Quantifying fitness distributions and phenotypic relationships in recombinant yeast populations

Ethan O. Perlstein*†‡, Eric J. Deeds*‡, Orr Ashenberg§, Eugene I. Shakhnovich§, and Stuart L. Schreiber†‡§

*Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138; §Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138; and †Howard Hughes Medical Institute, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142

Contributed by Stuart L. Schreiber, May 2, 2007 (sent for review August 31, 2006)

Studies of the role of sex in evolution typically involve a longitudinal comparison of a single ancestor to several intermediate descendants and to one terminally evolved descendant after many generations of adaptation under a given selective regime. Here we take a complementary, statistical approach to sex in evolution, by describing the distribution of phenotypic similarity in a population of yeast F_1 meiotic recombinants. By applying graph theory to fitness measurements of thousands of *Saccharomyces cerevisiae* recombinants treated with 10 mechanistically distinct, growth-inhibitory small-molecule perturbagens (SMPs), we show that the network of phenotypic similarity among F_1 recombinants exhibits a scale-free degree distribution. F_1 recombinants are often phenotypically unique and sometimes exceptional, and their fitness strengths are unevenly distributed across the 10 compound treatments. By contrast, highly phenotypically similar F_1 recombinants constitute failing hubs that display below-average fitness across all compound treatments and are candidate substrates for purifying selection. Comparison of the F_1 generation with the parental strains reveals that (*i*) there is a specialist more fit in any given single condition than any of the parents but (*ii*) only rarely are there generalists that exhibit greater fitness than both parental strains across a majority of conditions. This analysis allows us to evaluate and to gain better theoretical understanding of the costs and benefits of sex in the F_1 generation.

graph theory | meiotic recombination

Both theory and experiment suggest that (*i*) the benefits of sex (meiosis) outweigh its costs (1, 2) and (*ii*) organisms that reproduce sexually evolve faster than their asexual counterparts (3, 4). The essence of the Weismann hypothesis is that recombination, a combinatorial process by which sexual species access novel, beneficial arrangement of polymorphisms while at the same time sloughing off harmful mutations or allelic combinations, leads to an increase in the variance of fitness in sexually reproducing populations (5). However, sex is believed to doom favorable parental allelic combinations and incur additional costs (6). There have been many attempts to reconcile the costs and benefits of sex. Two recent studies (7, 8) illustrate the benefits of sex. We provide additional evidence for its benefits by describing a combined experimental/theoretic approach that reveals the degree distribution of phenotypic similarity of thousands of *Saccharomyces cerevisiae* F_1 meiotic recombinants derived from an outbred genetic cross. Our approach complements previous studies in two ways. First, we take fitness measurements of random meiotic recombinants in 10 diverse environmental conditions after an initial selection on a neutral medium. In contrast, other studies (9, 10) have compared an unfit ancestral strain with a fitter evolved strain after a continuous, multigenerational adaptation on a single selective medium that typically consists of an "off-the-shelf" stressor, e.g., nutrient starvation (11) or osmotic stress (4). Second, we randomly sample from the distribution of a single meiotic generation, the F_1 generation, which is composed of both potentially fit and unfit recombinants, rather than sample the fittest recombinants at different generational milestones. In this screening approach, we capture clones that would normally perish undetected under continuous selection.

The fitness measurements that we make are analyzed in a number of complementary ways. First, we construct a graph (network) that represents the phenotypic relationships among the recombinants: each meiotic recombinant is a node, and an edge is drawn between two nodes if they have similar composite fitness profiles across 10 environmental conditions. We find that the resulting network of phenotypic similarity among F_1 recombinants exhibits a "scale-free" topology in which the probability of finding a recombinant with a certain number of phenotypic neighbors (called the degree of that node and denoted k) follows a power-law distribution; i.e., $p(k) \sim k^{-\gamma}$. The value of the power-law exponent that we measure ($\gamma \sim 1.6$) indicates that both the average and variance of connectivities in this graph are not well defined; that is, it is impossible to identify a degree in the graph that characterizes most recombinants. To understand the origins of this scale-free network, we construct a simple model of recombination and show that it reproduces the scale-free topology of the observed phenotypic similarity network. Finally, we analyze the benefits and costs of sexual reproduction in this experimental population by comparing the fitness of all recombinants to the parental strains. This combined theoretical/experimental approach allows us to characterize the detailed phenotypic landscape produced by meiotic recombination in yeast.

