Grow diploid in rich media

BY/3S WT or BY/3S *HOS3/hos3Δ::KanMX*
MATa/α
CAN1/can1Δ::STE2prSpHIS5
HIS3/his3Δ

Sporulate in defined media with *HIS3* and LiAc

Select MATa haploids on YNB + canavanine

Pick WT segregants
MATa
can1Δ::STE2prSpHIS5

Replica plate onto YPD + G418

Pick *hos3Δ* segregants
hos3Δ::KanMX
MATa
can1Δ::STE2prSpHIS5

Figure S1. Generation of BYx3S segregants. Two distinct BY/3S diploid progenitors were used to obtain wild type and *hos3Δ* segregants respectively. Each diploid was generated and sporulated. The synthetic genetic array marker system was used to generate *MATa* segregants. Wild type segregants were collected directly from *MATa* selection plates. *MATa* selection plates were replica-plated onto G418 plates from which *hos3Δ::KanMX* segregants were selected.
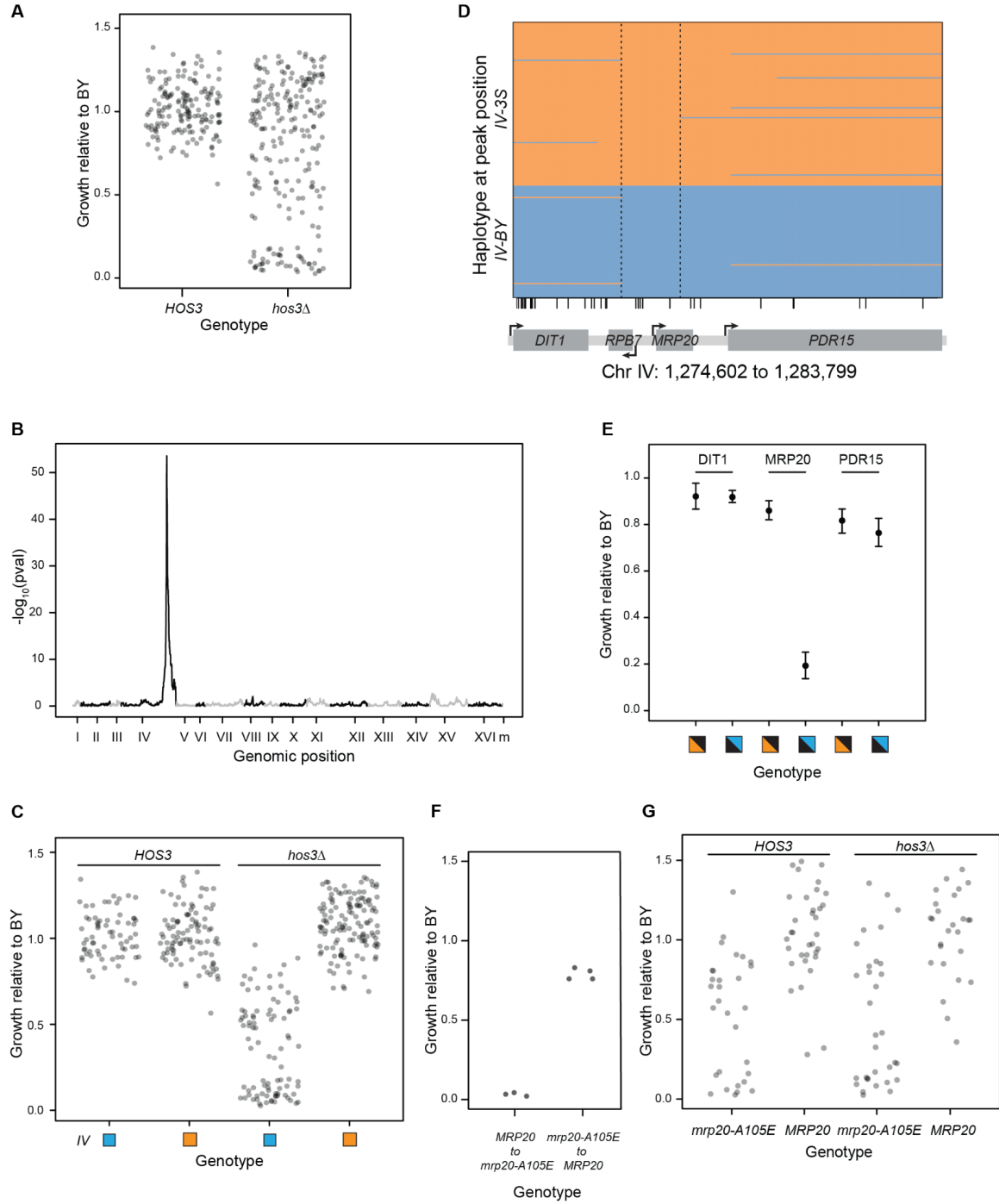
Figure S2. Identification of *mrp20-A105E*. (A) Wild type and hemizygous BY/3S diploids were generated and sporulated to produce *HOS3* and *hos3Δ* F2 BYx3S segregants. BYx3S *hos3Δ* segregants exhibited a large increase in phenotypic variability relative to wild type segregants. (B) Linkage mapping using the *HOS3* and *hos3Δ* segregants identified a single locus on Chromosome IV. The peak marker was from 1,277,231 to 1,277,959 and the confidence interval extended from position 1,272,164 to position 1,278,407, encompassing (from left to right) part of *URH1* and all of *DIT2*, *DIT1*, *RPB7*, and *MRP20*. (C) The BY allele of the Chromosome IV locus had a large effect in *hos3Δ* segregants, but no effect in *HOS3* segregants. (D) Recombination

