ELSEVIER

Journal of Molecular Catalysis B: Enzymatic 19–20 (2002) 13–19

JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC

www.elsevier.com/locate/molcatb

Metagenome—a challenging source of enzyme discovery

Patrick Lorenz [a,*], Christa Schleper [b,c]

[a] *B.R.A.I.N Aktiengesellschaft, Darmstaedterstrasse 34, 64673 Zwingenberg, Germany*
[b] *Technische Universität Darmstadt, Institut für Mikrobiologie und Genetik, Schnittspahnstrasse 10, 64287 Darmstadt, Germany*
[c] *Zentrum für molekulare Evolution und Biodiversität (ZEB), Schnittspahnstrasse 10, 64287 Darmstadt, Germany*

Received 8 November 2001; received in revised form 10 April 2002; accepted 24 April 2002

Abstract

Evolution over very long periods of time has generated a vast pool of physiologies and molecular adaptations to enable microorganisms to thrive in a wide range of living conditions. These solutions may be valuable tools in the rational design of technical processes. However, molecular ecological studies have proven that the vast majority of microorganisms cannot be cultivated in the laboratory and was therefore ignored when traditional screening procedures were applied to search for novel enzymes and bioactives.

Cultivation-independent approaches now open up the roads to analyze and screen the genetically and metabolically rich microbial communities in their entirety.

In these approaches DNA is directly isolated from environmental samples and cloned into suitable vectors to construct complex genomic libraries. These libraries can be analyzed for novel genes and pathways with sequence-based techniques or through screening proteins and drugs that are being produced in surrogate hosts.

The approach of directly cloning environmental DNA greatly enhances the opportunities to take full advantage of the enormous naturally occurring microbial resources.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Metagenome; Uncultivated biodiversity; Microorganism; Enzyme; Biocatalysis

1. Microbes as the largest natural source of molecular diversity

Life on earth is qualitatively shaped by microorganisms invisible to the naked eye. It is assumed that the earliest traceable forms of cellular life likely were autotrophic bacteria gaining energy from oxidizing inorganic substrates (H_2, H_2S or Fe^{2+}) and reducing electron acceptors like NO_3^-, SO_4^{2-}, S or CO_2. At a later stage such lithotrophs were complemented by phototrophic bacteria utilizing light to fix carbondioxide in the production of energy-rich organic material. This in turn fueled organotrophic food chains starting with bacteria and ending with eukaryotic organisms up to macroscopic animals [1].

By recycling carbon, phosphorous and nitrogen from dead organic matter, microbes form an important link in global biogeochemical cycles. Their versatility is also well reflected in their adaptations to extreme environments. Especially representatives of the archaea demonstrate the extreme physical limits of extant life, with organisms growing as high as 113 °C, at pH values around 0 or in saturated salt solutions.

The evolutionary headstart of bacterial life forms 3.5 billion years ago with exposure to an extreme range of selective pressures in different habitats has

* Corresponding author. Tel.: +49-6251-93310; fax: +49-6251-933111.
E-mail address: pl@brain-biotech.de (P. Lorenz).

1381-1177/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S1381-1177(02)00147-9

produced their unparalleled physiological and biochemical diversity. The enormous potential to generate mutations in the microbial world was illustrated by Whitman who calculated that with an estimated annual 10^{30} microbial cell divisions of the marine heterotrophic bacteria alone and an assumed mutation rate of 4×10^{-7} per gene $\times$ per division it takes only in the order of 1 h to accumulate 4 simultaneous mutations in a single shared gene [2]. With this historical perspective and proliferative capacity in mind it may not seem too daring then to predict that microorganisms have developed enzymes to deal with any conceivable biochemical entity showing at least a minimum resemblance of naturally available substrates.

Anticipating the anabolic and catabolic capacities of microorganisms it is obvious that human ingeniuity was tempted to instrumentalize microbes and their enzymes in technical processes. Human usage of microorganisms and microbial enzymes in the processing of natural materials has a long tradition, considering historical brewing, tanning or baking processes. This early exploitation of microbial biosynthetic capacity, however, was unwitting and facilitated by the ubiquity of microbes. Consequently nowadays enzymes are used as key active components in detergents, in various food/feed processes and the pulp and paper manufacturing just to mention a selection (see http://www.novozymes.com). The preference of synthetic organic chemists for enzyme catalysts over classical chemical catalysts largely stems from the fact that—built from L-amino acids—enzymes display chemo-, regio- and enantioselectivity. Additionally they are very efficient and environmentally compatible [3].

