.75 | 12.42 | 12.46 | 12.09 |
| N-45 | 37491486 | 38085308 | 44904 | 2.32 | 92.71 | 83.36 | 97.75 | 92.40 | 13.69 | 12.47 | 12.60 | 11.90 |
| Q31.4 | 0 | 0 | 0 | | 99.70 | 96.61 | 99.90 | 98.82 | 1.05 | 0.88 | 0.91 | 0.75 |
| Q32.3 | 14709963 | 14851989 | 17841 | 1.43 | 68.23 | 49.83 | 98.79 | 67.77 | 4.21 | 1.11 | 0.83 | 0.39 |
| Q59.1 | 14218580 | 14360798 | 17180 | 1.25 | 67.47 | 46.15 | 98.62 | 66.96 | 4.70 | 1.43 | 0.84 | 0.39 |
| Q62.5 | 15131650 | 15278084 | 19192 | 1.54 | 68.56 | 50.88 | 98.78 | 68.11 | 4.25 | 1.27 | 0.96 | 0.53 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Q69.8 | 0 | 0 | 0 | | 99.77 | 96.67 | 99.94 | 99.00 | 0.99 | 0.83 | 0.85 | 0.70 |
| Q74.4 | 0 | 0 | 0 | | 74.38 | 7.14 | 99.61 | 70.74 | 10.13 | 1.68 | 1.14 | 0.38 |
| Q89.8 | 10000133 | 10082699 | 11767 | 1.05 | 52.97 | 36.53 | 98.35 | 52.52 | 4.19 | 1.04 | 0.82 | 0.37 |
| Q95.3 | 16859478 | 17015684 | 20580 | 1.37 | 73.81 | 57.37 | 98.94 | 73.43 | 3.31 | 1.07 | 0.81 | 0.39 |
| S36.7 | 5432507 | 5475365 | 6180 | 2.03 | 34.59 | 21.36 | 97.82 | 33.96 | 4.91 | 1.34 | 1.53 | 0.42 |
| T21.4 | 15997317 | 16139255 | 18913 | 1.33 | 70.96 | 51.93 | 98.89 | 70.54 | 3.89 | 1.06 | 0.76 | 0.37 |
| UFRJ50791 | 8443837 | 8720036 | 9241 | 2.78 | 46.46 | 29.08 | 95.99 | 45.42 | 43.85 | 35.88 | 37.12 | 36.60 |
| UFRJ50816 | 13051390 | 13442447 | 15815 | 3.51 | 61.26 | 44.48 | 96.08 | 60.55 | 41.33 | 36.81 | 37.85 | 36.87 |
| UWOPS91.917.1 | 18989853 | 19444281 | 20307 | 3.79 | 73.48 | 55.72 | 95.84 | 70.58 | 42.33 | 37.10 | 36.96 | 23.37 |
| W7 | 0 | 0 | 0 | | 92.73 | 19.93 | 99.80 | 90.16 | 2.55 | 0.87 | 0.86 | 0.44 |
| Y6.5 | 11593491 | 11694293 | 14035 | 1.29 | 59.59 | 43.14 | 98.65 | 59.23 | 3.65 | 1.12 | 0.88 | 0.48 |
| Y7 | 13637595 | 13749127 | 17572 | 1.88 | 64.65 | 47.81 | 98.81 | 64.26 | 3.67 | 1.05 | 0.85 | 0.37 |
| Y8.1 | 0 | 0 | 0 | | 87.32 | 30.44 | 99.76 | 84.19 | 4.53 | 0.88 | 0.85 | 0.40 |
| Y8.5 | 0 | 0 | 0 | | 74.32 | 5.88 | 99.59 | 70.41 | 8.50 | 1.44 | 1.08 | 0.36 |
| Y9.6 | 0 | 0 | 0 | | 72.37 | 4.08 | 99.60 | 69.01 | 9.25 | 2.43 | 1.08 | 0.43 |
| YPS138 | 14478872 | 14880967 | 18003 | 2.61 | 67.21 | 51.04 | 96.16 | 66.59 | 40.38 | 36.84 | 37.98 | 37.27 |
| Z1.1 | 10879740 | 10986081 | 13006 | 1.18 | 55.57 | 39.59 | 98.61 | 54.93 | 3.74 | 1.13 | 0.86 | 0.41 |
| Z1 | 0 | 0 | 0 | 2.46 | 85.69 | 29.07 | 99.78 | 82.18 | 5.46 | 1.01 | 0.97 | 0.42 |

**Table S3.** Genome assembly statistics. Columns indicate: strain name; number of nucleotides from ABI reads of this strain that were aligned to reference nucleotides in the final assembly; the same, but counting inserted nucleotides too; number of ABI reads placed; percentage of ABI reads that were judged good quality but that didn't get placed, mostly because of some kind of ambiguity, uncertainty or clash; “S0”, percentage of reference positions that were aligned to a nucleotide of any quality in the strain (i.e. fraction of reference covered), including both ABI and Solexa reads; “S40”, as S0, but counting only positions of quality ≥40; “I0”, as “S0”,but from the imputed data; “I40”, as I0 but counting only positions of quality ≥40; then for each of S0, S40, I0 and I40, the number of differences per thousand whose value was different from the reference, out of all positions of the respective type.

| Chromosome | Position | SGD | SGRP | Quality |
|---|---|---|---|---|
| chr01 | 3836 | C | - | 44 |
| chr01 | 3981 | A | T | 45 |
| chr01 | 3982 | T | A | 45 |
| chr01 | 5244 | G | A | 57 |
| chr01 | 6454 | C | - | 36 |
| chr01 | 6756 | - | A | 47 |
| chr01 | 6756 | - | A | 44 |
| chr01 | 21531 | - | A | 63 |
| chr01 | 25341 | C | A | 93 |
| chr01 | 25498 | G | A | 93 |
| chr01 | 25507 | C | T | 90 |
| chr01 | 25510 | G | A | 84 |
| chr01 | 25513 | T | G | 84 |
| chr01 | 25516 | A | G | 78 |
| chr01 | 25585 | A | G | 64 |
| chr01 | 25612 | G | A | 93 |
| chr01 | 25711 | A | T | 62 |
| chr01 | 25712 | A | T | 83 |
| chr01 | 25713 | C | G | 76 |
| chr01 | 25714 | G | A | 78 |
| chr01 | 25765 | G | A | 93 |
| chr01 | 25766 | T | C | 93 |
| chr01 | 25769 | T | G | 93 |
| chr01 | 25779 | G | C | 93 |
| chr01 | 25789 | C | T | 93 |
| chr01 | 25793 | G | A | 93 |
| chr01 | 25794 | C | T | 93 |
| chr01 | 25795 | A | G | 93 |
| chr01 | 25798 | A | G | 93 |
| chr01 | 25799 | G | C | 91 |
| chr01 | 25801 | G | A | 93 |
| chr01 | 36120 | C | A | 38 |
| chr01 | 36814 | A | C | 57 |
| chr01 | 40231 | C | G | 53 |
| chr01 | 41240 | A | G | 42 |
| chr01 | 41664 | T | G | 45 |
| chr01 | 41700 | C | A | 60 |
| chr01 | 41703 | C | A | 66 |
| chr01 | 47821 | T | A | 39 |
| chr01 | 47826 | C | T | 39 |
| chr01 | 48772 | T | A | 72 |
| chr01 | 49904 | C | A | 57 |
| chr01 | 50327 | C | A | 60 |
| chr01 | 55746 | A | G | 52 |
| chr01 | 55954 | A | T | 53 |
| chr01 | 58196 | A | C | 60 |
| chr01 | 58247 | A | C | 60 |
| chr01 | 58424 | C | T | 66 |
| chr01 | 62767 | A | G | 39 |
| chr01 | 70794 | C | G | 60 |
| chr01 | 96740 | G | C | 48 |
| chr01 | 97025 | A | T | 51 |
| chr01 | 97026 | T | A | 51 |
| chr01 | 97678 | C | G | 51 |
| chr01 | 97679 | C | G | 51 |
| chr01 | 98350 | C | G | 39 |
| chr01 | 98351 | G | C | 39 |
| chr01 | 99565 | - | G | 39 |
| chr01 | 99841 | A | T | 42 |
| chr01 | 100399 | G | C | 54 |
| chr01 | 110470 | C | G | 66 |
| chr01 | 110471 | G | C | 63 |
| chr01 | 113702 | C | G | 50 |
| chr01 | 113703 | G | C | 50 |
| chr01 | 120442 | C | G | 57 |
| chr01 | 134125 | - | C | 54 |
| chr01 | 134852 | T | A | 52 |
| chr01 | 134854 | G | T | 49 |
| chr01 | 152189 | C | A | 48 |
| chr01 | 152190 | A | C | 48 |
| chr01 | 164040 | T | - | 42 |
| chr01 | 167048 | G | C | 75 |
| chr01 | 167087 | - | C | 66 |
| chr01 | 167134 | - | A | 78 |
| chr01 | 167139 | - | T | 72 |
| chr01 | 167144 | - | A | 63 |
| chr01 | 167551 | G | C | 69 |
| chr01 | 167802 | C | T | 48 |
| chr01 | 171984 | G | - | 45 |
| chr01 | 172041 | - | T | 54 |
| chr01 | 172042 | - | T | 57 |
| chr01 | 172172 | - | T | 54 |
| chr01 | 172434 | G | - | 51 |
| chr01 | 174173 | G | C | 52 |
| chr01 | 174187 | C | G | 54 |
| chr01 | 174188 | G | C | 57 |
| chr01 | 178256 | G | C | 54 |
| chr01 | 178638 | G | - | 36 |
| chr01 | 178647 | T | C | 66 |
| chr01 | 179683 | T | A | 53 |
| chr01 | 199804 | T | G | 54 |
| chr01 | 214048 | C | A | 44 |
| chr01 | 216010 | A | G | 35 |
| chr02 | 7510 | G | A | 53 |
| chr02 | 7914 | A | - | 52 |
| chr02 | 11053 | T | C | 50 |
| chr02 | 11308 | A | T | 84 |
| chr02 | 11309 | A | T | 84 |
| chr02 | 11345 | A | G | 50 |
| chr02 | 11379 | A | C | 50 |
| chr02 | 13102 | C | A | 44 |
| chr02 | 13553 | G | A | 54 |
| chr02 | 13966 | C | A | 39 |
| chr02 | 13982 | G | A | 42 |
| chr02 | 15332 | G | C | 50 |
| chr02 | 23913 | C | A | 48 |
| chr02 | 23947 | - | A | 60 |
| chr02 | 28909 | - | A | 63 |
| chr02 | 30655 | G | T | 36 |
| chr02 | 36312 | T | A | 55 |
| chr02 | 38067 | C | T | 52 |
| chr02 | 38210 | A | T | 57 |
| chr02 | 38728 | C | A | 44 |
| chr02 | 38908 | G | A | 54 |
| chr02 | 38917 | C | T | 53 |
| chr02 | 38920 | C | A | 53 |
| chr02 | 42377 | C | G | 72 |
| chr02 | 45495 | T | A | 48 |
| chr02 | 48369 | A | T | 57 |
| chr02 | 55145 | A | C | 50 |
| chr02 | 68154 | T | G | 49 |
| chr02 | 68155 | G | A | 49 |
| chr02 | 73450 | C | G | 72 |
| chr02 | 75208 | A | G | 66 |
| chr02 | 88493 | G | - | 52 |
| chr02 | 89277 | G | T | 56 |
| chr02 | 92679 | C | A | 57 |
| chr02 | 92916 | G | A | 54 |
| chr02 | 92918 | C | - | 55 |
| chr02 | 92919 | C | - | 55 |
| chr02 | 92920 | G | - | 55 |
| chr02 | 93622 | C | - | 38 |
| chr02 | 95345 | G | T | 48 |
| chr02 | 114868 | G | A | 60 |
| chr02 | 114870 | A | G | 54 |
| chr02 | 210433 | A | G | 42 |
| chr02 | 220413 | A | G | 48 |
| chr02 | 221294 | G | - | 42 |
| chr02 | 221295 | T | - | 42 |
| chr02 | 237021 | - | G | 45 |
| chr02 | 237839 | G | A | 48 |
| chr02 | 237892 | G | A | 54 |
| chr02 | 238116 | G | A | 46 |
| chr02 | 238133 | G | A | 57 |
| chr02 | 241396 | G | A | 48 |
| chr02 | 245111 | T | A | 48 |
| chr02 | 250039 | T | G | 39 |
| chr02 | 250403 | C | G | 42 |
| chr02 | 254495 | G | T | 63 |
| chr02 | 254532 | C | A | 78 |
| chr02 | 254616 | T | A | 42 |
| chr02 | 254719 | - | T | 36 |
| chr02 | 254720 | - | T | 39 |
| chr02 | 259762 | C | T | 52 |
| chr02 | 265902 | C | - | 55 |
| chr02 | 265905 | T | G | 55 |
| chr02 | 265922 | C | - | 54 |
| chr02 | 315071 | G | - | 44 |
| chr02 | 315270 | C | G | 36 |
| chr02 | 315271 | G | C | 36 |
| chr02 | 323835 | C | G | 39 |
| chr02 | 323836 | G | C | 42 |
| chr02 | 370774 | C | T | 51 |
| chr02 | 374011 | T | A | 51 |
| chr02 | 375272 | T | A | 54 |
| chr02 | 376497 | A | T | 72 |
| chr02 | 383173 | - | G | 45 |
| chr02 | 385360 | C | G | 39 |
| chr02 | 385361 | G | C | 39 |
| chr02 | 388774 | T | A | 45 |
| chr02 | 388775 | T | A | 78 |
| chr02 | 389421 | C | - | 42 |
| chr02 | 389423 | T | G | 66 |
| chr02 | 389425 | G | - | 38 |
| chr02 | 392557 | C | G | 49 |
| chr02 | 392558 | G | C | 48 |
| chr02 | 392568 | C | G | 44 |

