

## Metagenomic Analyses: Past and Future Trends

Carola Simon and Rolf Daniel

*Appl. Environ. Microbiol.* 2011, 77(4):1153. DOI:  
10.1128/AEM.02345-10.

Published Ahead of Print 17 December 2010.

---

Updated information and services can be found at:  
<http://aem.asm.org/content/77/4/1153>

---

*These include:*

**REFERENCES**

This article cites 148 articles, 64 of which can be accessed free at: <http://aem.asm.org/content/77/4/1153#ref-list-1>

**CONTENT ALERTS**

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), [more»](#)

---

Information about commercial reprint orders: <http://journals.asm.org/site/misc/reprints.xhtml>  
To subscribe to to another ASM Journal go to: <http://journals.asm.org/site/subscriptions/>

## MINIREVIEW

# Metagenomic Analyses: Past and Future Trends<sup>∇</sup>

Carola Simon<sup>1</sup> and Rolf Daniel<sup>1,2\*</sup>

*Abteilung Genomische und Angewandte Mikrobiologie<sup>1</sup> and Göttingen Genomics Laboratory,<sup>2</sup> Institut für Mikrobiologie und Genetik, Georg-August-Universität, Grisebachstr. 8, 37077 Göttingen, Germany*

**Metagenomics has revolutionized microbiology by paving the way for a cultivation-independent assessment and exploitation of microbial communities present in complex ecosystems. Metagenomics comprising construction and screening of metagenomic DNA libraries has proven to be a powerful tool to isolate new enzymes and drugs of industrial importance. So far, the majority of the metagenomically exploited habitats comprised temperate environments, such as soil and marine environments. Recently, metagenomes of extreme environments have also been used as sources of novel biocatalysts. The employment of next-generation sequencing techniques for metagenomics resulted in the generation of large sequence data sets derived from various environments, such as soil, the human body, and ocean water. Analyses of these data sets opened a window into the enormous taxonomic and functional diversity of environmental microbial communities. To assess the functional dynamics of microbial communities, metatranscriptomics and metaproteomics have been developed. The combination of DNA-based, mRNA-based, and protein-based analyses of microbial communities present in different environments is a way to elucidate the compositions, functions, and interactions of microbial communities and to link these to environmental processes.**

The total number of microbial cells on Earth is estimated to be  $10^{30}$  (126). Prokaryotes represent the largest proportion of individual organisms, comprising  $10^6$  to  $10^8$  separate genospecies (112). The genomes of these mainly uncultured species encode a largely untapped reservoir of novel enzymes and metabolic capabilities. Metagenomics bypasses the need for isolation or cultivation of microorganisms. Metagenomic approaches based on direct isolation of nucleic acids from environmental samples have proven to be powerful tools for comparing and for exploring the ecology (11) and metabolic profiling of complex environmental microbial communities (20, 125), as well as for identifying novel biomolecules by use of libraries constructed from isolated nucleic acids (18, 27, 42, 108, 116). In 1985, Pace et al. (84) were the first to propose the direct cloning of environmental DNA. This approach was used for cloning of DNA from picoplankton in a phage vector for subsequent 16S rRNA gene sequence analyses (103). The first successful function-driven screening of metagenomic libraries, termed zooblibraries by the authors, was conducted by Healy et al. (47).

The construction of metagenomic libraries and other DNA-based metagenomic projects are initiated by isolation of high-quality DNA that is suitable for cloning and covers the microbial diversity present in the original sample. DNA isolation, especially from extreme environments, is still a technological challenge. Reasons for this include the reluctance of many microorganisms present in these samples to lyse by protocols that have been developed mainly for DNA extraction from

mesophilic samples and the release of very stable nucleases upon cell lysis. Nevertheless, significant progress has been made, and various methods allowing the isolation of high-quality DNA from a variety of environments, i.e., soil (45, 87, 134, 139), marine picoplankton (117), contaminated subsurface sediments (1), groundwater (128), hot springs and mud holes in solfataric fields (94), surface water from rivers (145), glacier ice (109), Antarctic desert soil (48), and buffalo rumens (25), have been developed.

Initially, metagenomics was used mainly to recover novel biomolecules from environmental microbial assemblages. The development of next-generation sequencing techniques and other affordable methods allowing large-scale analysis of microbial communities resulted in novel applications, such as comparative community metagenomics, metatranscriptomics, and metaproteomics (14, 111). The correlation of the comprehensive data sets derived from these approaches with environmental parameters allows us to unravel complex ecosystem functions of microbial communities.

In this review, an overview of the different applications of metagenomics and an outline of the recent advances in this fast-developing field are given.

### BIOPROSPECTING OF METAGENOMES

In principle, the techniques for the recovery of novel biomolecules from environmental samples can be divided into two main approaches: function-based and sequence-based screening of metagenomic libraries (18, 27, 42). Both screening techniques comprise the cloning of environmental DNA and the construction of small-insert or large-insert libraries (Fig. 1). Subsequently, the resulting metagenomic libraries are used to transform a host, which is in most cases *Escherichia coli* (18, 42). As significant differences in expression modes between

\* Corresponding author. Mailing address: Institut für Mikrobiologie und Genetik der Georg-August-Universität, Grisebachstr. 8, 37077 Göttingen, Germany. Phone: 49-551-393827. Fax: 49-551-3912181. E-mail: rdaniel@gwdg.de.

<sup>∇</sup> Published ahead of print on 17 December 2010.

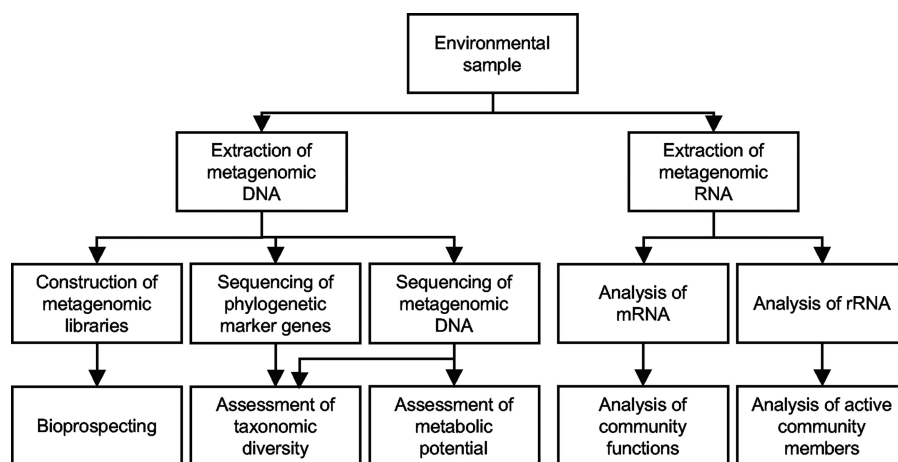


FIG. 1. Metagenomic analysis of environmental microbial communities based on nucleic acids.

different taxonomic groups of prokaryotes exist and only 40% of the enzymatic activities may be detected by random cloning in *E. coli* (32), additional hosts, such as *Streptomyces* spp. (136), *Thermus thermophilus* (3), *Sulfolobus solfataricus* (2), and diverse *Proteobacteria* (17), have been employed to expand the range of detectable activities in metagenomic screens.

