

# Genotype to Phenotype: A Complex Problem

Dowell, Ryan, et. al.

## Supporting Online Material

<b>Materials and Methods</b>	<b>2</b>
Genome . . . . .	2
Deletion Library . . . . .	3
<b>Supplemental Figures</b>	<b>7</b>
SFigure 1: Nucleotide differences between S288c and $\Sigma$ 1278b . . . .	7
SFigure 2: Summary of chromosomal comparison between S288c and $\Sigma$ 1278b . . . . .	8
<b>Supplemental Tables</b>	<b>12</b>
STable 2: Tetrad confirmation of strain specific essentials . . . . .	12
STable 3: Strains utilized . . . . .	24
STable 3: Hybrid tetrad analysis of $\Sigma$ 1278b specific essentials . . . .	25
<b>Data Files</b>	<b>26</b>
STable 1: Annotation of open reading frames and noncoding RNAs in $\Sigma$ 1278b . . . . .	26
STable 2: Heterozygous deletion collection for $\Sigma$ 1278b . . . . .	26
<b>Supplemental References</b>	<b>27</b>

Data files S1 and S2 are available at  
<http://mcdb.colorado.edu/labs1/dowelllab/pubs/DowellRyan/>

## Genome

Yeast strains used are listed in Table S2. Yeast cultures were grown as described (S1). Real-time PCR utilized the ABI 7500 system (Applied Biosystems, Foster City, CA) and was carried out with the appropriate enzymes and chemicals from Applied Biosystems as recommended by the supplier. CHEF chromosome separation was performed with a BioRad CHEF-DRII (BioRad, Hercules, CA) with the protocol supplied by BioRad.

**Sequence and Assembly** We produced whole genome shotgun sequence from two plasmid libraries (4kb and 10kb inserts) of the *Saccharomyces cerevisiae* strain  $\Sigma$ 1278b, sub-strain 10560-6B. Genomic DNA was isolated with the Qiagen Genomic-tip kit (Qiagen, Valencia, CA) following the manufacturers' protocol. Initial sequence was generated with the whole genome sequencing and assembly methodology utilized to sequence the RM11-1a strain (S2). The resulting 7.3X Arachne long read assembly contained 12.2 Mb in 111636 sequence reads, 357 contigs and 51 scaffolds.

In addition 20 million 36 nucleotide reads were generated using an Illumina Genome Analyzer located at the Whitehead Institute Genome Technology Core. Samples for Illumina sequencing were purified with the standard protocols outlined in their genomic DNA sample prep kit (Illumina, San Diego, CA). Three lanes of cluster generations were performed on an Illumina cluster station with 2pM sequencing libraries for each lane. These reads were assembled with Velvet (S3) (v0.6.03) with a coverage cutoff of 5 and a minimum length of 100 nts, resulting in 11.3 Mb in 5419 contigs.

The BlastZ (v7) (S4) and MUMer (v3.19) (S5) software packages were utilized to align the long read scaffolds to the S288c chromosomes [Saccharomyces Genome Database (SGD) March 2009; <http://www.yeastgenome.org/>]. Contour-clamped homogeneous electric field (CHEF) gel electrophoresis and site specific PCR were used to correct the misassembly of four scaffolds and ascertain their location, size, and boundaries. In addition, one scaffold had no clear S288c correspondence and was localized in  $\Sigma$ 1278b by CHEF gel. The short read scaffolds were then utilized to fill in gaps and correct poor quality segments within the chromosomes with a combination of BLAT (Nov 2006 (S6)), fsa (v1.07; (S7)), and manual inspection.

**Annotation** We used three methods to identify potential ORFs in the  $\Sigma 1278b$  sequence: (i) directly mapping S288c ORFs, (ii) identification of long open reading frames, and (iii) the genefinder GlimmerHMM (S8, S9). The S288c ORFs were mapped to  $\Sigma 1278b$  by identifying the best BLAT (Nov 2006 (S6)) hit utilizing the complete set of ORFs obtained from the *Saccharomyces* database (SGD: <http://www.yeastgenome.org/>; March 2009). GlimmerHMM (S8, S9) was trained on the non-mitochondrial S288c ORFs.

We identified the S288c orthologs within the  $\Sigma 1278b$  sequence by a combination of sequence identity and appropriate synteny (S10). The remaining potential  $\Sigma 1278b$  ORFs were compared to the non-redundant database (NCBI May 2008) by WU-Blast (v2.0; <http://blast.wustl.edu/>) to identify previously characterized genes not present in S288c. The gene names for the  $\Sigma 1278b$  genes with S288c homologs were annotated according to their S288c counterpart. The annotation of  $\Sigma 1278b$  genes absent from S288c is from a comparison to the non-redundant database (NCBI May 2008). Functional annotations, in particular GO associations, were taken from the S288c counterpart.

Noncoding RNAs were annotated by a combination of methods. tRNAscanSE v1.23 (S11) identified tRNAs within the  $\Sigma 1278b$  genome. Other RNAs features were identified by BLAT (Nov 2006 (S6)) from the S288c counterpart, taking into consideration synteny with surrounding ORF annotations.