| | | | | |
|---|---|---|---|---|
| chr02 | 392569 | G | C | 44 |
| chr02 | 426394 | C | G | 60 |
| chr02 | 426396 | G | C | 60 |
| chr02 | 432380 | T | C | 52 |
| chr02 | 433377 | C | G | 47 |
| chr02 | 433378 | G | C | 47 |
| chr02 | 437344 | A | G | 54 |
| chr02 | 439496 | T | G | 54 |
| chr02 | 447535 | C | G | 39 |
| chr02 | 447536 | G | C | 39 |
| chr02 | 456359 | C | T | 55 |
| chr02 | 481671 | T | G | 75 |
| chr02 | 481672 | G | T | 78 |
| chr02 | 481674 | T | G | 75 |
| chr02 | 481675 | G | T | 72 |
| chr02 | 488600 | T | A | 39 |
| chr02 | 488601 | C | T | 39 |
| chr02 | 513365 | A | G | 50 |
| chr02 | 514965 | A | C | 45 |
| chr02 | 527099 | A | G | 75 |
| chr02 | 547442 | A | T | 54 |
| chr02 | 561435 | C | - | 54 |
| chr02 | 623445 | C | - | 54 |
| chr02 | 625500 | T | C | 68 |
| chr02 | 627421 | T | G | 55 |
| chr02 | 627433 | A | C | 54 |
| chr02 | 627435 | C | T | 54 |
| chr02 | 627436 | C | T | 54 |
| chr02 | 627438 | G | A | 53 |
| chr02 | 627439 | G | A | 53 |
| chr02 | 627486 | A | T | 78 |
| chr02 | 627487 | T | A | 78 |
| chr02 | 631342 | G | C | 51 |
| chr02 | 631343 | C | G | 63 |
| chr02 | 631345 | G | T | 69 |
| chr02 | 631346 | G | T | 65 |
| chr02 | 631349 | A | C | 72 |
| chr02 | 631350 | A | C | 75 |
| chr02 | 631352 | C | G | 84 |
| chr02 | 631353 | G | C | 84 |
| chr02 | 631907 | G | A | 67 |
| chr02 | 631912 | G | A | 66 |
| chr02 | 631936 | G | A | 54 |
| chr02 | 631940 | C | A | 54 |
| chr02 | 631981 | A | G | 69 |
| chr02 | 633189 | T | A | 57 |
| chr02 | 634934 | C | T | 51 |
| chr02 | 635192 | A | G | 51 |
| chr02 | 635247 | A | G | 60 |
| chr02 | 635822 | C | G | 51 |
| chr02 | 636141 | A | G | 69 |
| chr02 | 636306 | A | T | 45 |
| chr02 | 636309 | A | G | 45 |
| chr02 | 739341 | A | T | 52 |
| chr02 | 740291 | C | G | 39 |
| chr02 | 740292 | G | C | 39 |
| chr02 | 743936 | G | C | 48 |
| chr02 | 743938 | C | G | 51 |
| chr02 | 754908 | T | C | 63 |
| chr02 | 757291 | A | - | 44 |
| chr02 | 774080 | C | G | 55 |
| chr02 | 774430 | C | G | 52 |
| chr02 | 779354 | T | A | 55 |
| chr02 | 779355 | A | C | 55 |
| chr02 | 779356 | C | T | 55 |
| chr02 | 780541 | C | A | 73 |
| chr02 | 781354 | G | C | 45 |
| chr02 | 786737 | - | C | 51 |
| chr02 | 786763 | - | A | 51 |
| chr02 | 792276 | - | T | 39 |
| chr02 | 793986 | A | C | 36 |
| chr02 | 793987 | C | A | 36 |
| chr03 | 101652 | A | T | 54 |
| chr03 | 101655 | T | A | 51 |
| chr03 | 110881 | - | A | 51 |
| chr03 | 152641 | G | A | 39 |
| chr03 | 250563 | A | T | 66 |
| chr03 | 275421 | A | G | 69 |
| chr04 | 24415 | C | A | 48 |
| chr04 | 27070 | C | T | 36 |
| chr04 | 30786 | G | A | 49 |
| chr04 | 108307 | T | A | 48 |
| chr04 | 119470 | G | A | 55 |
| chr04 | 119564 | T | A | 57 |
| chr04 | 121289 | G | A | 63 |
| chr04 | 130626 | T | A | 49 |
| chr04 | 215996 | - | C | 72 |
| chr04 | 277105 | C | G | 60 |
| chr04 | 277106 | G | C | 60 |
| chr04 | 392615 | A | G | 42 |
| chr04 | 396450 | T | G | 51 |
| chr04 | 396503 | C | G | 54 |
| chr04 | 542031 | A | - | 46 |
| chr04 | 544061 | G | A | 42 |
| chr04 | 544064 | - | G | 42 |
| chr04 | 569993 | - | G | 39 |
| chr04 | 578230 | T | - | 54 |
| chr04 | 620161 | A | C | 48 |
| chr04 | 620163 | - | A | 48 |
| chr04 | 757628 | G | C | 44 |
| chr04 | 802921 | - | T | 44 |
| chr04 | 1017381 | - | T | 48 |
| chr04 | 1063028 | C | A | 52 |
| chr04 | 1154954 | C | T | 45 |
| chr04 | 1176394 | - | C | 52 |
| chr04 | 1194954 | - | C | 42 |
| chr04 | 1253389 | T | C | 75 |
| chr04 | 1296176 | C | G | 66 |
| chr04 | 1305674 | G | - | 39 |
| chr04 | 1402297 | C | T | 45 |
| chr04 | 1433702 | A | T | 63 |
| chr04 | 1491660 | G | A | 48 |
| chr04 | 1491666 | G | C | 54 |
| chr04 | 1516808 | - | T | 47 |
| chr04 | 1516836 | - | A | 46 |
| chr04 | 1516838 | - | A | 44 |
| chr04 | 1519598 | C | G | 35 |
| chr04 | 1519662 | C | G | 60 |
| chr05 | 9168 | G | T | 48 |
| chr05 | 18079 | A | T | 42 |
| chr05 | 48384 | T | C | 51 |
| chr05 | 154530 | T | A | 66 |
| chr05 | 232634 | C | G | 45 |
| chr05 | 268857 | - | T | 57 |
| chr05 | 278525 | C | G | 69 |
| chr05 | 278526 | G | C | 69 |
| chr05 | 305258 | G | A | 75 |
| chr05 | 308627 | C | G | 48 |
| chr05 | 308984 | G | T | 45 |
| chr05 | 309047 | G | C | 36 |
| chr05 | 312197 | - | G | 75 |
| chr05 | 352390 | A | G | 54 |
| chr05 | 434284 | C | T | 54 |
| chr05 | 449959 | - | A | 54 |
| chr06 | 66443 | C | A | 66 |
| chr06 | 95000 | A | G | 69 |
| chr06 | 95410 | T | C | 45 |
| chr06 | 98798 | - | T | 48 |
| chr06 | 98798 | - | G | 39 |
| chr06 | 118584 | G | A | 50 |
| chr06 | 164824 | - | T | 49 |
| chr06 | 173057 | C | T | 75 |
| chr06 | 181041 | - | T | 81 |
| chr06 | 181041 | - | C | 78 |
| chr06 | 181041 | - | T | 78 |
| chr06 | 181041 | - | T | 78 |
| chr06 | 191312 | T | A | 36 |
| chr06 | 191388 | G | T | 42 |
| chr06 | 192394 | C | - | 38 |
| chr06 | 219387 | A | G | 54 |
| chr06 | 236216 | C | A | 42 |
| chr06 | 248160 | A | T | 39 |
| chr07 | 8152 | | A | 44 |
| chr07 | 76881 | T | C | 51 |
| chr07 | 89587 | G | - | 51 |
| chr07 | 89610 | G | - | 50 |
| chr07 | 89735 | G | C | 54 |
| chr07 | 89751 | T | - | 39 |
| chr07 | 95082 | G | - | 46 |
| chr07 | 95084 | A | - | 46 |
| chr07 | 95447 | T | A | 54 |
| chr07 | 95448 | A | T | 55 |
| chr07 | 98172 | G | A | 63 |
| chr07 | 125487 | - | A | 54 |
| chr07 | 125909 | C | G | 54 |
| chr07 | 125910 | G | C | 54 |
| chr07 | 131262 | G | - | 37 |
| chr07 | 190796 | A | - | 55 |
| chr07 | 208781 | - | A | 69 |
| chr07 | 230256 | C | T | 48 |
| chr07 | 275965 | C | G | 69 |
| chr07 | 286729 | T | - | 51 |
| chr07 | 289525 | - | G | 63 |
| chr07 | 289567 | - | G | 36 |
| chr07 | 319332 | - | G | 54 |
| chr07 | 384063 | C | G | 42 |
| chr07 | 384064 | G | C | 42 |
| chr07 | 384846 | C | G | 66 |
| chr07 | 384847 | G | C | 66 |
| chr07 | 386981 | C | G | 63 |
| chr07 | 386982 | G | C | 63 |
| chr07 | 392896 | - | A | 78 |
| chr07 | 392901 | A | - | 52 |
| chr07 | 397085 | C | G | 69 |
| chr07 | 397086 | G | C | 69 |
| chr07 | 397241 | A | C | 54 |
| chr07 | 397855 | A | - | 38 |
| chr07 | 397868 | A | - | 38 |

