Title: Idiosyncratic epistasis leads to global fitness-correlated trends

Authors: Christopher W. Bakerlee[†,1,2,3], Alex N. Nguyen Ba[†,1,2,4], Yekaterina Shulgina[3], Jose I. Rojas Echenique[1,6], and Michael M. Desai[*,1,2,5,7]

Affiliations:

[1]Department of Organismic and Evolutionary Biology, Harvard University; Cambridge, MA, USA.
[2]Quantitative Biology Initiative, Harvard University; Cambridge, MA, USA.
[3]Department of Molecular and Cellular Biology, Harvard University; Cambridge, MA, USA.
[4]Department of Cell and Systems Biology, University of Toronto; Toronto, Canada.
[5]NSF-Simons Center for Mathematical and Statistical Analysis of Biology, Harvard University; Cambridge, MA, USA.
[6]Department of Molecular Genetics, University of Toronto; Toronto, Canada.
[7]Department of Physics, Harvard University; Cambridge, MA, USA.

[†]These authors contributed equally to this work
[*]Corresponding author. Email: mdesai@oeb.harvard.edu

Abstract: Epistasis can dramatically affect evolutionary trajectories. In recent decades, protein-level fitness landscapes have revealed extensive idiosyncratic epistasis among specific mutations. In contrast, other work has found ubiquitous and apparently non-specific patterns of global diminishing-returns and increasing-costs epistasis among mutations across the genome. Here, we use a hierarchical CRISPR gene drive system to construct all combinations of 10 missense mutations from across the genome in budding yeast, and measure their fitness in six environments. We show that the resulting fitness landscapes exhibit global fitness-correlated trends, but that these trends emerge from specific idiosyncratic interactions. This provides the first experimental validation of recent theoretical work that has argued that fitness-correlated trends can emerge as the generic consequence of idiosyncratic epistasis.

One-Sentence Summary: A genome-spanning fitness landscape reveals how idiosyncratic genetic interactions lead to global epistatic patterns.

Main Text: Epistatic interactions have important consequences for the design and evolution of genetic systems (*1–3*). Significant work in recent decades has studied these interactions by measuring empirical fitness landscapes, most often at "shallow" depth for genome-scale studies (e.g., by quantifying pairwise but not higher order epistasis between all gene deletions or mutations) or at "narrow" breadth (such as complete landscapes at the scale of small select regions in single genes, for example by quantifying all orders of epistatic interactions among few amino acid residues) (*4–18*). These studies have often found many epistatic interactions among specific mutations at both lower (i.e., among few mutations) and higher orders (i.e., among many mutations). These reflect particular biological and physical interactions among the mutations involved; following recent work (*19, 20*) we refer to them as "idiosyncratic" epistasis, as they involve the specific details of these mutations. Overall, this body of work has highlighted the potential for epistasis to create historical contingency that tightly constrains the distribution of adaptive trajectories accessible to natural selection.

In contrast, other work examining adaptive trajectories that implicate loci across the genome has found patterns of apparently "global" epistasis, in which the fitness effect of a mutation varies systematically with the fitness of the genetic background on which it occurs (*21–28*). Typically, this manifests as either diminishing returns for beneficial mutations or increasing costs for deleterious mutations, with mutations having a less positive or more negative effect on fitter backgrounds. These consistent patterns of global epistasis may give rise to the dominant evolutionary trend of declining adaptability, and in contrast to the complexity of idiosyncratic interactions, they suggest that historical contingency could be less critical in constraining adaptive trajectories (*29*).

Despite their importance, these dual descriptions of epistasis have not been satisfactorily unified. In one view, global epistasis results from non-specific fitness-mediated interactions among mutations (*24*). Such interactions may for example emerge from the topology of metabolic networks, which generates overall patterns of diminishing returns and increasing costs that eclipse the specific details of idiosyncratic interactions (*30*). In contrast, other recent theoretical work has proposed an alternative view, hypothesizing that apparent fitness-mediated epistasis can instead emerge as the generic consequence of idiosyncratic interactions, provided they are sufficiently numerous and widespread (*19, 20*). These two models have substantially different implications for the structure of fitness landscapes, which in turn influence our expectations of the repeatability and predictability of evolution and of the effect of chance and contingency on adaptation at both the genotypic and phenotypic level. Thus, this dichotomy plays a central role in understanding of how epistasis affects evolutionary dynamics.

Thus far, however, empirical work has been unable to distinguish between these perspectives. The key difficulty is that testing these ideas requires both depth and breadth: we must analyze landscapes involving enough loci that we sample idiosyncratic interactions that can potentially give rise to overall fitness-mediated trends, and we must survey possible combinations of these mutations at sufficient depth to quantify the role of higher-order interactions (including potential "global" non-specific fitness-mediated interactions). Importantly, larger landscapes are also necessary to reduce the influence of measurement error on the inference of epistasis and analysis of fitness-correlated trends (see SI section 6.3). Achieving this depth and breadth is technically challenging, because it requires us to synchronize many mutations across the genome.

Here, we overcome this challenge by developing a novel method that exploits Cre-Lox recombination to create a combinatorially expanding CRISPR guide-RNA (gRNA) array in *Saccharomyces cerevisiae*, which allows us to iteratively generate mutations at distant loci via a gene drive mechanism (Fig. 1A). Briefly, strains of opposite mating type containing inducible Cre recombinase and *SpCas9* genes are mutated at one of two loci (*A* or *B*), and DNA encoding guide-RNAs (gRNAs) specific to the wild-type alleles at these loci are integrated into their genomes (Fig. S1). After mating to produce a diploid

heterozygous at *A* and *B*, we induce a gene drive to make the loci homozygous. This begins with expressing Cas9 and generating gRNA-directed double-strand breaks at the wild-type *A* and *B* alleles. These breaks are then repaired by the mutated regions of homologous chromosomes, making the diploid homozygous at these loci with at least 95% efficiency. Simultaneously, we express Cre to induce recombination that brings gRNAs into physical proximity on the same chromosome by way of flanking Lox sites, in a strategy similar to that described previously (*31*) (Fig. 1B). We then sporulate diploids and select haploids bearing the linked gRNAs from both parents. In parallel, we carry out this process with "pseudo-WT" versions of these loci, which contain synonymous changes that abolish gRNA recognition, but lack the non-synonymous change of interest. This creates a set of four strains, with all possible genotypes at loci *A* and *B*. Concurrently, we create separate sets of four strains with all possible genotypes at other pairs of loci (e.g., *C* and *D*).