Although the versatility and diversity of microbial enzymes is undoubtedly extremely high, their biotechnological potential has not been fully explored. Limitations in the application to a specific reaction type or process are seen more in the supply with suitable enzymes rather than in the nature of substrates or products [4]. This emphasizes the need to search for new enzymes in efficient screens covering natural and synthetic molecular diversities.

2. Limitations of cultivation

The concept of prokaryotic diversity has dramatically expanded within the past decade. It has been known for a long time that many microbial species are difficult to cultivate in the laboratory because of specialized growth requirements. Comparing direct microscopic cell counts to the number of colonies growing on nutrient agar (viable cell counts, Fig. 1), often showed that in natural samples even fewer than

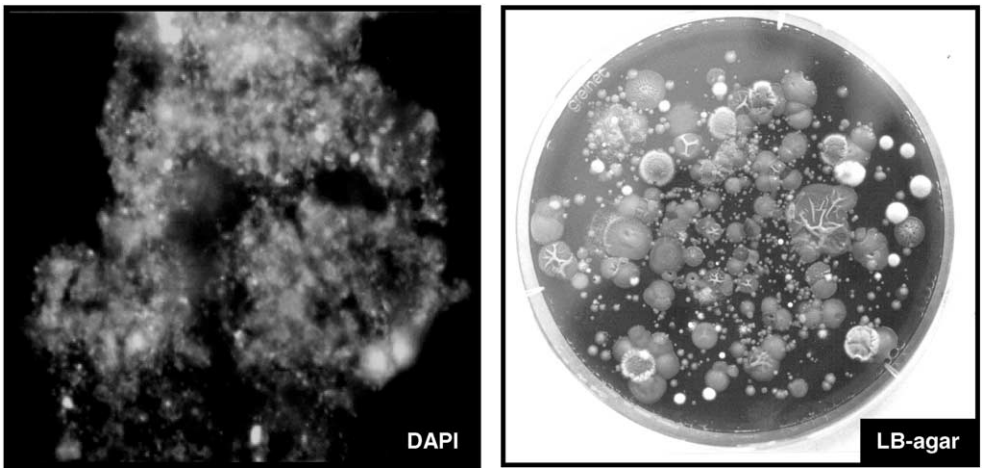

Fig. 1. Total vs. cultivatable microbial diversity. Microbial biodiversity as detected on a soil flake using fluorescence microscopy and a non-specific DNA stain (DAPI, 4′,6-Diamidino-2-phenylindole; left panel bright blue/white spots) is usually are 2–3 orders of magnitude higher than viable cell counts after cultivation in vitro (a plated soil suspension yields microbial colonies on enriched Luria–Bertani broth agar; right panel).

one cell in a thousand produced a colony ("the great plate count anomaly" [5]).

Modern molecular PCR-based studies in which DNA from environmental samples is directly amplified without culturing the corresponding microbes, has revealed that the uncultured species represent a spectacular diversity. The approach of cloning small subunit rRNA genes to identify uncultivated microbes and to map their phylogenetic relationships was first introduced by Pace and co-workers [6]. This approach has now been adopted to analyze the diversity of many different environments. All studies revealed an astonishing number of new microbial groups that had never been picked up by cultivation. Often, the new 16S RNA genes even represented new phyla within the bacterial or archaeal domain. [7]. In most habitats the fraction of uncultivated species exceeds 99% [8]. As the molecular ecological studies progress, it now becomes evident, that many of the novel lineages are widely distributed geographically and often represent abundant components of microbial populations [7,9]. There is no clear answer to the question, why so many and even abundant microorganisms have evaded cultivation. Some might require special handling and growth conditions that are hard or even impossible to mimik in the laboratory. Others might evade the enrichment in pure cultures because of their strict interdependence with other organisms. It becomes increasingly evident that not only more sophisticated techniques need to be developed in order to improve the enrichment and cultivation. But for the understanding of the ecological and biotechnological potential of most microorganisms, novel techniques that are independent of cultivation will be needed.

