Genetic architecture of a mutation's expressivity

Rachel Schell1, Joseph J. Hale1, Martin N. Mullis1, Takeshi Matsui1,2, Ryan Foree1, and Ian M. Ehrenreich1,*

1 Molecular and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA

2 Present address: Joint Initiative for Metrology in Biology, Stanford University, Stanford, CA 94305, USA

* Correspondence to: Ian M. Ehrenreich Phone #: 213 – 821 – 5349 Email: ian.ehrenreich@usc.edu

Abstract

Genetic background often influences the phenotypic consequences of mutations, resulting in variable expressivity. How standing genetic variants collectively cause this phenomenon is not fully understood. Here, we comprehensively identify loci in a budding yeast cross that impact the growth of individuals carrying a spontaneous missense mutation in the nuclear-encoded mitochondrial ribosomal gene *MRP20*. Initial results suggested that a single large effect locus influences the mutation's expressivity, with one allele causing inviability in mutants. However, further dissection revealed that in fact many additional loci shape the mutation's expressivity, collectively leading to a wide spectrum of mutational responses. These results exemplify how complex combinations of loci can produce a diversity of qualitative and quantitative responses to the same mutation.

One Sentence Summary Complex genetics shape a mutation's expressivity.

Main Text

Mutations frequently exhibit different effects in genetically distinct individuals (or 'background effects') (1-3). For example, not all people with the same disease-causing mutations manifest the associated disorder or exhibit identical symptoms. A commonly observed form of background effect among individuals carrying the same mutation is different degrees of response to that mutation (or 'variable expressivity') (4). Variable expressivity can arise due to a myriad of reasons, including genetic interactions (or epistasis) between a mutation and segregating loci (1), dominance (1), stochastic noise (5), and the environment (1).

The genetic architecture underlying expressivity has proven difficult to study, in part because natural populations harbor substantial genetic diversity, which can facilitate complex forms of epistasis between segregating loci and mutations (6-19). Mapping the loci involved in these genetic interactions is technically challenging. However, the introduction of mutations into diverse strains and crosses in the laboratory can enable the identification and genetic dissection of such background effects (6, 11, 16-18, 20).

In this paper, we use a series of controlled crosses in the budding yeast *Saccharomyces cerevisiae* to comprehensively characterize the genetic basis of a mutation's expressivity. We focus on a missense mutation in *MRP20*, an essential nuclear-encoded subunit of the mitochondrial ribosome (21). This mutation occurred by chance in a cross between the reference strain BY4716 ('BY') and a clinical isolate 322134S ('3S') while we were deleting another gene and was subsequently found to show variable

expressivity among BYx3S cross progeny. This presented an opportunity to determine how loci segregating in the BYx3S cross individually and collectively influence this mutation's expressivity.

A spontaneous mutation increases phenotypic variance in the BYx3S cross

221 haploid BYx3S segregants lacking the histone deacetylase Hos3 (22) were previously generated using standard approaches for transformation, gene deletion, and random spore isolation table (S1) (16). Relative to wild type segregants, the *hos3Δ* segregants exhibited substantially elevated variability in growth, especially on agar plates containing ethanol as the carbon source ('ethanol'; Fig. 1, A and B and table S2; Levene's test, $p = 1.4 \times 10^{-22}$). On ethanol, our focal condition hereafter, the *hos3Δ* segregants exhibited a non-normal distribution of phenotypes that quantitatively and qualitatively differed from what was seen among wild type segregants. A subset of the *hos3Δ* segregants barely grew ('inviable'), while no wild type segregants were inviable (fig. S1).

Although the increased phenotypic variance appeared to be due to the *HOS3* deletion, additional work showed that another mutation was in fact responsible. A linkage scan for growth among the *hos3Δ* segregants (16) detected a single locus on Chromosome IV (ANOVA, $p = 2.54 \times 10^{-54}$; Fig. 1C, supplementary text 1). This locus explained most of the phenotypic variance among *hos3Δ* segregants (ANOVA, $R^2 = 0.74$, $p = 4.3 \times 10^{-66}$) but had no effect in wild type segregants (ANOVA, $R^2 = 0$, $p = 0.91$; Fig 1. D). All inviable *hos3Δ* segregants carried the BY allele of the Chromosome IV locus.

We used reciprocal hemizygosity experiments and analysis of recombination breakpoints to resolve the Chromosome IV locus to five SNPs, one of which was a spontaneous mutation in the BY copy of *MRP20* (Fig. 1E and fig. S2). This mutation, which occurred in the *BY/3S HOS3/hos3Δ* diploid, results in an alanine to glutamine substitution at amino acid 105, in a core domain of Mrp20 that is conserved from bacteria to humans (21, 23). We validated the causality of *mrp20-105E* using nucleotide replacements in *hos3Δ* segregants (Fig. 1F).

To test whether the effect of *mrp20-105E* depended on the *HOS3* deletion, we dissected tetrads produced by a BY/3S *HOS3/hos3Δ MRP20/mrp20-105E* diploid and phenotyped the spores (table S1). Analysis of all four combinations of alleles at the two genes determined that *mrp20-105E* does not genetically interact with *HOS3*, as would be expected if the increased phenotypic variance required both mutations (ANOVA, $p = 0.76$; Fig. 1G). Thus, increased phenotypic variance among *hos3Δ* segregants does not involve *hos3Δ*, but rather is due to *mrp20-105E* alone.

A locus with a major effect on the expressivity of *mrp20-105E*

The high phenotypic variance among *mrp20-105E* segregants suggested this mutation shows variable expressivity. Loci contributing to this variable expressivity should be detectable through their genetic interactions with *MRP20*. To find such loci, we performed linkage scans for two-way epistasis with *MRP20*, using the *hos3Δ* segregants. We identified a single locus on Chromosome XIV (ANOVA, $p = 4.3 \times 10^{-16}$; Fig. 2A). Individuals with XIV^{BY} showed reduced growth among both *MRP20* and *mrp20-105E* segregants, but to a greater degree among the latter (Fig. 2B). The Chromosome XIV locus explained 79% of the phenotypic variance that was originally seen among *mrp20-105E* segregants (ANOVA, $p = 3.2 \times 10^{-31}$) and accounted for all observed cases of inviability (Fig. 2B).

