Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 6479–6483, July 1995
Genetics

Genetic footprinting: A genomic strategy for determining a gene's function given its sequence

(*Saccharomyces cerevisiae*/transposition/Ty*1*/selection/polymerase chain reaction)

VICTORIA SMITH*, DAVID BOTSTEIN*, AND PATRICK O. BROWN†‡

Departments of *Genetics and †Biochemistry, Stanford University, and ‡Howard Hughes Medical Institute, Stanford, CA 94305

Contributed by David Botstein, April 10, 1995

ABSTRACT This report describes an efficient strategy for determining the functions of sequenced genes in microorganisms. A large population of cells is subjected to insertional mutagenesis. The mutagenized population is then divided into representative samples, each of which is subjected to a different selection. DNA is prepared from each sample population after the selection. The polymerase chain reaction is then used to determine retrospectively whether insertions into a particular sequence affected the outcome of any selection. The method is efficient because the insertional mutagenesis and each selection need only to be performed once to enable the functions of thousands of genes to be investigated, rather than once for each gene. We tested this "genetic footprinting" strategy using the model organism *Saccharomyces cerevisiae.*

The *Saccharomyces cerevisiae* genome sequencing effort has produced seven complete chromosome sequences (GenBank release 86.0, December 1994). As a result, hundreds of putative genes of unknown function have been identified, and it is expected that thousands more will be identified as the complete sequence of the yeast genome is determined. Powerful techniques are available for determining the biological functions of yeast genes. These generally involve a gene disruption strategy, in which a mutation is created *in vitro* and introduced into the genome. The mutant strain's fitness is then tested under various physiological conditions. While this approach is highly effective for analyzing individual genes, applying it to thousands of genes would be a mammoth task.

We present an efficient experimental approach, designed to allow the biological roles of the thousands of genes in the *S. cerevisiae* genome to be studied economically. Insertional mutagenesis and selections are performed *en masse* in a large population of cells, in a manner that allows the effects of the mutations in any DNA sequence under any particular selection to be determined retrospectively using the polymerase chain reaction (PCR). Specifically, transposition of a marked Ty*1* transposable element is induced in a large population of cells, generating Ty*1* insertional mutations at diverse sites. The mutagenized population is then divided into representative samples, each of which is subjected to one of a large set of selections or fractionations. DNA is prepared from the selected cells. The recovery of cells carrying Ty*1* insertions at a particular site, following a particular selection, can be determined retrospectively by using an aliquot of DNA from the selected cells as the template for PCR amplification (1). A primer specific to the sequence under investigation and a second primer specific to the Ty*1* element are used, such that exponentially amplified products represent cells in which the sequence of interest is disrupted by a Ty*1* insertional mutation. A role for a particular sequence under a particular set of selective conditions is inferred from depletion of the corresponding PCR product bands: the "genetic footprint" (Fig. 1).

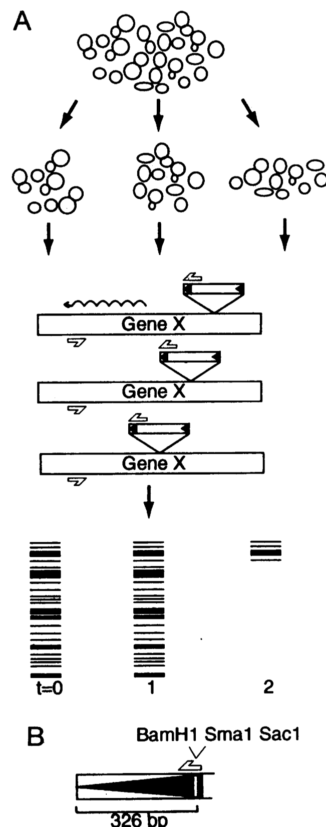

FIG. 1. Genetic footprinting strategy. (*A*) In this schematic outline, the sequence under investigation, "Gene X," was assumed to be dispensable for growth under condition 1 but important for cell growth under condition 2, as demonstrated by depletion of bands in lane 2. The depleted bands represent Ty*1* insertions in Gene X, which are absent as a result of loss of the cells bearing these insertions from the pool 2 cell population. The curved arrow represents the direction of transcription of Gene X. (*B*) Schematic representation of one of the δ elements of the Ty*1* transposon used in this study. The Ty*1*-specific oligonucleotide used as a primer for PCR corresponds to a sequence located 326 bp into the δ element, which includes a unique 16-bp polylinker sequence to distinguish it from endogenous elements. All PCR products corresponding to Ty*1* insertions are therefore >326 bp in size. For this reason, only the portion of the scan corresponding to products >320 bp (scanlines 1400 and up) is shown in Figs. 2–4.