3. Widening horizons: the metagenome approach

Novel approaches are being developed to access the genetic resources of the vast majority of microbial species that so far have escaped scientific scrutiny.

They involve the direct extraction of genomic DNA from environmental samples and therefore obviate the need for cultivation of organisms. The environmental DNA is subsequently cloned into appropriate gene vectors to construct complex genomic libraries.

Thus, the goal is to directly clone the collective genomes of all microorganisms present in a habitat at a certain time point: the "metagenome" [10]. The resulting complex metagenomic libraries serve as a basis for screening procedures to identify protein genes with specific probes based on DNA–DNA hybridization or PCR protocols. An alternative, and probably the more promising approach to identify fundamentally novel genes and pathways is based on activity screening. Here, the complex libraries are analyzed for novel phenotypic expression in surrogate host strains.

Although the advent of the PCR paved the way for directly cloning "*in vitro*-replicas" of environmental genes from low quantities of DNA [11] it took until 1991 for the first *sensu stricto* cloning of directly isolated metagenomic DNA to be reported [12]. Meanwhile complex genomic libraries that contain large DNA inserts of up to 40 kilobasepairs (kb) or even more than 100 kb in BAC (bacterial artificial chromosome) vectors have been constructed from marine environments. They have served to characterize genomic fragments of hitherto uncharacterized, abundant marine archaea and bacteria [13–15].

The environmental cloning approach has also been used successfully for soil microbial communities [16–19]. However, as the success of the metagenomic approach is crucially dependent on the purity and clonability of the extracted DNA, the presence of high molecular weight inhibitors such as humic and fulvic acids in soil samples makes this rich source of biodiversity a considerably more difficult substrate for preparing metagenomic libraries than aquatic samples.

The fundamentals in this area of research were laid down in pioneering work by Torsvik and co-workers [20,21]. DNA was purified from a bacterial fraction isolated from soil. Importantly this DNA could be degraded by "DNAseI", a sign of relative purity and prerequisite for subsequent molecular cloning procedures. Since then numerous protocols for isolating DNA from soil have been evaluated (see e.g. [22–24] for reviews and compilations).

They basically fall into two categories (Fig. 2).

In the first approach microbes are lysed in the context of the substratum (*in situ*). Soil is suspended directly in lysis buffer and successively treated with detergent and enzymes. Some protocols in this category involve vigorous mechanical agitation (e.g. bead beating). The resulting DNA is subsequently

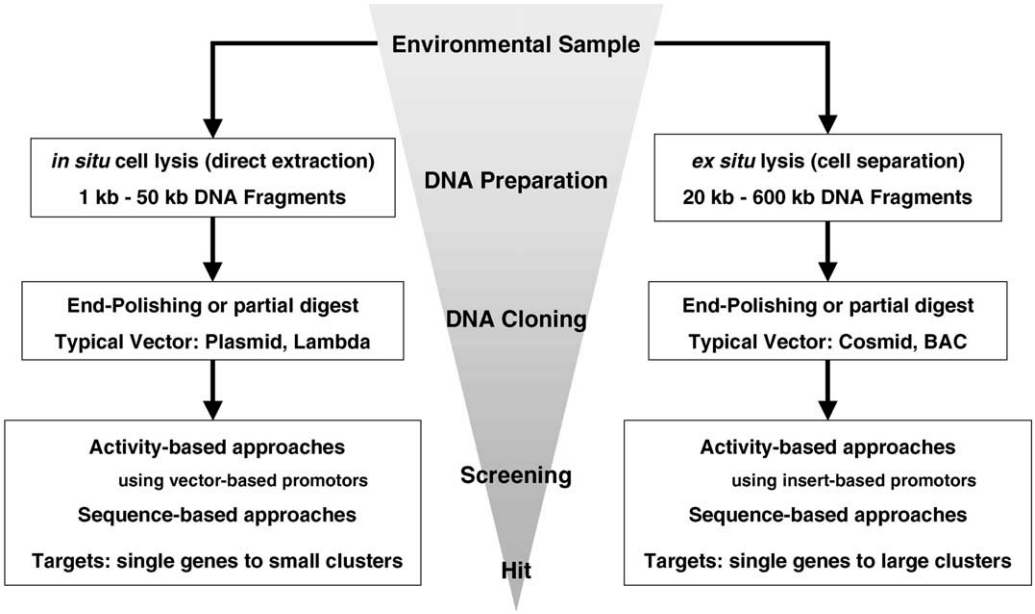

Fig. 2. Workflow in Metagenome cloning. DNA can be prepared differently depending on the size of the genomic target and the intended screening strategy.