The majority of differences between S288c and  $\Sigma 1278b$  excluding subtelomeric regions, were single nucleotide polymorphisms (SNPs) and small insertions or deletions (indels) distributed throughout the chromosomes. The  $\Sigma 1278b$  strain has an average SNP density of 3.2 SNPs per kilobase, as determined by alignments generated by fsa (v1.07; (S7)). Sequence comparison did not uncover any obvious duplication of genes essential in S288c.

## Deletion Library Construction

Deletion cassettes were PCR amplified such that they were flanked by 100-250 base pairs of S288c homology for each cassette. Primers were designed with Primer3 software (S12) with parameters set between 100-300bp beyond the START and STOP codons of each S288c open reading frame, with comparable melting temperatures and GC content.

Deletion cassettes were colony PCR amplified with Hi Fidelity Enzyme (Roche, Nutley, NJ) for 40 cycles. Each deletion cassette contains the kanamycin (KanMX) marker flanked by molecular barcodes and their common primers (S13) and thus each deletion allele and its corresponding molecular barcodes were transferred from the S288c deletion mutant collection to the  $\Sigma$ 1278b deletion mutant collection.

Forty-seven(47)  $\Sigma$ 1278b specific genes were deleted from the  $\Sigma$ 1278b genome. Primers were designed on the basis of the  $\Sigma$ 1278b genome sequence, with 50 base pair 5' tails of homology to the upstream and downstream of each specific gene to be deleted. Two unique molecular barcodes were assigned to each deletion mutant. PCR products (KanMX cassette + homologous DNA) were transformed with lithium acetate based transformation into strain YSWT3.

Transformants were recovered for 4 hours in YEPD liquid and then plated onto YEPD containing 200mg/ml G418. Colonies derived from a single transformation event were colony PCR confirmed by with primers that lie >350 base pairs upstream from START (KanMX internal primer sequence 5'-TCTGCAGCGAGGAGCCGTAAT-3').

### **Identifying Essential Genes by Random Spore Analysis (RSA)**

Haploid mutant strains were isolated by sporulating the diploid heterozygous deletion mutants for 4 days on solid sporulation medium. *MATa* meiotic progeny were germinated on haploid selection medium, SD-HIS/ARG/LYS+canavanine+thialysine (S14), which is minimal medium lacking histidine, arginine and lysine but included the toxic amino acid analogs canavanine and thialysine, which provides a counter-selection against heterozygous diploids. The lack of histidine selects for cells expressing STE2pr-sphis5, a construct that places the *S. pombe* his5 gene under the control of the *MATa* specific STE2 promoter. Following their germination, essential genes were identified by replica plating the haploid meiotic progeny from haploid selection medium to YEPD+G418, which selects for growth of haploid deletion mutant cells. Essential gene function was identified by the absence of viable colonies on the YEPD+G418 plate.

**Tetrad Confirmation** For genes determined to be essential for viability by RSA in the  $\Sigma$ 1278b background but non-essential for viability in S288c background (as defined by SGD), or genes determined to be non-essential for viability in the  $\Sigma$ 1278b background but essential for viability in the

S288c background, tetrad analysis was performed. Heterozygous diploid mutant strains of both backgrounds, S288c and  $\Sigma$ 1278b, were sporulated on solid sporulation medium for 1-2 weeks. Asci were digested with Zymolyase and tetrads were dissected onto YEPD and grown 4 days at 30°C. Plates were photographed and replica plated on YEPD+G418 to follow the segregation patterns of knockout alleles relative to fitness phenotype (See Table SS1). GO enrichments were calculated using SGD's GO Term Finder.

**Hybrid S288c/ $\Sigma$ 1278b Tetrad Dissection** A hybrid wild-type diploid strain (Y12868) was created crossing S288c *MATa* (Y1239) to  $\Sigma$ 1278b *MAT $\alpha$*  (Y3295) and zygotes were isolated with a tetrad dissecting microscope. To determine the naturally occurring synthetic lethality rate between wild type S288c and wild type  $\Sigma$ 1278b, 129 tetrads were dissected identifying 504 meiotic segregants, of which 6 failed to germinate (1.19% lethality). Hybrid mutant strains were created by crossing the *MATa* deletion mutants from the S288c collection to Y3295. Diploids were sporulated for 5 days and tetrads dissected on YEPD plates. Tetrad segregation pattern were tabulated for each hybrid deletion mutant. A chi-squared statistic ( $\chi^2$ ) was then utilized to test three separate hypothesis: (1) a single unlinked modifier explains the inheritance patterns (1:1:4 ratio expected); (2) three unlinked modifiers explain the inheritance patterns; and (3) complex genetics (many loci) make the inheritance patterns indistinguishable from empirically observed background, from the wild type vs wild type cross (Y1239 diploid). In all cases, a p-value was calculated for the  $\chi^2$  statistic using Microsoft Excel's CHIDIST function (Figure 4; Table SS3).