breakpoints in *hos3Δ* segregants delimited the Chromosome IV locus to five SNPs (small vertical black lines along the x-axis) in the *RPB7-MRP20* region of the chromosome. Dashed vertical lines show the window delimited by the recombination breakpoints. One of these variants was a spontaneous mutation in *MRP20*. Blue and orange respectively refer to the BY and 3S alleles of the locus. (E) Reciprocal hemizygosity analysis in a *hos3Δ* BY/3S diploid was conducted at closely linked non-essential genes and found that *MRP20* is the causal gene underlying the Chromosome IV locus. In these experiments the IV^{BY} allele includes the *mrp20-A105E* mutation and results in a substantial decrease in growth. Black triangles denote the absence of one allele and colored triangles indicate the alleles that are present. (F) The causality of *mrp20-A105E* was validated by engineering in segregants with *MRP20* (left) and *mrp20-A105E* (right). (G) Tetrad dissection of the original BY/3S *HOS3/hos3Δ MRP20/mrp20-A105E* diploid showed that increased variation was due to *mrp20-A105E*, not *hos3Δ*. Throughout the paper, blue and orange are used to denote BY and 3S genetic material, respectively. All growth data presented in the paper are measurements of colonies on agar plates containing rich medium with ethanol as the carbon source.
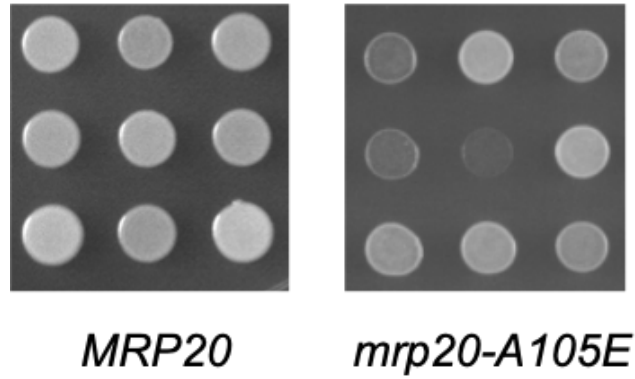
Figure S3. Representative _MRP20_ and _mrp20-A105E_ segregants on ethanol. Each colony is a genetically distinct BYx3S segregant grown on ethanol. A wide range of growth phenotypes was observed among _mrp20-A105E_ segregants, some of which were inviable in this condition.