By iterating this process, we can rapidly assemble an exponentially expanding, combinatorially complete genotype library. We mate separate sets of four genotypes bearing all combinations of mutations at two loci each in an all-against-all cross, drive their mutations, recombine their gRNAs, and sporulate to produce a 16-strain library bearing all 4-locus mutation combinations. Repeating these steps in a third cycle with two 4-locus libraries of opposite mating type yields a 256-strain 8-locus library, and a complete landscape of up to 16 mutations (2^{16} strains) can be constructed in just four cycles.

We sought to use this method to construct a complete fitness landscape that would shed light on the structure of epistasis: are fitness-correlated trends primarily the product of a global coupling of mutations via fitness, or do they emerge as the consequence of idiosyncratic epistasis? To do so, we surveyed studies of natural variation (e.g., (*32–36*)) and experimental evolution (e.g., (*37–39*)) to identify mutations relevant to adaptation in the laboratory strain. We selected a set of mutations that sample a wide range of cellular functions, such as membrane stress response, mitochondrial stability, and nutrient sensing. Our goal in making this choice was to maximize fitness variance while minimizing pathway-specific idiosyncratic interactions. We note that alternative choices of mutations, particularly if they were focused on a specific protein or pathway (or limited to those that accumulated along the line of descent in a single lineage), might exhibit very different patterns of epistasis, which would be characteristic of the particular details of that specific protein or pathway (or that specific adaptive trajectory). However, our goal here is to analyze potentially global patterns of epistasis among mutations across the genome that are relevant to adaptation and hence represent an overall fitness landscape for the laboratory strain.

We thus implemented our gene-drive system to construct a near-complete landscape spanning 10 missense mutations in 10 genes (including essential genes) on 8 chromosomes: *AKL1* (S176P), *BUL2* (L883F), *FAS1* (G588A), *MKT1* (D30G), *NCS2* (H71L), *PMA1* (S234C), *RHO5* (G10S), *RPI1* (E102D), *SCH9* (P220S), and *WHI2* (L262S) (Fig. 1C, Table S1). We found that a landscape of about this size is required to distinguish the two models (see SI section 6.3). Immediately before the final mating cycle, all strains were transformed with a unique DNA barcode next to the *LYS2* locus to enable high-throughput, sequencing-based competitive fitness assays (Fig. S2, S3). All strains in each replicate haploid library were genotyped at all 10 loci to confirm presence of the desired alleles (this also ensures presence in the diploid libraries). After excluding strains due to gene drive failure, 875 out of 1024 (85.4%) genotypes remained in at least one library (and 407 in both biological replicates). We also performed whole genome sequencing of 96 randomly selected strains to rule out pervasive aneuploidies or influential but spurious background mutations. One aneuploidy was identified, and 3 spurious background mutations were observed at >5% frequency. Subsequent analysis showed that these were unlikely to systematically influence our findings (Table S2, and SI section 5.1).

To obtain fitness landscapes, we conducted replicate bulk barcode-based fitness assays on both pooled haploid and homozygous diploid versions of the genotype library in 6 distinctly stressful media

environments: YPD + 0.4% acetic acid, YPD + 6 mM guanidium chloride, YPD + 35 μM suloctidil, YPD @ 37°C, YPD + 0.8 M NaCl, and SD + 10 ng/mL 4NQO (Fig. 1D). For each of 7 days, pools were allowed 7 generations of growth, and aliquots were sampled and sequenced at the barcode locus at generations 7, 14, 28, 42, and 49. We estimated the relative fitness of each genotype from changes in barcode frequencies through time, achieving consistent measurements across technical and biological replicates (Fig. 1E,F, S4). From these data, we inferred the background-averaged additive and epistatic effects of each mutation and combination of mutations, respectively (using LASSO regularization, see SI).

We found that our six environments yield substantially different landscapes, as demonstrated by the relatively low between-environment correlations of genotype fitnesses (Fig. 2A), the additive effects of each mutation (Fig. 2B), and the pairwise interactions between them (Fig. 2C). While haploid and homozygous diploid landscapes were largely correlated, there were several notable exceptions, particularly in the suloctidil environment (Fig. 2A,B). And while some pairwise interactions remain roughly constant in strength, even as the corresponding additive effects vary considerably (e.g., *RHO5* and *WHI2*), most wax and wane across environments (Fig. 2C). Nevertheless, the overall contribution from different epistatic orders shows some similarities across ploidies and environments (the magnitudes do differ; Fig. 2D), with additive and pairwise terms explaining most of the variance in the data, third-order terms contributing minorly, and the remaining orders making little difference, consistent with earlier studies (*40*). Across all epistatic orders, inferred effects were highly skewed, with a small number of terms explaining disproportionate variance (Fig. 2E).

We next sought to investigate potential patterns of global fitness-mediated epistasis. To do so, for each locus in each ploidy and environment, we plotted the fitness of a genotype with the mutated allele, φ_{Mut}, against the fitness of the same genotype with the WT allele, φ_{WT}. A regression slope, b, different from 1 in these plots signifies a fitness-correlated trend (FCT) (Fig. 3A, left; see SI). We note that some previous work has instead plotted the fitness effect of a mutation, $\Delta\varphi$, as a function of background fitness φ_{WT}. The advantage of our formulation here is that it does not privilege a specific allele as the "wild-type." Instead, regression in our plots translates intuitively when reversing direction to treat the reversion as the mutation: $b_{rev} = 1/b_{orig}$ by weighted-total least squares; see expanded discussion in the SI, Fig. S5-S8.

We found that FCTs are common in our landscapes: across all ploidies, environments, and loci, ~44% of regression slopes deviate substantially from 1 (i.e., $b \leq 0.9$ or $b \geq 0.9^{-1}$; these deviations are all significant; Fig. 3B, see histogram; Fig. S13 and S14). However, FCTs were not universal for fitness-affecting mutations: of the 49 examples across ploidies and environments of mutations with additive effects of magnitude ≥ 0.5%, 18 were associated with $0.9 < b < 0.9^{-1}$ (Fig. 3B).

