The cost of gene expression underlies a fitness trade-off in yeast

Gregory I. Lang[a,1], Andrew W. Murray[b,1,2], and David Botstein[a,1,2]

[a]Lewis-Sigler Institute for Integrative Genomics and the Department of Molecular Biology, Princeton University, Princeton, NJ 08544; and [b]FAS Center for Systems Biology and the Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

Contributed by David Botstein, February 12, 2009 (sent for review December 10, 2008)

Natural selection optimizes an organism's genotype within the context of its environment. Adaptations to one environment can decrease fitness in another, revealing evolutionary trade-offs. Here, we show that the cost of gene expression underlies a trade-off between growth rate and mating efficiency in the yeast *Saccharomyces cerevisiae*. During asexual growth, mutations that eliminate the ability to mate provide an ≈2% per-generation growth-rate advantage. Some strains, including most laboratory strains, carry an allele of *GPA1* (an upstream component of the mating pathway) that increases mating efficiency by ≈30% per round of mating at the cost of an ≈1% per-generation growth-rate disadvantage. In addition to demonstrating a trade-off between growth rate and mating efficiency, our results illustrate differences in the selective pressures defining fitness in the laboratory versus the natural environment and show that selection, acting on the cost of gene expression, can optimize expression levels and promote gene loss.

evolution | *GPA1* | mating pathway | *Saccharomyces cerevisiae*

A frequent observation in evolution is that traits not maintained by selection will be lost—this holds true at the morphological level and at the genetic level. Examples of gene loss include the loss of olfactory receptors in primates (1), the loss of pigmentation and vision in *Astyanax* cavefish (2), the loss of the galactose utilization pathway in yeast (3), and the degeneration of genes involved in carbon utilization during domestication of *Streptococcus thermophilus* (4). Such regressive evolution also occurs in laboratory populations; reduction in catabolic breadth and thermal tolerance is observed during long-term evolution in *Escherichia coli* (5–9), and sterility frequently arises during long-term asexual propagation of *Saccharomyces cerevisiae* (10).

Two mechanisms could account for gene loss during evolution. One possibility is that in the absence of selection, genes are lost because of the neutral accumulation of mutations. Alternatively, gene loss events could be driven by selection. The observation that many of these gene-loss events are repeatedly observed supports this hypothesis. Repeated loss of all or part of the *Rbs* operon (whose products catabolize ribose) in *E. coli* provides a selective advantage in minimal glucose media (8). Quantitative analysis of alleles leading to eye reduction in *Astyanax* indicates that selection, possibly against the energetic cost of vision, is responsible for eye degeneration in cavefish populations (11). These studies suggest that haploid yeast that are propagated for long periods without mating partners should become sterile. Previous studies, however, showed that lineages that evolved higher growth rates and lower mating efficiencies appeared to segregate these traits independently (10). Here, we set out to directly test whether selection drives yeast to become sterile by determining whether mutations conferring sterility provide a selective advantage.

Results

Sterility Increases Growth Rate by Eliminating Unnecessary Gene Expression. We tested the hypothesis that sterile strains generally have a growth-rate advantage by isolating sterile mutants and testing their fitness. Haploid, **a**-mating type (*MAT***a**) cells, arrest in G1 when exposed to the mating pheromone, alpha-factor (αF), and thus cannot form colonies on media containing αF. We initiated, from a single colony of haploid *MAT***a** cells, a large number of parallel cultures that were plated onto either rich media or rich media containing αF. On rich media, the vast majority of cells form colonies, but on αF, only the small fraction of cells that have acquired mutations in pheromone-induced signaling can form colonies. From each culture, we randomly chose a single alpha-factor resistant (αF^R) or unselected colony and measured its relative growth rate by using a FACS-based competitive growth-rate assay that can detect growth-rate differences as small as 0.5%. The growth-rate coefficient is a measure of the growth-rate advantage over wild type. Fig. 1*A* shows the growth-rate coefficients (s_g) for 27 unselected clones and 45 αF^R clones. As a control we measured the relative growth rates of 24 similarly selected mutants that were resistant to canavanine, a toxic arginine analog. In each case, several clones have a low growth rate ($s_g < -1\%$), suggesting that these strains have become mitochondrial deficient or have acquired a deleterious mutation. Excluding clones with $s_g < -1\%$, the growth-rate coefficients of the unselected clones follow a tight distribution (Fig. 1*A*, $s_g = 0.08\% \pm 0.35\%$) indistinguishable from the distribution of the canavanine-resistant mutants (Fig. 1*A*, $s_g = 0.36\% \pm 0.48\%$, $P > 0.05$, Wilcoxon rank sum test); however, the growth-rate coefficients of the αF^R mutants show greater variation and a positive growth-rate advantage (Fig. 1*A*, $s_g = 1.48\% \pm 0.85\%$, $P < 10^{-7}$, Wilcoxon).

It appears from these data that at least some sterile mutants have a clear growth-rate advantage over wild type. To determine whether all sterile strains have a similar advantage, and to determine the basis for any growth-rate advantage in the sterile strains, we used a combination of 4 methods: Phenotypic characterization of the spontaneous αF^R mutants, growth-rate assays on targeted gene deletions within the mating pathway, mapping of the mutations in the most fit sterile strains, and expression analysis on αF^R strains both with and without a growth-rate advantage.

The yeast mating pathway is one of the best studied mitogen-activated protein (MAP) kinase cascades (12). At the beginning of the pathway is a pheromone receptor (Ste2 in *MAT***a** or Ste3 in *MAT*α) that binds the cognate mating pheromone. Receptor stimulation activates a heterotrimeric G protein (consisting of Gpa1, Ste18, and Ste4), which in turn, activates a MAP kinase cascade (consisting of the MAP kinase kinase kinase, Ste11, the MAP kinase kinase, Ste7, the MAP kinases Fus3 and Kss1, and

Author contributions: G.I.L., A.W.M., and D.B. designed research; G.I.L. performed research; G.I.L. analyzed data; and G.I.L., A.W.M., and D.B. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: glang@princeton.edu, amurray@mcb.harvard.edu, or botstein@genomics.princeton.edu.

