

Full Length Research Paper

## Rapid approach for cloning bacterial single-genes directly from soils

Chenguang Zhu\*, Peipei Wang, Chunlei Shen, Wei Wang, Yuanping Tang, Bing Mei, Zhengkai Xu and Rentao Song

Shanghai Key Laboratory of Bio-energy Crops, School of Life Sciences, Shanghai University, Shanghai 200444, P. R. China.

Accepted 26 July, 2013

Obtaining functional genes of bacteria from environmental samples usually depends on library-based approach which is not favored as its large amount of work with small possibility of positive clones. A kind of bacterial single-gene encoding glutamine synthetase (GS) was selected as example to detect the efficiency of cloning strategy in this study. Five GS genes were directly cloned from soils using degenerate primers with two steps of nested polymerase chains reactions. The genes showed 94 to 99% amino acid identities to the homologs in the known database, and encoded proteins affiliated to GS I and GS II families, respectively. All the five genes could rescue the growth of *Escherichia coli* glutamine auxotroph mutant ET6017 in minimum medium (ammonium chloride was sole nitrogen source in this medium). This study develops one rapid approach for cloning bacterial single-genes directly from soils. Comparing with the conventional strategies for gene cloning from complex environmental samples, this method did not need making genomic library and isolating target genes from large amount of library clones. This approach distinctively demonstrates its advantages of rapidity and effectiveness particularly when it aims at cloning short single-genes that had known homologs in all kinds of nucleic acid databases.

**Key words:** Gene cloning, soil, glutamine synthetase, nested PCR, single-gene.

### INTRODUCTION

The fast development of modern biological science brings the urgent needs of cloning functional genes from various prokaryotic organisms in complex environments, as prokaryotes (mainly bacteria) exist in every part of our planet and demonstrate the huge diversity of gene resources that has far not yet to be developed and utilized by human societies (Cowan, 2000). Metagenome represents mixed genomes of multiple microorganisms from an environmental sample, and is a large gene resource that has been paid a lot of attention in recent years (Handelsman et al., 1998; Lorenz et al., 2002; Rajendhran and Gunasekaran, 2008; Cretoiu et al., 2012;

Ekkers et al., 2012). Soil is one of the complex environments, which appears to be a major habitat of microbes. Many works have reported that novel genes and proteins are discovered from the soil metagenomes by library-based enzyme activity screening or nucleotide sequence screening (Henne et al., 1999; Majernik et al., 2001; Brady et al., 2002; Voget et al., 2003; Daniel et al., 2004; Riesenfeld et al., 2004; Treusch et al., 2004; Tringe et al., 2005). But for cloning the bacterial single-genes which have known homologs in the public databases from soils, the library-based approach is not favored because of its large amount of work with small probability of positive

\*Corresponding author. E-mail: [cgzhu@shu.edu.cn](mailto:cgzhu@shu.edu.cn). Tel: 86-21-66135186-806. Fax: 86-21-66133225.

clones. Therefore, the more convenient approach for cloning such genes from environmental samples is keen to be developed.

Glutamine synthetase (GS), which produces glutamine from glutamate and ammonia, is the key enzyme in the system of regulating nitrogen assimilation of all the microbes (Shapiro and Stadtman, 1970). The reported coding region of GS genes from various bacteria is usually 1 to 2.3 kb (Patriarca et al., 1992; Kumada et al., 1993). The widely existence and suitable open reading frame (ORF) fragment length make the bacterial GS gene appropriate example for detecting the efficiency of various cloning methods. In this study, a convenient and effective approach to clone bacterial single-genes directly from the soils was developed, using GS gene as the target gene. This strategy was free of performing metagenomic library and screening target genes from the library. The advantage and potentially applicative field of this approach were also discussed in this work.

