

The magic and menace of metagenomics: prospects for the study of plant growth-promoting rhizobacteria

Johan H. J. Leveau

Received: 25 December 2006 / Accepted: 9 July 2007 / Published online: 21 September 2007
© KNPV 2007

Abstract This article aims to be a pragmatic primer into the field of metagenomics with special emphasis on the prospective contributions of metagenomics to the study of plant growth-promoting rhizobacteria (PGPR). After an introduction into the concepts and methodologies of metagenomics and a discussion of the numerous emerging variations on the basic theme, there will follow a short overview of the success stories in metagenomics (the ‘magic’ in the title of this review), a brief discussion about the technical problems and unrealistic expectations that are sometimes associated with metagenomics (the ‘menace’), and a shortlist of the lessons that can be learned by those that wish to explore the utility of metagenomics in the study of PGPR.

Keywords Metagenomics · PGPR · Plant growth-promoting rhizobacteria · Environmental shotgun sequencing

Metagenomics

The phrase ‘metagenome of the soil’ was first used by Handelsman et al. (1998) to describe the collective genomes of soil microflora. In the area of microbial

ecology, the term ‘metagenomics’ is now synonymous with the culture-independent application of genomics techniques to the study of microbial communities in their natural environments (Chen and Pachter 2005). Metagenomics arose in reaction to the observation that the majority of microorganisms on Earth resist life in captivity, i.e. they cannot be grown in broth or on plates in the laboratory. An often-cited estimate is that as much as 99% or more of microbial life remains unculturable, and therefore cannot be studied and understood in a way that microbial ecologists have become accustomed to over the past century. Metagenomics exploits the fact that while some microorganisms are culturable and others are not, all of them (i.e., 100%) are life-forms based on DNA as a carrier of genetic information. The metagenomic toolbox allows accessing, storing, and analyzing this DNA and thus can provide an otherwise hard-to-attain insight into the biology and evolution of environmental microorganisms, independent of their culturable status.

In modern metagenomics, three major and often-times overlapping directions can be recognized. The first is aimed at linking phylogeny to function. Once microbial ecologists got a satisfactory grasp on the issue of ‘who is out there?’ (Amann 2000), they set out to answer the question ‘who’s doing what out there?’. As one of many complementary methodologies, metagenomics can help answer that question in an indirect, culture-independent manner. One example

J. H. J. Leveau (✉)
Netherlands Institute of Ecology (NIOO-KNAW),
Heteren, The Netherlands
e-mail: j.leveau@nioo.knaw.nl

is through phylogenetic anchoring (Riesenfeld et al. 2004b), which involves the screening of large-insert environmental libraries for clones that carry phylogenetically informative genes (to reveal the ‘who’) and analyzing their flanking DNA for genes that reveal possible environmental functions of the DNA’s owner. A second trend in modern metagenomics involves its exploitation for the discovery of enzymes with novel, industrial and possibly exploitable properties. This aspect of metagenomics was well-reviewed recently by Lorenz and Eck (2005), who conclude that metagenomics “provides industry with an unprecedented chance to bring biomolecules into industrial application”. The third and most recent trend in metagenomics is the mass sequencing of environmental samples. The promise of this approach is to offer a more global or systems-biology view of the community under study. Indeed, in several instances has mass sequencing led to more complete assessments of genetic diversity and to first insights into the interactivities that occur in microbial communities (DeLong et al. 2006; Edwards et al. 2006; Gill et al. 2006; Hallam et al. 2004; Schmeisser et al. 2003; Tringe et al. 2005; Tyson et al. 2004; Venter et al. 2004).

Early pioneers in the field of metagenomics were Schmidt et al. (1991) who studied the phylogenetic diversity of an oligotrophic marine picoplankton community in the north central Pacific Ocean. Their original protocol involved the (1) isolation of bulk genomic DNA from picoplankton collected by tangential flow filtration, (2) fragmentation, size-fractionation (10–20 kb) and cloning of the mixed-population DNA into bacteriophage lambda, (3) screening of the resulting library of recombinant phages by hybridization with 16 S rDNA probes, (4) sequencing of PCR-amplified 16 S rDNA from positive clones, and (5) analysis of the DNA sequence from unique clones to database entries to reveal some of the uncultured diversity of picoplanktonic life in the Pacific Ocean. This series of subsequent steps (DNA isolation, cloning, library screening, sequencing of interesting clones, and DNA comparison) is in essence the classical metagenomic strategy as defined by Handelsman et al. (1998). This basic theme of metagenomics is also depicted schematically in Fig. 1. This figure will be used as a framework to describe the metagenomic methodology and illustrate the many variations that have evolved over the past few years. For good overviews on the subject of

metagenomics, several recent reviews are available (Daniel 2005; Deutschbauer et al. 2006; Green and Keller 2006; Handelsman 2004; Kowalchuk et al. 2007; Schloss and Handelsman 2005; Schmeisser et al. 2007; Streit and Schmitz 2004; Tringe and Rubin 2005; Ward 2006; Xu 2006). The present article presents an overview of metagenomics that is to some extent biased towards the discussion in the final section about how to capitalize on metagenomics as a tool in the study of PGPR.

Isolation of metagenomic DNA

The first and obviously most important step in any metagenomic approach is the isolation of DNA from

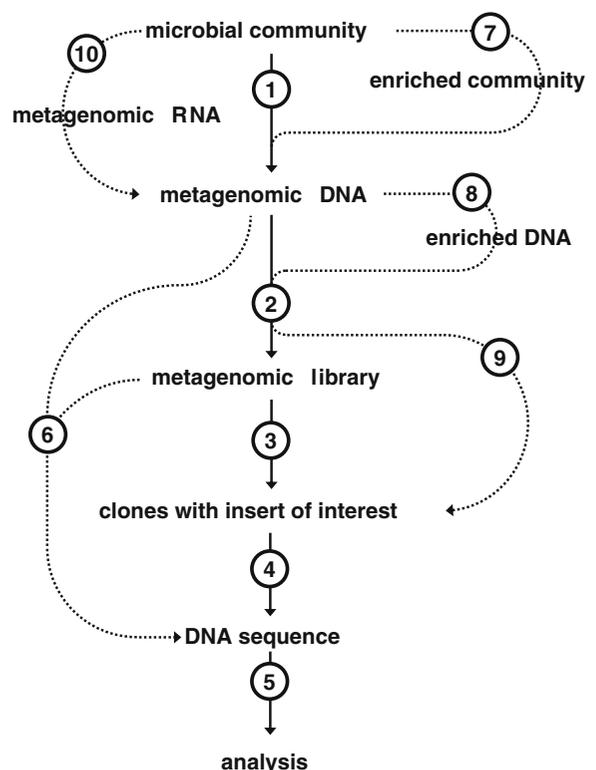


Fig. 1 Schematic representation of the classical metagenomic protocol (1–5) and variations on a theme (6–10). Each of the following steps is discussed in more detail in the text: (1) Isolation of metagenomic DNA; (2) Cloning of metagenomic DNA; (3) Metagenomic library screening; (4) DNA sequencing; (5) Sequence analysis; (6) Environmental shotgun sequencing; (7) Enrichment of a particular subpopulation; (8) Enrichment strategies at the DNA level; (9) Direct all-or-none selection for clones of interest; (10) Isolation of metagenomic RNA

the environment under study. Looking at the many protocols that have been published to date (for a fairly broad overview, see individual chapters in Kowalchuk et al. 2004), it becomes apparent that no single protocol is suitable for the extraction of DNA from all environments. Key issues to consider at this stage are the quantity, purity, integrity, and representativeness of the DNA after isolation. DNA can be extracted from microorganisms by lysis either directly in the environmental sample, or indirectly, i.e. after separation and concentration of the microbial cells from their environmental matrix. The latter is often inevitable or recommended for many environments but this may be for different reasons. Isolation of DNA directly from ocean water is not practical given the low microbial density, so some filtration step is usually performed to first concentrate cells (Fuhrman et al. 1988). Many soil types are notorious for the presence of contaminants such as polyphenolics that co-purify with DNA and can inhibit subsequent steps in the metagenomic process (Tebbe and Vahjen 1993). To prevent this, several groups have developed ways to first separate cells from the soil matrix, e.g. by application to a Nycodenz cushion (Lindahl and Bakken 1995). Others preferentially extract DNA directly from soil, mainly for reasons of increased DNA yield and lesser bias (see below). In these cases, a separate DNA purification step is often included to minimize contamination with unwanted soil substances (Zhou et al. 1996). For direct extraction/purification of environmental DNA from soil, many commercial kits are now also available, for example the SoilMaster™ DNA Extraction Kit (Epicentre, Madison, WI) and PowerMax™ Soil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA).

There is a growing recognition that different DNA extraction methods can yield different results in terms of microbial representation in the final DNA sample (Kozdroj and van Elsas 2000). For example, indirect DNA extraction from marine sediments resulted in a reduced observed microbial diversity at the DNA level than when direct lysis methods were used (Luna et al. 2006). As noted elsewhere (Tringe and Rubin 2005), the harsh lysis methods that are necessary to extract DNA from one organism will cause degradation of the DNA from another organism. A practical implication of this is that more than one type of extraction method should be used in the construction of a metagenomic library. Some DNA extraction

protocols have a requirement for the minimal amount of starting material, which can be a disadvantage for the isolation of metagenomic DNA from environments that permit only small sampling sizes. A related issue is the efficiency of DNA extraction. Chemical extraction is usually less effective than procedures that involve rigorous bead-beating (Miller et al. 1999). If DNA yield is a problem – for example, when it is less than the 0.5- to 4- μ g minimum to construct a library for shotgun sequencing (Abulencia et al. 2006) – one could consider the use of recently developed protocols for the amplification of whole genomes from environmental samples. One example is the use of Φ 29 DNA polymerase which by a process of multiple displacement can provide enough DNA from only five bacterial cells for subsequent analyses (Abulencia et al. 2006).

A final factor to consider is integrity of the isolated DNA which dictates how it can be used in the subsequent steps of the metagenomic protocol. Bead-beating-type procedures tend to fragment DNA and are therefore mostly used for the construction of small-insert libraries, e.g. for shotgun sequencing. In contrast, gentle chemical lysis of cells recovered from soil in agar plugs can produce very large clonable DNA with fragment sizes exceeding 1 Mbp (Berry et al. 2003). The latter is more interesting when the objective is to screen metagenomic libraries for phylogenetic anchors or for particular phenotypes (see “[Metagenomic library screening](#)”).

