Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Biotechnology

Functional metagenomics for enzyme discovery: challenges to efficient screening

Taku Uchiyama[1] and Kentaro Miyazaki[1,2]

Metagenomics has emerged as an alternative approach to conventional microbial screening that allows exhaustive screening of microbial genomes in their natural environments. Despite the potential usefulness of this approach, functional analysis of the metagenome is often problematic because of insufficient and biased expression of the cloned genes in *Escherichia coli*. This review highlights recent studies on the screening of metagenomic libraries and discusses some possible solutions for overcoming the expression problem in function-based screening.

Addresses
[1] Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan
[2] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

Corresponding author: Miyazaki, Kentaro (miyazaki-kentaro@aist.go.jp)

Current Opinion in Biotechnology 2009, **20**:616–622

This review comes from a themed issue on
Chemical biotechnology
Edited by Kazuya Watanabe and George Bennett

Available online 21st October 2009

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2009.09.010

Introduction

Enzymes are favorable catalysts for the development of environmentally benign industrial processes. To date, most industrially relevant enzymes are of microbial origin. Therefore, mining for microbial enzymes is a key step in the development of industrial bioprocesses, but we can cultivate less than 1% of environmental bacteria through standard laboratory techniques [1].

Metagenomics has appeared as an alternative approach to conventional microbial screening. By directly cloning environmental DNA (or metagenome) in a surrogate host, one can exhaustively investigate the metagenome, independent of the culturability of the source organisms [2,3]. This innovative technology is anticipated to accelerate the discovery process of novel useful genes from microorganisms [4–6]. In fact, various novel enzymes have been identified that have unique activities and/or sequences.

Currently, metagenomes are screened based on either function (activity) or sequence. Function-based screening is a straightforward way to obtain genes that have desired functions, but is often problematic, due primarily to the biased and insufficient expression of anonymous genomic fragments in *Escherichia coli*. By contrast, sequence-based screening is readily performed using either PCR-based or hybridization-based procedures, but the genes obtained through this approach are limited to those having homologies to the probe sequence and may not allow us to obtain novel genes. Various efforts have been made to solve these problems in the screening of useful genes from the metagenomes. This review highlights recent studies on the screening of metagenomic libraries and discusses possible solutions for overcoming these problems.

Functional screening of metagenomic libraries

Recent studies on function-based screening of metagenomic libraries are summarized in Table 1. The probability (hit rate) of identifying a certain gene depends on multiple factors that are inextricably linked to each other: the host–vector system, size of the target gene, its abundance in the source metagenome, the assay method, and the efficiency of heterologous gene expression in a surrogate host. In this section, we present an overview of trends in activity-based screening and discuss possible technical clues to overcome the low hit rate problem. The more fundamental problem — insufficient, biased expression of foreign genes in *E. coli* — will be discussed in the next section.

For the construction of a library, most researchers use *E. coli* as a surrogate host. They lack genes for homologous recombination (*recA*, *recBC*) and restriction (*mcrA*, *mcrBC*), which are useful for cloning variously modified foreign DNA into *E. coli*. Various types of *E. coli* strains are available as highly efficient competent cells from commercial sources. The choice of a vector depends largely on the length of the inserts. Plasmids are suitable for cloning smaller than 10-kb DNA fragments, and cosmids (25–35 kb), fosmids (25–40 kb), or BACs (100–200 kb) can be used to clone larger fragments. Despite the differences in preferred length of inserts, however, the targets are mainly single-component enzymes, even when using the latter vectors — a situation that has not changed over several years. Among these vectors, plasmids have high copy numbers and strong vector-borne promoters. Nevertheless, these apparent merits do not improve the hit rate significantly (Table 1). For example, Simon *et al.* [14•]