Depending on the desired insert size, metagenomic libraries have been constructed using plasmids (up to 15 kb), fosmids, cosmids (both up to 40 kb), or bacterial artificial chromosomes (>40 kb) as vectors. The choice of the vector system depends on the DNA quality, targeted genes, and screening strategy (18). Small-insert libraries can be employed for the identification of novel biocatalysts encoded by a single gene or a small operon, whereas large-insert libraries are required to recover large gene clusters, which code for complex pathways (18). Construction and screening of both types of metagenomic libraries have resulted in the identification of many novel biocatalysts, e.g., lipases/esterases (15, 26, 48, 49), cellulases (25, 47), chitinases (50), DNA polymerases (109), proteases (139), and antibiotics (95). To date, lipases/esterases are probably the biocatalysts which have been most frequently recovered from metagenomes.

In the following, a short overview of the two screening approaches, including recent examples, is given (for a detailed list, see references 27 and 107).

### FUNCTION-BASED SCREENING

Most of the screens for the isolation of genes encoding novel biomolecules are based on the metabolic activities of metagenomic-library-containing clones. As sequence information is not required, this is the only strategy that bears the potential to identify entirely novel classes of genes encoding known or novel functions (18, 27, 38, 42, 96). Three different function-driven approaches have been used to recover novel biomolecules: phenotypical detection of the desired activity (7, 38, 68), heterologous complementation of host strains or mutants (13, 95, 109, 137), and induced gene expression (128, 129, 141).

In most cases, phenotypical detection employs chemical dyes and insoluble or chromophore-bearing derivatives of enzyme substrates incorporated into the growth medium, where they

register the specific metabolic capabilities of individual clones. (27). A recent example of such an activity-driven screen targeted genes encoding bacterial  $\beta$ -D-glucuronidases, which are part of the human intestinal microbiome. These enzymes have putatively beneficial effects on human health (38). A metagenomic library comprising 4,600 clones derived from bacterial DNA extracted from pools of feces was screened using an *E. coli* strain which is deficient in  $\beta$ -D-glucuronidase activity. In this way, 19 positive clones, of which one exhibited strong  $\beta$ -D-glucuronidase activity after cloning of the corresponding gene into an expression vector, were detected (38). Another example for an activity-based screen by direct detection of phenotypes was published by Belouqui et al. (7). In order to identify novel glycosyl hydrolases, *E. coli* clones harboring metagenomic fosmid libraries derived from cellulose-depleting microbial communities of a fresh cast of earthworms were screened for their ability to hydrolyze *p*-nitrophenyl- $\beta$ -D-glucopyranoside and *p*-nitrophenyl- $\alpha$ -L-arabinopyranoside. Two of the recovered glycosyl hydrolases had no similarity to any known glycosyl hydrolases and represented two novel families of  $\beta$ -galactosidases/ $\alpha$ -arabinopyranosidases.

A different category of function-driven screens is based on heterologous complementation of host strains or mutants of host strains which require the targeted genes for growth under selective conditions. This technique allows a simple and fast screening of complex metagenomic libraries comprising millions of clones. Since almost no false positives occur, this approach is highly selective for the targeted genes (109). A recent example is the complementation-based screen of 446,000 clones containing a soil-derived metagenomic library for genes that confer resistance to tetracycline,  $\beta$ -lactams, or aminoglycoside antibiotics, which resulted in the identification of 13 different antibiotic-resistant clones (24a). Further examples for screens employing heterologous complementation include the identification of genes encoding lysine racemases (13), antibiotic resistance (21, 95), enzymes involved in poly-3-hydroxybutyrate metabolism (137), DNA polymerases (109), and  $\text{Na}^+/\text{H}^+$  antiporters (71).

In 2005, Uchiyama et al. (128) introduced a third type of activity-driven screen, which was termed substrate-induced

gene expression screening (SIGEX). This high-throughput screening approach employs an operon trap *gfp* expression vector in combination with fluorescence-activated cell sorting. The screen is based on the fact that catabolic-gene expression is induced mainly by specific substrates and is often controlled by regulatory elements located close to catabolic genes (128). To perform SIGEX, metagenomic DNA is cloned upstream of the *gfp* gene, thereby placing the expression of *gfp* under the control of promoters present in the metagenomic DNA. Clones influencing *gfp* expression upon addition of the substrate of interest are isolated by fluorescence-activated cell sorting (128). In this way, Uchiyama et al. (128) isolated aromatic-hydrocarbon-induced genes from a metagenomic library derived from groundwater. One drawback of this approach is the possible activation of transcriptional regulators by effectors other than the specific substrates (33).

A similar type of screen, designated metabolite-regulated expression (METREX), has been published by Williamson et al. (141). To identify metagenomic clones producing small molecules, a biosensor that detects small diffusible signal molecules that induce quorum sensing is inside the same cell as the vector harboring a metagenomic DNA fragment. The main component of the biosensor is a quorum-sensing promoter which controls the reporter *gfp* gene. When a threshold concentration of the signal molecule encoded by the metagenomic DNA fragments is exceeded, green fluorescent protein (GFP) is produced. Subsequently, positive clones are identified by fluorescence microscopy (141). Recently, Guan et al. (41) identified a new structural class of quorum-sensing inducers from the midgut microbiota of gypsy moth larvae by employing METREX. A monooxygenase homolog which produced small molecules that induced the activities of LuxR from *Vibrio fischeri* and CviR from *Chromobacterium violaceum* was detected.