Cloning of metagenomic DNA

After extraction and purification of metagenomic DNA, it is usually fragmented to the desired size (either enzymatically or by physical force) and size-separated, if necessary, e.g. on agarose gels (Liles et al. 2004). To carry the metagenomic DNA, a variety of vectors is available, including plasmids, bacteriophages, cosmids, fosmids, and bacterial artificial chromosomes (BACs). A common way to describe these vectors is by the size of DNA fragments that they can typically accommodate, i.e. 0.5–2, 7–10, 35–40, and 80–120 kb, respectively (Xu 2006). However, by this definition some vectors would qualify both as fosmids and as BACs. For example, Epicentre’s pCC1FOS and pCC1BAC are virtually identical but are sold as fosmid and BAC, respectively. The

difference in nomenclature lies not in the nature of the vector, but in the method that is used to introduce vector-ligated metagenomic DNA into the host strain, in this case *E. coli*. With pCC1FOS, this is mediated through lambda packaging, while pCC1BAC is transformed by electroporation. A limitation of the packaging procedure is that there is an upward size-selection against DNA inserts larger than ~40 kb. On the other hand, DNA inserts smaller than ~35 kb are also rejected, so that a typical fosmid library contains very few false-negatives, i.e. clones that lack an insert or contain very small fragments. With BACs, large DNA fragments can be cloned (Rondon et al. 2000), even as large as 1 Mbp (Berry et al. 2003), but it is also more common for a BAC library to be quite variable in insert size, with a strong selection for small insert sizes and only a relatively small subset of clones carrying very large fragments (Rondon et al. 2000).

Insert size is a critical matter in any metagenomic approach. For environmental shotgun sequencing (see “Environmental shotgun sequencing”), insert sizes of 1.5 to 3 kb are preferred, whereas larger inserts are needed if one wants to maximize on the phylogenetic anchoring approach or if one’s objective is to clone large operons e.g. those involved in the production of certain antibiotics (Handelsman et al. 1998). Especially in cases such as the latter, where metagenomic DNA of interest is only detected if it is expressed, the choice of host is crucial to success. *Escherichia coli* is the standard in most metagenomic applications, mainly because of ease of manipulation, but it has limited and biased ability in expressing heterologous genes. Gabor et al. (2004a) provided estimates that only 7% of coding sequences from representatives of the class Actinobacteria would actually be expressed in an independent manner in an *E. coli* background, compared to 73% of the genes from Firmicutes origin. Indeed, the use of alternative hosts, such as *Streptomyces* (Courtois et al. 2003; Wang et al. 2000), *Pseudomonas* (Martinez et al. 2004), or *Rhizobium leguminosarum* (Li et al. 2005) can reveal gene activities that would not have been picked up in an *E. coli* library.

A final consideration in the construction of a metagenomic DNA library is the number of clones that is needed to achieve the desired outcome. It has been estimated that 10^6 BAC clones with 100-kb inserts are required to represent the genomes of all the different prokaryotic species present in 1 g of soil

(Handelsman et al. 1998), assuming that all species are equally abundant. If this is not the case, $>10^{11}$ BAC clones, or 10 Tbp of DNA, would be required to achieve adequate genomic representation of rare species (Daniel 2005). Hence, cloning the entire microbial metagenome of an environment as complex as soil is not feasible. To illustrate a more modest goal, suppose one sets out to find bacterial antibiotic resistance genes in a metagenomic library. If one such gene occurs in every 100 bacterial genomes in the metagenome (with an average genome size of 5 Mbp), it would take at least 57,500 40-kb inserts to find a single such gene with a 99% probability. This relative low success rate has not only driven the interest in and application of automated handling systems (‘picking robots’) to construct and screen metagenomic libraries. It has also led to the exploration and development of variations on the classic metagenomic strategy in order to specifically increase the chances of finding clones of interest (see “Enrichment of a particular subpopulation,” “Enrichment strategies at the DNA level,” “Direct all-or-none selection for clones of interest,” and “Isolation of metagenomic RNA,” and the corresponding arrows in Fig. 1).

Metagenomic library screening

Two types of screening strategies can be distinguished. Those that are sequence-based capitalize on pre-existing DNA information to find clones in a metagenomic library that carry inserts with sequence similarity to a gene or locus of interest. This is achieved, for example, by colony or plaque hybridization with probes (Knietzsch et al. 2003a; Schmidt et al. 1991) or by PCR using a specific primer set (e.g. Courtois et al. 2003). Most common targets of these searches are the genes for subunits of the ribosome (e.g. Liles et al. 2003; Quaiser et al. 2002), most frequently the small subunit or 16 S rRNA gene. These genes serve as phylogenetic anchors to link the identity of a DNA’s owner to part of its biology and evolution by analysis of the DNA sequences flanking the rRNA genes. Special mention is also warranted for the screening method LIL-FISH, for large-insert library fluorescent in situ hybridization (Leveau et al. 2004). In essence, it is an activity screening that utilizes FISH to identify clones in a metagenomic library which heterologously express ribosomal RNA

genes. Its compatibility with FACS makes it an attractive method for high-throughput screening of libraries for phylogenetic anchors. A final screening strategy that still needs to be fully explored for its application in metagenomics but worthy of mention here is magnetic capture-hybridization (Jacobsen 1995). This technique involves the use of magnetic beads which are coated with a single-stranded DNA probe complementary to a target gene of interest and which can be used to identify fosmid clones of interest in a library (Hackl et al., unpublished data).

The second type of library screening is based on the expression of genes of interest through detection of their associated phenotypes in a clone library. The importance of a suitable host strain in this context has already been discussed. Preferably, the gene's phenotype is readily detectable and often involves an enzymatic activity on a plate or in a microtiter plate. Recently, several novel screening methods have been described. Two of these, SIGEX and METREX, are based on the use of green fluorescent protein as a reporter of gene activity. SIGEX stands for substrate-induced gene expression and was developed by Uchiyama et al. (2005). It allows a high-throughput screen for catabolic genes by exploitation of the fact that many such genes, e.g. those that code for the degradation of aromatic compounds are commonly induced by the substrate they are targeted to degrade. Hence, by coupling metagenomic DNA to a promoterless *gfp* gene, library clones carrying genes of interest can be readily identified as green fluorescent cells or colonies in the presence of the inducing substrate. The power of SIGEX lies in the fact that it is compatible with fluorescence-activated cell sorting, which makes library screening a matter of seconds or minutes. METREX (Williamson et al. 2005) is a similar strategy that helps identify functionally active library clones based on an intracellular screen for quorum-sensing inducers.

DNA sequencing

Once clones of interest are identified by a functional or sequenced-based screening, the next step usually involves the elucidation of the inserts' partial or complete DNA sequence. Typically, large-insert fosmid clones are subjected to shotgun or transposon sequencing. To get a typical 10 \times coverage of a 45-kb

fosmid (37-kb insert and 8-kb vector), 600 sequence reads of ~750 bp each are needed.

Most of modern sequencing takes place according to the original Sanger method (Sanger et al. 1977) of chain-termination by dideoxynucleotides and subsequent separation of differently sized DNA fragments by capillary gel electrophoresis. This method has proven to be very compatible with the high-throughput demand of many (meta)genomic projects. Recently, novel types of sequencing methods have been developed that are faster and have an even greater capacity. For example, pyrosequencing now allows the sequencing of 25 million bases in one 4-h run with an accuracy of 99.96% (Margulies et al. 2005). While this technology will not likely replace traditional sequencing, it has great utility for the environmental shotgun sequencing strategies discussed later (see “[Environmental shotgun sequencing](#)”). While there are some serious limitations to pyrosequencing – the main ones being the generation of relatively short (~100-bp) reads and the poor ability of most current DNA analysis programmes to assemble such short reads – it is an exciting new development that will undoubtedly revolutionize the field of microbial (meta)genomics.

Sequence analysis

There are many ways to analyze metagenomic DNA sequences (Chen and Pachter 2005). Gene-finding is supported by software applications such as GeneDB (Meyer et al. 2003), Artemis (Rutherford et al. 2000), Glimmer (Delcher et al. 1999), and FGenesB pipeline (www.softberry.com). These programmes use special algorithms to identify coding sequences, as well as other features such as promoters, terminators, operons, tRNA and rRNA. They can also provide a functional prediction of each identified putative gene based on sequence similarity of its predicted product. The programme MetaGene (Noguchi et al. 2006) is a prokaryotic gene-finding application that was designed specifically for the analysis of metagenomic datasets with many unassembled reads. Such datasets are typical for many types of environmental shotgun sequence projects (see “[Environmental shotgun sequencing](#)”). A crude approach to gene-finding is a BlastX-type of analysis, by which the DNA query sequence is first translated into protein sequences in

all six reading frames, after which these products are compared against the existing protein databases. This approach depends heavily on the availability of similar sequences in the database and novel genes might be missed in this way.

A particular problem with small-sized metagenomic DNA fragments is that they often lack phylogenetic markers in the shape of e.g. a 16 S rRNA gene. In those cases, parameters such as G+C content, BlastX scores, and codon usage frequencies can also be used as indicators of phylogenetic origin (Chen and Pachter 2005). Analysis of oligonucleotide frequencies is another promising approach to this problem, since these frequencies tend to exhibit species-specific patterns (Abe et al. 2005). TETRA (Teeling et al. 2004) is available as a web-service (www.megx.net/tetra) or stand-alone application that automates the task of comparing tetranucleotide frequencies. It computes correlation coefficients between patterns of tetranucleotide usage in DNA and works best with sequences of about 40 kb. Most recently, TETRA was used to group metagenomic DNA sequences from the marine worm *Olavius algarvensis* into four clusters representing four prokaryotic symbionts (Woyke et al. 2006).

Environmental shotgun sequencing

As already alluded to in the introduction, a recent development in metagenomics is environmental shotgun sequencing, also known as whole-genome sequencing (Chen and Pachter 2005). It usually involves the construction and end-sequencing of small-insert libraries from DNA directly isolated from the environment under study, although several variations exist. For example, the dataset of DeLong et al. (2006) consisted of end-sequences from a large-insert fosmid library, and Edwards et al. (2006) applied pyrosequencing which makes library construction redundant altogether. Examples of environments from which prokaryotic communities have been shotgun-sequenced are the Sargasso sea (Venter et al. 2004), the North Pacific ocean's surface to near-sea floor depths (DeLong et al. 2006), sunken whale carcasses (Tringe et al. 2005), deep-sea sediment (Hallam et al. 2004), an acid mine biofilm (Tyson et al. 2004), groundwater in banded iron formations of a subsurface mine (Edwards et al. 2006), the human distal gut (Gill et al. 2006), drinking-

water networks (Schmeisser et al. 2003), and an agricultural soil (Tringe et al. 2005).