[2]A.W.M. and D.B. contributed equally to this work.

This article contains supporting information online at www.pnas.org/cgi/content/full/0901620106/DCSupplemental.

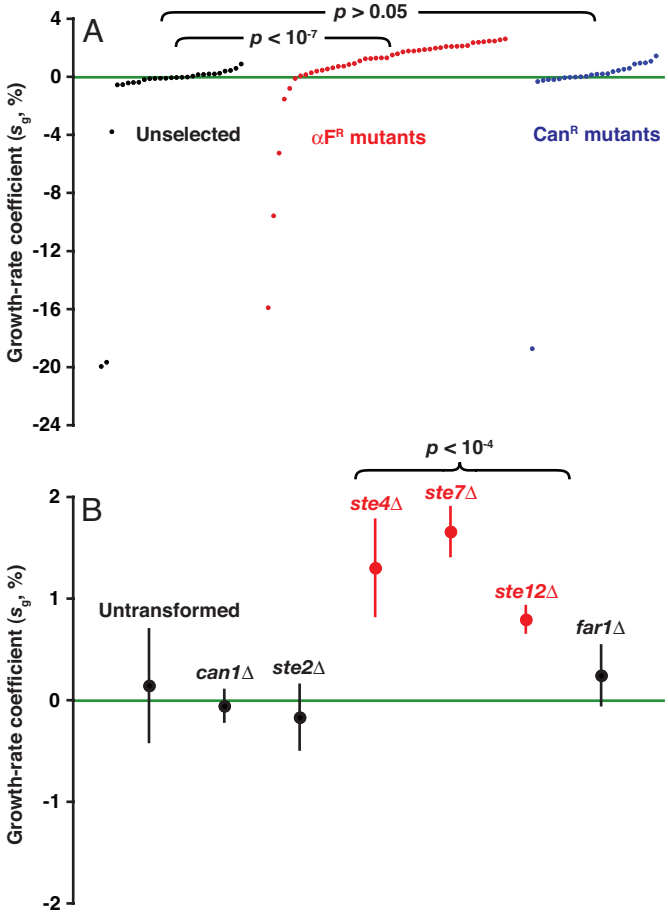

Fig. 1. A subset of αF^R mutations provides a competitive growth-rate advantage. (*A*) Spontaneous αF^R mutants have a greater variance of growth-rate coefficients (s_g) and a higher average growth rate than unselected clones ($P < 10^{-7}$, Wilcoxon rank sum test), whereas the distribution of s_g for Can^R mutants is similar to wild type ($P > 0.05$, Wilcoxon). The 8 clones with growth-rate disadvantages of at least 1% were excluded from the statistical analysis. (*B*) Targeted gene disruptions show that loss of the G_β subunit (Ste4), the MAP kinase kinase (Ste7), or the transcription factor (Ste12) increases growth rate ($P = 2.8 \times 10^{-5}$, 5.8×10^{-9}, 2.2×10^{-6}, respectively, *t* test); however, loss of the receptor (Ste2) or Far1 does not [$P = 0.23$, 0.03, respectively, *t* test (the value for *far1*Δ is not significant because we are testing 5 deletion mutants and thus require $P < 0.05/5 = 0.01$ to regard a result as significant)]. Values represent the mean and standard deviation of 5 to 7 independent gene deletions.

the scaffolding protein, Ste5) ultimately leading to a cell-cycle arrest (dependent on Far1) and the induction of mating genes through the transcription factor, Ste12 (13). Expression of mating pathway genes in the absence of pheromone is maintained by basal signaling through the pathway, which depends on the G protein, MAP kinase cascade, and Ste12, but is independent of the receptor or Far1.

We phenotypically characterized 41 of the 45 spontaneous αF^R mutants by determining the position of each mutation relative to Ste4 and Far1. Overexpression of Ste4 activates the signaling pathway in the absence of pheromone as does expression of a mutant form of Far1 encoded by *FAR1-22*. We transformed each αF^R mutant with plasmids containing galactose-inducible *STE4* or *FAR1-22*. Five of the 41 successfully transformed strains arrest after Ste4 overexpression indicating a mutation at or before Ste4 in the signal transduction pathway. These 5 mutations are found throughout the distribution of growth-rate coefficients indicating that a growth-rate advantage can be gained by losing signaling at multiple points in the mating pathway (Fig. S1*A*). All transformed strains arrest after overexpression of the dominant *FAR1-22* allele (Fig. S1*A*). The 4 αF^R strains not phenotypically characterized are among those that accumulate suppressor mutations and are biased toward the lower end of the growth-rate distribution (Fig. S1*B*, $P = 3.3 \times 10^{-4}$, Wilcoxon rank sum).

There are 2 possibilities for the growth-rate advantage observed for αF^R mutants: Elimination of basal transcription downstream of Ste12 or elimination of a transient Far1-dependent arrest because of inappropriate activation of the pathway. To distinguish between these possibilities, we measured the growth rate of several targeted gene deletions in the mating pathway. Deletion of the G_β subunit (Ste4), the MAP kinase kinase (Ste7), or the transcription factor (Ste12) increases growth rate relative to strains in which all 3 genes are intact (Fig. 1*B*) ($P = 2.8 \times 10^{-5}$, 5.8×10^{-9}, 2.2×10^{-6}, respectively, *t* test); however, deletion of the αF receptor (Ste2) or Far1 does not (Fig. 1*B*) [$P = 0.23$, 0.03, respectively, *t* test (the value for *far1*Δ is not significant because we are testing 5 deletion mutants and thus require $P < 0.05/5 = 0.01$ to regard a result as significant)]. This suggests that a growth-rate advantage exists for the subset of sterile strains that abolish basal signaling through the pathway (therefore eliminating basal expression of the mating genes), which depends on Ste4, Ste7, and Ste12, but not Ste2 or Far1 (14).