Table 1

Activity-based approach to screen enzymes in the period 2007 to early 2009

Environment	Target gene	Host, vector	Average size of insert (kb)	# of positives/# of screened clones	1 positive/Mb DNA screened	Assay technique	Reference
Compost	Esterase Amylase Phosphatase Dioxygenase Protease	*E. coli*, plasmid	3.2	14/21 000 13/31 967 38/31 967 2/31 967 0/31 967	1/4.8 1/7.9 1/2.7 1/51.1 0/102.3	Agar plate assay	[7•]
Compost (artificially contaminated with poly-lactic acid disks)	Esterase	*E. coli*, plasmid	2.5	3/40 000	1/33.3	Agar plate assay	[8]
Soil (nonspecified)	Xylanase	*E. coli*, plasmid	8.5	1/24 000	1/204	Agar plate assay	[9]
Loam soil	Oxygenase	*E. coli*, plasmid	5.5	1/65 000	1/357.5	Agar plate assay	[10]
Mining shaft, compost soil, sediments (sewage plant, sea, lake, and river)	Protease	*E. coli*, plasmid	4.0	4/389 000	1/389	Agar plate assay	[11]
Gypsy moth midgut	Production of *N*-acylhomoserine lactons	*E. coli*, plasmid	3.3	1/800 000	1/2640	Reporter assay	[12]
Rhizosphere soil from near side of acid mine drainage	Nickel resistance	*E. coli*, plasmid	2.5	13/726 500	1/139.7	Growth assay	[13]
Glacial ice	DNA polymerase I	*E. coli*, plasmid	4	230/230 000	1/4	Growth assay	[14•]
Ground water from crude-oil storage cavity	Aromatic hydrocarbon catabolic operon	*E. coli*, plasmid	7	35/152 000	1/30.4	Reporter assay	[15]
Intertidal flat sediment	Lipase	*E. coli*, fosmid	N.A.[a]	1/6000		Agar plate assay	[16]
Deep-sea sediment	Lipase	*E. coli*, fosmid	32.3	1/8823	1/285	Agar plate assay	[17]
Activated sludge treating coke plant wastewater	Extradiol dioxygenase	*E. coli*, fosmid	33	91/96 000	1/34.8	Cell lysate assay	[18•]
Soil (pasture)	Degradation of *N*-acylhomoserine lactons	*E. coli*, fosmid	35	1/10 121	1/354.2	Reporter assay	[19]
Activated sludge treating coke plant wastewater	Bleomycin resistance gene	*E. coli*, fosmid	33	3/96 000	1/1056	Growth assay	[20]
Forest rhizosphere soils	Fungal antagonism	*E. coli*, fosmid	35	1/113 700	1/3979.5	Growth inhibition	[21]
Human fecal from healthy parson and patients of Crohn's disease	Epithelial cell growth modulation	*E. coli*, fosmid	43	20/20 725 (inhibition) 30/20 725 (stimulation)	1/44.6 (inhibition) 1/29.7 (stimulation)	Cell lysate assay	[22]
Glacial ice	DNA polymerase I	*E. coli*, fosmid	36	20/4000	1/7.2	Growth assay	[14•]
Surface seawater	Esterase	*E. coli*, BAC	70	4/20 000	1/350	Agar plate assay	[23]
Pig fecal	Tetracycline-resistance gene	*E. coli*, BAC	15	10/9000	1/13.5	Growth assay	[24]
Rabbit cecum	Cellulase	*E. coli*, cosmid	35.1	11/32 500	1/103.7	Agar plate assay	[25]
Soil (wetland and sandbars)	4-Hydroxyphenylpyruvate dioxygenase	*E. coli*, cosmid	40	5/30 000	1/240	Liquid-based assay	[26]
Cow rumen	Mannanase/glucanase/xylanase	*E. coli*, phagemid	3	1/50 000	1/150	Agar plate assay	[27]
Soil	Degradation of *N*-acylhomoserine lactons	*E. coli*, phagemid	4.3	3/7392	1/10.60	Reporter assay	[28]
Oil-contaminated soil	Naphthalene dioxygenase	*Pseudomonas putida*, cosmid	25	2/24 000	1/300	Growth assay	[29]

[a] N.A., not available.

used both plasmid and fosmid vectors to clone DNA polymerase I from the same metagenome, but the hit rates (number of positives per total number of clones screened) were not very different between the two procedures (230/230 000 for plasmids versus 20/4000 for fosmids). In order to double the chances of transcribing the cloned insert fragments, Lämmle *et al.* [7•] developed a plasmid vector with a dual-orientation promoter and succeeded in increasing the hit rate. The source of the metagenome also affects the hit rate. Natural or artificial contamination [8] is an effective way to enrich the environments with target genes.