## MATERIALS AND METHODS

### Soil samples and bacterial strains

Six samples of soils were collected at different places (grassland, roots of tree, riverside, etc.) in the campus of Shanghai University. After collection, the soil samples were immediately performed with soil DNA extraction kit to isolate the metagenomic DNAs from these samples. *Escherichia coli* Top10 (stored in this laboratory) and ET6017 (provided by *E. coli* Genetic Stock Center, Yale University, USA) were respectively used for conventional DNA manipulation and functional analysis of cloned GS genes. *E. coli* was grown in Luria-Bertani (LB) medium or M9 medium (Sambrook and Russell, 2001) at 37°C. Ampicillin was used to screen *E. coli* transformants at concentration of 100 µg ml<sup>-1</sup>.

### Conventional genetic techniques

Extraction of bacterial plasmid DNA, restriction enzyme (Fermentas Life Sciences) digestion, DNA ligation, and transformation of *E. coli* cells were performed as described in the standard procedures (Sambrook and Russell, 2001). Soil metagenomic DNAs were extracted using the E.Z.N.A.<sup>TM</sup> (easy nucleic acid isolation) soil DNA kit (Omega Bio-tek, Inc.) according to the protocol provided by the kit manufacturer.

### Cloning the bacterial GS genes

The degenerate primers specific for GS genes from *Bacillus cereus*, *Rhizobium* spp. and *Pseudomonas* spp. (these bacteria are known to be widely distributed in various soil samples) were designed to amplify the positive gene fragments. For designing the primer pairs of GS genes from each bacterial species, the most abundant DNA sequences of known homologous GS genes from each bacterial species were chosen to do conservative analysis. The degenerate primers were designed based on the conservative analysis of these genes. The sequences of diverse primer pairs were listed as follows: For amplification of GS genes from *Pseudomonas* spp.:

Ps4-eIF: 5'-GCATCACCCAAATTC AAGGG-3'  
Ps4-F: 5'-ATGTCGAAGTCGGTTCAACT-3'

Ps4-R: 5'-TCAGCAGCTGTAGTAMAGCT-3'

For amplification of GS genes from *B. cereus*:

Bc1-eIF: 5'-ACTGATTCTGAAGGTGTTTA-3'  
Bc1-F: 5'-ATGGCTAGGTACACAAAAGA-3'  
Bc1-R: 5'-TTAGTAHAGAGACATATATT-3'

For amplification of GS genes from *Rhizobium* spp.:

Rz1-eIF: 5'-CGCAAAGTCCGWTACCGTCA-3'  
Rz1-F: 5'-ATGGCGACCGCAAGCGAAAT-3'  
Rz1-R: 5'-TCAGGCCGAATARTACATGT-3'  
Rz2-eIF: 5'-CTCCGGRTTGTCTACCTTC-3'  
Rz2-F: 5'-ATGACAAAATATAAGCTCGA-3'  
Rz2-R: 5'-TYARGCKGCAGCGGAAGCCG-3'

The nested PCR protocol was applied in this study. This protocol includes two steps of PCR amplification procedures. The primer pairs titled "eIF" and "R" were used in the first amplification procedure to amplify the whole ORF and part region (40 to 65 bps) upstream the initial codon of the ORF of the various GS genes. The amplified products were purified by AxyPrep PCR purification Kit (Axygen Biosciences) and were used as the template DNA for the second PCR amplification. The primer pairs titled "F" and "R" were used in the second amplification procedure to amplify the intact ORF of the various GS genes. The Ex Taq DNA polymerase (TaKaRa Co., Ltd) was used for the whole PCR procedures. The following amplification condition was applied as reference: 5 min at 94°C, 30 cycles of 50 s at 94°C, 55 s at 50°C, 90 s at 72°C, finally, 10 min at 72°C. The positive DNA fragments amplified from the second PCR procedure were purified by agarose gel purification using AxyPrep Gel Purification Kit (Axygen Biosciences), then were ligated with pMD18-T vector (TaKaRa Co., Ltd) to obtain the recombinant plasmids.