We identified the mutations in several αF^R mutants from the higher end of the growth-rate distribution by hybridizing genomic DNA to microarrays that cover the entire yeast genome (tiling arrays) and characterized their effect on basal expression downstream of Ste12 by using gene expression microarrays. For simplicity the 45 spontaneous αF^R mutants were numbered in order of growth-rate advantage, from αF^R-1 (highest) to αF^R-45 (lowest). We chose 7 αF^R strains for this analysis: αF^R-1, -2, -4, -7, -8, -17, and -20; our preliminary data analysis had suggested that these were the 7 αF^R mutants with the greatest growth-rate advantage. In strains αF^R-2, αF^R-8, and αF^R-20 we identified mutations in known mating genes: Ste11, Ste5, and Ste7, respectively. In strains αF^R-4, αF^R-7, and αF^R-17 we did not identify any mutations from the yeast tiling arrays; however, subsequent expression analysis suggested that αF^R-4 and αF^R-7 contain mutations in Ste7 and Ste4, respectively. Sequencing of these genes revealed coding changes in each gene resulting in the following protein modifications: $Ste11^{P656H}$, $Ste7^{E3ochre}$, $Ste4^{frameshift}$, $Ste5^{C198S}$, and $Ste7^{L7ochre}$ in αF^R-2, -4, -7, -8, and -20, respectively. Fig. 2*A* shows a mapping of growth-rate coefficients for the identified spontaneous αF^R mutants and gene deletions onto the mating pathway; 5 of the 6 faster-growing spontaneous αF^R mutants contain a single mutation that reduces basal-signaling-dependent gene expression. Strain αF^R-1 is the exception; it is the only instance where we found multiple mutations and mutations outside of the mating pathway. In this strain we found mutations in Apc1 (an essential component of the anaphase promoting complex) and Eds1 (an uncharacterized, putative zinc-cluster protein). Apc1 and its G1 cofactor Cdh1 play a conserved and critical role in maintaining a G1 arrest, and mutations in Apc1 result in premature entry into S phase (15, 16). This strain was selected for αF resistance, has an expression profile identical to that of canonical *ste* mutants, and arrests after overexpression of Far1. Because the pheromone signaling pathway is repressed outside G1, mutants that reduce the duration of G1 will reduce the basal expression of mating genes, as APC mutants have been shown to do (17).

We assayed for changes in gene expression for the 7 αF^R mutants from the upper end of the growth-rate distribution ($s_g > 0$), 3 from the lower end of the distribution ($s_g = 0$), and the targeted gene disruptions (Fig. 2*B*). The 7 spontaneous αF^R mutants from the upper end of the growth-rate distribution

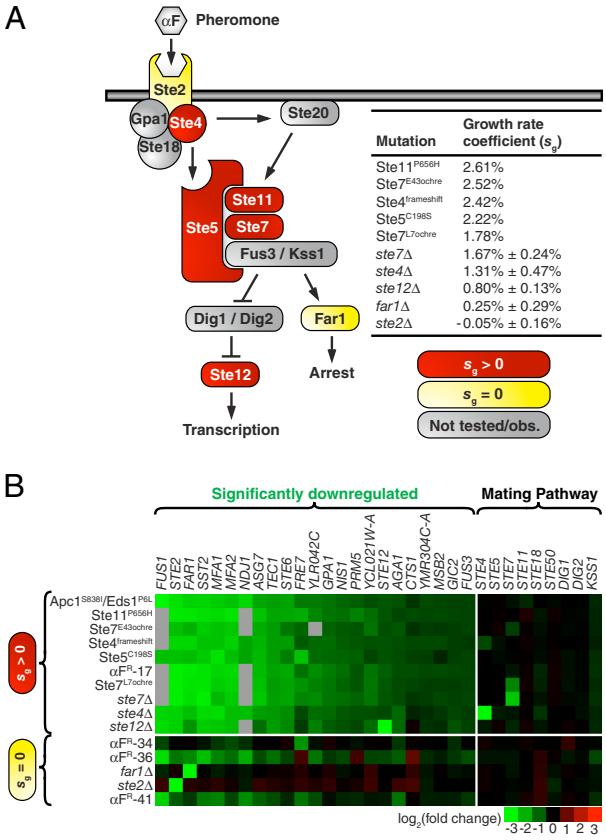

Fig. 2. The growth-rate advantage of αF^R mutants correlates with the elimination of gene expression. (*A*) A schematic of the pheromone-induced signaling pathway. In a *MAT***a** yeast cell, the mating pheromone, αF, binds to the receptor, Ste2. Pheromone-induced signaling activates a heterotrimeric G protein (consisting of Gpa1, Ste18, and Ste4), which in turn, activates a MAP kinase cascade (consisting of the MAP kinase kinase kinase, Ste11, the MAP kinase kinase, Ste7, the MAP kinases Fus3 and Kss1, and the scaffolding protein, Ste5) ultimately leading to a cell-cycle arrest dependent on Far1 and a transcriptional response through the transcription factor, Ste12 (13). Expression of mating pathway genes in the absence of pheromone is maintained by basal signaling through the pathway, which is independent of the receptor (Ste2) or Far1. Identification of the mutations in 5 spontaneous αF^R mutants with a growth-rate advantage using yeast tiling arrays (42) identified $Ste11^{P656H}$, $Ste5^{C198S}$, and $Ste7^{L7ochre}$ mutations in strains αF^R-2, αF^R-8, and αF^R-20, respectively. Using this method, we also identified $Apc1^{S838I}$ and $Eds1^{P6L}$ mutations in the strain αF^R-1. (*B*) αF^R strains with a growth-rate advantage reduce gene expression downstream of Ste12, whereas strains without a growth-rate advantage do not show the same reduction in gene expression. Strains are displayed in order of their growth-rate coefficient. The genes displayed are those whose expression changes significantly in the 7 spontaneous αF^R mutants with a competitive growth-rate advantage (Fig. S2) or known components of the mating pathway. Some of the apparent down-regulation in strains αF^R-36 and αF^R-41 may be an artifact because these strains acquire suppressor mutations that partially restore mating and αF arrest (Fig. S1*B*). From these arrays, the $Ste7^{E43ochre}$ and $Ste4^{frameshift}$ mutations were identified in strains αF^R-4 and αF^R-7, respectively, because the suppression of the expression of specific genes suggests nonsense-mediated decay of the transcripts.

