Accepted Manuscript

Contemporary molecular tools in microbial ecology and their application to advancing biotechnology

Mamoon Rashid, Ulrich Stingl

PII: S0734-9750(15)30038-0
DOI: doi: [10.1016/j.biotechadv.2015.09.005](https://doi.org/10.1016/j.biotechadv.2015.09.005)
Reference: JBA 6973

To appear in: *Biotechnology Advances*

Received date: 30 May 2015
Revised date: 19 September 2015
Accepted date: 20 September 2015

Please cite this article as: Rashid Mamoon, Stingl Ulrich, Contemporary molecular tools in microbial ecology and their application to advancing biotechnology, *Biotechnology Advances* (2015), doi: [10.1016/j.biotechadv.2015.09.005](https://doi.org/10.1016/j.biotechadv.2015.09.005)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**THIS VERSION WAS CAREFULLY EDITED FROM THE
PREVIOUS REVISED MANUSCRIPT**

**Contemporary molecular tools in microbial ecology and
their application to advancing biotechnology**

Mamoon Rashid and Ulrich Stingl

Red Sea Research Centre, King Abdullah University of Science and Technology, Thuwal
23955-6900, Saudi Arabia

Correspondence to: Uli.Stingl@kaust.edu.sa or Mamoon.Rashid@kaust.edu.sa

Table of Contents

ABSTRACT	4
1 INTRODUCTION	5
1.1 MICROBIAL ECOLOGY: HISTORICAL PERSPECTIVE	5
1.2 MOLECULAR METHODS AND TOOLS IN MICROBIAL ECOLOGY	5
1.3 APPLICATION OF MICROBIAL ECOLOGICAL TOOLS TO ADVANCING BIOTECHNOLOGY	7
1.4 WHY IS THIS REVIEW TIMELY?	8
2 MOLECULAR REVOLUTION IN MICROBIAL ECOLOGY.....	9
3 METAGENOMICS: TAPPING INTO THE UNKNOWNNS FOR BIOTECHNOLOGY	13
3.1 SEQUENCE-BASED SCREENING OF METAGENOMES FOR NOVEL ENZYMES	14
3.1.1 <i>Methods of sequence-based screening.....</i>	14
3.1.2 <i>Examples of novel enzymes identified by sequence-based metagenomics.....</i>	16
3.2 FUNCTION-BASED SCREENING FOR DETECTION OF NOVEL ENZYMES.....	16
3.2.1 <i>Functional complementation</i>	17
3.2.2 <i>High-Performance Thin Layer Chromatography (HPTLC)</i>	17
3.2.3 <i>Phenotypic MicroArray (PM).....</i>	17
3.2.4 <i>Community Isotope Array (CIArray).....</i>	17
3.2.5 <i>In vitro Compartmentalization-Fluorescence Activated Cell Sorting (IVC-FACS).....</i>	18
3.2.6 <i>Substrate-Induced Gene Expression (SIGEX)</i>	18
3.2.7 <i>Product-Induced Gene Expression (PIGEX).....</i>	19
3.2.8 <i>Examples of function-based screening of metagenomes for novel enzymes.....</i>	19
3.3 COMBINED SEQUENCE- AND FUNCTION-BASED SCREENINGS OF METAGENOMES FOR THE DETECTION OF NOVEL ENZYMES	19
3.4 NATURAL PRODUCTS THROUGH METAGENOMICS.....	21
3.4.1 <i>Natural products from free-living bacteria</i>	22
3.4.2 <i>Natural products from symbiotic bacteria</i>	22
3.4.3 <i>Examples of natural product discoveries through metagenomics.....</i>	23
3.4.4 <i>Drug discovery through metagenomics.....</i>	24
3.5 METAGENOMIC ENRICHMENT TECHNOLOGIES.....	25
3.6 BOTTLENECKS IN THE METAGENOMICS DISCOVERY PIPELINE	26

3.6.1 Sampling metagenomes: random or intuitive.....	26
3.6.2 DNA extraction.....	27
3.6.3 Choice of vectors.....	28
3.6.4 Choice of hosts	29
4 SINGLE-CELL GENOMICS: TAPPING THE UNTAPPED FOR APPLICATION IN BIOTECHNOLOGY.....	31
4.1 ANALYSES OF SINGLE-CELL GENOMES FOR BIOTECHNOLOGY.....	32
5 PROPOSITION: FUNCTIONAL SINGLE-CELL GENOMICS	33
6 CONCLUSION	34
FIGURE LEGENDS	59
TABLES	62

Abstract

Novel methods in microbial ecology are revolutionizing our understanding of the structure and function of microbes in the environment, but concomitant advances in applications of these tools to biotechnology are mostly lagging behind. After more than a century of efforts to improve microbial culturing techniques, about 70-80% of microbial diversity – recently called the “microbial dark matter” – remains uncultured. In early attempts to identify and sample these so far uncultured taxonomic lineages, methods that amplify and sequence ribosomal RNA genes were extensively used. Recent developments in cell separation techniques, DNA amplification, and high-throughput DNA sequencing platforms have now made the discovery of genes/genomes of uncultured microorganisms from different environments possible through the use of metagenomic techniques and single-cell genomics. When used synergistically, these metagenomic and single-cell techniques create a powerful tool to study microbial diversity. These genomics techniques have already been successfully exploited to identify sources for i) novel enzymes or natural products for biotechnology applications, ii) novel genes from extremophiles, and iii) whole genomes or operons from uncultured microbes. More can be done to utilize these tools more efficiently in biotechnology.

Keywords: Ecology, Biotechnology, Metagenomics, Single-cell genomics, Natural products, Enzymes

1 Introduction

1.1 Microbial ecology: historical perspective

The word *ecology* was coined in the late nineteenth century from the two Greek words *oikos* (household) and *logos* (law) to suggest that the study of the relations between organisms comes from “law of the household”. Ernst Haeckel (Haeckel 1866), a German biologist, identified ecology as a separate branch of science that deals with the interactions of living organisms with their biotic (living) and abiotic (non-living) environments. Microbial ecology is thus the study of microbes and their relationships with the (living and non-living) environment. Microbes shape the earth’s climate, directly and indirectly, because the evolution of the Earth’s atmosphere is tightly linked with the evolution of its biota (Kasting and Siefert 2002). Ecological research on microorganisms began in the late 1800s when microbiologists Sergei Winogradsky (Winogradsky 1887) and Martinus Beijerinck (Beijerinck 1888) demonstrated the roles played by microorganisms in the natural environment in sulphur oxidation, iron oxidation, and nitrogen fixation. However, the term “microbial ecology” was not used widely until the 1960s when public interest in environmental issues and in the essential roles of microorganisms in the Earth’s biosphere led to the emergence of the new field of “microbial ecology” (Xu 2006).

1.2 Molecular methods and tools in microbial ecology

Rapid progress and development in the field of microbial ecology owes much to the application of genomics and metagenomics methods. The basic genomic tools that have been applied to examine the natural microbial population and communities include DNA sequencing, polymerase chain reaction (PCR), DNA cloning systems, hybridization techniques like fluorescent in situ hybridization (FISH), and bioinformatics. The important technological advances in the field of molecular microbial ecology are shown in Figure 1, and the most important ones are described in subsequent sections of this manuscript.

Large-scale genome sequencing projects are now providing significant insights into the sizes and metabolic functions of the genomes of a wide variety of pure cultures of microorganisms (Abdallah, Rashid et al. 2012, Abdallah, Rashid et al. 2012, Jimenez-Infante, Ngugi et al. 2014, Reddy, Thomas et al. 2015, Salipante, Roach et al. 2015). The outcome of size analysis of sequenced prokaryotic genomes revealed that there is high variability in genome size between and within species. The genome sizes of sequenced prokaryotic genomes have been found to vary widely from the smallest known genome of the obligate archaeon parasite *Nanoarchaeum equitans* (490 kb) (Waters, Hohn et al. 2003, Das, Paul et al. 2006) to the largest known genome of the soil bacterium *Ktedonobacter racemifer* (13.6 Mb) (Chang, Land et al. 2011). Interestingly, 20-40% of the predicted open reading frames in sequenced microbial genomes do not have a known function (Xu 2006).

Novel methods, including rRNA-based techniques, metagenomic shotgun sequencing, metatranscriptomics, as well as phylogenetic and functional microarrays, have recently been used to unravel the genetic and functional diversity of environmental samples with complex microbial communities that can contain thousands of microbial species. The term metagenomics broadly refers to two different approaches. The first one relies on sequencing and analyzing the genetic material from microbial communities and is referred to as “sequence-based metagenomics”. The second approach relies on the heterologous gene expression of the cloned metagenomic fragments and is thus termed as “functional metagenomics”. These methods have been used successfully to map the genetic diversity and complexity present in natural microbial assemblages and also to identify many novel lineages of bacteria, viruses, archaea and microbial eukaryotes, as well as to predict the functional or metabolic diversity in these natural microbial populations without the need for culturing.

Recently, sequencing of amplified genomes of physically separated single cells has overcome some of the issues of metagenomics to reconstruct large non-chimeric gene fragments in order to vastly improve access to nearly complete genomes of uncultured microbial lineages.