To further resolve the Chromosome XIV locus, we crossed an *mrp20-105E* XIV^{BY} F_2 segregant and an *mrp20-105E* XIV^{3S} F_2 segregant (supplementary text 2; Fig. 2C, table S1). 361 F_3 progeny were genotyped

by low-coverage whole genome sequencing and phenotyped for growth. Linkage mapping with these data reidentified the Chromosome XIV locus at a p-value of 2.50 x 10^{-43} (ANOVA; Fig. 2D, fig. S3) and resolved it to a single SNP in the coding region of *MKT1* (Fig. 2E). This SNP, which encodes a glycine in BY and a serine in 3S at amino acid 30, was then validated by nucleotide replacement in *mrp20-105E* segregants (Fig. 2F). Notably, this specific SNP was previously shown to play a role in mitochondrial genome stability (24), suggesting epistasis between *MRP20* and *MKT1* involves mitochondrial dysfunction, impairing growth on non-fermentative carbon sources such as ethanol.

Epistasis between *MRP20* and *MKT1* differs in the cross parents

We attempted to validate the epistasis between *MKT1* and *MRP20* by introducing all four possible combinations of *MRP20* and *MKT1* into both BY and 3S (Fig. 3A). The *mrp20-105E* mutation affected growth in both parent strains (ANOVA, p = 4.3 x 10^{-24} and p = 4.0 x 10^{-4}). However, the magnitude of the effect differed between the two: *mrp20-105E* caused inviability in BY but had a more modest effect in 3S. In addition, *MKT1* influenced response to *mrp20-105E* in the 3S background (ANOVA, p = 0.01) but not in the BY background (ANOVA, p = 0 .99). The *MRP20-MKT1* interaction's sensitivity to genetic background implies that additional unidentified loci must also influence response to *mrp20-105E*.

Many additional loci affect the expressivity of *mrp20-105E*

To identify other loci underlying response to *mrp20-105E*, we generated two new BYx3S crosses (Fig. 3B, table S1). In both crosses, the BY and 3S parents were engineered to carry *mrp20-105E*. Further, one cross was engineered so that both parents carried $MKT1^{BY}$ and the other cross was engineered so that both parents carried $MKT1^{3S}$. By altering the parent strains in this manner, we increased the chance of detecting additional loci contributing to the variable expressivity of *mrp20-105E*. From these engineered crosses, 749 total segregants were obtained through tetrad dissection, genotyped by low-coverage genome sequencing, and phenotyped for growth on ethanol.

The new crosses exhibited a much broader range of phenotypes than was seen in the original experiments with *hos3Δ* segregants and lacked a distinct class of inviable segregants (Fig. 3C). In both the $MKT1^{BY}$ and $MKT1^{3S}$ crosses, *mrp20-105E* segregants ranged from inviable to nearly wild type. The distributions of phenotypes in the two crosses differed in a manner consistent with their *MKT1* alleles, with the mean of the $MKT1^{BY}$ segregants lower than the $MKT1^{3S}$ segregants (t-test, p = 4.8 x 10^{-34}). These data show that regardless of the *MKT1* allele present, additional loci can cause *mrp20-105E* to show a range of effects from lethal to benign.

We used the new crosses to map other loci contributing to response to *mrp20-105E*. Excluding *MKT1*, which explained 18% of the phenotypic variance in the new crosses, linkage mapping identified 16 new loci (Fig. 3D, fig. S4, and table S3). Of these loci, the BY allele was inferior at 10 and superior at six. These loci individually explained between 0.79% and 14% of the phenotypic variance. 13 of these loci resided on a subset of chromosomes but were distantly linked: four on Chromosomes XII, three on XIII, two on XIV, and four on XV. The three remaining loci were detected on Chromosomes IV, VII, and XI.

The Chromosome XIV locus contains multiple causal variants

Among the newly detected loci, the largest effect (14% phenotypic variance explained) was on Chromosome XIV. The position of maximal significance at this site was located two genes away from the end of *MKT1*, with a 99% confidence interval that did not encompass the causal variant in *MKT1* (table

S3). Thus, the originally identified large effect Chromosome XIV locus in fact represents multiple distinct closely linked nucleotides that both genetically interact with *MRP20* and occur in different genes (Fig. 3E).

The new locus on Chromosome XIV was delimited to two genes, one of which was *SAL1*, encoding a mitochondrial ADP/ATP transporter that physically interacts with Mrp20. A SNP in *SAL1* that segregates in this cross was previously linked to increased mitochondrial genome instability in BY (24), suggesting it is likely also causal in our study. For this reason, we refer to this additional Chromosome XIV locus as *'SAL1'*. We find no evidence for epistasis between *MKT1* and *SAL1* (ANOVA, p = 0.77).

Although the *MKT1-SAL1* locus had a large effect, it explained a minority of the phenotypic variance among *mrp20-105E* segregants in a model including all detected loci (32% for *MKT1-SAL1* vs. 36% for all other loci collectively). Thus, by enabling *MKT1* and *SAL1* to segregate independently through genetic engineering and examining a large number of *mrp20-105E* segregants with different *MKT1-SAL1* genotypes, we observed a greater diversity of mutational responses than was originally seen and detected many additional loci.

Aneuploidy also contributes to the variable expressivity of *mrp20-105E*

Despite the fact that the identified loci explain most of *mrp20-105E*'s variable expressivity, some individuals exhibited unexpectedly poor growth (Fig. 3F). This finding led to the identification of a Chromosome II duplication that reduced growth (ANOVA, 1.2×10^{-48}). The aneuploidy was common among *mrp20-105E* segregants, with a higher prevalence when $MKT1^{3S}$ was also present (Fisher's exact test, $p = 1.5 \times 10^{-43}$; table S4). The Chromosome II aneuploidy was not seen among wild type segregants. These data suggest that *mrp20-105E* increases the rate of aneuploidization and that genetic variation in *MKT1* influences the degree to which *mrp20-105E* segregants duplicate Chromosome II. The aneuploidy's contribution to phenotypic variation was relatively minor, explaining 5% of phenotypic variance among *mrp20-105E* segregants in a model also including all identified loci.

Multiple mechanisms underlie poor growth in the presence of *mrp20-105E*

Evidence suggests mitochondrial genome instability contributes to the variable expressivity of *mrp20-105E*. First, mitochondrial genome instability is known to cause poor growth on non-fermentative carbon sources, such as ethanol (25, 26). Second, the exact variants that segregate in our cross at *MKT1* and *SAL1* were previously linked to mitochondrial genome instability (24). Third, both Mrp20 and Sal1 function in the mitochondria (21, 27).