This depletion reflects the selective depletion of cells bearing Ty*1* insertions at that sequence.

The Ty*1* retrotransposon can insert at diverse sites in the nuclear genome, although certain sites are strongly preferred (2). Analysis of Ty*1* transposition into the *CAN1*, *LYS2*, and *URA3* loci indicates a strong preference for insertion into noncoding regions, especially 5′ of the gene, but Ty insertion into these coding sequences has also been observed (3–5). While the endogenous Ty*1* elements, present in most *S.*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

cerevisiae isolates, transpose infrequently, this rate can be increased 20- to 100-fold by overexpression of Ty*1* to yield five or more new insertions per genome after several generations (6). We used a Ty*1* element regulated by the *GAL1* promoter on a 2-μm plasmid (7) to test the feasibility of the genetic footprinting strategy on a set of 14 previously characterized genes.

MATERIALS AND METHODS

Induction of Ty*1* Transposition. The haploid yeast strain GRF167 [*mat*α, *ura3-52, his3-*Δ*200* (7)] containing the 2-μm plasmid PBTy1 (Ty*1* under the *GAL1* promoter, *URA3*) was grown to 1×10^7 cells per ml in 100 ml of synthetic complete (SC) medium lacking uracil [0.67% yeast nitrogen base (YNB)/0.5% Casamino acids/0.0026% adenine/0.008% tryptophan/2% glucose]. The cells were washed once with water and transferred to 800 ml of SC minus uracil, supplemented with 2% galactose, and incubated at 24°C. Cells (8×10^8) were transferred to fresh galactose medium each day over a course of 4 days. On the fourth day, DNA was prepared from 4×10^{10} cells (8), designated the time zero DNA sample. The remaining 1.3×10^{10} cells were stored in glycerol at −70°C in aliquots of 2×10^8 cells.

Selection. Cells (2×10^8) that had been subjected to Ty*1* insertional mutagenesis were cultivated for 15 population doublings in 100 ml of minimal medium (0.67% YNB/0.0022% uracil/0.0046% histidine/2% glucose) or for 5, 15, and 60 population doublings in 100 ml of rich medium (1% yeast extract/2% bactopeptone/0.008% tryptophan/2% glucose) at 24°C. DNA was prepared from each cell population (8).

PCR Analysis. PCR analysis to detect Ty*1* insertions at a particular sequence was performed on 1 μg of DNA using a gene-specific oligonucleotide labeled 5′ with fluorescein and an unlabeled Ty*1*-specific oligonucleotide. The reaction mixture also contained 10 mM Tris (pH 8.5), 50 mM KCl, 1.5 mM $MgCl_2$, 0.5 μM each oligonucleotide, 250 μM each dNTP, and 2 units of *Taq* DNA polymerase (Perkin–Elmer) in a 50-μl reaction mixture. The conditions were as follows: 93°C for 1 min, followed by 10 cycles of 92°C for 30 sec, 67°C for 45 sec, and 72°C for 2 min; then 20 cycles of 92°C for 30 sec, 62°C for 45 sec, and 72°C for 2 min. Five microliters of each reaction mixture was denatured by the addition of formamide and size fractionated on a 4.75% denaturing polyacrylamide gel. Fragments were detected using an Applied Biosystems model 373A automated DNA sequencer equipped with Genescan DNA fragment analysis software. The approximate sizes of DNA fragments were determined using the ROX-labeled size standards Genescan-1000 and Genescan-2500 (Applied Biosystems).

