

togenesis was used to create two "kinetic bottlenecks" in the reaction pathway of isocitrate dehydrogenase, greatly extending the lifetime of each of two transient intermediates, enabling their observation. In the present context, it is possible that the lattice contacts themselves have created such a kinetic bottleneck, enabling us to freeze-trap a conformational intermediate prior to RNA catalytic cleavage. Kinetic bottleneck mutants of the hammerhead RNA should allow capture of additional cleavage reaction intermediates.

REFERENCES AND NOTES

- C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **35**, 849 (1983); A. J. Zaugg, T. R. Cech, *Science* **231**, 470 (1986).
- H. W. Pley, K. M. Flaherty, D. B. McKay, *Nature* **372**, 111 (1994).
- W. G. Scott, J. T. Finch, A. Klug, *Cell* **81**, 991 (1995).
- O. C. Uhlenbeck, *Nature* **328**, 596 (1987); D. E. Ruffner, G. D. Stormo, O. C. Uhlenbeck, *Biochemistry* **29**, 10695 (1990).
- D. B. McKay, *RNA* **2**, 395 (1996).
- K. Moffat and R. Henderson, *Curr. Opin. Struct. Biol.* **5**, 656 (1995).
- W. Scott and A. Klug, *Trends Biochem. Sci.* **21**, 220 (1996).
- S. Dahm and O. C. Uhlenbeck, *Biochemistry* **30**, 9464 (1991); S. Dahm, W. Derrick, C. Uhlenbeck, *ibid.* **32**, 13040 (1993).
- J. B. Murray, A. K. Collier, J. R. P. Arnold, *Anal. Biochem.* **218**, 177 (1994); W. G. Scott et al., *J. Mol. Biol.* **250**, 327 (1995).
- T. Tushl et al., *Science* **266**, 785 (1994); S. T. Sigurdsson, T. Tushl, F. Eckstein, *RNA* **1**, 575 (1995); K. M. A. Amiri and P. J. Hagerman, *J. Mol. Biol.* **261**, 125 (1996).
- K. J. Hertel and O. C. Uhlenbeck, *Biochemistry* **34**, 1744 (1995).
- J.-H. Yang et al., *ibid.* **29**, 11156 (1990).
- H. van Tol et al., *Nucleic Acids Res.* **18**, 1971 (1990); G. Slim and M. J. Gait, *ibid.* **19**, 1183 (1991); M. Koizumi and E. Ohtsuka, *Biochemistry* **30**, 145 (1991).
- G. S. Bassi et al., *Nature Struct. Biol.* **2**, 45 (1995).
- J. B. Murray, C. J. Adams, J. R. P. Arnold, P. G. Stockley, *Biochem. J.* **311**, 487 (1995).
- J. M. Bolduc et al., *Science* **268**, 1312 (1995).
- Diffraction data were processed with DENZO [Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend*, L. Sawyer, N. Isaacs, S. Bailey, Eds. [Daresbury, U.K., and SERC, 1993], pp. 56–62]. Further details are described in Table 1.
- Molecular replacement was done with the AMoRe (automated molecular replacement) software distributed with CCP4 (19) and, as a probe, the dimer structure determined from the previous crystal form. The top translation function solution had an *R* factor of 53% and a correlation coefficient of 28. However, 10 cycles of rigid-body refinement in AMoRe reduced the *R* factor to 31%, and the correlation coefficient simultaneously increased to 86. Further rigid-body refinement followed by conventional positional refinement (Powell minimization) in X-PLOR 3.1 [A. T. Brünger, *X-PLOR 3.1: A System for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1993)] further reduced the *R* factor to 26%. The initial model of one hammerhead ribozyme molecule in the crystal asymmetric unit was further refined with a standard simulated annealing slow-cooling molecular dynamics protocol followed by conventional positional and restrained temperature factor refinement in X-PLOR 3.1 with data from 8.0 to 3.0 Å resolution and a modified RNA geometry parameter library [G. Parkinson, J. Vojtechovsky, L. Clowney, A. T. Brünger, H. M. Berman, *Acta Crystallogr.* **D52**, 57 (1996)]. Finally, the low resolution data were incorporated, a solvent mask was determined,

- and partial calculated structure factors were generated to model the bulk solvent contribution to the x-ray scattering amplitudes for further refinement within X-PLOR 3.1. The metal-bound crystal structures were refined identically (Table 1).
- Collaborative Computational Project, Number 4 [CCP4], *Acta Crystallogr.* **D50**, 760 (1994).
 - We thank P. Stockley, J. Finch, G. Varani, S. Price, K. Nagai, O. Uhlenbeck, and members of their research groups for advice; R. Sweet for help with data collec-

tion at Brookhaven synchrotron beamline X12C, and the Daresbury Laboratory for additional data collection; D. McKay and K. Flaherty for discussions and for providing us with an improved set of stereochemical parameters (18) for RNA refinement. Supported by the Medical Research Council of the United Kingdom and the American Cancer Society (W.G.S., PF-3970), and the NIH (B.L.S., GM-49857).

21 May 1996; accepted 26 October 1996

Functional Analysis of the Genes of Yeast Chromosome V by Genetic Footprinting

Victoria Smith, Karen N. Chou, Deval Lashkari, David Botstein, Patrick O. Brown*

Genetic footprinting was used to assess the phenotypic effects of Ty1 transposon insertions in 268 predicted genes of chromosome V of *Saccharomyces cerevisiae*. When seven selection protocols were used, Ty1 insertions in more than half the genes tested (157 of 268) were found to result in a detectable reduction in fitness. Results could not be obtained for fewer than 3 percent of the genes tested (7 of 268). Previously known mutant phenotypes were confirmed, and, for about 30 percent of the genes, new mutant phenotypes were identified.

The completion of the sequences of the genomes of several microorganisms is a watershed for the new science of genomics. The next important challenge is to determine, in an efficient and reliable way, something about the function of each gene in these genomes. The 12,057-kb nonrepetitive portion of the *S. cerevisiae* genome—the first completely sequenced eukaryotic genome—contains 6000 to 6500 predicted genes, of which fewer than half had previously been known. A still smaller fraction of the genes of yeast have been characterized experimentally with respect to biological function; indeed, previous work suggested that disruption of yeast genes resulted in a readily discernible phenotype only about 30% of the time (1). Here, we describe the results of genetic footprinting (2) as applied to 268 predicted protein-coding genes on chromosome V of *S. cerevisiae* (3).