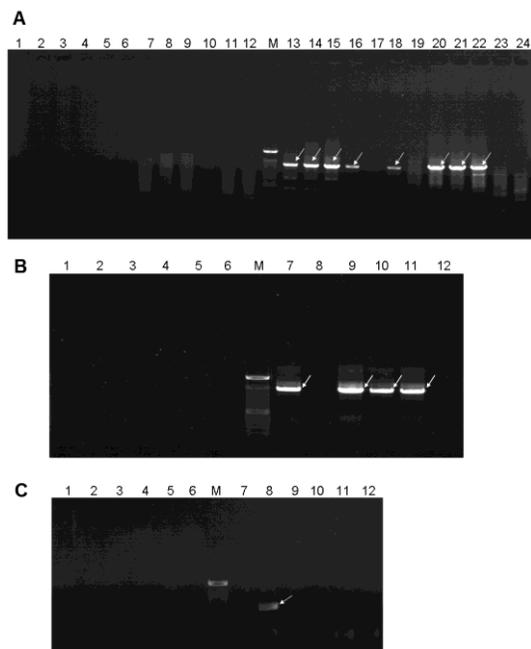
The positive clones were picked out by PCR detection using primer pairs titled "F" and "R". The nucleotide sequences of the positive clones were determined by double direction sequencing using M13 primer pair (forward primer 5'-CAGGAAACAGCTATGAC-3' and reverse primer 5'-GTTTCCCAGTCACGAC-3') in a MegaBACE 4500 DNA Analysis System (Amersham Biosciences).

### Functional complement of GS genes in *E. coli* GS mutant ET6017

The *E. coli* glutamine auxotroph mutant ET6017 [*araD139*]<sub>Br</sub> Δ(*argF-lac*)169 *flhD5301* Δ(*fruK-yeiR*)725(*fruA25*) *relA1* *rpsL150(strR)* Δ(*glnG-glnA*)229 *rha-10* *deoC1*] (provided by *E. coli* Genetic Stock Center, Yale University, USA) was used for the GS genes complement study. For expression in *E. coli*, the positive clones in which the inserted GS gene has the identical transcriptional direction with the *lac* promoter located at the pMD18-T, were picked out and sequenced to confirm their correctness of gene sequences. Then, the plasmids of these positive clones were isolated and transformed into *E. coli* strain ET6017 by electroporation. The recombinant *E. coli* clones were grown in M9 minimal medium as described by Sambrook and Russell (2001), supplemented with 100 µg ml<sup>-1</sup> of ampicillin. The OD<sub>600</sub> values of recombinant ET6017 cells were monitored by Cary 50 UV-visible spectrophotometer (Varian Inc., USA) to quantitatively analyze their growth status.

### Phylogenetic analysis of the cloned GSs

Sequences were compared with NCBI GenBank entries (<http://www.ncbi.nlm.nih.gov/>) using the protein-protein BLAST. For



**Figure 1.** The agarose gel electrophoresis of nested PCR products from six samples of soils. (A) Lanes 1 to 6, the first PCR product amplified by primer pairs Bc1-eIF and Bc1-R; lanes 7 to 12, the first PCR product amplified by primer pairs Rz1-eIF and Rz1-R; lanes 13 to 18, the second PCR product amplified by primer pairs Bc1-F and Bc1-R; lanes 19 to 24, the second PCR product amplified by primer pairs Rz1-F and Rz1-R; M, 100 bp DNA ladder. (B) Lanes 1 to 6, the first PCR product amplified by primer pairs Ps4-eIF and Ps4-R; lanes 7 to 12, the second PCR product amplified by primer pairs Ps4-F and Ps4-R; M, 100 bp DNA Ladder. (C) lanes 1 to 6, the first PCR product amplified by primer pairs Rz2-eIF and Rz2-R; lanes 7 to 12, the second PCR product amplified by primer pairs Rz2-F and Rz2-R; M, 100 bp DNA Ladder. The white arrows point the specific GS gene fragments.

sequence alignment of GSs, the software Clustal X 1.81 (Thompson et al., 1997) was used with the following alignment parameters: gap opening penalty 10, gap extension penalty 0.2, delay divergent sequences 30%, DNA transition weight 0.5, protein weight matrix Gonnet series and DNA weight matrix IUB. Phylogenetic tree was constructed by the neighbour-joining method of the MEGA 4 program (Tamura et al., 2007) using the bootstrap test based on 1000 replicates.