To determine the role of mitochondrial genome instability in the variable expressivity of *mrp20-105E*, we measured petite formation, a proxy for spontaneous mitochondrial genome loss (Fig. 4) (28). In addition to *MRP20* and *mrp20-105E* BY and 3S parent strains, 16 *MRP20* segregants and 42 *mrp20-105E* segregants were examined. Despite causing reduced growth in both parents, *mrp20-105E* only led to elevated mitochondrial genome instability in BY (t-test p = 0.013 in BY and p = 0.39 in 3S; Fig. 4A). This result indicates that mitochondrial genome instability explains part, but not all, of response to *mrp20-105E*. Supporting this point, *mrp20-105E* segregants exhibited increased mitochondrial genome instability relative to *MRP20* segregants (Wilcoxon rank-sum test p = 0.023), especially at lower levels of growth, but a subset of inviable segregants did not show elevated petite formation (Fig. 4B and C).

These results suggest that *mrp20-105E* can lead to poor growth through mitochondrial genome instability, but this is not the only mechanism determining response to *mrp20-105E*.

Genetic underpinnings of *mrp20-105E*'s expressivity in segregants and parents

We determined the extent to which our identified loci explained phenotypic variability among mutants. Modeling growth as a function of all identified loci and the aneuploidy accounted for the majority (78%) of the broad-sense heritability among *mrp20-105E* segregants (ANOVA, $p = 5.2 \times 10^{-188}$). Further, phenotypic predictions for segregants based on their genotypes were strongly correlated with their observed phenotypes ($r = 0.85$, $p = 4.4 \times 10^{-209}$; Fig. 5A).

The model was also effective for other genotypes that were not present in the new crosses, but had been generated throughout the course of this work. For instance, the model accurately predicted the phenotypes of *mrp20-105E* segregants in the original *hos3Δ* population ($r = 0.90$, $p = 1.6 \times 10^{-39}$ for *hos3Δ mrp20-105E* segregants), as well as for the cross parents (Fig. 5A). Moreover, the model explained both qualitative and quantitative variation within and between the two Chromosome XIV classes that were originally seen among *hos3Δ mrp20-105E* segregants. These results confirm that the variable expressivity of *mrp20-105E* is driven by many loci that collectively produce a spectrum of mutational responses.

Lastly, we examined the relationship between growth and the total number of detrimental alleles carried by *mrp20-105E* segregants (Fig. 5B). The number of detrimental alleles carried by a segregant showed a strong negative relationship with growth, which was not observed in wild type segregants (fig. S5). Further, regardless of genotype at *MKT1* and *SAL1*, the largest effect loci, growth ranged from lethal to benign in a manner dependent on the number of detrimental alleles present at other loci. These findings demonstrate that many segregating loci beyond the large effect *MKT1-SAL1* locus influence the expressivity of *mrp20-105E* and enable different genotypes in the cross to exhibit similar responses to the mutation.

Discussion

We have provided a detailed genetic characterization of the expressivity of a spontaneous mutation. Response to this mutation in a budding yeast cross is influenced by at least 18 genetic factors in total, with the largest effect due to two closely linked variants. However, at least 15 additional loci segregate and jointly exert larger effects than the largest two. Different combinations of alleles across these loci produce a continuous spectrum of mutational responses. Due to tight linkage between *MKT1* and *SAL1* in the original cross parents, the full extent of this continuum was not originally observed, leading to an initial understanding of the expressivity of the *mrp20-105E* mutation that was simplistic.

These findings also show how quantitative variation in mutational response can produce seemingly discrete outcomes. In part, whether responses appear qualitative depends on the configuration of mutationally responsive alleles in examined mutants. Approaches such as crossing of genetically engineered strains can be used to disrupt these configurations that mask the full extent of variation. However, another part of this expressivity is the tolerance of a system to quantitative variation in key processes, for example mitochondrial genome stability in the case of *mrp20-105E*. Our data suggest that these processes can only tolerate quantitative variation to a point, but also indicate that lethality to the same mutation may arise in different genetic backgrounds due to impairment of distinct processes.

Our results inform efforts to understand expressivity in other systems, including humans. For example, there is interest in determining why people who carry highly penetrant alleles known to cause disease do not develop pathological conditions (3, 29, 30). Such resilience, as observed here, may involve numerous loci. This speaks to the complicated and unexpected epistasis that can arise between mutations and segregating loci in genetically diverse populations (6-19). It also illustrates the importance of characterizing epistasis (31-41), including background effects, as these forms of genetic interactions are immediately relevant to evolution and disease, and may not emerge from studies that do not directly interrogate natural variation in genetically diverse populations.

References

1. J. H. Nadeau. *Nat Rev Genet* **2**, 165-174 (2001).
2. C. H. Chandler, S. Chari, I. Dworkin. *Trends Genet* **29**, 358-366 (2013).
3. J. D. Riordan, J. H. Nadeau. *Am J Hum Genet* **101**, 177-191 (2017).
4. A. J. F. Griffiths, S. R. Wessler, S. B. Carroll, J. Doebley, *Introduction to genetic analysis*. (Macmillian Publishers, New York, NY, ed. 12th, 2015).
5. A. Raj, S. A. Rifkin, E. Andersen, A. van Oudenaarden. *Nature* **463**, 913-918 (2010).
6. R. D. Dowell *et al. Science* **328**, 469 (2010).
7. S. Chari, I. Dworkin. *PLoS Genet* **9**, e1003661 (2013).
8. C. H. Chandler, S. Chari, D. Tack, I. Dworkin. *Genetics* **196**, 1321-1336 (2014).
9. A. B. Paaby *et al. Elife* **4**, (2015).
10. M. B. Taylor, I. M. Ehrenreich. *PLoS Genet* **10**, e1004324 (2014).
11. M. B. Taylor, I. M. Ehrenreich. *PLoS Genet* **11**, e1005606 (2015).
12. V. Vu *et al. Cell* **162**, 391-402 (2015).
13. M. B. Taylor, J. Phan, J. T. Lee, M. McCadden, I. M. Ehrenreich. *Nat Commun* **7**, 11669 (2016).
14. J. T. Lee, M. B. Taylor, A. Shen, I. M. Ehrenreich. *PLoS Genet* **12**, e1005929 (2016).
15. C. H. Chandler *et al. PLoS Genet* **13**, e1007075 (2017).
16. M. N. Mullis, T. Matsui, R. Schell, R. Foree, I. M. Ehrenreich. *Nat Commun* **9**, 3548 (2018).
17. J. Hou, G. Tan, G. R. Fink, B. J. Andrews, C. Boone. *Proc Natl Acad Sci U S A* **116**, 5045-5054 (2019).
18. J. T. Lee, A. L. V. Coradini, A. Shen, I. M. Ehrenreich. *Genetics* **211**, 1469-1482 (2019).
19. L. Parts *et al. Mol Syst Biol* **17**, e10138 (2021).
20. M. Galardini *et al. Mol Syst Biol* **15**, e8831 (2019).
21. K. Fearon, T. L. Mason. *J Biol Chem* **267**, 5162-5170 (1992).
22. A. A. Carmen *et al. Proc Natl Acad Sci U S A* **96**, 12356-12361 (1999).