Results and Discussion

Recombination and Fitness Measurements. We compare the sex-generated phenotypes of F_1 recombinants to the phenotypes of their two parental strains, as well as to the heterozygous diploid strain that bore them. In our study, we selected the heterozygous diploid strain XHS123 (12), the product of a cross between a laboratory strain and a clinical isolate (12). Heterozygosity is widespread in this cross, as evidenced by a comparison of the known genome sequences of both the laboratory strain and the clinical isolate: the latter displays an average difference to the former of 10^{-3} changes per synonymous site (13). XHS123 is heterozygous at the *HO* (*YDL227C*) gene, which encodes an endonuclease that initiates mating-type switching. After sporulation, four haploid segregants are formed, but two go on to self-diploidize after germination, i.e., become homozygous diploid. Nonuniform ploidy among F_1 recombinants may confound

Author contributions: E.O.P. and E.J.D. contributed equally to this work; E.O.P. and E.J.D. designed research; E.O.P. and E.J.D. performed research; E.O.P., E.J.D., E.I.S., and S.L.S. contributed new reagents/analytic tools; E.O.P., E.J.D., O.A., E.I.S., and S.L.S. analyzed data; and E.O.P., E.J.D., E.I.S., and S.L.S. wrote the paper.

The authors declare no conflict of interest.

‡To whom correspondence may be addressed. E-mail: perlst@fas.harvard.edu, deeds@fas.harvard.edu, or stuart_schreiber@harvard.edu.

Abbreviations: SMPs, small-molecule perturbagens; TBR, true benefit ratio; GC, giant component.

This article contains supporting information online at www.pnas.org/cgi/content/full/0704037104/DC1.

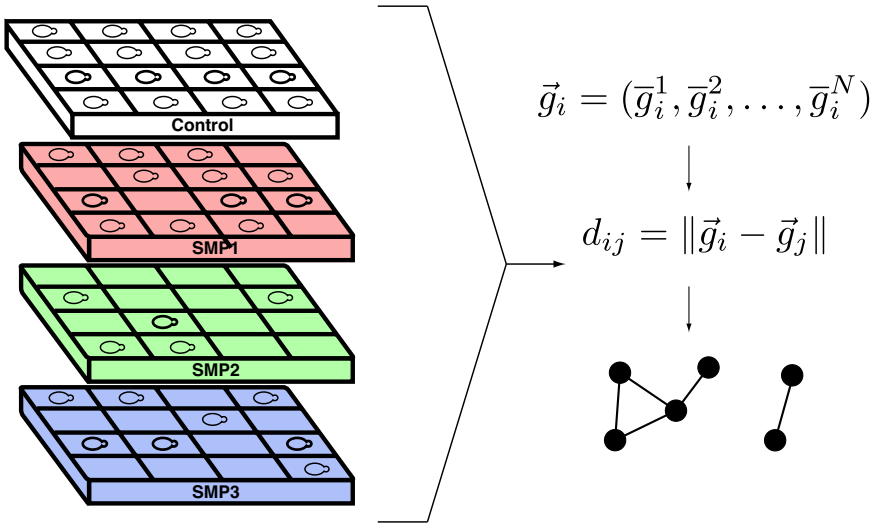

Fig. 1. Schematic depicting the generation of a network of phenotypic similarity. All XHS123 F_1 recombinants are replica-pinned from stock plates (not shown) into a control plate that contains DMSO (white) and into daughter plates, each of which contains a given SMP (red, green, and blue). The growth (i.e., fitness) of each recombinant is measured as described in the text. These growth values are normalized by the average growth of the entire plate, producing a vector of normalized growth values. The distance between the growth vectors of two recombinants is used to define the edges in the graph; similar recombinants are connected by edges, whereas dissimilar recombinants are not. Completely unconnected recombinants are ''orphans,'' whereas highly connected recombinants are ''hubs'' (not shown).

direct comparisons between our results and those of traditional studies of sex that involve direct competitions between strains with identical ploidy. We performed three independent sporulations of XHS123, and each time we harvested thousands of random F_1 meiotic recombinants ($n_1 = 1{,}932$, $n_2 = 2{,}657$, and $n_3 = 3{,}107$). Because XHS123 is URA3/ura3::kanMX4, one-half of F_1 meiotic recombinants inherit the wild-type *URA3* (orotidine-5′-phosphate decarboxylase) gene, i.e., are able to grow in the absence of uracil, whereas the other half of F_1 meiotic recombinants inherit the apposing kanMX4 cassette, which encodes a phosphotransferase that confers resistance to the toxic aminoglycoside geneticin (G418). 2:2 segregation of this artificially engineered Mendelian (i.e., monogenic) trait allowed us to identify, rapidly and reproducibly, recombinants that successfully underwent random assortment. We then assessed, in duplicate, the fitness of each F_1 recombinant in 10 different assays by using 384-well clear-bottom plates, wherein proliferation is measured by light absorbance (OD_{600}) after 48–96 h of benchtop incubation (Fig. 1). Our assays involve treatment with different small-molecule perturbagens (SMPs). Generally, the response of yeast to SMPs is complex (14) (i.e., polygenic), a fact that allows us to monitor the segregation of multiple quantitative trait loci (QTLs).