One characteristic that all environmental shotgun sequencing projects share is the incredible amount of data that is generated (Schloss and Handelsman 2005). The Sargasso Sea project alone produced 2 million sequence reads, or 1,600 Mbp of DNA sequence (Venter et al. 2004). This explosion in environmental sequence data led several authors (Handelsman 2005; DeLong 2004) to refer to metagenomics as 'megagenomics' or 'mega-metagenomics,' respectively. This 'genomics on a massive scale' poses considerable challenges for data assembly and annotation. Successful assembly of individual reads into contigs is inversely proportional to the complexity of the prokaryotic community from which the DNA originated. The acid mine biofilm studied by Tyson et al. (2004) represented a relatively simple community, with only three bacterial and three archeal lineages. From two members of the biofilm, *Ferroplasma* type II and *Leptospirillum* group II, near-complete genome sequences could be obtained after assembly. In more complex environments, this is typically not the case. For example, 50% of the reads from the Sargasso Sea could not be assembled, and for the agricultural soil, with an estimated species richness of >3,000 (Tringe et al. 2005), the percentage of unassembled reads even approached 100%. Interestingly, Tringe et al. (2005) recently questioned the need for assembly in such cases altogether and introduced the concept of environmental gene tags (EGTs), which takes a more gene-centric approach to the analysis of environmental sequencing. EGTs are in essence annotated individual reads from a metagenome shotgun project. Predicted genes on these EGTs are derived from individual members of the microbial community under study, and could potentially code for a habitat-specific adaptation. By comparison of EGT 'fingerprints' of different environments, Tringe et al. (2005) indeed observed emerging patterns of habitat-specific gene functions, for example, an over abundance of rhodopsin-like proteins in ocean surface waters, enzymes for the degradation of plant material in soil, and sodium transport proteins in marine environments. Basically, this approach represents an in silico version of subtractive hybridization (Schloss and Handelsman 2005) and presents one of several examples (Rodriguez-Brito et al. 2006) on how to tackle the challenges of comparative analysis of metagenomic datasets.

Enrichment of a particular subpopulation

This section is the first of four that describe some of the variations that have evolved from the basic metagenomic theme (Fig. 1) with the specific goal to increase the chance of finding a gene or function of interest. One of those variations is the subject of this section and involves community fractionation, i.e. the isolation of DNA only from a selected subpopulation of the environmental sample under study. One such type of selection is size. Using filters with different pore sizes, bacteriophage communities can be selectively enriched for metagenomic analyses (Edwards and Rohwer 2005). Similarly, size fractionation was used to select for consortia $> 3 \mu\text{m}$ from deep-sea sediments (Hallam et al. 2004). Other strategies for enrichment of subpopulations include affinity purification or differential lysis (Tringe and Rubin 2005).

A very effective type of selection at the community level is based on the application of selective pressure to the environmental sample under study in order to favour the growth of bacteria expressing a desired activity. In an early example of this strategy, Healy et al. (1995) readily recovered clones expressing cellulase and xylosidase activities from libraries of DNA isolated from a mix of thermophilic, anaerobic digesters, that were in continuous operation with lignocellulosic feedstocks for >10 years. In another case, Gabor et al. (2004b), DNA was isolated from enrichment cultures in which amides (either singly or as a mixture of aromatic and non-aromatic forms) were supplied as the sole nitrogen source to obtain a library enriched for amidases with different substrate specificities. Similarly, Entcheva et al. (2001) enriched for biotin-producing bacteria to isolate new biotin biosynthesis operons, whereas Knietsch et al. (2003b) pre-selected for utilization of glycerol and 1,2-propanediol to metagenomically identify genes encoding alcohol oxidoreductases.

Another clever selection procedure allows the isolation of DNA only from the live fraction of cells in a community. Nocker and Camper (2006) reported that propidium monoazide (PMA), like propidium iodide, is highly selective in penetrating only 'dead' bacterial cells where it intercalates in the DNA and can be covalently cross-linked by exposure to bright light. This process renders the DNA insoluble and results in its loss during subsequent genomic DNA extraction. Subjecting a bacterial population com-

prised of both live and dead cells to PMA treatment would thus result in selective removal of DNA from dead cells, which can be useful if one is interested only in analyzing the currently active, living fraction of a microbial community.

Enrichment strategies at the DNA level

A second kind of selection for genes or gene functions of interest involves selection at the level of metagenomic DNA after it has been isolated from the environment under study. The simplest application of this type of enrichment is the use of metagenomic DNA as a template in PCR (with degenerate primers, if needed) to amplify and clone genes of interest (Marchesi and Weightman 2003). A method described by Nesbø et al. (2005) allows for the enrichment only of metagenomic DNA that carries phylogenetic markers. Central in this approach is the restriction enzyme I-CeuI, which targets a 19-bp sequence that is conserved in 23 S rRNA of most bacteria. After isolation, fragmentation and end-repair, metagenomic DNA is digested with I-CeuI and ligated to vector pCC1.FOS.CeuI.23S, a derivative of pCC1FOS containing unique I-CeuI and blunt sites. To illustrate the effectiveness of this approach, fosmid libraries were constructed from anaerobic sediments sampled from Baltimore harbour, and found to be enriched by an approximate factor of 80 for clones carrying 23 S rRNA genes. An attractive side advantage of the method is that end-sequencing of clones in this library, which is relatively cheap, can provide instant information on the identity of the DNA's origin based on 23 S similarity.

Stable isotope probing is a technique by which to isolate from the metagenome pool specifically that DNA which is derived from organisms that can metabolize a particular substrate (Friedrich 2006). This is achieved by incorporation of ^{13}C - or ^{15}N -labelled substrates into biomass, including DNA, of the active subpopulation in the community under study. The labelled DNA (from the active microorganisms) is then separated from unlabelled DNA through density gradient centrifugation and used as starting material for the cloning into vectors. One example is the report by Dumont et al. (2006) which describes the application of $^{13}\text{CH}_4$ to forest soil to generate a BAC library enriched for methane monooxygenase genes. Schwarz et al. (2006) used ^{13}C -labelled glycerol on a sediment sample of the

Wadden Sea to enrich by 2.1- to 3.8-fold DNA fragments that carried genes encoding coenzyme B12-dependent glycerol dehydratases.

Galbraith et al. (2004) have described a novel method based on suppressive subtractive hybridization to specifically isolate from one metagenomic DNA sample (designated the ‘tester’) fragments that are absent in another metagenomic DNA sample (the ‘driver’). Applied to the study of microbial populations in the rumen, this approach revealed an unexpectedly large difference in archaeal community structure between steers fed identical diets. It should be noted that there are several drawbacks to this method: e.g. large amounts of subtractive sequences must be read to achieve an accurate sense of the degree of genetic diversity between two samples. However, the technique is very suitable for comparative purposes, able to zoom in on relatively small differences, and allows to possibly link those differences in genetic diversity to physical, chemical or biological dissimilarities between environmental samples.

Direct all-or-none selection for clones of interest

A very powerful instrument in the metagenomic toolbox is the use of conditional survival of the metagenomic host through functional complementation. In essence, one exploits the fact that the host that is being used needs a particular type of gene or gene cluster from the pool of metagenomic DNA fragments in order to survive an imposed condition. This concept has been applied by several groups. For example, Li et al. (2005) screened a metagenomic library for clones that could correct tryptophan auxotrophy in hosts *E. coli* and *R. leguminosarum*, and in doing so identified several different *trp* operons. Gabor et al. (2004b) were successful in finding metagenomic clones that complemented the leucine auxotrophy of the host strain *E. coli* TOP10 host by expression of amidase activity on a medium containing phenylacetyl-L-leucine or D-phenylglycine-L-leucine as the sole source of leucine. Similarly, Entcheva et al. (2001) used a biotin-auxotroph *E. coli* strain to pick up several clones with biotin biosynthesis operons.

The major advantage of this approach of complementation is its all-or-none character: only if a gene is present and expressed will the host survive, so that false-positives can be expected to be rare. False-negatives, on

the other hand, may be more frequent, depending on the host or range of hosts that is being used. Li et al. (2005) identified at least one set of *trp* genes that complemented *R. leguminosarum* but not *E. coli*, while Entcheva et al. (2001) noted that all of the biotin operons that were recovered with *E. coli* as a host had highest similarity to similar operons in *Enterobacteriaceae*. Thus, host choice greatly determines the success rate of this strategy in finding genes and gene functions of interest.

Isolation of metagenomic RNA

A final variation on the classical metagenomic theme starts with the isolation of RNA, not DNA, from the environment under study, followed by reverse transcription of this RNA, and cloning of the resulting cDNA. This approach can offer answers not to the question ‘who is out there?’ but to ‘who is active out there?’. It is a technically challenging approach, mostly because of RNA instability. Working protocols are available for the isolation of environmental prokaryotic RNA (Hurt et al. 2001) and cloning of short cDNA sequences (Poretzky et al. 2005). Grant et al. (2006) recently described procedures for stabilizing eukaryotic RNA in environmental samples in the field such that they can be transported back to the laboratory, the RNA isolated, and cDNA libraries made for subsequent sequencing and expression studies. Mills et al. (2004) were able to determine the composition of the metabolically-active fraction of microbial communities in marine sediments by rRNA extraction and reverse-transcription to obtain clonable complementary 16 S ribosomal DNA.

Metagenomic stories of success and words of caution

The present excitement about metagenomics has its roots in a number of clear success stories (Table 1). One is the discovery of a member of a new class of rhodopsins, encoded on a BAC-insert from an uncultivated marine α -Proteobacterium (Béjà et al. 2000b). Proteorhodopsin is a retinal-binding bacterial integral membrane protein that acts as a light-driven proton pump in surrogate host *E. coli* (Béjà et al. 2001). Subsequent studies have revealed the abun-

Table 1 A selection of success stories and milestones in the field of metagenomics