23. E. C. Koc *et al. J Biol Chem* **276**, 43958-43969 (2001).
24. L. N. Dimitrov, R. B. Brem, L. Kruglyak, D. E. Gottschling. *Genetics* **183**, 365-383 (2009).
25. K. A. Lipinski, A. Kaniak-Golik, P. Golik. *Biochim Biophys Acta* **1797**, 1086-1098 (2010).
26. G. S. Shadel. *Am J Hum Genet* **65**, 1230-1237 (1999).
27. B. Kucejova, L. Li, X. Wang, S. Giannattasio, X. J. Chen. *Mol Genet Genomics* **280**, 25-39 (2008).
28. B. Ephrussi, P. P. Slonimski. *Nature* **176**, 1207-1208 (1955).
29. R. Chen *et al. Nat Biotechnol* **34**, 531-538 (2016).
30. V. M. Narasimhan *et al. Science* **352**, 474-477 (2016).
31. O. Carlborg, C. S. Haley. *Nat Rev Genet* **5**, 618-625 (2004).
32. H. Shao *et al. Proc Natl Acad Sci U S A* **105**, 19910-19914 (2008).
33. T. F. Mackay. *Nat Rev Genet* **15**, 22-33 (2014).
34. T. F. Mackay, J. H. Moore. *Genome Med* **6**, 124 (2014).
35. M. L. Siegal, J. Y. Leu. *Annu Rev Ecol Evol Syst* **45**, 496-517 (2014).
36. M. B. Taylor, I. M. Ehrenreich. *Trends Genet* **31**, 34-40 (2015).
37. S. K. Forsberg, J. S. Bloom, M. J. Sadhu, L. Kruglyak, O. Carlborg. *Nat Genet* **49**, 497-503 (2017).
38. I. M. Ehrenreich. *Genetics* **206**, 531-535 (2017).
39. R. F. Campbell, P. T. McGrath, A. B. Paaby. *Trends Genet* **34**, 883-898 (2018).
40. M. Costanzo *et al. Science* **353**, aaf1420 (2016).
41. E. Kuzmin *et al. Science* **360**, aao1729 (2018).

Supplemental References

42. M. N. Mullis, T. Matsui, R. Schell, R. Foree, I. M. Ehrenreich. *Nat Commun* **9**, 3548 (2018).
43. A. H. Tong, C. Boone. *Methods Mol Biol* **313**, 171-192 (2006).
44. H. Li, R. Durbin. *Bioinformatics* **25**, 1754-1760 (2009).
45. H. Li *et al. Bioinformatics* **25**, 2078-2079 (2009).
46. M. B. Taylor, J. Phan, J. T. Lee, M. McCadden, I. M. Ehrenreich. *Nat Commun* **7**, 11669 (2016).
47. L. R. Rabiner. *Proc IEEE*, 257-286 (1989).
48. T. Benaglia, D. Cahauveau, D. R. Hunter, D. S. Young. *Journal of Statistical Software* **32**, 1-29 (2009).
49. T. Matsui, I. M. Ehrenreich. *PLoS Genet* **12**, e1006158 (2016).
50. G. A. Churchill, R. W. Doerge. *Genetics* **138**, 963-971 (1994).
51. L. M. Steinmetz *et al. Nature* **416**, 326-330 (2002).

52. R. D. Gietz, R. A. Woods. *Methods Enzymol* **350**, 87-96 (2002).

53. M. F. Laughery *et al. Yeast* **32**, 711-720 (2015).

54. K. Kannan *et al. Sci Rep* **6**, 30714 (2016).

55. J. R. Thompson, E. Register, J. Curotto, M. Kurtz, R. Kelly. *Yeast* **14**, 565-571 (1998).

56. D. G. Gibson *et al. Nat Methods* **6**, 343-345 (2009).

57. R. S. Sikorski, P. Hieter. *Genetics* **122**, 19-27 (1989).

58. L. N. Dimitrov, R. B. Brem, L. Kruglyak, D. E. Gottschling. *Genetics* **183**, 365-383 (2009).

Funding This work was funded by grants R01GM110255 and R35GM130381 from the National Institutes of Health to I.M.E., as well as Research Enhancement Fellowships from the University of Southern California Graduate School to R.S. and M.N.M.

Author Contributions R.S. and I.M.E. conceptualized this project. R.S., J.J.H., M.N.M., T.M., and R.F. performed the experiments. R.S. and I.M.E. analyzed the data. R.S. and I.M.E. wrote the paper. R.S., J.J.H., M.N.M., T.M., R.F., and I.M.E. edited the manuscript.

Competing Interests The authors declare no competing interests.

Acknowledgements We thank A. Dudley and G. Cromie for input at an early stage of this project, A. Mahableshwarkar and J. Sloan for assistance with experiments, and A. Coradini, I. Goldstein, and C. Hull for feedback during the implementation and writing phases of this work. We also thank the anonymous reviewers of this manuscript for their valuable feedback and Steven Finkel and his lab for allowing use of laboratory equipment.

Data Availability Statement All raw and processed data used in this work is publicly available. Processed data tables and all code used for analyses are included in supplementary Data S1-9 and will be uploaded to the Ehrenreich Lab GitHub page. Raw sequencing data will be available through Bioproject accession PRJNA739014 in the NCBI Short Read Archive.