1.3 Application of microbial ecological tools to advancing biotechnology

Over 3.8 billion years of evolution, prokaryotes have accumulated vast and diverse genetic and physiological characteristics that are essential for survival in their natural habitats. Some natural habitats have extreme physicochemical conditions similar to those used in industrial processes (Antunes, Ngugi et al. 2011). It has been determined that there are 10^6 - 10^8 different genospecies of prokaryotes (Simon and Daniel 2011), surpassing the number of isolated microbial species by a factor of about 10^4 . Almost three decades ago, the well-known phrase in microbial ecology of the “great plate count anomaly” was coined to describe this mismatch in numbers (Staley and Konopka 1985). In a variety of natural environments (aquatic, soil, and sediments), direct microscopic counts exceed viable-cell counts by several orders of magnitude. For example, the quantification of culturability (percentage of culturable bacteria in comparison to total cell counts) in samples from different ecological niches turned out to vary widely, but in general accounted for less than 1% (Amann, Ludwig et al. 1995), with marine surface waters being the exception, where recent advances in cultivation techniques have increased the culturability to up to 14% (Connon and Giovannoni 2002, Giovannoni and Stingl 2005, Giovannoni and Stingl 2007). In a recent effort to culture these previously uncultured microbes, the isolation chip (Ichip) has been developed (Nichols, Cahoon et al. 2010). Ichip is composed of several hundreds of miniature chambers containing approximately one environmental cell in each chamber. This chip is incubated in the corresponding natural environments and growth factors can diffuse through the semipermeable membranes covering the chambers enabling growth of uncultured microbes. The discovery of the new antibiotic teixobactin from previously uncultured bacteria proved the usefulness of the Ichip technology (Ling, Schneider et al. 2015). To understand the potential of the cultured minority, it is important to realize that 80% of the antibiotics used today are sourced from a single genus, *Streptomyces* (Procopio, Silva et al. 2012), and that the number of described and deposited strains is relatively low (Kyrpides, Hugenholtz et al. 2014). This suggests an enormous potential for bioprospecting among the *uncultured microbial majority*, referred to as “microbial dark matter” (Schmidt and Relman 1994). To harvest these untapped resources of novel

biocatalysts and secondary metabolites, culture-independent techniques have been developed and applied.

However, the concomitant application of these techniques to biotechnology lags behind. For example, the isolation of bacterial DNA from environmental samples was reported in 1978 (Torsvik and Goksoyr 1978). About a decade later, this DNA was used to amplify and sequence ribosomal RNA genes to study the evolution and phylogeny of natural microbial populations (Olsen, Lane et al. 1986, Pace, Stahl et al. 1986, Pace, Olsen et al. 1986). About another decade later, the same DNA was used to identify/isolate genes encoding various enzymes like cellulases, amylases, lipases, etc. (Healy, Ray et al. 1995, Handelsman, Rondon et al. 1998, Rondon, August et al. 2000).

The methods used to isolate, clone, sequence, and functionally screen DNA from environmental samples and then to heterologously express selected clones have collectively been called “metagenomics” (Handelsman, Rondon et al. 1998). The potential of metagenomics to discover new molecules with diverse and desired functions was realized by the biotech industry (Lorenz and Eck 2005) and the development of detergent proteases subtilisins (Maurer 2004), esterases/lipases, polyketide synthases, antibiotics, chitinases and xylanases are clear examples (Lorenz, Liebeton et al. 2002, Schloss and Handelsman 2003). The application of single-cell genome sequencing has led to the discovery and first analyses of nearly complete genomic data of novel phylogenetic branches, the so-called “microbial dark matter” (Rinke, Schwientek et al. 2013), and concomitant advances in using single-cell genomics for advancing biotechnology should follow soon (Huang, Song et al. 2015).

1.4 Why is this review timely?

There are numerous review articles describing the impact of genomics and metagenomics methods on microbial ecology (Xu 2006, Binga, Lasken et al. 2008, Aziz, Breitbart et al. 2010, Gilbert and Dupont 2011, Rittmann and Herwig 2012, Su, Lei et al. 2012, Thompson, Field et al. 2013, Reid 2014, Tseng and Tang 2014). There is also a plethora of papers that discuss the application of metagenomics in the field of biotechnology to discover novel enzymes, drug molecules, processes and applications (Lorenz, Liebeton et al. 2002, Schloss and Handelsman 2003, Cowan, Meyer et al. 2005, Lorenz and Eck

2005, Kennedy, Marchesi et al. 2007, Brady, Simmons et al. 2009, Steele, Jaeger et al. 2009, Piel 2011, Iqbal, Feng et al. 2012, de Castro, Sartori da Silva et al. 2013, Hicks and Prather 2014). Literature that reviews the development of genomic or metagenomic methods in microbial ecology and the application of these methods to biotechnology is, on the contrary, scarce (Handelsman 2004, Steele and Streit 2005, Hentschel, Piel et al. 2012). This review adds to and updates this scarce literature.

2 Molecular revolution in microbial ecology

In their classical review in 1965, Zuckerkandl and Pauling discussed the fundamental properties of biological macromolecules that could potentially form the basis of molecular phylogeny (Zuckerkandl and Pauling 1965). The molecule of choice should be “*universal*” in physical space and “*conserved*” in functional and sequence spaces across all living beings, with a sequence “*less prone to mutation*”. RNA-RNA and RNA-DNA hybridization competition experiments and the thermal stability of these hybrids (Moore and McCarthy 1967, Bendich and McCarthy 1970, Pace and Campbell 1971) showed that the 16S rRNA molecule was more stable and more conserved than mRNA and less prone to mutational changes. Analyses of the respective gene could thus be used as a tool to calculate homologies among distant organisms. Oligonucleotide cataloguing of products of T1 ribonuclease digests of 5S rRNA (Sogin, Sogin et al. 1972) and 16S rRNA (Woese, Fox et al. 1975, Fox, Pechman et al. 1977) molecules further proved the conservation (mostly at the 3' terminal) of the primary structure of ribosomal RNA, thus facilitating its use as a phylogenetic marker. Comparative analysis of these oligonucleotide catalogues of 16S rRNA molecules further resulted in the first phylogenetic studies of prokaryotes (Woese and Fox 1977, Fox, Stackebrandt et al. 1980), establishing 16S rRNA as the molecule of choice for evolutionary studies. With the realization of the enormous potential in 16S rRNA studies, together with the advent of DNA sequencing methods (Sanger, Nicklen et al. 1977, Smith, Sanders et al. 1986), much of the focus was on determining their complete nucleotide sequences (Brosius, Palmer et al. 1978, Carbon, Ebel et al. 1981) because sequences would allow quantitative inferences of phylogenetic relationships (Zuckerkandl and Pauling 1965, Fitch and Margoliash 1967). By that time, the principle of using 16S rRNA genes for characterization of microorganisms had gained

wide acceptance and much simpler sequencing methods were needed. A major breakthrough in this context was the development of a method to rapidly identify partial 16S rRNA sequences from a cellular RNA preparation without the need for isolation of the molecule and cloning of its gene (Lane, Pace et al. 1985).

These historical developments in microbial ecology and evolution studies revolved around 16S rRNA molecules and genes from cultured microorganisms. **Figure 1** shows a timeline plot of the important events in the development of 16S rRNA-based methods, metagenomics, and single-cell techniques. One of the first revolutions in microbial ecology happened when 16S rRNA genes were sequenced from natural microbial communities (which were neither cultivated nor characterized previously in the laboratory) for phylogenetic and diversity analyses (Olsen, Lane et al. 1986, Pace, Stahl et al. 1986). The next step was then the separation of single 16S rRNA genes from the environment and thus the development of 16S rRNA gene clone libraries from natural microbial populations (Weller and Ward 1989) using polymerase chain reaction (PCR) (Mullis and Faloona 1987, Saiki, Gelfand et al. 1988). The advent of PCR was a remarkable breakthrough in molecular biology. Together, these improvements enabled some landmark studies on phylogenetic diversity of bacterioplankton from the Sargasso Sea (Giovannoni, Britschgi et al. 1990) and of a cyanobacterial mat from Octopus Spring in Yellowstone National Park (Ward, Weller et al. 1990), which were not possible previously with traditional culture techniques. Novel microbial lineages were discovered, and a surprising diversity of uncultured microbes was revealed from these environments. One of the daunting tasks facing microbial ecologists then (and still today) was to link this diversity to ecosystem functions and biogeochemical processes. Although most key findings were centered on 16S rRNA (Pace 1997, Rappe and Giovannoni 2003), this molecule alone does not provide biochemical and metabolic information on uncultured microbes as, in most cases, phylogeny and microbial physiology are not very closely linked.

