31

GENETIC FOOTPRINTING AND FUNCTIONAL MAPS OF THE YEAST GENOME

Tracy Ferea, Barbara Dunn, David Botstein, and Patrick Brown

Upon completion of the *Saccharomyces cerevisiae* genome it became clear that many of the genes obtained by the sequencing effort did not have a function inferable from their sequence. Approximately 44% of the genes have been characterized either experimentally or by their strong similarity to proteins or protein families with known functions (Cherry *et al.*, 1998; Mewes *et al.*, 1997). The majority of the genome currently consists of functionally uncharacterized open reading frames (ORFs). In *S. cerevisiae,* as with other organisms, gene function has typically been inferred by generating mutations in an individual gene and then testing fitness under a variety of conditions to detect a phenotype that suggests a function. Genetic footprinting is a highly parallel genomic approach that was developed to meet the challenge of determining gene function more efficiently (Smith *et al.*, 1996).

Genetic footprinting employs an insertional mutagenesis strategy (Fig. 1). A thorough mutagenesis is accomplished by transiently driving the expression of the Ty1 transposon from a regulatable promoter. The mutated population is then tested for fitness under a variety of selective growth conditions. Fitness is determined for each ORF retrospectively using PCR to compare the Ty1 insertion pattern at

PCR Applications

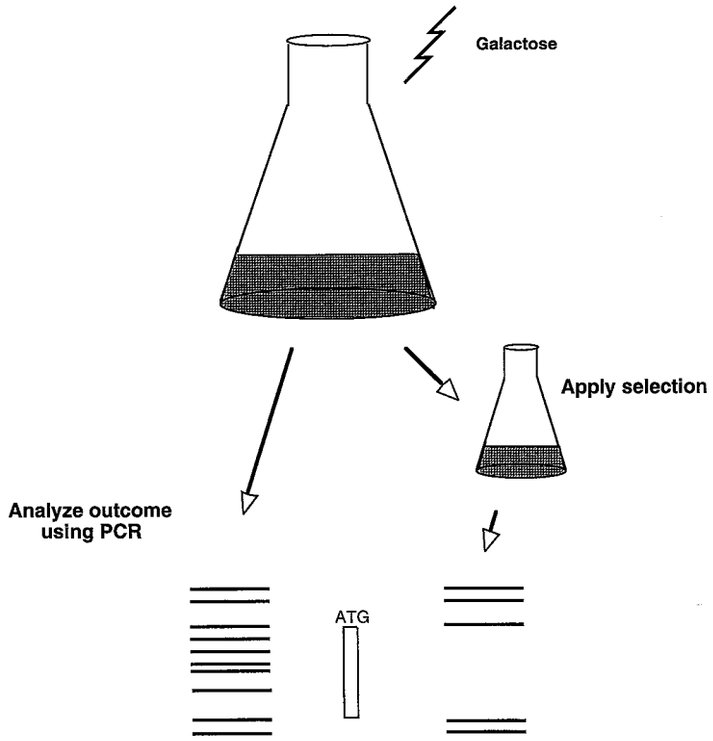

Figure 1 Mutagenesis is triggered by growing cells on medium containing galactose. Galactose induces the expression of the Ty1 transposon from the GAL1 promoter and subsequent insertion into the genome. For each gene, PCR is employed to compare the Ty1 insertion patterns both before and after growth on selective medium. Insertion into the coding region of a gene important for growth on selective medium results in a loss of cells containing those insertions from the growing population and a concomitant loss of PCR products.

each gene in the mutated population both before and after selection. When insertion in a gene results in a growth disadvantage, cells carrying insertions in that gene become underrepresented as the population expands. The relative depletion of these cells from the population results in a concomitant loss of their PCR products. PCR products from the selected population and the original mutagenized population are then compared. A growth disadvantage upon selection is seen as a "footprint" or loss of PCR products resulting from insertion within regions of the gene essential for function or expression. The pattern of PCR products from outside that region remains intact.