By partitioning background genotypes by the presence or absence of specific mutations, we can determine whether FCTs are truly "global" (i.e., they transcend these partitions and any corresponding idiosyncratic interactions; Fig. 3A, middle), or are instead fundamentally idiosyncratic (i.e., they emerge from regression across partitions shifted in φ_{Mut} versus φ_{WT} space by sparse interactions with specific background loci; Fig. 3A, right). When we partitioned FCTs by the presence or absence of interacting mutations in the background, we found several instances where the idiosyncratic model clearly explains the fitness-correlated trend. For example, the effect of the G10S mutation in *RHO5* at 37°C exhibits a clear FCT ($b = 0.76$) (Fig. 3C). However, we can partition points by the presence of interacting *WHI2* and *AKL1* alleles in the background. Doing so shows that pairwise interactions with these alleles cause systematic shifts in φ_{10S} vs φ_{10G} space, with each partition assuming a slope near 1. Thus, over a range of background fitnesses, a FCT in the effect of the G10S emerges from these specific idiosyncratic interactions (Fig. 3C, S11). In the case of the homozygous *AKL1* S176P mutation in suloctidil, we observe a similar decomposition of a FCT ($b = 1.29$) when partitioning genotypes according to the

presence of three interacting loci in the background (*MKT1*, *RHO5*, and *WHI2*) (Fig. 3D, S11). However, in other cases it is less clear whether the FCT can be partitioned in this way, and since deeper partitions tend to reduce background fitness variance and limit our confidence in regression slopes, a different approach is required to characterize the extent to which idiosyncratic terms cause FCTs across our data.

To investigate this question, we therefore analyzed the effect of removing specific idiosyncratic epistatic terms on the overall fitness-correlated trends. To do so, for each focal locus (in each ploidy and environment) we first calculated the weighted sum of squared errors (*41*) of observed fitnesses from the global regression line ($SSE_{b=global}$) and from a fitted line of slope 1 ($SSE_{b=1}$, which corresponds to no FCT). We then set the largest epistatic term to zero and recalculated the expected fitness of each resulting genotype (assuming all other terms and residuals are non-zero), again obtaining both $SSE_{b=global}$ and $SSE_{b=1}$. If the fitness-correlated trend arose from a global effect, we expect that $SSE_{b=global}$ would be less than $SSE_{b=1}$ even as terms are removed. Instead, we found that, after removing the effect of just a few terms, a regression with a fixed slope of b=1 typically fit the data better than the b=global FCT slope (Fig. 3E, S11, with FCT threshold set to $b \leq 0.9$ or 0.8)), approaching the fit of an unconstrained regression that minimizes SSE (i.e., the final slope approaches 1, Fig. S10). This indicates that the apparent FCT arises from these few idiosyncratic interactions, even for global slopes very different from 1. While we also documented cases where b=global fit the data better than b=1 even after removing many terms, we expect most if not all these instances may be due to measurement error, since they tend to arise in ploidies and environments where the data is noisier (Fig. S17).

To further evaluate whether idiosyncratic interactions between these mutations are sufficient to generate FCTs, we performed the converse analysis, this time with genotype fitnesses as predicted by our model of additive and idiosyncratic epistatic terms. Instead of removing the effects of epistatic terms one at a time, we first stripped from the model all interactions involving the focal locus, yielding perfectly linear points of slope 1 when plotting φ_{Mut} vs φ_{WT}. We then added interactions one by one to our fitness prediction, from largest to smallest, and examined the resulting slopes. As shown in Fig. 3F for the haploid *PMA1* S234C mutation in 4NQO, adding just a handful of terms associated with 3 background loci recapitulates a strong FCT. Repeating this analysis with all our mutations shows that, on average, just 4 idiosyncratic interactions (primarily pairwise) are sufficient to recapitulate the full-model FCTs (a slope within 0.01 of the global slope, Fig. 3G, orange; see SI), which is far lower than the total number of inferred terms (median of 53) but represents on average 89% of the potential variance explained that could have been added (Fig. S12). Thus, while fitness-correlated trends are real and likely have important biological consequences, our data demonstrate that apparent fitness-mediated epistasis can readily emerge from remarkably few low-order idiosyncratic interactions.

Since the landscapes we study here have no natural polarization (i.e., neither allele is the assumed wildtype), we cannot comment directly on why earlier studies of global epistasis have more commonly found negative than positive FCTs (when plotting $\Delta\varphi$ versus φ_{WT}). However, this distribution of FCT directions is important because it may underly the ubiquitous trend of declining adaptability observed across laboratory evolution experiments (*29*). The observed bias towards negative trends may arise from asymmetries in the average sign of epistatic interactions between mutations away from extant high-fitness genotypes relative to their reversions, which theory has predicted should arise from idiosyncratic interactions (*19*, *20*). In addition, we note that choosing polarizations at random will lead to more negative than positive FCTs across the full parameter space (see extended discussion in the SI).

Regardless of the cause of any asymmetry in the direction of fitness-correlated trends, our results support recent theoretical arguments that fitness-mediated epistasis can emerge as the generic consequence of widespread idiosyncratic interactions, rather than reflecting a global fitness-mediated coupling of mutations. Indeed, at least in our system, we see that fitness-correlated trends can arise even from a

relatively small number of low-order interactions. We note that landscapes involving other types of variation (e.g., within a single protein or pathway or along the line of descent in a single lineage (*21*)) may exhibit different patterns, though we may expect these scenarios to involve an even stronger role for idiosyncratic interactions. More generally, we emphasize that idiosyncratic epistasis and global fitness-mediated effects are not mutually exclusive, and while fitness-correlated trends can be explained by the former in our system, in other cases both effects may contribute. However, our results suggest that nonspecific global epistasis may not be the primary driver of patterns of declining adaptability in laboratory evolution experiments, and this has general implications for the ways in which epistasis constrains evolutionary trajectories.

References and Notes

1. D. J. Kvitek, G. Sherlock, Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet*. **7**, e1002056 (2011).

2. P. A. Romero, F. H. Arnold, Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol*. **10**, 866–876 (2009).

3. K. M. Flynn, T. F. Cooper, F. B.-G. Moore, V. S. Cooper, The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS Genet*. **9**, e1003426 (2013).