In addition to ecological and evolutionary interest in this uncultured microbial majority, interest in harvesting these microbes for novel enzymes or natural products of biotechnological importance grew. One of the first examples of functional screening of gene clones used DNA isolated from a laboratory enrichment culture to help in the direct

isolation of cellulases (Healy, Ray et al. 1995). With the advent of large-insert cloning vectors like Bacterial Artificial Chromosome (BAC) (Shizuya, Birren et al. 1992), the concept of cloning genomic DNA of mixed microbial populations was put forth and the term “metagenome” was coined (Handelsman, Rondon et al. 1998). This was indeed the beginning of new era in microbial ecology that promised the potential of the discovery of natural products/enzymes from uncultured microbes in the soil (Handelsman, Rondon et al. 1998). Subsequently, metagenomic clone libraries were prepared and the genetic and functional diversity of soil microbiota was studied (Rondon, August et al. 2000) with the identification of lipase, amylase, and nuclease. Similar reports also came from marine microbial ecologists (Beja, Suzuki et al. 2000). The power of metagenomics became clear when genomic analysis of naturally occurring marine bacterioplankton revealed the presence of rhodopsin genes (Beja, Aravind et al. 2000, Beja, Spudich et al. 2001) in the γ -proteobacterial clade SAR86 [reviewed in: (Fuhrman, Schwalbach et al. 2008, Miyake and Stingl 2011)]. With this discovery of bacterial rhodopsin, a new type of phototrophy was described in oceanic surface waters, where the rhodopsin protein presumably acts as a light-driven proton pump. Before this seminal study, light energy was thought to enter marine ecosystem only through chlorophyll molecules. Marine viral metagenomes were also studied and researchers inferred that much (over 65 %) of the viral diversity was uncharacterized (Breitbart, Salamon et al. 2002). One issue with metagenomics that became apparent after the initial hype was that most assemblies of the rather short sequences failed to reproduce complete or nearly complete microbial genomes (Steward and Rappé 2007). Among the main reasons for this is the enormous and unexpected micro-heterogeneity, especially in marine bacterioplankton. One of the first attempts to reconstruct nearly complete microbial genomes from an environmental sample with low complexity was achieved in 2004, leading to the identification of carbon and nitrogen fixation pathways (Tyson, Chapman et al. 2004). In this study, the metabolic potential of a reconstructed genome of *Leptospirillum* group II was analysed. *Leptospirillum* has the genes needed to fix carbon through the Calvin-Benson-Bassham cycle using a type II ribulose 1,5-bisphosphate carboxylase-oxygenase and to fix nitrogen through nitrogenase. One of the most comprehensive marine metagenomic studies so far was initiated through the Global Ocean Sampling (GOS) expedition, with sampling sites from the North

Atlantic to the South Pacific oceans along a transect of several thousand kilometers. The environmental genomic content from the Sargasso Sea was published as a pilot project under the GOS survey and included more than a million novel genes comprising ~150 novel bacterial phylotypes (Venter, Remington et al. 2004). The final results of the GOS expedition revealed the distribution and relative abundance of hundreds of bacterial genomes present in ocean surface waters across the world (Rusch, Halpern et al. 2007, Yooseph, Sutton et al. 2007, Yooseph, Neelson et al. 2010). The GOS study used traditional Sanger sequencing of metagenomic and genomic clone libraries using a whole-genome shotgun strategy. These large-scale genome-sequencing projects prompted the development of more sophisticated, higher-throughput, cheaper and faster next-generation sequencing technologies (NGS) (Margulies, Egholm et al. 2005, Shendure, Porreca et al. 2005). In comparison with Sanger sequencing, NGS technologies offered very deep sequencing coverage that was able to capture even rare or low abundance (<1% relative abundance) organisms from metagenomic samples (Albertsen, Hugenholtz et al. 2013, Sims, Sudbery et al. 2014). As anticipated, the techniques were soon applied in microbial ecology, e.g. to study deep-mine microbial communities (Edwards, Rodriguez-Brito et al. 2006).

Although metagenomics was very useful, it remained inefficient in handling natural genomic heterogeneity and cross-strain assemblies could not be avoided (Rinke, Lee et al. 2014). Also, taxonomic assignments of metagenomic reads/contigs remain challenging. The recent breakthrough of single-cell genome sequencing (SCGC), in which sequencing the genomic material of a single physically separated uncultured microbial cell is possible, alleviated many bottlenecks in metagenomics (Kvist, Ahring et al. 2007, Marcy, Ouverney et al. 2007). In this technology, DNA from a single cell is isolated and amplified through multiple displacement amplification (MDA) in a sufficient quantity to be sequenced by the current day genome sequencers. But, challenges of uneven amplification of genomic regions and chimera formation in the obtained single cell genomes need to be overcome. Like metagenomics, also SCGC exploits the genetic and metabolic diversity of uncultured environmental microorganisms, but at a single-cell level. A landmark paper in single-cell genomics recently reported the genomes of 201

archaeal and bacterial cells belonging to 29 major uncultivated branches of the tree of life (“microbial dark matter”) (Rinke, Schwientek et al. 2013).

3 Metagenomics: Tapping into the unknowns for biotechnology

The application of genomics to the mixed genetic material from environmental microbial assemblages is called “*metagenomics*”. Some of the earliest examples of these practical approaches were published even before the term “metagenome” was coined. In a study on the phylogenetic diversity of marine picoplankton, environmental DNA was fragmented and cloned into bacteriophage lambda before being screened for 16S rRNA gene sequences (Schmidt, DeLong et al. 1991). This was the first example of the use of environmental DNA for phylogenetic purposes. Technologies for environmental sampling, DNA isolation, clone library preparation and screening/sequencing of the clone libraries have matured, and by 2015 the metagenomics approach has been applied to virtually every type of microbial niches (e.g., soil, seawater, freshwater, penguin feces, yak rumen, pitcher plants), diverse host-microbe symbiotic associations (lichens, invertebrate-microbes, vertebrate-microbes), extreme environments (brine pools, acid mines, hot springs, etc.) for a better understanding of the *structure* and *function* of the underlying uncultured microbial majority. After the establishment of next-generation sequencing (NGS) technologies, the pile of metagenomic data started growing enormously due to the speed, lower cost and higher throughput of metagenomics methods. Scientists observed this enormous increase in metagenomic datasets and realized the need for better and faster analysis methods. As noted in an editorial in *Nature*: “*Metagenomics sprang from advances in sequencing technology, and continued improvements are providing data in quantities unimaginable a few years ago. But without concerted efforts, the amount of data will quickly outpace the ability of scientists to analyze it*” (Editorial 2009). A general schematic presentation of the steps involved in a typical metagenomic discovery pipeline is presented in Figure 2 (along with the single-cell amplified genome (SAG) screening method that is described below).

Apart from the applications in microbial ecology, metagenomic methods have been used for biotechnological applications, but there are challenges and obstacles associated with it (Council 2007, Chistoserdova 2010, Grotzinger, Alam et al. 2014).

Sequence-based metagenomics suffers from the fact that the amount of sequence data accumulating in the databases is just outpacing their timely analysis. This results in the need of continuous refinements of bioinformatics methods to search for biocatalytic genes in the heap of data, and the need of a robust data management and storage system. Similarly, functional metagenomics also has challenges of developing better and more sensitive assay methods for selection of positive clones for a wide variety of biocatalysts and natural products, improved heterologous gene expression of these positive clones, and finally optimizing lead products at the production stage to be used for biotechnological purposes. The advantages and disadvantages of sequence- and function-based metagenomics are listed in Table 1.

3.1 Sequence-based screening of metagenomes for novel enzymes

In most early metagenomic studies, the assignment of taxonomic information to the retrieved sequences was difficult. *DeLong and co-workers* (Stein, Marsh et al. 1996) presented one of the early examples of a sequence-based approach where a 40-kilobase-pair fragment was cloned from a marine planktonic community and completely sequenced. A 16S rRNA gene belonging to Crenarchaeota as well as other functional enzymes such as RNA helicase and glutamate semialdehyde aminotransferase were identified. The same group later used a similar approach to identify a bacterial rhodopsin in which a 130 kb fragment containing a rRNA operon from an uncultivated marine γ -proteobacterium (SAR86 group) was revealed to encode a putative rhodopsin (proteorhodopsin) gene (Beja, Aravind et al. 2000). A new type of photo(hetero)trophy was described using proteorhodopsin gene variants in the ocean (Beja, Spudich et al. 2001). These examples indicated the potential of the sequence-based metagenomic approach that links the phylogeny of the genomic fragments of uncultivated microorganisms to functional aspects, because both the rRNA gene and functional genes were on the same fragment (Handelsman 2004).

3.1.1 Methods of sequence-based screening

PCR and probe-based methods

The most common and conventional nucleotide sequence-based approaches are PCR amplification of target gene(s) or oligonucleotide probe-based hybridization experiments (Knietzsch, Bowien et al. 2003, Daniel 2005, Lawley and Tannock 2012). PCR has been used extensively to retrieve partial sequences of target genes for both taxonomic diversity as well as functional studies (Lorenz, Liebeton et al. 2002, Cowan, Meyer et al. 2005). These methods rely on prior knowledge of “*known*” primer or probe sequences used to detect the “*unknown*” diversity of genes present in environmental samples, thus missing potentially new enzymes due to a possible sequence divergence in the target region.

Metagenomic Microarray (MGA)

Since the large number of clones in clone libraries demands rapid methods of screening, the use of microarrays in this context introduced a promising high-throughput screening (HTS) method for sequenced-based metagenomics (Sebat, Colwell et al. 2003, Park, Kang et al. 2008). There is plethora of examples of metagenome microarray (MGA) studies; a few recent ones are (Park, Kim et al. 2012, Guo, Yin et al. 2013, Jacquiod, Demaneche et al. 2014). Briefly, clones from metagenomic libraries are arrayed on a glass slide as probes and the labeled target DNA (ribosomal genes, functional genes, community DNA) is hybridized to detect signals. The quantitative nature of microarrays using MGA provides an advantage over PCR-based screening. In comparison to next generation sequencing used for microbial diagnosis in metagenomic samples, the MGA has significant advantages in cost and processing time. Still, there are challenges associated with this technology concerning sensitivity, specificity and quantification of the assay. Moreover, the identification of rare or novel uncultivated species in metagenomic samples remains difficult.

Shotgun metagenomics

With the advent of the faster, higher-throughput, and cheaper NGS technologies, metagenomic shotgun sequencing has become very handy and routinely produces large metagenomic datasets (Simon and Daniel 2009, Hentschel, Piel et al. 2012, Di Bella, Bao et al. 2013, Ferreira, Siam et al. 2014, Sharpton 2014). In the shotgun sequencing approach, already applied earlier to cultured microbes and human genomes, metagenomic DNA is randomly sheared and sequenced in short reads. The reads are then analyzed or assembled to produce longer genomic sequences. Shotgun metagenomic sequence data

provides an opportunity to simultaneously investigate two important aspects of the microbial community, *taxonomic diversity* and *biological function*.