Genetic footprinting is efficient because both the mutagenesis and the selection processes are performed on large populations of cells. Therefore, these processes need to be performed only once and then PCR can be employed to analyze every gene for which a unique primer has been designed. When performed at the genome scale such an analysis is the equivalent of a saturation mutagenesis. The comprehensive nature and efficiency of the genetic footprinting method should expedite the genomic scale mapping of gene function.

Plasmid and Strain Construction

The plasmid pPBTy1 was constructed from two plasmids, pVIT41 and pJEF1562, obtained from Jef Boeke (Johns Hopkins, Boston). A 7.5-kb *Apa*I/*Xho*I fragment of pVIT41 was isolated. This fragment contained the Ty1 transposable element and a unique sequence complementary to the PBTy1R1 oligonucleotide (5′ AGAGCTCCCGGGATCCTCTACTAAC 3′). This unique sequence serves as a molecular marker distinguishing the transposable element from endogenous transposons. After endonuclease digestion a 6.5-kb *Xho*I/*Apa*I fragment of pJEF1562 was isolated. The key features of the 6.5-kb fragment are the 2-μm-based plasmid backbone for high copy number, the *URA3* gene for use as a selectable marker, and the regulatable *Gal1* promoter. Ligation of the two fragments generated the plasmid pPBTy1.

A uracil auxotroph of the haploid yeast strain X2180-1A, a derivative of S288C, was selected on 5-fluoroorotic acid after transformation and homologous integration of a linear *Hin*dIII fragment of the *URA3* gene with an internal deletion of the 250-bp *Eco*RV/*Stu*I fragment. Effective induction of Ty1 transposition requires the strain to be GAL$^+$; therefore, the galactose permease gene (*GAL2*) which is defective in many strains of the S288C background was replaced by two-step gene replacement (Guthrie and Fink, 1991). Briefly, a YIp5 vector containing a 2.5-kb *Eco*RI/*Hin*dIII fragment of the *GAL2* gene was linearized within the *GAL2* coding region with *Bgl*II. The linear fragment with *GAL2* ends, for integration, also contained the *URA3* gene for selection. Integration at the *gal2* locus generated a duplication with a wild-type copy of the gene, flanked by the mutant copy. Plasmid excision was selected on 5-fluoroorotic acid and colonies were screened for the wild-type *GAL2* phenotype on

YP galactose medium containing antimycin A (10 μg/ml). Transformation with pPBTY1 completed the construction of the strain DBY7282 employed in the mutagenesis procedure.

Mutagenesis

In *S. cerevisiae,* the mutagenesis is performed by transiently expressing a marked Ty1 transposable element under the control of the regulatable *GAL1* promoter (Boeke *et al.*, 1985), from the high copy number plasmid pPBTy1. Exponential growth on SC galactose medium (Rose *et al.*, 1990) at 25°C for 20 generations induces sufficient transposition to obtain footprints for >97% of genes (Smith *et al.*, 1996). The efficiency of mutagenesis can be monitored by determining the number of canavanine-resistant mutants produced by insertion into the arginine permease gene *CAN1* (Rose *et al.*, 1990). Typically, a population with a measured mutant frequency of approximately 2×10^{-5} *can1*^{R} mutants/cell indicates sufficient transposition to obtain good footprinting results for other genes in the genome.

After a large-scale mutagenesis, DNA is isolated from at least 10^{12} cells. This cell number should yield enough DNA to analyze the effectiveness of the mutagenesis for nearly every gene in the genome. The remaining cells are frozen as glycerol stocks for future selection experiments in aliquots of $\geq 2 \times 10^{8}$ cells.

Selection

Typically a selection is initiated and maintained with $\geq 2 \times 10^{8}$ mutagenized cells to help prevent stochastic loss of mutants from the population (i.e., by random genetic drift). Cells are grown under the selective condition for 15–60 generations. The cells containing insertions into a gene important for growth under a particular selective conditions are depleted from the population. For example, when the overall population has doubled 15 times, a mutant growing at a rate of 80% of wild-type will have doubled only 12 times. Therefore, at 15 generations the mutant will be represented at $2^{12}/2^{15}$ or one-eighth of its original abundance in the mutagenized population. Thus, even more subtle growth defects can be detected by extending the number of generations the population is grown under selective pressure.