Supplementary Materials

Materials and Methods

Supplementary Text S1-S2

Figures S1-S5

Tables S1-S4

Data S1-S9

Figure Legends

Fig. 1. Identification of *mrp20-105E*. (A) Wild type and hemizygous BY/3S diploids were generated and sporulated to produce *HOS3* and *hos3Δ* F2 BYx3S segregants. A spontaneous mutation in *MRP20* was also present in a subset of BYx3S *hos3Δ* segregants. (B) BYx3S *hos3Δ* segregants exhibited a large increase in phenotypic variability relative to wild type segregants. (C) 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*. (D) The BY allele of the Chromosome IV locus had a large effect in *hos3Δ* segregants, but no effect in *HOS3* segregants. (E) 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*. (F) The causality of *mrp20-105E* was validated by engineering in segregants with *MRP20* (left) and *mrp20-105E* (right). (G) Tetrad dissection of the original BY/3S *HOS3/hos3Δ MRP20/mrp20-105E* diploid showed that increased variation was due to *mrp20-105E*, 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.

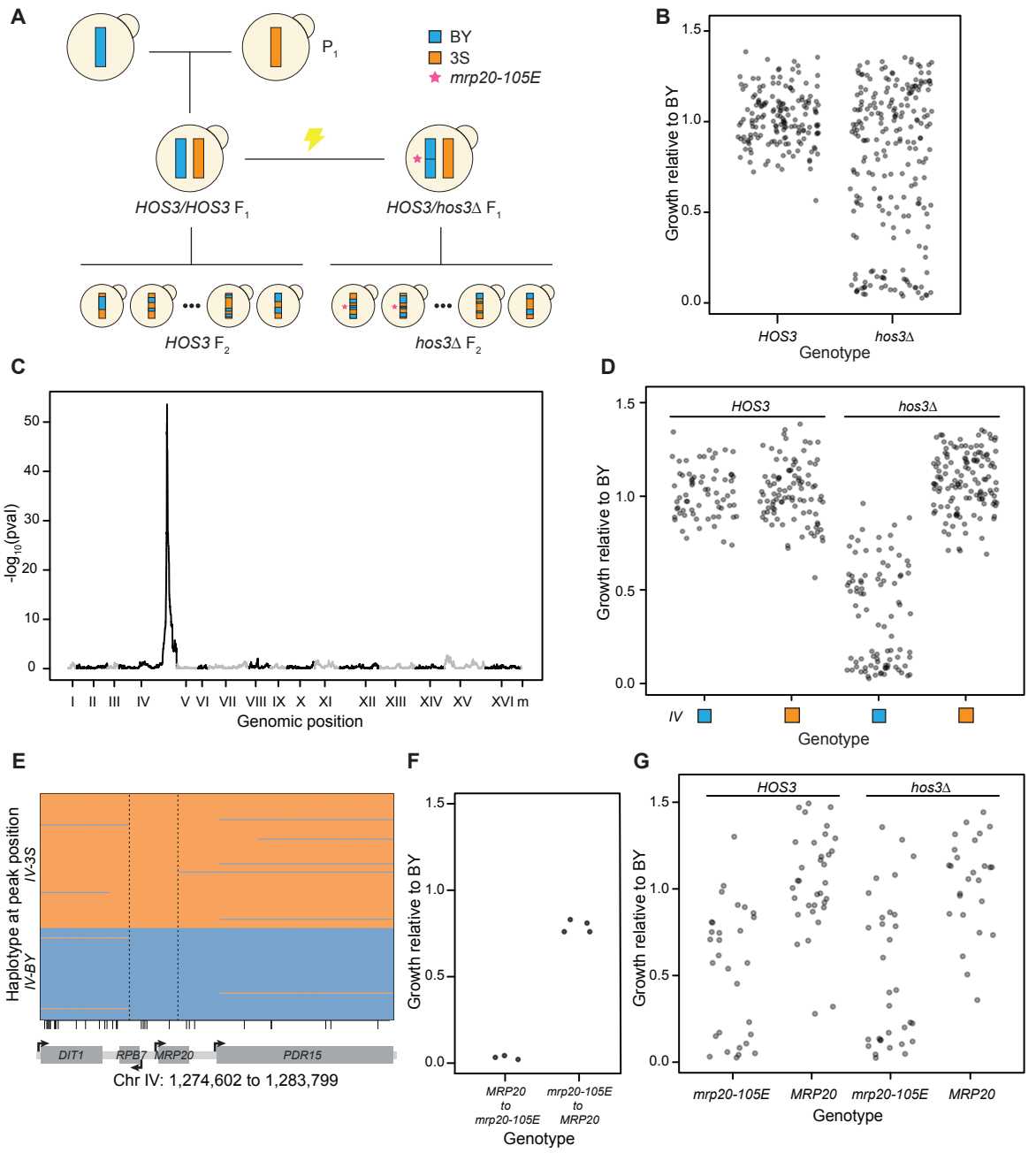
Fig. 2. Epistasis between *MRP20* and *MKT1* appears to mostly explain response to *mrp20-105E*. (A) Linkage mapping in the *hos3Δ mrp20-105E* BYx3S F_2 segregants identified a locus on Chromosome XIV that shows a two-way genetic interaction with *MRP20*. (B) The Chromosome XIV locus had effects in both *MRP20* and *mrp20-105E* segregants but had a greater effect among *mrp20-105E* segregants. (C) To identify the causal gene, we crossed two *mrp20-105E* F_2 segregants that differed at the Chromosome XIV locus and gathered a panel of F_3 segregants. (D) Linkage mapping in the F_3 segregants identified the Chromosome XIV locus at high resolution, with a peak at position 467,219. Tick marks denote every 100,000 bases along the chromosome. (E) Recombination breakpoints in the F_3 segregants delimited the Chromosome XIV locus to a single SNP in *MKT1* at position 467,219. Vertical dashed line highlights the delimited causal polymorphism, while small vertical lines along the x-axis indicate different SNPs in the window that is shown. (F) Engineering the BY allele into *mrp20-105E* XIV^{3S} segregants changed growth (left), while substitutions at the nearest upstream and downstream variants did not (right).