Theoretical population genetics suggests that loss-of-function mutations are predicted to accumulate to high levels in a population for genes with a single and closely linked synthetic lethal partner because linkage prevents clearance of these mutations through mating and meiotic recombination (S15). To test for the possibility that the segregation patterns we observed are caused by the tight linkage of a conditional essential gene to a single second gene that causes its lethal phenotype, we examined two conditional essential genes in greater detail. We transformed Y12868 with either *mto1 $\Delta$ ::KanMX* or *pep1 $\Delta$ ::KanMX* and then sporulated the resultant heterozygous deletion mutant. Sequencing proximal to the integrated deletion cassette allowed us to determine the parental locus (S288c or  $\Sigma$ 1278b) into which the deletion cassette

integrated. If a synthetic lethal partner present only in  $\Sigma 1278b$  were tightly linked to the conditionally essential gene, equivalent to having a single tightly linked suppressor in the S288c genetic background, then the deletion allele integrated into the  $\Sigma 1278b$  chromosome should yield only inviable progeny, whereas the deletion allele integrated into the S288c chromosome should yield only viable spores. Between 60 and 70 tetrads were dissected for each mutant and segregation patterns of lethality and G418-resistance were scored. For both MTO1 and PEP12, we found that the deletion alleles integrated into both S288c and S1278b generated relatively few inviable spores. These data, along with the segregation of inviability, show that conditional essentiality is most often a consequence of complex synthetic lethality.

## Supplemental Figures:

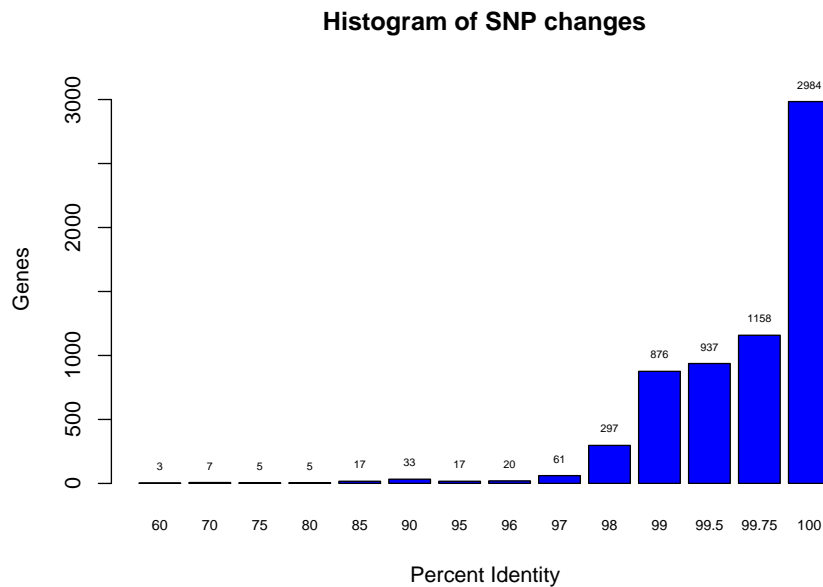


Figure S1: **Histogram showing nucleotide differences between ORFs in S288c and  $\Sigma$ 1278b** Of those genes less than 90% identical, nearly one half are contained within subtelomeric regions. The percent identity labels indicate the lower range of the bar, with the 100 % bin containing only those genes that are absolutely identical between the strains (no SNPs or indels). Bins are chosen to emphasize the fact that 94% of all genes are 99% identical or better. Genes containing N's in the  $\Sigma$ 1278b genome are excluded. Pairwise percent identity is calculated as the number of identical nucleotides divided by length of the shorter sequence on alignments generated by ClustalW (v1.83 (S16)).