4. M. Costanzo, B. VanderSluis, E. N. Koch, A. Baryshnikova, C. Pons, G. Tan, W. Wang, M. Usaj, J. Hanchard, S. D. Lee, V. Pelechano, E. B. Styles, M. Billmann, J. van Leeuwen, N. van Dyk, Z.-Y. Lin, E. Kuzmin, J. Nelson, J. S. Piotrowski, T. Srikumar, S. Bahr, Y. Chen, R. Deshpande, C. F. Kurat, S. C. Li, Z. Li, M. M. Usaj, H. Okada, N. Pascoe, B.-J. San Luis, S. Sharifpoor, E. Shuteriqi, S. W. Simpkins, J. Snider, H. G. Suresh, Y. Tan, H. Zhu, N. Malod-Dognin, V. Janjic, N. Przulj, O. G. Troyanskaya, I. Stagljar, T. Xia, Y. Ohya, A.-C. Gingras, B. Raught, M. Boutros, L. M. Steinmetz, C. L. Moore, A. P. Rosebrock, A. A. Caudy, C. L. Myers, B. Andrews, C. Boone, A global genetic interaction network maps a wiring diagram of cellular function. *Science*. **353** (2016), doi:10.1126/science.aaf1420.

5. A. D. Norris, X. Gracida, J. A. Calarco, CRISPR-mediated genetic interaction profiling identifies RNA binding proteins controlling metazoan fitness. *Elife*. **6** (2017), doi:10.7554/eLife.28129.

6. M. A. Horlbeck, A. Xu, M. Wang, N. K. Bennett, C. Y. Park, D. Bogdanoff, B. Adamson, E. D. Chow, M. Kampmann, T. R. Peterson, K. Nakamura, M. A. Fischbach, J. S. Weissman, L. A. Gilbert, Mapping the Genetic Landscape of Human Cells. *Cell*. **174**, 953-967.e22 (2018).

7. K. S. Sarkisyan, D. A. Bolotin, M. V. Meer, D. R. Usmanova, A. S. Mishin, G. V. Sharonov, D. N. Ivankov, N. G. Bozhanova, M. S. Baranov, O. Soylemez, N. S. Bogatyreva, P. K. Vlasov, E. S. Egorov, M. D. Logacheva, A. S. Kondrashov, D. M. Chudakov, E. V. Putintseva, I. Z. Mamedov, D. S. Tawfik, K. A. Lukyanov, F. A. Kondrashov, Local fitness landscape of the green fluorescent protein. *Nature*. **533**, 397–401 (2016).

8. C. A. Olson, N. C. Wu, R. Sun, A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr Biol*. **24**, 2643–2651 (2014).

9. S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D. S. Tawfik, Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*. **444**, 929–932 (2006).

10. N. C. Wu, J. Otwinowski, A. J. Thompson, C. M. Nycholat, A. Nourmohammad, I. A. Wilson, Major antigenic site B of human influenza H3N2 viruses has an evolving local fitness landscape. *Nat Commun*. **11**, 1233 (2020).

11. V. O. Pokusaeva, D. R. Usmanova, E. V. Putintseva, L. Espinar, K. S. Sarkisyan, A. S. Mishin, N. S. Bogatyreva, D. N. Ivankov, A. V. Akopyan, S. Y. Avvakumov, I. S. Povolotskaya, G. J. Filion, L. B. Carey, F. A. Kondrashov, An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLoS Genet*. **15**, e1008079 (2019).

12. J. I. Jiménez, R. Xulvi-Brunet, G. W. Campbell, R. Turk-MacLeod, I. A. Chen, Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc Natl Acad Sci U S A*. **110**, 14984–14989 (2013).

13. B. Madan, B. Zhang, K. Xu, C. W. Chao, S. O'Dell, J. R. Wolfe, G.-Y. Chuang, A. S. Fahad, H. Geng, R. Kong, M. K. Louder, T. D. Nguyen, R. Rawi, A. Schön, Z. Sheng, R. Nimrania, Y. Wang, T. Zhou, B. C. Lin, N. A. Doria-Rose, L. Shapiro, P. D. Kwong, B. J. DeKosky, Mutational fitness landscapes reveal genetic and structural improvement pathways for a vaccine-elicited HIV-1 broadly neutralizing antibody. *Proc Natl Acad Sci U S A*. **118** (2021), doi:10.1073/pnas.2011653118.

14. N. C. Wu, L. Dai, C. A. Olson, J. O. Lloyd-Smith, R. Sun, Adaptation in protein fitness landscapes is facilitated by indirect paths. *Elife*. **5** (2016), doi:10.7554/eLife.16965.

15. C. Bank, S. Matuszewski, R. T. Hietpas, J. D. Jensen, On the (un)predictability of a large intragenic fitness landscape. *Proc Natl Acad Sci U S A*. **113**, 14085–14090 (2016).

16. J. Domingo, G. Diss, B. Lehner, Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature*. **558**, 117–121 (2018).

17. J. Teyra, A. Ernst, A. Singer, F. Sicheri, S. S. Sidhu, Comprehensive analysis of all evolutionary paths between two divergent PDZ domain specificities. *Protein Sci*. **29**, 433–442 (2020).

18. T.-L. V. Lite, R. A. Grant, I. Nocedal, M. L. Littlehale, M. S. Guo, M. T. Laub, Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *Elife*. **9** (2020), doi:10.7554/eLife.60924.

19. D. M. Lyons, Z. Zou, H. Xu, J. Zhang, Idiosyncratic epistasis creates universals in mutational effects and evolutionary trajectories. *Nat Ecol Evol*. **4**, 1685–1693 (2020).

20. G. Reddy, M. M. Desai, Global epistasis emerges from a generic model of a complex trait. *Elife*. **10** (2021), doi:10.7554/eLife.64740.

21. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*. **332**, 1193–1196 (2011).

22. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*. **332**, 1190–1192 (2011).

23. L. Perfeito, A. Sousa, T. Bataillon, I. Gordo, Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution*. **68**, 150–162 (2014).

24. S. Kryazhimskiy, D. P. Rice, E. R. Jerison, M. M. Desai, Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*. **344**, 1519–1522 (2014).