We subjected a large population of haploid yeast cells ($\sim 10^{11}$ cells) to mutagenesis by transiently inducing transposition of a marked Ty1 transposable element. DNA was extracted from a portion of this culture (the "time-zero" DNA). Other representa-

tive samples of this population were subjected to one of several selections (Table 1). DNA was extracted from the cells recovered after each selection. The presence and relative abundance of cells carrying Ty1 insertions within a gene of interest was assessed for each of these samples by means of a polymerase chain reaction (PCR) (2). In general, for each gene, we surveyed a minimum of 500 to 900 base pairs (bp) of coding sequence, along with 400 to 600 bp of upstream sequence. Smaller genes (<700 bp) were analyzed in their entirety, along with several hundred base pairs of sequence flanking the start and stop codons. A growth disadvantage to cells carrying insertions in a gene, under a particular selection, was reflected by the loss or depletion of the PCR products representing those insertions (the "genetic footprint") when DNA samples from the selected population were compared with the time-zero DNA samples. The method not only detects severe growth disadvantages, but also sensitively measures moderate reductions in fitness.

For each predicted protein-coding sequence on chromosome V, a color was assigned on the basis of whether a particular selection protocol resulted in a perceptible depletion of the PCR products representing insertions in that coding sequence (Fig. 1). Overall, we were able to obtain satisfactory genetic footprinting data for 261 (97%) of the predicted protein-coding genes (4, 5). This total includes six putative genes contained in repeated telomeric sequences (boxed in Fig. 1): cells with mutations in any of these six genes appeared wild type for

V. Smith and K. N. Chou, Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA.

D. Lashkari and D. Botstein, Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA.

P. O. Brown, Howard Hughes Medical Institute and Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: pbbrown@cmgm.stanford.edu

growth in all selections. However, because a unique priming site was found for only one of these genes, they are excluded from the following discussion. Of the remaining 255 genes, we could detect a phenotype that distinguished the corresponding mutants from wild-type cells for 157 genes (61.6%). For 98 genes (38.4%), Ty1 insertions resulted in a reduction in growth of less than 5 to 10%, or a decrease in mating proficiency (as

an α) of less than ~50%; that is, the mutants were indistinguishable from the wild type for growth or mating, within the sensitivity of our assays.

The most common mutant phenotype we found among all the genes analyzed was a general growth disadvantage for mutant cells under all selections. By analyzing multiple time points using the least stringent selection for growth—growth in rich-glu-

cose medium at 25°C (Table 1)—we could estimate the growth rates of each mutant. We divided the results into three categories: Q1, Q2, and Q3. Q1 indicates that mutant cells grew at no more than 75% of the growth rate of the overall population (for example, after 15 population doublings the mutants had doubled no more than 12 times, and thus were represented at no more than $1/2^3 = 1/8$ of their abundance in the

