Problems with metagenomic screening

To the editor:

The paper by Kazuya Watanabe and colleagues published in the January issue (*Nat. Biotechnol.* **23**, 88–93, 2005) is framed both in the paper itself and in the accompanying News & Views by Jo Handelsman (*Nat. Biotechnol.* **23**, 38–39, 2005) as an astute, general genetic procedure for fishing out catabolic genes from environmental metagenomes. The proposed method exploits the concept of guilt-by-association by assuming that genes determining transcriptional regulators of pathways for metabolism of many chemical species lie adjacent to the very catabolic genes they regulate. On this basis, genes encoding potentially interesting enzymes can be captured through the fluorescence-activated cell sorting of substrate-induced *Escherichia coli* cells bearing random green fluorescent protein fusions to metagenomic DNA fragments. Although the approach has potential in selecting transcriptional factors that respond to given chemical compounds, for several reasons it is likely to work in only a very few instances in environmental screens for novel catabolic enzymes, as I outline below.

The first problem is that transcriptional regulators may not be associated with operons controlled *in cis*. Although the long-evolved regulatory network of *E. coli*[1] does contain such gene arrangements, they are much less common in recently evolved metabolic pathways, where the most interesting and newest enzymes are to be found.

A second, more important, caveat is that catabolic promoters for xenobiotic-degrading operons may not be regulated and thus are expressed constitutively or semi-constitutively. It is very unlikely, therefore, that these types of genes would show up in the screen of Watanabe and colleagues.

A more troubling difficulty with the procedure is that transcriptional regulators are frequently activated by effectors that are not substrates of the pathways that they regulate. Conversely, substrates for interesting enzymes may not always induce expression of the corresponding genes; in fact, there is a considerable promiscuity in how substrates induce enzymes[2–4], which is likely to endow the screening system described by Watanabe and colleagues with considerable noise from false positives and false negatives.

A final problem (which is shared by all *E. coli*-based strategies for surveying metagenomic libraries) is the uncertainty about expression of the genes encoded by environmental DNA sequences in the surrogate bacterial host. Although genes born by Gram-negative bacteria for catabolism of long-existing, natural hydrocarbons are likely to be picked up with this procedure, the most recently evolved genes encoding metabolic enzymes for more recent xenobiotics are likely to be missed because of the suboptimal evolution of their regulatory circuits in the short time frame[2]. Needless to say, the strategy of Watanabe and colleagues is also blind to archaeal DNA, fungal DNA and any other DNA encoding transcriptional factors that cannot work in concert with the RNA polymerase of *E. coli*. New procedures for screening the whole enzymatic contents of microbial metagenomes are still badly needed to tackle these challenges in environmental biotechnology.

Victor de Lorenzo

Centro Nacional de Biotecnología, CSIC, Campus de Cantoblanco, Madrid 28049, Spain.
e-mail: vdlorenzo@cnb.uam.es

1. Warren, P.B. & ten Wolde, P.R. *J. Mol Biol.* **342**, 1379–1390 (2004).
2. de Lorenzo, V. & Perez-Martin, J. *Mol. Microbiol.* **19**, 1177–1184 (1996).
3. Cases, I & de Lorenzo, V. *EMBO J.* **20**, 1–11 (2001).
4. Tropel, D. & van der Meer, J.R. *Microbiol. Mol. Biol. Rev.* **68**, 474–500 (2004).

Watanabe and colleagues respond:

de Lorenzo outlines some possible limitations of the substrate-induced gene expression screening (SIGEX) method. Although we discussed many of the points in his letter in our paper, it is worth reiterating them to emphasize the challenges faced in designing a metagenomic screen.

Possible limitations of SIGEX include the following: first, the expression of catabolic genes is not always induced by their substrates; second, regulatory genes and catabolic genes are not always close to each other; and third, genes in organisms distantly related to a SIGEX host may not be obtained. We consider that enough genes not affected by these limitations are present in the metagenome to enable SIGEX to be of significant utility in finding new enzyme activities. Because the environment harbors a huge repertoire of catabolic genes, we can reasonably deduce that there exist in the natural gene pool at least some genes of interest that are detectable, despite SIGEX's limitations.

Metagenomics is a recently evolved approach for tackling natural genetic diversity. Novel genes have been obtained using existing metagenome screening methods (that is, sequence-based and enzyme activity–based methods). Similar to SIGEX, however, these methods can fish up only a small fraction of genes from the natural gene pool. The point that de Lorenzo makes, that SIGEX is blind to archaeal and fungal sequences, is in fact true of all metagenomic analysis methods described to date.