As with any selection the parameters of the experiment must be carefully designed. Therefore, pH, temperature, and oxygenation of the culture should be carefully controlled. Careful review of the components of the medium is also essential. The population should be maintained at a sufficient number of cells to avoid stochastic loss of individual mutations during the selection process. Unless the purpose of the experiment is to test the effects of mutations on cell viability during various phases of growth, exponential growth should be maintained throughout the selection process.

Large-Scale Yeast DNA Isolations

Large-scale DNA preparations from 1×10^{12} cells can be performed by scaling up the yeast DNA isolation procedure described by Rose *et al.*, (1990). Only slight modifications to the procedure are necessary.

1. Grow 10 liters of cells to approximately 1×10^{8} cells/ml. Centrifuge in a Sorvall GS3 rotor at 3000 rpm for 5 min (3 liters at a time). Discard the supernatant.
2. Wash cells by resuspending pellets in a total of 100 ml of 1.0 *M* sorbitol. Pool cells and centrifuge in GS3 rotor at 3000 rpm for 5 min. Discard the supernatant.
3. Resuspend the cells in 225 ml of 1.0 *M* sorbitol, 0.1 *M* Na_2 EDTA (pH7.5), and 14 m*M* β-mercaptoethanol.
4. Add 7.5 ml of a 2.5 mg/ml solution of zymolyase-100T (ICN Immunobiologicals, P. O. Box 1200, Lisle, IL 60532) dissolved in 1.0 *M* sorbitol, 0.1 *M* Na_2 EDTA (pH 7.5), and 14 m*M* β-mercaptoethanol. Incubate in a 2-liter flask at 37°C for 60 min, shaking gently (50 rpm).
5. Centrifuge the cells in a GSA rotor for 5 min at 3000 rpm. Discard the supernatant.
6. Resuspend the cell pellet in a total of 375 ml of 50 m*M* Tris–Cl (pH 7.4) and 20 m*M* Na_2 EDTA.
7. Add 37.5 ml of 10% SDS and mix.

8. Incubate at 65°C for 30 min in a 1-liter flask at 100 rpm.
9. Add 112.5 ml of 5 *M* potassium acetate and store on ice for 60 min.
10. Centrifuge in a GSA rotor at 11,000 rpm for 10 min. Transfer supernatant to a fresh tube and repeat centrifugation.
11. Transfer the supernatant to a fresh GSA tube and add 2 vol of 95% ethanol at room temperature. Mix and centrifuge at 11,000 rpm for 15 min at room temperature.
12. Discard the supernatant. Dry the pellet and then resuspend in 225 ml of 10 m*M* Tris–Cl (pH 7.4) and 1 m*M* EDTA. This may take several hours.
13. Add 1.25 ml of a 10 mg/ml solution of RNase A and incubate at 37°C for 30 min.
14. Add 1 vol of 100% isopropanol and gently mix. Remove the precipitate by stiring with a glass Pasteur pipet. The precipitate should wind around the pipet like a loose "cocoon" of fibers. Air-dry cocoon.
15. Wash with 70% ETOH, followed with a 95% ETOH wash. Resuspend in 10 m*M* Tris–Cl (pH 7.4) and 1 m*M* EDTA to a concentration of 500 ng/μl. Clarify the DNA solution by centrifugation in a Sorvall SS-34 rotor at 10,000 rpm for 15 min.

PCR

To examine a representative population of cells for each gene 1 μg of DNA (approximately 10^8 genomes) is used per PCR reaction. One primer is an oligonucleotide (PBTy1R1) complementary to a marked region of the *Ty1* element, that is not present in the endogenous element. The other is a gene-specific oligonucleotide that is fluorescently labeled at the 5′ end with FAM. A nearly complete set of primers for the *S. cerevisiae* genome is available from Research Genetics (2130 Memorial Parkway, Huntsville, AL 35801; *www.resgen.com*). The gene-specific primers were chosen to typically survey approximately 700 bp of coding region with a calculated T_M of 72°C by using the Design Primers program (*http://genome-www.stanford.edu/Saccharomyces/*). As a control for spurious products PCR analyses are also performed on DNA isolated from cells

(DBY7282) that have not been induced for Ty1 transposition, i.e., they were grown on SC glucose medium.