The following are the 10 SMPs used in our study (not counting the DMSO control treatment that captures intrinsic growth differences between recombinants): (*i*) dimethyl sulfoxide (DMSO), the vehicle (solvent) in which most SMPs are resuspended; (*ii*) geneticin (G418); (*iii*) complete synthetic media lacking uracil (CSM-URA); (*iv*) hydrogen peroxide (H_2O_2); (*v*) SP600125; (*vi*) tunicamycin; (*vii*) calcimycin; (*viii*) menadione; (*ix*) diphenyleneiodonium; (*x*) doxorubicin; and (*xi*) alpha factor (αF). Briefly, H_2O_2 (15) and menadione (16) are both inducers of oxidative stress, although their specific mechanisms of action vary. SP600125 (17) is a putative Jun N-terminal kinase (JNK) inhibitor, and diphenyleneiodonium (18) is a flavoprotein inhibitor. Tunicamycin (19) is an inhibitor of ALG7, UDP-*N*-acetyl-glucosamine-1-P transferase, an evolutionarily conserved gene involved in protein glycosylation. Calcimycin (20) is a calcium-ion ionophore. Doxorubicin (21) is a topoisomerase II inhibitor. Finally, αF is a 13-amino acid peptide that binds a G protein coupled receptor (GPCR) that triggers the mitogen-activated protein kinase (MAPK)-regulated mating response pathway in haploid cells of the MAT**a** mating type (22).

Phenotypic Similarity Graph. We began by implementing a pairwise measure of phenotypic similarity between each of the F_1 recombinants. This measure treats each recombinant as a vector that is composed of the normalized growth of the recombinant in the presence of each SMP (see *Methods*). The phenotypic distance between a pair of recombinants is the Euclidian distance between their growth vectors (Fig. 1). This procedure produces a matrix of distances, and we analyze the distribution of phenotypic similarity in this system by defining a cutoff in phenotypic distance. This cutoff is used to define an unweighted graph, in which nodes are recombinants, and edges are placed between two nodes if the distance between them is less than the cutoff distance. Each recombinant is characterized by the number of neighbors it possesses at a given cutoff (called the degree of that node and denoted k). Degree k represents the size of the F_1 subpopulation that is phenotypically ''closer'' to a given recombinant than the cutoff. Analysis of the resultant networks from three independent sporulation experiments reveals that each iteration exhibits the well known scale-free topology across a variety of cutoff values, indicated by the fact that the probability of finding a recombinant with k phenotypic neighbors is well fit by a power-law function; i.e., $p(k) \sim k^{-\gamma}$. The representative degree distribution displayed in Fig. 2*A* exhibits a power-law exponent of 1.64. The nature of this degree distribution is robust both to bootstrapping of the data and to removing one or two entire conditions from the set of SMPs used to create the phenotype vectors. Discussion of how cutoffs are chosen, and how they affect the degree distribution of the network, can be found in *Methods*. Similar topologies have been observed in networks of both biological (23–27) and nonbiological origin, and the ubiquity of these networks, along with some of the general features of networks that display this topology, have been discussed at length in the literature in recent years.

The scale-free topology that we observe in the network of phenotypic similarity has a number of implications for the fitness landscape produced by recombination in this cross. First, and

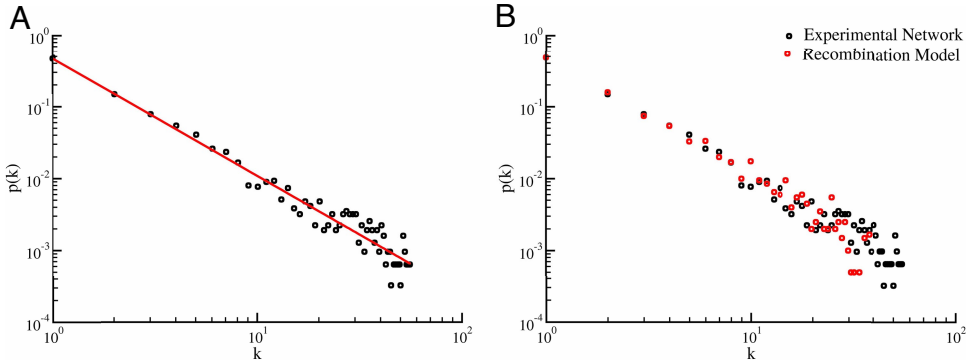

Fig. 2. Scale-free degree distribution of a representative network of phenotypic similarity composed of 3,107 F_1 recombinants. (*A*) Empirical degree distribution fit by power law with exponent 1.64. (*B*) Direct comparison of results observed in experimental network and results generated by recombination model.