| | | | | |
|---|---|---|---|---|
| chr07 | 404477 | - | G | 51 |
| chr07 | 404526 | - | G | 60 |
| chr07 | 413089 | G | - | 49 |
| chr07 | 413366 | G | C | 41 |
| chr07 | 413367 | C | G | 44 |
| chr07 | 413409 | C | G | 38 |
| chr07 | 413410 | - | C | 39 |
| chr07 | 413784 | A | - | 51 |
| chr07 | 413969 | G | - | 52 |
| chr07 | 418633 | C | - | 39 |
| chr07 | 441004 | - | T | 36 |
| chr07 | 598535 | C | T | 63 |
| chr07 | 607109 | T | G | 44 |
| chr07 | 610095 | A | G | 72 |
| chr07 | 630688 | C | T | 39 |
| chr07 | 700197 | - | A | 63 |
| chr07 | 783805 | A | C | 42 |
| chr07 | 794412 | - | C | 40 |
| chr07 | 795986 | A | G | 60 |
| chr07 | 795987 | A | G | 60 |
| chr07 | 795988 | A | G | 60 |
| chr07 | 795989 | A | G | 63 |
| chr07 | 796569 | A | C | 54 |
| chr07 | 796570 | C | A | 54 |
| chr07 | 924485 | - | A | 45 |
| chr07 | 924491 | - | C | 54 |
| chr07 | 938725 | T | - | 40 |
| chr07 | 958924 | C | - | 53 |
| chr07 | 999271 | C | T | 45 |
| chr07 | 999367 | A | C | 57 |
| chr07 | 999677 | T | - | 53 |
| chr07 | 1006285 | - | C | 44 |
| chr07 | 1032373 | G | A | 36 |
| chr07 | 1033108 | C | T | 52 |
| chr07 | 1038054 | - | C | 48 |
| chr07 | 1039740 | A | - | 38 |
| chr07 | 1041869 | C | T | 66 |
| chr07 | 1042082 | A | G | 75 |
| chr07 | 1042184 | A | G | 75 |
| chr07 | 1042307 | A | G | 39 |
| chr07 | 1042312 | A | G | 48 |
| chr07 | 1042562 | G | A | 39 |
| chr08 | 8358 | - | G | 72 |
| chr08 | 18567 | - | G | 54 |
| chr08 | 62687 | G | A | 55 |
| chr08 | 102564 | - | C | 54 |
| chr08 | 133275 | - | T | 42 |
| chr08 | 240687 | G | A | 53 |
| chr08 | 369889 | - | A | 66 |
| chr08 | 369987 | T | - | 52 |
| chr08 | 417057 | G | A | 60 |
| chr08 | 423723 | - | C | 60 |
| chr09 | 18865 | C | - | 44 |
| chr09 | 23186 | - | G | 42 |
| chr09 | 23214 | T | G | 44 |
| chr09 | 23253 | C | - | 44 |
| chr09 | 23264 | C | - | 44 |
| chr09 | 23284 | C | - | 38 |
| chr09 | 128403 | - | C | 51 |
| chr09 | 128410 | - | T | 60 |
| chr09 | 128410 | - | C | 60 |
| chr09 | 203638 | A | C | 51 |
| chr09 | 248836 | A | T | 44 |
| chr09 | 248836 | - | A | 44 |
| chr09 | 318692 | T | A | 54 |
| chr09 | 333322 | - | C | 54 |
| chr09 | 439155 | T | G | 51 |
| chr09 | 439328 | C | T | 51 |
| chr10 | 76241 | T | C | 45 |
| chr10 | 81927 | A | T | 44 |
| chr10 | 84958 | G | A | 39 |
| chr10 | 89005 | C | A | 66 |
| chr10 | 90365 | A | C | 72 |
| chr10 | 96057 | G | C | 60 |
| chr10 | 97250 | - | T | 66 |
| chr10 | 97253 | - | G | 66 |
| chr10 | 97489 | C | G | 52 |
| chr10 | 99469 | C | G | 48 |
| chr10 | 99572 | A | - | 36 |
| chr10 | 99767 | G | C | 84 |
| chr10 | 99777 | C | G | 81 |
| chr10 | 99778 | G | C | 78 |
| chr10 | 99792 | A | T | 66 |
| chr10 | 99793 | A | T | 63 |
| chr10 | 102276 | C | G | 45 |
| chr10 | 102610 | A | C | 63 |
| chr10 | 102642 | A | G | 45 |
| chr10 | 102643 | A | C | 48 |
| chr10 | 110898 | T | - | 44 |
| chr10 | 111287 | G | - | 49 |
| chr10 | 111646 | C | - | 44 |
| chr10 | 113842 | G | A | 55 |
| chr10 | 123309 | T | C | 38 |
| chr10 | 123314 | T | C | 39 |
| chr10 | 126896 | T | - | 54 |
| chr10 | 129111 | T | C | 51 |
| chr10 | 171997 | C | G | 69 |
| chr10 | 171998 | G | C | 69 |
| chr10 | 179434 | C | G | 60 |
| chr10 | 179436 | G | C | 60 |
| chr10 | 195503 | - | C | 39 |
| chr10 | 200219 | A | G | 51 |
| chr10 | 204551 | G | T | 54 |
| chr10 | 204552 | T | G | 54 |
| chr10 | 205348 | C | A | 47 |
| chr10 | 297727 | T | G | 45 |
| chr10 | 407437 | - | A | 57 |
| chr10 | 414288 | C | G | 66 |
| chr10 | 414289 | G | C | 66 |
| chr10 | 421511 | G | A | 36 |
| chr10 | 438634 | - | A | 42 |
| chr10 | 599597 | A | - | 48 |
| chr10 | 599640 | A | - | 46 |
| chr10 | 599715 | A | - | 51 |
| chr10 | 599750 | A | T | 52 |
| chr10 | 625484 | G | A | 54 |
| chr10 | 627373 | T | C | 63 |
| chr10 | 627528 | C | G | 75 |
| chr10 | 627529 | C | G | 78 |
| chr10 | 629484 | T | A | 50 |
| chr10 | 670287 | T | G | 49 |
| chr10 | 676914 | G | A | 52 |
| chr10 | 687986 | A | G | 42 |
| chr10 | 708358 | T | A | 60 |
| chr10 | 708442 | A | T | 66 |
| chr10 | 713836 | C | G | 57 |
| chr10 | 716506 | C | A | 57 |
| chr10 | 717779 | G | A | 84 |
| chr10 | 722289 | G | T | 60 |
| chr10 | 724288 | T | G | 72 |
| chr10 | 724996 | G | T | 69 |
| chr10 | 727116 | C | G | 42 |
| chr10 | 733259 | G | A | 72 |
| chr10 | 733992 | C | T | 75 |
| chr10 | 734777 | G | A | 69 |
| chr11 | 68457 | - | C | 63 |
| chr11 | 68457 | - | C | 63 |
| chr11 | 68471 | - | G | 60 |
| chr11 | 69325 | T | C | 51 |
| chr11 | 69326 | G | T | 51 |
| chr11 | 69494 | G | - | 39 |
| chr11 | 192315 | A | T | 63 |
| chr11 | 192316 | T | A | 63 |
| chr11 | 197105 | C | G | 72 |
| chr11 | 197106 | G | C | 72 |
| chr11 | 199356 | - | T | 50 |
| chr11 | 199359 | - | T | 48 |
| chr11 | 199368 | - | T | 51 |
| chr11 | 199377 | A | T | 50 |
| chr11 | 199378 | T | A | 50 |
| chr11 | 242811 | T | A | 48 |
| chr11 | 253006 | A | T | 72 |
| chr11 | 253007 | T | A | 75 |
| chr11 | 316656 | C | T | 78 |
| chr11 | 322663 | A | T | 51 |
| chr11 | 335190 | C | T | 48 |
| chr11 | 340150 | A | T | 54 |
| chr11 | 357929 | T | C | 45 |
| chr11 | 359605 | - | G | 51 |
| chr11 | 393438 | C | G | 53 |
| chr11 | 393439 | G | C | 53 |
| chr11 | 457775 | C | T | 42 |
| chr11 | 463435 | G | T | 54 |
| chr11 | 479596 | C | T | 54 |
| chr11 | 509993 | C | G | 54 |
| chr11 | 509994 | G | C | 54 |
| chr11 | 610650 | C | G | 53 |
| chr11 | 618236 | G | A | 45 |
| chr11 | 618237 | A | G | 48 |
| chr12 | 32902 | - | A | 54 |
| chr12 | 185062 | G | A | 51 |
| chr12 | 187139 | T | - | 52 |
| chr12 | 187393 | A | - | 39 |
| chr12 | 187424 | A | - | 44 |
| chr12 | 192416 | C | T | 52 |
| chr12 | 193483 | A | C | 69 |
| chr12 | 210771 | A | T | 57 |
| chr12 | 426841 | A | G | 48 |
| chr12 | 426842 | - | A | 54 |
| chr12 | 725938 | G | A | 60 |
| chr12 | 750224 | C | T | 36 |
| chr12 | 762846 | A | T | 84 |
| chr12 | 767026 | T | C | 87 |
| chr12 | 767027 | C | T | 84 |
| chr12 | 822959 | - | G | 72 |
| chr12 | 828902 | C | T | 66 |
| chr12 | 878172 | G | - | 41 |
| chr12 | 924691 | - | G | 63 |

| | | | | |
|---|---|---|---|---|
| chr12 | 1032534 | G | - | 40 |
| chr12 | 1038770 | C | - | 51 |
| chr12 | 1039433 | T | A | 42 |
| chr13 | 34473 | T | G | 51 |
| chr13 | 283467 | - | G | 60 |
| chr13 | 321234 | - | C | 54 |
| chr13 | 448332 | G | A | 51 |
| chr13 | 480878 | - | C | 50 |
| chr13 | 672128 | T | C | 50 |
| chr13 | 680935 | T | C | 54 |
| chr13 | 794225 | G | C | 36 |
| chr13 | 808976 | A | T | 38 |
| chr13 | 809197 | A | G | 39 |
| chr13 | 864682 | G | T | 45 |
| chr14 | 7478 | - | T | 44 |
| chr14 | 12986 | - | T | 48 |
| chr14 | 25775 | G | A | 54 |
| chr14 | 129279 | A | - | 47 |
| chr14 | 132573 | T | C | 60 |
| chr14 | 189081 | C | G | 57 |
| chr14 | 189082 | G | C | 57 |
| chr14 | 274732 | G | C | 59 |
| chr14 | 276356 | A | G | 50 |
| chr14 | 278945 | T | G | 47 |
| chr14 | 290429 | T | G | 57 |
| chr14 | 299120 | C | G | 69 |
| chr14 | 304165 | G | T | 51 |
| chr14 | 306231 | C | A | 54 |
| chr14 | 307728 | T | G | 57 |
| chr14 | 308753 | G | C | 48 |
| chr14 | 315278 | T | G | 66 |
| chr14 | 359024 | C | T | 50 |
| chr14 | 374768 | A | C | 75 |
| chr14 | 374810 | A | C | 51 |
| chr14 | 377889 | G | T | 60 |
| chr14 | 383566 | A | G | 57 |
| chr14 | 560911 | A | T | 54 |
| chr14 | 560914 | - | A | 54 |
| chr14 | 766523 | C | T | 44 |
| chr14 | 784223 | G | T | 93 |
| chr15 | 8476 | - | T | 66 |
| chr15 | 35765 | G | C | 42 |
| chr15 | 36013 | T | A | 42 |
| chr15 | 36056 | G | A | 60 |
| chr15 | 36119 | G | C | 36 |
| chr15 | 36149 | G | A | 51 |
| chr15 | 49829 | G | T | 60 |
| chr15 | 50069 | G | T | 39 |
| chr15 | 50081 | T | C | 45 |
| chr15 | 50430 | G | C | 36 |
| chr15 | 50431 | G | C | 36 |
| chr15 | 50767 | C | T | 54 |
| chr15 | 56036 | C | G | 66 |
| chr15 | 57699 | A | T | 48 |
| chr15 | 58601 | C | G | 81 |
| chr15 | 63708 | C | T | 54 |
| chr15 | 64173 | G | C | 45 |
| chr15 | 66022 | A | T | 60 |
| chr15 | 69084 | C | T | 45 |
| chr15 | 79297 | C | G | 42 |
| chr15 | 79304 | A | T | 63 |
| chr15 | 79305 | T | A | 63 |
| chr15 | 84121 | T | C | 54 |
| chr15 | 186243 | A | C | 48 |
| chr15 | 186979 | G | C | 57 |
| chr15 | 190052 | C | G | 66 |
| chr15 | 190053 | G | C | 69 |
| chr15 | 210453 | A | C | 42 |
| chr15 | 210486 | A | C | 44 |
| chr15 | 219000 | - | G | 42 |
| chr15 | 280393 | A | G | 60 |
| chr15 | 343142 | - | A | 66 |
| chr15 | 389237 | T | G | 48 |
| chr15 | 462780 | - | G | 53 |
| chr15 | 570495 | A | G | 48 |
| chr15 | 571534 | - | T | 51 |
| chr15 | 588318 | - | T | 57 |
| chr15 | 588362 | - | A | 36 |
| chr15 | 611023 | G | T | 53 |
| chr15 | 611024 | C | G | 53 |
| chr15 | 611036 | A | G | 47 |
| chr15 | 816965 | - | T | 51 |
| chr15 | 874911 | C | T | 60 |
| chr15 | 877795 | - | G | 51 |
| chr15 | 889947 | G | C | 63 |
| chr15 | 892173 | - | C | 48 |
| chr15 | 980884 | G | A | 48 |
| chr15 | 980884 | - | G | 48 |
| chr16 | 128039 | T | G | 57 |
| chr16 | 191943 | C | T | 69 |
| chr16 | 347528 | - | G | 39 |
| chr16 | 347759 | - | C | 84 |
| chr16 | 520376 | - | C | 60 |
| chr16 | 520378 | C | G | 60 |
| chr16 | 523639 | C | T | 48 |
| chr16 | 560302 | - | T | 72 |
| chr16 | 642955 | G | A | 51 |
| chr16 | 642995 | C | T | 57 |
| chr16 | 727933 | G | A | 54 |
| chr16 | 759398 | T | G | 54 |
| chr16 | 769390 | - | C | 42 |
| chr16 | 778303 | - | G | 66 |
| chr16 | 778377 | - | C | 66 |
| chr16 | 778863 | T | A | 49 |
| chr16 | 819447 | - | T | 54 |
| chr16 | 890342 | T | A | 53 |

**Table S4.** List of changes to the S288c genome sequence submitted to SGD. Columns are: chromosome and offset, from the SGD reference of October 2$^{nd}$ 2007; SGD nucleotide or gap; corrected nucleotide or gap; and PHRED quality score of correction.

| Lineage | Polymorphic Sites | Polymorphic Private | Polymorphic Private but in Mosaics | Monomorphic Private | Monomorphic Private but in Mosaics | Private SNPs | % Monomorphic | Number of Strains |
|---|---|---|---|---|---|---|---|---|
| Wine/EU | 32608 | 13104 | 12833 | 149 | 24820 | 40906 | 85.9 | 10 |
| North American | 681 | 187 | 76 | 8263 | 3130 | 11656 | 99.7 | 2 |
| Sake | 22548 | 8035 | 5266 | 1862 | 7064 | 22227 | 90.2 | 3 |
| Malaysian | 55312 | 1437 | 243 | 20453 | 4490 | 26623 | 76.1 | 3 |
| West African | 57940 | 147 | 257 | 2227 | 23154 | 25785 | 74.9 | 2 |
| Mosaics | 173622 | 56556 | - | 0 | - | 56556 | 24.9 | 16 |
| Non-mosaics | 175561 | 57604 | - | 0 | - | 57604 | 24.5 | 20 |

**Table S5.** Analysis of SNP distributions among *S. cerevisiae* lineages for the 231117 diallelic sites across the genome.

| Species | Population | No. of Strains | Avg. % TE | SD |
|---|---|---|---|---|
| *S. cerevisae* | global * | 30 | 1.91 | 0.611 |
| *S. cerevisae* | Wine/European ** | 8 | 1.32 | 0.158 |
| *S. cerevisae* | clean *** | 14 | 1.61 | 0.542 |
| *S. cerevisae* | mosaic | 16 | 2.17 | 0.556 |
| *S. paradoxus* | global | 27 | 1.57 | 0.477 |
| *S. paradoxus* | UK | 10 | 1.39 | 0.130 |

**Table S6.** Estimates of Ty transposable element (TE) abundance in *S. cerevisae* and *S. paradoxus* populations. Values are averages and standard deviations (SD) across strains.

