Sorting out metagenomes

Jo Handelsman

An ingenious screening method exploiting substrate induction of gene expression provides a rapid means of identifying new catabolic pathways in unculturable microorganisms.

© 2005 Nature Publishing Group http://www.nature.com/naturebiotechnology

Many of the microorganisms on earth are shrouded from our curious eyes because of their reluctance to grow in pure culture. The recognition that readily cultured microorganisms represent only 0.1% to 1% of the organisms in most habitats[1] has fueled interest in methods for studying microbial life that are not predicated on pure cultures. The major impediment of such methods has been the onerous process of screening clones for new activities. In this issue, Uchiyama *et al.*[2] present an innovative screen that rapidly sorts thousands of environmental DNA clones to identify those with a desired metabolic activity. The screen relies on a reporter system to trap genes encoding biodegradative pathways, and is based on the prediction that operons encoding these pathways are likely to be induced by the substrate of the pathway.

Metagenomics, the genomic analysis of an assemblage of microorganisms, provides a window on the riches of the uncultured world. This culture-independent genomic analysis requires cloning DNA that has been extracted directly from an environmental sample[3–5]. The advent of metagenomics was accompanied by promises of great biotechnological applications as well as fundamental discovery. Although it has revealed fundamental knowledge[6,7] and applications[4,8] of the microbial world that could not be attained by culture-based analysis, the promise of metagenomics as a source of new technology has not been fully realized, primarily because of the challenges in screening for desired activities.

Screening of metagenomic libraries has traditionally followed two paths: sequence-based and function-based screening (**Fig. 1**). Some libraries have been screened by hybridization or PCR to detect genes with homology to known genes. This approach has been fruitful, but is necessarily limited to discovery of genes in previously described families.

The alternative is functional analysis, which enables researchers to discover entirely new classes of genes for useful functions but requires expression of genes from exotic organisms in tame bacteria such as *Escherichia coli*. Functional analysis has depended largely on low-throughput screens, such as visual detection of growth inhibition of indicator bacteria, and selection, such as resistance to antibiotics[9] or complementation of necessary biochemical pathways[8]. Selection is a powerful way to detect a rare clone among many, but most phenotypes of biotechnological interest do not lend themselves to selection. Low-throughput screens can detect a wider range of activities, but their utility is limited by the low frequency of active clones, which is typically less than 10^{-4} in any one assay.

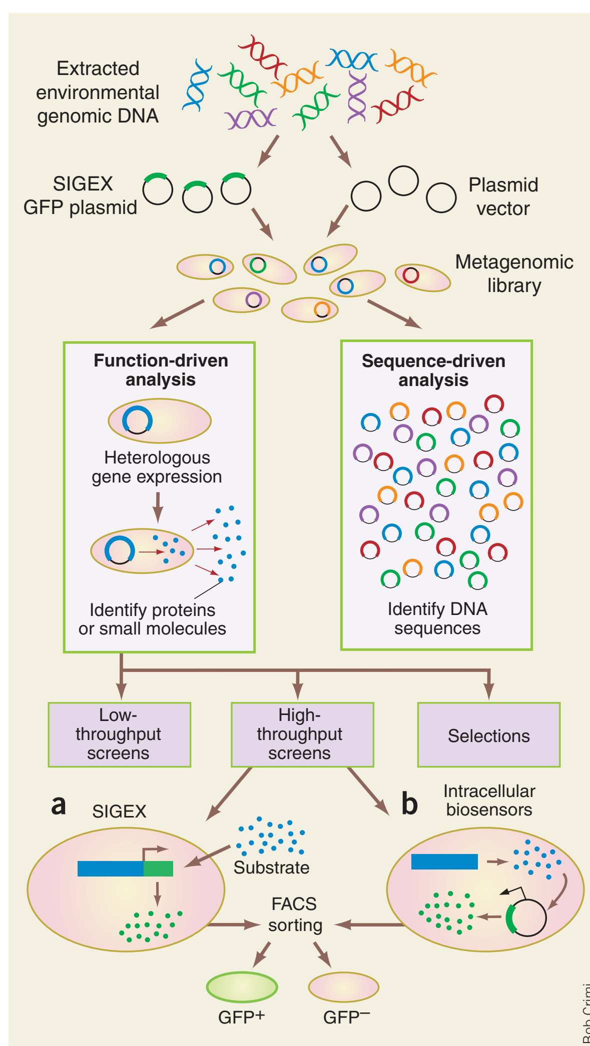

Figure 1 Scheme for constructing and screening metagenomic libraries. DNA is extracted directly from environmental samples, cloned into an appropriate vector and transformed into a bacterial host. Transformants of interest can be identified based on DNA sequence (by PCR or hybridization) or based on an expressed function. Functional analysis can be with low-throughput screens (each clone tested separately for a desired feature), high-throughput screens (rapid, usually automated method to separate desirable candidates from the other clones), or by selection (in which only the desirable clones grow). (**a**,**b**) Two examples of high-throughput screens are SIGEX (**a**) and an intracellular biosensor (**b**). SIGEX exploits the principle that catabolic genes are often substrate-induced by fusing a promoterless *gfp* to the metagenomic DNA and identifying clones in which GFP production is induced by the substrate of interest[2]. An intracellular biosensor detects biologically active small molecules. GFP expression is dependent on the presence of a small molecule that activates a regulator (an example is a quorum-sensing promoter activated by acyl homoserine lactones or other quorum-sensing inducing molecules)[10].