significantly decrease the expression of 23 genes but did not significantly increase the expression of any other genes (Fig. S2). The 3 spontaneous αF^R mutants from the lower end of the growth-rate distribution and the *far1* and *ste2* deletions do not decrease the expression of these genes to the degree seen in strains with a growth-rate advantage, consistent with the hypothesis that the growth-rate advantage is the result of elimination of basal expression of the mating pathway genes; the apparent down-regulation of some mating genes observed in strains αF^R-36 and αF^R-41 could be an artifact because of the acquisition of suppressor mutations that partially restore mating. We conclude that the 2% growth-rate advantage in the most-fit sterile strains is not the result of loss of signaling through the mating pathway per se, but rather the elimination of unnecessary expression from 23 genes related to the mating pathway, supporting the hypothesis that selection, acting on the cost of gene expression, can promote gene loss.

Selection for Mating Efficiency Can Increase the Cost of Gene Expression. Although expressing mating genes decreases growth rate, strains derived from the best characterized laboratory strain, S288c, carry an allele of the G_α subunit (*GPA1*) that increases basal expression of these genes (18). [The strain used in these experiments, W303, is a mosaic genome comprised of ≈85% S288c (Fig. S3*A*). *GPA1* is located on the left arm of chromosome VIII, in a region of predominantly non-S288c descent; however, there is evidence of recombination breakpoints 8 kb upstream and 14 kb downstream of this gene consistent with the idea that the derived allele confers a selective advantage under laboratory conditions (Fig. S3*B*).] The S288c allele of *GPA1* contains a T at position 1,406 in place of the ancestral G (found in multiple wild *S. cerevisiae* isolates and the other sequenced *sensu stricto* species: *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. bayanus*) resulting in the substitution of isoleucine for serine at amino acid 469, which is located in the C-terminal region of the protein thought to interact with the Ste2 receptor; other mutations in this region constitutively activate the mating pathway (19).

To test the effects of the 2 *GPA1* alleles on fitness, we introduced the wild-type *GPA1* allele into the laboratory strain and measured the effect of this single mutation on growth rate and mating efficiency relative to the laboratory strain. We isolated 3 individual transformants of each allele (wild-type *GPA1* and *GPA1*-G1406T), which are isogenic except for the *GPA1* locus. Expression analysis shows that the wild-type *GPA1* allele reduces expression of genes downstream of Ste12 relative to the *GPA1*-G1406T allele, although these genes are not reduced to the levels seen for sterile mutations (*ste7*Δ, *ste4*Δ, and *ste12*Δ), which eliminate signaling through the mating pathway (Fig. 3*A*). To determine whether reducing the expression of the mating genes by restoring the wild-type *GPA1* results in a growth-rate advantage, we measured the relative growth rate by using the FACS-based growth-rate assay and by direct competitions. Strains with the wild-type *GPA1* allele have a significant growth-rate advantage over strains with the *GPA1*-G1406T allele in both assays (Fig. 3*B* and Fig. S4, $s_g = 0.92\% \pm 0.35\%$, for the FACS-based assay, $P = 2.57 \times 10^{-6}$, *t* test; $s_g = 0.84\% \pm 0.15\%$ for the direct competition, $P < 0.001$; the values for s_g are in good agreement between these 2 assays, $P > 0.10$, *t* test). Consistent with the hypothesis that the growth-rate advantage is the result of reduction in gene expression, we find that, like expression levels, the growth-rate advantage of the wild-type *GPA1* allele replacement strains is greater than that observed for strains that do not affect signaling (*far1*Δ and *ste2*Δ) and less than that observed for strains that eliminate signaling (*ste7*Δ, *ste4*Δ, and *ste12*Δ) through the mating pathway.

Among the genes up-regulated by the *GPA1*-G1406T allele are the pheromone genes (*MFA1* and *MFA2*); it has been demonstrated that strains producing more pheromone are more attractive to a mating partner (20). To determine whether the 1% growth-rate disadvantage of strains carrying the *GPA1*-G1406T allele is offset by an increase in mating efficiency, we measured competitive mating efficiency of strains carrying the 2 alleles of *GPA1* by using an assay analogous to the competitive growth-rate assay: *MAT***a** cells carrying either the wild-type or G1406T allele of *GPA1* were mixed and mated to a limited number of