Oligonucleotides. Oligonucleotides were designed using the programs OLIGO 4.0 (National Biosciences, Plymouth, MN) and PRIMER (Whitehead Institute for Biomedical Research, Cambridge, MA) and have a calculated melting temperature of 72°C. They were labeled at the 5′ end with 5(6)-carboxyfluorescein and were obtained from Operon Technologies (Alameda, CA). The following primer sequences were used to generate the data shown (5′ → 3′): *TRP1*, GTAAAAGTCAACCCCCTGCGATGT; *MSH1*, ATTGGCTTCATTATCTGCCACGTC; *CLN2*, AAGACCTGACCATCACCACAGTAATG; *ILV2*, GGCATCTGCCATTGGAGTAACGAC; *SNP1*, CTGCCTCTTTCTATGTCGACTATGCA; *LEU2*, TCTTTGCACTTCTGGAACGGTGTA; *ADE2.1*, GTACAAAGGACGATCCTTCAGTACTTC; *ADE2.2*, ACAACCGGGAAAAGATTTGATTGC; *LYS2*, TGGATGATGCGGTGATCAGAGAG; *PHO5*, CTGCAAGGTGATGTTGAATTGGTC; *CAN1*, ACTTTGATGGAAGCGACCCAGAAC; *RNR3*, TGAGGACGTTCTGCCACTTCAC; *CLN1*, ATGAACCAATTCAGAAAGGCGTGG; PBTY1R1 (Ty*1*-specific primer), AGAGCTCCCGGGATCCTCTACTAAC.

RESULTS

Analysis at *ADE2*. A representative analysis, at the *ADE2* locus, is shown in Fig. 2. The fluorescently labeled PCR products were detected using an automated DNA sequencer and displayed using the program GENESCAN (Applied Biosystems). Each PCR product is represented as a peak, whose area is proportional to the amount of product detected. PCR analysis of the DNA sample designated "time zero" (Fig. 2*B*), which was isolated immediately after induction of transposition ceased, yielded numerous peaks of varying intensities. Repetition of the PCR using a second *ADE2*-specific labeled oligonucleotide, which primes at a site 207 bp upstream of the first oligonucleotide (Fig. 2*A*), confirmed that these peaks resulted from Ty*1* insertions at *ADE2*. The second primer generated the same pattern of PCR products, shifted by 207 bp. PCR analysis using the same primers on DNA isolated from cells grown in complete medium but not induced for Ty*1* transposition gave no product peaks (not shown). PCR analysis of a DNA sample taken from cells cultivated for 15 population doublings in minimal medium after Ty*1* mutagenesis yielded peaks representing insertions upstream of the start codon, but all peaks representing Ty*1* insertions in the coding region were absent (Fig. 2*C*). This result implies that disruption of *ADE2* by Ty*1* insertions impaired cell growth in minimal medium. PCR products representing Ty*1* insertions in the coding region were still visible on analysis of DNA isolated from cells after 15 population doublings in rich medium (Fig. 2*E*), but they were reduced in number and intensity compared to the time zero sample. PCR analysis of DNA isolated from cells after 60 population doublings in rich medium produced a footprint very similar to that observed in minimal medium (Fig. 2*F*), indicating that disruption of *ADE2* was detrimental to growth even in rich medium. Supplementing the medium with extra adenine to 20 mg/liter did not alleviate this effect. Analysis of DNA isolated from cells after only 5 population doublings in rich medium following Ty*1* mutagenesis produced a pattern more similar to the time-zero pattern (Fig. 2*D*).

Analysis at 12 Loci. Results for 12 of the 14 test genes are displayed in Fig. 3 in vertical format. The analyses of three DNA samples are shown for each gene: the time zero sample (0) and samples isolated from cells after 15 population doublings in minimal (M) or rich (R) medium, following Ty*1* mutagenesis. The same yeast strain, grown in complete medium but not induced for Ty*1* expression, provided control DNA for each PCR. None of the bands shown in Fig. 3 was generated in the control reactions (not shown). At least two different gene-specific oligonucleotides corroborated the results for each gene shown in Fig. 3. For several genes, notably those giving rise to some faint PCR product bands (e.g., *SNP1, ADE2, TRP1*), an aliquot of the PCR product was digested with *Bam*HI or *Sma* I—enzymes that cleave in the Ty*1* priming site (Fig. 1*B*). All of the PCR product bands shifted as expected upon digestion, indicating that these bands were not products of premature termination during the PCR. For the remaining 2 of the 14 test genes, *ARO1* and *ARO2*, the PCR data obtained were too low in signal and too high in Ty-independent background to be meaningfully interpreted (not shown).