The PCR reaction volume is 50 μl and reactions are set up in 96-well plates.

1 μg DNA
1.5–2.5 m*M* MgCl
250 μ*M* each dNTP
50 m*M* KCl
10 m*M* Tris–HCl,pH 8.3
2 U *Taq* polymerase
0.64 μ*M* gene-specific oligonucleotide
0.38 μ*M* Ty1-specific oligonucleotide

The 50-μl PCR reaction mix is centrifuged to the bottom of the plate in a Beckman GS6 centrifuge set at 22°C, 600 rpm, for 1 min. The plate is then loaded into a preheated 93°C Perkin–Elmer Model 9700 thermocycler.

Thermocycling parameters

Denature at 93°C for 1 min
10 cycles of 92°C for 30 sec, 67°C for 45 sec, 72°C for 2 min
20 cycles of 92°C for 30 sec, 62°C for 45 sec, 72°C for 2 min

Analysis of PCR Products

PCR products are analyzed on Applied Biosystems DNA sequencing machines, Models 373 or 377, equipped with GeneScan DNA analysis software and with the XL upgrade. The advantage of the XL upgrade is that it expands the gel capacity to 64 samples per run. Notched microtrough gel plates with a 12-cm well-to-read distance for the ABI 373 and 6 cm for the ABI 377 are recommended (CBS Scientific, P. O. Box 856, Del Mar, CA 92014). The notches increase sample volume capacity, facilitating loading. A denaturing polyacrylamide gel matrix with slightly different compositions is used depending on the model of ABI machine: 4% Gene-Page Plus for the ABI 377 and 6% for the ABI 373 (Amresco, Inc., 30175 Solon Industrial Parkway, Solon, OH 44139).

Five microliters of each PCR product and 2 μl of loading buffer

are combined in wells of a 96-well plate. Loading buffer is made beforehand and stored at −20°C in covered 96-well plates. Centrifuge at 1000 rpm to spin samples to bottom of plate and then denature at 95°C for 10 min. The denatured sample, now approximately 1 μl in volume, is loaded directly from the hot block onto the polyacrylamide gel using an eight-channel Hamilton syringe. Loading is staggered by loading every other lane and running the gel 2–5 min prior to the second loading; this staggers the molecular weight markers to facilitate lane tracking. Running parameters are 16 W for 8 hr for the ABI 373 or 640 V for 4 hr for the ABI 377.

Loading buffer

2 ml loading dye: 30 mg/ml dextran blue and 50 m*M* EDTA (pH 8)

1.8 ml GeneScan-2500 TAMRA molecular weight markers (Applied Biosystems, 850 Lincoln Center Dr., Foster City, CA 94404)

10 ml formamide

Interpretation

ABI GeneScan analysis software is employed to display electropherograms of the PCR products. Lane tracking is adjusted, and then the local Southern method (Southern, 1979) is used to examine the reciprocal relationship between mobility and fragment length to determine the molecular weights of PCR products relative to the GeneScan 2500 molecular weight standards. The accuracy of automated size calling should always be visually assessed. Although some PCR-to-PCR variability is seen in the pattern produced by genes that have relatively few insertions, the complex patterns of insertions found for most genes are well conserved even after selective growth. For example, electropherograms from the *MET10* gene (which encodes a protein necessary for methionine prototrophy) are shown in Fig. 2. The mutagenized cells (time 0) have a dense pattern of PCR products resulting from multiple independent insertions into the gene within the analyzed population. Growth on medium containing the nonfermentable carbon source lactate results in a pattern similar to that of the mutagenized population, suggesting that this gene is not essential for growth under this condition. However,