most striking, the observed power-law exponent of the degree distribution is <2 for the entire range of cutoffs that yield robust scale-free networks (see *Methods*). This finding indicates that the distribution does not exhibit a well defined average or variance (28). If we consider the case of an infinite population of recombinants (assuming that the power law we observe would still hold), the average connectivity in that graph would diverge, i.e., tend toward infinity. This is not to say that it is impossible to calculate a mean and variance for the degree distributions described above; this finding simply indicates that, as the recombinant population size increases, the average and standard deviation in phenotypic connectivity within that population will increase as well. This behavior is quite unlike other common probability distributions (i.e., Gaussian or Poisson); in the observed scale-free regime the majority of nodes, or recombinants, do not share a characteristic (average) value or fitness profile. A related implication of this finding is that phenotypic "hubs" (nodes with many phenotypically similar neighbors) are found at a much higher frequency than one would expect to observe in a Gaussian distribution with a similar sample average and standard deviation.

Cumulatively, these observations pose a number of interesting questions. First, we address the origin of a scale-free distribution of phenotypic similarity. One possibility is that it arises as a consequence of the physical nature of recombination. We tested this hypothesis by creating and implementing a simple model of recombination (described in detail in *Methods*). This model involves polygenic traits that represent "resistance" or "sensitivity" to a given set of conditions (here we set the growth of resistant strains to 1 and sensitive strains to 0). The loci for these traits are randomly arrayed on a single chromosome, and each polygenic trait (corresponding to resistance in a given condition) corresponds to the same number of loci (e.g., each trait exhibits three loci). Each locus has two alleles, one that contributes to resistance and one that contributes to sensitivity. We model total resistance using an "all-or-none" principle; for a recombinant to be considered resistant to a particular condition, all of the loci that correspond to that trait must have the resistant value. Recombination occurs between two parents that have completely opposite resistance profiles. Thus, if parent A is resistant to a particular condition (with all of the loci for that condition set to the resistant allele), parent B is sensitive to that condition (with all loci set to the sensitive allele). Recombination occurs at nearly random points on the chromosome, and depending on the frequency of recombination in the resulting population, we observe a scale-free distribution of phenotypic similarity in the simulated F_1 generation (Fig. 2*B*). Although this model is very simple and lacks much of the complexity of the system that we experimentally test (such as continuously varying quantitative traits with complex allelic dependencies), it does indicate that this observation may represent a simple consequence of the process of recombination for a large set of polygenic traits and, thus, could describe the phenotypic landscape one might expect to result from recombination.

Second, we address the consequences of a scale-free distribution of phenotypic similarity for the evolution of sexually reproducing organisms, which involves a comparison of the overall phenotypic performance of F_1 recombinants to the performances of the parental strains that bore them. If we consider each of the parental strains, which we will denote "lab" for BY4741 (haploid) and BY4743 (diploid), "clinical" for YJM789 (haploid) and YAG040 (diploid), and "het" for XHS123, we find that two of them exhibit low connectivity (k_{BY4741}; $k_{YJM789} = 1$), whereas the rest have no neighbors in the network ($k = 0$). Although it is not surprising that the recombinants differ from each of their parental strains, the fact that the heterozygote does not represent a hub indicates that the cumulative dissipating effects of recombination (i.e., independent segregation and assortment of genomic loci) results in few F_1 recombinants that are phenotypically similar to the heterozygous state. Next, we generated fitness distributions of F_1 recombinants treated with each SMP. Analysis of these growth distributions indicates that the process of recombination generally produces a bimodal distribution of growth in the F_1 population (Fig. 3*B*) and that transgressive segregation, wherein meiotic recombinants display trait values (e.g., resistance) greater than either parent, is commonplace. The large, less fit peak in these distributions corresponds to those recombinants that closely resemble the sensitive parental strain, whereas the small, fitter peak corresponds to those organisms that are more similar to the resistant parent.

In every case there is a sizeable subpopulation of recombinants that grow *at least* as well as the resistant parental strain or heterozygote [Fig. 3*B* and supporting information (SI) Fig. 5]. Although it is unclear to what extent the differences in normalized growth among the resistant individuals in our assay would translate to significant differences in fitness and growth rates that might influence the outcome of competition experiments, our results indicate that a large fraction of the F_1 population would compete favorably with the parental and heterozygote strains (and perhaps even out-compete them). This assertion is further corroborated by the fact that the difference in normalized growth between the most fit parental strain and the most fit recombinant is generally greater than the standard deviation in normalized growth between individual replicates (data not shown) and among all recombinants and parental strains in DMSO (≈0.2 normalized growth units; see Fig. 3*A*). Given the unclear relationship between the growth measurements we use and the performance of the recombinants in competition experiments, we did not perform an ANOVA analysis to establish