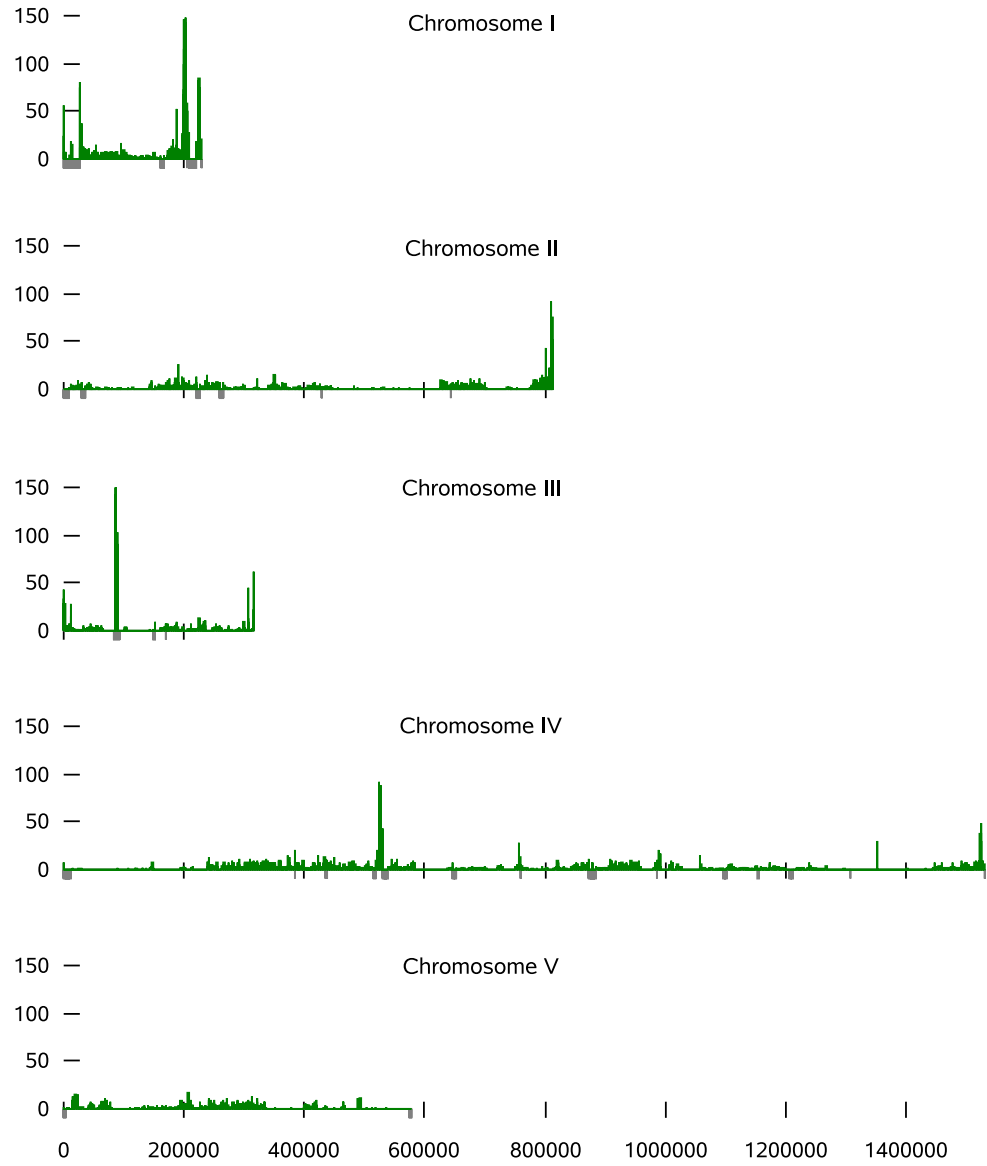


Figure S2: Continued next page



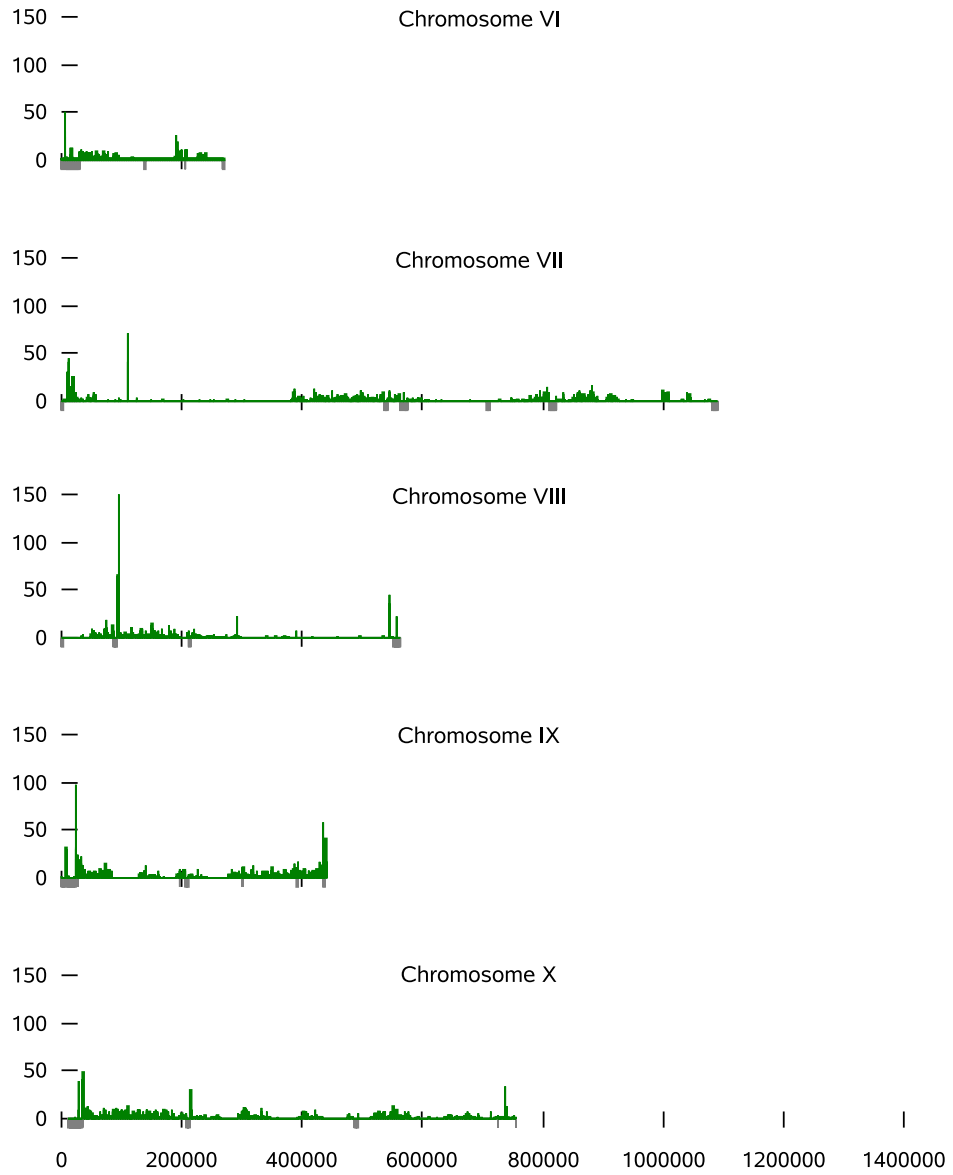


Figure S2: Continued next page

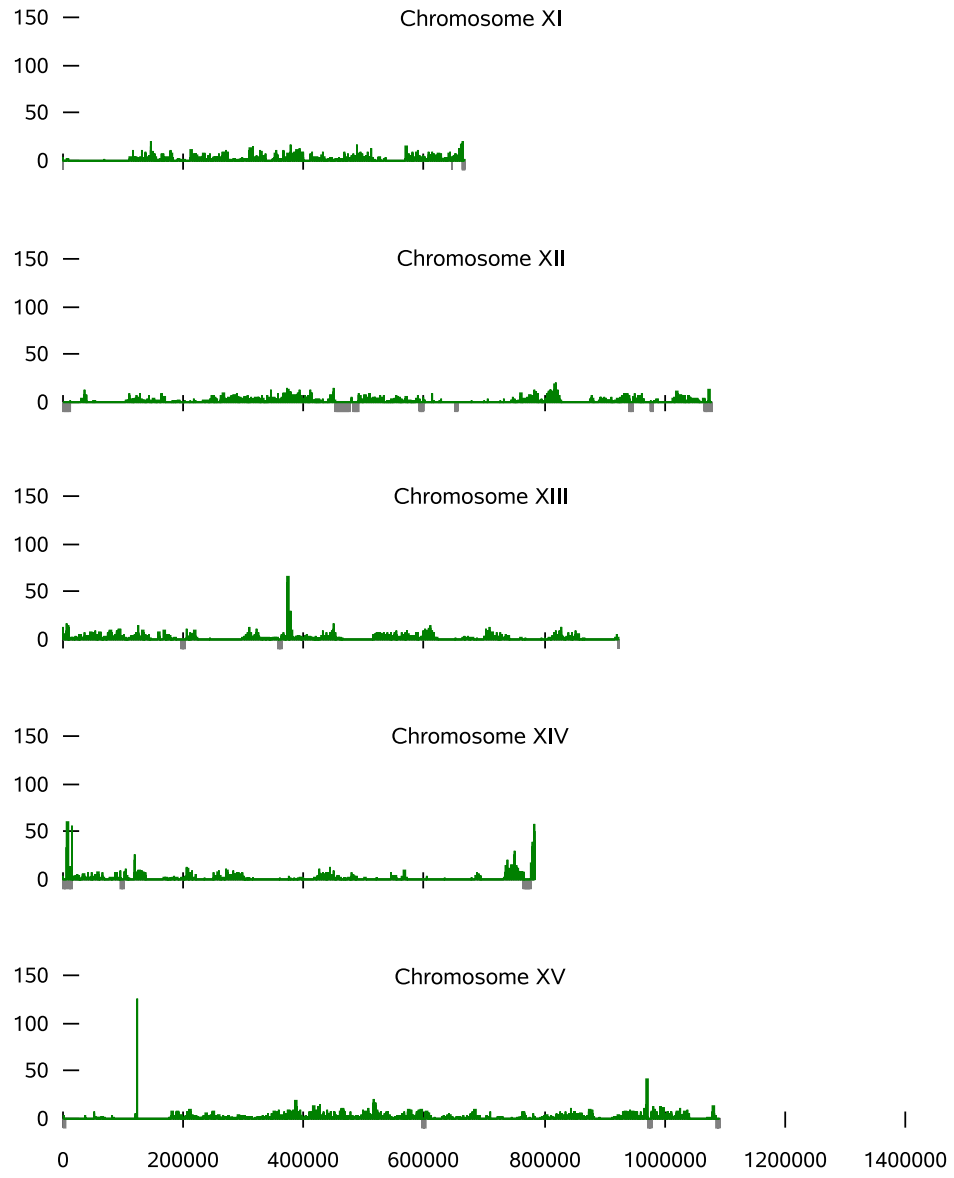


Figure S2: Continued next page

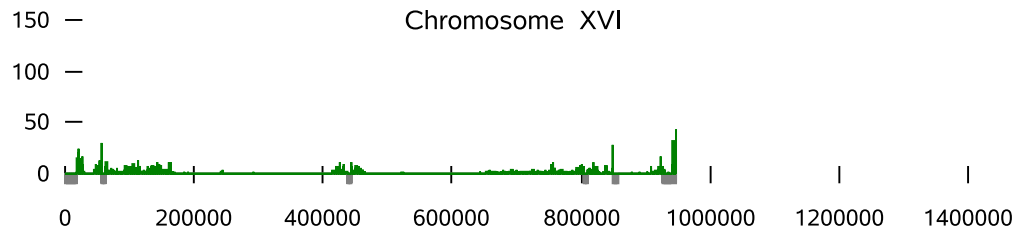


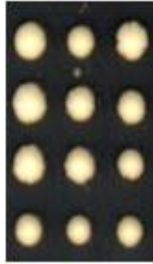







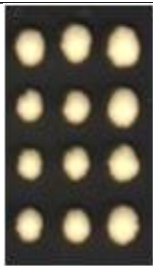

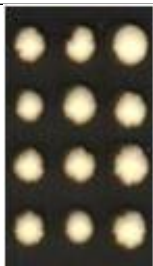

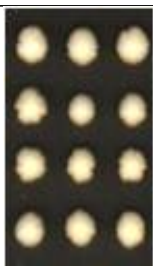







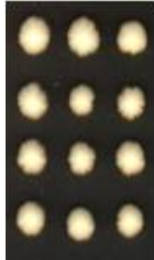





Figure S2: **Chromosomal SNP comparison between S288c and  $\Sigma$ 1278b** Graphs comparing S288C (x-axis) chromosomes to  $\Sigma$ 1278b. The Y-axis indicates the number of SNPs per window, with rolling windows of length 500, on alignments generated by fsa (v1.07 (S7)). Regions of large structural differences (insertions, deletions, and translocations) are indicated by dark grey boxes below the zero axis.








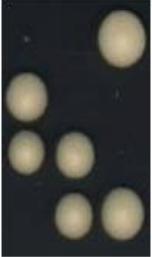