* Excludes 6 potential clonemates: NCYC110, UWOPS03-461.4, UWOPS05-217.3, YPS128, YJM975, YJM981

** Excludes 2 potential clonemates: YJM975, YJM981

*** Includes Wine/European strains

| | No. strains | No. sites analysed | No. strains per site* | $\theta_\pi$ x1000 | $\theta_s$ x1000 | Tajima's D |
|---|---|---|---|---|---|---|
| *S. paradoxus* (UK) | 18 | 11,116,410 | 7.17<br>(0.06) | 1.02<br>(0.03) | 1.01<br>(0.02) | 0.10<br>(0.03) |
| *S. cerevisiae* (global)** | 32 | 11,657,564 | 14.96<br>(0.17) | 5.65<br>(0.3) | 5.93<br>(0.2) | -0.23<br>(0.05) |
| *S. cerevisiae* (Wine/European)*** | 9 | 8,682,981 | 5.07<br>(0.02) | 1.04<br>(0.04) | 1.11<br>(0.05) | -0.55<br>(0.04) |

**Table S7.** Population genomic estimates of mutational diversity.

Values are averages across the 16 chromosomes; values in parentheses are standard errors.

*average number of strains per site with a nucleotide with q>40.

** Excluding 6 potential clonemates: NCYC110, UWOPS03-461.4, UWOPS05-217.3, YPS128, YJM975, YJM981

***Excluding 2 potential clonemates, YJM975, YJM981

We have also analysed nucleotide diversity separately for each chromosome. In the global sample of *S. cerevisiae* there is a significant negative correlation between the length of the chromosome and the amount of variation, in particular for shorter chromosomes (Kendall's $\tau$ = -0.52, p=0.008). This appears to be because there is more variation in the subtelomeric regions extending 30kb from each end of each chromosome, and these regions make up a larger proportion of shorter chromosomes. If these regions are excluded, the correlation is no longer significant.

## Supplementary text

### Strain selection

Strain selection was subject to certain constraints to avoid technical difficulties with sequencing as well as maximize future utility. Except for the laboratory strains S288c, W303 and the baking strains, the isolates selected were diploid. As nearly 10% of *Saccharomyces sensu stricto* isolates appear to be hybrids between different species[56], we selected strains that behaved as diploid non-hybrids in meiosis and test crosses. We obtained a single spore isolate from the strains in order to avoid the problem of heterozygosity within the diploid genome, which would cause problems with SNP calling and assemblies. Each single spore isolate sequenced, as well as the other three spores from the same meiosis, have been stored and are available at NCYC[71].

### Strain description

The reference genome was sequenced by a large consortium of laboratories from around the world and various libraries were used from isogenic derivatives of S288c[42]. Individual chromosomes were sequenced at different times. The origins of S288c are described in Johnston and Mortimer[48]. W303 has a chequered past described by Rodney Rothstein in a short document at SGD[72]. S288c is on a long branch of the neighbour-joining tree and therefore has few large blocks similar to any of the clean lineages except for regions similar to the European cluster. W303 is clearly derived from S288c as described and shows numerous recombination breakpoints along the chromosomes compared to S288c. When not similar to S288c it is similar to other lineages studied here in many cases. For example on chromosome 2 on the left arm there is a region similar to the West African lineage while on the right arm there is a region from the European cluster and there are also regions derived from the sake lineage.

The other laboratory strains sequenced here are very interesting as they are widely used in various studies, particularly meiotic recombination studies. SK1 is first mentioned in Kane and Roth[49] but no information is given as to its origins. It has been extensively engineered to make it a useful genetic organism[73]. Y55 is first mentioned by Halvorson[52] where it is attributed to Winge. It has been extensively engineered for use in genetic studies[74]. SK1 is generally thought to be from North America while Y55 was from a grape in France. Both strains have useful meiotic properties and it is interesting to see that they both are on the West African lineage. They are clearly both recent recombinants with other lineages. The non West African segments of Y55 are mostly derived from the European cluster while those from SK1 are generally unlike any of the clean lineages with the exception of segments from the sake lineage. It may be of interest to note that Roth went through the Halvorson laboratory, where he may have obtained SK1, at the time many sporulation studies were being done with Y55 and other strains.

The clean lineages exhibit varying degrees of similarities with the West African and North American lineages being very similar within while the other lineages, although more similar within, exhibit similarities consistent with the branch lengths in the NJ tree. This is true for the sake, Malaysian and European/Wine cluster lineages. The rest of the strains are mosaics with no large tracts of similarity to the clean lineages or each other. They exhibit many small regions of similarity to many other strains but mostly with the European/Wine cluster. This is consistent with recombination between several lineages sometime in the past.

**Genome Assembly**

Reference-based genome assemblies were created for each strain in a series of steps. First, each read was aligned to the reference genome for the relevant species (S288c or CBS432). As this approach cannot deal with large indels or with sequences not present in the reference genome, we developed an iterative parallel alignment assembling tool, PALAS (see Supplementary Methods), to introduce insertions that were allowed to share material between related strains. Two versions of each strain sequence were produced, a partial assembly derived just from data collected from that strain, and a more complete assembly using an imputation process to infer the most likely sequence of the strain taking into account data from related strains. In both cases confidence estimates are given for each base call. Even using PALAS, some sequences such as the subtelomeric regions, which structurally are highly polymorphic[75], could not be reliably assembled. Because of their extreme AT-richness, the mitochondrial genome sequences are also incomplete.

The available sequence for the type strain of *S. paradoxus*[37] is not complete and so we sequenced a single spore isolate from our version of CBS432 to 4.3X coverage with ABI. Although the two sequences are closely related and are part of the same population, there are numerous SNP differences between them. This could be due to the fact that we used a single spore isolate which would not have any heterozygosity, while the assembly created by Kellis *et al*[37] has the other allele at some SNP locations. Alternatively, the difference may come from the strains actually not being the same as they were obtained from separate culture collections (Northern Regional Research Laboratory, NRRL, and Centraalbureau voor Schimmelcultures, CBS). Our version of CBS432 is identical to that in the CBS collection over specific regions of SNP differences confirmed by sequencing of an independent sample from CBS. In order to circumvent this issue, we first produce a *S. paradoxus* reference sequence with the greatest contiguity by collecting together all the reads from the UK isolates of *S. paradoxus*, the reads from the Broad project[37], and artificial reads created by shredding the Broad contigs. We assembled these reads using Phusion[36] into 608 contigs, which we then aligned to the *S. cerevisiae* reference to place them on chromosomes. We used the resulting sequence as the starting point for the reference-based assemblies for the individual *S. paradoxus* strains. We later confirm and updated this *S. paradoxus* reference genome by deep parallel paired-end sequencing producing an 80X coverage of a single, CBS432, isolate. We also aligned this sequence and the contigs for *S. bayanus, S. kudriavzevii* and *S. mikatae* (downloaded from SGD[76]) to create cross-species multiple alignments used in subsequent analyses. Further details of both procedures are in the supplementary methods.

**SGD and SGRP discrepancies**

We extracted the single-nucleotide differences (including single-base indels) between our (SGRP) version of S288c and the SGD reference. For every such position, we determined whether other strains sequenced at that position (a) all agreed with our version, (b) all agreed with SGD, or (c) were split or absent. We found that for positions where our value had a quality of 35 or more, there was a strong excess of (a) (480 cases) over (b) (18 cases), whereas for quality less than 35, the converse was true (17 (a)s to 81 (b)s for qualities 30 to 34 inclusive). We therefore concluded that the large majority of (a) cases with quality 35 or more were valid corrections, and submitted these to SGD. The (c) cases were more evenly distributed among different qualities, so that while some of them probably do denote errors in SGD, it is not clear which ones they are, so they were not submitted.

**Ty element abundance**

We estimated the overall level of Ty element abundance for each strain directly from ABI sequencing reads, since a set of dispersed features like Ty elements is sampled proportionally with light shotgun sequencing coverage and because large insertions like Ty elements present a challenge for reference-based genome assembly. The proportion of Ty sequences is typically less than 3% in all strains in both species, with the highest abundance observed in the laboratory strain S288c (3.53%) (Fig. S7). Globally, we find that the proportion of Ty sequence per strain is higher among strains in *S. cerevisiae* relative to *S. paradoxus* (Table S6; Wilcoxon Test, p = 0.02275). Interestingly, *S. paradoxus* strains from South America (UFRJ50791 and UFRJ50816), which are partially reproductively isolated from other *S. paradoxus* lineages and have been previously known as a separate species (*S. cariocanus*)[55,77], have the highest Ty abundance for this species. This correlates well with the increased rate of rearrangement due to reciprocal translocations in the South America *S. paradoxus* lineage, and supports the hypothesis that a burst of rearrangements occurred in this lineage due to increased Ty activity [78]. Levels of variation in Ty abundance are similar in the Wine/European *S. cerevisae* population and in the UK *S. paradoxus* population, and levels of variation in Ty abundance for the global *S. cerevisae* sample are substantially higher than the Wine/European *S. cerevisae* population (Table S6). Finally, we find that the mosaic strains of *S. cerevisae* have higher Ty abundance than the clean lineages (Table S6; Wilcoxon Test, p=0.006896), suggesting that hybridization may have led to an increase in Ty element activity in this species.

**rDNA variability**

Sequence coverage was sufficient to ensure that each position in the 9.1kb rDNA repeat was covered many times in each strain (average 140). A complete analysis of this dataset is given elsewhere (unpublished) but we report here an interesting correlation between intragenomic rDNA sequence variation and genome mosaicism (Fig. S6). By comparing the number of rDNA reads against total overall reads (Methods in SI) we estimated rDNA copy number for each *S. cerevisiae* strain to range between 54 (K11) and 511 (YJM981). The estimates were generally in good agreement with previous estimates[79,80]. The exceptionally high count for strain YJM981 appears to be linked to an unusual karyotype resulting in a much larger than average overall genome size (results not shown). A link between rDNA copy number and genome size has been reported previously[81]. We also investigated the possibility of a link between rDNA copy number and genome mosaicism but found them to be unrelated. However, when polymorphic sites in the rDNA repeats from individual strains were enumerated (Supplementary Methods), mosaic genomes were found to have significantly more variable positions than the clean lineages ($p=1.3 \times 10^{-6}$ under Mann-Whitney U rank test). We further characterised this variation in terms of number of substitutions per position and found the vast majority of variant sites to be in the 0-10% minor allele frequency range (Fig. S6). In contrast to Ganley and Kobayashi[82] who found only four polymorphic sites in strain RM11-1A, we identified 518 polymorphic sites of which 156 are resolved to a complete SNP (relative to the S288c rDNA consensus sequence) in at least one strain. A possible reason for the increased polymorphism within mosaic strains is that differences in the rDNA sequence between the parents of the mosaic may not have yet been resolved by gene conversion[83], suggesting the mosaics are relatively recent.

**Evidence for selection on non-coding regions**

Non-coding regions contain many functional sequences including regulatory sequences and non-coding RNA genes. We computed the derived allele frequency spectrum for the SNPs that we identified in non-coding regions (Fig. S8A). We found that there is strong evidence for purifying selection in all classes of non-coding regions in yeast (compared to synonymous sites), and that tRNA genes show a skew in their allele frequency distribution comparable to amino acid altering mutations in protein coding regions.

## Evidence for selection on synonymous sites

In yeast, genes with high expression levels exhibit high codon bias[84] and this can be measured by the codon-adaptation index (CAI[85]). We computed the derived allele frequency spectrum for the silent polymorphisms in genes with high levels of codon bias (average CAI>0.6) and found an excess of polymorphism at both low and high frequency (Fig. 3B, s*). This suggests the action of both positive and negative selection on polymorphism at these sites. To test this we identified polymorphisms in which the derived allele was either a preferred or un-preferred codon, as defined by the CAI, and compared their allele frequencies in genes with high codon bias to those in the rest of the genome. Consistent with both positive and negative selection acting on synonymous sites in highly expressed genes, we observe both an excess of low-frequency (DAF<20%) SNPs that created un-preferred codons (38% vs. 51%, $p<10^{-6}$) and an excess of high-frequency (DAF>20%) SNPs that create preferred codons (67% vs. 54%, p<0.01 Fisher's Exact Test). Codon bias in *S. cerevisiae* appears to be maintained by both purifying and positive selection, as suggested by the mutation-selection-drift model[86].