Origin of metagenomic DNA	Total DNA cloned/sequenced	Average insert size (kb)	Type of screening (sequence-based, activity, shotgun sequencing)	Number of clones screened/sequenced	Number of positive clones (%) / sequence assembly information	Success and/ or milestone	Reference
Concentrated picoplankton (2×10^{12} cells from 8 m^3 of Pacific Ocean water)	$1-2 \times 10^8$ kb in 10^7 clones	10–20	Sequenced-based (Southern hybridization and PCR) for ribosomal RNA genes	3.2×10^4	38 (0.1%)	First real application of the metagenomics principles	Schmidt et al. 1991
Two thermophilic, anaerobic digesters	>24 Mbp	1.6–10.5	Screening for a) carboxymethylcellulase, b) 4-methylumbelliferyl β -D-cellobiopyranoside activity	15,000	(a) 12 (0.08%), (b) 11 (0.07%)	First example of screening metagenomic libraries for activity	Healy et al. 1995
soil from a meadow, sugar beet field, and valley of the river Nieme (Germany)	>4.65 Mbp	5–8	Screening of <i>E. coli</i> library clones on tetrazolium indicator plates containing 4-hydroxybutyrate	930,000	5 (0.0005%)	First report on the isolation of a biocatalyst from soil-derived libraries	Henne et al. 1999
Soil from the West Madison Agricultural Research Station in Madison, WI	100 Mbp in library SL1, 1 Gbp in library SL2	27–45	(a) PCR screening for rRNA genes, (b) DNase, lipase, amylase and antibacterial activity, (c) hemolytic activity, (d) cellulase, chitinase, esterase, keratinase, protease, hemolytic activity/ siderophore production	(a) 3,648 (SL1), (b) 3,648 (SL1), (c) 24,576 (SL2), (d) 3,648 (SL1)	(a) 7 (0.2%), (b) 12 total (0.3%); antibacterial clone SL1-36C7 was completely sequenced, (c) 29 (0.1%), (d) 0 (<0.03%)	Key paper by the group that gave the field its name (Handelsman et al. 1998), demonstrating the enormous potential of large-insert metagenomic libraries for finding new activities	Rondon et al. 2000
Unspecified	Unspecified	Unspecified	Screening of library clones with HPLC-ESIMS to identify transformants producing novel compounds	1,020	2 (0.2%); clone 436-s4-5b1 produced 4 types of terragin and 446-s3-102g1 another type	One of the first reports of the use of a non- <i>E. coli</i> host (i.e. <i>Streptomyces lividans</i>) for screening of metagenomic libraries	Wang et al. 2000

Table 1 (continued)

Origin of metagenomic DNA	Total DNA cloned/sequenced	Average insert size (kb)	Type of screening (sequence-based, activity, shotgun sequencing)	Number of clones screened/sequenced	Number of positive clones (%) / sequenced assembly information	Success and/or milestone	Reference
Natural microbial assemblage collected from surface waters off the Californian coast	468 Mbp	60–80	PCR screening for the presence of (a) archaeal 23 S ribosomal RNA, (b) SAR11-like rRNA, and (c) bacterial 16S-ITS-23S sequences	6,240	(a) 1 (0.02%), (b) 3 (0.05%), (c) ≥ 22 ($\geq 0.4\%$); several clones were fully sequenced and on one (BAC clone 31A08) the proteorhodopsin gene was found	Discovery of a bacterial rhodopsin gene and a previously unrecognized mode of light-driven energy generation in bacteria	Béjà et al. 2000a, 2000b
Soil (source or type not specified)	Unspecified	Unspecified	Screening for growth inhibition of <i>Bacillus subtilis</i> , indicating the production of antibacterial activity	700,000	6.5 (0.009%); one clone, cosmid CSL12, was fully sequenced	First instance of isolating the genetic information for antibiotic production from a metagenomic library	Brady and Clardy 2000
Soil from a meadow, sugar beet field, and valley of the river Nieme (Germany)	>7.4 Gbp	5–8	Screening for growth rescue of an antiporter-deficient <i>E. coli</i> host strain by metagenomic DNA fragments	1,480,000	2 (0.0001%)	First example of using functional complementation in metagenomics to discover new genes (coding for antiporter activity in this case)	Majermik et al. 2001
200 l of surface seawater from S cripps Pier (SP) and the Mission Bay (MB)	Unspecified	0.7	Shotgun sequencing	1,934 sequences	3.5% of sequences could be assembled	One of the first shotgun sequencing approaches, showing tremendous, previously uncharacterized, viral diversity	Breitbart et al. 2002

Upper layer (0–5 cm) of a calcareous grassland near Darmstadt (Germany)	0.9 Gbp	32.5–43.5	PCR on pooled clones using archaea-specific 16S rDNA primers followed by colony hybridization of clones within positive pool with labeled PCR product	25,278	1 (0.004%); clone 2914 was fully sequenced	First full sequence of an archaeal metagenomic clone	Quaiser et al. 2002
Microbial biomass on the surface of the rubber-coated valves in the drinking water networks in the state of North Rhine-Westphalia (Germany)	Unspecified	1–5	Shotgun sequencing	5,000 sequences	2,496 high-quality reads representing a total of >2 Mbp sequence	First report of shotgun sequencing of a microbial community	Schmeisser et al. 2003
Biofilm growing on the surface of flowing acid mine at Iron Mountain, CA	76.2 Mbp	3.2 kb	Shotgun sequencing	103,462 reads with average length of 0.7 kb	Assembly into 10.83 Mbp	Near-complete genomes of dominant members of a simple microbial community	Tyson et al. 2004
Surface water samples from Sargasso sea	1.63 Gbp	2–6	Shotgun sequencing	2 million reads with average length of 0.8-kb	Assembly into 1 Gbp of nonredundant sequence	'Megagenomics': unprecedented sequence-based discovery of 148 new bacterial phylotypes and 1.2 million novel genes	Venter et al. 2004
Soil from the West Madison Agricultural Research Station in Madison, WI	5.4 Gbp	2.7–45	Screening for acquired resistance in host strain <i>E. coli</i> to (a) kanamycin, (b) tetracycline, apramycin, butirosin, tobramycin, and (c) nalidixic acid	1,186,200	(a) 5 (0.0004%), (b) 1 each (0.00008%), (c) 0 (<0.00008%)	Shows the use of positive selection to isolate rare gene activities in a metagenomic library	Riesenfeld et al. 2004a

Table 1 (continued)

Origin of metagenomic DNA	Total DNA cloned/sequenced	Average insert size (kb)	Type of screening (sequence-based, activity, shotgun sequencing)	Number of clones screened/sequenced	Number of positive clones (%) / sequence assembly information	Success and/or milestone	Reference
Surface soil (0–10 cm) from a farm in Waseca County, MN	100 Mbp	2.4 kb	Shotgun sequencing	149,085 reads	100% of reads unassembled	Metagenomic sequencing of complex microbial community (soil); first demonstration of quantitative comparative metagenomics	Tringe et al. 2005
Unspecified forest soil	>23 Mbp	10–30	Screening by hybridization to <i>pmoA</i> gene (particulate methane monooxygenase)	2,300	1 (0.04%)	Proof-of-concept for the use of stable isotope probing to generate metagenomic library from specific subpopulation	Dumont et al. 2006
Sub-surface banded iron formation from Soudan Iron Mine in Minnesota	70 Mbp	n/a	Shotgun pyrosequencing	723,013 reads with average length of 0.1 kb	No assembly	First application of pyrosequencing to environmental samples	Edwards et al. 2006
Fecal material from two healthy human subjects	Unspecified	2–3	Shotgun sequencing	139,521 reads with average length of 0.9 kb	33.8 Gbp assembled sequence plus 45 Gbp singletons	Metagenomic analysis of the human distal gut microbiome	Gill et al. 2006

dance and diversity of this type of protein in ocean waters (Venter et al. 2004). This discovery clearly illustrates the utility of metagenomics to reveal unsuspected biological functions. Another story of success is the Sargasso Sea project (Venter et al. 2004). It demonstrated, unambiguously, the muscle of metagenomics: by sheer brute force, 1 Gbp of nonredundant sequence was generated, two to three orders of magnitude more than the human genome (Venter et al. 2001). From <2 cubic meters of ocean water, an unprecedented microbial diversity was uncovered, including 148 previously unknown bacterial phylotypes and 1.2 million previously unknown genes. The Sargasso Sea study greatly stimulated the discussion on the limits and the future of metagenomics and still serves as a reference point for many of the environmental shotgun sequencing projects that followed in its footsteps. A third tale of metagenomic fame is that of the acid mine drainage biofilm (Tyson et al. 2004). It actually makes several interesting points, most of which are a direct consequence of the low-level complexity of the community under study. Due to the latter, the genomes of two members in the consortium could be reconstructed to near completion, enabling a true systems-biology approach to studying the interplay of microbial metabolism and mineral dissolution (Allen and Banfield 2005). A proteomic analysis of this community showed that biofilm polymer production and nitrogen fixation appeared to be partitioned among community members and that an abundant cytochrome-like protein might be essential to the production of acid mine drainage (Ram et al. 2005). The metagenome sequence also revealed single-nucleotide polymorphisms which prompted addressing the biofilm community in terms of population genomics and evolutionary ecology (Whitaker and Banfield 2006). Furthermore, based on clues from its metagenomic sequence, one of the previously unculturable consortium members, *Leptospirillum ferrodiazotrophum*, could be grown in the laboratory (Tyson et al. 2005), reducing the number of ‘unculturable’ bacteria by one.

Many of the stories of metagenomic success also contain important words of caution. In hindsight and in words of DeLong (Sreenivasan 2001), the discovery of the proteorhodopsin gene was to a large degree a matter of serendipity, or luck. Since it was one of the very first successful cases to demonstrate the ability of metagenomics to link phylogeny and function, the bar was put high for any future

endeavour in this direction. In practice, while the chance of finding novel genes on any metagenomic fosmid- or BAC-insert is not very low, it is much less probable to find a gene coding for a never-seen-before prokaryotic life-style such as is the capture of light for energy. Unrealistic expectations may also arise from a false sense of probability of finding a particular gene of interest in a metagenome library. Even without considering potential problems of expressing heterologous DNA in a surrogate host or insufficient homology to identify clones using PCR or hybridization, it remains a laborious and often underestimated task to screen through many thousands of clones, especially in the case of metagenomic DNA from environments with high microbial complexity. For example, it took 1,186,200 clones containing a collective 5.4 Gbp of soil DNA to identify nine unique clones that conferred resistance to aminoglycoside antibiotics and one clone expressing resistance to tetracycline (Riesenfeld et al. 2004a). A hit rate of one interesting gene per 5–5,000 Mbp of cloned metagenomic DNA is not uncommon for activity-based screenings (Lorenz and Eck 2005) and should probably be considered a normal operating range in any metagenomic screening.

A final consideration in this section is the recent argument (Oremland et al. 2005) that the predictive and interpretative value of metagenomics is limited by the validity of database entries. Many genes are listed with unknown function and for many genes that have a function assigned based on sequence similarity, this function still needs to be validated experimentally. In this context, it is of extreme value to have culturable representatives available from the environment under study. The Sargasso Sea study and other massive-scale shotgun projects clearly show the rapid progress in sequencing capacity, but do not contribute greatly to our ability to assign gene function. Metagenomics is more than a descriptive science; it should be appreciated as a methodology that is complementary to conventional approaches in testing hypotheses on the composition, diversity and functionality of microbial communities.

Metagenomics and the study of plant growth-promoting rhizobacteria

It was mentioned earlier in the introduction that the first and most important step in a metagenomic

approach is the isolation of DNA from the environment under study. This is not quite right: it is probably more important to first ask whether it is warranted or wise to invest in a metagenomics approach at all and for what purpose. The remainder of this article will be an attempt to address this question in general terms for the study of plant growth-promoting rhizobacteria (PGPR).