Enzyme activities are usually assayed on agar plates supplemented with substrates. By cultivating a metagenomic library on the plates, one can identify positive clones through visual screening for the appearance of a clear zone (halo) or color. Screening requires no special devices and can be performed at high throughput, but the signals are often faint, which could be a reason for the common low hit rates.

To improve the sensitivity of agar plate-based screening, some alternative approaches are applied. One strategy is to use cell lysates for screening. By growing a library of cells in 96-well plates, followed by preparing lysates by either chemical [18•] or physical [22] procedures and mixing with substrates, the sensitivity can be improved dramatically. Even when using agar plates, if the target activities can be linked to the survivability of the host organisms, the screen (or selection) becomes highly sensitive and high throughput. This method is applied frequently to screening for resistance genes to toxic compounds, such as antibiotics [20,24] or heavy metals [13]. One can also apply this method to screen for essential genes using hosts that lack the target genes [14•,29]. A third approach is based on the reporter assay [12,15,19,28]. By linking target biological events to the expression of a reporter gene, screens can be conducted based solely on the expression of more sensitive reporter genes, such as green fluorescent protein [12,15,30], β-galactosidase [28,31,32], and tetracycline-resistance gene [32].

Problems with heterologous gene expression in *E. coli* and possible solutions

Since the metagenome is a complex mixture of genomes from a diverse array of microorganisms, the host's genetic machineries often cannot recognize transcriptional/translational signals in the metagenome and fail to decode certain genetic information therein. Currently, the extent of incidental transcription arising from natural regulatory sequences in cloned DNA remains largely unknown and whether we get a hit from the metagenome depends largely on luck.

Gabor *et al.* [33] presented a set of formulas that describe the chance of isolating a gene by random expression cloning. Thirty-two complete genome sequences of prokaryotic organisms were analyzed for the presence of expression signals functional in *E. coli* hosts using bioinformatics tools. They found significant differences in the predicted expression modes between distinct taxonomic groups of organisms and suggested that ~40% of the enzymatic activities may be readily recovered by random cloning in *E. coli*.

The translational initiation codon is biased toward AUG in *E. coli* (91%), but some organisms favor GUG and UUG as well. Villegas and Kropinski [34•] analyzed reported microbial genomes and found no tendency between initiation codon usage and the %[G + C] of the genomes. Taking the biased usage of AUG in *E. coli*, non-AUG initiation codons may not be recognized efficiently.

Warren *et al.* [35•] compared the transcription levels of foreign DNA in *E. coli*. They used programmable microarrays and RT-PCR to measure the transcription of *Haemophilus influenzae*, *Pseudomonas aeruginosa*, and human DNA propagating in *E. coli* as BACs. At least half of all *H. influenzae* genes were transcribed in *E. coli*. Highly transcribed genes frequently had promoter regions similar to *E. coli* s70 (RpoD)-binding sites. Compared to *H. influenzae*, a smaller proportion of *P. aeruginosa* genes were transcribed in *E. coli*, and punctuated transcription of human DNA was observed.

Synthetic genes can also be used as models to investigate the factors affecting heterologous gene expression in *E. coli*. Recently, Kudla *et al.* [36••] synthesized 154 variants of GFPs having altered (yet synonymous) codons over the entire sequence (226 of 240 amino acids). They expressed them in *E. coli* and compared the GFP expression levels, which varied by 250-fold across the library. They also compared the mRNA expression levels, mRNA degradation patterns, and bacterial growth rates. They found that the codon bias did not correlate with gene expression. Rather, the stability of mRNA folding near the ribosomal binding site accounted for more than half of the observed variation in protein levels. They concluded that mRNA folding and the associated rates of translation initiation played a predominant role in shaping the expression levels of individual genes, whereas codon bias influenced the global translation efficiency and cellular fitness.