Table S1: **Tetrad confirmation of strain specific essentials.** Conditional essentials were defined by tetrads in which both deletion bearing spores failed to germinate after 4 days in one strain, but both germinated when made in the other background. The suppression of the lethal ranged from excellent (growth indistinguishable on YPD from wild type) to partial.

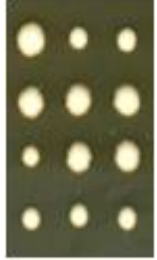

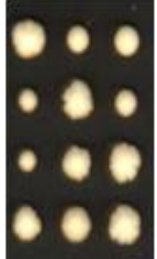



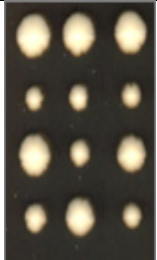



$\Sigma$ 1278b Specific Essentials:

Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>aat2</i> $\Delta$		
<i>bem1</i> $\Delta$		
<i>fmp27</i> $\Delta$		

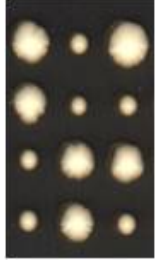

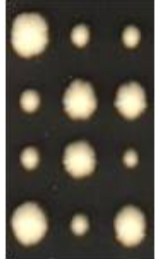







Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>lea1</i> $\Delta$		
<i>mcm22</i> $\Delta$		
<i>mto1</i> $\Delta$		
<i>pep12</i> $\Delta$		
<i>pep7</i> $\Delta$		






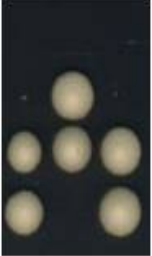

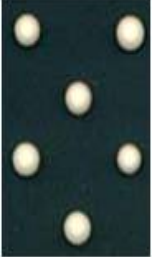


Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>sbh2</i> $\Delta$		
<i>ski7</i> $\Delta$		