phenol/chloroform-extracted and size-fractionated on a conventional agarose gel. The advantages of this approach are high DNA yields and a relatively low chance of an extraction bias due to failure of lysis of microorganisms with strong cell walls. However extraction efficiency has its price as the resulting DNA fragments are rather small (1 kb up to 50 kb) due to mechanical shearing (Fig. 3 left panel). Such DNA will be suitable primarily for cloning into plasmids or lambda vectors, carrying strong promotors to facilitate gene expression from the cloned fragment [25].

In the second approach microorganisms are physically separated from the substratum before lysis *ex situ*. This approach limits interferences of inhibiting organic and inorganic substrate components with cell lysis and DNA extraction procedures and can be elaborated to significantly reduce mechanical stress on released genomic DNA.

Consequently very large DNA fragments (20 kb up to >500 kb, see Fig. 3 right panel) with the potential to carry entire gene-clusters can be prepared for cloning into Cosmid or BAC vectors. Genomic libraries carrying such large inserts are particularly suited for sequence homology-based screening approaches while expression analysis will have to rely on insert-borne promotors.

An advantage of this method may be a reduction of contaminating non-microbial and free soil DNA in the preparation. The disadvantages however are the loss of microorganisms during preparation of the bacterial fraction due to failing separation from the substratum and consequently low DNA recoveries. In addition when avoiding mechanical stress for the sake of increased DNA fragment size, the microbial species having robust cell walls may escape lysis and be underrepresented in the metagenomic DNA pool.

The method of choice to be used in a metagenomic approach depends on the goals of the project: the size of the genomic target, its assumed organismic affiliation and the method of hit identification (see below). It is also important to realize the dimensions of the proposed undertaking. What is the size of a metagenome? Taking soil which can be home to several thousand bacterial species [26] and assuming average genome sizes of 4 megabasepairs the metagenome would comprise in the order of 10^{10} bp. The size of a DNA library representing the entire metagenome then would have to be 4.6×10^{10} bp [27] contained

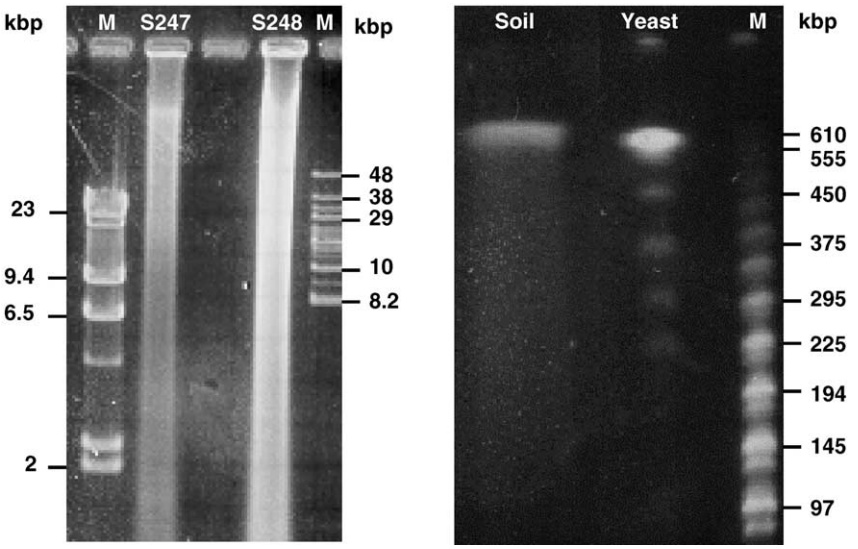

Fig. 3. Different DNA isolation procedures produce different fragment sizes.DNA extracted from microbial cells in situ generally is smaller and more heterogenous in size (1–50 kbp; left panel comparing soils S247 and S248) than DNA extracted from purified microbial pellets (>600 kbp; right panel comparing DNA extracted from soil with yeast genomic DNA). Metagenomic DNA and size markers (M in kbp) were separated by pulsed-field gel electrophoresis (PFGE) and visualized with ethidium bromide.

either in 10^7 plasmid (5 kb insert) or near 10^6 BAC (100 kb insert) clones [10]. These hugh numbers however are based on an idealized assumption that all species are equally represented in the isolated metagenomic DNA and the corresponding library. If the goal is to have a representation of the complete naturally occurring metagenome in the library, then the number of clones probably needs to be considerably higher. However, using normalization procedures (an attempt to enrich the DNA of low abundance organisms in the metagenomic DNA pool e.g. by separating DNA on the basis of its A–T content [28]) it may be possible to improve the representation of rare genomes in a library.