25. M. S. Johnson, A. Martsul, S. Kryazhimskiy, M. M. Desai, Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science*. **366**, 490–493 (2019).

26. X. Wei, J. Zhang, Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. *Mol Biol Evol*. **36**, 1008–1021 (2019).

27. S. Schoustra, S. Hwang, J. Krug, J. A. G. M. de Visser, Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. *Proc Biol Sci*. **283** (2016), doi:10.1098/rspb.2016.1376.

28. R. C. MacLean, G. G. Perron, A. Gardner, Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in Pseudomonas aeruginosa. *Genetics*. **186**, 1345–1354 (2010).

29. A. Couce, O. A. Tenaillon, The rule of declining adaptability in microbial evolution experiments. *Front Genet*. **6**, 99 (2015).

30. S. Kryazhimskiy, Emergence and propagation of epistasis in metabolic networks. *Elife*. **10** (2021), doi:10.7554/eLife.60200.

31. A. N. Nguyen Ba, I. Cvijović, J. I. Rojas Echenique, K. R. Lawrence, A. Rego-Costa, X. Liu, S. F. Levy, M. M. Desai, High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*. **575**, 494–499 (2019).

32. H. Sinha, L. David, R. C. Pascon, S. Clauder-Münster, S. Krishnakumar, M. Nguyen, G. Shi, J. Dean, R. W. Davis, P. J. Oefner, J. H. McCusker, L. M. Steinmetz, Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics*. **180**, 1661–1670 (2008).

33. S. W. Doniger, H. S. Kim, D. Swain, D. Corcuera, M. Williams, S.-P. Yang, J. C. Fay, A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet*. **4**, e1000183 (2008).

34. E. X. Kwan, E. Foss, L. Kruglyak, A. Bedalov, Natural polymorphism in BUL2 links cellular amino acid availability with chronological aging and telomere maintenance in yeast. *PLoS Genet*. **7**, e1002250 (2011).

35. J. S. Bloom, I. M. Ehrenreich, W. T. Loo, T.-L. V. Lite, L. Kruglyak, Finding the sources of missing heritability in a yeast cross. *Nature*. **494**, 234–237 (2013).

36. D. M. Wloch-Salamon, K. Tomala, D. Aggeli, B. Dunn, Adaptive Roles of SSY1 and SIR3 During Cycles of Growth and Starvation in Saccharomyces cerevisiae Populations Enriched for Quiescent or Nonquiescent Cells. *G3 (Bethesda)*. **7**, 1899–1911 (2017).

37. B. Szamecz, G. Boross, D. Kalapis, K. Kovács, G. Fekete, Z. Farkas, V. Lázár, M. Hrtyan, P. Kemmeren, M. J. A. Groot Koerkamp, E. Rutkai, F. C. P. Holstege, B. Papp, C. Pál, The genomic landscape of compensatory evolution. *PLoS Biol*. **12**, e1001935 (2014).

38. L. M. Kohn, J. B. Anderson, The underlying structure of adaptation under strong selection in 12 experimental yeast populations. *Eukaryot Cell*. **13**, 1200–1206 (2014).

39. S. Venkataram, B. Dunn, Y. Li, A. Agarwala, J. Chang, E. R. Ebel, K. Geiler-Samerotte, L. Hérissant, J. R. Blundell, S. F. Levy, D. S. Fisher, G. Sherlock, D. A. Petrov, Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell*. **166**, 1585-1596.e22 (2016).

40. D. M. Weinreich, Y. Lan, J. Jaffe, R. B. Heckendorn, The Influence of Higher-Order Epistasis on Biological Fitness Landscape Topography. *J Stat Phys*. **172**, 208–225 (2018).

41. D. York, N. M. Evensen, M. L. Martínez, J. De Basabe Delgado, Unified equations for the slope, intercept, and standard errors of the best straight line. *American Journal of Physics*. **72**, 367–375 (2004).

Acknowledgments: We thank the Bauer Core facility at Harvard, Gautam Reddy, Boris Shraiman, and members of the Desai lab for experimental assistance and comments on the manuscript. We thank Artur Rego-Costa for his help in creating Fig. 2C. Computational work was performed on the Odyssey cluster supported by the Research Computing Group at Harvard University.

Funding:

Department of Defense (DoD), National Defense Science & Engineering Graduate (NDSEG) Fellowship Program (CWB)
Natural Sciences and Engineering Research Council of Canada, Postdoctoral fellowship, Discovery Grant RGPIN-2021-02716, and Discovery Launch supplement DGECR-2021-00117 (ANNB)
National Human Genome Research Institute of the National Institutes of Health award F31-HG010984 (YS)
National Science Foundation grant PHY-1914916 (MMD)
National Institutes of Health grant GM104239 (MMD)

Author contributions:

Conceptualization: ANNB, CWB, YS, JIR, MMD
Methodology: ANNB, CWB, YS, JIR
Investigation: CWB, ANNB
Formal analysis: CWB, ANNB
Visualization: CWB, ANNB
Funding acquisition: MMD
Resources: MMD
Supervision: MMD
Writing – original draft: CWB, ANNB, MMD
Writing – review & editing: CWB, ANNB, MMD

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data described in the paper are presented in the supplementary materials. Raw sequencing data will be made publicly available at the NCBI Sequence Read Archive (accession no. TBD), and all analysis code will be made available from Github. Strains used in this study are available upon request.