#### Nucleotide sequence accession numbers

The nucleotide sequences of GS genes cloned from soil samples were deposited in the GenBank data library under accession number JX017366 ~ JX017370.

## RESULTS AND DISCUSSION

### Optimization of PCR procedures, cloning and sequence analysis of GS genes

At the beginning of this study, we tried to use the primer

pairs ‘-F’ and ‘-R’ of various GS genes to rapidly clone the ORF of these genes from soils, but the results showed that it was hard to obtain the clear and optimal PCR fragments directly from soil DNAs (data not shown). This was probably due to the existence of humic compounds in the soil DNA templates together with the low abundance of target bacterial genome templates among the soil metagenomes, which resulted in the poor amplification of target genes. As PCR was a kind of sensitive enzymatic reaction, the purity and specialty of template DNA seemed to be very important. Therefore, we designed a series of nested primers termed as ‘-eIF’. Utilization of primer pairs ‘-eIF’ and ‘-R’ could ensure that the integrate ORF of various GS genes had been amplified from the soil, although the concentration of the amplified specific DNA fragments was too low to directly do the subcloning experiments. The target ORF of the GS genes was obviously amplified in the second PCR reaction (Figure 1) because after purification of the amplified products of the first PCR reaction, the contaminants including humic compounds, nucleases and soil genomes had been almost removed, and subsequently, the amplified DNA fragments from the first step of PCR became to be the predominant template DNA in the reactionary system of the second PCR. This could be the major reason why the clear and abundant DNA fragments were abundantly amplified in the second PCR process.

The amplified fragments in the second PCR procedure were purified and sequenced. Altogether five GS genes were cloned by nested PCR. These genes were respectively termed as *grassB4*, *raadR6*, *grassP1*, *riverP6* and *grassRz2*. Amino acid homology analysis revealed that the *grassB4*, *raadR6*, *grassP1* and *riverP6* encoded proteins with highest amino acid identities respectively to the glutamine synthetases type I from *Bacillus cereus* Rock4-18 (accession no. EEL59774, 98% identity), *Rhizobium leguminosarum* bv. *viciae* 3841 (accession no. CAK07882, 99% identity), *Pseudomonas fluorescens* SBW25 (accession no. CAY46625, 96% identity) and *Pseudomonas putida* GB-1 (accession no. ABZ00983, 94% identity). The *grassRz2* encoded protein with highest amino acid identity to the glutamine synthetase type II from *Rhizobium etli* CFN 42 (accession no. ABC91870, 99% identity) (Table 1). These results indicate that all five genes cloned in this study belonged to bacterial GS gene families.

### Phylogenetic analysis of cloned GS proteins

After homology analysis, phylogenetic tree of diverse GS proteins from bacteria, green alga, higher plants and animals, together with five GS proteins obtained in this study was performed to investigate the evolutionary relationship of theirs. The result indicate that *GrassB4*, *RoadR6*, *GrassP1* and *RiverP6* were closely affiliated to the GS I family which was widely distributed among bacteria (Figure 2). The *GrassRz2* was affiliated to the GS