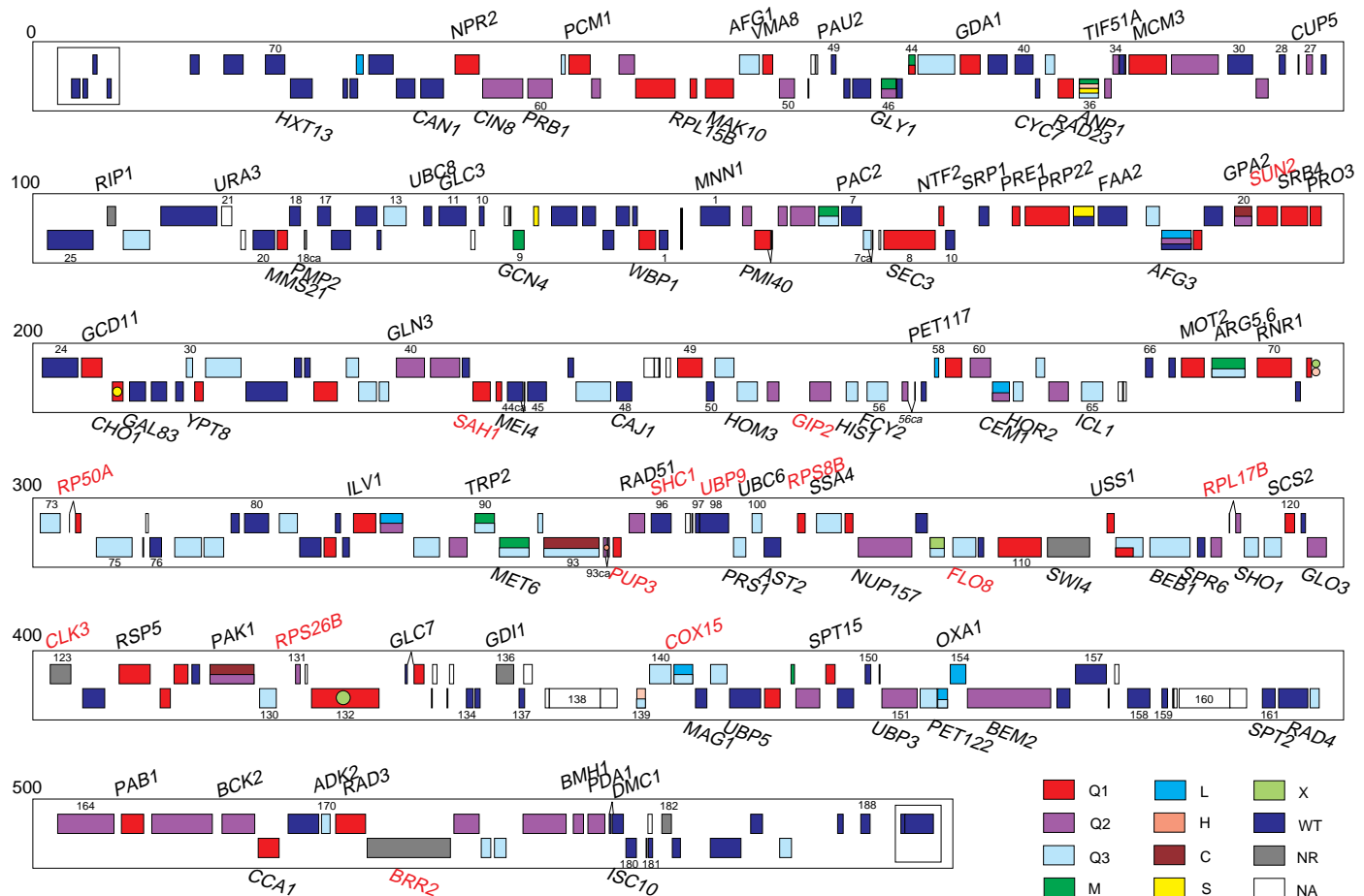


Fig. 1. Representation of chromosome V incorporating genetic footprinting data. Genes are represented by boxes of the appropriate size, which are placed on the upper or lower half of each panel to indicate their sense orientation on the chromosome [upper, w (Watson); lower, c (Crick)]. The scale on the left is in kilobases. Genes are numbered with their systematic genome sequencing numbers (left and right, starting from the centromere), as defined for chromosome V (3); the centromere is represented by the vertical black line at 149 kb, and at least every tenth gene is numbered. Names of genes that have been described in publications are indicated in black adjacent to the box representing the corresponding gene on the map. Names of genes for which some preliminary data (or the name of a person who can be contacted for more information) are available in the SacchDB database (<http://www-genome.stanford.edu>) and MIPS database (<http://www.mips.biochem.mpg.de>), but not yet published, are indicated in red. Each box is colored to summarize the results obtained to date from genetic footprinting. Mutant phenotypes: Q1, all selections, <75% population growth rate; Q2, all selections, 75 to 85% population growth rate; Q3, all selections, 85 to <100% population growth rate. Genes in which mutations gave distinctive phenotypes under specific selections are coded according to the selection(s) affected, as follows: M, minimal medium; L, lactate medium; H, high temperature; C, caffeine; S, high salt; X,

mating; WT, wild type for all selections; NR, no result, or results too ambiguous to interpret; and NA, not analyzed. See the expanded legend at Science's Web site (<http://www.sciencemag.org>) for (i) an estimate of the degree of the growth defect for mutants with reduced fitness in specific selections; (ii) possible additional, more subtle phenotypes (for example, a growth difference of <5% in rich medium, or 10% in other selections), which were observed for 11.6% of the genes; and (iii) other observations regarding particular genes or particular analyses (such as alternative interpretations of the footprints observed, or weak data for any specific selection). The only significant failure rate for specific selections was for mate DNA: usable data were not obtained for nine of the genes (3.4%) for which usable data were otherwise obtained with other selections. This slightly higher rate of PCR failure may be attributable to the presence of the extra (diploid) genome. For simplicity, the boxes are completely colored, even though for many genes only a portion of the full coding sequence was analyzed (almost always the portion encoding the predicted NH₂-terminus of the encoded protein). A minimum of 700 bp was analyzed for each gene, and the full coding sequence of many genes was analyzed. Genes contained in full-length Ty elements were not analyzed. *URA3* was not analyzed because the strain used in this study is a *ura3-52* mutant. Other white boxes represent solo δ , τ , and σ elements, and RNA genes.

time-zero population). This category includes genes whose product is absolutely required for vegetative cell growth ("essential" genes). The Q2 and Q3 categories include genes for which mutant cells were at more subtle growth disadvantages, growing at apparent rates of 75 to 85% and 85 to <100% of the population growth rate, respectively (Fig. 1 and Table 2). None of the phenotypes in these classes was a consequence of the freezing of cells in glycerol after mutagenesis, or of their subsequent resuscitation (6). These data were confirmed by analysis of DNA isolated from three independent rich-medium selections. In addition, the data obtained from the other selections corroborated the general growth defects observed in rich-glucose medium. For >90% of the genes in the Q2 and Q3 categories, growth deficits were reproducibly detected to within $\pm 5\%$ variability in independent samples.

Fifty-one genes (20%) fell into the Q1 category; 14 of these were previously undescribed (Table 2 and Figs. 1 and 2). In addition to several known essential genes, this category included *ILV1* and *PRO3*, mutations of which are known to result in auxotrophy. As described previously for *ade2* insertions (2), genetic footprinting revealed considerable growth defects for *ilv1* and *pro3* mutant cells even in rich medium thought to contain adequate quantities of the relevant nutrients [the inability of *pro3* mutants to grow in rich medium was reported previously (7)]. *MOT2*, which encodes a transcriptional repressor, also fell in this class; although null mutants were viable at 25°C, their growth defect was clearly exposed by genetic footprinting analysis (8). Other genes previously reported to be non-essential, such as *VMA8* (encoding the vacuolar H-adenosine triphosphatase subunit), *GDA1* (encoding guanosine diphosphatase

of the Golgi membrane), and *MMS21* (encoding a DNA repair protein) (9), were found to fall in the Q1 category, which indicated a substantial growth disadvantage of the corresponding mutant strains.

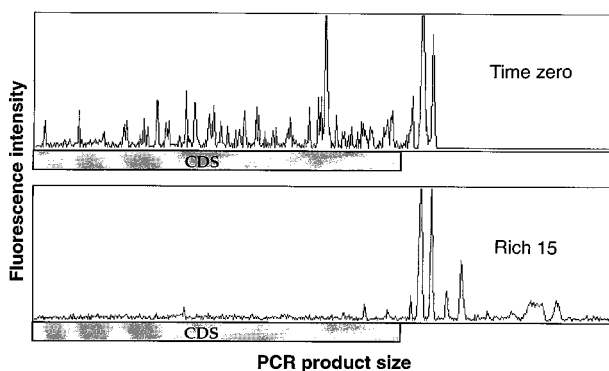
Some selection against cells with mutations in essential genes may occur during the 4-day Ty1 mutagenesis. It is possible that this factor accounted for some cases in which we failed to obtain an interpretable result (3% of the genes analyzed) (10). Most essential genes were not excluded from analysis in this way, as we were reliably able to obtain PCR products representing Ty1 insertions in almost all of the previously identified essential genes on chromosome V with the use of DNA from the time-zero cells.