Numerous Ty*1* insertions were detected upstream to and within the coding region of most of the test genes. Most had complex Ty*1* insertion patterns that were well conserved among all DNA samples for which no selective loss was expected. In some cases, such as *CAN1* and *CLN1*, PCR analysis produced results that were variable between DNA samples. For example, different low molecular weight bands (at 400–700 bp) were detected upon PCR analysis at *CAN1* when the three DNA samples were compared (lanes 0, M, and R in Fig. 3). PCRs using different aliquots of the same DNA sample produced similar variation (data not shown). This probably reflected a low frequency of Ty*1* insertion at these

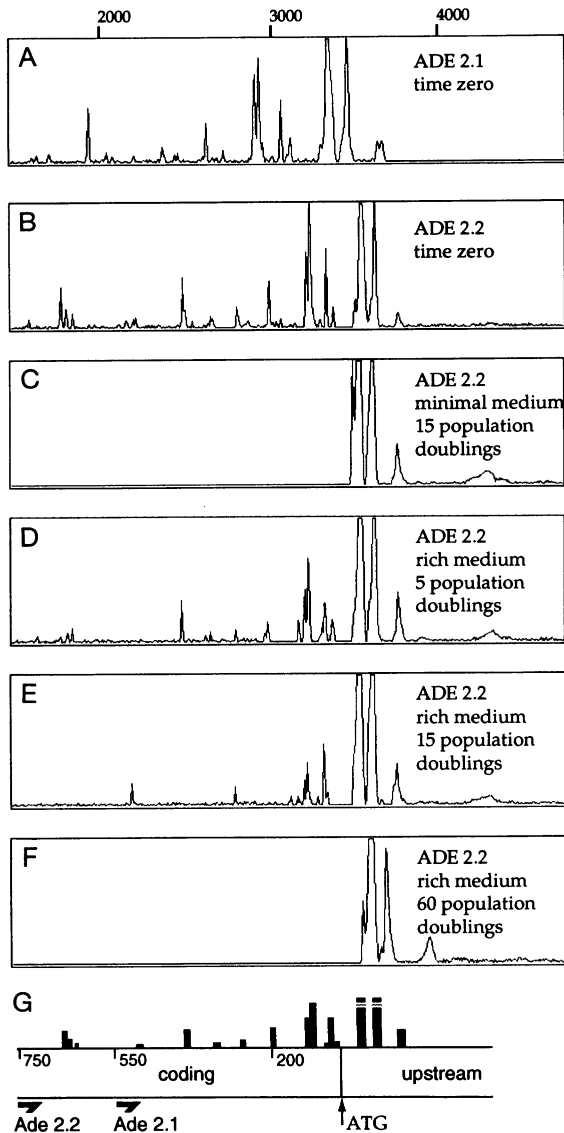

FIG. 2. Genetic footprinting at the *ADE2* locus. PCR analysis of Ty*1* insertions at the *ADE2* locus. These data were obtained using the program GENESCAN (Applied Biosystems). Fluorescence intensity (*y* axis) is plotted against scanline number (*x* axis). The scanline number refers to the time at which each DNA molecule was detected during electrophoresis. Thus, scanline number is related to molecular weight. As the resolution of the gel decreases with increasing molecular weight, peaks detected after scanline 3000 may represent two or more PCR products that are within 20 bp in size. The most intense peaks (between scanlines 3300 and 3650) are shown off scale in this plot, to allow clearer visualization of the lower-intensity peaks. (*A* and *B*) PCR analysis of time zero DNA, using the oligonucleotide Ade2.1 (*A*) and 2.2 (*B*). (*C–F*) PCR analysis using Ade2.2 on DNA isolated from cells after 15 population doublings in minimal medium (*C*) or after 5 (*D*), 15 (*E*), and 60 (*F*) population doublings in rich medium, following Ty*1* insertional mutagenesis. The analysis shown in *F* was electrophoresed on a different gel to the analyses in *A–E* and thus the relative positions of peaks are slightly altered. (*G*) Locations of Ade2 oligonucleotides and Ty*1* insertions shown schematically against the *ADE2* gene. The oligonucleotide Ade2.1 primes 207 bp upstream from Ade2.2. Ty*1* insertions are represented by vertical bars. The area of the bars represents peak area. The two largest peaks are off scale in this figure, as they are 5- to 7-fold more intense than the next largest peak. The numbers represent nucleotides in the *ADE2* gene, where the A of the start ATG is nucleotide number 1.