The term PGPR (Kloepper and Schroth 1978) refers to those plant root ('rhizosphere')-associated bacteria that are capable of stimulating plant growth, e.g. by improving plant nutrition, by the production of plant growth regulators or by preventing the attack of pathogenic microorganisms. PGPR vary in their degree of intimacy with the plant, from intracellular, i.e. existing inside root cells, to extracellular, i.e. free-living in the rhizosphere (Gray and Smith 2005). Some PGPR are commercially available as inoculants and have applicability for example in agriculture, forest regeneration, and phytoremediation of soils (Lucy et al. 2004).

There is a clear potential for metagenomics to contribute to the study of microbial communities of the rhizosphere, in particular PGPR. Possible contributions include (1) the discovery of novel plant-growth promoting genes and gene products, and (2) the characterization of (not-yet-)culturable PGPRs. Before discussing these in more detail below, it is worth noting that in practical terms, the application of metagenomics to PGPR in the rhizosphere greatly benefits from previous advances in DNA isolation and library construction from other environments. Rhizosphere soil poses more or less the same challenges as bulk soil, for which several metagenomic success stories have been published (Daniel 2005). Probably the biggest obstacle in the construction of a metagenomic library from rhizosphere soil DNA is the relative low availability of starting material. To 1 cm of root typically adheres only 20 mg of soil (Jacobsen 2004), so one needs (depending on the plant species under investigation) 50 to 500 cm of root material in order to apply a DNA extraction method that requires 1 to 10 g of soil. Several protocols have also been developed for the isolation of metagenomic bacterial DNA from inside plant material. For example, Jiao et al. (2006) describe an indirect method based on enzymatic hydrolysis of plant tissues to release associated microorganisms for subsequent DNA isolation and cloning. While optimized for leaves and seeds, this

method seems readily adaptable for use with root material, and thus of great use to the metagenomic exploration of microorganisms in the rhizosphere.

Novel plant growth-promoting genes and gene products

For many of the traits or mechanisms known to be plant growth-promoting (Bloemberg and Lugtenberg 2001), in vitro activity assays have been described and are, at least in theory, exploitable for gain-of-function screenings of a metagenomic library from rhizosphere DNA. For example, antibiotic activity towards (plant-pathogenic) bacteria or fungi can be assessed by testing whole-cell library clones or their extracts in a medium- or high-throughput manner for performance in confrontation assays. Many of such assays have been described using a variety of indicator strains, including some of the most important soil borne pathogens, e.g. the bacteria *Erwinia* (Emmert et al. 2004) and *Xanthomonas* (Rangarajan et al. 2003), the fungi *Fusarium* (Chin-A-Woeng et al. 1998; Kim et al. 2006), and *Rhizoctonia* (Kim et al. 2006; Rangarajan et al. 2003), and the fungus-like oomycetes *Phytophthora* (Kim et al. 2006) and *Pythium* (Kim et al. 2006; Rajendran et al. 1998). Production of the plant hormone indole 3-acetic acid (IAA) by metagenomic library clones can be measured using high-pressure liquid chromatography or colorimetric assays (Bric et al. 1991; Omer et al. 2004; Radwan et al. 2002; Leveau and Lindow 2005), while cytokinins and their metabolites are detectable in supernatants by e.g. immunoaffinity chromatography (Timmusk et al. 1999). Genes for nitrogen fixation are retrievable with the use of nitrogen-free media (Ding et al. 2005; Hashidoko et al. 2002; Tejera et al. 2005). Similarly, genes for the utilization of particular rhizosphere exudates could be recovered using an all-or-none complementation selection for growth on minimal medium containing these exudates as sole source of energy, carbon and/or nitrogen. In a similar approach, clones expressing 1-aminocyclopropane 1-carboxylate (ACC) deaminase, a plant growth-promoting enzyme that lowers plant ethylene levels (Glick et al. 1998), could be selected for by using ACC as sole source of nitrogen, as described previously (Holguin and Glick 2003; Shaharoon et al. 2006). The activities of lytic enzymes are most easily identified

through clear zones around colonies on solid media, as has been documented e.g. for several biocontrol chitinases (Basha and Ulaganathan 2002; Gohel et al. 2004; Kobayashi et al. 2002; Leveau et al. 2006). Assays based on halo formation are also available to identify PGPR-related phenotypes such as solubilization of mineral phosphate (Rodriguez et al. 2000) and siderophore production (Lee et al. 2003).

For the functional screening of library clones for PGPR functions, the use of alternative hosts seems very promising and rational, for several reasons. First, there is an abundant availability of phylogenetically diverse culturable PGPRs (Vessey 2003, Lucy et al. 2004) which could improve the probability of finding genes of interest, especially those that are not expressed in *E. coli* and whose full activity requires a specific PGPR background. The host role could also be played by several of the numerous defined mutants of PGPR that carry a knockout in one or several genes contributing to a particular PGPR phenotype. Such mutants could be useful to screen libraries for heterologous genes and gene functions by a functional complementation approach. Proof-of-principles for such an approach are available in the literature, e.g. single-gene complementation of a mutant of *Burkholderia* sp. strain PsJN in quinolinate phosphoribosyltransferase (QAPRTase) activity (Wang et al. 2006), a mutant of *Pseudomonas putida* WCS358 unable to produce the antibiotic pseudobactin 358 (Devescovi et al. 2001), and a mutant of *Pseudomonas chlororaphis* PCL1391 impaired in the production of the antifungal secondary metabolite phenazine-1-carboxamide (PCN; Girard et al. 2006). Especially interesting in this respect is the use of mutants that lack one or more genes in a multi-gene pathway for the production of antibiotics by enzymes such as polyketide synthases (Staunton and Weissman 2001) and non-ribosomal peptide synthases (Raaijmakers et al. 2006). The modularity underlying such proteins allows for a strategy of combinatorial complementation (Coeffet-Le Gal et al. 2006), possibly leading to the discovery of antimicrobials with new structures and new target specificities (Wenzel and Müller 2005).

Activity screenings such as the ones described above have the potential to retrieve never-before-seen genes with PGPR activity from the metagenomic pool. In contrast, the metagenomic harvest from sequence-based approaches such as PCR and/or Southern hybridizations will inevitably uncover only genes that match the specificity of the primers and/or

probes that were used to find them. Nevertheless, screening rhizosphere DNA for PGPR-related genes by PCR (e.g. Juraeva et al. 2006; Sato et al. 1997) or Southern hybridization (e.g. Blaha et al. 2005; Shah et al. 1998) has several advantages in a metagenomic setting. Most importantly, sequencing the flanking regions of such genes on large-insert fosmid or BAC clones could provide insight into the identity of their owner, on the genetic context of these PGPR genes and possibly on the mechanisms of their regulation.

Highly complementary to activity- and sequence-based screenings, a third approach to finding novel plant growth-promoting genes and gene functions is through comparative metagenomics. For one, the rhizosphere can be viewed as an environment that in comparison to the bulk soil is enriched in particular types of microorganisms, including PGPR. Indeed, there is ample evidence that the microbial diversity as measured by phylogenetic markers such as ribosomal RNA genes can differ dramatically between bulk and rhizosphere soil (Costa et al. 2006a; Sanguin et al. 2006). EGT fingerprinting by shotgun sequencing (Tringe and Rubin 2005) or suppressive subtractive hybridization (Galbraith et al. 2004) of bulk and rhizosphere soil compartments could reveal differences in the type of gene adaptations that each compartment selects for. It is expected that genes with PGPR-like functions would be enriched in the rhizosphere library. Similarly, comparison of the genomic diversity of disease-suppressive and non-suppressive soils (Weller et al. 2002) could expose genetic factors that contribute to or are predictive of the suppressiveness towards e.g. pathogenic microorganisms or nematodes.

Characterization of (not-yet-)culturable PGPRs

Due to a historical bias to study those microorganisms that can be grown in the laboratory, there is limited knowledge on the abundance and activity of not-yet-culturable PGPR. However, there are several examples of their existence and contribution to plant health, e.g. *Pasteuria penetrans*, a not-yet-culturable bacterium parasitic to plant-pathogenic nematodes (Fould et al. 2001), the nitrogen fixing activity by viable-but-not-culturable *Azoarcus* grass endophytes (Hurek et al. 2002), and the obligate biotrophism of arbuscular mycorrhizal (AM) fungi (Millner and

Wright 2002). Bacteria belonging to the *Acidobacteria* and *Verrucomicrobia* are in many rhizospheres among the most abundant, difficult-to-culture representatives (Buckley and Schmidt 2003; Gremion et al. 2003). However, it is not clear if and how their abundance is correlated to their contribution towards plant health. A phylogenetic anchoring approach, e.g. using the previously described I-Ceul method (Nesbø et al. 2005) to construct a library over-represented in DNA fragment harbouring 23 S rRNA genes, in combination with a PCR-based screening of this library with Acidobacterial or Verrucomicrobial primers would allow a (partial) insight into the genomes of these bacteria beyond the limited dataset that currently exists for these classes of bacteria (Liles et al. 2003; Quaiser et al. 2003; Wagner and Horn 2006) and into their possibly beneficial effect on plant growth. Major progress is being made in the development of new cultivation techniques (Joseph et al. 2003; Kaerberlein et al. 2002; Stevenson et al. 2004; Zengler et al. 2002), which offers the prospect that many more of the formulated hypotheses based on metagenomic analysis of the rhizosphere and its PGPR constituency (see below) will become testable as the number of culturable representatives from the rhizosphere steadily increases.

An analysis of the rhizosphere by comparative metagenomics holds the promise to reveal several important questions regarding the unculturable fraction of the rhizosphere community. For one, it could expose what actually constitutes this fraction from a comparison of metagenomic DNA isolated directly from rhizosphere to DNA isolated from all the colonies forming on solid media after plating from that same rhizosphere (i.e. the culturable fraction). One could expect a phylogenetic analysis of these two libraries to show differences, based on previous observations (e.g. Sliwinski and Goodman 2004; Costa et al. 2006b), and with large-scale DNA sequencing of both libraries, a start could be made to contrast the genetic diversity of the two populations. Furthermore, by comparison of the functions enriched for in a library from rhizosphere soil versus one from bulk soil, the degree of the selection in each of the compartments for particular microbial activities, specifically those with PGPR relevance, can be estimated. A shotgun sequencing approach for unlocking the unculturable diversity of rhizosphere bacteria, including PGPR, has not yet been reported. Recently, a study was published (Erkel et al. 2006),

describing the use of metagenomic sequencing to reconstruct the 3.18-Mbp genome of rice cluster I (RC-I) Archaea with origin in the rice rhizosphere. DNA for the shotgun library was isolated from a methanogenic enrichment culture using rice paddy soil as an inoculum. While RC-I Archaea do not necessarily qualify as PGPR, the study shows that shotgun sequencing in combination with a prior enrichment strategy towards an originally complex rhizosphere population allows the metagenomic analysis of rhizobacteria with a particular function of interest.