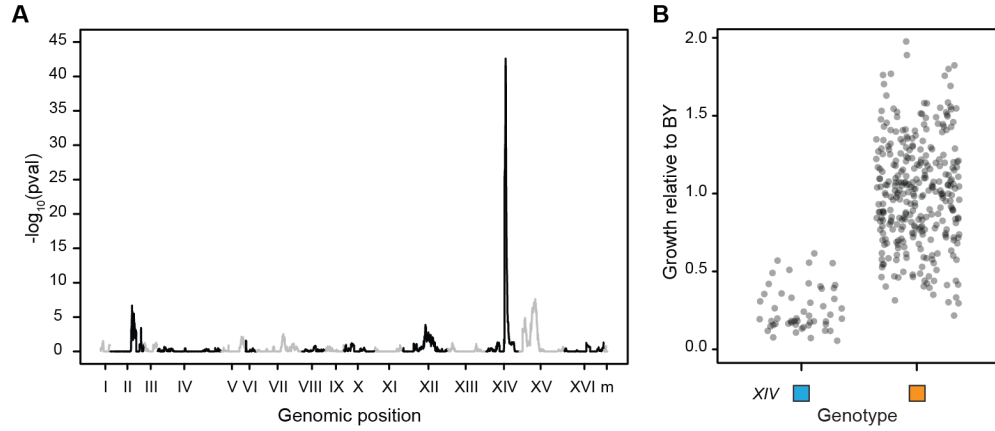
Figure S4. Linkage mapping in the F$_3$ panel more finely resolves the Chromosome XIV locus. The model *growth ~ locus + error* was used. The genome-wide significance plot of the *locus* term is shown in (**A**) and the relationship between genotype at the Chromosome XIV locus are shown in (**B**). The peak and 99% confidence interval solely included the position 467,219.
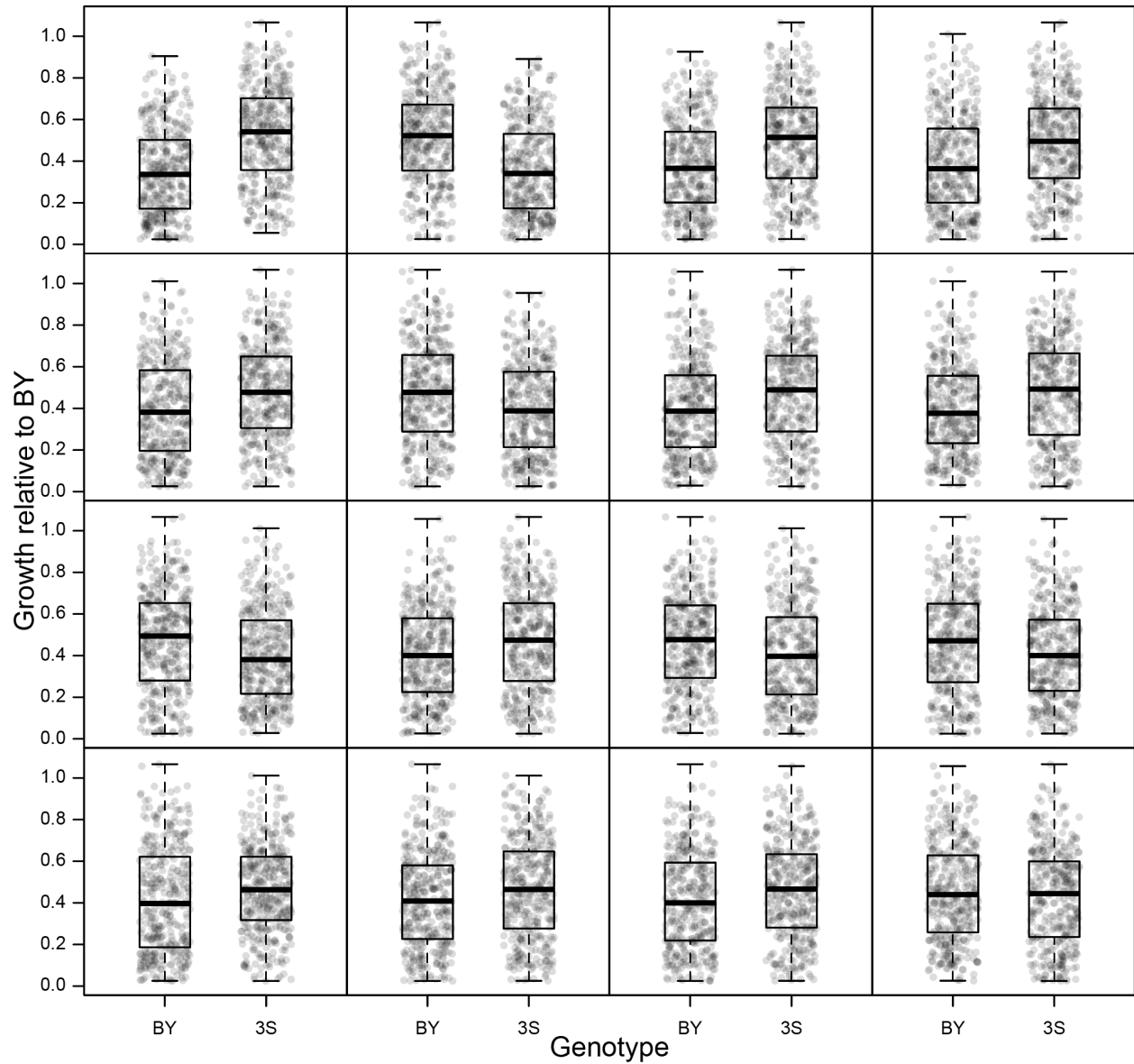
Figure S5. Growth effects of loci detected in BY x 3S *mrp20-A105E* crosses. The relationship between genotype is shown at each of the 16 loci detected among BYx3S *mrp20-A1015E* segregants shown in Figure 5C. Effects are shown from greatest to least effect size, left to right, top to bottom.
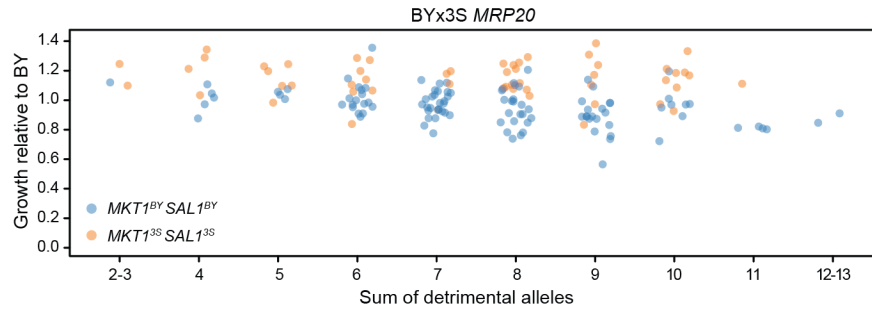
Figure S6. Loci affecting expressivity of *mrp20-A105E* show minimal effects in *MRP20* segregants. Growth relative to the sum of detrimental alleles is shown for *MRP20* segregants. While predictions for *MRP20* segregants correlated with observed growth (Pearson's r = 0.70, p = 9.6 x 10^{-25}), the cumulative effects of loci differed between *mrp20-A105E* and *MRP20* segregants (ANOVA, observedGrowth ~ predictedGrowth*MRP20; interaction term p = 2.8 x 10^{-23}). This is likely, in part, due to the fact that wild type segregants exhibited a narrower range of phenotypes which did not include inviable segregants.