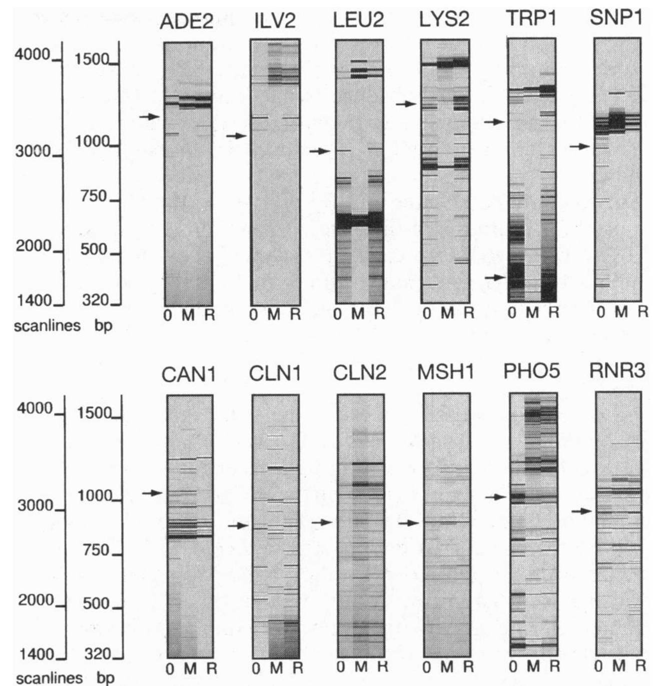

FIG. 3. Genetic footprint of 12 genes. The Ty insertion pattern of each gene from DNA isolated from time zero cells (0) and cells grown for 15 generations in minimal (M) or rich (R) medium is displayed. PCR products were size fractionated on denaturing polyacrylamide gels and detected using an Applied Biosystems Model 373A automated DNA sequencer equipped with GENESCAN software. The leftmost vertical scale shows scanline number and approximate size in bp is indicated on the adjacent scale. The location of the start ATG for each gene is indicated by an arrow. Bands of lower molecular weight (below the arrow) correspond to Ty*1* insertions within the coding sequence, while bands above the arrow represent Ty*1* insertions upstream of the start ATG. The second (lower) arrow on *TRP1* indicates the location of the stop codon. The gene-specific oligonucleotides used as primers for PCR correspond to sequences 500–900 bp downstream from the initiation codon of the genes (as in Fig. 2*G*). This representation of the data was generated by plotting all data between scanlines 1400 and 4200, and creating an image from a single vertical column of pixels from selected lanes, using the numeric computation and visualization software package MATLAB (The Math Works, Natick, MA). This image amounts to a vertical representation of the trace file, where peak intensity is mapped linearly to a gray scale (darker = more intense).

sites. Estimates of the approximate number of Ty*1* insertions represented by each PCR product were made for *CAN1* by observing the relative number of canavanine-resistant colonies (*can1* mutants) generated by Ty*1* induction. Colony PCR was used to estimate the fraction of *can*-resistant colonies that were detectable using the *CAN1*-specific primer. We estimate that, for *CAN1*, the PCR product bands in the 300- to 1500-bp size range represented an aggregate of 550–650 cells bearing detectable Ty*1* insertions at *CAN1* in a pool of 7×10^7 cells (≈ 1 μg DNA). The strongest bands therefore probably represented >100 cells bearing Ty*1* insertions at *CAN1*, while weaker bands probably represented only 1–10 cells.

The variation in peak intensities between different PCR products in each sample probably reflected the variable abundance of cells bearing Ty*1* insertions at those locations. For example, cells bearing Ty*1* insertions in *ADE2* were less fit in rich medium, and the reduced signal of the PCR product bands corresponding to Ty*1* insertions in the *ADE2* coding sequence reflects the depletion of these *ade2* mutants from the population. However, variation in peak intensity was also apparent in unselected populations. In the time zero PCR analysis, the peaks corresponding to Ty*1* insertions upstream of the *ADE2* start codon were 5- to 10-fold more intense than peaks representing insertions in the coding sequence. This trend was also apparent in many of the other test genes and probably reflected a general preference for Ty*1* insertion upstream of

coding sequences. It is unlikely to be an artifact of the PCR, as the most intense peaks were typically larger PCR products, whereas amplification biases should favor smaller PCR products. However, in the absence of an internal standard to control for variations in signal that may reflect amplification biases, we cannot precisely correlate peak intensity with cell number.

Auxotrophic Test Genes. Fig. 3 *Upper* shows data for 5 genes required for synthesis of different amino acids (*ADE2, ILV2, LEU2, LYS2, TRP1*). In each case, almost every band corresponding to a Ty*1* insertion within or immediately adjacent to the coding region was depleted in DNA samples isolated from cells after 15 population doublings in minimal medium.