Conclusions

In summary, the tools of metagenomics offer many openings into a broadened view of the rhizosphere in general and of PGPR and their activities in particular. Several existing assays for PGPR activity have been listed here and proposed to have immediate utility for the screening of large-insert DNA libraries for gain-of-function phenotypes. The discovery of novel PGPR activities, either by functional screening or based on DNA sequence information, will add enormously to our understanding of the mechanistic variation that exists in PGPR phenotypes. It will also benefit our ability to improve existing PGPR, by adding to the pool of exploitable PGPR genes and utilization of this pool to develop PGPRs with enhanced performance (Downing and Thomson 2000; Glick and Bashan 1997; Holguin and Glick 2003; Timms-Wilson et al. 2004). The use of metagenomics in parallel with established or novel molecular approaches to the study of PGPR, such as genome sequencing of new PGPR isolates (Jeong et al. 2006) and transcriptional profiling of PGPR (Mark et al. 2005; Wang et al. 2005) will undoubtedly lead to the discovery of novel mechanisms of PGPR activity, new types of PGPR identity and a fresh look on the biology and practical application of PGPR.

References

- Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S., & Ikemura, T. (2005). Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. *DNA Research*, *12*, 281–290.
- Abulencia, C. B., Wyborski, D. L., Garcia, J. A., Podar, M., Chen, W. Q., & Chang, S. H., et al. (2006). Environmental

- whole-genome amplification to access microbial populations in contaminated sediments. *Applied and Environmental Microbiology*, *72*, 3291–3301.
- Allen, E. E., & Banfield, J. F. (2005). Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology*, *3*, 489–498.
- Amann, R. (2000). Who is out there? Microbial aspects of biodiversity. *Systematic and Applied Microbiology*, *23*, 1–8.
- Basha, S., & Ulaganathan, K. (2002). Antagonism of *Bacillus* species (strain BC121) towards *Curvularia lunata*. *Current Science*, *82*, 1457–1463.
- Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., & Nguyen, L. P., et al. (2000a). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, *289*, 1902–1906.
- Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M., & DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, *411*, 786–789.
- Béjà, O., Suzuki, M. T., Koonin, E. V., Aravind, L., Hadd, A., & Nguyen, L. P., et al. (2000b). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology*, *2*, 516–529.
- Berry, A. E., Chiocchini, C., Selby, T., Sosio, M., & Wellington, E. M. H. (2003). Isolation of high molecular weight DNA from soil for cloning into BAC vectors. *FEMS Microbiology Letters*, *223*, 15–20.
- Blaha, D., Sanguin, H., Robe, P., Nalin, R., Bally, R., & Moenne-Loccoz, Y. (2005). Physical organization of phytoeneficial genes *nifH* and *ipdC* in the plant growth-promoting rhizobacterium *Azospirillum lipoferum* 4V(I). *FEMS Microbiology Letters*, *244*, 157–163.
- Bloemberg, G. V., & Lugtenberg, B. J. J. (2001). Molecular basis of plant growth promotion and biocontrol by rhizobacteria. *Current Opinion in Plant Biology*, *4*, 343–350.
- Brady, S. F., & Clardy, J. (2000). Long-chain *n*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *Journal of the American Chemical Society*, *122*, 12903–12904.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., & Mead, D., et al. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences USA*, *99*, 14250–14255.
- Bric, J. M., Bostock, R. M., & Silverstone, S. E. (1991). Rapid in situ assay for indoleacetic-acid production by bacteria immobilized on a nitrocellulose membrane. *Applied and Environmental Microbiology*, *57*, 535–538.
- Buckley, D. H., & Schmidt, T. M. (2003). Diversity and dynamics of microbial communities in soils from agroecosystems. *Environmental Microbiology*, *5*, 441–452.
- Chen, K., & Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Computational Biology*, *1*, 106–112.
- Chin-A-Woeng, T. F. C., Bloemberg, G. V., van der Bij, A. J., van der Drift, K. M. G. F., Schripsema, J., & Kroon, B., et al. (1998). Biocontrol by phenazine-1-carboxamide-producing *Pseudomonas chlororaphis* PCL1391 of tomato root rot caused by *Fusarium oxysporum* f. sp. *radicis-lycopersici*. *Molecular Plant–Microbe Interactions*, *11*, 1069–1077.
- Coiffet-Le Gal, M. F., Thurston, L., Rich, P., Miao, V., & Baltz, R. H. (2006). Complementation of daptomycin *dptA* and *dptD* deletion mutations in trans and production of hybrid lipopeptide antibiotics. *Microbiology*, *152*, 2993–3001.
- Costa, R., Gotz, M., Mroczek, N., Lottmann, J., Berg, G., & Smalla, K. (2006a). Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. *FEMS Microbiology Ecology*, *56*, 236–249.
- Costa, R., Salles, J. F., Berg, G., & Smalla, K. (2006b). Cultivation-independent analysis of *Pseudomonas* species in soil and in the rhizosphere of field-grown *Verticillium dahliae* host plants. *Environmental Microbiology*, *8*, 2136–2149.
- Courtois, S., Cappellano, C. M., Ball, M., Francou, F. X., Normand, P., & Helyncck, G., et al. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology*, *69*, 49–55.
- Daniel, R. (2005). The metagenomics of soil. *Nature Reviews Microbiology*, *3*, 470–478.
- Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*, *27*, 4636–4641.
- DeLong, E. F. (2004). Microbial population genomics and ecology: The road ahead. *Environmental Microbiology*, *6*, 875–878.
- DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., & Frigaard, N. U., et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, *311*, 496–503.
- Deutschbauer, A. M., Chivian, D., & Arkin, A. P. (2006). Genomics for environmental microbiology. *Current Opinion in Biotechnology*, *17*, 229–235.
- Devescovi, G., Aguilar, C., Majolini, M. B., Marugg, J., Weisbeek, P., & Venturi, V. (2001). A siderophore peptide synthetase gene from plant growth-promoting *Pseudomonas putida* WCS358. *Systematic and Applied Microbiology*, *24*, 321–330.
- Ding, Y., Wang, J., Liu, Y., & Chen, S. (2005). Isolation and identification of nitrogen-fixing bacilli from plant rhizospheres in Beijing region. *Journal of Applied Microbiology*, *99*, 1271–1281.
- Downing, K., & Thomson, J. A. (2000). Introduction of the *Serratia marcescens chiA* gene into an endophytic *Pseudomonas fluorescens* for the biocontrol of phytopathogenic fungi. *Canadian Journal of Microbiology*, *46*, 363–369.
- Dumont, M. G., Radajewski, S. M., Miguez, C. B., McDonald, I. R., & Murrell, J. C. (2006). Identification of a complete methane monoxygenase operon from soil by combining stable isotope probing and metagenomic analysis. *Environmental Microbiology*, *8*, 1240–1250.
- Edwards, R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., & Peterson, D. M., et al. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics*, *7*, 57.
- Edwards, R. A., & Rohwer, F. (2005). Viral metagenomics. *Nature Reviews Microbiology*, *3*, 504–510.
- Emmert, E. A. B., Klimowicz, A. K., Thomas, M. G., & Handelsman, J. (2004). Genetics of Zwittermicin A

- production by *Bacillus cereus*. *Applied and Environmental Microbiology*, 70, 104–113.
- Entcheva, P., Liebl, W., Johann, A., Hartsch, T., & Streit, W. R. (2001). Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied and Environmental Microbiology*, 67, 89–99.
- Erkel, C., Kube, M., Reinhardt, R., & Liesack, W. (2006). Genome of Rice Cluster I archaea - the key methane producers in the rice rhizosphere. *Science*, 313, 370–372.
- Fould, S., Dieng, A. L., Davies, K. G., Normand, P., & Maitelle, T. (2001). Immunological quantification of the nematode parasitic bacterium *Pasteuria penetrans* in soil. *FEMS Microbiology Ecology*, 37, 187–195.
- Friedrich, M. W. (2006). Stable-isotope probing of DNA: Insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. *Current Opinion in Biotechnology*, 17, 59–66.
- Fuhrman, J. A., Comeau, D. E., Hagstrom, A., & Chan, A. M. (1988). Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Applied and Environmental Microbiology*, 54, 1426–1429.
- Gabor, E. M., Alkema, W. B. L., & Janssen, D. B. (2004a). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology*, 6, 879–886.
- Gabor, E. M., de Vries, E. J., & Janssen, D. B. (2004b). Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental Microbiology*, 6, 948–958.
- Galbraith, E. A., Antonopoulos, D. A., & White, B. A. (2004). Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: The rumen as a model. *Environmental Microbiology*, 6, 928–937.
- Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., & Samuel, B. S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, 312, 1355–1359.
- Girard, G., Barends, S., Rigali, S., van Rij, E. T., Lugtenberg, B. J. J., & Bloemberg, G. V. (2006). Pip, a novel activator of phenazine biosynthesis in *Pseudomonas chlororaphis* PCL1391. *Journal of Bacteriology*, 188, 8283–8293.
- Glick, B. R., & Bashan, Y. (1997). Genetic manipulation of plant growth-promoting bacteria to enhance biocontrol of phytopathogens. *Biotechnology Advances*, 15, 353–378.
- Glick, B. R., Penrose, D. M., & Li, J. P. (1998). A model for the lowering of plant ethylene concentrations by plant growth-promoting bacteria. *Journal of Theoretical Biology*, 190, 63–68.
- Gohel, V., Megha, C., Vyas, P., & Chhatpar, H. S. (2004). Strain improvement of chitinolytic enzyme producing isolate *Pantoea dispersa* for enhancing its biocontrol potential against fungal plant pathogens. *Annals of Microbiology*, 54, 503–515.
- Grant, S., Grant, W. D., Cowan, D. A., Jones, B. E., Ma, Y. H., & Ventosa, A., et al. (2006). Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. *Applied and Environmental Microbiology*, 72, 135–143.
- Gray, E. J., & Smith, D. L. (2005). Intracellular and extracellular PGPR: Commonalities and distinctions in the plant–bacterium signaling processes. *Soil Biology & Biochemistry*, 37, 395–412.
- Green, B. D., & Keller, M. (2006). Capturing the uncultivated majority. *Current Opinion in Biotechnology*, 17, 236–240.
- Gremion, F., Chatzinotas, A., & Harms, H. (2003). Comparative 16S rDNA and 16S rRNA sequence analysis indicates that *Actinobacteria* might be a dominant part of the metabolically active bacteria in heavy metal-contaminated bulk and rhizosphere soil. *Environmental Microbiology*, 5, 896–907.
- Hallam, S. J., Putnam, N., Preston, C. M., Detter, J. C., Rokhsar, D., & Richardson, P. M., et al. (2004). Reverse methanogenesis: Testing the hypothesis with environmental genomics. *Science*, 305, 1457–1462.
- Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, 68, 669–685.
- Handelsman, J. (2005). Metagenomics or megagenomics? *Nature Reviews Microbiology*, 3, 457–458.
- Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry & Biology*, 5, R245–R249.
- Hashidoko, Y., Tada, M., Osaki, M., & Tahara, S. (2002). Soft gel medium solidified with gellan gum for preliminary screening for root-associated, free-living nitrogen-fixing bacteria inhabiting the rhizosphere of plants. *Bioscience Biotechnology and Biochemistry*, 66, 2259–2263.
- Healy, F. G., Ray, R. M., Aldrich, H. C., Wilkie, A. C., Ingram, L. O., & Shanmugam, K. T. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Applied Microbiology and Biotechnology*, 43, 667–674.
- Henne, A., Daniel, R., Schmitz, R. A., & Gottschalk, G. (1999). Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Applied and Environmental Microbiology*, 65, 3901–3907.
- Holguin, G., & Glick, B. R. (2003). Transformation of *Azospirillum brasilense* Cd with an ACC deaminase gene from *Enterobacter cloacae* UW4 fused to the Tet(r) gene promoter improves its fitness and plant growth promoting ability. *Microbial Ecology*, 46, 122–133.
- Hurek, T., Handley, L. L., Reinhold-Hurek, B., & Piche, Y. (2002). *Azoarcus* grass endophytes contribute fixed nitrogen to the plant in an unculturable state. *Molecular Plant–Microbe Interactions*, 15, 233–242.
- Hurt, R. A., Qiu, X. Y., Wu, L. Y., Roh, Y., Palumbo, A. V., & Tiedje, J. M., et al. (2001). Simultaneous recovery of RNA and DNA from soils and sediments. *Applied and Environmental Microbiology*, 67, 4495–4503.
- Jacobsen, C. S. (1995). Microscale detection of specific bacterial-DNA in soil with a magnetic capture-hybridization and PCR amplification assay. *Applied and Environmental Microbiology*, 61, 3347–3352.
- Jacobsen, C. S. (2004). Purification of microbial genes from soil and rhizosphere by magnetic capture hybridization and subsequent amplification of target genes by PCR. In: G. A. Kowalchuk, F. J. De Bruijn, I. M. Head, A. D. L. Akkermans,