Table S1. Programs and settings used for sequence analyses. This table includes the programs used to map and process sequencing reads in bash, and additional programs and functions beyond base R that were used in subsequent analyses included in this paper.

Program / Library / Function	Version	Setting	Language	Reference(s)
bwa	0.7.7-r44	default	bash	(Li and Durbin 2009)
samtools	1.9	default	bash	(Li *et al.* 2009)
HMM, initHMM()	1.0	default	R	(Rabiner 1989)
mixtools, normalmixEM()	1.2.0	default	R	(Benaglia *et al.* 2009)
car, leveneTest()	3.0.6	default	R	(Fox and Weisberg 2018)
sommer, A.mat(), mmer()	4.0.9	default	R	(Henderson 1975; Endelman and Jannink 2012; Covarrubias-Pazaran 2016)

Table S2. Crosses and segregant populations examined in this study. All BY x 3S crosses and segregant populations examined in this study are listed. Note, at times different methods of obtaining segregants, either random spores or tetrad dissection were employed.

Diploid	Cross	Segregants	Method	Publication	Total
A	BY x 3S	wild type F_2	random spores	Mullis, et al, 2018	164
B	BY *mrp20-A105E hos3∆* x 3S *MRP20 HOS3*	*MRP20 hos3∆* F_2	random spores	Mullis, et al, 2018	131
		mrp20-A105E hos3∆ F_2		Mullis, et al, 2018	90
		MRP20 hos3∆ F_2	tetrad dissection	this paper	27
		mrp20-A105E hos3∆ F_2		this paper	32
		MRP20 HOS3 F_2		this paper	34
		mrp20-A105E HOS3 F_2		this paper	30
C	BYx3S *mrp20-A105E XIV^{BY}* F_2 x BYx3S *mrp20-A105E XIV^{3S}* F_2	*mrp20-A105E* F_3	random spores	this paper	361
D	BY *mrp20-A105E MKT1^{BY}* x 3S *mrp20-A105E MKT1^{BY}*	*mrp20-A105E MKT1^{BY}* F_2	tetrad dissection	this paper	353
E	BY *mrp20-A105E MKT1^{3S}* x 3S *mrp20-A105E MKT1^{3S}*	*mrp20-A105E MKT1^{3S}* F_2	tetrad dissection	this paper	396

Table S3. Loci other than *MKT1* that influence growth in *mrp20-A105E* segregants. These loci were detected by mapping *growth ~ locus* in *mrp20-A105E MKT^{BY}* F$_2$ and in *mrp20-A105E MKT^{3S}* F$_2$ individuals shown in Figure 5-8 and Figure S4. Confidence intervals are reported as 2 LOD drops around the peak position at a locus.