In 2010, Uchiyama and Miyazaki (129) introduced another screen based on induced gene expression, termed product-induced gene expression (PIGEX). In this reporter assay system, enzymatic activities are also detected by the expression of *gfp*, which is triggered by product formation. In order to screen for amidases, the benzoate-responsive transcriptional regulator BenR is used as a sensor. Recombinant *E. coli* strains harboring the sensor and 96,000 metagenomic clones derived from activated sludge were cocultured in microtiter plates in the presence of the substrate benzamide. In response to benzoate production by the metagenomic clones, the sensor cells fluoresced. In this way, three novel genes encoding amidases were identified.

### SEQUENCE-BASED SCREENING

The application of sequence-based approaches involves the design of DNA probes or primers which are derived from conserved regions of already-known genes or protein families. In this way, only novel variants of known functional classes of proteins can be identified. Nevertheless, this strategy has led to the successful identification of genes encoding novel enzymes, such as dimethylsulfoniopropionate-degrading enzymes (131), dioxygenases (80, 118, 148), nitrite reductases (6), [Fe-Fe]-hydrogenases (102), [NiFe] hydrogenases (75), hydrazine oxidoreductases (67), chitinases (50), and glycerol dehydratases (61).

For example, Bartossek et al. (6) detected genes encoding homologs of copper-dependent nitrite reductases (NirK) in ammonia-oxidizing archaea derived from different environments, such as soil, sediment, freshwater, chicken manure, and invertebrates, by using a PCR-based approach. Based on deduced amino acid sequences of NirK proteins from bacteria and two archaeal homologs, different sets of degenerated primers for the amplification of *nirK*-related genes from archaea were designed and used for amplification. In this way, the authors demonstrated that archaeal nitrite reductases are ubiquitous and contribute to the global biogeochemical nitrogen cycle. In order to analyze the diversity and abundance of herbicide-degrading dioxygenases encoded by *tfdA*-like genes in soil, a quantitative kinetic PCR assay was designed by Zaprasis et al. (148). A total of 437 *tfdA*-like sequences were identified by employing five different primer sets, which targeted conserved regions of *tfdA*. Approximately  $1.0 \times 10^6$  to  $65 \times 10^6$  copies of novel *tfdA*-like genes per gram of dry soil were calculated. This indicated the presence of unknown herbicide-degrading dioxygenases in soil (148).

In order to gain comprehensive insights into the available sequence space of the genes of interest, PCR-based screening approaches have been combined with large-scale pyrosequencing of amplicons. The thereby-collected sequence information can subsequently be used to design probes which are suitable to recover full-length versions of the target genes. This approach was introduced by Iwai et al. (54). The authors termed it gene-targeted metagenomics (GT-metagenomics). It was applied to recover genes encoding aromatic dioxygenases from polychlorinated-biphenyl-contaminated soil samples. The authors employed a PCR primer set that was directed against a 524-bp conserved region which confers substrate specificity to biphenyl dioxygenases. Totals of 2,000 and 604 sequences were retrieved from the 5' and 3' ends of the PCR products, respectively. Based on alignments, the sequences were assigned to 22 (5'-end) and 3 (3'-end) novel clusters that did not include previously known sequences. Thus, a larger variety of genes putatively involved in the carbon cycle were detected than previously assumed (54). To assess the diversity of bacterial genes involved in the demethylation of dimethylsulfoniopropionate in marine environments, a similar approach has been applied. Varaljay et al. (131) amplified *dmdA* genes, encoding dimethylsulfoniopropionate demethylase, from composite free-living coastal bacterioplankton DNA by employing different primer pairs which targeted 10 different clades and subclades of DmdA. Subsequently, the sequences of the resulting PCR products were determined by pyrosequencing. With >90% nucleotide sequence identity, approximately 62,000 sequences were assigned to more than 700 clusters of environmental *dmdA* sequences.

### MINING OF METAGENOMES FROM EXTREME ENVIRONMENTS

To date, the majority of biomolecules are derived from metagenomic libraries which have been constructed from temperate soil samples (70, 111). However, extreme environments, such as solfataric hot springs (94), Urania hypersaline basins (27), glacier soil (147), glacial ice (109), and Antarctic/Arctic soil (15, 48, 55), represent an almost untapped reservoir of novel

biomolecules with biotechnologically valuable properties. Although the diversity of microbial communities present in most extreme habitats is likely to be low, these environments are nevertheless an interesting source for novel biocatalysts that are active under extreme conditions (116).

Recently, a number of metagenomic libraries derived from the above-mentioned extreme habitats have been constructed. The majority of these libraries have been mined for novel lipases/esterases. Rhee et al. (94) constructed large-insert fosmid libraries from environmental samples originating from solfataric hot springs in Indonesia. Function-driven screening resulted in the identification of a novel esterase, which was classified as a new member of the hormone-sensitive lipase family. This enzyme exhibited a high temperature optimum and high thermal stability. Additionally, Ferrer et al. (28) constructed a metagenomic library derived from the brine seawater interface of Urania hypersaline basins. Five novel esterases which showed no significant amino acid sequence similarity to known esterases were identified. All of these enzymes displayed habitat-specific properties, such as a preference for high hydrostatic pressure and salinity.

Samples of an extreme environment were also used to isolate the first metagenome-derived DNA-modifying enzymes by a function-based approach. Small-insert and large-insert metagenomic libraries derived from glacier ice were constructed (109). An *E. coli* mutant that carries a cold-sensitive lethal mutation in the 5'-3' exonuclease domain of the DNA polymerase I was employed as a host for the metagenomic libraries. Only recombinant *E. coli* strains complemented by a gene conferring DNA polymerase activity were able to grow. Nine novel DNA polymerases or domains typical of these enzymes were identified and exhibited only weak similarities to known genes.

#### ASSESSMENT OF TAXONOMIC AND FUNCTIONAL DIVERSITY OF MICROBIAL COMMUNITIES

Microbial diversity in environments such as soil, sediment, or water has been assessed by analysis of conserved marker genes, e.g., 16S rRNA genes (72, 77). In addition, large databases of reference sequences, such as Greengenes (22), SILVA (92), or Ribosomal Database Project II (RDP II) (16), provide an important and useful resource for rRNA gene-based classification of microorganisms. In addition, other conserved genes, such as *recA* or *radA* and genes encoding heat shock protein 70, elongation factor Tu, or elongation factor G (132), have been employed as markers for phylogenetic analyses.