Insertion mutations in 99 genes (38.8%) resulted in more subtle quantitative growth defects in all selections. These genes were divided into two classes (Q2 and Q3, Fig. 1 and Table 2) on the basis of the estimated growth rate disadvantage of mutants. Most of the genes in these groups (58%) had not been previously characterized. Moreover, for many of the previously characterized genes in these categories, the growth disadvantage we observed had not previously been reported. An example of such a gene is *SSA4*, for which we found a Q3 mutant growth defect. This gene is a member of a large family of genes that encode apparently interchangeable cytoplasmic heat shock proteins (11). We also observed a Q2 growth disadvantage for mutants of *RAD51* (encoding a RecA-like DNA repair protein), *PRB1* (encoding vacuolar protease B), and *GLN3* (encoding a positive nitrogen-regulatory protein) (12).

Spurious PCR products (those unrelated to real Ty1 insertions in the gene of interest) would not be expected to be depleted under a selection that requires the gene's

Table 1. Selections used in genetic footprinting analysis. The time points at which DNA samples were isolated correspond to the given numbers of population doublings in each selection after Ty1 mutagenesis. Each gene was analyzed using DNA (1 μ g) isolated at a primary time point for each selection; the pattern of PCR products was compared with that obtained with the time-zero DNA sample. Secondary time points were used to confirm potential growth defects or to resolve ambiguities identified by the primary analysis. Analysis of both primary and secondary time points was useful for confirmation of general quantitative growth defects as well as for identification of growth defects in specific selections. As a control to identify any nonspecific PCR products, each gene-specific primer was also used with the Ty1-specific primer for a PCR that used DNA isolated from cells in which Ty1 transposition had not been induced. Ty1 transposition mutagenesis was performed as described (2). At least 4×10^5 cells were transferred to the appropriate medium for each selection, and cell density was maintained at 1×10^5 to 3×10^7 cells per milliliter (or less) over the course of each selection. In some cases, cell density was allowed to reach 5×10^7 to 1×10^8 cells per milliliter for harvesting at the final time point. All selections were performed at 25°C, except high-temperature growth (36.5°C) and mating (30°C). The media for all selections contained 2% glucose, except rich-lactate medium (2% lactate). The pH of the media was ~ 5.5 . Standard rich medium, with supplements for auxotrophies of the host strain, was used for the rich-medium, lactate, high-temperature, caffeine, and high-salt selections (1% yeast extract, 2% bactopectone, 0.008% tryptophan, 0.0026% adenine, 0.0022% uracil, and 0.0046% histidine). The medium for the caffeine selection contained, in addition, 6 mM caffeine, and the high-salt selection medium included 0.9 M NaCl. The minimal medium used for selection lacked amino acids and nucleotides, except as required by the auxotrophy of the strain (0.67% yeast nitrogen base, 0.0022% uracil, and 0.0046% histidine). For the mating selection, mutagenized cells (*his3 HIS4*) were grown in rich-glucose medium for 4 hours, then mixed with a threefold excess of cells of opposite mating type (*HIS3 his4*), collected on a filter, and incubated on rich-glucose medium at 30°C for 7 hours. Cells were then washed into SC-histidine liquid medium and incubated at 30°C for 16 hours (~ 10 population doublings) to select the HIS⁺ diploid products of successful mating. To compensate for the presence of the extra haploid genome, we used 2 μ g of DNA from the selected diploids for each PCR.

Fig. 2. *YER083c*, an example of a novel gene identified as important for growth under all selections (Q1), including rich-glucose medium. These data were generated by analyzing 10 μ l of each PCR reaction by denaturing polyacrylamide gel electrophoresis, as described (2). In this and subsequent figures, the most intense peaks (at the right of each trace) are shown off-scale to allow clearer visualization of the lower-intensity peaks; the shaded box under each trace represents the coding sequence of the gene (CDS), and peaks to the right of this box correspond to insertions upstream of the start ATG codon. For most genes on chromosome V, these peaks tended to be more intense than peaks corresponding to insertions in the coding sequence, reflecting a preference for Ty1 insertion into non-coding regions. In this example, all peaks corresponding to insertions in the coding sequence of *YER083c* were depleted after 15 population doublings in rich medium (Rich 15).



sequence of the gene (CDS), and peaks to the right of this box correspond to insertions upstream of the start ATG codon. For most genes on chromosome V, these peaks tended to be more intense than peaks corresponding to insertions in the coding sequence, reflecting a preference for Ty1 insertion into non-coding regions. In this example, all peaks corresponding to insertions in the coding sequence of *YER083c* were depleted after 15 population doublings in rich medium (Rich 15).

| Selection | Primary time point (population doublings) | Secondary time points (population doublings) |
|---------------------|---|--|
| Rich medium | 18 | 5, 12, 15, 23 |
| Rich medium | 51 | 56, 60 |
| Minimal medium | 18 | 10, 15 |
| Rich-lactate medium | 18 | 11 |
| Caffeine | 18 | 12 |
| High temperature | 18 | 10 |
| High salt | 10 | |
| Mating | 15 | |

Downloaded from www.sciencemag.org on November 12, 2009

activity, leading to underestimation of the selective disadvantage of the mutants. A similar behavior would be expected for PCR products that represent actual insertions within a gene that do not impair its function. PCR products corresponding to tolerated Ty1 insertions were frequently observed immediately adjacent to the 5' and 3' boundaries of coding sequences and were even observed within coding sequences (13). Because we analyzed multiple time points, we were usually able to determine whether these anomalous products were consistently present and omitted them from the quantitative analysis. The 50- to 60-generation time points were particularly useful for this purpose (14).