## McDonald-Kreitman tests for adaptive evolution

One of the most exciting applications of population genomics is to systematically identify mutations that may have been the targets of positive natural selection that and hence may underlie the adaptive differences between species. Comparisons of divergence to diversity in different classes of sites, such as the McDonald-Kreitman test (M-K[87]) provide a powerful means to do so, and are relatively robust to demographic complexity[88]. The M-K test compares the ratio non-synonymous to synonymous differences in polymorphism to that in divergence. An excess of amino acid differences fixed between species is interpreted as evidence that positive selection has acted to fix the amino acid differences between species. As noted above, there was a large excess of amino-acid replacement polymorphism at low allele frequencies (Fig. S8B, inset), indicating weakly deleterious alleles segregating in the population. We therefore excluded SNPs with minor allele frequency less than 20%[89] and performed McDonald-Kreitman tests on the 1105 genes for which there were at least 5 SNPs, to test the hypotheses that there was an excess of fixed amino acid differences between *S. cerevisiae* and *S. paradoxus*. Overall, we found a skew in the distribution of the M-K ratio towards values less than one (Fig. S8b), indicating either pervasive purifying selection on the differences between species, or reduction in the efficacy of selection within the current *S. cerevisiae* population[87]. In the positive tail (M-K ratio>1) of this distribution lie genes enriched for amino-acid changes between species, candidates for adaptive differences. However, none of these were significant after a correction for multiple testing.

**References**

31. Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A. & Voytas, D.F. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Res* 8, 464-78 (1998).
32. Fingerman, E.G., Dombrowski, P.G., Francis, C.A. & Sniegowski, P.D. Distribution and sequence analysis of a novel Ty3-like element in natural Saccharomyces paradoxus isolates. *Yeast* 20, 761-70 (2003).
33. Minichiello, M.J. & Durbin, R. Mapping trait loci by use of inferred ancestral recombination graphs. *American Journal of Human Genetics* 79, 910-922 (2006).
34. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17, 368-76 (1981).
35. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* (2008).
36. Mullikin, J.C. & Ning, Z. The phusion assembler. *Genome Res* 13, 81-90 (2003).
37. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E.S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241-54 (2003).
38. Ning, Z.M., Cox, A.J. & Mullikin, J.C. SSAHA: A fast search method for large DNA databases. *Genome Research* 11, 1725-1729 (2001).
39. Notredame, C., Higgins, D.G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302, 205-217 (2000).
40. Deutschbauer, A.M. et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169, 1915-1925 (2005).
41. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402 (1997).
42. Goffeau, A. et al. Life with 6000 genes. *Science* 274, 546, 563-7 (1996).
43. Mewes, H.W. et al. Overview of the yeast genome. *Nature* 387, 7-65 (1997).
44. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792-7 (2004).
45. Brachmann, C.B. et al. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14, 115-32 (1998).
46. Warringer, J. & Blomberg, A. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in Saccharomyces cerevisiae. *Yeast* 20, 53-67 (2003).
47. Warringer, J., Ericson, E., Fernandez, L., Nerman, O. & Blomberg, A. High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc Natl Acad Sci U S A* 100, 15724-9 (2003).
48. Mortimer, R.K. & Johnston, J.R. Genealogy of principal strains of the yeast genetic stock center. *Genetics* 113, 35-43 (1986).
49. Kane, S.M. & Roth, R. Carbohydrate metabolism during ascospore development in yeast. *J Bacteriol* 118, 8-14 (1974).
50. Rothstein, R.J. A genetic fine structure analysis of the suppressor 3 locus in Saccharomyces. *Genetics* 85, 55-64 (1977).
51. Rothstein, R.J., Esposito, R.E. & Esposito, M.S. The effect of ochre suppression on meiosis and ascospore formation in Saccharomyces. *Genetics* 85, 35-54 (1977).
52. Tauro, P. & Halvorson, H.O. Effect of gene position on the timing of enzyme synthesis in synchronous cultures of yeast. *J Bacteriol* 92, 652-61 (1966).
53. Bell, P.J., Higgins, V.J. & Attfield, P.V. Comparison of fermentative capacities of industrial baking and wild-type yeasts of the species Saccharomyces cerevisiae in different sugar media. *Lett Appl Microbiol* 32, 224-9 (2001).
54. Gerke, J.P., Chen, C.T. & Cohen, B.A. Natural isolates of Saccharomyces cerevisiae display complex genetic variation in sporulation efficiency. *Genetics* 174, 985-97 (2006).
55. Liti, G., Barton, D.B. & Louis, E.J. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. *Genetics* 174, 839-50 (2006).

56. Liti, G., Peruffo, A., James, S.A., Roberts, I.N. & Louis, E.J. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the Saccharomyces sensu stricto complex. *Yeast* 22, 177-92 (2005).
57. Lodder, J. & kreger-van Rij, N.J.W. *The Yeasts, a Taxonomic Study*, 220 (Amsterdam, 1952).
58. Van Der Walt, J.P. Saccharomyces Vafer and S. Inconspicuus Spp.N. *Antonie Van Leeuwenhoek* 31, 187-92 (1965).
59. McCullough, M.J., Clemons, K.V., Farina, C., McCusker, J.H. & Stevens, D.A. Epidemiological investigation of vaginal Saccharomyces cerevisiae isolates by a genotypic method. *J Clin Microbiol* 36, 557-62 (1998).
60. Naumov, G.I., Serpova, E.V. & Naumova, E.S. A genetically isolated population of Saccharomyces cerevisiae in Malaysia. *Mikrobiologiia* 75, 245-9 (2006).
61. Fay, J.C. & Benavides, J.A. Evidence for domesticated and wild populations of Saccharomyces cerevisiae. *PLoS Genet* 1, 66-71 (2005).
62. Sniegowski, P.D., Dombrowski, P.G. & Fingerman, E. Saccharomyces cerevisiae and Saccharomyces paradoxus coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Res* 1, 299-306 (2002).
63. Martini, A.V. & Martini, A. Three newly delimited species of Saccharomyces sensu stricto. *Antonie Van Leeuwenhoek* 53, 77-84 (1987).
64. Johnson, L.J. et al. Population Genetics of the Wild Yeast Saccharomyces paradoxus. *Genetics* 166, 43-52 (2004).
65. Koufopanou, V., Hughes, J., Bell, G. & Burt, A. The spatial scale of genetic differentiation in a model organism: the wild yeast Saccharomyces paradoxus. *Philos Trans R Soc Lond B Biol Sci* (2006).
66. Naumov, G.I. Genetic-Basis for Classification and Identification of the Ascomycetous Yeasts. *Studies in Mycology*, 469-475 (1987).
67. Iurkov, A.M. First isolation of the yeast Saccharomyces paradoxus in Western Siberia. *Mikrobiologiia* 74, 533-6 (2005).
68. Naumov, G.I., Naumova, E.S. & Sniegowski, P.D. Differentiation of European and Far East Asian populations of Saccharomyces paradoxus by allozyme analysis. *Int J Syst Bacteriol* 47, 341-4 (1997).
69. Vaughan Martini, A. Saccharomyces paradoxus comb. nov., a Newly Separated Species of the Saccharomyces sensu stricto Complex Based upon nDNA/nDNA Homologies. *System. Appl. Microbiol.* 12, 179-182 (1989).
70. Naumov, G.I., Naumova, E.S., Hagler, A.N., Mendonca-Hagler, L.C. & Louis, E.J. A new genetically isolated population of the Saccharomyces sensu stricto complex from Brazil. *Antonie Van Leeuwenhoek* 67, 351-5 (1995).
71. Pope, G. Saccharomyces Genome Resequencing Project Strains. (2008). http://www.ncyc.co.uk/sgrp.html
72. http://wiki.yeastgenome.org/index.php/CommunityW303.html.
73. Bishop, D.K., Park, D., Xu, L. & Kleckner, N. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* 69, 439-56 (1992).
74. McCusker, J.H. & Haber, J.E. Cycloheximide-resistant temperature-sensitive lethal mutations of Saccharomyces cerevisiae. *Genetics* 119, 303-15 (1988).
75. Liti, G. & Louis, E.J. Yeast evolution and comparative genomics. *Annu Rev Microbiol* 59, 135-53 (2005).
76. Hong, E.L. et al. Gene Ontology annotations at SGD: new data sources and annotation methods. *Nucleic Acids Res* 36, D577-81 (2008).
77. Naumov, G.I., James, S.A., Naumova, E.S., Louis, E.J. & Roberts, I.N. Three new species in the Saccharomyces sensu stricto complex: Saccharomyces cariocanus, Saccharomyces kudriavzevii and Saccharomyces mikatae. *Int J Syst Evol Microbiol* 50 Pt 5, 1931-42 (2000).
78. Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G. & Louis, E.J. Chromosomal evolution in Saccharomyces. *Nature* 405, 451-4 (2000).
79. Kobayashi, T., Heck, D.J., Nomura, M. & Horiuchi, T. Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev* 12, 3821-30 (1998).

80. Petes, T.D. Yeast ribosomal DNA genes are located on chromosome XII. *Proc Natl Acad Sci U S A* 76, 410-4 (1979).

81. Prokopowich, C.D., Gregory, T.R. & Crease, T.J. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46, 48-50 (2003).

82. Ganley, A.R. & Kobayashi, T. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res* 17, 184-91 (2007).

83. Ohta, T. Some models of gene conversion for treating the evolution of multigene families. *Genetics* 106, 517-28 (1984).

84. Coghlan, A. & Wolfe, K.H. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. *Yeast* 16, 1131-45 (2000).

85. Sharp, P.M. & Li, W.H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15, 1281-95 (1987).

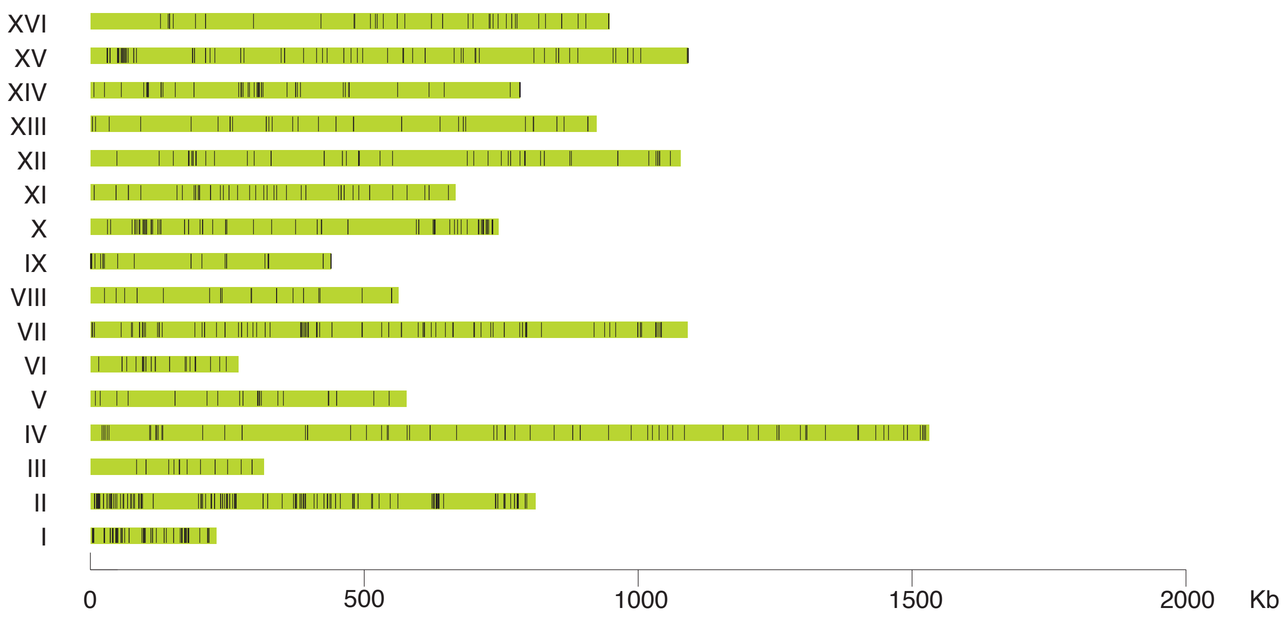
86. Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129, 897-907 (1991).