Supplementary Materials

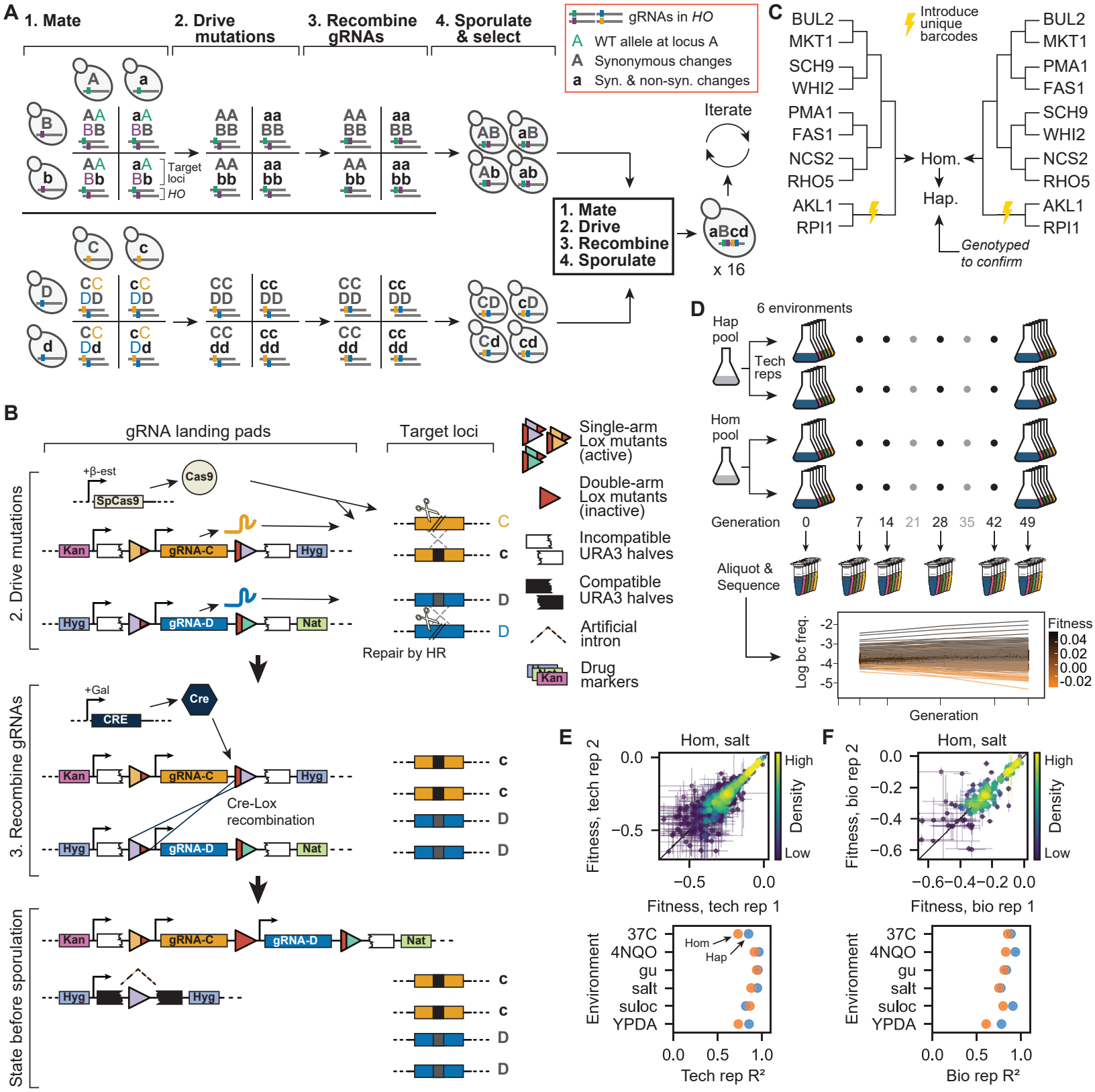
Materials and Methods
Supplementary Text
Figs. S1 to S17
Table S1
References (*1–41*)
Data S1 to S3