When a gene is important for cell growth, the dilution of the corresponding mutant cells as the general cell population expands leaves progressively fewer targets for PCR amplification. Reduced competition for PCR reagents could then lead to more efficient amplification of these remaining DNA targets. For example, PCR products representing insertions upstream of an important coding region were often more abundant among the products amplified with the use of DNA samples from the corresponding selection than among the products of the time-zero PCR. We tried to take this effect into account by normalizing the signal produced in different PCRs relative

to insertions upstream of the coding regions, under the assumption that these upstream insertions had no effect on fitness. To provide a more robust normalization, we developed an independent internal standard for the PCR reactions. A library of Sau 3A restriction fragments of yeast genomic DNA was cloned into a vector carrying the marked Ty1 primer sequence, such that PCR amplification of this library with the Ty1-specific primer and any labeled gene-specific primer resulted in a predictable pattern of products (15). DNA from the library was mixed at a fixed concentration with DNA samples from selected cells and time-zero cells, respectively. The mixtures were subjected to a second set of PCR reactions for each of 63 genes that were quantitatively important for growth (16). The intensities of the PCR products from the different selected DNAs were then normalized using the library-specific peaks (Fig. 3). The growth rate defect of the mutant cells was estimated by comparing the normalized signals from the different time points. These data facilitated the grouping of genes into the Q classifications and were also useful in resolving ambiguous results concerning general or specific growth defects.

Ty1 insertion mutations in a few genes resulted in growth defects that were discernible only in one of the selection protocols. Mutations in several other genes resulted in general growth defects as well as more severe specific ones. Genes required for growth in minimal medium include four previously characterized genes involved in amino acid

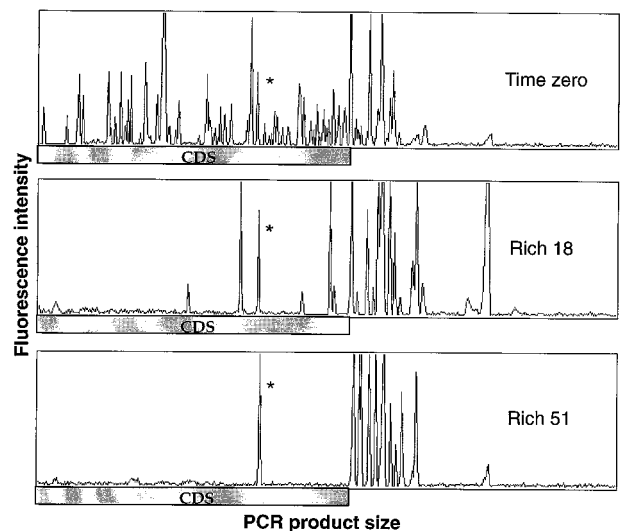
biosynthesis [GLY1, TRP2, MET6, and ARG5,6 (17)] and two novel genes, YER006w and YEL044w (18). (Novel genes are genes that have not been previously characterized and have not had a putative function assigned on the basis of compelling homology to a characterized gene.) Mutations in GLY1, ARG5,6, MET6, TRP2, and YER006w also produced more subtle growth defects in rich medium (Q2 for GLY1, Q3 for the others). GCN4, which encodes a transcriptional activator of amino acid biosynthetic genes, displayed a Q2 phenotype only in minimal medium (19). Mutations in YER146w produced a Q3 growth defect only in minimal medium. Mutations in ANP1, which encodes a protein involved in retention of glycosyltransferases in the Golgi apparatus (20), produced numerous phenotypes: Q1 (possibly Q2) in minimal medium, in high-salt medium, and at high temperature, and Q3 in all other selections.

Genes known to be required for respiration (PET117, PET122, CEM1, OXA1, and AFG3) behaved as expected, showing at least a Q1 defect in rich-lactate medium (21, 22). YER141w was also important for growth on lactate, consistent with its identification as COX15 (23). Mutations in most of these genes resulted in less severe general growth defects in all other selections as well. One novel gene, YER087w, was important for growth in lactate medium (Q1 in lactate) and was quantitatively important for growth in rich medium (Q2). Mutants of YEL066w

Table 2. Numbers of genes giving mutant phenotypes in each selection. The percentage of genes associated with each phenotype was calculated with respect to the 255 nontelomeric genes on chromosome V for which reliable data were obtained. Some genes were associated with more than one phenotype. For 98 genes (38.4%), Ty1 insertion mutants were indistinguishable from wild-type cells in any of the selections we used. R refers to genes for which mutants with general severe growth defects (Q1 or Q2) were able to grow at notably improved rates under particular selections. Many of the phenotypes we detected for previously characterized genes are, in fact, novel phenotypes (not listed). These new phenotypes were typically Q2 or Q3 (sometimes Q1) growth disadvantages for mutants of genes previously described as wild type for vegetative growth.

| Selection | Number of genes (%) | Number of novel genes |
|-----------------------|---------------------|-----------------------|
| Rich: Q1 (<75%) | 51 (20.0) | 14 |
| Rich: Q2 (75 to 85%) | 44 (17.3) | 22 |
| Rich: Q3 (85 to 100%) | 55 (21.6) | 35 |
| Minimal | 9 (3.5) | 3 |
| Lactate | 8 (3.1) | 1 |
| Caffeine | 3 (1.2) | 1 |
| High temperature | 2 (0.8) | 1 |
| High temperature R | 1 (0.4) | 1 |
| High salt | 3 (1.2) | 2 |
| High salt R | 1 (0.4) | 0 |
| Mating | 1 (0.4) | 1 |
| Mating R | 2 (0.8) | 2 |

Fig. 3. Quantitative depletion of cells with mutations in *BEM2*, analyzed by normalization with the Sau 3A library DNA. In this example, the coding sequence matches the region from which peaks were depleted. The peak in each tracing that represents the PCR product derived from the library DNA is marked with an asterisk. This peak increases in intensity at the time points corresponding to 18 and 51 population doublings (Rich 18 and Rich 51), as the reduced number of template DNA molecules corresponding to Ty1 insertions in *BEM2* results in decreased competition for PCR reagents. The area under the control peak was used as the basis for normalizing the total signal obtained from the various selection time point DNAs, relative to the time-zero sample, in the region undergoing depletion. In this example, 8% of the normalized signal remains at 18 population doublings, corresponding to a mutant growth rate of ~80% of the population growth rate. These data are consistent with published data for *BEM2*; The *bem2* mutants are viable at 26°C but have a slower growth rate (37). For each analysis in which library DNA was used, one PCR (and corresponding gel lane) contained library DNA alone, to allow identification of library-specific peaks. The library DNA, isolated from $\sim 1.5 \times 10^6$ independent colonies, was estimated to represent ~75 to 80% of Sau 3A sites; hence, the 8 pg of library DNA added to each PCR would represent, on average, ~30 molecules per peak.



had a moderate growth defect (Q2) only in lactate medium.