The employment of next-generation sequencing technologies, such as pyrosequencing of 16S rRNA gene amplicons, provided unprecedented sampling depth compared to traditional approaches, such as denaturing gradient gel electrophoresis (DGGE) (81), terminal restriction fragment length polymorphism (T-RFLP) analysis (29, 123), or Sanger sequencing of 16S rRNA gene clone libraries (113). However, the intrinsic error rate of pyrosequencing may result in the overestimation of rare phylotypes. Each pyrosequencing read is treated as a unique identifier of a community member, and correction by assembly and sequencing depth, which is typically applied during genome projects, is not feasible (51, 64).

To assess microbial community composition, the rRNA

gene-based approach employed is increasingly complemented or replaced by shotgun sequencing of microbial community DNA (Fig. 1). Direct sequencing of metagenomic DNA has been proposed to be the most accurate approach for assessment of taxonomic composition (135). The major advantage of this approach is the avoidance of bias introduced by amplification of phylogenetic marker genes. In 2004, two landmark publications described the application of shotgun sequencing to assess the compositions and functions of microbial populations of an acid mine drainage biofilm (127) and the Sargasso Sea (132). In addition, large-scale sequencing of the low-diversity acid mine drainage biofilm allowed genome reconstruction of the dominant bacterial species (127). Nearly complete genomes of *Leptospirillum* group II and *Ferroplasma* type II organisms and partial genomes of three other microorganisms were recovered. In addition, the reconstruction of main metabolic pathways provided insights into survival strategies of microbes living in an extreme environment.

The introduction of next-generation sequencing platforms, such as the Roche 454 sequencer (73), the SOLiD system of Applied Biosystems (9), and the Genome Analyzer of Illumina, had a big impact on metagenomic research (9). The advances in throughput and cost reduction have increased the number and size of metagenomic sequencing projects, such as the *Sorcerer II* Global Ocean Sampling (GOS) project (12, 98) and the metagenomic comparison of 45 distinct microbiomes and 42 viromes (24). The analysis of the resulting large data sets allowed the exploration of the taxonomic and functional biodiversity and of the system biology of diverse ecosystems (111).

A crucial step in the taxonomic analysis of large metagenomic data sets is called binning. Within this step, the sequences derived from a mixture of different organisms are assigned to phylogenetic groups according to their taxonomic origins. Depending on the quality of the metagenomic data set and the read length of the DNA fragments, the phylogenetic resolution can range from the kingdom to the genus level (146). Currently, two broad categories of binning methods can be distinguished: similarity-based and composition-based approaches. The similarity-based approaches classify DNA fragments based on sequence homology, which is determined by searching reference databases using tools like the Basic Local Alignment Search Tool (BLAST) (52, 78). Examples of bioinformatic tools employing similarity-based binning are the Metagenome Analyzer (MEGAN) (52), CARMA (62), or the sequence ortholog-based approach for binning and improved taxonomic estimation of metagenomic sequences (Sort-ITEMS) (79). CARMA assigns environmental sequences to taxonomic categories based on similarities to protein families and domains included in the protein family database (Pfam) (30), whereas MEGAN and Sort-ITEMS classify sequences by performing comparisons against the NCBI nonredundant and NCBI nucleotide databases (101). One pitfall of these approaches is that taxonomic classification of the metagenomic data sets relies on the use of reference databases that contain sequences of known origin and gene function. To date, the common databases are biased toward model organisms or readily cultivable microorganisms. This is a major limitation for taxonomic classification of microbial communities in ecosystems, as up to 90% of the sequences of a metagenomic data

set may remain unidentified due to the lack of a reference sequence (53). In contrast, composition-based binning methods analyze intrinsic sequence features, such as GC content (10), codon usage (5), or oligonucleotide frequencies (59, 100), and compare these features with reference genome sequences of known taxonomic origins. Tools such as PhyloPythia (76), TETRA (121, 122), and the taxonomic composition analysis method (TACO) (23) allow direct classification of short single reads.

Recently, Web-based metagenomic annotation platforms, such as the metagenomics RAST (mg-RAST) server (78), the IMG/M server (74), or JCVI Metagenomics Reports (METAREP) (39) have been designed to analyze metagenomic data sets. Via generic interfaces, the uploaded environmental data sets can be compared to both protein and nucleotide databases, such as the Gene Ontology (GO) database (4), the Clusters of Orthologous Groups (COG) database (120), and the Pfam (30), NCBI (101), SEED (83), and Kyoto Encyclopedia of Genes and Genomes (KEGG) (57) databases. In this way, multiple metagenomic data sets derived from various environments can be compared at various functional and taxonomic levels (39). Recent examples of metagenomic surveys of whole microbial communities include those studying the hindgut microbiota of a wood-feeding higher termite (138), glacier ice (110), sludge communities subjected to enhanced biological phosphorus removal (34), a biogas plant microbial community (63), and Minnesota farm soil (125).

### METATRANSCRIPTOMICS

Metagenomics provides information on the metabolic and functional capacity of a microbial community. However, as metagenomic DNA-based analyses cannot differentiate between expressed and nonexpressed genes, it fails to reflect the actual metabolic activity (114). Recently, sequencing and characterization of metatranscriptomes have been employed to identify RNA-based regulation and expressed biological signatures in complex ecosystems (Fig. 1) (46). So far, metatranscriptomic studies of microbial assemblages *in situ* are rare. This is due to difficulties associated with the processing of environmental RNA samples (149). Technological challenges include the recovery of high-quality mRNA from environmental samples (99), short half-lives of mRNA species (91), and separation of mRNA from other RNA species (114).

Until recently, metatranscriptomics had been limited to the microarray/high-density array technology (46, 86, 140) or analysis of mRNA-derived cDNA clone libraries (90, 119). These approaches have produced significant insights into the gene expression of microbial communities but have limitations. A microarray gives information about only those sequences for which it was designed. The detection sensitivities are not equal for all imprinted sequences, as results are dependent on the chosen hybridization conditions. Low-abundance transcripts are often not detected. Although transcript cloning avoids some of these problems through random amplification and sequestering of mRNA fragments, it introduces other biases associated with the cloning system and the host of the libraries. The limitations of both approaches can be circumvented by application of direct cDNA sequencing employing next-generation sequencing technologies. This provides affordable access

to the metatranscriptome and allows whole-genome expression profiling of a microbial community. In addition, direct quantification of the transcripts is feasible (31, 66, 91, 106, 130). Leininger et al. (66) were the first employing pyrosequencing to unravel active genes of soil microbial communities. In this way, the activity and importance of ammonia-oxidizing archaea in soil ecosystems have been shown (66). Other metatranscriptomic studies employing direct sequencing of cDNA have targeted the ocean surface waters from the North Pacific subtropical gyre (31), coastal waters of a fjord close to Bergen, Norway (36, 91, 106), a phytoplankton bloom in the Western English Channel (37), and soil samples from a sandy lawn (130).