Table 1
Vector systems used in the discovery of enzyme activities in metagenome libraries

Vector	Enzyme activity	Host	DNA source	Reference
Plasmid	4-Hydroxy butyrate DH	*E. coli*	Soil	[16]
Plasmid	Lipase/esterase	*E. coli*	Soil	[25,30,33]
Plasmid	Protease	*E. coli*	Soil	[31]
Plasmid	Amylase	*E. coli*	Soil	[30]
Plasmid	Oxygenase	*E. coli*	Soil	[30]
Lambda	Chitinase	*E. coli*	Sea water	[29]
Cosmid	Oxygenase*	*E. coli*	Soil	[32]
	L-amino acid oxidase*	*E. coli*	Soil	[32]
	Hydroxylase*	*E. coli*	Soil	[32]
Cosmid/fosmid	Protease	*E. coli*	Soil	[30]
BAC	DNAse	*E. coli*	Soil	[17]
	Amylase	*E. coli*	Soil	[17]
	Lipase	*E. coli*	Soil	[17]

An asterisk "∗" indicates predicted enzyme-ORF activities in phenotypically active gene clusters.

4. Outcome of metagenomic screening campaigns

The generation, handling and analysis of comprehensive metagenome libraries call for extreme efficiencies of DNA extraction and cloning at the front end and high throughput screening procedures downstream. Therefore ultracompetent *Escherichia coli* will be the initial cloning host of choice and robotics seem mandatory. Libraries can be screened by hybridization or PCR-techniques on the basis of sequence similarity or conserved motifs. In this way it is possible to identify novel variants of distinct protein families or of known functional classes of proteins. Yet, more attractive for the identification of novel enzymes and drugs is a screening based on activities. In this case identification of hits is dependent on successful transcription and translation of the protein genes in the surrogate host. If small DNA inserts are cloned, external transcriptional promoters from the vector might be used for expression. Large insert libraries, instead, will depend on the expression from internal signatures from the original donor. This sequence-independent approach generally allows for the identification of entirely novel enzymatic activities or specificities.

As most microbial species and the majority of the microbial biomass are unknown, most genes identified in a metagenomic library will be novel. The wealth of new genes from hitherto uncultured species is such that chances of a molecular *deja vu* are low even when skimming only the surface of the unknown diversity. A situation that will naturally change, however, with a growing body of sequence data. Reports from the literature and own experience (Table 1) show that even screening comparatively small metagenomic libraries constructed in vectors relying on vector-borne and insert-borne promotors for expression produces novel hits with often very low sequence similarity to entries in public databases. Metagenomic activity screens have revealed enzymes like chitinase [29], amylase, DNAse [17], esterase/lipase [17,25,30], protease [31], oxygenase [30] and antimicrobial activities potentially involving multigene expression [17,18,32]. Using a genetic complementation approach Henne and co-workers isolated genes encoding cytoplasmic 4-hydroxybutyrate dehydrogenase from the environment [16]. The large diversity of biochemical properties to be expected from screening