Chromosome	Peak Position(s)	Confidence Interval	P-value
4	615,114	565,313 to 639,037	6.6×10^{-7}
7	140,806 to 142,398	86,297 to 164,025	1.4×10^{-6}
11	171,261 to 171,457	150,554 to 264,387	2.1×10^{-6}
12	448,368	444,569 to 513,765	7.5×10^{-9}
12	679,600	660,371 to 701,793	4.8×10^{-10}
12	116,314 116,980	85,385 to 141,900	3.2×10^{-5}
12	1,022,895 to 1,022,933	1,013,592 to 1,059,611	1.1×10^{-8}
13	612,355	594,280 to 652,003	2.8×10^{-5}
13	833,534	812,150 to 892,748	4.6×10^{-7}
13	449,844 to 451,590	436,342 to 465,619	3.2×10^{-5}
14	299,515	295,050 to 312,454	2.0×10^{-12}
14	473,648	468,488 to 478,701	4.8×10^{-34}
15	185,793	167,270 to 189,700	3.4×10^{-5}
15	343,484 to 343,921	340,625 to 363,553	2.4×10^{-24}
15	627,315 to 628,209	553,072 to 717,181	2.7×10^{-5}
15	885,437 to 885,914	869,837 to 916,094	5.0×10^{-6}

Table S4. Candidate genes at loci in Table S3. These loci additively affect the expressivity of *mrp20-A105E* mutation. Recombination delimits each peak to between one candidate (12 loci), two candidate genes (3 loci), or three genes (1 locus). Location of the delimited SNPs is included.

Chromosome	Peak Position(s)	Candidate Gene(s)	Position of Peak SNP(s)
4	615,114	*AFR1*	Coding
7	140,806 to 142,398	*HOS2, YGL193C,* and *IME4*	Promoter and Coding
11	171,261 to 171,457	*SDH1*	Promoter
12	116,314 116,980	*BPT1*	Promoter and Coding
12	448,368	*RNH203*	Promoter
12	679,600	*BOP2*	Coding
12	1,022,895 to 1,022,933	*ECM7*	Coding
13	449,844 to 451,590	*YMR090W* and *NPL6*	Promoter
13	612,355	*ECM5*	Coding
13	833,534	*AEP2*	Coding
14	299,515	*PBR1*	Coding
14	472,584 to 473,648	*SAL1* and *PMS1*	Promoter and Coding
15	185,793	*BRX1*	3' UTR
15	343,484 to 343,921	*YOR008C-A* and *TIR4*	Promoter
15	627,315 to 628,209	*ISN1*	Promoter and Coding
15	885,437 to 885,914	*ISW2*	Promoter and Coding

Table S5. Presence of Chromosome II duplication differs among BY x 3S crosses. We observed a Chromosome II duplication in crosses fixed for *mrp20-A105E*. Among these two crosses, the cross fixed for *MKT1³S* had much higher prevalence of the aneuploidy relative to the cross fixed for *MKT1^{BY}*.

Diploid	Cross	% wild type	% aneuploid
A	BY x 3S	100	0
D	BY *mrp20-A105E MKT1^B* x 3S *mrp20-A105E MKT1^{BY}*	94	5.9
E	BY *mrp20-A105E MKT1³S* x 3S *mrp20-A105E MKT1³S*	51	49