Despite the advantages and usefulness of metagenomic sequencing, it also suffers from caveats and challenges (Council 2007). Firstly, the size and complexity of the sequence data pose analytic and informatics challenges. Additionally, it is difficult to trace the taxonomic source of short reads from a complex community. The presence of unwanted host DNA or environmental contamination in metagenomic sequencing data is another important issue (Schmieder and Edwards 2011, Degnan and Ochman 2012). Lastly, it is relatively costly to generate metagenomes compared to amplicon sequences in case of complex communities or where the contamination is a prominent issue.

3.1.2 Examples of novel enzymes identified by sequence-based metagenomics

Next generation sequencing technologies have had a big impact on the identification of functional genes from every kind of environmental microbial communities. Enzymes identified by sequence-based metagenomics approaches are compiled from the existing literature and presented in Table 2.

3.2 Function-based screening for detection of novel enzymes

Functional metagenomics relies on the expression of the gene(s) present in clone libraries in heterologous hosts to produce certain phenotypes, such as color, inhibition zones, or fluorescence. Generally, an assay is designed to detect the enzymatic activity either in agar colonies or crude cell lysates by alteration of chromophores or fluorophores (Bornscheuer 2002, Goddard and Reymond 2004, Andexer, Guterl et al. 2006). Most agar-plate-based methods suffer from low sensitivity due to the diffusion of the soluble products. Crude-cell-lysate-based methods usually have low throughput and are less reproducible (Felczykowska, Dydecka et al. 2014). Lipases/esterases constitute a major fraction of enzymes derived from metagenomic screening simply due to the availability of easy high-throughput screening methods (Taupp, Mewis et al. 2011, Reyes-Duarte, Ferrer et al. 2012). Apart from these direct screening methods, more sophisticated high-throughput screening technologies have recently been developed.

3.2.1 Functional complementation

In a functional complementation assay, a bacterial strain missing the activity of a specific enzyme (or unable to utilize a certain substrate due to non-functioning of the respective enzyme) is screened for restoration of the function in the presence of a metagenomic clone library. The metagenomic clone that complements the desired function in the mutant strain is isolated and sequenced for the identification of coding gene(s) (Wang, Meek et al. 2006). In the functional complementation of a constructed dehydratase-negative *Escherichia coli* strain, two positives were obtained out of 560,000 tested clones (Knietsch, Bowien et al. 2003).

3.2.2 High-Performance Thin Layer Chromatography (HPTLC)

Metagenome extract thin-layer chromatography analysis (META) was developed as a high-performance method to detect glycosyltransferase (GT) and other flavonoid-modifying activities (Rabausch, Juergensen et al. 2013). This method is sensitive enough to detect even 4 ng of modified flavonoid molecules. The technology was validated with a control library of 1920 fosmid clones from a single *Bacillus cereus* isolate and then applied on 38,000 metagenomic clones identifying two novel uridine diphosphate (UDP) glycosyltransferase (UGT) genes.

3.2.3 Phenotypic MicroArray (PM)

PM employs a microtitre-plate-based assay in which each well of the microtitre plate contains a different set of conditions and tests for a specific phenotype. PM has been used to identify gene functions, pathogenicity, metabolic capacity and drug targets. This method led to the identification of a novel gene involved in salt-tolerance *sdtR* (σ^{54} -dependent transcriptional regulator) from a human gut microbiome (Culligan, Marchesi et al. 2014).

3.2.4 Community Isotope Array (CIArray)

In a CIArray, the assimilation of radiolabeled isotopes by a specific microorganism in a community is detected and identified. This method has been used to identify the microorganisms that degraded phenol under anoxic conditions from an activated sludge (Tourlousse, Kurisu et al. 2013). The fosmid clones were spotted on an electropositive nylon membrane to constitute the CIArray. The DNA isolated from the ^{14}C -phenol-amended sample was used as hybridization probes on this CIArray to detect clones that

facilitated isotope assimilation. A positive clone was identified as belonging to a marine γ - proteobacterium only distantly related to cultured organisms.

3.2.5 *In vitro* Compartmentalization-Fluorescence Activated Cell Sorting (IVC-FACS)

Cloning and expression-based screening methods have some disadvantages. Sometimes the target genes encoding novel enzymes constitute a very small proportion (less than 0.01%) of the total DNA extracted from the environmental sample and may thus be overlooked during the functional screening procedure. Another possible problem with functional metagenomic screening is the heterologous gene expression in surrogate hosts (Council 2007, Gabor, Liebeton et al. 2007). *In vitro* compartmentalization works with water-in-oil emulsion, where a water phase is dispersed in oil to form aqueous microdroplets. Each of these microdroplets (with ~5 femtoliter volume) generally enables a single gene to transcribe and translate in a cell-free manner (Griffiths and Tawfik 2006) and also detects single enzyme molecules (Griffiths and Tawfik 2003). The ease of determination and regulation of droplet content and the large number of droplets ($>10^{10}$ per milliliter emulsion) makes IVC a suitable ultra-high-throughput screening technology. In the first application of this technology, genes encoding methyltransferases were selected from a 10^7 -fold excess of genes encoding another enzymes (Tawfik and Griffiths 1998). Application of IVC together with FACS in metagenomics, as an advantage over traditional functional screening, has been reviewed in detail by (Ferrer, Belouqui et al. 2009).

3.2.6 Substrate-Induced Gene Expression (SIGEX)

This method is based on the knowledge that catabolic gene expression is induced in the presence of a specific substrate and controlled by regulatory elements located in close proximity to catabolic genes. This knowledge was exploited and an operon-trap green fluorescent protein (GFP)-expression vector was constructed for shotgun cloning of metagenomic DNA. In the presence of the substrate, the cloned catabolic gene is expressed resulting in GFP expression, thus facilitating the high-throughput selection of positive clones in liquid cultures by fluorescence-activated cell sorting (Uchiyama, Abe et al. 2005). The SIGEX method was developed and tested first time to clone the phenol-degradation operon (pox operon) from *Ralstonia eutropha*. The method was also used to

clone aromatic hydrocarbon-induced genes from a groundwater metagenome library. Catabolic genes that are distant from the transcriptional regulator cannot be cloned using SIGEX. Also, fragments cloned using this method may contain partial genes (Uchiyama and Watanabe 2008).

3.2.7 Product-Induced Gene Expression (PIGEX)

Products of an enzymatic reaction carried out in a heterologous host can induce gene expression in another sensor cell; the corresponding assay is called product-induced gene expression (PIGEX). This method was applied to identify amidases from metagenomic clone libraries (Uchiyama and Miyazaki 2010). A specialized sensor for an *Escherichia coli* strain was constructed with the benzoate-responsive transcriptional activator (benR), upstream of the gene for GFP. The presence of benzoate, but not benzamide, activates the benR-GFP gene cassette and the sensor cells fluoresce. First, 96,000 metagenomic clones were grown in a 96-well plate format in LB medium containing benzamide. These wells were then co-cultivated with sensor cells and analyzed for any fluorescence. Eleven amidases were identified from 143 fluorescent cells out of which three were novel enzymes (Uchiyama and Miyazaki 2010).

3.2.8 Examples of function-based screening of metagenomes for novel enzymes

Examples of enzymes identified through functional-metagenomic approaches are listed in Table 3.

3.3 Combined sequence- and function-based screenings of metagenomes for the detection of novel enzymes

Several studies have been published that apply sequence- and function-based screenings of metagenomes simultaneously. In such cases, one or more of the bioactivity assays (section 3.2) and the homology-based searches (section 3.1) are performed independently on the same metagenomic sample/dataset. The combined method offers higher chances of identification of the desired enzymes. For e.g., if the enzyme of interest did not show a high activity in functional assays (due to any reasons not known), PCR or oligo-based methods might still identify genes for the same enzyme class in the same dataset.

Several recent papers used this combined approach. Antarctica soil metagenomes were screened to identify 14 lipase/esterase-, 14 amylase-, 3 protease-, and 11 cellulase-producing clones by both functional and sequence-based screening (Berlemont, Pipers et al. 2011). In a similar study, the RhaB1 enzyme belonging to a new subclass of bacterial B type α -L-rhamnosidases of the Glycosyl Hydrolases 78 (GH78) family was identified in the metagenomes of elephant feces using a combination of sequence- and function-based analysis (Rabausch, Ilmberger et al. 2014). In another study, three metagenomes of insect gut symbionts were generated to compare the taxonomic and metabolic diversity of gut microbiomes to the diet and life history of their insect hosts (Shi, Xie et al. 2013). To exploit the insect gut symbiosis for biotechnology applications, four biomass-degrading enzymes including one endoglucanase and one xylanase were cloned and characterized. Similarly, sequence-based identification and functional characterization of a novel cellulase gene from a horse feces metagenome revealed a bifunctional cellulolytic enzyme (Chandrasekharaiah, Thulasi et al. 2012). The gut metagenome of the endangered Iberian Lynx revealed the dominance (28% of total GH) of alpha-amylases with low enzymatic activity and 1.5% of beta-xylosidases with significant enzyme activity (Alcaide, Messina et al. 2012). In an innovative study of 268 gigabases of metagenomic DNA from microbes adherent to plant fibers incubated in cow rumen, 27,755 putative CAZymes were identified and 90 candidate proteins were expressed, of which more than half were enzymatically active against cellulosic substrates (Hess, Sczyrba et al. 2011). In this study, 15 uncultured microbial genomes were assembled and validated using complementary single-cell genome sequencing. Deep sequencing and functional screening of a viral metagenome from a hot spring revealed a highly thermostable DNA polymerase (pol) enzyme with inherent reverse transcriptase activity in another study. The benchmarking of this enzyme with other RT-PCR systems proved its potential to be a potent RT-PCR enzyme in research and diagnostics (Moser, DiFrancesco et al. 2012). High-throughput functional screening using fluorogenic and chromogenic substrates combined with automated handling and genetically modified expression of the host resulted in identification of many plant polymer (cellulose, hemicellulose, chitin, starch and protein) decomposing enzymes (Nyyssonen, Tran et al. 2013). This paper describes the amalgamation of both functional screening and sequence-based analysis with

metagenomic DNA to establish a link between microbial functionality and community composition.