**Table 1.** Protein homologs of five GS genes cloned from the soil metagenomes.

Gene	ORF length (bps)	Protein length	Function	% protein identity (number of residues)	Organism	Accession number
<i>grassB4</i>	1335	444	Glutamine synthetase	98(436)	<i>Bacillus cereus</i> Rock4-18	EEL59774
				98(435)	<i>Bacillus cereus</i> Rock1-3	EEL21814
				98(435)	<i>Bacillus cereus</i> Rock3-28	EEL33362
<i>raodR6</i>	1410	469	Glutamine synthetase	99(465)	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	CAK07882
				99(465)	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM1325	ACS56202
				99(464)	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2304	ACI55023
<i>grassP1</i>	1407	468	Glutamine synthetase	96(451)	<i>Pseudomonas fluorescens</i> SBW25	CAY46625
				96(450)	<i>Pseudomonas fluorescens</i> WH6	EFQ65563
				95(445)	<i>Pseudomonas brassicacearum</i> subsp. <i>brassicacearum</i> NFM421	AEA66503
<i>riverP6</i>	1407	468	Glutamine synthetase	94(440)	<i>Pseudomonas putida</i> GB-1	ABZ00983
				94(440)	<i>Pseudomonas</i> sp. TJI-51	EGB97092
				94(438)	<i>Pseudomonas fluorescens</i> Pf-5	AA95795
<i>grassRz2</i>	1041	346	Glutamine synthetase	99(342)	<i>Rhizobium etli</i> CFN 42	ABC91870
				98(340)	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2304	ACI56112
				98(340)	<i>Rhizobium etli</i> CNPAF512	EGE56659

II family that is widely distributed among various eukaryotes and several soil bacteria. All the five GS proteins cloned in this study were obviously different with the GS III proteins judging from the evolutionary distances in the tree (Figure 2).

#### Functional complement of five GS genes in *E. coli* GS mutant ET6017

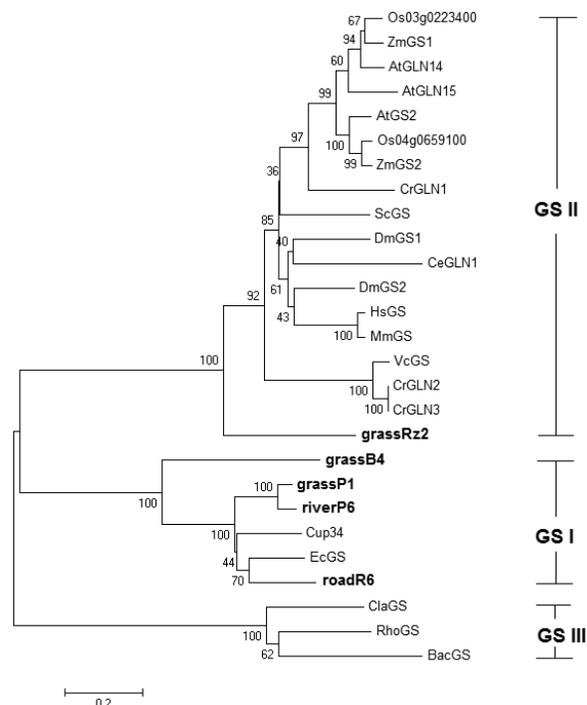
Strain ET6017 lacked the ability of synthesizing glutamine via ammonium. When this strain was inoculated into the M9 medium which contained the ammonium chloride as the sole nitrogen source, it could not grow. In this study, in order to confirm the exact function of five GS genes, the *E.*

*coli* GS mutant ET6017 was chosen to be the receptor strain. After the GS genes were respectively transformed into strain ET6017, the recombinant cells all grew well in the M9 medium according to their OD<sub>600</sub> values, however, the ET6017 cells bearing the pMD18-T vector failed to grow continuously in the M9 medium (Figure 3). This result confirms that all the GS genes could be expressed in the ET6017 cells and demonstrated the enzymatic functions of glutamine synthetase.