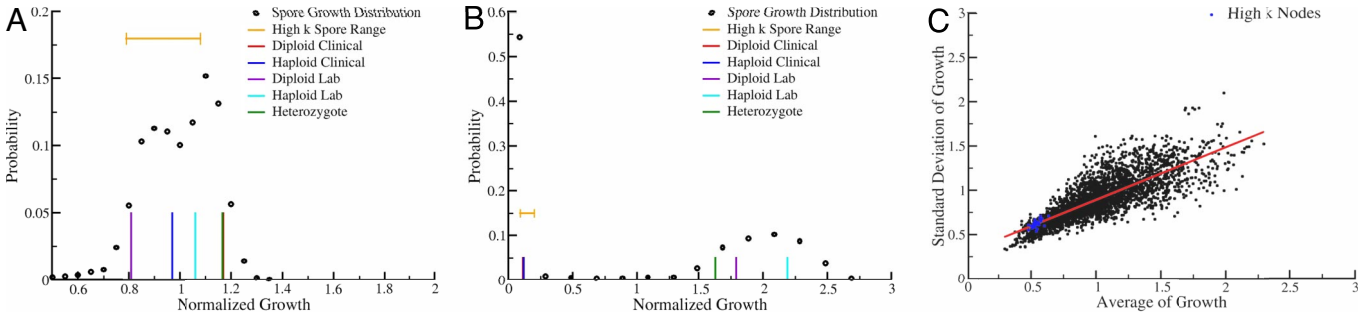

Fig. 3. Representative bimodal fitness distribution revealing phenotypically failing hubs. (*A*) Fitness distribution of 3,107 F_1 recombinants grown in dimethyl sulfoxide (DMSO). The bracketed orange line depicts the range of high *k* nodes. Vertical colored lines denote the fitness of parental strains, as shown in the inset key. (*B*) Fitness distribution of 3,107 F_1 recombinants grown in the small-molecule perturbagen (SMP) calcimycin. (*C*) Plot of average growth across all 11 conditions (10 SMPs + DMSO) versus standard deviation in growth between all 11 conditions. Each black point corresponds to an F_1 recombinant, blue points correspond to high *k* nodes, and the red line depicts a linear regression. The correlation coefficient is 0.8.

the likelihood that these differences are significant; a more detailed comparison of fitness distributions and differences among the parental strains and recombinants is left to future work. Nevertheless, we find that the hubs in the phenotypic network invariably correspond to the less fit regions of these distributions, indicating that these phenotypically similar recombinants sample similar unfit areas of the landscape (Fig. 3*C*). This finding represents a direct observation of the efficiency of purifying selection in recombination; these hubs comprise a "sink" for unfit allele combinations that would clearly be outcompeted by their more fit brethren in any of the conditions tested.

Costs and Benefits of Sexual Reproduction. To quantify phenotypic performance and to compare recombinants across all 10 assays, we introduced the true benefit ratio (TBR). We calculate the TBR for a given recombinant by dividing its normalized growth in each condition by the growth of each parental strain (het, clinical, and lab), a process that yields a ratio for each parent; the TBR is the minimum of these ratios. The TBR represents the absolute benefit that recombination has provided to an F_1 recombinant. As is clear from Fig. 4*A*, in every condition we observe a number of recombinants with TBR ≥ 1. However, we find that the majority of recombinants are phenotypic failures that do not exhibit TBR > 1 in any condition, indicating that unfavorable combinations of alleles (which collect in these recombinants) have a high probability of being eliminated from the population via purifying selection in a wide variety of conditions (10). Most F_1 recombinants are specialists: they have a TBR ≥ 1 in one or a few conditions, whereas only a few generalists have high TBRs in many conditions (Fig. 4*B*). Not surprisingly, generalists tend to be phenotypic orphans; they represent rare combinations of resistant alleles across a variety of conditions. The hubs of the phenotypic network are universally "losers"; indeed, none of the top five *k* recombinants in the network exhibits a single TBR > 1. We observe the same general relationship between the number of conditions to which a recombinant is resistant and connectivity within the context of our simple model (Fig. 4*C*), indicating that the above observation may reflect general features of the process of recombination. It is important to note that the model does not exhibit quite the same level of separation between resistance and sensitivity in high-connectivity organisms (compare *A* and *C* in Fig. 4), which may result from the fact that this model of recombination does not consider important features such as epistasis and complex dependence of resistance on allele combinations. Despite this fact, the nodes with the 50 highest connectivities in the model are resistant to <5 of the 25 model conditions.