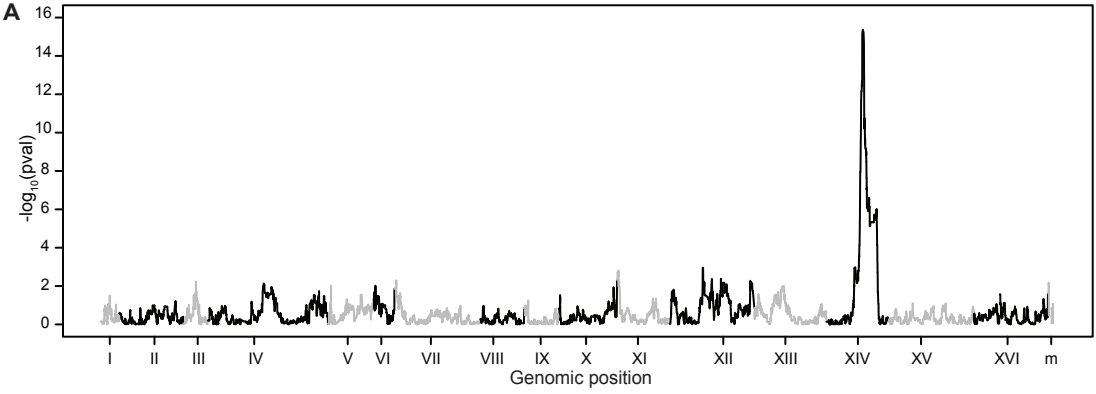
Fig. 3. Additional loci govern response to the mutation. (A) We engineered all combinations of *MRP20* and *MKT1* into the BY and 3S cross parents. Expected phenotypes are shown as shaded boxes denoting 95% confidence interval based on the originally obtained segregant phenotypes. (B) We generated BY x 3S crosses in which all segregants carried *mrp20-105E*. Two crosses were performed: one in which all segregants carried $MKT1^{BY}$ and one in which all segregants carried $MKT1^{3S}$. Tetrads were dissected and spores were phenotyped for growth on ethanol. (C) Each of the new crosses showed increased variance in growth that extended from inviable to wild type, differing from the more qualitative phenotypes seen among the original *hos3Δ mrp20-105E* segregant population. (D) Linkage mapping identified a total of 16 loci that influenced growth. After four iterations of a forward regression, no additional loci were identified. (E) Inviable segregants were present among all *mrp20-105E MKT1 SAL1* genotype classes. (F) Aneuploid individuals with duplicated Chromosome II showed reduced growth. Aneuploid individuals were not evenly detected across the different *MKT1-SAL1* genotype classes.

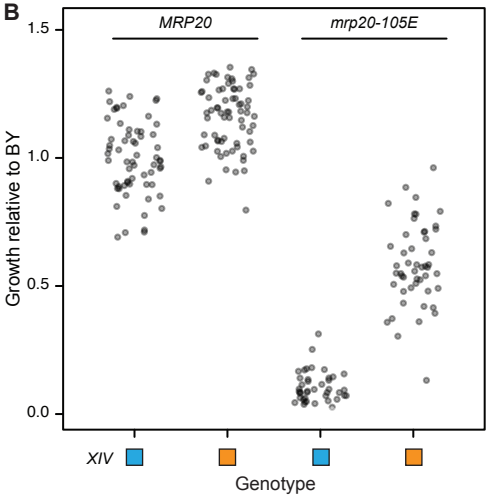
Fig. 4. Mitochondrial genome instability partially underlies the expressivity of *mrp20-105E*. We measured petite formation frequency, which estimates the proportion of cells within a clonal population capable of respiratory growth. Higher petite frequency is a proxy for greater mitochondrial genome instability. (A) We examined *MRP20* and *mrp20-105E* versions of the BY and 3S parent strains. For each, average values and 95% bootstrapped confidence intervals are shown. BY showed elevated mitochondrial genome instability in the presence of *mrp20-105E*, while 3S showed no change. (B) We examined 16 BYx3S *MRP20* segregants. These segregants were randomly selected and spanned the

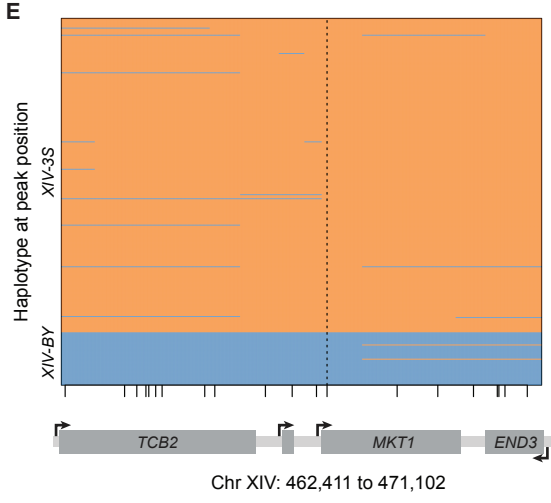
range of growth values for *MRP20* segregants. (C) 45 BYx3S *mrp20-105E* segregants. Poorer growing segregants tended to exhibit higher mitochondrial genome instability, though some exhibited wild type levels of mitochondrial genome instability. The gray dashed line indicates the threshold used to call inviability.

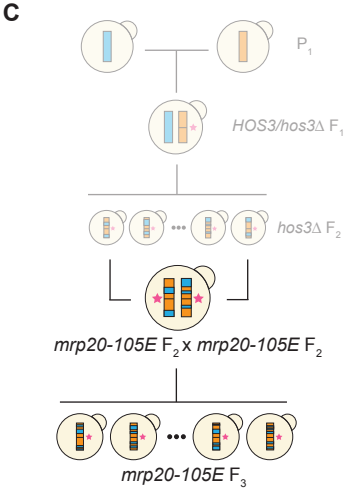
Fig. 5. Detected loci quantitatively and qualitatively explain mutant phenotypes. (A) We fit a linear model accounting for the effects of all detected loci and the aneuploidy on the growth of *mrp20-105E* segregants. This model not only explained the growth of the new BY x 3S *mrp20-105E* crosses generated in this paper, but also accurately predicted the phenotypes of the mutant parents and previously generated segregants. (B) We examined growth relative to the sum of detrimental alleles carried by a segregant. This relationship shows how collections of loci produce a quantitative spectrum of phenotypes, including instances of qualitative phenotypic responses. This relationship explains the full range of responses, from inviable to wild type growth, across *MKT1-SAL1* genotypes. The gray dashed line indicates the threshold used to call inviability.