#### Significance and applicative potential of this approach

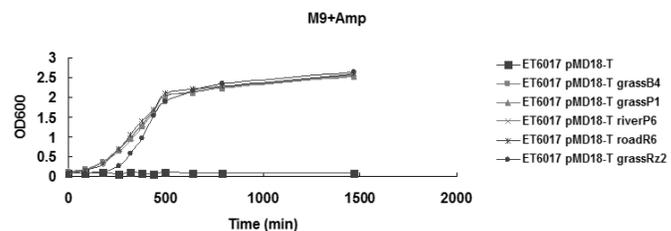
Soil is a complex environment which is a major

resource of microbial diversity (Gans et al., 2005). Current opinion indicates that more than 99% of the microbes present in many natural environments such as soil are not culturable. Therefore, analysis of nucleic acids directly extracted from environmental samples facilitates the process of mining unknown genes (proteins) without the need for microbial isolation and cultivation (Schloss and Handelsman, 2003). Presently, the conventional approach of cloning genes from soil metagenomes is to construct the clone library (using vectors such as BAC, cosmid, etc.) and isolate genes from the library (Rondon et al., 2000; Rajendhran and Gunasekaran, 2008). However, the existence of contaminants including



**Figure 2.** Phylogenetic relationship between the five cloned GS proteins and partial homologs of known organisms. The scale bar represents the substitution frequencies per amino acid residue. The values of bootstrap for the clade in 1000 replicates are indicated by numbers at nodes. Os04g0659100 and Os03g0223400, GS from *Oryza sativa Japonica* (NP\_001054133 and NP\_001049424); ZmGS1 and ZmGS2, GS from *Zea mays* (BAA03433, NP\_001105725); AtGLN14, AtGLN15 and AtGS2, GS from *Arabidopsis thaliana* (NP\_568335, NP\_175280, NP\_198413); CrGLN1, CrGLN2 and CrGLN3, GS from *Chlamydomonas reinhardtii* (EDP07598, Q42689, XP\_001692927); DmGS1 and DmGS2, GS from *Drosophila melanogaster* (NP\_722606, NP\_727525); EcGS, GS from *Escherichia coli* str. K-12 substr. MG1655 (NP\_418306); HsGS, GS from *Homo sapiens* (NP\_001028216); ScGS, GS from *Saccharomyces cerevisiae* S288c (NP\_015360); MmGS, GS from *Mus musculus* (NP\_032157); CeGLN1, GS from *Caenorhabditis elegans* (NP\_509012); VcGS, GS from *Volvox carteri* f. *nagariensis* (XP\_002947417); Cup34, GS from *Cupriavidus metallidurans* CH34 (ABF08941); ClaGS, GS from *Clostridium acetobutylicum* ATCC824 (AAK80605); RhoGS, GS from *Rhodococcus erythropolis* PR4 (BAH34552); BacGS, GS from *Bacteroides fragilis* YCH46 (BAD47790).

humic compounds, nucleases, together with the difficulty of screening functional genes from the library, makes the approach based on genomic library analysis hard work and low efficient. In this study, a strategy for cloning bacterial single-genes directly from soil samples was developed without microbial isolation and cultivation. Based on this method, the target genes could be easily obtained to meet the subsequent needs for functional research. The practical premise of this method was to



**Figure 3.** The growth curve of the pMD18-T based recombinant plasmids in *Escherichia coli* ET6017. The mutant ET6017 was transformed with pMD18-T, pMD18-T grassB4, pMD18-T grassP1, pMD18-T riverP6, pMD18-T roadR6, pMD18-T grassRz2. The transformants were cultivated in liquid M9 medium (100 µg ml<sup>-1</sup> of ampicillin was supplemented). The growth curve was measured in the 1500 min.

know which species of microbes living in the target soil sample and the homologs of target genes could be exactly found in the known gene database to conduct the sequence comparison and design the degenerate primers.

Different with the genomic library approach which was usually used for mining unknown genes and studying the function of large gene clusters, our method was favorable to be utilized in cloning short single-genes (no longer than 2 kbs) having known information of their homologs. Compared with the library screening approach, this method demonstrated the advantages of rapidity, effectiveness and convenience.