In conclusion, recombination simultaneously generates both phenotypic specialists and generalists (substrates for positive selection) while also removing particularly unfit combinations of alleles via purifying selection in multiple uncorrelated conditions. Although generalists are statistically rare (that is, they will only be reliably observed in large F_1 populations spawned by outbred crosses), they represent the chance that recombination will produce exceptional individuals that are robust to many different environments. In each environmental condition, we observe at least one recombinant that is more fit than either of the parental strains. Moreover, the frequency of recombinants that are as resistant to a given SMP as the most resistant parental

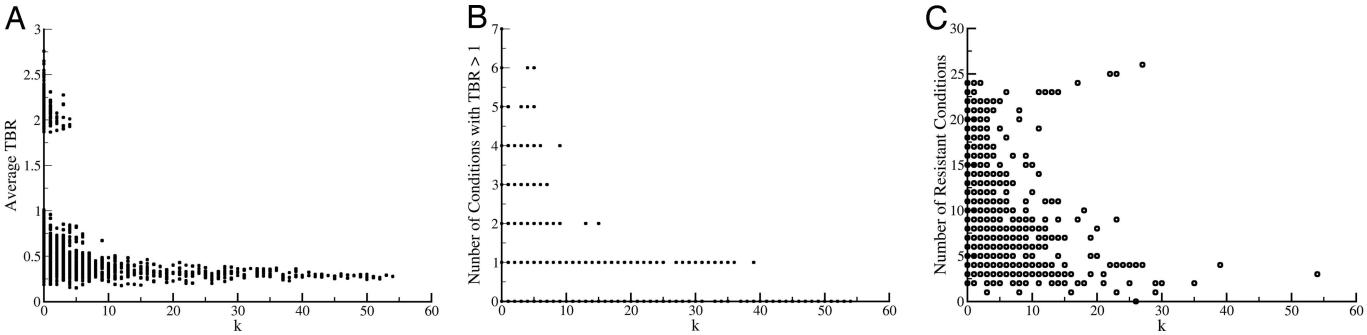

Fig. 4. Calculation of the true-benefit ratio enables a cost-benefit analysis of sex in the F_1 generation. (*A*) Average TBR of a spore (calculated across all 10 conditions) vs. connectivity of that spore in the network with degree distribution shown in Fig. 2. (*B*) Plot similar to that in *A* but displaying the number of conditions in which a given spore exhibits a TBR >1 vs. the connectivity of the spore. (*C*) Number of model conditions to which a recombinant is resistant vs. the connectivity of that recombinant in our simple model of recombination.

strain is several orders of magnitude more probable than what would be accessible by mutation. This study, therefore, not only reinforces the established notion that recombination increases variance but also demonstrates that this variance diverges with increasing population size (29). In other words, the potential for phenotypic diversity is greater under the observed scale-free regime than it would be under, for example, a Gaussian regime. Finally, the process of recombination also results in a large number of recombinants that exhibit comparatively low growth rates across many conditions. These unfit strains tend to cluster in certain regions of the fitness landscape, resulting in failing phenotypic hubs embedded in a scale-free phenotypic similarity network. The topology of this network implies that these unfit hubs represent a significant portion of the population and thus expose deleterious allelic combinations to purifying selection.

Methods

Yeast Strains. YAG040 (MAT**a**/MATα HO/hoD::hphMX4), YJM789 (MATα hoD:hisG lys2 LYS5 cyh2), and XHS123 (MAT**a**/MATα URA3/ura3D::kanMX4) were generous gifts from J. H. McCusker (Duke University, Durham, NC). BY4741 (MAT**a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and BY4743 (MAT**a**/MATα *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 MET15Δ0/met15Δ0 LYS2Δ0/lys2Δ0 ura3Δ0/ura3Δ0*) were obtained from American Type Culture Collection.

Random F_1 Meiotic Recombinant Generation and Fitness Measurement. Culturing and sporulation of yeast, as well as media formulation, was done as described (30). F_1 meiotic recombinants were originally derived as random spores (not by tetrad dissection), and they were then either manually arrayed or robotically arrayed by a QBot colony picker (no. X8000; Genetix, Boston, MA) from dabs of single colonies grown on agar trays (no. X6021; Genetix) into individual wells of 384-well plates containing rich media to make strain stock plates. NUNC 384-well, clear-bottom, untreated, sterile plates (no. 62409-604; VWR Scientific, West Chester, PA) containing rich media and either vehicle solvent (DMSO) or a given SMP were inoculated with yeast from strain stock plates by using sterile polypropylene 384-pin replicators (no. X5050; Genetix). Inoculated plates were grown without agitation on the bench top at ambient temperature conditions for 48–96 h and then vortexed on a standard table-top vortexer (VWR Scientific) for 10–30 sec before measurement in a SpectraMax plate reader (Molecular Devices, Sunnyvale, CA) set to 600-nm emission. The growth of each XHS123 meiotic recombinant in rich media containing a given SMP was compared with its growth in media containing DMSO. Meiotic recombinants that were unable to grow in DMSO were eliminated from further study.