3.4 Natural products through metagenomics

Nature has provided humans with very useful products known as “natural products” that are of therapeutic and biotechnological importance (Hunter 2008). They are so important that almost half of the therapeutic drugs available on the market have been derived from them (Newman, Cragg et al. 2000, Newman, Cragg et al. 2003). The ecological functions of bioactive compounds are diverse and include protecting the host from predators, pathogens and competitors, and signaling messages for the presence of food, mates and enemies (Gershenzon and Dudareva 2007). Although terrestrial organisms are supposed to be the richest source of natural products, marine ecosystems might also prove to be an important domain for bioprospecting (Putz and Proksch 2010) due to the fact that over 70% of the earth is covered by oceans and large parts of these oceans still remain unexplored or inaccessible. Marine sponges, in this context, are well known for their vast chemical and microbial diversity (Hochmuth, Niederkruger et al. 2010, Hentschel, Piel et al. 2012) and are an important source of useful bioactive compounds (Noro, Kalaitzis et al. 2012). It has long been thought that the host synthesizes these products, but recent findings suggest bacterial origins for many (Piel 2002, Piel, Hofer et al. 2004, Zimmermann, Engeser et al. 2009). Secondary metabolites inside microbial cells are synthesized by biosynthetic gene clusters (BGCs). Most often, bacteria producing natural products live in symbiotic relationship with eukaryotes like fungi, marine invertebrates, insects or nematodes (Crawford and Clardy 2011). The vast majority of bacterial symbionts have not been successfully cultured (Webster and Hill 2001, Webster, Wilson et al. 2001) creating a major bottleneck in the development of marine-derived drugs due to the limited concentrations of the natural products in situ. Metagenomics not only deciphers the genetic background of the chemical ecology of host-microorganism symbioses, but also enables the possibility of biotechnological production of natural products in heterologous hosts.

3.4.1 Natural products from free-living bacteria

Unlike the discovery of biocatalysts/enzymes (primary metabolites), where a single gene/enzyme is the focus, the identification of natural products (secondary metabolites) may involve many genes in the form of gene clusters. The general design principle applied to identifying gene cluster for natural biosynthetic pathway is more or less similar to that applied for identification of discrete enzymes using sequence- and function-based screening methods. The most common heterologous hosts used for the construction of metagenomic clone libraries have been *Streptomyces* (Wang, Graziani et al. 2000) and *Escherichia coli* (Lorenz and Eck 2005, Brady 2007). A more recent study of *Ralstonia metallidurans* showed that using alternate hosts might increase the number of positive clones in the metagenomic screening procedure (Craig, Chang et al. 2009).

3.4.2 Natural products from symbiotic bacteria

Recently, a growing number of natural products, originally thought to originate from the invertebrates from where they were isolated, have been found to be produced by bacterial symbionts (Piel, Butzke et al. 2005, Piel 2006, Piel 2009). Since most of these symbiotic bacteria are resistant to cultivation, metagenomic approaches to whole organisms have been widely used to identify/clone the corresponding gene clusters responsible for synthesis of natural products (Piel 2011). Cultivation-dependent methods to explore natural products from symbioses have been covered elsewhere (Piel 2004, Piel 2009). There are two general approaches to obtaining natural products from symbiotic bacteria using metagenomics (Brady, Simmons et al. 2009).

Chemistry-directed approaches

Knowledge of the chemical structure of a natural product might be useful to identify or clone its biosynthetic genes using metagenomic techniques. In principle, these approaches are analogous to pathway-cloning from cultured bacteria. In some cases, the symbiotic bacterial strain is known as the producer of the chemical and this information is used to clone the candidate gene cluster from the metagenome. This constitutes the “candidate strain approach” and many bioactive molecules have been identified using this approach, including bryostatin, pederin, and cyanobactin (Lopanik, Shields et al. 2008, Zimmermann, Engeser et al. 2009). Another chemistry-directed approach relies on

“chemical or biochemical homology” and identifies the homology of the respective genes in the metagenomes using degenerate primers or probes. The first example of this approach is the discovery of an onnamide (Piel, Hui et al. 2004) gene cluster from the marine sponge of the genus *Theonella*. Other examples of the chemical-homology approach are palmerolide, barbamide, and dysidenin (Chang, Flatt et al. 2002, Ridley, John Faulkner et al. 2005, Schmidt, Donia et al. 2012).

Gene-directed approaches

Biosynthetic genes for natural products have also been cloned from metagenomes of symbiotic systems without using chemical guidance. One of the very first examples of this approach is the study of polyketide biosynthetic pathways in lichens using molecular genetic techniques such as PCR, library construction and heterologous expression (Miao, Coeffet-LeGal et al. 2001). Recently, many polyketide synthases (PKS) and nonribosomal peptide synthase (NRPS) genes have been cloned from symbiotic associations using this approach (Kim and Fuerst 2006, Fieseler, Hentschel et al. 2007, Hochmuth and Piel 2009).

3.4.3 Examples of natural product discoveries through metagenomics

Examples of natural products, where the gene loci were identified or cloned by metagenomics, are listed in Table S1. Progress in metagenomic applications to natural product research up to 2005 was summarized in a review (Piel, Butzke et al. 2005). Not all the PKS homologs in sponge metagenomes are pharmacologically relevant, but the ones with cis-AT (Acyl Transferase) or trans-AT or NRPS and PKS are. A comprehensive study of metagenomes from 20 demosponge species from the world's oceans showed that only 8% of the PKS sequences belonged to families usually involved in the production of bioactive polyketides (Fieseler, Hentschel et al. 2007). A phylogenetic analysis of PKS sequences (using degenerate PCR primers) revealed that there are few amplicons belonging to the cis-AT and trans-AT type that could be easily distinguished from the FAS (Fatty Acid Synthase) type. These amplicon sequences were used as oligonucleotide probes in the screening of metagenomic libraries for cloned PKS clusters producing bioactive peptides (Fieseler, Hentschel et al. 2007).

3.4.4 Drug discovery through metagenomics

The cultured minority of the soil microbial community, 0.3 % (Amann, Ludwig et al. 1995), has been a remarkable source of bioactive natural products for decades, and one of the best examples are *Actinomycetes* (Brady and Clardy 2000, Marris 2006). The uncultured majority is thought to be at least equally useful in this context and culture-independent or metagenomic approaches are being used increasingly to exploit the great chemical and structural diversity of the soil that has remained hidden until recently. Researchers from Sean Brady's laboratory working on bioactive small molecules started a mega-project of profiling secondary metabolites from different soil microbiomes from the 50 US states and the pilot results were summarized recently (Charlop-Powers, Owen et al. 2014). This initiated a big citizen science project of searching for drugs from dirt and a dedicated website has been created for active participation (<http://www.drugsfromdirt.org/>). The land has been studied for thousands of years for natural products, but similar studies from the oceans were started only half a century ago (O'Hanlon 2006). It has been speculated that the marine environment will provide a new wave of chemical structures not found in microbes from traditional terrestrial habitats (Marris 2006, Hopwood 2007, Montaser and Luesch 2011). The natural products from marine organisms over the period of 28 years from 1985 to 2012 were statistically analyzed for their bioactivity, temporal trends, chemical structures and species distribution in a recent review article (Hu, Chen et al. 2015). This article reported on 15,000 chemical substances including 4196 bioactive natural products from marine sources. The number of bioactive compounds discovered from marine sources has decreased in the last decade, while the number of natural products developed from marine sources is increasing each year. The most important classes of chemical structures are terpenes, alkaloids, ketals and sterides in decreasing order of abundance among the natural products. The distribution of bioactivities revealed that most of the compounds are anticancerous (56%), followed by anti-bacterial and anti-fungal (13% and 5%, respectively). The major sources of natural products in the marine environment are marine invertebrates (~ 75%). With this enormous amount of chemical and bioactive diversity, marine natural products might serve as a treasure trove for new drug leads for therapeutic use.

Surprisingly, not a single natural product identified and characterized using metagenomic approaches has been approved as a drug molecule though efforts are being made globally along this line. Nonetheless, the natural products whose biosynthetic loci have been identified/cloned using metagenomic approaches might be suitable candidates for the preclinical drug discovery pipeline and are listed in Table S1. Recently, a biosynthetic gene cluster (NRPS pathway) for Yondelis (Et-743), an approved anticancer drug, was cloned and analyzed using metagenomic sequencing of the total DNA from a tunicate-microbial consortium (Rath, Janto et al. 2011). This molecule is obtained in low quantity from the tunicate *Ecteinascidia turbinata* and generated for clinical use by a cumbersome semisynthetic method. The work by Rath et al. (2011) laid the foundation for direct production of the drug and its analogues using metagenomic approaches. The scope of metagenomics, together with metabolic engineering and other contemporary technologies, in the drug discovery pipeline is broad and can be envisaged by the proportion of small molecule compounds and biologicals in the total number of approved drugs by the US FDA. Thus, metagenomics can be used either to clone the biosynthetic gene clusters for already approved drugs or to discover entirely new chemical entities from diverse natural resources. Interesting review articles about the possible contribution and application of metagenomics in the drug discovery pipeline include (Schmidt and Donia 2010, Iqbal, Feng et al. 2012, Schmidt, Donia et al. 2012).