<i>ski8</i> $\Delta$		
<i>vps16</i> $\Delta$		
<i>ydl089w</i> $\Delta$		








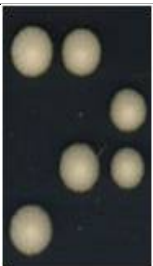


Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>yhr009c</i> $\Delta$		
<i>ykr075c</i> $\Delta$		
<i>ypr015c</i> $\Delta$		
<i>zuf1</i> $\Delta$		
<i>cys3</i> $\Delta$		


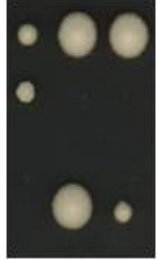








Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>cys4</i> $\Delta$		
<i>rps10a</i> $\Delta$		
<i>npl3</i> $\Delta$		
<i>lsm6</i> $\Delta$		
<i>pho88</i> $\Delta$		





Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>pho90</i> $\Delta$		
<i>adk1</i> $\Delta$		
<i>arp5</i> $\Delta$		
<i>ies6</i> $\Delta$		
<i>ost4</i> $\Delta$		


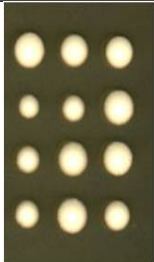

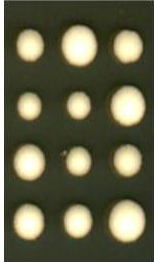
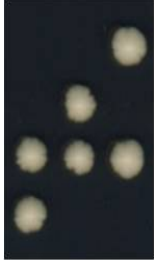
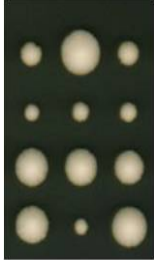
Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>snt309</i> $\Delta$		
<i>ydr241w</i> $\Delta$		
<i>lsm7</i> $\Delta$		
<i>swi6</i> $\Delta$		
<i>tma108</i> $\Delta$		