- J. D. Van Elsas (Eds.), *Molecular microbial ecology manual* (pp. 181–188). Dordrecht, The Netherlands: Kluwer.
- Jeong, H., Kim, J. F., Park, Y. K., Kim, S. B., Kim, C., & Park, S. H. (2006). Genome snapshot of *Paenibacillus polymyxa* ATCC 842. *Journal of Microbiology and Biotechnology*, *16*, 1650–1655.
- Jiao, J. Y., Wang, H. X., Zeng, Y., & Shen, Y. M. (2006). Enrichment for microbes living in association with plant tissues. *Journal of Applied Microbiology*, *100*, 830–837.
- Joseph, S. J., Hugenholz, P., Sangwan, P., Osburne, C. A., & Janssen, P. H. (2003). Laboratory cultivation of widespread and previously uncultured soil bacteria. *Applied and Environmental Microbiology*, *69*, 7210–7215.
- Juraeva, D., George, E., Davranov, K., & Ruppel, S. (2006). Detection and quantification of the *nifH* gene in shoot and root of cucumber plants. *Canadian Journal of Microbiology*, *52*, 731–739.
- Kaeblerlein, T., Lewis, K., & Epstein, S. S. (2002). Isolating ‘uncultivable’ microorganisms in pure culture by a simulated natural environment. *Science*, *296*, 1127–1129.
- Kim, J., Kim, E., Kang, Y., Choi, O., Park, C. S., & Hwang, I. (2006). Molecular characterization of biosynthetic genes of an antifungal compound produced by *Pseudomonas fluorescens* MC07. *Journal of Microbiology and Biotechnology*, *16*, 450–456.
- Klopper, J. W., & Schroth, M. N. (1978). Plant growth-promoting rhizobacteria in radish. In: Proceedings of the 4th International Conference on Plant Pathogenic Bacteria (pp. 879–882), INRA, Angers, France.
- Knietsch, A., Bowien, S., Whited, G., Gottschalk, G., & Daniel, R. (2003a). Identification and characterization of coenzyme B-12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Applied and Environmental Microbiology*, *69*, 3048–3060.
- Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., & Daniel, R. (2003b). Construction and screening of metagenomic libraries derived from enrichment cultures: Generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Applied and Environmental Microbiology*, *69*, 1408–1416.
- Kobayashi, D. Y., Reedy, R. M., Bick, J., & Oudemans, P. V. (2002). Characterization of a chitinase gene from *Stenotrophomonas maltophilia* strain 34S1 and its involvement in biological control. *Applied and Environmental Microbiology*, *68*, 1047–1054.
- Kowalchuk, G. A., de Bruijn, F. J., Head, I. M., Akkermans, A. D. L., & van Elsas, J. D. (2004). *Molecular microbial ecology manual*. Dordrecht, The Netherlands: Kluwer.
- Kowalchuk, G. A., Speksnijder, A. G. C. L., Zhang, K., Goodman, R. M., & van Veen, J. A. (2007). Finding the needles in the metagenome haystack. *Microbial Ecology*, *53*, 475–485.
- Kozdroj, J., & van Elsas, J. D. (2000). Application of polymerase chain reaction-denaturing gradient gel electrophoresis for comparison of direct and indirect extraction methods of soil DNA used for microbial community fingerprinting. *Biology and Fertility of Soils*, *31*, 372–378.
- Lee, E. T., Lim, S. K., Nam, D. H., Khang, Y. H., & Kim, S. D. (2003). Pyoverdinin(2112) of *Pseudomonas fluorescens* 2112 inhibits *Phytophthora capsici*, a red-pepper blight-causing fungus. *Journal of Microbiology and Biotechnology*, *13*, 415–421.
- Leveau, J. H. J., Gerards, S., de Boer, W., & van Veen, J. A. (2004). Phylogeny-function analysis of (meta)genomic libraries: Screening for expression of ribosomal RNA genes by large-insert library fluorescent in situ hybridization (LIL-FISH). *Environmental Microbiology*, *6*, 990–998.
- Leveau, J. H. J., Gerards, S., Fritsche, K., Zondag, G., & van Veen, J. A. (2006). Genomic flank-sequencing of plasmid insertion sites for rapid identification of functional genes. *Journal of Microbiological Methods*, *66*, 276–285.
- Leveau, J. H. J., & Lindow, S. E. (2005). Utilization of the plant hormone indole-3-acetic acid for growth by *Pseudomonas putida* strain 1290. *Applied and Environmental Microbiology*, *71*, 2365–2371.
- Li, Y. G., Wexler, M., Richardson, D. J., Bond, P. L., & Johnston, A. W. B. (2005). 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. *Environmental Microbiology*, *7*, 1927–1936.
- Liles, M. R., Manske, B. F., Bintrim, S. B., Handelsman, J., & Goodman, R. M. (2003). A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Applied and Environmental Microbiology*, *69*, 2684–2691.
- Liles, M. R., Williamson, L. L., Handelsman, J., & Goodman, R. M. (2004). Isolation of high molecular weight genomic DNA from soil bacteria for genomic library construction. In G. A. Kowalchuk, F. J. De Bruijn, I. M. Head, A. D. L. Akkermans, & J. D. Van Elsas (Eds.) *Molecular microbial ecology manual* (pp. 839–849). Dordrecht, The Netherlands: Kluwer.
- Lindahl, V., & Bakken, L. R. (1995). Evaluation of methods for extraction of bacteria from soil. *FEMS Microbiology Ecology*, *16*, 135–142.
- Lorenz, P., & Eck, J. (2005). Metagenomics and industrial applications. *Nature Reviews Microbiology*, *3*, 510–516.
- Lucy, M., Reed, E., & Glick, B. R. (2004). Applications of free living plant growth-promoting rhizobacteria. *Antonie van Leeuwenhoek*, *86*, 1–25.
- Luna, G. M., Dell’Anno, A., & Danovaro, R. (2006). DNA extraction procedure: A critical issue for bacterial diversity assessment in marine sediments. *Environmental Microbiology*, *8*, 308–320.
- Majernik, A., Gottschalk, G., & Daniel, R. (2001). Screening of environmental DNA libraries for the presence of genes conferring Na⁺(Li⁺)/H⁺ antiporter activity on *Escherichia coli*: Characterization of the recovered genes and the corresponding gene products. *Journal of Bacteriology*, *183*, 6645–6653.
- Marchesi, J. R., & Weightman, A. J. (2003). Comparing the dehalogenase gene pool in cultivated α -halocarboxylic acid-degrading bacteria with the environmental metagenome pool. *Applied and Environmental Microbiology*, *69*, 4375–4382.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., & Bembem, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*, 376–380.
- Mark, G. L., Dow, J. M., Kiely, P. D., Higgins, H., Haynes, J., & Baysse, C., et al. (2005). Transcriptome profiling of bacterial responses to root exudates identifies genes involved in