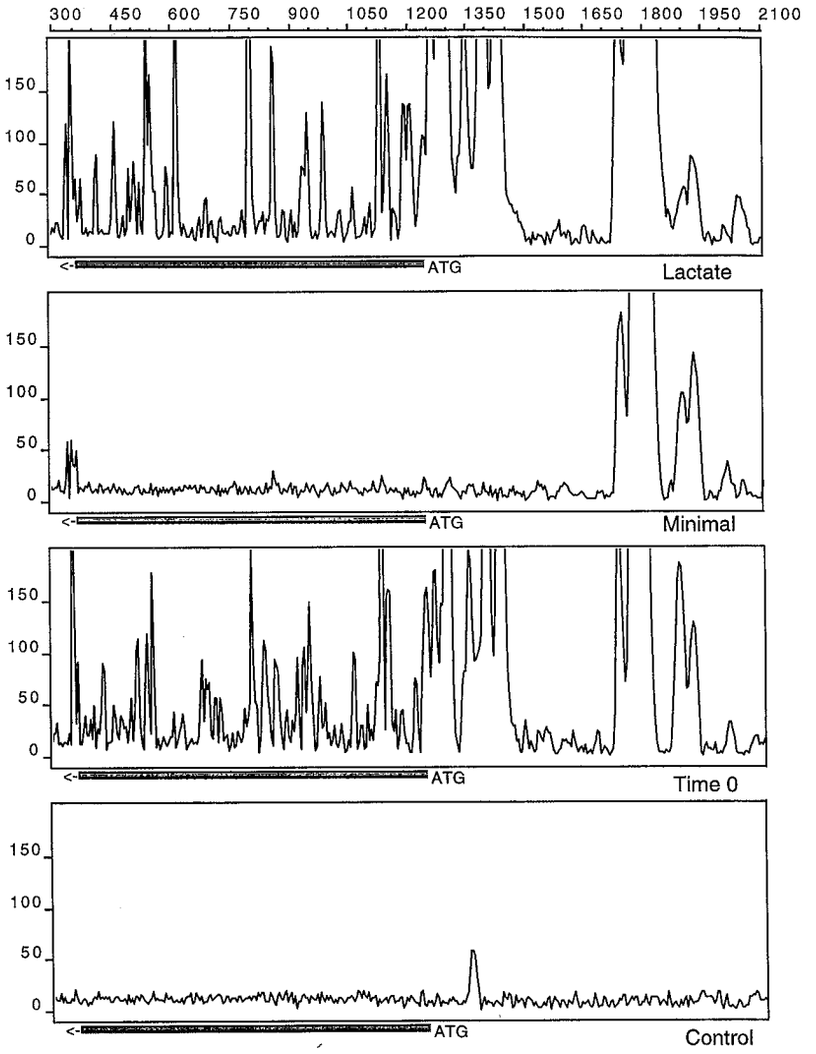

Figure 2 Electropherograms depicting the Ty1 insertion pattern at the *MET10* locus. Relative fluorescent intensity is indicated on the *y* axis and molecular weight on the *x* axis. Each electropherogram contains PCR products from cells grown as follows: the initial mutagenized population (time 0), unmutagenized cells (control), or cells from the mutagenized population after 21 population doublings on lactate or minimal selective medium. The area of the trace emcompassing the start codon (ATG) and first 1246 bases of the 3107-bp *Met10* coding region is depicted with a bar under each electropherogram. Note that peaks representing insertions in the *MET10* coding region are depleted after growth on minimal medium. The lack of significant peaks in the control indicates the specificity of the oligonucleotide primers.

growth on minimal medium that lacks methionine results in a loss of PCR products from both the coding region and the immediate upstream vicinity of the gene. This "genetic footprint" indicates that this gene is essential for growth in minimal medium.

The footprinting data should be interpreted with appropriate caution. Improper interpretation of the data can result when comparing improperly scaled electropherograms. It is important to normalize the data by scaling the peaks upstream of the coding region to the same height, the underlying assumption being that insertions far enough upstream of the coding region do not cause a selective growth disadvantage. Normalization compensates for differences in baseline subtraction, quantities loaded, and slight differences in PCR effectiveness between samples. Inadequate extension during PCR, resulting in low-amplitude signal for products greater than approximately 500 bp, can also make interpretation difficult. If PCR products large enough to survey outside of the coding region are not obtained, the PCR should be repeated. Peaks with relative fluorescent intensity of <50 should not be used for analysis or scaling because increasing their scale also increases the background signal resulting in the inability to distinguish a footprint among the noise. Repeating the PCR, or designing a new primer, usually results in peaks of higher intensity. Peaks that are too intense because they are located near integration hot spots can also be difficult to interpret. Intense peaks can cause exhaustion of the PCR reagents resulting in partially extended products that can mask a footprint. This is easily remedied by repeating the PCR with fewer cycles.