Mutants of three genes were found to have growth defects (Q1) in medium containing 6 mM caffeine. Two of these, *PAK1* and *GPA2* (24), encode a protein kinase and a guanosine triphosphate-binding regulatory protein, respectively; the third, *YER093c*, was a novel gene (Fig. 4). Mutations in all three of these genes also resulted in a less severe general growth disadvantage (Q2 or Q3) (Fig. 4). One novel gene, *YER139c*, was identified as important for growth at high temperature. Two novel genes, *YEL008w* and *YER014w*, were found to be important for growth in high-salt medium.

There were some examples of unexpected survival in particular selections. *CHO1*, which encodes a phosphatidylserine synthetase, is important for phosphatidylcholine synthesis (25). By genetic footprinting analysis, *cho1* mutants were at a severe growth disadvantage in rich medium (Q1) and all other selections, except rich medium containing high salt (0.9 M NaCl). Cells containing insertions in the *CHO1* coding sequence were abundantly represented after selection in the high-salt medium, indicating that the *cho1* mutants were not only viable, but grew at least as well as the general population of cells in this medium. This unexpected result has no obvious explanation (26), but it highlights the value of applying

genetic footprinting analysis to all genes, even the apparently well-characterized ones. Similarly, mutants in the novel genes *YER093ca* (Q2) and *YER072w* (Q1) showed general growth defects at 25°C but grew at improved rates at 36.5°C (Fig. 1).

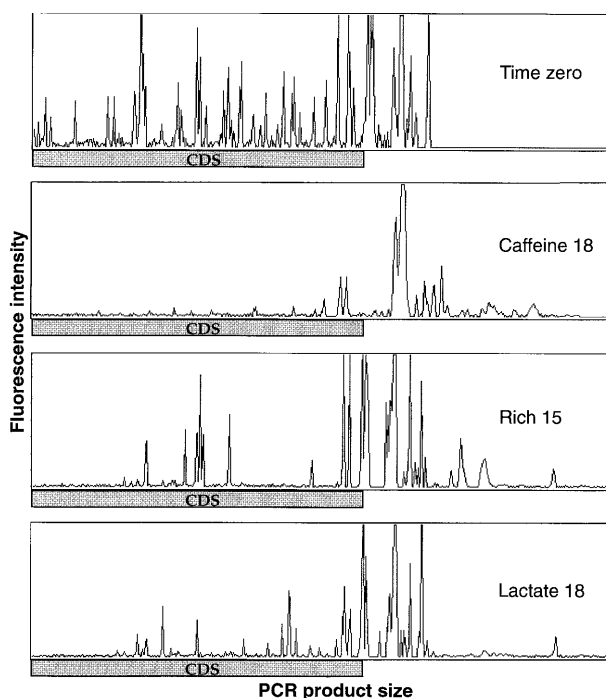
The ability to mate was also used as a selection protocol. Selection for diploids was done after Ty1-mutagenized *MAT α* cells were mated with excess *MAT a* partners. One novel gene on chromosome V, *YER107c*, was required for mating. Mutations in this gene also resulted in a Q3 growth defect in all selections. Conservative interpretation of mating data was necessary in cases of mutants with more severe general growth defects (Q1 and Q2), as these mutants were more susceptible to inadvertent selection during the mating procedure (27). PCR products representing detrimental insertions in mutants with the most severe growth defects (Q1) were frequently (but not always) absent upon analysis of the DNA from the selected diploid products of a successful mating. There were two notable exceptions. Peaks corresponding to insertions in the analyzed coding sequence of *YER132c* were barely detectable after PCR analysis of the time-zero DNA and were depleted in all other selections. However, these peaks were detected upon PCR analysis of DNA isolated from the diploid products of mating, which suggested that these

severely growth-impaired mutants were nonetheless able to mate. Similarly, peaks corresponding to insertions in the coding sequence of *YER072w* (Q1) were detected upon PCR analysis of the successful maters. Mutants with less severe growth defects (Q2 and Q3) were generally able to mate (28).

Genetic footprinting has allowed us to determine that an unexpectedly large fraction of the genes of *S. cerevisiae* make detectable contributions to fitness under standard laboratory conditions. This finding may be attributable in part to the essentially quantitative nature of the data obtained with this method. Unlike standard gene disruption methods, genetic footprinting provides an estimate of the fitness of mutants relative to the population as a whole. The estimates obtained for the fraction of genes that are "essential" is governed by this feature: the Q1 category includes not only essential genes but genes for which mutants have severe growth defects but might still manage a visible colony. The number observed here (20%) is indeed higher than that reported in other studies estimating the proportion of "essential" genes (2). Mutations in another 39% of the nontelomeric genes on chromosome V resulted in general growth defects. These results suggest that despite the apparent redundancy in the yeast genome, more than half of all yeast genes contribute detectably to competitive fitness. Even redundant genes can be important for vegetative growth. For example, on chromosome V, *YER074w* (*RP50A*) and *YER131w* (*RPS26b*), which encode predicted ribosomal proteins, have nearly perfect homologs on other chromosomes. Mutants of *YER074w* and *YER131w* were nonetheless at substantial growth disadvantages (Q1 and Q2) in rich medium (29).