Supplementary Data Legends

Dataset S1. Segregant genotype table
Chromosome and position columns refer to the chromosome and position of each genetic variant used in this study, The mitochondria is referred to as chromosome 17. Each additional column contains the genetic information for a given segregant. Each segregant was named by type (F_2 and F_3), the diploid from which it originated (A, B, C, D, E defined in table S2), whether that segregant was wild type or mutant at *MRP20* ('MRP20' or 'mrp20'), and was randomly numbered from one through the total number of that segregant type. Segregants originating from diploid B contained additional information pertaining to whether that segregant was wild type or knockout at *HOS3* ('HOS3' or 'hos3Δ') and whether that segregant was obtained by random spore prep or tetrad dissection ('random' or 'dissected'). Note, that for *hos3Δ* segregants obtained by random spore preparation of diploid B, the BY allele at *MRP20* contained the *mrp20-A105E* mutation. Each genetic variant used in this study is presented as a row, whereby the haplotype information for each segregant is denoted as 0 for BY or 1 for 3S respectively. A value of 'NA' indicates a site that lacked coverage for which a haplotype was not called. In the BY x 3S crosses fixed for *mrp20-A105E* (diploids D and E), a third heterozygous state (2) was used to denote heterozygosity for individuals with the chromosome II duplication event.

Data S2. Segregant phenotype table
Each segregant's growth in ethanol is presented. A single growth measurement is reported, which is the mean value of three biological replicates of growth normalized to on plate BY controls.

Data S3. Reciprocal hemizygosity experiments
Segregants that were used in reciprocal hemiygosity experiments that delimied the Chromosome IV allele to *MRP20* are included. The segregants mated together for each hemizygous diploid are listed under 'Parent1' and 'Parent2' columns. The 'Gene' column lists the gene at which reciprocal hemizygosity was engineered. The 'LossOfFunction' column indicates which allele, BY or 3S (encoded as 0 or 1) was engineered to be non-functional. The 'Ethanol' columns contains a growth value normalized to on plate BY controls. Each biological replicate is included separately and denoted by the 'Replicate' column to enable calculation of confidence intervals.

Data S4. Cloning experiments
Each segregant and parent strain used for cloning causal nucleotides at *MRP20* and *MKT1* are included. The 'Type' column denotes whether the engineered strain is a segregant or parent ('segregant' or 'parent'), the 'MRP20' column describes whether that strain is wild type or *mrp20-A105E* (denoted as 'MRP20' or 'mrp20'), and the 'MKT1' column describes whether that strain was BY or 3S (encoded as 0 or 1) at the causal SNP at position 467,219. The 'Gene' column lists the gene at which engineering had occurred, including ('MRP20', 'MKT1', 'MRP20andMKT1', and 'WT' which denotes parental control samples. The 'Edit' column explains the type of engineering that was performed in segregants. Thus, cloning experiments at *MRP20* in segregants are described as 'fromMuttoWT' or 'fromWTtoMut'. Similarly cloning experiments at *MKT1* in segregants are described as 'from3StoBY-1SNP', 'from3StoBYcandidate', and 'from3StoBY+1SNP' for strain engineering at the nearest upstream, causal, and nearest downstream SNPs. This column is not relevant for parental cloning and therefore NA is reported in those cells. Lastly, the 'Ethanol' column contains a single growth measurement which is the mean value of three biological replicates of growth normalized to on plate BY controls.

Data S5. Petite frequency
Each strain used in petite frequency assays, either segregants described in Data S1-S3 or parental strains is reported in the 'Sample' column. The 'Type' column denotes whether the given strain is a segregant or parent ('segregant' or 'parent'), and the 'MRP20' column describes whether or not the strain is wild type or *mrp20-A105E* (denoted as 'MRP20' or 'mrp20'). A comma separated list of Image J reported colony sizes (in^2) is also included in the 'ColonySizes' column. The largest observed petite colony among wild type parental strains was 0.001 in^2 and was thus used to designate *petites* from total colonies. The final 'Frequency' column reports the petite frequency, or the number colonies at or below this threshold relative to the total number of colonies times 100.

Data S6. Individuals with Chromsome II duplication event

Each segregant from BY x 3S crosses that were engineered at *mrp20-A105E* and *MKT1* (croses D and E) in which the Chromosome II duplication was observed is listed under the 'Sample' column, The average normalized coverage acrpss chromosome II is reported in the 'AverageChr2Coverage' column, and the 'Chr2C' column denotes if that segregant was determined to be WT or Aneuploid (encoded as 0 or 1).

Data S7. Genetic mapping analysis code

All code used for linkage mapping and multiple testing correction is included as '.R' file to be used in R programming language.

Data S8. Statistical analysis code

All code used for statistical analyses is included as '.R' file used to be used in R programming language.

Data S9. Figure plotting code

All code used for plotting data shown in main text and manuscript is included as '.R' file to be used in R programming language.