GENETICS

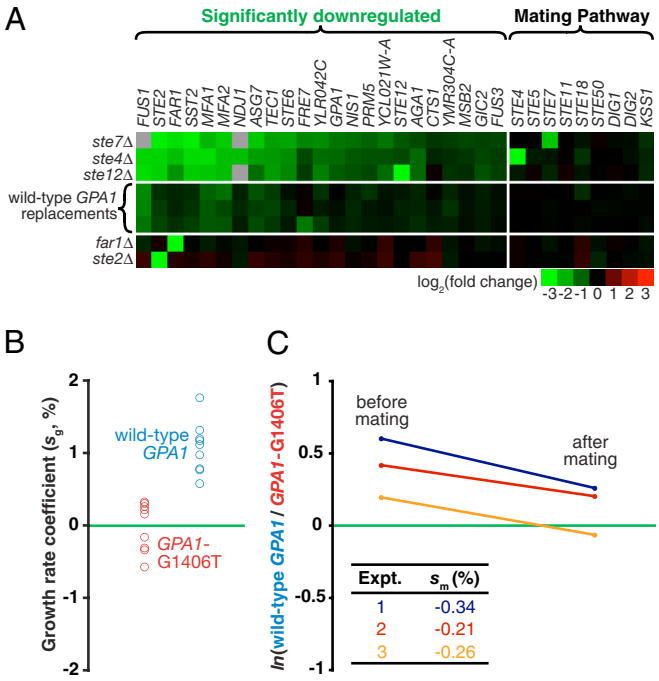

Fig. 3. A trade-off between growth rate and mating efficiency. (*A*) The wild-type allele of *GPA1* down-regulates genes in the mating pathway producing an expression profile intermediate to that of deletions eliminating basal signaling (*ste7*Δ, *ste4*Δ, and *ste12*Δ) and those not affecting signaling (*far1*Δ and *ste2*Δ). Shown are 3 independent wild-type *GPA1* allele replacement strains. (*B*) Wild-type *GPA1* allele replacement strains have a growth-rate advantage relative to the *GPA1*-G1406T allele strains ($s_g = 0.92\% \pm 0.35\%$ and $-0.17\% \pm 0.34\%$ for the wild-type *GPA1* allele and the *GPA1*-G1406T allele, respectively, $P < 2.6 \times 10^{-6}$, *t* test). The points represent 3 independent measurements for each of 3 independent transformants of each *GPA1* allele. (*C*) Wild-type *GPA1* allele replacement strains have a mating disadvantage relative to the *GPA1*-G1406T allele strains ($s_m = -27.2\% \pm 6.5\%$). *MAT***a** strains carrying each allele were mixed and allowed to compete for a limiting number of *MAT*α cells. The mating coefficients (s_m) were calculated as the change in the natural logarithm of the ratio of the 2 alleles: $s_m = ln(\text{wild-type } GPA1/GPA1\text{-G1406T})_{postmating} - ln(\text{wild-type GPA1}/GPA1\text{-G1406T})_{premating}$.

*MAT*α cells. The ratio of the 2 *GPA1* alleles before and after mating was determined and the mating coefficient (s_m) of strains carrying the wild-type *GPA1* allele was calculated as the change in the natural logarithm of the ratio of the 2 alleles. The mating coefficients are all negative indicating that strains carrying the wild-type *GPA1* allele, which have a growth-rate advantage, have a disadvantage in mating relative to the *GPA1*-G1406T allele (Fig. 3*C*, $s_m = -27.2\% \pm 6.5\%$).

Discussion

In bacteria, gratuitous gene expression reduces growth rate (21–26). In the most carefully studied case, deregulation of the *lac* operon can reduce growth rate by ≈10% by diverting ribosomes from the synthesis of other proteins (25). This finding implies that any gene slows cell growth in proportion to how strongly it is expressed. We suspect that the cost of gene expression is not specific to bacterial enzymes or genes in the yeast mating pathway, but rather reflects a universal cost of gene expression and that this cost must be borne in all environments where the gene is expressed. In environments where the protein's expression increases fitness, this cost is offset by larger benefits, but it is never zero. Previous attempts to demonstrate a general fitness advantage from the elimination of dispensable genes in yeast have been unsuccessful (27, 28). Because 97% proteins are expressed at levels <0.1% of total cell protein (29), the growth-rate advantage for the majority of single gene deletions is below the limit of detection by current assays. However, because mating depends on the expression of 23 genes, mutations that eliminate signaling through the mating pathway eliminate expression of a number of genes and produce a measurable growth-rate advantage.

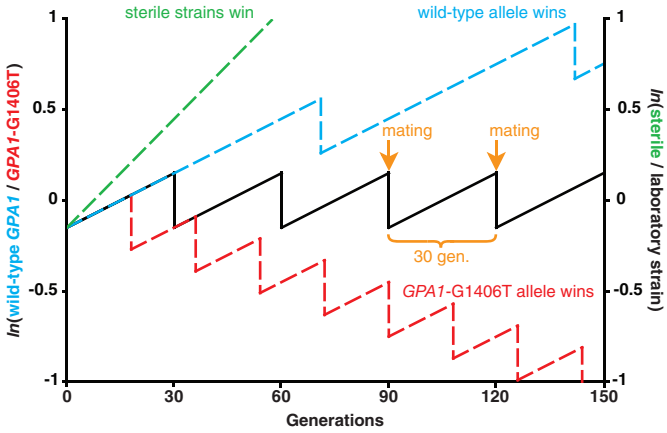

Fig. 4. A schematic of the evolutionary dynamics between the 2 alleles of *GPA1*. The wild-type *GPA1* allele has growth-rate advantage of ≈1% per generation, but a mating disadvantage of ≈30% per round of mating compared with the *GPA1*-G1406T allele. If these 2 strains are mixed and propagated in a regime where 1 round of mating occurs every 30 generations, these 2 strains would be equally fit (black trace). If mating occurs more frequently than every 30 generations, the *GPA1*-G1406T allele will win the competition (red trace). Conversely, if more than 30 generations pass between rounds of mating, the wild-type allele of *GPA1* will win (blue trace). During long-term evolution, strains are typically propagated asexually. Under such a circumstance, sterile strains, which eliminate basal signaling through the mating pathway, will outcompete mating-proficient strains (green trace, on the secondary *y* axis to demonstrate the ≈2% advantage versus strains carrying the *GPA1*-G1406T allele).

Given that gene expression is costly, it is surprising that some strains carry a mutation in *GPA1*, an upstream component of the mating pathway, which increases expression of the mating genes; this polymorphism is one of the strongest *trans*-acting regulatory polymorhphisms between laboratory and wild strains (18). Our results provide a plausible explanation for the existence of the *GPA1*-G1406T allele: This mutation increases the basal expression of genes in the mating pathway thus increasing mating efficiency by ≈30% at a growth-rate cost of ≈1% per generation. In the laboratory, cells are mated en masse, a condition that strongly selects for cells that produce more pheromone (20). For these *GPA1* variants, where we have determined the effect on growth rate and mating efficiency, we can determine under what conditions each allele will be favored. If these 2 strains are mixed and propagated in a regime where 1 round of mating occurs every 30 generations, these 2 strains would be equally fit (Fig. 4, black trace). If mating is less frequent than every 30 generations, the wild-type allele is favored, whereas if mating is more frequent than every 30 generations, the *GPA1*-G1406T allele is favored (Fig. 4). During long-term evolution, strains are typically propagated asexually. In this regime, sterile strains, which eliminate basal signaling through the mating pathway, will outcompete mating-proficient strains (Fig. 4, green trace).