There were some exceptions. In *LYS2*, a band near the 5′ end of the coding sequence was still visible after growth in minimal medium, possibly representing a Ty*1* insertion that did not completely abolish the gene's function. A few low molecular weight bands representing Ty*1* insertions close to the *TRP1* stop codon (within 150 bp) also survived selection in minimal medium. One very intense cluster of bands, mapping ≈260 bp from the start codon, was detected with all *LEU2*-specific primers. Unlike the other bands representing Ty*1* insertions well within the coding region, this cluster was not depleted during growth in minimal medium. Although this band cluster was not produced by any control PCR, its anomalous intensity—it was severalfold more intense than any other bands observed in these experiments—together with its paradoxical indifference to growth in minimal medium suggest that it may be an artifact.

Analysis at the *ILV2* locus did not reproducibly yield detectable PCR products corresponding to Ty*1* insertions within the 707 bp of coding region that we surveyed. However, a cluster of four bands representing insertions immediately 5′ to the *ILV2* initiation codon was depleted after growth in minimal medium, whereas the bands representing insertions further upstream were unaffected.

An Essential Gene. This analysis used a haploid yeast strain. The question therefore arises whether Ty*1* insertions can be detected in a gene essential for vegetative growth. In a simple case, we expect Ty*1* insertions in an essential gene to be underrepresented by $2/N$ (for large N), where N is the number of population doublings during the period of Ty*1* induction. *SNP1*, which encodes the yeast homologue of the human 70-kDa U1 small nuclear ribonucleoprotein, is essential for cell growth on glucose medium: *snp1* spores are able to germinate but cannot divide beyond the 20 to 35 cell stage (9). Ty*1* insertions at 20 sites in a 765-bp region of the coding sequence were reproducibly detected in a time zero DNA sample. Nineteen of these bands were undetectable after five or more population doublings in rich or minimal medium.

Nonauxotrophic Test Genes. Fig. 3 *Lower* shows data for six genes whose function was not required for growth in either rich or minimal medium. For each of these genes, there was a sufficient density and distribution of Ty*1* insertions to indicate that neither *CAN1* (arginine permease), *CLN1* and *CLN2* (G_1 cyclins with overlapping functions), *PHO5* (acid phosphatase), *RNR3* (a nonessential subunit of ribonucleotide reductase), nor *MSH1* (*Mut S* homologue) was essential for 15 population doublings in rich or minimal medium. This was the expected result for five of these six genes (10–14).

Analysis at *MSH1*. Cells with mutations in the sixth of these genes, *MSH1*, develop a petite phenotype: they are respiratory-deficient and are thus unable to grow on a nonfermentable carbon source (14). Depletion of bands in the coding region might therefore be expected upon extended growth in either rich or minimal medium. We did not observe such a depletion after 15 population doublings. Indeed, cells carrying Ty*1* insertions in the coding region of *MSH1* were recovered even after 10 population doublings in medium containing a nonfermentable carbon source (not shown). However, after 60 population doublings in rich medium, we observed substantial depletion of PCR products corresponding to insertions in or adjacent to the coding region (Fig. 4*A*). Significant depletion of these bands was also observed after 18 population doublings in medium containing a nonfermentable carbon source (lactate, not shown). *MSH1* is a homolog of the bacterial *muts* gene, which encodes an enzyme essential for DNA replication fidelity. The phenotype therefore presumably reflects the accumulation of lethal mutations in the mitochondrial genome. As *msh1* spores are identical to wild type in their ability to germinate and grow to hundreds to thousands of cells (14), it is likely that 15 population doublings were insufficient for *msh1* cells to accumulate sufficient mutations in the mitochondrial genome to result in a significant growth disadvantage.

Selection in Rich Medium After 60 Population Doublings. For most of the tested genes, PCR analysis of DNA isolated from cells after 60 population doublings in rich medium produced a pattern of bands that was very similar to the pattern seen after 15 population doublings, as shown in Fig. 4*B* for *CLN2*. However, like *MSH1*, *ADE2* (Fig. 2*F*) and *TRP1* (not shown) showed substantial depletion of bands corresponding to insertions in the coding region. In the case of *ADE2*, these data reflected the growth disadvantage of *ade2* mutants in rich medium, which probably resulted from the production of a toxic intermediate in *ade2 ADE3* cells. The result seen with *TRP1* may similarly reflect a growth disadvantage even in tryptophan-containing medium, or it may be a consequence of the lability of the tryptophan in the growth medium.