- microbe–plant interactions. *Proceedings of the National Academy of Sciences USA*, 102, 17454–17459.
- Martinez, A., Kolvek, S. J., Yip, C. L. T., Hopke, J., Brown, K. A., & MacNeil, I. A., et al. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Applied and Environmental Microbiology*, 70, 2452–2463.
- Meyer, F., Goesmann, A., McHardy, A. C., Bartels, D., Bekel, T., & Clausen, J., et al. (2003). GenDB – An open source genome annotation system for prokaryote genomes. *Nucleic Acids Research*, 31, 2187–2195.
- Miller, D. N., Bryant, J. E., Madsen, E. L., & Ghiorse, W. C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology*, 65, 4715–4724.
- Millner, P. D., & Wright, S. F. (2002). Tools for support of ecological research on arbuscular mycorrhizal fungi. *Symbiosis*, 33, 101–123.
- Mills, H. J., Martinez, R. J., Story, S., & Sobecky, P. A. (2004). Identification of members of the metabolically active microbial populations associated with *Beggiatoa* species mat communities from Gulf of Mexico cold-seep sediments. *Applied and Environmental Microbiology*, 70, 5447–5458.
- Nesbø, C. L., Boucher, Y., Dlutek, M., & Doolittle, W. F. (2005). Lateral gene transfer and phylogenetic assignment of environmental fosmid clones. *Environmental Microbiology*, 7, 2011–2026.
- Nocker, A., & Camper, A. K. (2006). Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. *Applied and Environmental Microbiology*, 72, 1997–2004.
- Noguchi, H., Park, J., & Takagi, T. (2006). MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research*, 34, 5623–5630.
- Omer, Z. S., Tombolini, R., Broberg, A., & Gerhardson, B. (2004). Indole-3-acetic acid production by pink-pigmented facultative methylotrophic bacteria. *Plant Growth Regulation*, 43, 93–96.
- Oremland, R. S., Capone, D. G., Stolz, J. F., & Fuhrman, J. (2005). Whither or wither geomicrobiology in the era of “community metagenomics”. *Nature Reviews Microbiology*, 3, 572–578.
- Poretzky, R. S., Bano, N., Buchan, A., LeClerc, G., Kleikemper, J., & Pickering, M., et al. (2005). Analysis of microbial gene transcripts in environmental samples. *Applied and Environmental Microbiology*, 71, 4121–4126.
- Quaiser, A., Ochsenreiter, T., Klenk, H. P., Kletzin, A., Treusch, A. H., & Meurer, G., et al. (2002). First insight into the genome of an uncultivated crenarchaeote from soil. *Environmental Microbiology*, 4, 603–611.
- Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S. C., Treusch, A. H., & Eck, J., et al. (2003). Acidobacteria form a coherent but highly diverse group within the bacterial domain: Evidence from environmental genomics. *Molecular Microbiology*, 50, 563–575.
- Raaijmakers, J. M., de Bruijn, I., & de Kock, M. J. D. (2006). Cyclic lipopeptide production by plant-associated *Pseudomonas* spp.: Diversity, activity, biosynthesis, and regulation. *Molecular Plant–Microbe Interactions*, 19, 699–710.
- Radwan, T., Mohamed, Z. K., & Reis, V. M. (2002). Production of indole-3-acetic acid by different strains of *Azospirillum* and *Herbaspirillum* spp. *Symbiosis*, 32, 39–53.
- Rajendran, N., Rompf, A., Marahel, M. A., & Jahn, D. (1998). Isolation and characterization of a transposon mutant of *Pseudomonas fluorescens* AU63 deficient in antifungal activity against *Pythium ultimum*. *Letters in Applied Microbiology*, 27, 152–157.
- Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., & Blake, R. C., et al. (2005). Community proteomics of a natural microbial biofilm. *Science*, 308, 1915–1920.
- Rangarajan, S., Saleena, L. M., Vasudevan, P., & Nair, S. (2003). Biological suppression of rice diseases by *Pseudomonas* spp. under saline soil conditions. *Plant and Soil*, 251, 73–82.
- Riesenfeld, C. S., Goodman, R. M., & Handelsman, J. (2004a). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology*, 6, 981–989.
- Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004b). Metagenomics: Genomic analysis of microbial communities. *Annual Reviews in Genetics*, 38, 525–552.
- Rodriguez, H., Gonzalez, T., & Selman, G. (2000). Expression of a mineral phosphate solubilizing gene from *Erwinia herbicola* in two rhizobacterial strains. *Journal of Biotechnology*, 84, 155–161.
- Rodriguez-Brito, B., Rohwer, F., & Edwards, R. A. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics*, 7, 162.
- Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., & Liles, M. R., et al. (2000). Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, 66, 2541–2547.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., & Rajandream, M. A. (2000). Artemis: Sequence visualization and annotation. *Bioinformatics*, 16, 944–945.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA Sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 5463–5467.
- Sanguin, H., Remenant, B., Dechesne, A., Thioulouse, J., Vogel, T. M., & Nesme, X., et al. (2006). Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities. *Applied and Environmental Microbiology*, 72, 4302–4312.
- Sato, M., Watanabe, K., Yazawa, M., Takikawa, Y., & Nishiyama, K. (1997). Detection of new ethylene-producing bacteria, *Pseudomonas syringae* pvs. *cannabina* and *sesami*, by PCR amplification of genes for the ethylene-forming enzyme. *Phytopathology*, 87, 1192–1196.
- Schloss, P. D., & Handelsman, J. (2005). Metagenomics for studying unculturable microorganisms: Cutting the Gordian knot. *Genome Biology*, 6, 229.
- Schmeisser, C., Steele, H., & Streit, W. R. (2007). Metagenomics, biotechnology with non-culturable microbes. *Applied Microbiology and Biotechnology*, 75, 955–962.
- Schmeisser, C., Stockigt, C., Raasch, C., Wingender, J., Timmis, K. N., & Wenderoth, D. F., et al. (2003). Metagenome

- survey of biofilms in drinking-water networks. *Applied and Environmental Microbiology*, 69, 7298–7309.
- Schmidt, T. M., Delong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S ribosomal-RNA gene cloning and sequencing. *Journal of Bacteriology*, 173, 4371–4378.
- Schwarz, S., Waschkwitz, T., & Daniel, R. (2006). Enhancement of gene detection frequencies by combining DNA-based stable-isotope probing with the construction of metagenomic DNA libraries. *World Journal of Microbiology & Biotechnology*, 22, 363–367.
- Shah, S., Li, J. P., Moffatt, B. A., & Glick, B. R. (1998). Isolation and characterization of ACC deaminase genes from two different plant growth-promoting rhizobacteria. *Canadian Journal of Microbiology*, 44, 833–843.
- Shaharoon, B., Arshad, M., & Zahir, Z. A. (2006). Effect of plant growth-promoting rhizobacteria containing ACC-deaminase on maize (*Zea mays* L.) growth under axenic conditions and on nodulation in mung bean (*Vigna radiata* L.). *Letters in Applied Microbiology*, 42, 155–159.
- Sliwinski, M. K., & Goodman, R. M. (2004). Comparison of crenarchaeal consortia inhabiting the rhizosphere of diverse terrestrial plants with those in bulk soil in native environments. *Applied and Environmental Microbiology*, 70, 1821–1826.
- Sreenivasan, A. (2001). Earth's anonymous inhabitants. In: Science Notes, an online magazine produced by the graduate students in the Science Communication Program at the University of California, Santa Cruz (URL to the article: <http://scicom.ucsc.edu/SciNotes/0101/bacteria.html>).
- Staunton, J., & Weissman, K. J. (2001). Polyketide biosynthesis: A millennium review. *Natural Product Reports*, 18, 380–416.
- Stevenson, B. S., Eichorst, S. A., Wertz, J. T., Schmidt, T. M., & Breznak, J. A. (2004). New strategies for cultivation and detection of previously uncultured microbes. *Applied and Environmental Microbiology*, 70, 4748–4755.
- Streit, W. R., & Schmitz, R. A. (2004). Metagenomics – The key to the uncultured microbes. *Current Opinion in Microbiology*, 7, 492–498.
- Tebbe, C. C., & Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant-DNA from bacteria and a yeast. *Applied and Environmental Microbiology*, 59, 2657–2665.
- Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., & Glockner, F. O. (2004). TETRA: A web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics*, 5, 163.
- Tejera, N., Lluch, C., Martinez-Toledo, M. V., & Gonzalez-Lopez, J. (2005). Isolation and characterization of *Azotobacter* and *Azospirillum* strains from the sugarcane rhizosphere. *Plant and Soil*, 270, 223–232.
- Timms-Wilson, T. M., Kilshaw, K., & Bailey, M. J. (2004). Risk assessment for engineered bacteria used in biocontrol of fungal disease in agricultural crops. *Plant and Soil*, 266, 57–67.
- Timmusk, S., Nicander, B., Granhall, U., & Tillberg, E. (1999). Cytokinin production by *Paenibacillus polymyxa*. *Soil Biology & Biochemistry*, 31, 1847–1852.
- Tringe, S. G., & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, 6, 805–814.
- Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., & Chang, H. W., et al. (2005). Comparative metagenomics of microbial communities. *Science*, 308, 554–557.
- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., & Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428, 37–43.
- Tyson, G. W., Lo, I., Baker, B. J., Allen, E. E., Hugenholtz, P., & Banfield, J. F. (2005). Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp nov from an acidophilic microbial community. *Applied and Environmental Microbiology*, 71, 6319–6324.
- Uchiyama, T., Abe, T., Ikemura, T., & Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*, 23, 88–93.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., & Sutton, G. G., et al. (2001). The sequence of the human genome. *Science*, 291, 1304–1351.
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., & Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304, 66–74.
- Vessey, J. K. (2003). Plant growth-promoting rhizobacteria as biofertilizers. *Plant and Soil*, 255, 571–586.
- Wagner, M., & Horn, M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Current Opinion in Biotechnology*, 17, 241–249.
- Wang, K., Conn, K., & Lazarovits, G. (2006). Involvement of quinolinate phosphoribosyl transferase in promotion of potato growth by a *Burkholderia* strain. *Applied and Environmental Microbiology*, 72, 760–768.
- Wang, G. Y. S., Graziani, E., Waters, B., Pan, W., Li, X., & McDermott, J., et al. (2000). Novel natural products from soil DNA libraries in a Streptomyces host. *Organic Letters*, 2, 2401–2404.
- Wang, Y. Q., Ohara, Y., Nakayashiki, H., Tosa, Y., & Mayama, S. (2005). Microarray analysis of the gene expression profile induced by the endophytic plant growth-promoting rhizobacteria, *Pseudomonas fluorescens* FPT9601-T5 in *Arabidopsis*. *Molecular Plant–Microbe Interactions*, 18, 385–396.
- Ward, N. (2006). New directions and interactions in metagenomics research. *FEMS Microbiology Ecology*, 55, 331–338.
- Weller, D. M., Raaijmakers, J. M., Gardener, B. B. M., & Thomashow, L. S. (2002). Microbial populations responsible for specific soil suppressiveness to plant pathogens. *Annual Review of Phytopathology*, 40, 309–348.
- Wenzel, S. C., & Müller, R. (2005). Recent developments towards the heterologous expression of complex bacterial

- natural product biosynthetic pathways. *Current Opinion in Biotechnology*, 16, 594–606.
- Whitaker, R. J., & Banfield, J. F. (2006). Population genomics in natural microbial communities. *Trends in Ecology & Evolution*, 21, 508–516.
- Williamson, L. L., Borlee, B. R., Schloss, P. D., Guan, C. H., Allen, H. K., & Handelsman, J. (2005). Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Applied and Environmental Microbiology*, 71, 6335–6344.
- Woyke, T., Teeling, H., Ivanova, N. N., Huntemann, M., Richter, M., & Gloeckner, F. O., et al. (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature*, 443, 950–955.
- Xu, J. P. (2006). Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. *Molecular Ecology*, 15, 1713–1731.
- Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., & Short, J. M., et al. (2002). Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 15681–15686.
- Zhou, J. Z., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, 62, 316–322.