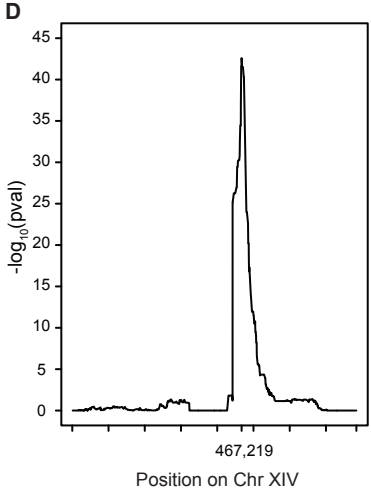

A
P1
BY
3S
mrp20-105E
HOS3/HOS3 F1
HOS3/hos3Δ F1
HOS3 F2
hos3Δ F2
B
Growth relative to BY
1.5
1.0
0.5
0.0
HOS3
hos3Δ
Genotype
C
-log10(pval)
50
40
30
20
10
0
I II III IV V VI VII VIII IX X XI XII XIII XIV XV XVI m
Genomic position
D
HOS3
hos3Δ
Growth relative to BY
1.5
1.0
0.5
0.0
IV
Genotype
E
Haplotype at peak position
IV-3S
IV-BY
DIT1
RPB7
MRP20
PDR15
Chr IV: 1,274,602 to 1,283,799
F
Growth relative to BY
1.5
1.0
0.5
0.0
MRP20 to mrp20-105E
mrp20-105E to MRP20
Genotype
G
HOS3
hos3Δ
Growth relative to BY
1.5
1.0
0.5
0.0
mrp20-105E
MRP20
mrp20-105E
MRP20
Genotype

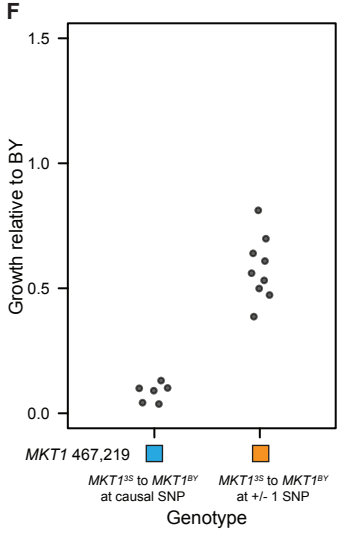

A
-log10(pval)
16
14
12
10
8
6
4
2
0
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI
m
Genomic position

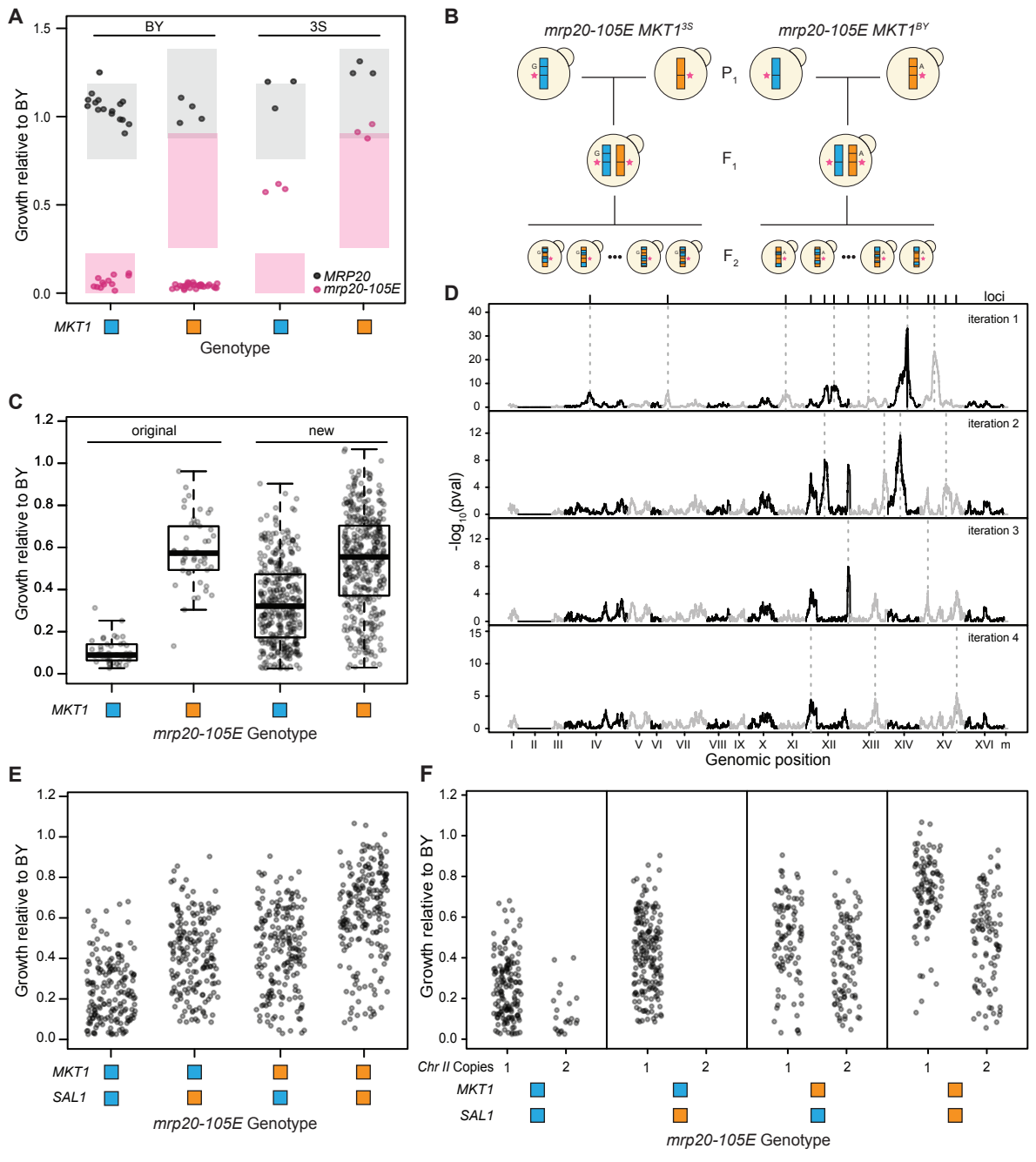

B
MRP20
mrp20-105E
Growth relative to BY
1.5
1.0
0.5
0.0
XIV
Genotype

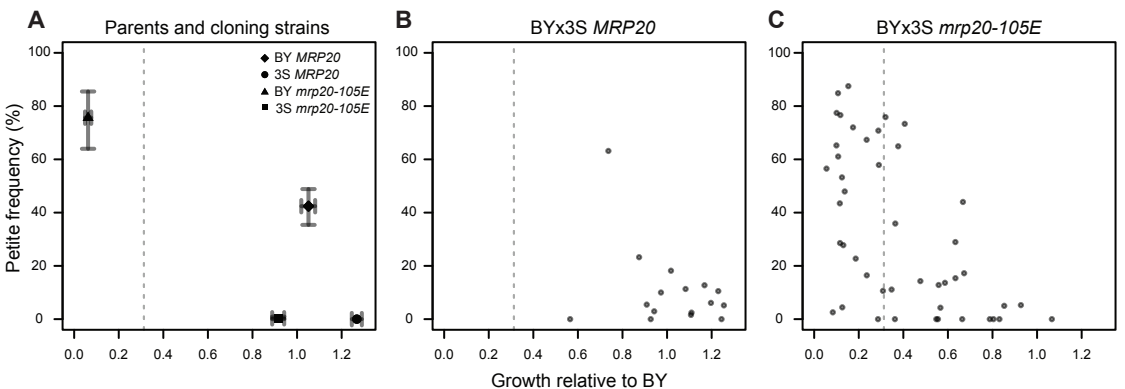

E
Haplotype at peak position
XIV-3S
XIV-BY
TCB2
MKT1
END3
Chr XIV: 462,411 to 471,102