87. McDonald, J.H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351, 652-4 (1991).

88. Fay, J.C. & Wu, C.I. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genomics Hum Genet* 4, 213-35 (2003).

89. Fay, J.C., Wyckoff, G.J. & Wu, C.I. Positive and negative selection on the human genome. *Genetics* 158, 1227-34 (2001).

a



b

```
SGD       GTTATCAATACAAATCCGGGCGCCAGAACCTCAATCTTAGCGGCAGCAAATCCGTTGTATGGTAGAATT
          | ||||||||||| |  |  ||||||||||||||||||||||||||||||||||||||||||||||  |
SGRP      GGTATCAATACAACTTTGAACGCCAGAACCTCAATCTTAGCGGCAGCAAATCCGTTGTATGGTAGATAT
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
RM11      GGTATCAATACAACTTTGAACGCCAGAACCTCAATCTTAGCGGCAGCAAATCCGTTGTATGGTAGATAT
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||| || ||||||||||||
CBS432    GGTATCAATACAACTTTGAACGCCAGAACATCAATCTTAGCGGCAGCAAATCCCTTATATGGTAGATAT
```

```
SGD                 VINTNPGARTSILAAANPLYGRI
                     |||    |||||||||||||||
SGRP                GINTTLNARTSILAAANPLYGRY
                    |||||||||||||||||||||||
S. stricto          GINTTLNARTSILAAANPLYGRY
```

c

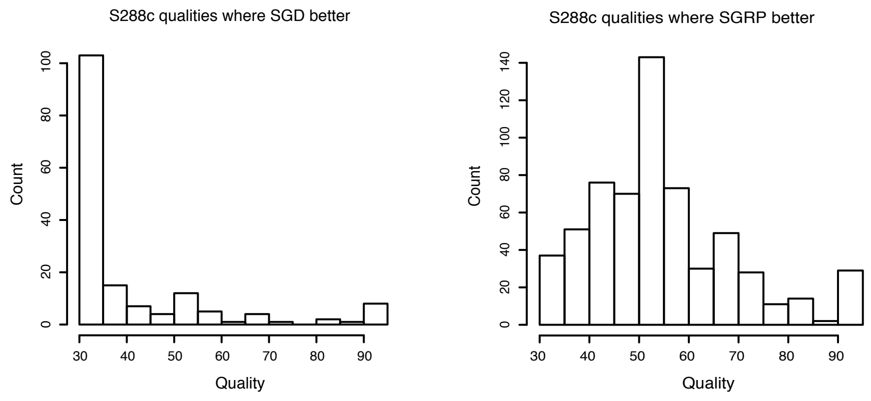


Figure S1

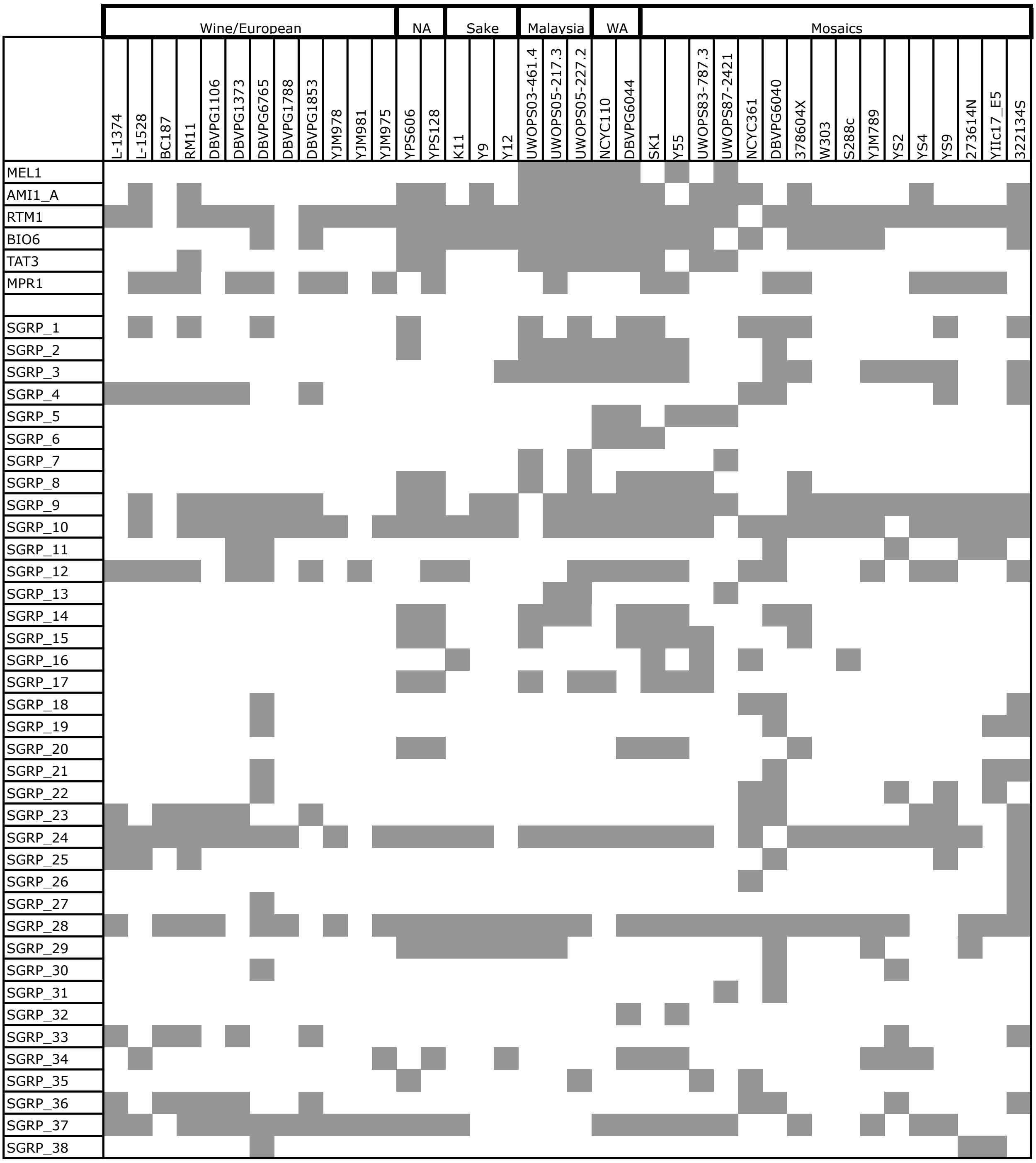

Wine/European
NA
Sake
Malaysia
WA
Mosaics
L-1374
L-1528
BC187
RM11
DBVPG1106
DBVPG1373
DBVPG6765
DBVPG1788
DBVPG1853
YJM978
YJM981
YJM975
YPS606
YPS128
K11
Y9
Y12
UWOPS03-461.4
UWOPS05-217.3
UWOPS05-227.2
NCYC110
DBVPG6044
SK1
Y55
UWOPS83-787.3
UWOPS87-2421
NCYC361
DBVPG6040
378604X
W303
S288c
YJM789
YS2
YS4
YS9
273614N
YIIc17_E5
322134S
MEL1
AMI1_A
RTM1
BIO6
TAT3
MPR1
SGRP_1
SGRP_2
SGRP_3
SGRP_4
SGRP_5
SGRP_6
SGRP_7
SGRP_8
SGRP_9
SGRP_10
SGRP_11
SGRP_12
SGRP_13
SGRP_14
SGRP_15
SGRP_16
SGRP_17
SGRP_18
SGRP_19
SGRP_20
SGRP_21
SGRP_22
SGRP_23
SGRP_24
SGRP_25
SGRP_26
SGRP_27
SGRP_28
SGRP_29
SGRP_30
SGRP_31
SGRP_32
SGRP_33
SGRP_34
SGRP_35
SGRP_36
SGRP_37
SGRP_38

Figure S2

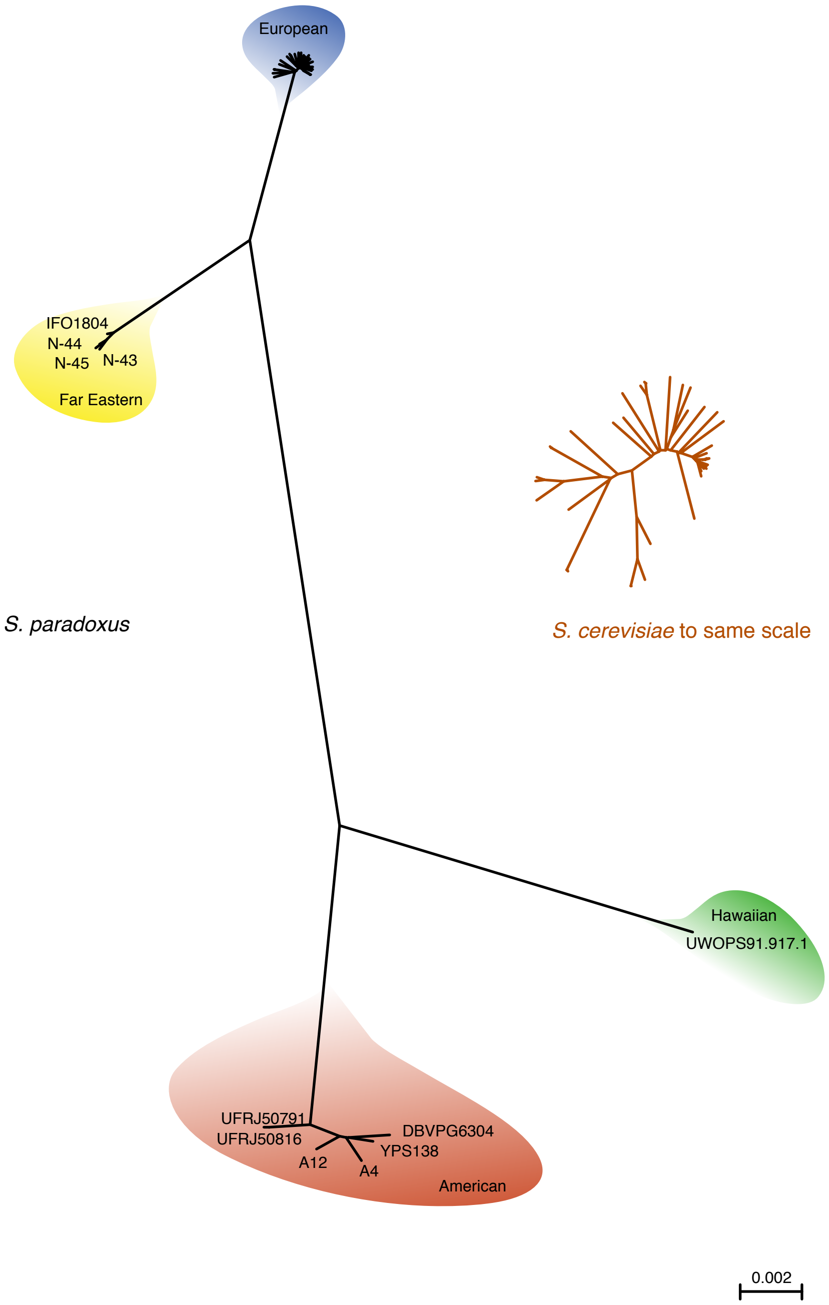

European
IFO1804
N-44
N-45
N-43
Far Eastern
S. paradoxus
S. cerevisiae to same scale
Hawaiian
UWOPS91.917.1
UFRJ50791
UFRJ50816
A12
A4
DBVPG6304
YPS138
American
0.002

Figure S3

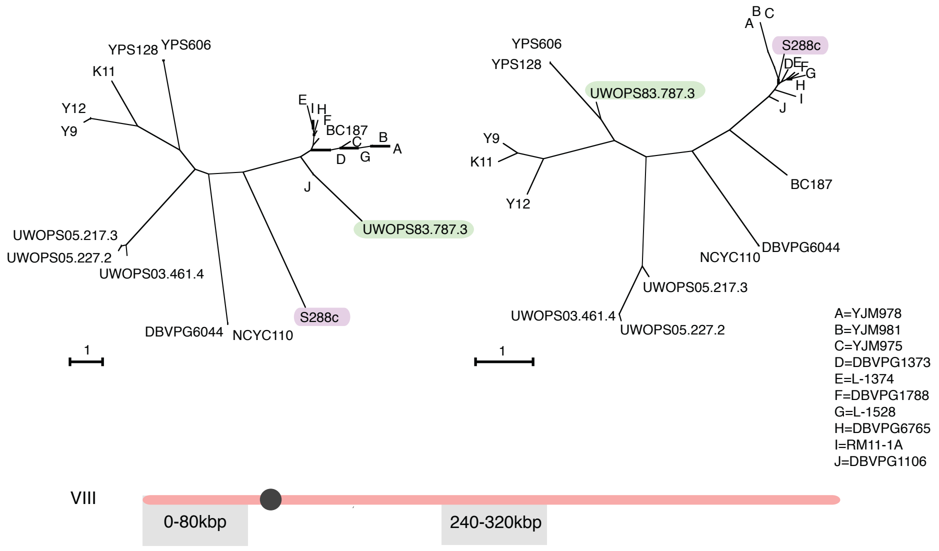

YPS128
YPS606
K11
Y12
Y9
E
I
H
F
BC187
C
B
A
D
G
J
UWOPS83.787.3
UWOPS05.217.3
UWOPS05.227.2
UWOPS03.461.4
S288c
DBVPG6044
NCYC110
1
B
C
A
S288c
YPS606
YPS128
DE
F
G
H
I
J
UWOPS83.787.3
Y9
K11
Y12
BC187
NCYC110
DBVPG6044
UWOPS05.217.3
UWOPS03.461.4
UWOPS05.227.2
1
A=YJM978
B=YJM981
C=YJM975
D=DBVPG1373
E=L-1374
F=DBVPG1788
G=L-1528
H=DBVPG6765
I=RM11-1A
J=DBVPG1106
VIII
0-80kbp
240-320kbp