We anticipate that simply extending the methods described above to the whole genome will enable us to assign mutant phenotypes to more than half of the genes. In our work to date, the more specialized selections yielded fewer mutant phenotypes, but the number obtained is of consequence nonetheless. Mutants of 11 novel genes were found to have specific growth defects or growth advantages in specialized selections. Extrapolating this result to the rest of the genome predicts that these specialized selections alone would identify specific mutant phenotypes for another 250 to 300 novel genes, in addition to any new discoveries for previously characterized genes. Incorporation of additional selection protocols should increase the fraction of genes for which a specific role can be inferred. However, even when mutations in a gene produced no discernible phenotype under any of the selections we used, the result was still informative: we learned that the gene un-

Fig. 4. *YER093c*, an example of a novel gene important for growth in the presence of 6 mM caffeine (Q1). *YER093c* mutants are also at a more subtle growth disadvantage in all other selections (Q3). In this example, there was almost complete depletion of all peaks corresponding to insertions in the coding sequence of *YER093c* after growth in caffeine medium (Q1), and substantial but not complete depletion of these peaks in other selections (Q3; data for rich-glucose and rich-lactate medium shown). In the absence of caffeine, cells with Ty1 insertions in the *YER093c* coding region grew at ~85 to 90% of the population rate. The Q3 depletion manifests as a reduced total number of peaks after 18 population doublings, rather than as a systematic decrease in signal of all peaks detected in the time-zero DNA. The individual peaks detected at 18 population doublings vary between different DNAs and also between independent PCRs of the same DNA sample (not shown). This pattern is typical of the data generated by genetic footprinting for many genes that are not favored targets for Ty1 transposition, and reflects sampling of the cell population. As discussed previously (2), the lower intensity peaks corresponding to insertions in the coding sequence of many such genes may represent as few as 1 to 10 cells.



der study was dispensable for growth in each of the media tested, to a resolution of 5 to 10% of the population growth rate (30).

Genetic footprinting provides an efficient and economical way to take advantage of the genomic sequences of microorganisms so as to learn about the functions of the newly identified genes. Thanks to the high degree of functional conservation among eukaryotic proteins, we can expect that the information gathered in this way will also, in many cases, provide important clues to the functions of the cognate human genes.