Compared with the conventional clone strategy from the bacterial strains, this protocol did not need to isolate the bacterial strains. It was postulated that this approach could attribute to provide rich bacterial genes for many genetic engineering improvements. Moreover, the primers designed in this work were also inferred to be used to clone the same GS genes not only from soils, but also from other complex environmental samples containing identical bacterial species.

## ACKNOWLEDGEMENTS

This work was supported by the grants from the National Natural Sciences Foundation of China (30871278, 30970242), the Ministry of Agriculture of China (2013ZX08003005-002), and Shanghai Municipal Science and Technology Committee (10DZ2271800).

## REFERENCES

- Brady SF, Chao CJ, Clardy J (2002). New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* 124:9968-9969.
- Cowan DA (2000). Microbial genomes-the untapped resource. *Trends Biotechnol.* 18:14-16.
- Cretoiu MS, Kielak AM, Al-Soud WA, Sørensen SJ, van Elsas JD (2012). Mining of unexplored habitats for novel chitinases-*chiA* as a helper gene proxy in metagenomics. *Appl. Microbiol. Biotechnol.* 94:1347-1358.

- Daniel R (2004). The soil metagenome - a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* 15:199-204.
- Ekkers DM, Cretoiu MS, Kielak AM, Elsas JD (2012). The great screen anomaly - a new frontier in product discovery through functional metagenomics. *Appl. Microbiol. Biotechnol.* 93:1005-1020.
- Gans J, Wolinsky M, Dunbar J (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Sci.* 309:1387-1390.
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5: R245-R249.
- Henne A, Daniel R, Schmitz RA, Gottschalk G (1999). Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* 65: 3901-3907.
- Kumada Y, Benson DR, Hillemann D, Hosted TJ, Rochefort DA, Thompson CJ, Wohlleben W, Tateno Y (1993). Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proc. Natl. Acad. Sci. USA.* 90:3009-3013.
- Lorenz P, Liebeton K, Niehaus F, Eck J (2002). Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr. Opin. Biotechnol.* 13:572-577.
- Majernik A, Gottschalk G, Daniel R (2001). Screening of environmental DNA libraries for the presence of genes conferring Na<sup>+</sup>(Li<sup>+</sup>)/H<sup>+</sup> antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J. Bacteriol.* 183:6645-6653.
- Patriarca EJ, Chiurazzi M, Manco G, Riccio A, Lamberti A, De Paolis A, Rossi M, Defez R, Iaccarino M (1992). Activation of the *Rhizobium leguminosarum glnII* gene by NtrC is dependent on upstream DNA sequences. *Mol. Gen. Genet.* 234:337-345.
- Rajendhran J, Gunasekaran P (2008). Strategies for accessing soil metagenome for desired applications. *Biotechnol. Adv.* 26: 576-590.
- Riesenfeld CS, Goodman RM, Handelsman J (2004). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* 6:981-989.
- Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66:2541-2547.
- Sambrook J, Russell JW (2001). *Molecular cloning, a laboratory manual.* Cold Spring Harbor Laboratory Press, New York, USA.
- Schloss PD, Handelsman J (2003). Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* 14:303-310.
- Shapiro BM, Stadtman ER (1970). The regulation of glutamine synthesis in microorganisms. *Annu. Rev. Microbiol.* 24:501-524.
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596-1599.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876-4882.
- Treusch AH, Kletzin A, Raddatz G, Ochsenreiter T, Quaiser A, Meurer G, Schuster SC, Schleper C (2004). Characterization of large-insert DNA libraries from soil for environmental genomic studies of archaea. *Environ. Microbiol.* 6:970-980.
- Tringe SG, Von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005). Comparative metagenomics of microbial communities. *Sci.* 308:554-557.
- Voget S, Leggewie C, Uesbeck A, Raasch C, Jaeger KE, Streit WR (2003). Prospecting for novel biocatalysts in a soil metagenome. *Appl. Environ. Microbiol.* 69:6235-6242.