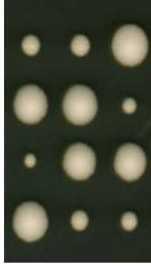
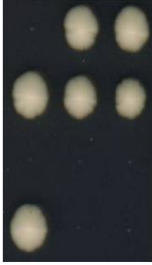
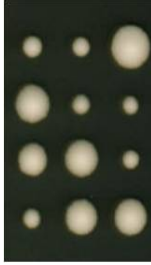
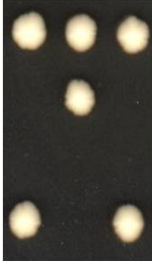
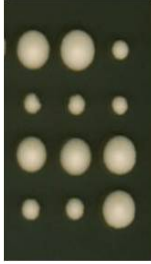


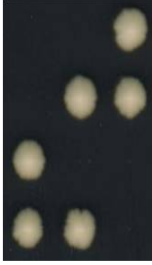
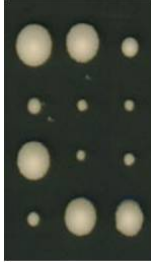
Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>vps34</i> $\Delta$		
<i>vps75</i> $\Delta$		
<i>uaf30</i> $\Delta$		
<i>pop2</i> $\Delta$		
<i>ctk1</i> $\Delta$		

Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>cyc8</i> $\Delta$		
<i>gon7</i> $\Delta$		
<i>cdc40</i> $\Delta$		
<i>cgr1</i> $\Delta$		
<i>rom2</i> $\Delta$		

Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>utr1</i> $\Delta$		

S288c Specific Essentials:

Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>plp2</i> $\Delta$		
<i>ret2</i> $\Delta$		
<i>pfy1</i> $\Delta$		

Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>rho3</i> $\Delta$		
<i>srp101</i> $\Delta$		
<i>srp102</i> $\Delta$		
<i>srp21</i> $\Delta$		
<i>srp14</i> $\Delta$		


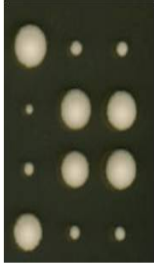

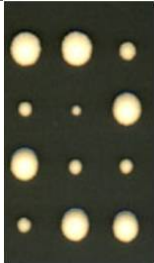
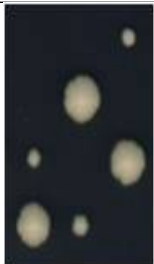



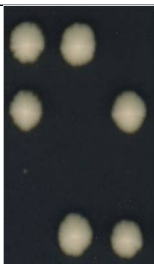
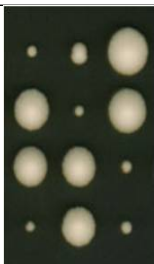
Gene	S288c Tetrads	$\Sigma$ 1278b Tetrads
<i>srp72</i> $\Delta$		
<i>uso1</i> $\Delta$		
<i>yml6</i> $\Delta$		
<i>srp68</i> $\Delta$		
<i>ubc1</i> $\Delta$		

Table S2: Strains utilized

Strain	Genotype	Background	Reference
10560-6B	<i>MAT<math>\alpha</math> ura3-52 trp1::hisG leu2::hisG his3::hisG</i>	$\Sigma$ 1278b	Fink lab strain collection
YSWT3	<i>MAT<math>\alpha</math>/<math>\alpha</math> can1<math>\Delta</math>::<i>STE2pr-sphis5/CAN1</i> lyp1<math>\Delta</math>::<i>STE3pr-LEU2/LYP1</i> his3::hisG his3::hisG leu2<math>\Delta</math>/leu2<math>\Delta</math> ura3<math>\Delta</math>/ura3<math>\Delta</math></i>	$\Sigma$ 1278b	this study
Y1239	<i>MAT<math>\alpha</math> his3<math>\Delta</math>1 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0 met15<math>\Delta</math>0</i>	S288c	Rosetta strain BY4741a
Y3295	<i>MAT<math>\alpha</math> ura3<math>\Delta</math> leu2<math>\Delta</math> his3::hisG</i>	$\Sigma$ 1278b	Microbia strain MT1562
Y12868	<i>MAT<math>\alpha</math>/MAT<math>\alpha</math> his3<math>\Delta</math>1/his3::hisG leu2<math>\Delta</math>0/leu2<math>\Delta</math>0 ura3<math>\Delta</math>0/ura3<math>\Delta</math>0 met15<math>\Delta</math>0/MET15</i>	S288c x $\Sigma$ 1278b	this study



**Table S3: Hybrid tetrad analysis of 18  $\Sigma$ 1278b specific essentials**

Hybrid mutant strains were created by crossing the *MATa* deletion mutants from the S288c collection to  $\Sigma$ 1278b wild type (Y3295). Tetrads were dissected and scored for segregation patterns (parental ditype 2:2; nonparental ditype 4:0; and tetratypes 3:1). A chi-squared p-value was then determined to test three hypothesis: (1) a single unlinked modifier explains the inheritance patterns (single gene p-value) when a 1:1:4 ratio is anticipated; (2) three unlinked modifiers explain the inheritance patterns (three gene p-value) when a 1:163:53 ratio is anticipated; (3) that complex genetics (multiple loci) make the inheritance patterns indistinguishable from the empirically observed background (wild type p-value). All 18 cases reject the null hypothesis (p-value < 0.01) of a single gene modifier. The null hypothesis of three modifiers is rejected by three genes (LEA1, FMP27, and YPR015C). The null hypothesis of inheritance indistinguishable from background is rejected in 5 cases. Finally, the observed wild-type frequencies also reject the single and three gene hypothesis.