REFERENCES AND NOTES

- M. G. Goebel and T. D. Petes, *Cell* **46**, 983 (1986); S. G. Oliver *et al.*, *Nature* **357**, 38 (1992).
- V. Smith, D. Botstein, P. O. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6479 (1995).
- F. Dietrich *et al.*, *Nature*, in press. In subsequent releases, additional telomeric sequences were added, making the full length of chromosome V 574,860 bp [updated in the Martins Institute for Protein Sequences (MIPS) Yeast Genome Database]. The number of predicted genes is derived from Dietrich *et al.* and from the MIPS database. Most of the possible additional and alternative open reading frames noted in the MIPS database were also covered by the primers used in this study. One to three independent gene-specific primers were used to analyze each gene. Primers were designed with the use of the program PRIMER (Whitehead Institute for Biomedical Research, Cambridge, MA) with a specified melting temperature of 69° to 73°C. Primers were synthesized by means of a 96-well array synthesizer [D. A. Lashkari, S. P. Hunnicke-Smith, R. Norgren, R. W. Davis, T. Brennan, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7912 (1995)] and labeled at their 5' end with 5-carboxyfluorescein (Applied Biosystems or Pharmacia). Approximately 85% of primers produced usable data. Alternative primers were synthesized for any gene for which the first primer failed to produce usable data. Each labeled gene-specific primer was used in a PCR with an unlabeled Ty1-specific primer, as described (2).
- Although genetic footprinting analysis of chromosome V had a high success rate, there were some failures. For seven genes (*RIP1*, *PMP2*, *SWI4*, *CLK3*, *GDI1*, *BRR2*, and *YER182w*), PCR reactions that used two different gene-specific primers failed to produce enough amplified products to allow meaningful data interpretation. In some cases, products corresponding to insertions upstream of the start ATG codon were readily detected, but products representing insertions in the coding sequence were not reproducibly detected. This may reflect a low frequency of Ty1 insertion in the coding sequences of these genes. It is also possible that Ty1 insertions in essential genes may not be detected, even in the time-zero sample, because of selection during the mutagenesis (10). However, at least some of the genes for which no insertions could be detected are known from previous work not to be essential for vegetative growth, so the latter hypothetical explanation cannot account for all failures resulting from insufficient signal. In almost all cases, when two independent primers for any particular gene both produced Ty1-dependent signal upon PCR analysis, a very similar distribution of PCR products was obtained. In three cases, however, two independent primers for the same gene produced apparently credible data yet gave different results by genetic footprinting analysis (*YEL044w*, *AFG3*, and *RSP5*; Fig. 1). On the basis of the relative frequency of these cases of discordant data produced by two gene-specific primers relative to the frequency of concordant result, we estimate that each analysis using a single primer has a 2 to 3% chance of producing unreliable data.
- Cautions were necessary in interpreting data for genes that had very strong local preferences for Ty1 insertion upstream of the start ATG codon. On chromosome V, these genes were generally located in the immediate vicinity of tRNA genes. Preferred upstream sites for Ty1 transposition frequently generated a cluster of PCR products with 100 to 1000 times the signal typically observed. This raises the possibility of exhaustion of PCR reagents at a stage earlier than 30 cycles, because of the increased number of initial template molecules. It is thus possible that some of the less abundant, smaller DNA products observed in these cases, which would ordinarily be assumed to represent insertions in the coding sequence of the gene, could be artifactual. In cases of this kind, PCR was repeated with only 23 cycles in an attempt to minimize this possible artifact. In addition, new primers that did not encompass the preferred site were also used; typically, these primers were located in the first 100 bp of the gene and were directed downstream, toward the stop codon. These "reverse" primers typically allowed a more reliable survey of products corresponding to insertions in coding sequences of genes with very strong upstream site preferences for Ty1 insertion.
- Typically, after mutagenesis, aliquots of 2×10^8 cells were stored in 25% glycerol at -80°C . After one mutagenesis, cells were immediately transferred to rich medium for growth, without intervening storage. The DNA isolated from these cells was analyzed with primers specific to every gene for which a quantitative growth defect was detected. No cases were found in which storage in glycerol, freezing and re-suscitation, or both accounted for the apparent growth defect, although possible contributions of these procedures to some phenotypes are noted in Fig. 1.
- M. C. Brandriss, *J. Bacteriol.* **138**, 816 (1979).
- K. Irie, K. Yamaguchi, K. Kawase, K. Matsumoto, *Mol. Cell. Biol.* **14**, 3150 (1994).
- L. A. Graham, K. J. Hill, T. H. Stevens, *J. Biol. Chem.* **270**, 15037 (1995); C. Abejón *et al.*, *J. Cell Biol.* **122**, 307 (1993); S. Prakash and L. Prakash, *Genetics* **87**, 229 (1977).
- Selection against the mutants during the mutagenesis procedure itself may account for our failure to detect insertions in most of the coding sequence of *GDI1*, an essential gene [M. D. Garrett, J. E. Zahner, C. M. Cheney, P. L. Novick, *EMBO J.* **13**, 1718 (1994)]. Factors that may affect the degree to which mutants are lost during the period of mutagenesis, and thus our ability to detect insertions in an essential gene, include the integrity of the mutant cell (rapid cell lysis would limit recovery of the cell's DNA) and the stability and turnover of residual RNA and protein produced before Ty1 insertion.
- M. Werner-Washburne, D. E. Stone, E. A. Craig, *Mol. Cell. Biol.* **7**, 2568 (1987).
- A. Shinohara, H. Ogawa, T. Ogawa, *Cell* **69**, 457 (1992); C. M. Moehle, M. W. Aynardi, M. R. Kolodny, F. J. Park, E. W. Jones, *Genetics* **115**, 255 (1987); A. P. Mitchell and B. Magasanik, *Mol. Cell. Biol.* **4**, 2758 (1984).
- These observations presumably reflect the fact that the Ty1 element could supply promoter or terminator functions and could even provide an ATG start codon [J. D. Boeke, D. J. Garfinkel, C. A. Styles, G. R. Fink, *Cell* **40**, 491 (1985)].
- For example, for a gene in which mutations reduce the growth rate to 90% of the population rate, 29% of the signal corresponding to insertions affecting gene function would remain after 18 population doublings, but only 3% would remain after 50 population doublings.
- There are 1628 Sau 3A sites in the 569,202-bp sequence of chromosome V investigated in this analysis. Thus, a Sau 3A site occurs, on average, every 350 bp. In this analysis, for 600 to 900 bp of coding sequence, two or three peaks resulting from Sau 3A sites are expected, assuming a random distribution.
- We analyzed 63 genes for which quantitative mutant growth deficits were detected in rich medium (42%) independently by library DNA normalization and normalization to upstream peaks; the remainder were analyzed only by normalization to upstream peaks. Variations in intensities of library peaks were observed. In some cases, library peaks could not be identified because of the density of Ty1-specific peaks. The library DNA was most useful as a normalization standard for genes that contained Sau 3A sites within the region undergoing depletion. Most of the genes analyzed (94%) satisfied this criterion.
- J. B. McNeil *et al.*, *J. Biol. Chem.* **269**, 9155 (1994); G. Nass and K. Poralla, *Mol. Gen. Genet.* **147**, 39 (1976); M. Masselot and H. de Robichon-Szulmajster, *ibid.* **139**, 121 (1975); C. Boonchird, F. Messenguy, E. Dubois, *ibid.* **226**, 154 (1991).
- YEL044w* may be important for growth under all selections (Q1). The two primers used to analyze this gene gave conflicting data.
- A. P. Mitchell and B. Magasanik, *Mol. Cell. Biol.* **4**, 2767 (1984).
- R. E. Chapman and S. Munro, *EMBO J.* **13**, 4896 (1994).
- J. E. McEwen, K. H. Hong, S. Park, G. T. Preciado, *Curr. Genet.* **23**, 9 (1993); J. D. Ohmen, B. Kloeckener-Gruissem, J. E. McEwen, *Nucleic Acids Res.* **16**, 10783 (1988); A. Harrington, C. J. Herbert, B. Tung, G. S. Getz, P. P. Slonimski, *Mol. Microbiol.* **9**, 545 (1993); N. Bonnefoy, F. Chalvet, P. Hamel, P. P. Slonimski, G. Dujardin, *J. Mol. Biol.* **239**, 201 (1994).
- E. Guelin, M. Rep, L. A. Grivell, *Yeast* **10**, 1389 (1994).
- GenBank accession numbers YSCCOX15A and L38643.
- S. Thiagalingam, K. W. Kinzler, B. Vogelstein, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6062 (1995); M. Nakafuku *et al.*, *ibid.* **85**, 1374 (1988).
- K. D. Atkinson *et al.*, *J. Bacteriol.* **141**, 558 (1980).
- Perhaps an alternative synthesis or salvage pathway is activated when cells are exposed to 0.9 M NaCl, possibly as part of a general process in which cell wall structure is altered to enhance viability in the high-osmolarity medium.
- Approximately four to five population doublings elapsed during the mixing and plating of cells to allow mating before selection for diploid (mated) cells.
- Depending on the severity of the growth defect, the number of peaks corresponding to detrimental insertions detected in the PCR analysis of DNA from successful maters typically ranged between the number observed in the time-zero analysis and the number observed at 18 population doublings.
- Explanations for this type of phenomenon have been reported previously. For example, *RNR1* and *RNR3* are closely related in sequence, and both encode large subunits of ribonucleotide reductase. However, *RNR1* is constitutively expressed and is essential for vegetative growth, whereas *RNR3* is induced in response to exposure to DNA-damaging agents [S. J. Elledge and R. W. Davis, *Genes Dev.* **4**, 4740 (1990)].
- After 60 population doublings, mutants with a 5% growth rate disadvantage will have doubled only 57 times, and thus will be underrepresented by a factor of $2^3 = 8$. Thus, by examining cells grown for 60 population doublings in rich-glucose medium, we could readily detect a 5% deficit in growth rate in this medium. Because selections for growth in other media involved fewer population doublings (typically 15 to 18), we may only have been able to recognize growth deficits of $\geq 10\%$ in those selections (15).
- Y.-J. Kim, L. Francisco, G.-C. Chen, E. Marcotte, C. S. M. Chan, *J. Cell Biol.* **127**, 1381 (1994).
- We thank R. Norgren for assistance with data analysis and the generation of Fig. 1, and L. McAllister and K. Davis for discussions. Supported by NIH grant 1P01 HG00205-05 (to R. Davis, P.O.B., and D.B.) and by the Howard Hughes Medical Institute (P.O.B.). Supported in part by grant LT-141/93 from the Human Frontier Science Program Organization to V.S. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

31 July 1996; accepted 7 November 1996