SUPPLEMENTAL INFORMATION - MATERIALS AND METHODS

Genome Sequencing Analyses

The genomes of the JAY270 parent strain and of 56 haploids derived from 14 complete JAY270 tetrads were sequenced using the Illumina short read whole genome sequencing platform. Genome sequencing data associated with this study is available in the Sequence Read Archive (SRA) database under study number SRP082524.

JAY270 draft HetSNP map construction

We developed a map of heterozygous SNPs in JAY270 using a high stringency approach that would identify only high confidence sites. This was a conservative approach that is therefore likely to be missing some loci, but it is unlike to contain any false calls. We took the reads from 44 haploid spores from 11 complete JAY270 tetrads and applied two parallel analyses. These 11 tetrads had been sequenced at the same time and had uniform coverage and read lengths. The data from the additional 3 tetrads were not as homogeneous so they were not used for this aspect of the work.

In the first analysis we combined all haploid reads to simulate an ultra deep coverage sequencing dataset from the JAY270 diploid. The data from each haploid was first processed to include only high quality reads (Q>30), and the ends were trimmed to obtain 90 nt reads. Next we determined the haploid with the lowest number of reads within each tetrad. All the reads from this haploid and an equal number of random reads from each of its three sibling haploids were selected for the next phase. This ensured an equal number of reads contributed by each haploid within each tetrad. Finally, we combined all processed and intra-tetrad number-adjusted reads from the 44 haploids to generate the simulated JAY270 sequencing dataset. This set was composed of ~129 million reads for a mean depth coverage of ~800 reads per base. We aligned these reads to the *S. cerevisiae* S288c reference genome and independently called out SNPs using GATK (McKenna et al. 2010) and Samtools (Li et al. 2009), limiting the analysis to SNPs with coverage higher than 200 and allele frequency between 0.4 and 0.6. We then obtained a list of 13,594 candidate HetSNPs found by both approaches, all had allele frequencies close to 0.5.

For the second analysis we aligned the reads from each individual haploid to the reference genome and called the SNPs using GATK (McKenna et al. 2010), identifying 18,201 sites. Next we aligned the calls from each group of four haploids belonging to the same tetrad and determined the segregation ratio for each of the SNPs within each tetrad.

We then took the 13,075 sites that were discovered in both approaches and filtered to improve the confidence of the heterozygous calls. We retained only the sites that had a Mendelian 2:2 segregation in at least 9 of the 11 tetrads for sites located in central regions of chromosomes (defined by the first and last genes annotated as essential in SGD). We used a stricter filter of all 11 tetrads displaying 2:2 for sites located at distal regions to avoid confounding effects from sites present at subtelomeric repeated gene families. The segregation filtering resulted in 12,197 sites, which we then narrowed down manually (mostly by removing subtelomeric sites) to arrive at the final list of 12,023 high confidence HetSNPs shown in Fig. S1. Note that this draft list is limited only to allelic sites that have one nucleotide that matches the S288c reference and the other that is a variant present in JAY270. It does not contain sites in which two nucleotide variants are present at the same site, nor short nucleotide insertion or deletions (with the exception of *ace2-A7*; Chr12_405,714), nor larger structural variants. A comprehensive high-quality and phased genome assembly of the JAY270 diploid genome will be described elsewhere.

HetSNP phasing.

The availability of genome sequencing data from multiple complete tetrads allowed us to deduce the phasing association between the JAY270 HetSNPs. To do so, we initially arbitrarily assigned the S288c reference bases to one phased haplotype and all alternative bases to the other haplotype. Next we aligned the genotypes of 56 haploids along the HetSNP list and determined the positions of crossover events between the two haplotypes within each of the respective 14 tetrads. Considering that meiotic crossovers are very unlikely to occur at exactly the same position in different tetrads, we could make corrections to the arbitrary phasing to minimize the number of crossovers. In most cases, apparent 4-chromatid double crossovers were observed at

the same interval in all tetrads, indicating an error in the arbitrary phasing. A simple correction at those sites resulted in the much more likely scenario of no crossovers at that interval in any of tetrads. In cases where actual crossovers occurred in one or a few of the tetrads, they typically were 2-chromatid single crossover events that we could also clearly identify and correct the arbitrary phasing accordingly to minimize the number of crossovers. We did the analysis and phasing corrections manually over three sequential iterations arriving at the phased HetSNP list shown in Fig. S1. For most chromosomes where the physical distance between consecutive HetSNPs was short we were able to unambiguously deduce a single linkage group. In a few cases, either at long intervals delimited by distant consecutive HetSNPs, or sites of possible meiotic recombination hot spots, there was ambiguity in the phase calling, so we broke down the respective chromosome in multiple linkage groups. Overall for the 15 chromosomes with heterozygosity in them, 10 yielded a single linkage group, 4 had two linkage groups, and 1 had three linkage groups. No phasing could be done for Chr01 since it was fully homozygous. Once the phasing was completed and the haplotypes were defined, we arbitrarily named one of them maternal (M) and the other paternal (P) to facilitate the subsequent LOH tract analyses.

SUPPLEMENTAL INFORMATION - REFERENCES

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16):2078-2079.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis ToolKit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20 (9):1297-1303.