All of these studies indicate clearly that the expression of metagenomes can be severely biased in *E. coli*. One possible solution for the global expression of foreign genes in *E. coli* may be to engineer the transcription/translation machinery of *E. coli*. Bernstein *et al.* [37•] evolved one of the ribosome proteins, S1 (coded by *rpsA*) and succeeded in obtaining variants able to recognize genes of *Rhodopseudomonas palustris*, which has a high [G + C] content, with 12 times greater efficiency. They concluded that the variant S1 protein became able to

Table 2

Sequence-based approach to screen enzymes in the period 2007 to early 2009

Environment	Target gene	Method	Number of retrieved clones	Identity to known enzyme	Reference
3-Chlorobenzoate enrichment	Benzoate 1,2-dioxygenase, chlorocatechol 1,2-dioxygenase	Degenerate PCR	2	74, 88%	[46]
Sediment from hot spring	Pullulanase	Degenerate PCR	1	48%	[47]
1,2-Dichloroethane enrichment	Reductive dehalogenase	Degenerate PCR	1	98%	[48]
Bioreactors treating gold-bearing concentrates	Sulfur oxygenase reductase	Degenerate PCR	2	48, 53%	[49]
Deep-sea sediment	Alkane hydroxylase	Degenerate PCR	2	55, 56%	[50]
4-Chlorobenzoate enrichment	4-Chlorobenzoyl-CoA dehalogenase	Degenerate PCR	2	77%	[51]
Marine sponge	Related polyketide synthesis	Degenerate PCR	3	50%	[52]
Marine sponge	Related polyketide synthesis	Degenerate PCR	5	43%	[53]
Deep-sea hydrothermal vents	Integron gene casette	Degenerate PCR	46	N.A.[a]	[57]
Tar pond	Integron gene casette	Degenerate PCR	708	N.A.[a]	[58]
Grassland soil	Nitrite reductase, nitrous oxide reductase	Probe hybridization	9	75–84%	[54]
Sequence database	Methyl halide transferase	*In silico* data mining	89	Average 28%	[59••]

[a] N.A., not available.

recognize the 5′-untranslated region including the ribosome-binding site. Several studies have reported altered translational profiles in *E. coli* through engineering ribosome proteins [38–40]. On the basis of an Actinomycetes platform, Ochi's research group reported an altered expression profile as revealed by the appearance of the novel antibiotic, piperidamycin, which was not produced in native strains [41•]. The piperidamycin-producing mutants all carried mutations in RNA polymerase and/or the ribosomal protein S12. In addition to the transcription/translation matter, the folding of the nascent polypeptide into the proper form is also critical for its activity. Therefore, the coexpression of molecular chaperones may also be effective in enhancing this final step [42].

Some researchers use hosts other than *E. coli*. More precisely, they use *E. coli* for library construction and a second host for functional screening. By using shuttle vectors, this approach is effective for retaining the size of libraries while complementing the low-level gene expression of the metagenome. Wang *et al.* [43] used an *E. coli–Sinorhizobium* shuttle vector and screened for clones that utilize D-3-hydroxybutyrate in a *bdhA*$^-$ mutant of *Sinorhizobium meliloti* and *atoC*$^-$ mutant of *E. coli*, both of which lacked the ability to utilize D-3-hydroxybutyrate. Li *et al.* [44] used an *E. coli–Rhizobium* shuttle vector and screened for tryptophan synthetic pathway genes using two different *trp*$^-$ mutants of *E. coli* and *Rhizobium leguminosarum*. In both studies, most of the clones selected using non-*E. coli* hosts failed to complement the deficiency of *E. coli* (and vice versa), indicating different expression profiles depending on the hosts and demonstrating the advantage of using a second host for functional screening. To this end, it is important to develop novel shuttle vectors that have extended host ranges. Ono *et al.* [29] used a cosmid, which contained both colE1 and RK2 replicons so that it can replicate in a range of Gram-positive and Gram-negative bacteria. Aakvik *et al.* [45] also used the RK2 replicon to modify fosmid and BAC vectors so that they were stably maintained in alternate hosts.