The success of different *GPA1* alleles in the laboratory versus in wild strains reflects differences in the selective pressures defining fitness in these 2 environments. A similar phenomenon has been observed in *E. coli* where the laboratory strain B carries an allele of the arginine repressor, *argR*, that results in deregulation of the arginine biosynthesis pathway, providing a selective advantage under conditions where arginine availability fluctu-

ates rapidly (30). The *argR*^{B} allele appears to be a naturally occurring variant that is beneficial under laboratory conditions (31). To follow up on the possibility that the *GPA1* mutation arose during laboratory cultivation, we traced the lineage of S288c back to the wild diploid strain, EM93, which was isolated from a rotting fig near Merced, California in 1938 (32). We sequenced the *GPA1* genes in EM93, and found, somewhat to our surprise, that EM93 is homozygous for the *GPA1*-G1406T allele. Thus, we are left with some ambiguity as to evolutionary origin of this allele: If one discounts a strain mix-up, the mutation appears to have arisen in the wild, raising the question as to how much of laboratory domestication is the result of de novo mutation versus the selection of favorable combinations of naturally occurring alleles.

Gene loss is an important process in evolution. Extensive gene loss in protomitochondria and *Mycobacterium leprae* may have fostered the transition from facultative to obligate intracellular parasites (33), reciprocal gene loss after whole-genome duplication has reinforced species barriers by establishing Dobzhansky-Muller incompatibilities (34), and the loss of key developmental regulators early in vertebrate evolution has been suggested to have played a role in the establishment of modern phyla (35). Here, we provide evidence for a general cost of gene expression and find that elimination of the expression of 23 genes results in a 2% growth-rate advantage. Assuming that each of these genes contributes equally, the growth-rate advantage attained by eliminating a single dispensable gene is $<0.1\%$. The fate of mutations whose selection coefficient is $>1/N$ is dominated by selection; therefore, for population sizes greater than $\approx 10^3$, such as panmictic microbial populations, selection will oppose unnecessary gene expression, but for small or subdivided populations, drift will dominate for all but the small fraction of strongly expressed genes. Selection for sterile strains during long-term evolution and for the *GPA1*-G1406T allele supports the hypothesis that selection can optimize the level of gene expression to balance the cost of protein production and the demand for protein function, and argues that proteins that do not increase fitness will be lost.

Materials and Methods

Strains, Plasmids, and Media. The strains used in this experiment are derived from the base strain, DBY15084, a haploid yeast strain derived from the W303 background with genotype *MAT*a, *ade2-1*, *CAN1*, *his3-11*, *leu2-3,112*, *trp1-1*, *URA3*, *bar1*Δ::*ADE2*, and hmlα Δ::*LEU2*. The full genotypes of all strains used in this study are described in Table S1. Selection for αF- and canavanine-resistant clones was performed by spot plating on plates containing 10 μg/mL αF or 60 μg/mL l-canavanine as described in ref. 36. Integrative and plasmid transformations were performed by using standard yeast procedures (37). Primers used in this study are described in Table S2. Plasmids pTCN112 and pTCN113 were used for Gal overexpression of *STE4* and the dominant *FAR1-22*, respectively (38). Each strain was transformed with pTCN112 and pTCN113 by using standard techniques (37) and 2 transformants each were tested for growth on media containing 2% galactose and raffinose. pGIL025 was constructed by cloning a 2-kb fragment of *GPA1* from a strain carrying the wild-type allele of *GPA1*, Gpa1^{S469I}, marked by NatMX into pFA6. The insert consists of 388 bp from the 3′ end of *GPA1* and 128 bp homology after the NatMX marker. pGIL025 was used to transform DBY15084 to generate *GPA1* wild-type allele replacement strains and control strains where the NatMX reporter is integrated downstream of the *GPA1*-G1406T allele. To generate strains for the mating assays DBY15097 was crossed to DBY15098 to restore *BAR1*.

Competitive Growth-Rate Assays. Growth rate was determined by using 2 assays: A FACS-based competitive growth-rate assay (39, 40) and direct competition. For the FACS-based assay, cells were grown to mid-log phase and mixed ≈1:1 with a fluorescently labeled reference strain expressing a Cwp2-YFP fusion protein (40). Cultures were immediately diluted 1:500 into 10 mL of prewarmed YPD (yeast extract, peptone, 2% dextrose) and a 1-mL sample was spun down, resuspended in phosphate buffered saline with 0.1% Tween, and stored at 4 °C. Dilutions were repeated after 12 and 24 h. Samples collected at 0, 12, 24, and 36 h were prepared for FACS by vortex and sonication. The number of YFP-positive (reference) and non-fluorescent (experimental) cells was determined by using an LSRII flow cytometer (BD Biosciences) counting 30,000 total cells for each sample. At each time point a subset of the samples were counted by using a particle counter (Beckman-Coulter) to determine the number of generations between each sample point. The growth-rate coefficient (s_g) of each strain relative to the reference was calculated as the rate of the change in the *ln* ratio of experimental to reference versus generations (41). Of the 162 fitness assays performed in this study, only 13 had an R^2 value for the correlation between *ln* ratio versus generations of <0.925 (Fig. S5). In all cases, 1 of the 4 data points deviated from the trend, most often the result of 1 strain entering a lag phase after mixing. Once the offending data point was removed, s_g and R^2 values were recalculated for each experiment. The reference strain has an ≈3% disadvantage compared with wild type, therefore the values of s_g were normalized by using the equation $s_{g,\ \text{Exp vs. WT}} = (1 + s_{g,\ \text{Exp vs. Ref}})/(1 + s_{g,\ \text{WT vs. Ref}}) - 1$. The unselected clones and the spontaneously occurring αF and canavanine-resistant mutants were normalized to the median s_g of the unselected clones. The targeted gene disruptions were normalized to the median s_g of the *can1*Δ::KanMX4 gene disruption. The *GPA1* replacement strains were normalized to the median s_g of strains carrying the NatMX marker downstream of the S288c version of *GPA1*.