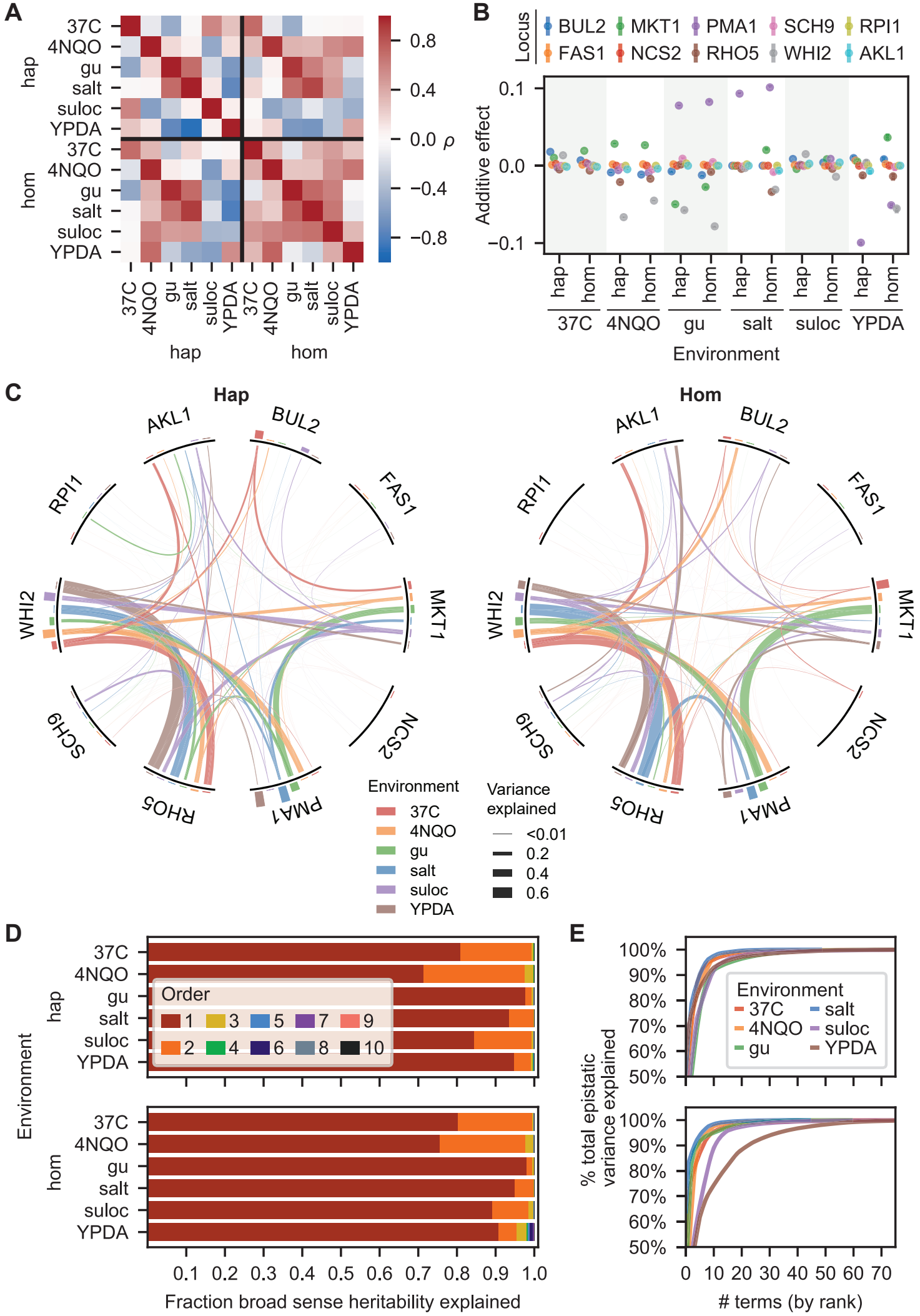
Figure Captions

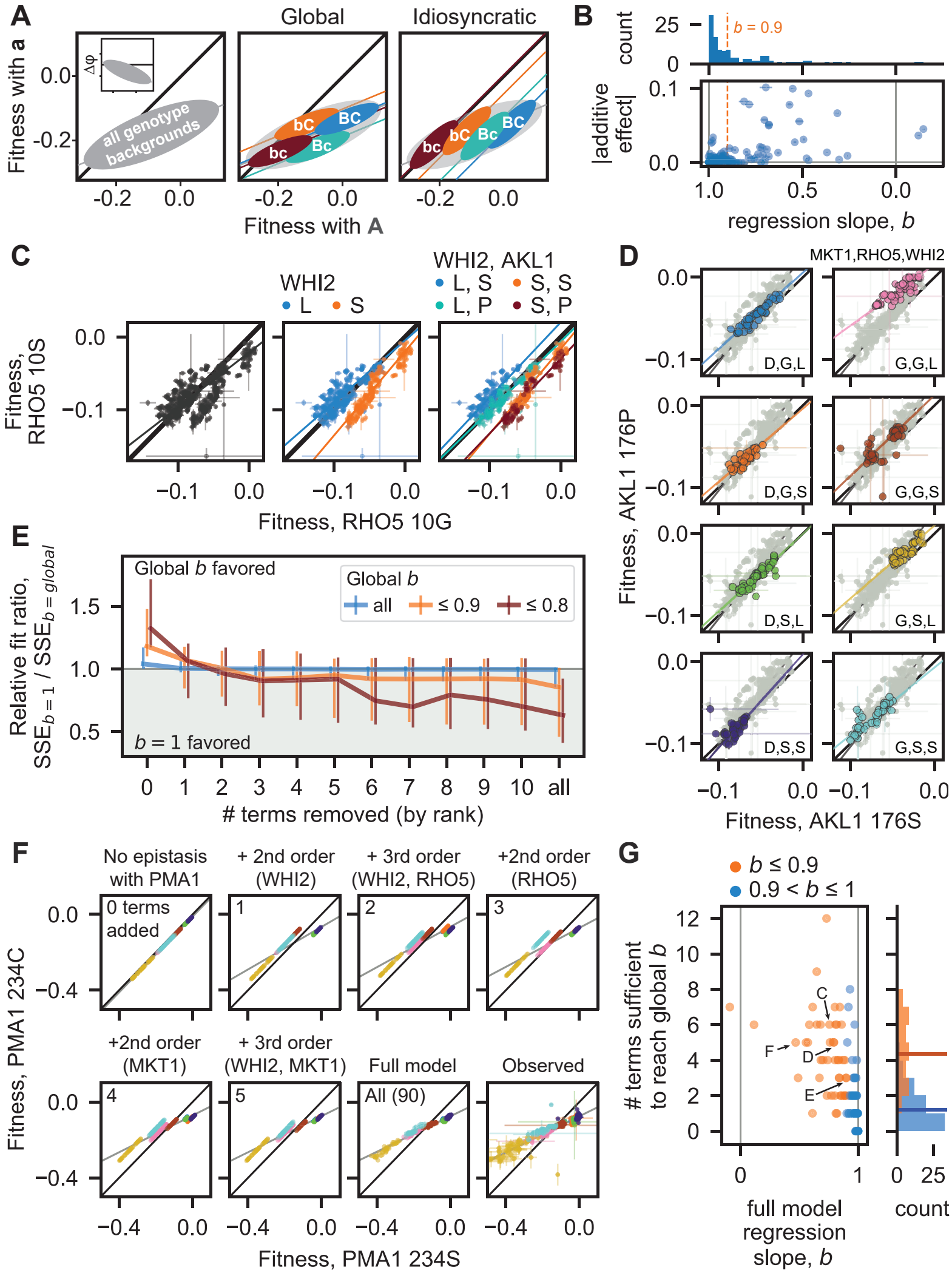
Fig. 1. Recombining CRISPR-gene drive system. (A) Experimental design. Strains of opposite mating type carrying known mutations and corresponding guide-RNAs (gRNAs) mate to form heterozygous diploids. Cas9 expression "drives" these mutations, and site-specific recombination links gRNAs. Homozygous diploids are sporulated, haploids with linked gRNAs are selected, and the process repeats, incorporating exponentially increasing numbers of mutations. **(B)** Recombining gene drive system. gRNAs targeting heterozygotic loci are flanked by selection markers and two of three orthogonal Lox sites (colored triangles), which are inactivated through recombination (red triangles). Cas9 "drives" targeted mutations, while Cre-Lox recombination brings like markers to the same chromosome and activates a *URA3* gene interrupted by an artificial intron. Following sporulation, the chromosome with gRNAs is selected using the markers of interest while the other is counterselected using 5-FOA. **(C)** Cross design. A complete fitness landscape is produced in parallel by distinct cross designs that yield final homozygous diploids and haploids in biological replicates with unique DNA barcodes. **(D)** Bulk-fitness assays. Pooled strains are assayed in replicate for competitive fitness in several environments by sequencing barcodes to obtain strain frequencies over time. **(E)** Repeatability of technical replicate competitive fitness measurements. **(F)** Repeatability of biological replicate competitive fitness measurements.