Figure S4

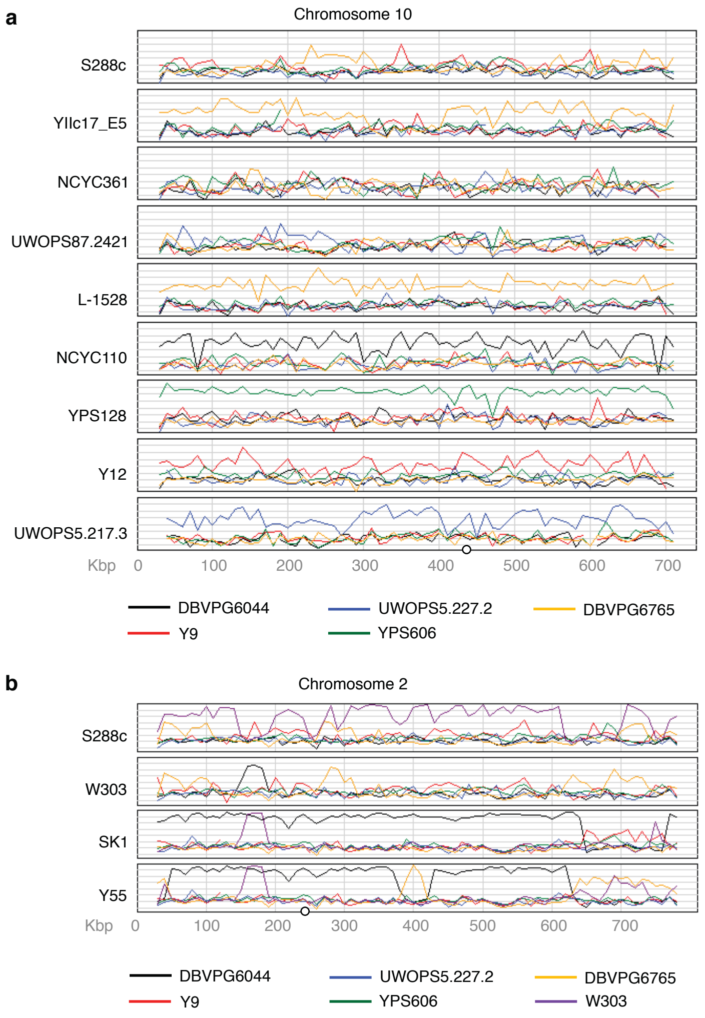

a
Chromosome 10
S288c
YIlc17_E5
NCYC361
UWOPS87.2421
L-1528
NCYC110
YPS128
Y12
UWOPS5.217.3
Kbp
0
100
200
300
400
500
600
700
DBVPG6044
UWOPS5.227.2
DBVPG6765
Y9
YPS606
b
Chromosome 2
S288c
W303
SK1
Y55
Kbp
0
100
200
300
400
500
600
700
DBVPG6044
UWOPS5.227.2
DBVPG6765
Y9
YPS606
W303

Figure S5

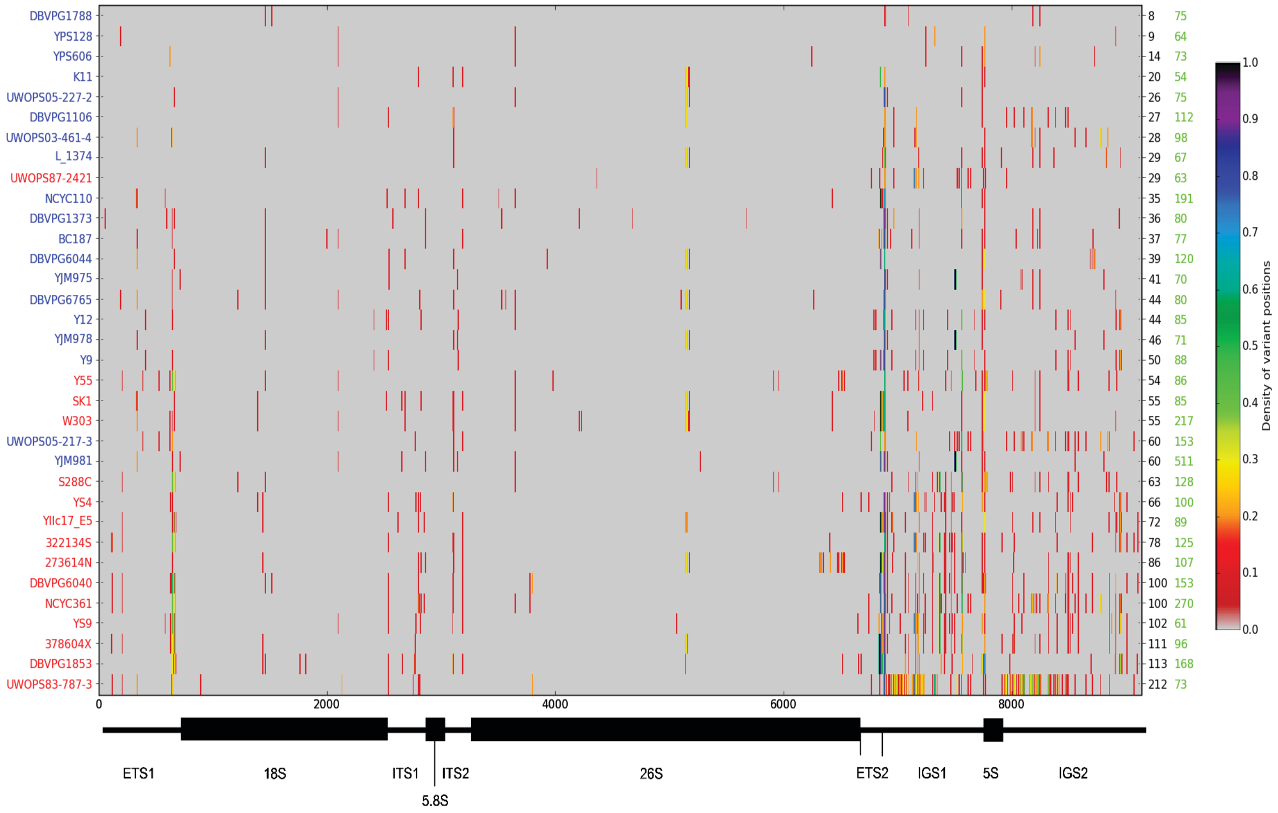

DBVPG1788
YPS128
YPS606
K11
UWOPS05-227-2
DBVPG1106
UWOPS03-461-4
L_1374
UWOPS87-2421
NCYC110
DBVPG1373
BC187
DBVPG6044
YJM975
DBVPG6765
Y12
YJM978
Y9
Y55
SK1
W303
UWOPS05-217-3
YJM981
S288C
YS4
YIIc17_E5
322134S
273614N
DBVPG6040
NCYC361
YS9
378604X
DBVPG1853
UWOPS83-787-3
8 75
9 64
14 73
20 54
26 75
27 112
28 98
29 67
29 63
35 191
36 80
37 77
39 120
41 70
44 80
44 85
46 71
50 88
54 86
55 85
55 217
60 153
60 511
63 128
66 100
72 89
78 125
86 107
100 153
100 270
102 61
111 96
113 168
212 73
0
2000
4000
6000
8000
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Density of variant positions
ETS1
18S
ITS1
ITS2
5.8S
26S
ETS2
IGS1
5S
IGS2

Figure S6

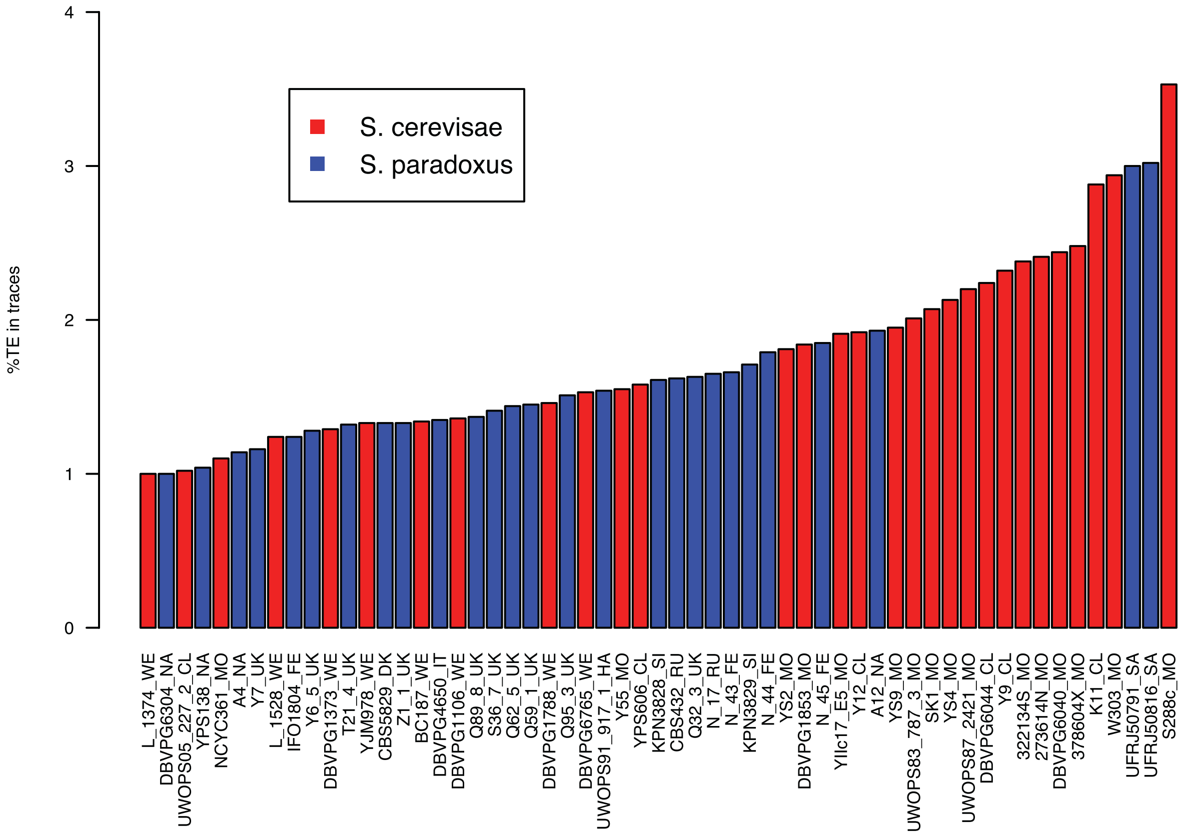

%TE in traces
0
1
2
3
4
🟥 S. cerevisae
🟦 S. paradoxus
L_1374_WE
DBVPG6304_NA
UWOPS05_227_2_CL
YPS138_NA
NCYC361_MO
A4_NA
Y7_UK
L_1528_WE
IFO1804_FE
Y6_5_UK
DBVPG1373_WE
T21_4_UK
YJM978_WE
CBS5829_DK
Z1_1_UK
BC187_WE
DBVPG4650_IT
DBVPG1106_WE
Q89_8_UK
S36_7_UK
Q62_5_UK
Q59_1_UK
DBVPG1788_WE
Q95_3_UK
DBVPG6765_WE
UWOPS91_917_1_HA
Y55_MO
YPS606_CL
KPN3828_SI
CBS432_RU
Q32_3_UK
N_17_RU
N_43_FE
KPN3829_SI
N_44_FE
YS2_MO
DBVPG1853_MO
N_45_FE
YIIc17_E5_MO
Y12_CL
A12_NA
YS9_MO
UWOPS83_787_3_MO
SK1_MO
YS4_MO
UWOPS87_2421_MO
DBVPG6044_CL
Y9_CL
322134S_MO
273614N_MO
DBVPG6040_MO
378604X_MO
K11_CL
W303_MO
UFRJ50791_SA
UFRJ50816_SA
S288c_MO

Figure S7

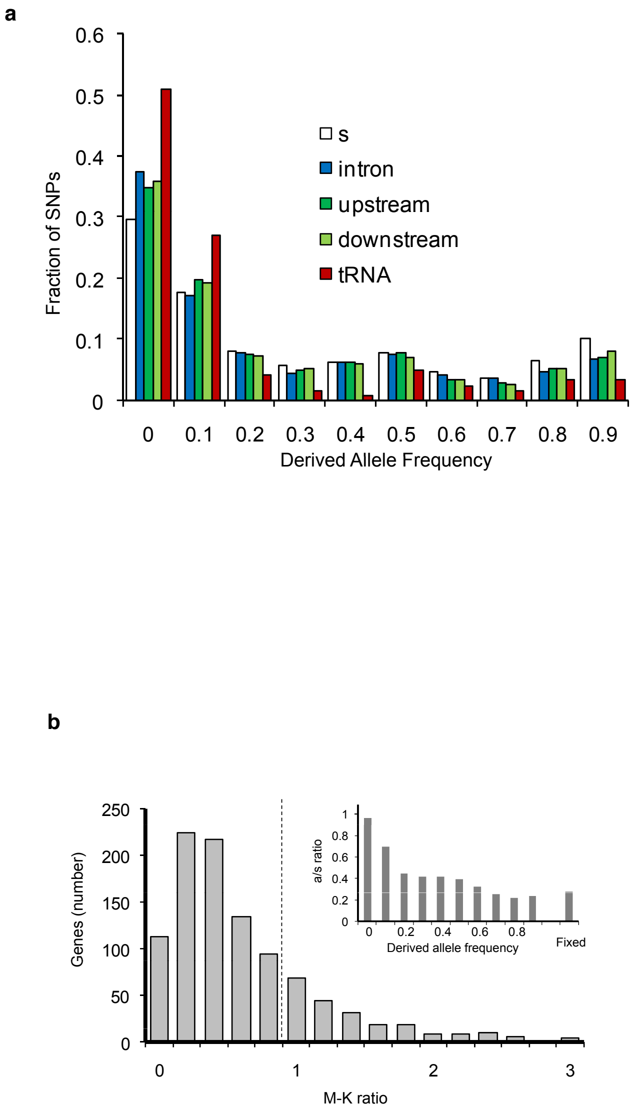

a
Fraction of SNPs
0.6
0.5
0.4
0.3
0.2
0.1
0
□ s
■ intron
■ upstream
■ downstream
■ tRNA
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Derived Allele Frequency
b
Genes (number)
250
200
150
100
50
0
0
1
2
3
M-K ratio
a/s ratio
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
Derived allele frequency
Fixed