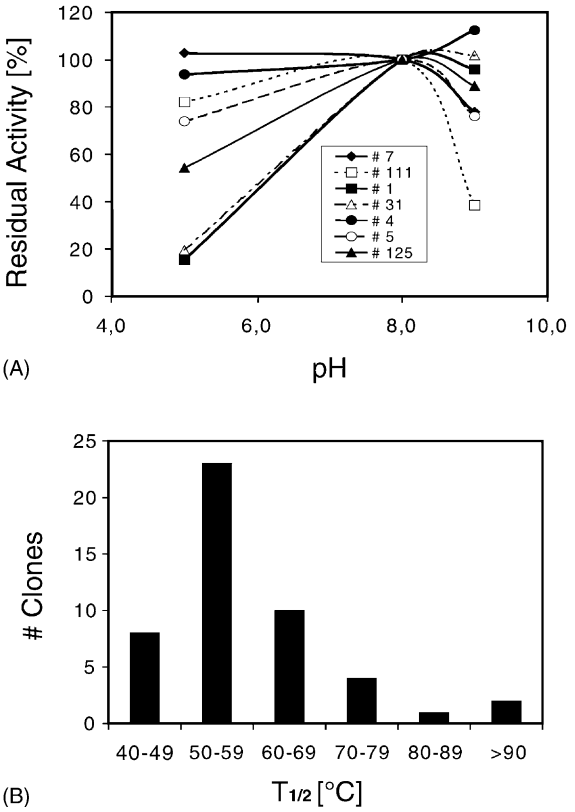

Fig. 4. Biochemical diversity of metagenomic esterase/lipase clones from a mesophilic habitat.A metagenome library was prepared in pUC18 from DNA directly extracted from a pH 8 soil sampled near Heidelberg, Germany. Of around 50,000 primary *E. coli* colonies approximately 120 tested esterase/lipase-positive on tributyrine agar. Enzyme activities of selected clones were measured as initial rates of *p*-nitrophenol butyrate hydolysis (increase in OD_{410}). (A) pH stability: extracts from positive clones were preincubated at 30 °C for 30 min at the pH-values indicated. Subsequently hydrolysis was measured in 10 mM Tris–HCl, pH 8.0 (activity after pH 8.0 incubation normalized to 100%); (B) thermal stability: extracts from 48 positive clones were incubated for 30 min in 10 mM Tris–HCl, pH 8.0 at temperatures between 40 and 95 °C. After cooling to 30 °C activity on *p*-nitrophenol butyrate was assayed as above. Shown are number of clones falling into ranges of half-maximal inactivation temperatures ($T_{1/2}$).

metagenomic libraries is exemplified in a study with esterases/lipases. In Fig. 4 pH- and thermostabilities of esterases/lipases are shown that were isolated from a plasmid library prepared from DNA extracted from an alkaline soil sampled near Heidelberg, Germany. A great range of pH- and thermostability optima was found within 120 analyzed clones.

Obviously these initial results are promising and prospects of seemingly boundless diversity are tempting. Yet methodological problems will set the limits of this approach. Sequence-based techniques by nature are conservative and expression screens that have the potential to reveal enzymes covering entirely novel sequence-space are hampered with all known problems of heterologous gene expression. Using a variety of different hosts (e.g. *Streptomyces lividans* and *Bacillus* sp.) in addition to *E. coli* should significantly boost the success rate in heterologous expression screens.

5. Conclusion

The microbial world seems to offer the greatest natural resource of molecular diversity. The classical approach of cultivating and characterising isolates on the strain level prior to screening and gene isolation is valid and powerful yet severely restricted in scope. It needs to be complemented by direct cloning approaches targeting environmental DNA. Thereby it is possible to access the so far uncharacterized majority of microbes that in any habitat may constitute up to 99% of the entire population. These new dimensions of genomic coverage likely promise a quantum leap in novel enzyme sequence space to be discovered.