Gene	Total Tetrads	Parental ditype (2:2)	Nonparental ditype (4:0)	Tetratype (3:1)	single gene p-value	three gene p-value	wild type p-value
<i>bem1</i> $\Delta$	116	32	22	62	2E-3	0	1E-191
<i>ski7</i> $\Delta$	89	26	21	42	3E-4	0	8E-133
<i>lea1</i> $\Delta$	62	1	50	11	1E-40	0.22	1E-4
<i>fmp27</i> $\Delta$	59	1	49	9	2E-41	0.12	3E-3
<i>ypr015c</i> $\Delta$	49	1	41	7	3E-35	0.08	0.02
<i>sbh2</i> $\Delta$	61	5	47	9	1E-35	3E-18	3E-5
<i>pep7</i> $\Delta$	64	4	53	7	1E-44	7E-12	0.01
<i>tma108</i> $\Delta$	61	3	50	8	2E-41	4E-7	0.01
<i>zwf1</i> $\Delta$	66	3	55	8	1E-46	7E-7	0.02
<i>mto1</i> $\Delta$	89	5	77	7	9E-69	3E-14	0.07
<i>ski8</i> $\Delta$	61	2	53	6	9E-48	3E-4	0.27
<i>ykr075c</i> $\Delta$	59	0	57	2	1E-59	8E-4	0.38
<i>yhr009c</i> $\Delta$	61	2	56	3	1E-54	2E-5	0.27
<i>aat2</i> $\Delta$	58	2	51	5	8E-47	1E-4	0.47
<i>ydl089w</i> $\Delta$	60	2	53	5	6E-49	1E-4	0.52
<i>pep12</i> $\Delta$	47	1	45	1	1E-46	6E-4	0.60
<i>vps16</i> $\Delta$	63	1	60	2	1E-61	3E-4	0.67
<i>mcm22</i> $\Delta$	72	1	67	4	2E-66	7E-4	0.87
wild type	129	3	119	7	5E-116	6E-8	-

All data files can be downloaded from  
<http://mcdb.colorado.edu/labs1/dowelllab/pubs/DowellRyan/>

**Data File S1: Annotation of open reading frames and noncoding RNAs in  $\Sigma 1278b$**

The file contains the  $\Sigma 1278b$  annotation in tab-delimited format with 9 columns: orfname, gene name, chromosome, strand, start, end, number of exons, exon starts (separated by commas), exon ends (separated by commas). The orfname utilizes the S288c ortholog when available and otherwise a  $\Sigma 1278b$  specific systematic name.

**Data File S2: Heterozygous deletion collection for  $\Sigma 1278b$**

The file contains the  $\Sigma 1278b$  deletion collection in tab-delimited format with 7 columns: index, orfname, gene name, set identifier, row and column location, uptag sequence, andowntag sequence. Tag sequences are given as 5' to 3'.

## Supplemental References

- S1. F. Sherman, G. Fink, J. Hicks, *Methods in yeast genetics*, vol. 263 (Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, 1991).
- S2. *Saccharomyces cerevisiae* RM11-1a sequencing project, [www.broad.mit.edu](http://www.broad.mit.edu).
- S3. D. Zerbino, E. Birney, *Genome Research* **18** (2008).
- S4. M. Blanchette, et al., *Genome research* **14**, 708 (2004).
- S5. A. Delcher, A. Phillippy, J. Carlton, S. Salzberg, *Nucleic acids research* **30**, 2478 (2002).
- S6. W. Kent, *Genome Research* pp. GR-2292R (2002).
- S7. R. K. Bradley, et al., *PLoS Comput Biol* **5**, e1000392+ (2009).
- S8. S. Salzberg, M. Pertea, A. Delcher, M. Gardner, H. Tettelin, *Genomics* **59**, 24 (1999).
- S9. W. Majoros, M. Pertea, S. Salzberg, *Bioinformatics (Oxford, England)* **20**, 2878 (2004).
- S10. M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E. Lander, *Nature* **423**, 241 (2003).
- S11. T. M. Lowe, S. R. Eddy, *Nucleic acids research*. **25**, 955 (1997).
- S12. S. Rozen, H. Skaletsky, *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Humana Press, Totowa, NJ, 2000), chap. Primer3 on the WWW for general users and for biologist programmers, pp. 365–386.
- S13. G. Giaever, et al., *Nature* **418**, 387 (2002).
- S14. A. H. Van Tong, et al., *Science* **303**, 808 (2004).
- S15. P. Phillips, *Genetics* **149**, 1167 (1998).
- S16. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic acids research*. **22**, 4673 (1994).