Fig. 2. Fitness landscapes. (A) Correlation in observed fitness (upper right) and predicted fitness (from inferred model, lower left, see SI section 5.1) across ploidies and environments. **(B)** Background-averaged additive effect of each locus across ploidies and environments. Error bars represent 95% confidence intervals. **(C)** Background-averaged pairwise epistatic effects between loci across ploidies and environments. Weights of edges connecting loci represent the proportion of pairwise variance explained by each interaction. Heights of bars on the perimeter correspond to the proportion of additive variance explained by each locus in each environment. **(D)** Variance partitioning of broad-sense heritability from additive and epistatic orders across ploidies and environments. **(E)** Cumulative distribution of the epistatic variance explained by rank-ordered epistatic terms of all orders.

Fig. 3. Fitness-correlated trends (FCTs). (A) Schematic contrasting how global or idiosyncratic epistasis could produce FCTs. Inset shows FCT analyzed as the effect of a mutation ($\Delta\varphi$) on backgrounds of different fitnesses. **(B)** Histogram and scatterplot of regression slopes, b, between φ_{Mut} and $\varphi_{WT,}$ and corresponding absolute additive effects of mutations. Polarity adopted such that $b \leq 1$. Total error bar length is twice the standard error of the slope. **(C)** Fitness effect of *RHO5* mutation (G10S) (φ_{Mut} versus φ_{WT}) in all haploid backgrounds at 37°C (left) and partitioned by genotypes at *WHI2* (L262S) (middle) and *WHI2* and *AKL1* (S176P) (right). Initial $SSE_{b=1}$ / $SSE_{b=global}$ is 1.21. **(D)** Fitness effect of *AKL1* mutation in all homozygote backgrounds in the suloctidil environment, partitioned by genotypes at *MKT1* (D30G), *RHO5*, and *WHI2*. Initial $SSE_{b=1}$ / $SSE_{b=global}$ is 1.31. **(E)** Median relative fit ratio between regressions with fixed slope of b=1 and b=global, as function of number of epistatic terms removed from observed phenotypes. Vertical lines represent interquartile range. Polarity adopted such that $b \leq 1$. **(F)** Inferred fitness effect of *PMA1* S234C mutation in 4NQO environment across all haploid backgrounds. Epistatic terms interacting with *PMA1* are completely removed from genotype fitnesses, then added back sequentially (from largest to smallest). Bottom-right: full-model (inferred) and observed genotype fitnesses, respectively. Grey line is regression slope. **(G)** Scatterplot and histograms of FCT regression slopes for all data, and number of epistatic terms sufficient to recapitulate them. Horizontal lines in histogram indicate means. Arrows, letters indicate populations presented in previous panels. Polarity adopted such that $b \leq 1$.


A
1. Mate
2. Drive mutations
3. Recombine gRNAs
4. Sporulate & select
gRNAs in HO
A WT allele at locus A
A Synonymous changes
a Syn. & non-syn. changes
Target loci
HO
Iterate
1. Mate
2. Drive
3. Recombine
4. Sporulate
aBcd
x 16
B
gRNA landing pads
Target loci
2. Drive mutations
+β-est
SpCas9
Cas9
Kan
gRNA-C
Hyg
gRNA-D
Nat
Repair by HR
Single-arm Lox mutants (active)
Double-arm Lox mutants (inactive)
Incompatible URA3 halves
Compatible URA3 halves
Artificial intron
Drug markers
3. Recombine gRNAs
+Gal
CRE
Cre
Cre-Lox recombination
State before sporulation
C
BUL2
MKT1
SCH9
WHI2
PMA1
FAS1
NCS2
RHO5
AKL1
RPI1
Introduce unique barcodes
Hom.
Hap.
Genotyped to confirm
D
6 environments
Hap pool
Tech reps
Hom pool
Generation
0
7
14
21
28
35
42
49
Aliquot & Sequence
Log bc freq.
Fitness
0.04
0.02
0.00
-0.02
E
Hom, salt
Fitness, tech rep 2
Fitness, tech rep 1
High
Density
Low
Environment
37C
4NQO
gu
salt
suloc
YPDA
Hom
Hap
Tech rep R²
F
Hom, salt
Fitness, bio rep 2
Fitness, bio rep 1
Bio rep R²


A
37C
4NQO
gu
salt
suloc
YPDA
hap
hom
0.8
0.4
0.0 ρ
−0.4
−0.8
B
Locus
BUL2
MKT1
PMA1
SCH9
RPI1
FAS1
NCS2
RHO5
WHI2
AKL1
Additive effect
0.1
0.0
−0.1
Environment
C
Hap
Hom
AKL1
BUL2
FAS1
MKT1
NCS2
PMA1
RHO5
SCH9
WHI2
RPI1
Environment
37C
4NQO
gu
salt
suloc
YPDA
Variance explained
<0.01
0.2
0.4
0.6
D
Order
1
2
3
4
5
6
7
8
9
10
Fraction broad sense heritability explained
E
Environment
% total epistatic variance explained
100%
90%
80%
70%
60%
50%
terms (by rank)


A
Global
Idiosyncratic
Fitness with a
Fitness with A
all genotype backgrounds
bc
bC
Bc
BC
B
count
b = 0.9
|additive effect|
regression slope, b
C
WHI2
L
S
WHI2, AKL1
L, S
S, S
L, P
S, P
Fitness, RHO5 10S
Fitness, RHO5 10G
D
MKT1,RHO5,WHI2
D,G,L
G,G,L
D,G,S
G,G,S
D,S,L
G,S,L
D,S,S
G,S,S
Fitness, AKL1 176P
Fitness, AKL1 176S
E
Global b favored
Global b
all
≤ 0.9
≤ 0.8
b = 1 favored
Relative fit ratio, SSE_{b=1} / SSE_{b=global}
terms removed (by rank)
all
F
No epistasis with PMA1
+ 2nd order (WHI2)
+ 3rd order (WHI2, RHO5)
+2nd order (RHO5)
0 terms added
+2nd order (MKT1)
+ 3rd order (WHI2, MKT1)
Full model
Observed
All (90)
Fitness, PMA1 234C
Fitness, PMA1 234S
G
b ≤ 0.9
0.9 < b ≤ 1
terms sufficient to reach global b
full model regression slope, b
count