Mining for useful enzyme genes based on sequence homology

Recent sequence-based screenings of metagenomes are listed in Table 2. The common procedure is to use a set of degenerate oligonucleotide primers that are designed based on the consensus amino acid sequences [46–54]. Since this method only targets part of the internal gene sequence, one needs to clone the flanking regions to obtain the whole gene. Some methods have been developed for efficient gene walking [55,56]. If one uses a metagenome library as a template, tedious gene walking steps can be omitted [50–54].

In some cases, one can design primers outside the target gene sequence. For example, one of the mobile genetic elements, integron, carries gene cassettes. Genes contained in the cassettes are highly variable, but are commonly flanked by *attC* (or 59-be) sequences that are required for the translocation and recombination of integrons. By designing primers based on the *attC* sequences, it is possible to clone various genes included in the cassettes [57,58].

A most radical, yet straightforward approach has been reported recently by Voigt's group [59••]. Methyl halides are valuable industrial and agricultural chemicals, and they can also be used as precursors for the synthesis of more complex hydrocarbons. In an attempt to obtain such value-added methyl halides, they screened for methyl halide transferases in the public NCBI sequence database *in silico*. They collected 89 functionally uncharacterized

proteins possibly having the activities from various organisms (not restricted to microorganisms), of which only one gene was annotated as a methyl halide transferase and 55% were annotated as methyltransferases. Next, they synthesized all of the candidate genes with codons optimal for *E. coli* and *Saccharomyces cerevisiae* and expressed them in the hosts. Unexpectedly, 94% of the proteins showed methyl halide transferase activities with various halide (Cl, Br, I) specificities.

Future prospects

As has been described, recent activity-based screenings of metagenomes share a common problem because of the insufficient, biased expression in *E. coli*. Unlike the overexpression of a specific gene, the unbiased (or less biased) expression of foreign genes at sufficient levels requires a totally different approach. Ribosome engineering, as well as targeted engineering of factors related to transcription/translation, will provide possible solutions. As for the sequence-based approach, taking advantage of progress in synthetic biology, we will be able to take a more radical approach — to retrieve all possible candidate genes from the database, synthesize the genes, and test the activities. A vast and increasing volume of uncharacterized proteins exists in the public sequence databases that are rich, unexplored genetic resources, along with genes still unexplored in natural environments. Not a simple BLAST search, but rather structure-guided functional prediction, will be a key technology for obtaining novel enzymes.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Amann RI, Ludwig W, Schleifer KH: **Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation**. *Microbiol Rev* 1995, **59**:143-169.

2. Healy FG, Ray RM, Aldrich HC, Wilkie AC, Ingram LO, Shanmugam KT: **Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose**. *Appl Microbiol Biotechnol* 1995, **43**:667-674.

3. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM: **Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products**. *Chem Biol* 1998, **5**:245-249.

4. Beloqui A, de María PD, Golyshin PN, Ferrer M: **Recent trends in industrial microbiology**. *Curr Opin Microbiol* 2008, **11**:240-248.

5. Ferrer M, Beloqui A, Timmis KN, Golyshin PN: **Metagenomics for mining new genetic resources of microbial communities**. *J Mol Microbiol Biotechnol* 2009, **16**:109-123.

6. Steele HL, Jaeger KE, Daniel R, Streit WR: **Advances in recovery of novel biocatalysts from metagenomes**. *J Mol Microbiol Biotechnol* 2009, **16**:25-37.

7. • Lämmle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H, Rupp S: **Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning**. *J Biotechnol* 2007, **127**:575-592.

The authors used a plasmid vector with a dual-orientation promoter system in an attempt to double the chances of transcribing the insert fragments, and demonstrated the effect as revealed by an increase in the hit rate.

8. Mayumi D, Akutsu-Shigeno Y, Uchiyama H, Nomura N, Nakajima-Kambe T: **Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome**. *Appl Microbiol Biotechnol* 2008, **79**:743-750.

9. Hu Y, Zhang G, Li A, Chen J, Ma L: **Cloning and enzymatic characterization of a xylanase gene from a soil-metagenomic library with an efficient approach**. *Appl Microbiol Biotechnol* 2008, **80**:823-830.