In addition to the FACS-based relative growth-rate assay, we used a direct competition to determine the difference in growth rate between strains carrying the *GPA1*-G1406T and the wild-type allele of *GPA1*. This assay was performed exactly as the relative growth-rate assay except that the ratio of the 2 strains was determined by quantitative sequencing. Briefly, genomic DNA was prepared from the mixed sample at each time point and the *GPA1* allele was amplified and sequenced in both directions. From the sequencing chromatogram, we calculated the ratio of 2 alleles by using peak height as a proxy for abundance. The growth-rate coefficient (s_g) of the wild-type *GPA1* allele was calculated as the rate of the change in the *ln* ratio of the wild-type *GPA1* allele to the *GPA1*-G1406T allele versus generations.

Yeast Tiling Arrays and Mutation Identification. Genomic DNA was prepared from 20 mL of saturated YPD culture by using QIAGEN Genomic-Tip 100/G and Genomic DNA buffers according to the manufacturer's instructions (QIAGEN). Five micrograms of total genomic DNA were fragmented by sonication (30 0.5-s pulses) by using a microtip sonicator (Misonix) and cleaned by using Zymo Cleanup kit according to the manufacturer's instructions (Zymo Research). Five hundred nanograms of fragmented DNA were labeled by using BioPrime labeling kit according to the manufacturer's instructions with half-volume reactions (Invitrogen). Hybridization to Affymetrix yeast tiling arrays and identification of single nucleotide polymorphisms (SNPs) by using the SNPScanner algorithm were done as described in ref. 42. Seven αF^{R} mutants, along with the reference strain DBY15084 were each hybridized to an array. The strain used in this study contains ≈8,000 SNPs relative to the S288c strain on which the array is based (43). The output files from SNPScanner were viewed by using the Integrated Genome Browser software (Affymetrix) and scanned by eye for SNPs present in the αF^{R} strains but absent from the reference strain. Among the 7 strains tested, 9 SNPs were identified and a 600-bp region centered at each SNP was sequenced from the reference strain and 2 clones from the αF^{R} strain in which the SNP was identified. Four of the 9 SNPs were present only in the clone used for array analysis, 5 SNPs were identified in both clones, and none of the SNPs were identified in the reference strain.

Yeast Gene Expression Microarrays. Cells were harvested from 10-mL YPD cultures at mid log (≈2×10^7 cells per mL) by vacuum filtration onto a 25-mm nylon membrane, snap frozen in liquid nitrogen, and stored at −80 °C. RNA preparation, labeling, hybridization, and data acquisition were performed as described in ref. 44. Briefly, crude RNA was extracted by the acid phenol lysis protocol and cleaned by using a QIAGEN RNeasy column. Cy3- and Cy5-labeled CTP were incorporated by using the Agilent Low RNA Input Fluorescent Linear Amplification kit. Each Cy5-labeled sample was mixed with the common Cy3-labeled reference strain (DBY15084) and hybridized to an Agilent yeast gene expression 4×44-k array (Agilent Technologies). Arrays were scanned by using an Agilent DNA microarray scanner and analyzed by using Agilent's Feature Extraction software. Each Agilent yeast gene expression 4×44 k array contains 7 identical probes for each *S. cerevisiae* gene included on the array. Combining the final processed intensities from 7 of the fittest of the spontaneous αF^{R} mutants gives us 49 estimates for the hybridization intensity in the Cy5 (experimental) and Cy3 (reference) channel. *P* values were calculated by using a 2-tailed *t* test of the Cy5 and Cy3 hybridization intensities for each

GENETICS

gene. Significant changes reported in Fig. 3 were determined by eye from a volcano plot of log_2(red/green) versus P value and correspond to a 1.5-fold change in expression and $P < 10^{-31.5}$ (Fig. S2).

Competitive Mating Assays. Competitive mating assays were performed by competing *MAT***a** cells carrying the *GPA1*-G1406T allele with cells carrying the wild-type allele of *GPA1* (marked with NatMX) for a limited number of *MAT*α cells. Each strain was grown to mid-log phase ($\approx 10^7$ cells per mL) then 5×10^6 cells of each *MAT***a** strain added to 10 mL of YPD + ADE. A sample was diluted, sonicated, and plated to single colonies on -leu media. *MAT*α mating tester (2×10^6 cells) was added to the *MAT***a** mixture and filtered onto a 25-mm 0.45-μm nylon filter. Cells were mated on a YPD plate at 30°C. After 5 h, the filters were washed and the cells were diluted, sonicated, and plated to single colonies onto minimal media to select for diploids. Colonies on the -leu and minimal plates were replica plated to YPD plates containing ClonNat to determine the ratio of strains carrying the wild-type and G1406T alleles before and after mating. The mating coefficient (s_m) was calculated as the change in the natural logarithm of the ratio of the 2 alleles: $s_m = ln$(wild-type *GPA1*/*GPA1*-G1406T)$_{postmating} - ln$(wild-type *GPA1*/*GPA1*-G1406T)$_{premating}$.

Notebook. The complete laboratory notebook describing these experiments is available at http://www.genomics.princeton.edu/glang/notebooks.htm.