References

[1] N.R. Pace, Science 276 (1997) 734.
[2] W.B. Whitman, D.C. Coleman, W.J. Wiebe, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 6578.
[3] K. Faber, Biotransformations in Organic Chemistry—A Textbook, Springer, Berlin, Heidelberg, New York, 2001.
[4] A. Schmid, J.S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, Nature 409 (2001) 258.
[5] J.T. Staley, A. Konopka, Annu. Rev. Microbiol. 39 (1985) 321.
[6] N.R. Pace, D.A. Stahl, D.J. Lane, G.J. Olsen, Adv. Microb. Ecol. 9 (1986) 1.
[7] P. Hugenholtz, B.M. Goebel, N.R. Pace, J. Bacteriol. 180 (1998) 4765.
[8] R.I. Amann, W. Ludwig, K.H. Schleifer, Microbiol. Rev. 59 (1995) 143.
[9] E.F. DeLong, Curr. Opin. Genet. Dev. 8 (1998) 649.
[10] J. Handelsman, M.R. Rondon, S.F. Brady, J. Clardy, R.M. Goodman, Chem. Biol. 5 (1998) R245.
[11] R.J. Steffan, J. Goksoyr, A.K. Bej, R.M. Atlas, Appl. Environ. Microbiol. 54 (1988) 2908.
[12] T. Schmidt, E. DeLong, N. Pace, J. Bacteriol. 173 (1991) 4371.
[13] J.L. Stein, T.L. Marsh, K.Y. Wu, H. Shizuya, E.F. DeLong, J. Bacteriol. 178 (1996) 591.
[14] C. Schleper, E.F. DeLong, C.M. Preston, R.A. Feldman, K.Y. Wu, R.V. Swanson, J. Bacteriol. 180 (1998) 5003.
[15] O. Beja, M.T. Suzuki, E.V. Koonin, L. Aravind, A. Hadd, L.P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S.B. Jovanovich, R.A. Feldman, E.F. DeLong, Environ. Microbiol. 2 (2000) 516.
[16] A. Henne, R. Daniel, R.A. Schmitz, G. Gottschalk, Appl. Environ. Microbiol. 65 (1999) 3901.
[17] M.R. Rondon, P.R. August, A.D. Bettermann, S.F. Brady, T.H. Grossman, M.R. Liles, K.A. Loiacono, B.A. Lynch, I.A. MacNeil, C. Minor, C.L. Tiong, M. Gilman, M.S. Osburne, J. Clardy, J. Handelsman, R.M. Goodman, Appl. Environ. Microbiol. 66 (2000) 2541.
[18] I.A. MacNeil, C.L. Tiong, C. Minor, P.R. August, T.H. Grossman, K.A. Loiacono, B.A. Lynch, T. Phillips, S. Narula, R. Sundaramoorthi, A. Tyler, T. Aldredge, H. Long, M. Gilman, D. Holt, M.S. Osburne, J. Mol. Microbiol. Biotechnol. 3 (2001) 301.
[19] A. Quaiser, T. Ochsenreiter, H.-P. Klenk, A. Kletzin, A.H. Treusch, G. Meurer, J. Eck, C. Sensen, C. Schleper, Environ. Microbiol. (2002) in press.
[20] V. Torsvik, J. Goksoyr, Soil Biol. Biochem. 10 (1978) 7.
[21] V. Torsvik, Soil Biol. Biochem. 12 (1980) 15.
[22] J.T. Trevors, J.D. vanElsas, Nucleic Acids in the environment—Methods and Applications, Springer, Berlin, Heidelberg, New York, 1995.
[23] M. Krsek, E.M. Wellington, J. Microbiol. Meth. 39 (1999) 1.
[24] S. Courtois, A. Frostegard, P. Goransson, G. Depret, P. Jeannin, P. Simonet, Environ. Microbiol. 3 (2001) 431.
[25] A. Henne, R.A. Schmitz, M. Bomeke, G. Gottschalk, R. Daniel, Appl. Environ. Microbiol. 66 (2000) 3113.
[26] V. Torsvik, J. Goksoyr, F.L. Daae, Appl. Environ. Microbiol. 56 (1990) 782.
[27] B. Seed, R.C. Parker, N. Davidson, Gene 19 (1982) 201.
[28] A.M. Johnson, J.P. Dubey, J.B. Dame, Aust. J. Exp. Biol. Med. Sci. 64 (1986) 351.
[29] M.T. Cottrell, J.A. Moore, D.L. Kirchman, Appl. Environ. Microbiol. 65 (1999) 2553.
[30] P. Lorenz, K. Liebeton, F. Niehaus, J. Eck, H. Zinke, in: Proceedings of the Fifth International Symposium on Biocatalysis and Biotransformation, 2–7 September 2001, Darmstadt, Germany, p. 379.
[31] R. Gupta, Q.K. Beg, P. Lorenz, Appl. Microbiol. Biotechnol. 59 (2002) 15.
[32] S.F. Brady, C.J. Chao, J. Handelsman, J. Clardy, Org. Lett. 3 (2001) 1981.
[33] K. Liebeton, P. Lorenz, F, Niehaus, T. Ochsenreiter, A. Qaiser, C. Schleper, J. Eck, H. Zinke, Conference Proceedings of the Jahrestagung der Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM) (2002), March 24–27 Göttingen, Germany, p. 88.