3.5 Metagenomic enrichment technologies

The low abundance of certain genes or genomes in metagenomic samples seriously affects downstream functional screening procedures to identify rare and novel biocatalysts (Cowan, Meyer et al. 2005). Pre-enrichment of the metagenomic samples significantly enhances the screening hit rate. A simple example of enrichment at whole-cell level is the Sargasso Sea genome sequencing project, where size-selective filters were applied to remove eukaryotic cells (Venter, Remington et al. 2004). Another common enrichment method at nucleic acid level is Stable Isotope Probing (SIP) where a growth substrate labeled with a stable isotope (^{13}C , ^{15}N etc.) is fed to the enrichment culture or soil sample and links the respective microbial function to a specific taxon or

subpopulation via selective recovery of labeled DNA (Radajewski, McDonald et al. 2003, Kalyuzhnaya, Lapidus et al. 2008). Suppressive subtractive hybridization (SSH) (Galbraith, Antonopoulos et al. 2004), differential expression analysis (DEA) (Green, Simons et al. 2001), phage display (Cramer and Suter 1993, Ciric, Moon et al. 2014), affinity capture (Demidov, Bukanov et al. 2000), and microarrays (Wu, Thompson et al. 2001) are other technologies that might be used for metagenomic enrichment at various stages of sampling, nucleic acid extraction and clone library preparation.

3.6 Bottlenecks in the metagenomics discovery pipeline

While metagenomics has been used widely to identify novel enzymes or natural products, there are challenges and issues associated with it. (Thomas, Gilbert et al. 2012, Leis, Angelov et al. 2013).

3.6.1 Sampling metagenomes: random or intuitive

Abiotic factors such as trophic levels, biogeochemical fluxes, and physicochemical properties inherent to any natural ecosystem influence the composition of microbial communities. Knowledge of the environmental parameters underlying samples may aid in the selection of suitable screen targets and substrates for metagenomic functional screening (Taupp, Mewis et al. 2011). For example, clone libraries from the bovine rumen metagenome are a logical source for carbohydrate active enzymes (CAZymes) (Patel, Patel et al. 2014). The rumen microbiome provides a rich source of enzymes that digest complex plant oligosaccharides given the natural diet of the host. In another study, enzymes catabolizing ethanol as a substrate were identified based on knowledge of ethanol production as a bi-product in anaerobic digestors (Wexler, Bond et al. 2005). Another method is to look for the desired enzymes in those natural environment where the conditions are most favorable for that particular enzyme. Obvious examples are cold-adapted (Berlemont, Pipers et al. 2011, Jansson and Tas 2014, Martinez-Martinez, Lores et al. 2014, Tchigvintsev, Tran et al. 2014) and heat-adapted (Xia, Ju et al. 2013, Elleuche, Schroder et al. 2014, Schroder, Elleuche et al. 2014) enzymes that have been identified and isolated from extremely cold and hot habitats, respectively. The isolated extremozymes have definite industrial applications (Elleuche, Schroder et al. 2014).

3.6.2 DNA extraction

The quality and yield of DNA extracts from a metagenomic sample are very crucial for downstream analyses. The very first consideration in metagenomic DNA extraction is DNA size. If the downstream analysis includes next-generation sequencing, amplicon sequencing (PCR amplification-based sequencing), and small-insert clone libraries, then harsh extraction methods, leading to substantially sheared but highly pure DNA, can be used. For large-insert metagenomic libraries, alternative extraction methods are adopted to obtain intact DNA of high molecular weight (Liles, Williamson et al. 2009). The DNA extraction method should be sensitive enough so that the metagenomic DNA captures the complete microbial diversity of the sample. The factors that affect the DNA extraction sensitivity in context of the commonly used bead beating method are the beating time as well as speed, volume and temperature of the buffer, and types and amount of used beads. Different treatments reflect significant differences in the ratio of band intensities of restriction fragment length polymorphism patterns (RFLP) of bacterial and archaeal communities. But, this difference was not observed in the case of eukarya and high-GC Gram-positive bacteria. Failing to take this into account could lead to over- and under-representation of some microbial taxa in the downstream analysis (Liles, Manske et al. 2003, Feinstein, Sul et al. 2009). The second consideration in metagenomic DNA extraction is DNA yield from two viable methods: 1) direct extraction and 2) indirect extraction. In the direct extraction method, DNA is isolated from the environmental sample, which has many advantages such as short processing time and greater DNA yield (Ogram, Sayler et al. 1987). The major disadvantage of the direct extraction method is the higher proportion of extracellular non-bacterial DNA (Tsai and Olson 1991, Tebbe and Vahjen 1993). In the indirect extraction method, the microbial cells are separated before cell lysis and DNA extraction and purification, resulting in less non-bacterial DNA (Osborn and Smith 2005). Comparative benchmarking of direct and indirect extraction methods showed that the indirect method yielded 10- to 100-fold lower amounts of DNA than did direct procedures, but revealed a much higher bacterial diversity (Gabor, de Vries et al. 2003). The disadvantages of indirect methods are the longer processing time and lower DNA yields. One recent study demonstrated quantitatively the effect of DNA extraction methods on the microbial diversity of the

sampled soil and concluded that the yield and diversity of DNA greatly depends on the extraction method and suggested that the use of a novel extraction method may result in having DNA from a microbial taxon that otherwise would not be identified (Inceoglu, Hoogwout et al. 2010).

3.6.3 Choice of vectors

Generally, the choice of the cloning vector depends on the degree of fragmentation of the metagenomic DNA after isolation and purification. Small inserts (<20 kb) are cloned into a plasmid, medium inserts from 20 to 40 kb into fosmids and cosmids, and large-inserts up to ~200 kb into BAC vectors. The advantages of small-insert clone libraries include having fewer genes per insert to express, high copy numbers, and good transcription under the strong promoters of the plasmids. Together, these features allow the detection of even weakly expressed enzymes. Large-insert libraries tend to contain and express multiple genes or entire operons by the native promoters present on the insert itself. The stability of large-insert vectors inside the host can be a problem though. To overcome this, fertility (F)-factor-based origin of replication from *Escherichia coli* was developed as a single-copy origin of replication for fosmid, cosmid (Kim, Shizuya et al. 1992) and BAC (Shizuya, Birren et al. 1992) vectors, rendering a high degree of stability inside the host. BAC systems could maintain inserts as large as 300 kb stably in *E. coli* cells for as many as 100 generations. Despite their high stability, BACs suffer from low DNA yield due to the single-copy construct. Development of conditionally amplifiable BACs switching from single-copy to high-copy vectors could be a solution with high yields of DNA obtained from BAC clones, which would retain all the advantages of the single-copy BAC clone (Wild, Hradecna et al. 2002). More sophisticated commercial cloning vectors are available to meet specific needs such as pCC1FOS from Epicentre (Bohnke and Perner 2014) and pJAZZ from Lucigen [<http://lucigen.com/pJAZZ-OK/>].

Even if the metagenomic DNA is cloned successfully, it does not guarantee its expression in a specific heterologous host and thus genes or pathways might remain undetected. To maximize the discovery of novel enzymes and secondary metabolites by functional metagenomics, alternate host-vector systems have been successfully developed and implemented (McMahon, Guan et al. 2012, Liebl, Angelov et al. 2014). The vector component should have a broad range and be capable of shuttling metagenomic inserts to

multiple hosts for expression screening. Very early examples of the application of shuttle cosmid or BAC vectors are those where metagenomic libraries were first produced in *E. coli* and transferred to other hosts such as *Streptomyces lividans* and *Pseudomonas putida* for functional screening of natural products leading to drug discovery (Courtois, Cappellano et al. 2003, Martinez, Kolvek et al. 2004). A plasmid RK2-based broad-host range vector (pRS44/pTA44) was used to transfer metagenomic libraries to a variety of bacterial species like *E. coli*, *Pseudomonas fluorescens* and *Xanthomonas compestris* (Aakvik, Degnes et al. 2009). In a more recent and comprehensive study, broad-host-range cosmid environmental DNA libraries were screened in parallel for small molecules across six different proteobacterial hosts (Craig, Chang et al. 2010). In examples presented here, there was no or very little overlap of phenotypes and clones produced by different hosts expressing the same metagenomic library, suggesting the use of alternate host-vector systems for increasing the success rate in metagenomic functional screenings.