DISCUSSION

Genetic footprinting is an efficient and economical strategy for functional analysis of DNA sequences on a genomic scale. The Ty*1* transposable element was suitable as an insertional mu-

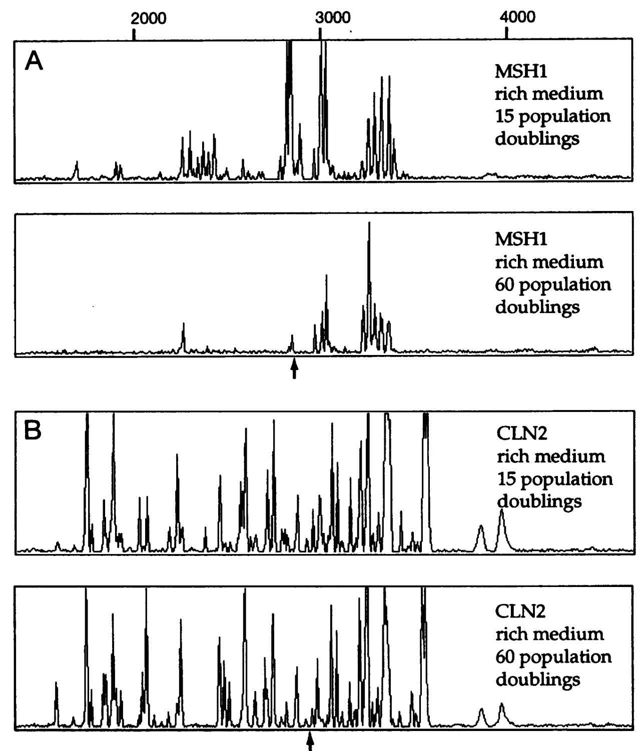

FIG. 4. PCR analysis of DNA isolated from cells after 15 and 60 population doublings in rich medium following Ty*1* mutagenesis. (*A*) *MSH1*. (*B*) *CLN2*. Fluorescence intensity (y axis) is plotted against scanline (x axis). The arrows represent the locations of the start ATGs. The most intense peaks are shown off scale to allow clearer visualization of the lower-intensity peaks.

tagen for this analysis in *S. cerevisiae.* As previous work has suggested, Ty*1* transposition was sufficiently random to generate a range of useful insertion mutations within, upstream, and downstream of coding sequences. However, we did not observe an even distribution of Ty*1* insertions at each nucleotide but, instead, detected strong site preferences for Ty*1* insertion that were conserved for each test gene. Thus, the distinctive pattern of insertions that was observed for each of the test genes shown in Fig. 3 was reliably reproduced in different PCRs on independent DNA samples. Consistent data were also obtained from cells mutagenized in independent Ty*1* induction experiments.

A DNA sample of 1 μg, which corresponds to $\approx 7 \times 10^7$ haploid yeast genomes, gave a sufficient representation of Ty*1* insertions at 11 of the 12 test genes presented, such that interpretation of the functional importance of that region was possible. The genetic footprint at the *ILV2* locus matched the result expected for this gene, but had its function not been known, inferences would have been made with caution, due to the low number of Ty*1* insertions detected at this locus and their absence from the coding region. While occasional nonlethal Ty*1* insertions or PCR artifacts may generate products that do not behave as expected for a simple disruption of a required function, we easily obtained data sufficient to allow accurate conclusions to be drawn regarding the functions of 11 of the 14 genes we surveyed. For the essential *SNP1* gene, and for most of the auxotrophic genes analyzed after 15 population doublings in minimal medium, a clear genetic footprint was observed—at least 90% of all discrete product bands corresponding to insertions in the coding region were completely depleted, while most peaks corresponding to insertions further upstream were retained in all three DNA samples. Active transcription of a gene did not appear to be a strict requirement for that gene to serve as a target for Ty*1* transposition. We observed a rich pattern of Ty*1* insertions in *RNR3*, a gene that is not constitutively expressed but is induced in response to DNA damage (13). However, we have not yet investigated other genes that may be more tightly transcriptionally inactive during normal vegetative growth.

This experimental approach to the analysis of gene function was sensitive to quantitative differences in growth rate. For example, the competitive disadvantage of *ade2* mutants in rich medium was clearly reflected by the reduced relative intensity of peaks representing insertions in the coding region after 15 population doublings. A sufficient number of population doublings under selective conditions, or cycles of selection, could expose even a subtle disadvantage for a particular mutant in the heterogeneous cell population. For example, strains with mutations in either of the two yeast genes encoding hydroxymethylglutaryl-CoA reductase grow as wild type on rich or minimal medium plates, but their distinct growth disadvantages are evident when they are cocultivated with the wild-type parent strain in liquid medium (15). It may therefore be possible to recognize and distinguish between genes that superficially appear to be functionally redundant.