10. van Hellemond EW, Janssen DB, Fraaije MW: **Discovery of a novel styrene monooxygenase originating from metagenome**. *Appl Environ Microbiol* 2007, **73**:5832-5839.

11. Waschkowitz T, Rockstroh S, Daniel R: **Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries**. *Appl Environ Microbiol* 2009, **75**:2506-2516.

12. Guan C, Ju J, Borlee BR, Williamson LL, Shen B, Raffa KF, Handelsman J: **Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota**. *Appl Environ Microbiol* 2007, **73**:3669-3676.

13. Mirete S, de Figueras CG, Gonzàlez-Pastor JE: **Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage**. *Appl Environ Microbiol* 2007, **73**:6001-6011.

14. • Simon C, Herath J, Rockstroh S, Daniel R: **Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice**. *Appl Environ Microbiol* 2009, **75**:2964-2968.

The authors amplified trace metagenomic DNA using the multiple displacement amplification (MDA) method and successfully retrieved DNA polymerases by the complementation of the *polA* mutant of *E. coli*. Despite their success, some clones contained repeated sequences derived from MDA artifacts, which is a common problem of *in vitro* amplification.

15. Uchiyama T, Abe T, Ikemura T, Watanabe K: **Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes**. *Nat Biotechnol* 2005, **23**:88-93.

16. Kim EY, Oh KH, Lee MH, Kang CH, Oh TK, Yoon JH: **Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases**. *Appl Environ Microbiol* 2009, **75**:257-260.

17. Jeon JH, Kim JT, Kim YJ, Kim HK, Lee HS, Kang SG, Kim SJ, Lee JH: **Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome**. *Appl Microbiol Biotechnol* 2009, **81**:865-874.

18. • Suenaga H, Ohnuki T, Miyazaki K: **Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds**. *Environ Microbiol* 2007, **9**:2289-2297.

Using cell lysate-based screening in 96-well plates instead of conventional screening on agar plates, the authors retrieved 91 catechol 2,3-dioxygenase genes that included numerous novel subfamily genes.

19. Riaz K, Elmerich C, Moreira D, Raffoux A, Dessaux Y, Faure D: **A metagenomic analysis of soil bacteria extends the diversity of quorum-quenching lactonases**. *Environ Microbiol* 2008, **10**:560-570.

20. Mori T, Mizuta S, Suenaga H, Miyazaki K: **Metagenomic screening for bleomycin resistance genes**. *Appl Environ Microbiol* 2008, **74**:6803-6805.

21. Chung EJ, Lim HK, Kim JC, Choi GJ, Park EJ, Lee MH, Chung YR, Lee SW: **Forest soil metagenome gene cluster involved in antifungal activity expression in *E. coli***. *Appl Environ Microbiol* 2008, **74**:723-730.

22. Gloux K, Leclerc M, Iliozer H, L'Haridon R, Manichanh C, Corthier G, Nalin R, Blottière HM, Doré J: **Development of high-throughput phenotyping of metagenomic clones from the human gut microbiome for modulation of eukaryotic cell growth**. *Appl Environ Microbiol* 2007, **73**:3734-3737.

23. Chu X, He H, Guo C, Sun B: **Identification of two novel esterases from a marine metagenomic library derived from South China Sea**. *Appl Microbiol Biotechnol* 2008, **80**:615-625.

24. Kazimierczak KA, Scott KP, Kelly D, Aminov RI: **Tetracycline resistome of the organic pig gut**. *Appl Environ Microbiol* 2009, **75**:1717-1722.

25. Feng Y, Duan CJ, Pang H, Mo XC, Wu CF, Yu Y, Hu YL, Wei J, Tang JL, Feng JX: **Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases**. *Appl Microbiol Biotechnol* 2007, **75**:319-328.

26. Lee CM, Yeo YS, Lee JH, Kim SJ, Kim JB, Han NS, Koo BS, Yoon SH: **Identification of a novel 4-hydroxyphenylpyruvate dioxygenase from the soil metagenome**. *Biochem Biophys Res Commun* 2008, **370**:322-326.