Recently, Shi et al. (106) showed the involvement of small RNAs (sRNAs) in many environmental processes, such as carbon metabolism and nutrient acquisition, by comparison of metatranscriptomic data sets from the Hawaii Ocean Time-series station ALOHA (58). They found that a large proportion of cDNA sequences were not homologous to known genes encoding proteins. Almost a third of these unassigned cDNA sequences showed similarities to intergenic regions of microbial genomes in which sRNA molecules are encoded. Thirteen known sRNA families were identified in the metatranscriptomic data set by searching the RNA family database Rfam (35). In addition, a large fraction of the metatranscriptomic data set could not be assigned to any known sRNA family, but these unassigned reads exhibited a high nucleotide identity to intergenic regions found in microbial genome sequences. These sequences displayed characteristic conserved secondary structures and were often flanked by potential regulatory elements. This indicated the presence of so-far-unrecognized putative sRNA molecules and provided evidence for the importance of sRNAs for the regulation of microbial gene expression in response to changing environmental parameters (106).

### METAPROTEOMICS

The proteomic analysis of mixed microbial communities is a new emerging research area which aims at assessing the immediate catalytic potential of a microbial community. In 2004, Wilmes and Bond (143) coined the term “metaproteomics” as a synonym for large-scale characterization of the entire protein complement of environmental microbiota at a given time point. In this landmark study, the proteins produced by a microbial community derived from activated sludge were analyzed by two-dimensional polyacrylamide gel electrophoresis and mass spectrometry. Highly expressed proteins, such as an outer membrane protein and an acetyl coenzyme A acyltransferase, were identified. These enzymes putatively originated from an uncultured polyphosphate-accumulating *Rhodocyclus* strain that was dominant in the activated sludge (143).

So far, one of the most comprehensive metaproteomic studies has been conducted by Ram et al. (93), who analyzed the gene expression, key activities, and metabolic functions of a natural acid mine drainage microbial biofilm by mass spectrometry. In this way, more than 2,000 proteins from the five most abundant microorganisms were identified. In addition, 357 unique and 215 novel proteins were detected. One highly expressed novel protein was capable of iron oxidation, a process central to acid mine drainage formation. This study and other studies provided comprehensive insights into microbial

communities that exhibited a relatively low complexity (21, 40, 69, 93), i.e., communities derived from a continuous-flow bioreactor fed with cadmium (65), activated sludge (85, 142–144), and the phyllosphere (19). In addition, metaproteomic analyses of microbial communities displaying a high complexity, such as communities present in the hindguts of termites (138), sheep rumens (124), human fecal samples (60, 133), human saliva samples (97), marine samples (56, 115), dissolved organic matter from lake and forest soil (105), and contaminated soil and groundwater (8), were carried out but at a lower resolution (for a review, see reference 104). Nevertheless, it is a daunting task to detect and identify all proteins produced by a complex environmental microbial community. Challenges for metaproteomic analyses include uneven species distribution, the broad range of protein expression levels within microorganisms, and the large genetic heterogeneity within microbial communities (104). Despite these hurdles, metaproteomics has a huge potential to link the genetic diversity and activities of microbial communities with their impact on ecosystem function.

### CONCLUSIONS

Metagenomics is one of today's fastest-developing research areas. Since 1998, when the term "metagenomics" was coined by Handelsman et al. (43), great progress has been made. In the beginning, metagenomics was driven mainly by the search for novel biomolecules in microbial communities derived from temperate environments. The development of improved DNA isolation methods, cloning strategies, and screening techniques allowed the assessment and exploiting of microbial assemblages from extreme and inhospitable environments, such as solfataric hot springs, hypersaline basins, and ice.

With the launch of next-generation sequencing techniques, more and increasingly complex environmental sequence data sets were produced, which in turn led to the development of various bioinformatic tools for the analysis and comparison of these data sets with respect to taxonomic and metabolic diversity. To date, more than 210 different metagenomes have been sequenced from a large variety of environments, such as soil, global oceans, the human gut, and feces (39). Metagenomics is now also applied to medical or forensic investigations. In 2009, the Human Microbiome Project (88) was founded. This initiative aims to map microbial communities that are associated with the human gut, mouth, skin, or vagina. In addition, extinct species, such as the woolly mammoth (89) and Neanderthals (82), have been analyzed by metagenomic approaches. To take full advantage of this large amount of information, improved integrated analysis tools and comprehensive databases are needed.

In recent years, studies of the gene expression and protein production of microbial communities emerged to complement DNA-based metagenomic analyses. Metatranscriptomics and metaproteomics are approaches that have the potential to allow us to understand the functional dynamics of microbial communities. In combination with metagenome DNA analysis, these approaches offer significant promise to advance the measurement and prediction of the *in situ* microbial responses, activities, and productivity of microbial consortia. In addition, analyses of the thereby-generated comprehensive data sets

have an unprecedented potential to shed light on ecosystem functions of microbial communities and evolutionary processes.