Figure S8

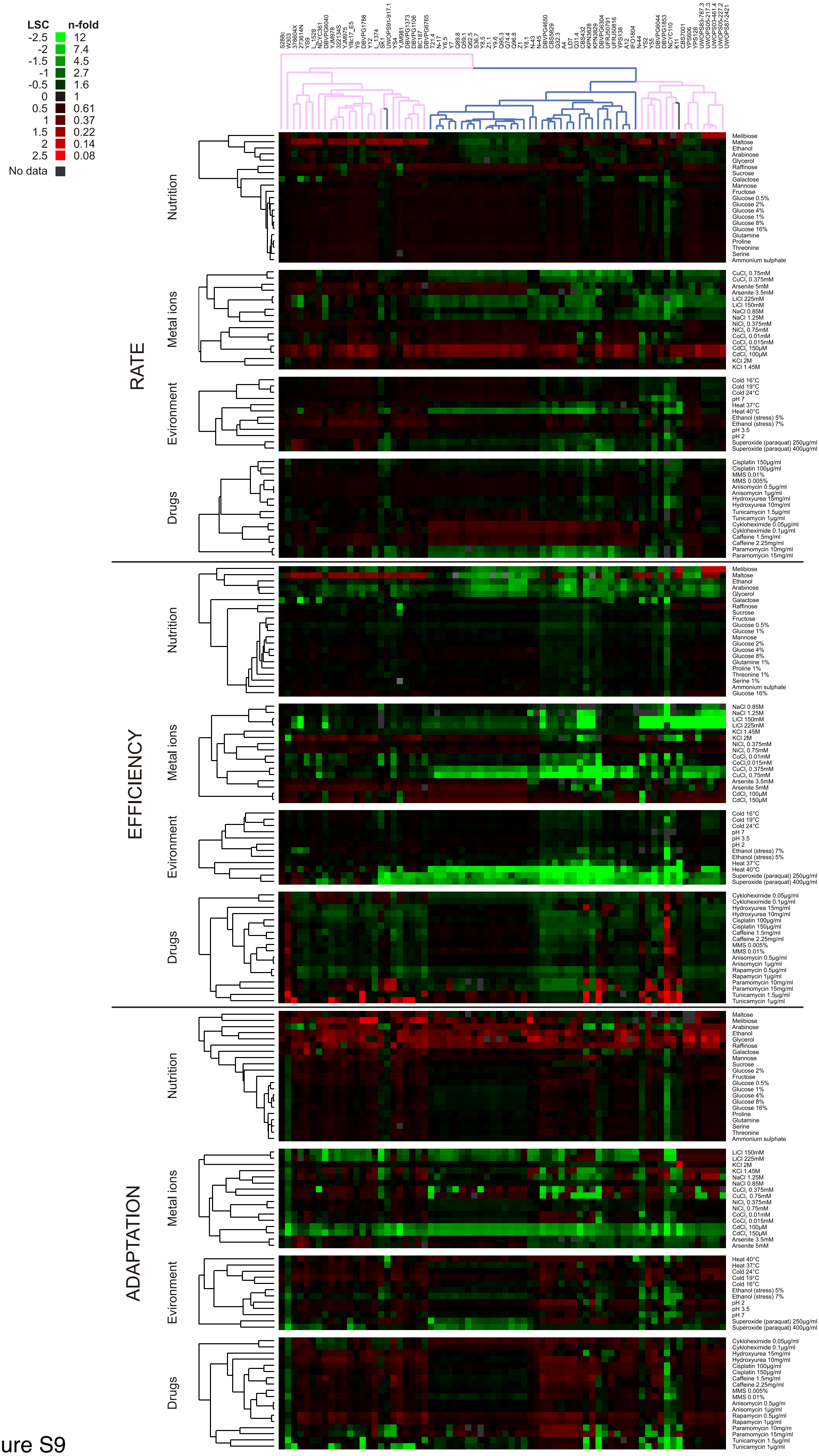

LSC n-fold
-2.5 12
-2 7.4
-1.5 4.5
-1 2.7
-0.5 1.6
0 1
0.5 0.61
1 0.37
1.5 0.22
2 0.14
2.5 0.08
No data
S288c
W303
378604X
273614N
YS9
L_1528
NCYC361
DBVPG6040
YJM978
322134S
YJM975
YIIc17_E5
Y9
DBVPG1788
Y12
L_1374
SK1
UWOPS91-917.1
YS4
YJM981
DBVPG1373
DBVPG1106
BC187
DBVPG6765
T21.4
N-17
Y6.5
Y7
Q89.8
Q59.1
Q62.5
S36.7
Y8.5
Z1.1
Y9.6
Q95.3
Q74.4
Q96.8
Z1
Y8.1
N-43
N-45
DBVPG4650
CBS5829
Q32.3
A4
LD7
Q31.4
CBS432
KPN3828
KPN3829
DBVPG6304
UFRJ50791
UFRJ50816
YPS138
A12
IFO1804
N-44
YS2
Y55
DBVPG6044
DBVPG1853
NCYC110
K11
CBS7001
YPS606
YPS128
UWOPS83-787.3
UWOPS05-217.3
UWOPS03-461.4
UWOPS05-227.2
UWOPS87-2421
RATE
Nutrition
Melibiose
Maltose
Ethanol
Arabinose
Glycerol
Raffinose
Sucrose
Galactose
Mannose
Fructose
Glucose 0.5%
Glucose 2%
Glucose 4%
Glucose 1%
Glucose 8%
Glucose 16%
Glutamine
Proline
Threonine
Serine
Ammonium sulphate
Metal ions
CuCl2 0.75mM
CuCl2 0.375mM
Arsenite 5mM
Arsenite 3.5mM
LiCl 225mM
LiCl 150mM
NaCl 0.85M
NaCl 1.25M
NiCl2 0.375mM
NiCl2 0.75mM
CoCl2 0.01mM
CoCl2 0.015mM
CdCl2 150µM
CdCl2 100µM
KCl 2M
KCl 1.45M
Evironment
Cold 16°C
Cold 19°C
Cold 24°C
pH 7
Heat 37°C
Heat 40°C
Ethanol (stress) 5%
Ethanol (stress) 7%
pH 3.5
pH 2
Superoxide (paraquat) 250µg/ml
Superoxide (paraquat) 400µg/ml
Drugs
Cisplatin 150µg/ml
Cisplatin 100µg/ml
MMS 0.01%
MMS 0.005%
Anisomycin 0.5µg/ml
Anisomycin 1µg/ml
Hydroxyurea 15mg/ml
Hydroxyurea 10mg/ml
Tunicamycin 1.5µg/ml
Tunicamycin 1µg/ml
Cykloheximide 0.05µg/ml
Cykloheximide 0.1µg/ml
Caffeine 1.5mg/ml
Caffeine 2.25mg/ml
Paramomycin 10mg/ml
Paramomycin 15mg/ml
EFFICIENCY
Nutrition
Melibiose
Maltose
Ethanol
Arabinose
Glycerol
Galactose
Raffinose
Sucrose
Fructose
Glucose 0.5%
Glucose 1%
Mannose
Glucose 2%
Glucose 4%
Glucose 8%
Glutamine 1%
Proline 1%
Threonine 1%
Serine 1%
Ammonium sulphate
Glucose 16%
Metal ions
NaCl 0.85M
NaCl 1.25M
LiCl 150mM
LiCl 225mM
KCl 1.45M
KCl 2M
NiCl2 0.375mM
NiCl2 0.75mM
CoCl2 0.01mM
CoCl2 0.015mM
CuCl2 0.375mM
CuCl2 0.75mM
Arsenite 3.5mM
Arsenite 5mM
CdCl2 100µM
CdCl2 150µM
Evironment
Cold 16°C
Cold 19°C
Cold 24°C
pH 7
pH 3.5
pH 2
Ethanol (stress) 7%
Ethanol (stress) 5%
Heat 37°C
Heat 40°C
Superoxide (paraquat) 250µg/ml
Superoxide (paraquat) 400µg/ml
Drugs
Cykloheximide 0.05µg/ml
Cykloheximide 0.1µg/ml
Hydroxyurea 15mg/ml
Hydroxyurea 10mg/ml
Cisplatin 100µg/ml
Cisplatin 150µg/ml
Caffeine 1.5mg/ml
Caffeine 2.25mg/ml
MMS 0.005%
MMS 0.01%
Anisomycin 0.5µg/ml
Anisomycin 1µg/ml
Rapamycin 0.5µg/ml
Rapamycin 1µg/ml
Paramomycin 10mg/ml
Paramomycin 15mg/ml
Tunicamycin 1.5µg/ml
Tunicamycin 1µg/ml
ADAPTATION
Nutrition
Maltose
Melibiose
Arabinose
Ethanol
Glycerol
Raffinose
Galactose
Mannose
Sucrose
Glucose 2%
Fructose
Glucose 0.5%
Glucose 1%
Glucose 4%
Glucose 8%
Glucose 16%
Proline
Glutamine
Serine
Threonine
Ammonium sulphate
Metal ions
LiCl 150mM
LiCl 225mM
KCl 2M
KCl 1.45M
NaCl 1.25M
NaCl 0.85M
CuCl2 0.375mM
CuCl2 0.75mM
NiCl2 0.375mM
NiCl2 0.75mM
CoCl2 0.01mM
CoCl2 0.015mM
CdCl2 100µM
CdCl2 150µM
Arsenite 3.5mM
Arsenite 5mM
Evironment
Heat 40°C
Heat 37°C
Cold 24°C
Cold 19°C
Cold 16°C
Ethanol (stress) 5%
Ethanol (stress) 7%
pH 2
pH 3.5
pH 7
Superoxide (paraquat) 250µg/ml
Superoxide (paraquat) 400µg/ml
Drugs
Cykloheximide 0.05µg/ml
Cykloheximide 0.1µg/ml
Hydroxyurea 15mg/ml
Hydroxyurea 10mg/ml
Cisplatin 100µg/ml
Cisplatin 150µg/ml
Caffeine 1.5mg/ml
Caffeine 2.25mg/ml
MMS 0.005%
MMS 0.01%
Anisomycin 0.5µg/m
Anisomycin 1µg/ml
Rapamycin 0.5µg/ml
Rapamycin 1µg/ml
Paramomycin 10mg/m
Paramomycin 15mg/ml
Tunicamycin 1.5µg/ml
Tunicamycin 1µg/ml

Figure S9

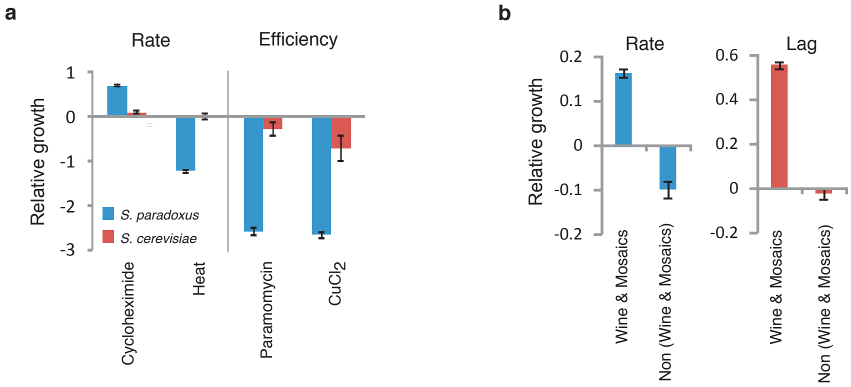

a
Rate
Efficiency
Relative growth
1
0
-1
-2
-3
🟦 *S. paradoxus*
🟥 *S. cerevisiae*
Cycloheximide
Heat
Paramomycin
$CuCl_2$
b
Rate
Lag
Relative growth
0.2
0.1
0
-0.1
-0.2
0.6
0.4
0.2
0
-0.2
Wine & Mosaics
Non (Wine & Mosaics)
Wine & Mosaics
Non (Wine & Mosaics)

Figure S10