27. Palackal N, Lyon CS, Zaidi S, Luginbühl P, Dupree P, Goubet F, Macomber JL, Short JM, Hazlewood GP, Robertson DE, Steer BA: **A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut**. *Appl Microbiol Biotechnol* 2007, **74**:113-124.

28. Schipper C, Hornung C, Bijtenhoorn P, Quitschau M, Grond S, Streit WR: **Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa***. *Appl Environ Microbiol* 2009, **75**:224-233.

29. Ono A, Miyazaki R, Sota M, Ohtsubo Y, Nagata Y, Tsuda M: **Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches**. *Appl Microbiol Biotechnol* 2007, **74**:501-510.

30. Uchiyama T, Watanabe K: **Substrate-induced gene expression (SIGEX) screening of metagenome libraries**. *Nat Protoc* 2008, **3**:1202-1212.

31. Mohn WW, Garmendia J, Galvao TC, de Lorenzo V: **Surveying biotransformations with *à la carte* genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into *lacZ*-based phenotypes**. *Environ Microbiol* 2006, **8**:546-555.

32. van Sint Fiet S, van Beilen JB, Witholt B: **Selection of biocatalysts for chemical synthesis**. *Proc Natl Acad Sci U S A* 2006, **103**:1693-1698.

33. Gabor EM, Alkema WB, Janssen DB: **Quantifying the accessibility of the metagenome by random expression cloning techniques**. *Environ Microbiol* 2004, **6**:879-886.

34. • Villegas A, Kropinski AM: **An analysis of initiation codon utilization in the Domain Bacteria — concerns about the quality of bacterial genome annotation**. *Microbiology* 2008, **154**:2559-2561.
The authors analyzed 620 complete bacterial genomes and found no relationship between initiation codon (ATG, GTG and TTG) usage and the %[G + C] of the genomes.

35. • Warren RL, Freeman JD, Levesque RC, Smailus DE, Flibotte S, Holt RA: **Transcription of foreign DNA in *Escherichia coli***. *Genome Res* 2008, **18**:1798-1805.
This paper clearly demonstrates that the degree of expression of genes from phylogenetically diverse organisms in *E. coli* is lower than for organisms closely related to *E. coli*, indicating the importance of engineering *E. coli* to broaden the range of expressible genes.

36. •• Kudla G, Murray AW, Tollervey D, Plotkin JB: **Coding-sequence determinants of gene expression in *Escherichia coli***. *Science* 2009, **324**:255-258.
The authors compared the expression of 154 variants of GFPs with altered (yet synonymous) codons over the entire sequence and concluded that mRNA folding and the associated rates of translation initiation played a predominant role in shaping the expression levels of individual genes, whereas codon bias influenced the global translation efficiency and cellular fitness.

37. • Bernstein JR, Bulter T, Shen CR, Liao JC: **Directed evolution of ribosomal protein S1 for enhanced translational efficiency of high GC *Rhodopseudomonas palustris* DNA in *Escherichia coli***. *J Biol Chem* 2007, **282**:18929-18936.
This paper outlines the future direction of *E. coli* engineering to broaden the range of expressible genes.

38. O'Connor M, Gregory ST, Dahlberg AE: **Multiple defects in translation associated with altered ribosomal protein L4**. *Nucleic Acids Res* 2004, **32**:5750-5756.

39. Das G, Dineshkumar TK, Thanedar S, Varshney U: **Acquisition of a stable mutation in *metY* allows efficient initiation from an amber codon in *Escherichia coli***. *Microbiology* 2005, **151**:1741-1750.

40. Maar D, Liveris D, Sussman JK, Ringquist S, Moll I, Heredia N, Kil A, Bläsi U, Schwartz I, Simons RW: **A single mutation in the IF3 N-terminal domain perturbs the fidelity of translation initiation at three levels**. *J Mol Biol* 2008, **383**:937-944.

41. • Hosaka T, Ohnishi-Kameyama M, Muramatsu H, Murakami K, Tsurumi Y, Kodani S, Yoshida M, Fujie A, Ochi K: **Antibacterial discovery in Actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12**. *Nat Biotechnol* 2009, **27**:462-464.
This paper reports on the possibility of engineering the transcriptional/translational machinery to alter the expression profile.