3.6.4 Choice of hosts

The choice of the cloning and expression hosts depends on many factors such as the vector systems used for cloning the metagenomes and the phylogenetic and genetic relatedness of the host with the microbial population inhabiting the metagenomic sampling site. The criteria for ideal host-vector systems are four, two related to cloning vectors and two to hosts (Angelov, Mientus et al. 2009). Cloning vectors should be easily transferable and maintainable in the new host in order not to lose or modify the inserted metagenomic DNA. The host should be genetically accessible to add or delete certain genomic elements and easily cultured and screened for desired function for selection of metagenomic clones. Although *Escherichia coli* has been used for decades as the most common expression host for functional metagenomics and has been useful in the discovery of novel enzymes, the frequency of the positive clones as well as the structural diversity of identified small molecules has remained very low (Piel 2011). For example, in an antibacterial screen, one out of every 10,000 to 20,000 metagenomic cosmid clones exhibited antibacterial activity (Brady, Chao et al. 2004). Another example reported one positive clone with antifungal activity out of 113,700 fosmid clones when *E. coli* was used as a host (Chung, Lim et al. 2008). The quantification of screening sensitivity (number of positives/number of screened clones) in *E. coli* hosts was reviewed by

Uchiyama and Miyazaki (Uchiyama and Miyazaki 2009). Their results show generally low hit rates in functional screenings. An *in silico* study to assess the ability of an *E. coli* system to express genes from 32 taxonomically diverse genomes showed that only 40% of the genes could be expressed (Gabor, Alkema et al. 2004). Another impressive study on the quantification of transcription of foreign DNA (*Haemophilus influenza*, *Pseudomonas aeruginosa*, and human) in *E. coli* showed that nearly 50% of *H. influenza* genes and a much smaller proportion of *P. aeruginosa* genes are transcribed in *E. coli* (Warren, Freeman et al. 2008). The transcription of human DNA in *E. coli* was widespread, but punctuated. In 2007, an attempt to experimentally determine barriers of horizontal gene transfer was conducted by cloning about 250,000 genes from 79 prokaryotic genomes into *E. coli*. The study found that toxicity of the foreign genes for the host was the major factor inhibiting cloning into *E. coli* (Sorek, Zhu et al. 2007). The reasons for low or no heterologous gene expression in *E. coli* host are numerous and include different codon usages, improper promoter recognition, lack of initiation factors, improper protein folding, lack of essential co-factors and biosynthetic precursors, enzymatic breakdown of the gene product, toxicity of the gene product, inability of the host to secrete the gene product, and lack of posttranslational modification of the *apo* proteins in the host (Uchiyama and Miyazaki 2009, Piel 2011, Ekkers, Cretioiu et al. 2012). These authors also suggested possible solutions to the problems associated with heterologous gene expressions in *E. coli*. To circumvent most, if not all, of these expression-related issues, non-*E. coli* hosts have been developed. A summary of alternate expression hosts is presented in Table S2. Apart from using alternate hosts, *E. coli* strains have also been improved to increase metagenomic gene expression. In some studies, heterologous sigma factors, capable of recognizing heterologous promoters, were expressed in *E. coli* cells enabling higher expression of metagenomic DNA (Stevens, Conway et al. 2013, Gaida, Sandoval et al. 2015). The insertion of a constitutive promoter at the multiple cloning site (MCS) of the plasmid vector in expression host showed a significantly higher expression of the gene placed downstream to the MCS region (Frebourg and Brison 1988). This approach has been applied to produce engineered *Streptomyces* and *Sulfolobus* species to enhance heterologous gene expression in these model systems (Guan, Cui et al. 2015, Hwang, Choi et al. 2015).

4 Single-cell genomics: Tapping the untapped for application in Biotechnology

Advances in microfluidics-based cell separation techniques together with whole genome amplification of genetic material isolated from a single cell laid the foundation for the single-cell genomics technique (Shapiro, Biezuner et al. 2013). The amount of DNA isolated from a single cell, which is in the picogram range, is not sufficient for DNA sequencing methods. Many of the early developments in this field have made it possible to amplify the plasmid and phage DNA (Dean, Nelson et al. 2001, Detter, Jett et al. 2002) as well as human genomic DNA from clinical samples (Dean, Hosono et al. 2002, Hosono, Faruqi et al. 2003, Lage, Leamon et al. 2003) using Phi 29 DNA polymerase and exonuclease-resistant random hexamer primers, and the technology called “multiple displacement amplification” (MDA). The MDA technique was initially used to amplify human genomic DNA in microgram quantities from precious samples and clinical specimens for genetic testing (Lasken and Egholm 2003). Subsequently, it was applied to pre-implantation genetic diagnosis of single gene effects by whole genome amplification from single or small numbers of lymphocytes and blastomeres isolated from embryos (Handyside, Robinson et al. 2004). Not only limited to human developmental studies, MDA was used to successfully amplify genomic DNA from a single bacterium by about 5 billion fold (Raghunathan, Ferguson et al. 2005).

The first application of MDA and single-cell genomics in the field of microbial ecology was the publication of a partial genome sequence of an uncultured archaeon from a soil sample (Kvist, Ahring et al. 2007). Subsequently, the genomes of cells of the rare and uncultivated candidate phylum TM7 were published from two ecological niches, one from a human mouth sample (Marcy, Ouverney et al. 2007), and the other from a soil sample (Podar, Abulencia et al. 2007). Although, single-cell genomics proved to be a great novel technology to study genomes from individual cells, the coverage of the genome sequenced after MDA represented no more than ~70% (Marcy, Ouverney et al. 2007, Woyke, Xie et al. 2009). The fine-tuning in MDA methods coupled with post-amplification normalization to mitigate large amounts of variation in sequencing coverage resulted in a high-throughput single-cell genome sequencing pipeline enabling recovery as high as 95% of a *Prochlorococcus* genome (Rodrigue, Malmstrom et al.

2009). So far, single-cell genomics has been used comprehensively to discover many uncultured rare bacterial lineages over the years (Lasken and McLean 2014, Kamanda Ngugi, Blom et al. 2015) as listed in Table 4. The largest study to date using single-cell sequencing at the US Department of Energy Joint Genome Institute generated partial single-cell amplified genomes (SAGs), ranging from ~150 kbp to 2.4 Mbp in size, belonging to 29 major uncharted branches in the evolutionary tree (Rinke, Schwientek et al. 2013). In this study, many novel candidate bacterial phyla along with the diverse archaeal groups were sequenced and their metabolic potentials were predicted to unravel unexpected metabolic features like the opal stop codon UGA coding for glycine, OP11 bacteria using the archaeal pathway for purine biosynthesis, and nanoarchaeal genome encoding complete bacteria-like sigma factors. The outstanding contribution of this novel genomic technique was recognized when “single-cell genomics” was regarded by “Nature Methods” as the method of the year in 2013 (Chi 2014).

4.1 Analyses of single-cell genomes for biotechnology

Many single-cell genomes have been annotated and various enzymes have been identified through bioinformatics, suggesting their metabolic potential and possible ecological roles in the natural environment (see Table 4). One of the challenges in sequenced-based screening approaches is that genes for completely novel enzymes are mostly likely overlooked. In fact, especially when working with sequence data from so far uncultured lineages, as often the case for SAGs, a large percentage of the open reading frames do not encode for genes with significant homologies to those in existing databases. As single-cell genomes become more and more affordable and commonplace in microbiology labs through reduced sequencing costs, we see an opportunity there to use the amplified DNA for functional screening methods as well. So far, there are only a few important studies utilizing single-cell MDA products for high-throughput screening in microorganisms (Lasken 2012). But, to the best of our knowledge, none of the single-cell MDA products have been screened using functional screening procedures like shotgun cloning to discover novel genes/enzymes for biotechnological applications. In contrast, functional metagenomics has yielded a remarkable number of such products of industrial importance (Lorenz and Eck 2005). Although single-cell amplified DNA might be useful

for functional screening owing to its predetermined taxonomic affiliations and genetic homogeneity, which would help in choosing the right expression host system for the functional screening, it has a few inherent potential issues, including amplification bias and chimera formation (Lasken and McLean 2014).

The first attempt to screen MDA products for onnamide and polytheonamide biosynthetic gene clusters using PCR and cloning from SAGs and metagenomic assemblies of a sponge symbiont, *Candidatus Entotheonella*, was demonstrated very recently from the marine sponge *Theonella swinhoei* (Wilson, Mori et al. 2014). This study marks the beginning of function-based screening to identify secondary metabolites or natural products from single cells isolated from environmental samples.

5 Proposition: Functional single-cell genomics

The application of metagenomics has resulted in a lot of novel enzymes or secondary metabolites for potential industrial use (Iqbal, Feng et al. 2012, Lee and Lee 2013, Lewin, Wentzel et al. 2013, Lopez-Lopez, Cerdan et al. 2013). The practical amalgamation of other “omics” technologies like metatranscriptomics and metaproteomics with metagenomics is emerging as a powerful tool set to study complex microbial communities to help to identify biological molecules for biotechnological purposes (de Castro, Sartori da Silva et al. 2013). In contrast to functional metagenomics, functional single-cell genomics has not been exploited yet, except for very few examples (Woyke and Jarett 2015) for the identification of novel biological molecules from the uncultured microbial majority inhabiting natural environments. We believe that “functional single-cell genomics” will undoubtedly increase the scope of identifying novel enzymes or molecules. The obvious advantages of single-cell genomics (over metagenomics) are the genetic homogeneity of the sequences obtained (due to the sequences coming from a single organism) and taxonomic assignment of the genomic sequences.

Like any other technology, single cell genomics also suffers from shortcomings. One prominent issue is the amplification bias towards certain genomic regions during the MDA reaction leading to uneven coverage of the sequenced single cell genome. The reason for this bias was initially thought to be related with GC content (Pinard, de Winter et al. 2006), but many other studies provided evidence for random over-amplification

resulting from stochastic priming and amplification in the beginning of the MDA reaction (Zhang, Martiny et al. 2006, Marcy, Ishoey et al. 2007). Another major challenge in using MDA reactions is the formation of chimeric sequences joining two non-contiguous regions of a genome (Lasken and Stockwell 2007). It has been shown that 2-4% of sequencing reads in the 454-FLX single-cell libraries are chimeric (Rodrigue, Malmstrom et al. 2009). The challenges associated with whole genome amplification using MDA and single cell sequencing together with the strategies to minimize those bottlenecks have been studied recently (Rodrigue, Malmstrom et al. 2009, Lasken 2012, Lasken and McLean 2014). After the successful round of MDA, if the SAG DNA needs to be cloned for a potential functional screening (as we propose in this manuscript) certain points should be considered carefully. One obvious pitfall is the redundancy of the DNA sequence owing to the whole-genome amplification, which will result in redundant clones in functional screening. Another limiting factor is the shorter MDA product size, for e.g. using Illustra Genomiphi V2 DNA amplification kit (Kumar, Rech et al. 2007) an average of 10 kb of products are obtained which is less likely to contain biosynthetic gene clusters coding for natural products. But, this is still an average value and many MDA products might have the bigger size in order to contain large biosynthetic gene loci. Moreover there will not be any effect on screening MDA products for single enzymes, which generally have smaller sizes.