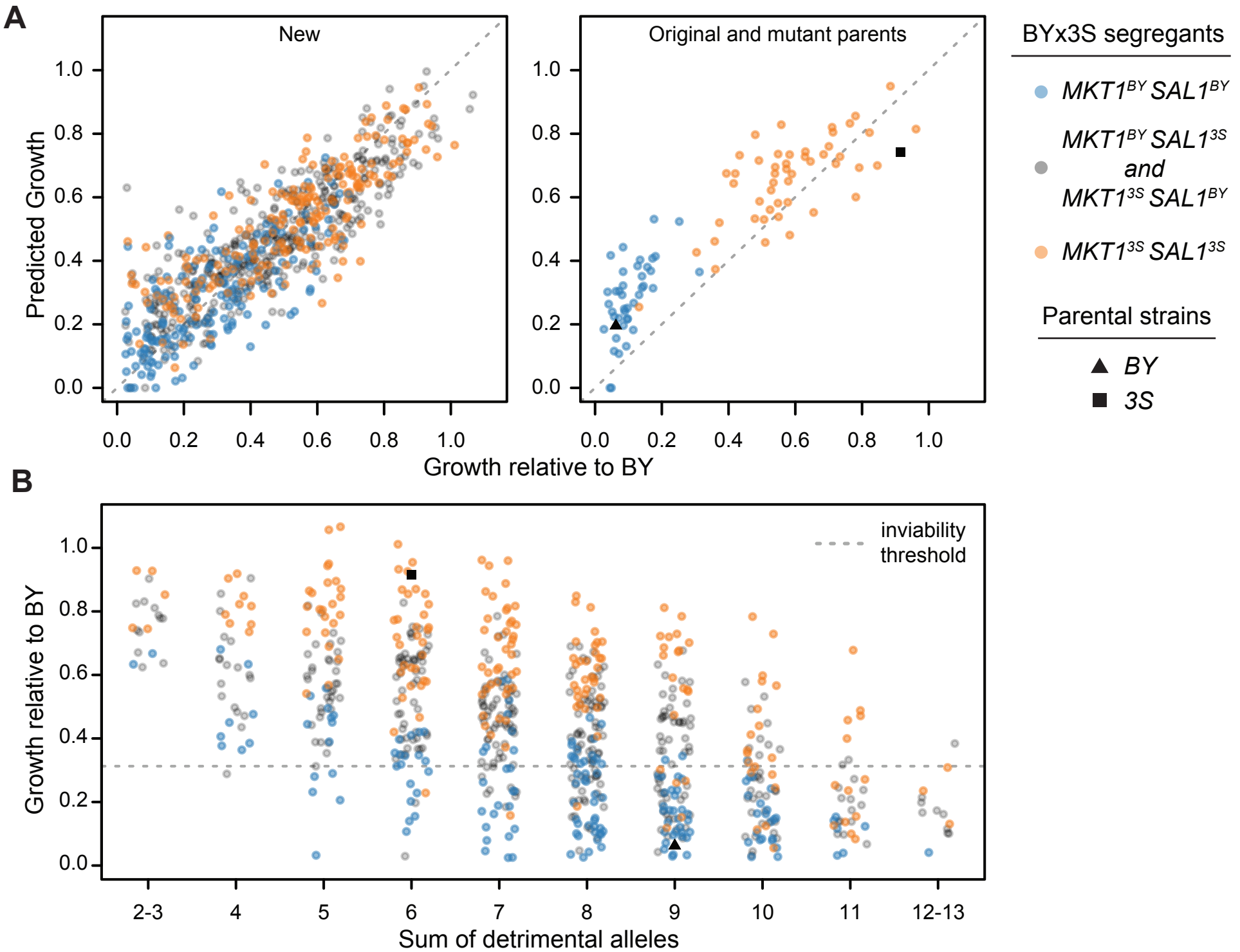

C
P1
HOS3/hos3Δ F1
hos3Δ F2
mrp20-105E F2 x mrp20-105E F2
mrp20-105E F3


D
-log10(pval)
45
40
35
30
25
20
15
10
5
0
467,219
Position on Chr XIV


F
Growth relative to BY
1.5
1.0
0.5
0.0
MKT1 467,219
MKT1 3S to MKT1 BY at causal SNP
MKT1 3S to MKT1 BY at +/- 1 SNP
Genotype


A
BY
3S
Growth relative to BY
MRP20
mrp20-105E
MKT1
Genotype
B
mrp20-105E MKT1^3S
mrp20-105E MKT1^BY
P1
F1
F2
C
original
new
Growth relative to BY
MKT1
mrp20-105E Genotype
D
loci
iteration 1
iteration 2
iteration 3
iteration 4
-log10(pval)
Genomic position
E
Growth relative to BY
MKT1
SAL1
mrp20-105E Genotype
F
Growth relative to BY
Chr II Copies
MKT1
SAL1
mrp20-105E Genotype


A
Parents and cloning strains
BY MRP20
3S MRP20
BY mrp20-105E
3S mrp20-105E
Petite frequency (%)
B
BYx3S MRP20
C
BYx3S mrp20-105E
Growth relative to BY


A
New
Original and mutant parents
Predicted Growth
Growth relative to BY
0.0
0.2
0.4
0.6
0.8
1.0
BYx3S segregants
MKT1BY SAL1BY
MKT1BY SAL13S and MKT13S SAL1BY
MKT13S SAL13S
Parental strains
BY
3S
B
inviability threshold
Growth relative to BY
Sum of detrimental alleles
2-3
4
5
6
7
8
9
10
11
12-13