SMPs. SMPs were individually purchased as powder stocks from various vendors in either 10-mg or smaller quantities, resuspended in DMSO in glass vials as stock solutions, and used at the following final concentrations: 25 μg/ml calcimycin (no. CA-100; Biomol, Plymouth Meeting, PA), 25 μg/ml doxorubicin (no. GR-319; Biomol), 25 μg/ml SP600125 (no. 420119; EMD Biosciences, San Diego, CA), 2 μg/ml tunicamycin (no. CC-104; Biomol), 5 μg/ml menadione (no. M5625; Sigma–Aldrich, St. Louis, MO), 25 μg/ml diphenyleneiodonium (no. D2926; Sigma–Aldrich), 0.015% hydrogen peroxide (no. 31642; Sigma–Aldrich), 10 μM alpha factor (Y1001; Zymo Research, Orange, CA), 200 mg/ml G418 (no. 11811-031; Gibco, Carlsbad, CA), and 0.77 g/liter complete supplemental mixture without uracil (CSM-URA) (no. 4511-212; Qbiogene, Carlsbad, CA). See the National Cancer Institute-sponsored public database ChemBank (http://chembank.broad.harvard.edu) for complete structural annotations and mechanisms of action.

Graph Theoretic Analysis and TBR. The graph theoretic analysis was conducted by first normalizing the growth of each well on each plate by dividing the raw growth measurement for each well by the average growth of the plate. Normalization of this type is required to take into account the fact that each plate was not allowed the same time in which to grow and that plates analyzed later will have, in general, higher optical density (i.e., total growth) values compared with plates analyzed earlier. This normalization produces a smaller overall variance between replicates (i.e., between identical individuals) compared with other normalizations (i.e., median growth, average parental growth on that plate, etc.). The normalized growth for each recombinant was then averaged across the two replicates for each condition. This normalized average growth across all 11 conditions represents an 11-dimensional vector that is used to create a matrix of distances between the F_1 recombinants. In this case we employ a simple Euclidean measure of distance.

In the graph that we created, each recombinant represents a node, and the edges between these nodes are constructed from the distance matrix by defining a cutoff, d_c, such that two recombinants i and j are only linked by an edge if $d_{ij} < d_c$. At a given cutoff, the graph can be easily clustered by using a standard depth-first search algorithm. The graph can be characterized by the size of its largest cluster (the giant component, GC). As the distance cutoff is decreased (i.e., as the cutoff becomes more stringent), the size of the GC undergoes a transition from fully connected, in which every node belongs to the GC, to completely unconnected, in which every node is an orphan. This transition has been well characterized (31) and allows us to standardize our analysis between different experiments and between the experimental and computational results. Given that the actual distance values in each separate matrix (and the distance measure itself) do not have the same inherent scale when comparing between experiments or between experiment and theory, this transition allows us to perform a responsible comparison between the graphs; that is, we cannot fit the distance cutoff independently to observe a particular behavior in each case. In the work and graphs described in the text, d_c is set so that the GC contains one-fourth of the nodes in each graph. Cutoffs that yield a GC containing between approximately one-eighth and one-half of the number of nodes in the graph yield degree distributions that are well fit by power-law functions with exponents between -1 and -2. At more stringent cutoffs, the graph is very sparse (i.e., there are very few edges) and maximal degrees are small, making it difficult to robustly fit degree distributions in this cutoff region to power-law functions. At more permissive cutoffs, the graph becomes very dense and the degree distributions are no longer well fit by power-law functions. In this range of cutoffs, many phenotypically unrelated nodes become connected, obviating the scale-free behavior.

The TBR is calculated in a straightforward manner. First, the parental growth values are normalized and averaged for each plate and replicate in each condition. The overall growth of each parental strain was determined by averaging the growth of that strain across all of the plates in each condition. The TBR for each recombinant in each condition was obtained by dividing the normalized growth of that recombinant in that condition by the overall growth value of each parental strain in that condition. The minimum such ratio was taken to be the TBR.

Computational Model of Recombination. To asses whether the scale-free network of phenotypic relatedness could be achieved based simply on the physical nature of recombination, rather than on other biological or evolutionary factors, we created a simple computational model of recombination. This model considers a set of polygenic traits where resistance to a given model condition is determined by more than one locus in the model organism. In this model, every such trait corresponds to

the same number of loci (i.e., all of the resistances are quantitative traits with the same number of genetic determinants). Each locus exhibits two alleles, one that contributes to resistance (denoted "r") and one that does not (denoted "s"). Resistance to a particular condition is achieved only if all of the loci that correspond to that trait exhibit "r" alleles in a particular organism; that is, resistance requires that all of the alleles that determine that trait have the resistant value. Resistant organisms display a growth of 1 in the respective model condition and sensitive organisms display a growth of 0 in the respective model condition.