These initial experiments used only the simplest of selections. Practical selections can be designed, however, to identify genes that are important for diverse biochemical, developmental, regulatory, and signaling pathways as well as ability to adapt to environmental alterations. For example, large populations of cells can be selected for tolerance to doses of x-ray or UV radiation, ability to mate and sporulate, or ability to progress through specific stages of the cell cycle. Selections that would be too difficult, expensive, or laborious to apply to testing the fitness of thousands of individual mutants, one at a time, could nevertheless be used to investigate thousands of genes, since each selection would only need to be applied once to a diverse population of mutant cells. Our current efforts are directed at providing timely preliminary information about the functions of putative genes identified by sequencing the *S. cerevisiae* genome. This genomic strategy cannot substitute for a comprehensive investigation of individual genes and pathways. The information it can provide regarding the function of a new DNA sequence should nonetheless be useful in directing the attention of biologists interested in those particular functions in yeast or in homologous sequences or processes in other organisms.

The genetic footprinting strategy should be applicable to diverse microorganisms for which genomic sequencing is practical and a convenient method is available for insertional mutagenesis. To allow analysis of recessive insertional mutations, the organism must be able to be cultured as a haploid or conveniently homozygosed. The organism must be sufficiently small that a large number of independent insertional mutants can be obtained and subjected to a range of selections *en masse.* Because each selection would only need to be performed once to provide the DNA required to analyze a large number of genes, the genetic footprinting approach may be particularly useful in studying microorganisms that are difficult or dangerous to cultivate.

We thank Vit Lauermann and Jef Boeke for providing the Ty*1* element marked with a polylinker, prior to publication, and Jef Boeke for many helpful discussions and suggestions. We thank Sue Klapholz for advice and assistance with initial experiments and many helpful discussions, Richard Norgren for data analysis using MATLAB, Karen Chou for technical assistance, and Deval Lashkari for help with oligonucleotide labeling and analysis. We thank Ron Davis for supporting this project in the Stanford Center for DNA Sequencing and Technology. V.S. is supported by Human Frontiers Science Program Organization Fellowship LT-141/93. This work was also supported by National Institutes of Health Grant 1P01 HG00205-05 to Ron Davis, P.O.B., and D.B., and by the Howard Hughes Medical Institute. P.O.B. is an Assistant Investigator of the Howard Hughes Medical Institute.

1. Pryciak, P. M. & Varmus, H. E. (1992) *Cell* **69,** 769–780.
2. Ji, H., Moore, D. P., Blomberg, M. A., Braiterman, L. T., Voytas, D. F., Natsoulis, G. & Boeke, J. D. (1993) *Cell* **73,** 1007–1018.
3. Wilke, C. M., Heidler, S. H., Brown, N. & Liebmann, S. W. (1989) *Genetics* **123,** 655–665.
4. Natsoulis, G., Thomas, W., Roughmann, M. C., Winston, F. & Boeke, J. D. (1989) *Genetics* **123,** 269–279.
5. Eibel, H. & Philippsen, P. (1984) *Nature (London)* **307,** 386–388.
6. Boeke, J. D., Eichinger, D. J. & Natsoulis, G. (1991) *Genetics* **129,** 1043–1052.
7. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. (1985) *Cell* **40,** 491–500.
8. Struhl, K., Stinchcomb, D. T., Scherer, S. & Davis, R. W. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1035–1039.
9. Smith, V. & Barrell, B. G. (1991) *EMBO J.* **10,** 2627–2634.
10. Wheelan, W. L., Gocke, E. & Manney, T. R. (1979) *Genetics* **91,** 35–51.
11. Hadwiger, J. A., Wittenberg, C., Richardson, H. E., Lopes, M. de B. & Reed, S. I. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6255–6259.
12. Rogers, D. T., Lemire, J. M. & Bostian, K. A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2157–2161.
13. Elledge, S. J. & Davis, R. W. (1990) *Genes Dev.* **4,** 740–751.
14. Reenan, R. A. G. & Kolodner, R. D. (1992) *Genetics* **132,** 975–985.
15. Basson, M. E., Moore, R. L., O'Rear, J. & Rine, J. (1987) *Genetics* **117,** 645–655.