### ACKNOWLEDGMENTS

Financial support from the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

### REFERENCES

1. **Abulencia, C. B., et al.** 2006. Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl. Environ. Microbiol.* **72**:3291–3301.
2. **Albers, S.-V., et al.** 2006. Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl. Environ. Microbiol.* **72**:102–111.
3. **Angelov, A., M. Mientus, S. Liebl, and W. Liebl.** 2009. A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Syst. Appl. Microbiol.* **32**:177–185.
4. **Ashburner, M., et al.** 2000. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* **25**:25–29.
5. **Bailly-Bechet, M., A. Danchin, M. Iqbal, M. Marsili, and M. Vergassola.** 2006. Codon usage domains over bacterial chromosomes. *PLoS Comput. Biol.* **2**:e37.
6. **Bartossek, R., G. W. Nicol, A. Lanzen, H. P. Klenk, and C. Schleper.** 2010. Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ. Microbiol.* **12**:1075–1088.
7. **Belouqui, A., et al.** 2010. Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. *Appl. Environ. Microbiol.* **76**:5934–5946.
8. **Benndorf, D., G. U. Balcke, H. Harms, and M. von Bergen.** 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J.* **1**:224–234.
9. **Bentley, D. R.** 2006. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**:545–552.
10. **Bentley, S. D., and J. Parkhill.** 2004. Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* **38**:771–792.
11. **Biddle, J. F., S. Fitz-Gibbon, S. C. Schuster, J. E. Brechley, and C. H. House.** 2008. Metagenomic signatures of the Peru Margin seafloor biosphere show a genetically distinct environment. *Proc. Natl. Acad. Sci. U. S. A.* **105**:10583–10588.
12. **Biers, E. J., S. Sun, and E. C. Howard.** 2009. Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Appl. Environ. Microbiol.* **75**:2221–2229.
13. **Chen, I. C., V. Thiruvengadam, W. D. Lin, H. H. Chang, and W. H. Hsu.** 2010. Lysine racemase: a novel non-antibiotic selectable marker for plant transformation. *Plant Mol. Biol.* **72**:153–169.
14. **Chistoserdova, L.** 2010. Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol. Lett.* **32**:1351–1359.
15. **Cieśliński, H., et al.** 2009. Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library. *Pol. J. Microbiol.* **58**:199–204.
16. **Cole, J. R., et al.** 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res.* **31**:442–443.
17. **Craig, J. W., F.-Y. Chang, J. H. Kim, S. C. Obiajulu, and S. F. Brady.** 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse *Proteobacteria*. *Appl. Environ. Microbiol.* **76**:1633–1641.
18. **Daniel, R.** 2005. The metagenomics of soil. *Nat. Rev. Microbiol.* **3**:470–478.
19. **Delmotte, N., et al.** 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **106**:16428–16433.
20. **DeLong, E. F., et al.** 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**:496–503.
21. **Denef, V. J., et al.** 2009. Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environ. Microbiol.* **11**:313–325.
22. **DeSantis, T. Z., et al.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**:5069–5072.
23. **Diaz, N. N., L. Krause, A. Goemann, K. Niehaus, and T. W. Nattkemper.** 2009. TACO: taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *BMC Bioinformatics* **10**:56.
24. **Dinsdale, E. A., et al.** 2008. Functional metagenomic profiling of nine biomes. *Nature* **452**:629–632.
- 24a. **Donato, J. J., et al.** 2010. Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. *Appl. Environ. Microbiol.* **76**:4396–4401.

25. Duan, C. J., et al. 2009. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* **107**:245–256.
26. Elend, C., et al. 2006. Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl. Environ. Microbiol.* **72**:3637–3645.
27. Ferrer, M., A. Beloqui, K. N. Timmis, and P. N. Golyshin. 2009. Metagenomics for mining new genetic resources of microbial communities. *J. Mol. Microbiol. Biotechnol.* **16**:109–123.
28. Ferrer, M., et al. 2005. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem. Biol.* **12**:895–904.
29. Fierer, N., and R. B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* **103**:626–631.
30. Finn, R. D., et al. 2010. The Pfam protein families database. *Nucleic Acids Res.* **38**:D211–D222.
31. Frias-Lopez, J., et al. 2008. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. U. S. A.* **105**:3805–3810.
32. Gabor, E. M., W. B. Alkema, and D. B. Janssen. 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* **6**:879–886.
33. Galvão, T. C., W. W. Mohn, and V. de Lorenzo. 2005. Exploring the microbial biodegradation and biotransformation gene pool. *Trends Biotechnol.* **23**:497–506.
34. García Martín, H., et al. 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat. Biotechnol.* **24**:1263–1269.
35. Gardner, P. P., et al. 2009. Rfam: updates to the RNA families database. *Nucleic Acids Res.* **37**:D136–D140.
36. Gilbert, J. A., et al. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS One* **3**:e3042.
37. Gilbert, J. A., et al. 2009. Potential for phosphonoacetate utilization by marine bacteria in temperate coastal waters. *Environ. Microbiol.* **11**:111–125.
38. Gloux, K., et al. 2010. Microbes and Health Sackler Colloquium: a metagenomic  $\beta$ -glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1000066107.
39. Goll, J., et al. 2010. METAREP: JCVI metagenomics reports—an open source tool for high-performance comparative metagenomics. *Bioinformatics* **26**:2631–2632. doi:10.1093/bioinformatics/btq455.
40. Goltsman, D. S., et al. 2009. Community genomic and proteomic analyses of chemoautotrophic iron-oxidizing “*Leptospirillum rubrum*” (group II) and “*Leptospirillum ferrodiazotrophum*” (group III) bacteria in acid mine drainage biofilms. *Appl. Environ. Microbiol.* **75**:4599–4615.
41. Guan, C., et al. 2007. Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota. *Appl. Environ. Microbiol.* **73**:3669–3676.
42. Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**:669–685.
43. Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**:R245–R249.
44. Reference deleted.
45. Hårdeman, F., and S. Sjöling. 2007. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol. Ecol.* **59**:524–534.
46. He, S., et al. 2010. Metatranscriptomic array analysis of ‘*Candidatus* Accumulibacter phosphatis’-enriched enhanced biological phosphorus removal sludge. *Environ. Microbiol.* **12**:1205–1217.
47. Healy, F. G., et al. 1995. Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl. Microbiol. Biotechnol.* **43**:667–674.
48. Heath, C., X. P. Hu, S. C. Cary, and D. Cowan. 2009. Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from Antarctic desert soil. *Appl. Environ. Microbiol.* **75**:4657–4659.
49. Henne, A., R. A. Schmitz, M. Bömeke, G. Gottschalk, and R. Daniel. 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl. Environ. Microbiol.* **66**:3113–3116.
50. Hjort, K., et al. 2010. Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS Microbiol. Ecol.* **71**:197–207.
51. Huse, S. M., J. A. Huber, H. G. Morrison, M. L. Sogin, and D. M. Welch. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **8**:R143.
52. Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. MEGAN analysis of metagenomic data. *Genome Res.* **17**:377–386.
53. Huson, D. H., D. C. Richter, S. Mitra, A. F. Auch, and S. C. Schuster. 2009. Methods for comparative metagenomics. *BMC Bioinformatics* **10**(Suppl. 1):S12.
54. Iwai, S., et al. 2010. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *ISME J.* **4**:279–285.
55. Jeon, J. H., J. T. Kim, S. G. Kang, J. H. Lee, and S. J. Kim. 2009. Characterization and its potential application of two esterases derived from the arctic sediment metagenome. *Mar. Biotechnol. (NY)* **11**:307–316.
56. Kan, J., T. E. Hanson, J. M. Ginter, K. Wang, and F. Chen. 2005. Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline Syst.* **1**:7.
57. Kanehisa, M., et al. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**:D480–D484.
58. Karl, D. M., and R. Lukas. 1996. The Hawaii Ocean Time-series (HOT) program: background, rationale and field implementation. *Deep-Sea Res. II* **43**:129–156.
59. Karlin, S., J. Mrázek, and A. M. Campbell. 1997. Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.* **179**:3899–3913.
60. Klaassens, E. S., W. M. de Vos, and E. E. Vaughan. 2007. Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl. Environ. Microbiol.* **73**:1388–1392.
61. Knietzsch, A., S. Bowien, G. Whited, G. Gottschalk, and R. Daniel. 2003. Identification and characterization of coenzyme B<sub>12</sub>-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl. Environ. Microbiol.* **69**:3048–3060.
62. Krause, L., et al. 2008. Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res.* **36**:2230–2239.
63. Kröber, M., et al. 2009. Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. *J. Biotechnol.* **142**:38–49.
64. Kunin, V., A. Engelbrektsen, H. Ochman, and P. Hugenholtz. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12**:118–123.
65. Lacerda, C. M., L. H. Cho, and K. F. Reardon. 2007. Metaproteomic analysis of a bacterial community response to cadmium exposure. *J. Proteome Res.* **6**:1145–1152.
66. Leininger, S., et al. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**:806–809.
67. Li, M., Y. Hong, M. G. Klotz, and J. D. Gu. 2010. A comparison of primer sets for detecting 16S rRNA and hydrazine oxidoreductase genes of anaerobic ammonium-oxidizing bacteria in marine sediments. *Appl. Microbiol. Biotechnol.* **86**:781–790.
68. Liaw, R. B., M. P. Cheng, M. C. Wu, and C. Y. Lee. 2010. Use of metagenomic approaches to isolate lipolytic genes from activated sludge. *Bioreour. Technol.* **101**:8323–8329.
69. Lo, I., et al. 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446**:537–541.
70. Lorenz, P., and J. Eck. 2005. Metagenomics and industrial applications. *Nat. Rev. Microbiol.* **3**:510–516.
71. Majerník, A., G. Gottschalk, and R. Daniel. 2001. Screening of environmental DNA libraries for the presence of genes conferring Na<sup>+</sup>(Li<sup>+</sup>)/H<sup>+</sup> antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J. Bacteriol.* **183**:6645–6653.
72. Manichanh, C., et al. 2008. A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. *Nucleic Acids Res.* **36**:5180–5188.
73. Margulies, M., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
74. Markowitz, V. M., et al. 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36**:D534–D538.
75. Maróti, G., et al. 2009. Discovery of [NiFe] hydrogenase genes in metagenomic DNA: cloning and heterologous expression in *Thiocapsa roseopersicina*. *Appl. Environ. Microbiol.* **75**:5821–5830.
76. McHardy, A. C., H. G. Martín, A. Tsirigos, P. Hugenholtz, and I. Rigoutsos. 2007. Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Methods* **4**:63–72.
77. McHardy, A. C., and I. Rigoutsos. 2007. What’s in the mix: phylogenetic classification of metagenome sequence samples. *Curr. Opin. Microbiol.* **10**:499–503.
78. Meyer, F., et al. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**:386.
79. Monzoorul Haque, M., T. S. Ghosh, D. Komanduri, and S. S. Mande. 2009. SORT-ITEMS: sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics* **25**:1722–1730.
80. Morimoto, S., and T. Fujii. 2009. A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Appl. Microbiol. Biotechnol.* **83**:389–396.
81. Muyzer, G., E. C. de Waal, and A. G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59**:695–700.