ACKNOWLEDGMENTS. We thank M. Desai, L. Kruglyak, and members of the Botstein and Murray labs for comments on the manuscript; J. Shapiro and L. Kruglyak (Princeton University, Princeton, NJ) for providing the wild-type *GPA1* allele; B. Tilton and C. DeCoste for assistance with flow cytometry; and A. Ward, D. Gresham, J. Buckles, and D. Storton for assistance with microarrays. This work was supported by National Institute of General Medical Sciences Centers of Excellence Grants GM068763 (to A.W.M.) and GM071508 (to D.B.) and National Institutes of Health Grants GM43987 (to A.W.M.) and GM046406 (to D.B).

1. Gilad Y, Przeworski M, Lancet D (2004) Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2:E5.
2. Protas ME, et al. (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet* 38:107–111.
3. Hittinger CT, Rokas A, Carroll SB (2004) Parallel inactivation of multiple *GAL* pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci USA* 101:14144–14149.
4. Bolotin A, et al. (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* 22:1554–1558.
5. Sleight SC, Wigginton NS, Lenski RE (2006) Increased susceptibility to repeated freeze-thaw cycles in *Escherichia coli* following long-term evolution in a benign environment. *BMC Evol Biol* 6:104.
6. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407:736–739.
7. Cooper VS, Bennett AF, Lenski RE (2001) Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. *Evolution* 55:889–896.
8. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J Bacteriol* 183:2834–2841.
9. Bennett AF, Lenski RE (2007) An experimental test of evolutionary trade-offs during temperature adaptation. *Proc Natl Acad Sci USA* 104(Suppl 1):8649–8654.
10. Zeyl C, Curtin C, Karnap K, Beauchamp E (2005) Antagonism between sexual and natural selection in experimental populations of *Saccharomyces cerevisiae*. *Evolution* 59:2109–2115.
11. Protas M, et al. (2007) Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. *Curr Biol* 17:452–454.
12. Herskowitz I (1995) MAP kinase pathways in yeast: For mating and more. *Cell* 80:187–197.
13. Roberts CJ, et al. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* 287:873–880.
14. Hagen DC, McCaffrey G, Sprague GF, Jr (1991) Pheromone response elements are necessary and sufficient for basal and pheromone-induced transcription of the FUS1 gene of *Saccharomyces cerevisiae*. *Mol Cell Biol* 11:2952–2961.
15. Bashir T, Pagano M (2004) Don't skip the G1 phase: How APC/CCdh1 keeps SCFSKP2 in check. *Cell Cycle* 3:850–852.
16. Tanaka-Matakatsu M, Thomas BJ, Du W (2007) Mutation of the Apc1 homologue shattered disrupts normal eye development by disrupting G1 cell cycle arrest and progression through mitosis. *Dev Biol* 309:222–235.
17. Irniger S, Nasmyth K (1997) The anaphase-promoting complex is required in G1 arrested yeast cells to inhibit B-type cyclin accumulation and to prevent uncontrolled entry into S-phase. *J Cell Sci* 110(Pt 13):1523–1531.
18. Yvert G, et al. (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 35:57–64.
19. Hirsch JP, Dietzel C, Kurjan J (1991) The carboxyl terminus of Scg1, the G alpha subunit involved in yeast mating, is implicated in interactions with the pheromone receptors. *Genes Dev* 5:467–474.
20. Jackson CL, Hartwell LH (1990) Courtship in *S. cerevisiae*: Both cell types choose mating partners by responding to the strongest pheromone signal. *Cell* 63:1039–1051.
21. Andrews KJ, Hegeman GD (1976) Selective disadvantage of non-functional protein synthesis in *Escherichia coli*. *J Mol Evol* 8:317–328.
22. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436:588–592.
23. Koch AL (1983) The protein burden of *lac* operon products. *J Mol Evol* 19:455–462.
24. Novick A, Weiner M (1957) Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci USA* 43:553–566.
25. Stoebel DM, Dean AM, Dykhuizen DE (2008) The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. *Genetics* 178:1653–1660.
26. Zamenhof S, Eichhorn HH (1967) Study of microbial evolution through loss of biosynthetic functions: Establishment of ''defective'' mutants. *Nature* 216:456–458.
27. MacLean RC (2007) Pleiotropy and *GAL* pathway degeneration in yeast. *J Evol Biol* 20:1333–1338.
28. Sliwa P, Korona R (2005) Loss of dispensable genes is not adaptive in yeast. *Proc Natl Acad Sci USA* 102:17670–17674.
29. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
30. Suiter AM, Banziger O, Dean AM (2003) Fitness consequences of a regulatory polymorphism in a seasonal environment. *Proc Natl Acad Sci USA* 100:12782–12786.
31. Merlo LM, Sadowsky MJ, Ferguson JA, Dean AM (2006) The argRB of *Escherichia coli* is rare in isolates obtained from natural sources. *Gene* 376:240–247.
32. Mortimer RK, Johnston JR (1986) Genealogy of principal strains of the yeast genetic stock center. *Genetics* 113:35–43.
33. Cole ST, et al. (2001) Massive gene decay in the leprosy bacillus. *Nature* 409:1007–1011.
34. Scannell DR, et al. (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440:341–345.
35. De Robertis EM (2008) Evo-devo: Variations on ancestral themes. *Cell* 132:185–195.
36. Lang GI, Murray AW (2008) Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178:67–82.
37. Sherman F, Fink G, Lawrence C (1974) *Methods in Yeast Genetics* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
38. Norman TC, et al. (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285:591–595.
39. Thompson DA, Desai MM, Murray AW (2006) Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. *Curr Biol* 16:1581–1590.
40. Desai MM, Fisher DS, Murray AW (2007) The speed of evolution and maintenance of variation in asexual populations. *Curr Biol* 17:385–394.
41. Hartl D (2000) *A Primer of Population Genetics* (Sinauer Associates, Sunderland, MA).
42. Gresham D, et al. (2006) Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. *Science* 311:1932–1936.
43. Schacherer J, et al. (2007) Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS ONE* 2:e322.
44. Brauer MJ, et al. (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell* 19:352–367.