The search for high-throughput screens that do not require a selectable phenotype has led to a focus on phenotypes, such as pigments, that are readily visible, and to the use of fluorescence-activated cell sorting, for example, to detect expression of certain types of genes by regulation of a fluorescent biosensor present in the same cell as the metagenomic DNA (**Fig. 1** (ref. 10)). High-throughput screens will be a critical tool for discovery within metagenomic libraries.

Uchiyama *et al.* have designed a clever, high-throughput screen for catabolic pathways, designated ‘substrate-induced gene expression

Jo Handelsman is in the Department of Plant Pathology, 1630 Linden Drive, University of Wisconsin, Madison, Wisconsin 53706, USA.
e-mail: joh@plantpath.wisc.edu

screening' or SIGEX. The screen is based on the observation that genes encoding the enzymes for catabolic pathways are commonly arranged in operons that are induced by the pathways' substrates. The authors exploit this feature in an 'operon trap' in which the metagenomic DNA is cloned upstream of the *gfp* gene, thereby placing green fluorescent protein (GFP) expression under the control of promoters in the metagenomic DNA. They seek clones in which GFP expression is regulated by the substrate of interest (**Fig. 1**).

The power of the screen is illustrated with a search for genes involved in the degradation of naphthalene and benzoate. All clones that fluoresce in the absence of the substrate are discarded, and the remaining clones are exposed to naphthalene or benzoate; those clones that fluoresce are captured by fluorescence-activated cell sorting. From a library of 152,000 clones with inserts averaging 7 kb, Uchiyama *et al.* swiftly identified 58 clones regulated by benzoate and 4 by naphthalene. Some of the clusters identified by SIGEX encode proteins with an obvious connection to benzoate or naphthalene degradation, such as a putative benzoate dioxygenase, and others have no apparent connection with the targeted function.

The potential of SIGEX is demonstrated by a detailed functional analysis of clone BZO71. The clone had no activity in a benzoate transformation assay, but one of its genes, encoding a P450 homolog, caused the chemical transformation of 4-hydroxybenzoate when expressed from a highly active promoter. Clone BZO71 would not have been detected by sequence-based screening because the protein it encodes is the first P450 associated with transformation of 4-hydroxybenzoate. It would not have been detected by a functional expression assay because protein expression was poor in the original clone before it was engineered to boost protein expression. Although the original clone was not optimized for expression, the promoter activity in the metagenomic DNA was sufficiently modulated by benzoate to be detected by SIGEX.

SIGEX adds a significant technology to the field of metagenomics. Its reliance on promoter activity, rather than on sequence similarity or phenotypic expression, circumvents many of the limitations of previous metagenomic screens. The high-throughput capacity of SIGEX makes it possible to screen large libraries rapidly (almost as efficiently as a selection). These advantages catapult metagenomic analysis into a new sphere, providing access to genes that were silent or too rare to be revealed by other methods.

Although SIGEX will lead to the discovery of many genes, these genes must share certain features that limit the method's utility. The promoter that responds to the substrate is necessarily upstream of the catabolic gene or genes. To be detected, a clone must contain enough of an operon to be useful in subsequent functional studies, but it cannot contain the terminal part of the operon if it carries a transcription terminator that will prevent GFP expression. Substantial redundancy in the library will increase the chance of obtaining a fragment of DNA that meets these stringent requirements.

Perhaps the most significant constraint on SIGEX is the requirement that regulatory machinery that recognizes the promoter and substrate be present and functional in the host cell. The host cell must also be able to take up the substrate. These requirements limit discovery of new catabolic genes to those whose regulatory elements are similar to those of *E. coli*. Thus, it is no surprise that more than 60% of the genes identified by Uchiyama *et al.* have significant sequence similarity to genes from Proteobacteria, the phylum in which *E. coli* resides. The authors wisely suggest that conducting the SIGEX screen in other host species will overcome this barrier and broaden the capacity of SIGEX for discovery of novel genes.

The SIGEX template allows infinite variation in the functions to be discovered, the regulatory molecules to be employed and the reporter proteins to be detected. The high-throughput nature of the screen and of related screens that it will spawn makes SIGEX a landmark in metagenomic analysis. Sorting through massive libraries representing complex communities that contain thousands of species is now within reach, providing an opportunity for a new assessment of the value of metagenomics to biotechnology.

1. Pace, N.R. *Science* **276**, 734–740 (1997).
2. Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. *Nat. Biotechnol.* **23**, 88–93 (2005).
3. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. & Goodman, R.M. *Chem. Biol.* **5**, R245–249 (1998).
4. Healy, F.G. *et al. Appl. Microbiol. Biotechnol.* **43**, 667–674 (1995).
5. Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. & DeLong, E.F. *J. Bacteriol.* **178**, 591–599 (1996).
6. Beja, O. *et al. Science* **289**, 1902–1906 (2000).
7. Tyson, G.W. *et al. Nature* **428**, 37–43 (2004).
8. Henne, A., Daniel, R., Schmitz, R.A. & Gottschalk, G. *Appl. Environ. Microbiol.* **65**, 3901–3907 (1999).
9. Diaz-Torres, M.L. *et al. Antimicrob. Agents Chemother.* **47**, 1430–1432 (2003).
10. Handelsman, J. *Micro. Molec. Biol. Rev.* **68**, 669–685 (2004).

© 2005 Nature Publishing Group http://www.nature.com/naturebiotechnology