42. Ferrer M, Chernikova TN, Timmis KN, Golyshin PN: **Expression of a temperature-sensitive esterase in a novel chaperone-based *Escherichia coli* strain**. *Appl Environ Microbiol* 2004, **70**:4499-4504.

43. Wang C, Meek DJ, Panchal P, Boruvka N, Archibald FS, Driscoll BT, Charles TC: **Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants**. *Appl Environ Microbiol* 2006, **72**:384-391.

44. Li Y, Wexler M, Richardson DJ, Bond PL, Johnston AWB: **Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes**. *Environ Microbiol* 2005, **7**:1927-1936.

45. Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L, Völker U, Ellingsen TE, Valla S: **A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species**. *FEMS Microbiol Lett* 2009, **296**:149-158.

46. Morimoto S, Fujii T: **A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking**. *Appl Microbiol Biotechnol* 2009, **83**:389-396.

47. Tang K, Kobayashi RS, Champreda V, Eurwilaichitr L, Tanapongpipat S: **Isolation and characterization of a novel thermostable neopullulanase-like enzyme from a hot spring in Thailand**. *Biosci Biotechnol Biochem* 2008, **72**:1448-1456.

48. Marzorati M, de Ferra F, Van Raemdonck H, Borin S, Allifranchini E, Carpani G, Serbolisca L, Verstraete W, Boon N, Daffonchio D: **A novel reductive dehalogenase, identified in a contaminated groundwater enrichment culture and in *Desulfitobacterium dichloroeliminans* strain DCA1, is linked to dehalogenation of 1,2-dichloroethane**. *Appl Environ Microbiol* 2007, **73**:2990-2999.

49. Chen ZW, Liu YY, Wu JF, She Q, Jiang CY, Liu SJ: **Novel bacterial sulfur oxygenase reductases from bioreactors treating gold-bearing concentrates**. *Appl Microbiol Biotechnol* 2007, **74**:688-698.

50. Xu M, Xiao X, Wang F: **Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment**. *Extremophiles* 2008, **12**:255-262.

51. Chae JC, Song B, Zylstra GJ: **Identification of genes coding for hydrolytic dehalogenation in the metagenome derived from a denitrifying 4-chlorobenzoate degrading consortium**. *FEMS Microbiol Lett* 2008, **281**:203-209.

52. Fieseler L, Hentschel U, Grozdanov L, Schirmer A, Wen G, Platzer M, Hrvatin S, Butzke D, Zimmermann K, Piel J: **Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges**. *Appl Environ Microbiol* 2007, **73**:2144-2155.

53. Fisch KM, Gurgui C, Heycke N, van der Sar SA, Anderson SA, Webb VL, Taudien S, Platzer M, Rubio BK, Robinson SJ *et al.*:

Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat Chem Biol* 2009, **5**:494-501.

54. Demanèche S, Philippot L, David MM, Navarro E, Vogel TM, Simonet P: **Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach**. *Appl Environ Microbiol* 2009, **75**:534-537.

55. Uchiyama T, Watanabe K: **Improved inverse PCR scheme for metagenome walking**. *BioTechniques* 2006, **41**:183-188.

56. Yamada K, Terahara T, Kurata S, Yokomaku T, Tsuneda S, Harayama S: **Retrieval of entire genes from environmental DNA by inverse PCR with pre-amplification of target genes using primers containing locked nucleic acids**. *Environ Microbiol* 2008, **10**:978-987.

57. Elsied H, Stokes HW, Nakamura T, Kitamura K, Fuse H, Maruyama A: **Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents**. *Environ Microbiol* 2007, **9**:2298-2312.

58. Koenig JE, Sharp C, Dlutek M, Curtis B, Joss M, Boucher Y, Doolittle WF: **Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the Sydney tar ponds**. *PLoS ONE* 2009, **4**:e5276.

59. •• Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA: **Synthesis of methyl halides from biomass using engineered microbes**. *J Am Chem Soc* 2009, **131**:6508-6515.

Through *in silico* screening of public sequence databases and chemical synthesis of the retrieved gene sequences, one can instantly generate gene libraries, including those that are functionally uncharacterized.