Application to Other Organisms

The genetic footprinting strategy is applicable to other organisms with minimal requirements: (i) a suitable insertional mutagen that can be used to generate enough independent mutations to saturate the genome; (ii) DNA sequence information sufficient to choose specific primers; (iii) a genome small enough that a population of mutagenized cells, sufficient to include a representative sample of mutations in each gene, can be present within every PCR reaction. Growth as a haploid or homozygous diploid is desirable if recessive phenotypes are to be analyzed. With the explosion of microbial sequencing projects, these requirements could easily be met for an

increasing number of microorganisms. Modifications of the procedure permit footprinting of viral genomes or cloned DNA at a high resolution (Singh *et al.*, 1997).

Functional Maps

The ability to survey an entire genome for the contributions of each gene to cellular fitness under a variety of selective conditions makes it possible to generate comprehensive functional maps for an organism. The entire repertoire of genes essential for a specific regulatory mechanism, signal transduction, or biochemical pathway can be mapped with appropriately designed experiments. We initiated this process by examining several selective conditions that have been widely used in analysis of *S. cerevisiae* gene function. This has resulted in a functional map for the 268 genes that reside on chromosome V of *S. cerevisiae* (Smith *et al.*, 1996). A detectable mutant phenotype was found for 62% of these genes after analyzing only seven selections. It is only a matter of time before a complete functional map of the genome is made. With 31% of yeast genes having mammalian homologs (Botstein *et al.*, 1997) among the still rapidly growing Genbank entries, it is easy to imagine the impact of a eukaryotic functional map on the understanding of cognate genes in other organisms.

Acknowledgments

We thank Rebecca Koskela, Sandra Metzner, Edward Chung, and Katja Schwartz for critical reading of the manuscript.

References

Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. Ty elements transpose through an RNA intermediate. *Cell* **40**(3), 491–500.

Botstein, D., Chervitz, S. A., and Cherry, J. M. (1997). Yeast as a model organism. *Science* **277**(5330), 1259–1260.

Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A., Dwight, S. S., Hester, E. T., Jia, Y., Juvik, G., Roe, T., Schroeder, M., Weng, S., and Botstein, D. (1998). SGD: Saccharomyces Genome Database. *Nucleic Acids Res.* **26**(1), 73–79.

Guthrie, C., and Fink, G. R. (1991). *Methods in Enzymology: Guide to Yeast Genetics and Molecular Biology*, Vol. 194, Academic Press, San Diego.

Mewes, H. E., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maieri, A., Oliver, S. G., Pfeiffer, F., and Zollner, A. (1997). Overview of the yeast genome. *Nature*, **387** (Suppl.), 7–65.

Rose, M. D., Winston, F., and Hieter, P. (1990). *Methods in Yeast Genetics: A Laboratory Course Manual.* Cold Spring Harbor Laboratory Press, Plainview, NY.

Singh, I. R., Crowley, R. A., and Brown, P. O. (1997). High-resolution functional mapping of a cloned gene by genetic footprinting. *Proc. Natl. Acad. Sci. U.S.A.* **94**(4), 1304–1309.

Smith, V., Botstein, D., and Brown, P. O. (1995). Genetic footprinting: A genomic strategy for determining a gene's function given its sequence. *Proc. Natl. Acad. Sci. U.S.A.* **92**(14), 6479–6483.

Smith, V., Chou, K. N., Lashkari, D., Botstein, D., and Brown, P. O. (1996). Functional analysis of the genes of yeast chromosome V by genetic footpringing. *Science* **274**(5295), 2069–2074.

Southern, E. M. (1979). Measurement of DNA length by gel electrophoresis. *Anal. Biochem.* **100**(2), 319–323.