In this model we consider organisms with a single chromosome. Loci are arranged on this chromosome completely randomly and without respect to which trait each locus contributes (i.e., the correspondence between loci and traits is completely random and uncorrelated). Two parental strains are constructed to have completely dimorphic resistance profiles; that is, if parent A is resistant to condition α, then parent B will be sensitive to condition α. For simplicity, this is achieved by setting all of the alleles that contribute to resistance in condition α to "r" in parent A and to "s" in parent B. Resistance in each condition is assigned randomly to one of the parents. Recombination between the A and B genomes occurs at random points along the chromosome. The algorithm moves down a particular chromosome, and at every position there is an independent, constant probability (p_{cross}) that a recombination event will occur. One individual recombinant is chosen at random and placed in the model F_1 recombinant population.

The results reported in Figs. 2*B* and 4*C* were obtained from a single realization of the model in which 3,000 independent recombinants were generated according to the above procedure. These results are based on 50 conditions (and thus a total of 50 traits) each of which is based on 3 alleles for a total of 150 alleles, which are randomly placed on a single chromosome 1,500 "units" in length. For the run described in the figures, p_{cross} is set to 0.0033. The realization displayed in the text is representative of the ensemble of graphs created by using this set of parameters (data not shown). We observe similar scale-free networks and k vs. resistance behavior using different parameter sets (data not shown), but we leave a systematic analysis of this model and its dependencies to future work.

We thank Gopal Ramachandran for helping us to conceive this work; Drs. Andrew Murray, David Altshuler, and Dan Hartl for their comments on the manuscript; and Dr. David Spiegel for original artwork. This work was supported by the National Institute of General Medicine Sciences (to S.L.S.). S.L.S. is an Investigator of the Howard Hughes Medical Institute, and E.J.D. was supported by a Howard Hughes Medical Institute predoctoral fellowship.

1. Barton NH, Charlesworth B (1998) *Science* 281:1986–1990.
2. Burt A (2000) *Evolution (Lawrence, Kans)* 54:337–351.
3. Kaltz O, Bell G (2002) *Evolution (Lawrence, Kans)* 56:1743–1753.
4. Greig D, Borts RH, Louis EJ (1998) *Proc R Soc London Ser B* 265:1017–1023.
5. Weismann A (1904) *The Evolution Theory* (Arnold, London).
6. Maynard Smith J (1978) *The Evolution of Sex* (Cambridge Univ Press, Cambridge, UK).
7. Goddard MR, Godfary HC, Burt A (2005) *Nature* 434:636–640.
8. Paland S, Lynch M (2006) *Science* 311:990–992.
9. Cooper TF, Rozen DE, Lenski RE (2003) *Proc Natl Acad Sci USA* 100:1072–1077.
10. Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D (2002) *Proc Natl Acad Sci USA* 99:16144–16149.
11. Brauer MJ, Saldanha AJ, Dolinski K, Botstein D (2005) *Mol Biol Cell* 16:2503–2517.
12. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW (2002) *Nature* 416:326–330.
13. Gu Z, David L, Petrov D, Jones T, Davis RW, Steinmetz LM (2005) *Proc Natl Acad Sci USA* 102:1092–1097.
14. Perlstein EO, Ruderfer DM, Ramachandran G, Haggarty SJ, Kruglyak L, Schreiber SL (2006) *Chem Biol* 13:319–327.
15. Moradas-Ferreira P, Costa V (2000) *Redox Rep* 5:277–285.
16. Shapira M, Segal E, Botstein D (2004) *Mol Biol Cell* 15:5659–5669.
17. Bennett BL, Sasaki DT, Murray BW, O'Leary EC, Sakata ST, Xu W, Leisten JC, Motiwala A, Pierce S, Satoh Y, *et al.* (2001) *Proc Natl Acad Sci USA* 98:13681–13686.
18. Hancock JT, Jones OT (1987) *Biochem J* 242:103–107.
19. Barnes G, Hansen WJ, Holcomb CL, Rine J (1984) *Mol Cell Biol* 4:2381–2388.
20. Friis J, Szablewski L, Christensen ST, Schousboe P, Rasmussen L (1994) *FEMS Microbiol Lett* 123:33–36.
21. Meczes EL, Marsh KL, Fisher LM, Rogers MP, Austin CA (1997) *Pharmacology* 39:367–375.
22. Reed SI (1991) *Curr Opin Genet Dev* 1:391–396.
23. Jeong H, Tombor B, Albert R, Oltvai ZN, Barbasi AL (2000) *Nature* 407:651–654.
24. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) *Science* 297:1551–1555.
25. Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabasi AL (2004) *Nature* 427:839–843.
26. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M (2004) *Nature* 430: 88–93.
27. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, *et al.* (2000) *Nature* 403:623–627.
28. Albert R, Barabási AL (2002) *Rev Mod Phys* 74:47–97.
29. Colegrave N (2002) *Nature* 420:664–666.
30. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, eds (1996) *Current Protocols in Molecular Biology* (Wiley, New York), Vol 2.
31. Albert R, Barabási AL (2000) *Phys Rev Lett* 85:5234–5237.