82. Noonan, J. P., et al. 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**:1113–1118.
83. Overbeek, R., et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**:5691–5702.
84. Pace, N. R., D. J. Stahl, D. J. Lane, and G. J. Olsen. 1985. Analyzing natural microbial populations by rRNA sequences. *ASM News* **51**:4–12.
85. Park, C., and R. F. Helm. 2008. Application of metaproteomic analysis for studying extracellular polymeric substances (EPS) in activated sludge flocs and their fate in sludge digestion. *Water Sci. Technol.* **57**:2009–2015.
86. Parro, V., M. Moreno-Paz, and E. González-Toril. 2007. Analysis of environmental transcriptomes by DNA microarrays. *Environ. Microbiol.* **9**:453–464.
87. Pathak, G. P., A. Ehrenreich, A. Losi, W. R. Streit, and W. Gärtner. 2009. Novel blue light-sensitive proteins from a metagenomic approach. *Environ. Microbiol.* **11**:2388–2399.
88. Peterson, J., et al. 2009. The NIH Human Microbiome Project. *Genome Res.* **19**:2317–2323.
89. Poinar, H. N., et al. 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**:392–394.
90. Poretsky, R. S., et al. 2005. Analysis of microbial gene transcripts in environmental samples. *Appl. Environ. Microbiol.* **71**:4121–4126.
91. Poretsky, R. S., et al. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ. Microbiol.* **11**:1358–1375.
92. Pruesse, E., et al. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**:7188–7196.
93. Ram, R. J., et al. 2005. Community proteomics of a natural microbial biofilm. *Science* **308**:1915–1920.
94. Rhee, J. K., D. G. Ahn, Y. G. Kim, and J. W. Oh. 2005. New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. *Appl. Environ. Microbiol.* **71**:817–825.
95. Riesenfeld, C. S., R. M. Goodman, and J. Handelsman. 2004. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* **6**:981–989.
96. Riesenfeld, C. S., P. D. Schloss, and J. Handelsman. 2004. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**:525–552.
97. Rudney, J. D., H. Xie, N. L. Rhodus, F. G. Ondrey, and T. J. Griffin. 2010. A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. *Mol. Oral Microbiol.* **25**:38–49.
98. Rusch, D. B., et al. 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**:e77.
99. Saleh-Lakha, S., et al. 2005. Microbial gene expression in soil: methods, applications and challenges. *J. Microbiol. Methods* **63**:1–19.
100. Sandberg, R., et al. 2001. Capturing whole-genome characteristics in short sequences using a naïve Bayesian classifier. *Genome Res.* **11**:1404–1409.
101. Sayers, E. W., et al. 2009. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **37**:D5–D15.
102. Schmidt, O., H. L. Drake, and M. A. Horn. 2010. Hitherto unknown [Fe-Fe]-hydrogenase gene diversity in anaerobes and anoxic enrichments from a moderately acidic fen. *Appl. Environ. Microbiol.* **76**:2027–2031.
103. Schmidt, T. M., E. F. DeLong, and N. R. Pace. 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**:4371–4378.
104. Schneider, T., and K. Riedel. 2010. Environmental proteomics: analysis of structure and function of microbial communities. *Proteomics* **10**:785–798.
105. Schulze, W. X., et al. 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* **142**:335–343.
106. Shi, Y., G. W. Tyson, and E. F. DeLong. 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459**:266–269.
107. Simon, C., and R. Daniel. 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Appl. Microbiol. Biotechnol.* **85**:265–276.
108. Simon, C., and R. Daniel. 2010. Construction of small-insert and large-insert metagenomic libraries. *Methods Mol. Biol.* **668**:39–50.
109. Simon, C., J. Herath, S. Rockstroh, and R. Daniel. 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl. Environ. Microbiol.* **75**:2964–2968.
110. Simon, C., A. Wiesz, A. W. Strittmatter, and R. Daniel. 2009. Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Appl. Environ. Microbiol.* **75**:7519–7526.
111. Sjöling, S., and D. A. Cowan. 2008. Metagenomics: microbial community genomes revealed, p. 313–332. *In* R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday (ed.), *Psychrophiles: from biodiversity to biotechnology*. Springer-Verlag, Berlin, Germany.
112. Sleator, R. D., C. Shortall, and C. Hill. 2008. Metagenomics. *Lett. Appl. Microbiol.* **47**:361–366.
113. Sogin, M. L., et al. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proc. Natl. Acad. Sci. U. S. A.* **103**:12115–12120.
114. Sorek, R., and P. Cossart. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.* **11**:9–16.
115. Sowell, S. M., et al. 2009. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J.* **3**:93–105.
116. Steele, H. L., K. E. Jaeger, R. Daniel, and W. R. Streit. 2009. Advances in recovery of novel biocatalysts from metagenomes. *J. Mol. Microbiol. Biotechnol.* **16**:25–37.
117. Stein, J. L., T. L. Marsh, K. Y. Wu, H. Shizuya, and E. F. DeLong. 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**:591–599.
118. Sul, W. J., et al. 2009. DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Appl. Environ. Microbiol.* **75**:5501–5506.
119. Tartar, A., et al. 2009. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol. Biofuels* **2**:25.
120. Tatusov, R. L., et al. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**:22–28.
121. Teeling, H., A. Meyerdieck, M. Bauer, R. Amann, and F. O. Glöckner. 2004. Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ. Microbiol.* **6**:938–947.
122. Teeling, H., J. Waldmann, T. Lombardot, M. Bauer, and F. O. Glöckner. 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.
123. Tiedje, J. M., S. Asuming-Brempong, K. Nüsslein, T. L. Marsh, and S. J. Flynn. 1999. Opening the black box of soil microbial diversity. *Appl. Soil Ecol.* **13**:109–122.
124. Toyoda, A., W. Iio, M. Mitsumori, and H. Minato. 2009. Isolation and identification of cellulose-binding proteins from sheep rumen contents. *Appl. Environ. Microbiol.* **75**:1667–1673.
125. Tringe, S. G., et al. 2005. Comparative metagenomics of microbial communities. *Science* **308**:554–557.
126. Turnbaugh, P. J., and J. I. Gordon. 2008. An invitation to the marriage of metagenomics and metabolomics. *Cell* **134**:708–713.
127. Tyson, G. W., et al. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**:37–43.
128. Uchiyama, T., T. Abe, T. Ikemura, and K. Watanabe. 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat. Biotechnol.* **23**:88–93.
129. Uchiyama, T., and K. Miyazaki. 10 September 2010. Product-induced gene expression (PIGEX): a product-responsive reporter assay for enzyme screening of metagenomic libraries. *Appl. Environ. Microbiol.* doi:10.1128/AEM.00464-10.
130. Ulrich, T., et al. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3**:e2527.
131. Varaljay, V. A., E. C. Howard, S. Sun, and M. A. Moran. 2010. Deep sequencing of a dimethylsulfoniopropionate-degrading gene (*dmdA*) by using PCR primer pairs designed on the basis of marine metagenomic data. *Appl. Environ. Microbiol.* **76**:609–617.
132. Venter, J. C., et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**:66–74.
133. Verberkmoes, N. C., et al. 2009. Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* **3**:179–189.
134. Voget, S., H. L. Steele, and W. R. Streit. 2006. Characterization of a metagenome-derived halotolerant cellulase. *J. Biotechnol.* **126**:26–36.
135. von Mering, C., et al. 2007. Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**:1126–1130.
136. Wang, G.-Y.-S., et al. 2000. Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* **2**:2401–2404.
137. Wang, C., et al. 2006. Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl. Environ. Microbiol.* **72**:384–391.
138. Warnecke, F., et al. 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560–565.
139. Waschkowitz, T., S. Rockstroh, and R. Daniel. 2009. Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl. Environ. Microbiol.* **75**:2506–2516.
140. Weckx, S., et al. 2010. Community dynamics of sourdough fermentations as revealed by their metatranscriptome. *Appl. Environ. Microbiol.* **76**:5402–5408.
141. Williamson, L. L., et al. 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl. Environ. Microbiol.* **71**:6335–6344.
142. Wilmes, P., et al. 2008. Community proteogenomics highlights microbial

- strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J.* **2**:853–864.
143. **Wilmes, P., and P. L. Bond.** 2004. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ. Microbiol.* **6**:911–920.
144. **Wilmes, P., M. Wexler, and P. L. Bond.** 2008. Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS One* **3**:e1778.
145. **Wu, C., and B. Sun.** 2009. Identification of novel esterase from metagenomic library of Yangtze River. *J. Microbiol. Biotechnol.* **19**:187–193.
146. **Yang, B., et al.** 2010. Unsupervised binning of environmental genomic fragments based on an error robust selection of l-mers. *BMC Bioinformatics* **11**(Suppl. 2):S5.
147. **Yuhong, Z., et al.** 2009. Lipase diversity in glacier soil based on analysis of metagenomic DNA fragments and cell culture. *J. Microbiol. Biotechnol.* **19**:888–897.
148. **Zapras, A., Y. J. Liu, S. J. Liu, H. L. Drake, and M. A. Horn.** 2010. Abundance of novel and diverse *tfdA*-like genes, encoding putative phenoxyalkanoic acid herbicide-degrading dioxygenases, in soil. *Appl. Environ. Microbiol.* **76**:119–128.
149. **Zhou, J., and D. K. Thompson.** 2002. Challenges in applying microarrays to environmental studies. *Curr. Opin. Biotechnol.* **13**:204–2207.