6 Conclusion

Most metagenome-based enzyme discoveries resulted from functional metagenomics (Ferrer, Beloqui et al. 2009, Tuffin, Anderson et al. 2009). GHs and lipase/esterases constitute the highest proportion among those enzymes (Simon and Daniel 2009, Steele, Jaeger et al. 2009) possibly due to the rapid and easy plate-based detection assays for these two types of enzymes (Taupp, Mewis et al. 2011). The lack of availability of assay methods or more specifically of high-throughput screening methods for enzymes other than hydrolases/lipases has impeded the discovery of more biotechnologically important enzymes. The combination of metagenomics with metagenomic enrichment technologies is powerful enough to increase the screening hit rate (Verastegui, Cheng et al. 2014). The heterologous expression of metagenomic DNA has always been one of the main

bottlenecks in functional screening for useful enzymes and natural products. The development and use of broad-range host vectors and alternate hosts (see Table S2) have addressed this issue considerably. An increasing number of biosynthetic gene loci are being cloned from metagenomic samples both from symbiotic microorganisms as well as from environmental samples, which will provide new lead molecules for the pre-clinical phases in the drug discovery pipeline (Rath, Janto et al. 2011).

The significant advances (in terms of ever reducing cost, shorter sequencing time, easy availability) in high-throughput sequencing methods have generated enormous amounts of metagenomic sequence data from virtually every habitat on earth. With these large datasets, scientists were able to decipher the microbial community structure and function of many environmental niches that were not explored earlier. One of the challenges now is how to store, process and distribute the data among the scientific community. Although considerable efforts in terms of web-based metagenomic resources are being made, we are still far away from a satisfactory solution (Sharpton 2014). The functional annotation of the sequence data critically depends upon the reference genome sequence database and novel sequences coming from the unexplored environmental samples create another challenge to functional annotation especially when there are no hits from the existing reference databases. This problem could be partly alleviated by sequencing of novel reference genomes/genes from the uncharted branches in tree of life using metagenomics (Venter, Remington et al. 2004) and single cell genomics (Rinke, Schwientek et al. 2013).

The advent of single cell genomics was indeed a great leap in studying uncultivated microbial majority compared to metagenomics. The obvious advantages of single cell genomics over metagenomics are: i) it does not require error-prone assembly or binning methods to reconstruct a microbial genome from a complex community as metagenomics, ii) it preserves the taxonomic identification of the single cell and iii) it enables characterization of microbes one cell at a time with the help of other omics technologies like transcriptomics, proteomics, and metabolomics. Many genomes are routinely sequenced now by this technology and the amount of sequence data is increasing. However, single cell genomics will soon benefit from other technologies like single molecule real-time sequencing (SMRT) (Eid, Fehr et al. 2009, Korlach, Bjornson

et al. 2010) and nanopore based single-molecule sequencing (McNally, Singer et al. 2010) to bypass the amplification step in the existing single cell genomics protocol. The single cell sequence data so far generated necessitates functional validation in order to understand the ecological role of the single cells as well as selecting the desired phenotypes for biotechnology application. Although few discoveries (Martinez-Garcia, Brazel et al. 2012, Wilson, Mori et al. 2014) have marked the beginning of the proposed idea, there is an enormous potential in the coming years for function-driven single cell genomics to provide enzymes or biomolecules of industrial importance.

Acknowledgement: This study was sponsored by King Abdullah University for Science and Technology (KAUST).

References

- Aakvik, T., K. F. Degnes, R. Dahlsrud, F. Schmidt, R. Dam, L. Yu, U. Volker, T. E. Ellingsen and S. Valla (2009). "A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species." *FEMS Microbiol Lett* **296**(2): 149-158.
- Abdallah, A. M., M. Rashid, S. A. Adroub, M. Arnoux, S. Ali, D. van Soolingen, W. Bitter, A. Pain and H. Elabdalaoui (2012). "Complete genome sequence of Mycobacterium phlei type strain RIVM601174." *J Bacteriol* **194**(12): 3284-3285.
- Abdallah, A. M., M. Rashid, S. A. Adroub, H. Elabdalaoui, S. Ali, D. van Soolingen, W. Bitter and A. Pain (2012). "Complete genome sequence of Mycobacterium xenopi type strain RIVM700367." *J Bacteriol* **194**(12): 3282-3283.
- Albertsen, M., P. Hugenholtz, A. Skarszewski, K. L. Nielsen, G. W. Tyson and P. H. Nielsen (2013). "Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes." *Nat Biotechnol* **31**(6): 533-538.
- Alcaide, M., E. Messina, M. Richter, R. Bargiela, J. Peplies, S. A. Huws, C. J. Newbold, P. N. Golyshin, M. A. Simon, G. Lopez, M. M. Yakimov and M. Ferrer (2012). "Gene sets for utilization of primary and secondary nutrition supplies in the distal gut of endangered Iberian lynx." *PLoS One* **7**(12): e51521.
- Amann, R. L., W. Ludwig and K. H. Schleifer (1995). "Phylogenetic identification and in situ detection of individual microbial cells without cultivation." *Microbiol Rev* **59**(1): 143-169.
- Andexer, J., J. K. Guterl, M. Pohl and T. Eggert (2006). "A high-throughput screening assay for hydroxynitrile lyase activity." *Chem Commun (Camb)*(40): 4201-4203.
- 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**(3): 177-185.
- Antunes, A., D. K. Ngugi and U. Stingl (2011). "Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes." *Environ Microbiol Rep* **3**(4): 416-433.
- Aziz, R. K., M. Breitbart and R. A. Edwards (2010). "Transposases are the most abundant, most ubiquitous genes in nature." *Nucleic Acids Res* **38**(13): 4207-4217.
- Bao, L., Q. Huang, L. Chang, Q. Sun, J. Zhou and H. Lu (2012). "Cloning and characterization of two beta-glucosidase/xylosidase enzymes from yak rumen metagenome." *Appl Biochem Biotechnol* **166**(1): 72-86.
- Beijerinck, M. W. (1888). "The root-nodule bacteria." *Bot. Zeitung* **46**: 725-804.
- Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich and E. F. DeLong (2000). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." *Science* **289**(5486): 1902-1906.
- Beja, O., E. N. Spudich, J. L. Spudich, M. Leclerc and E. F. DeLong (2001). "Proteorhodopsin phototrophy in the ocean." *Nature* **411**(6839): 786-789.
- Beja, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich, R. A. Feldman and E. F. DeLong (2000).

"Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage." *Environ Microbiol* **2**(5): 516-529.

Bendich, A. J. and B. J. McCarthy (1970). "Ribosomal RNA homologies among distantly related organisms." *Proc Natl Acad Sci U S A* **65**(2): 349-356.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, E. C. M. Chiara, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G. D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin and A. J. Smith (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." *Nature* **456**(7218): 53-59.

Berlemont, R., D. Pipers, M. Delsaute, F. Angiono, G. Feller, M. Galleni and P. Power (2011). "Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements." *Rev Argent Microbiol* **43**(2): 94-103.

Binga, E. K., R. S. Lasken and J. D. Neufeld (2008). "Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology." *ISME J* **2**(3): 233-241.

Bohnke, S. and M. Perner (2014). "A function-based screen for seeking RubisCO active clones from metagenomes: novel enzymes influencing RubisCO activity." *ISME J*.

- Bornscheuer, U. T. (2002). "Methods to increase enantioselectivity of lipases and esterases." Curr Opin Biotechnol **13**(6): 543-547.
- Brady, S. F. (2007). "Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules." Nat Protoc **2**(5): 1297-1305.
- Brady, S. F., C. J. Chao and J. Clardy (2004). "Long-chain N-acyltyrosine synthases from environmental DNA." Appl Environ Microbiol **70**(11): 6865-6870.
- Brady, S. F. and J. Clardy (2000). "Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA." Journal of the American Chemical Society **122**(51): 12903-12904.
- Brady, S. F., L. Simmons, J. H. Kim and E. W. Schmidt (2009). "Metagenomic approaches to natural products from free-living and symbiotic organisms." Nat Prod Rep **26**(11): 1488-1503.
- Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam and F. Rohwer (2002). "Genomic analysis of uncultured marine viral communities." Proc Natl Acad Sci U S A **99**(22): 14250-14255.
- Brosius, J., M. L. Palmer, P. J. Kennedy and H. F. Noller (1978). "Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli." Proc Natl Acad Sci U S A **75**(10): 4801-4805.
- Bunterngsook, B., P. Kanokratana, T. Thongaram, S. Tanapongpipat, T. Uengwetwanit, S. Rachdawong, T. Vichitsoonthonkul and L. Eurwilaichitr (2010). "Identification and characterization of lipolytic enzymes from a peat-swamp forest soil metagenome." Biosci Biotechnol Biochem **74**(9): 1848-1854.
- Campbell, J. H., P. O'Donoghue, A. G. Campbell, P. Schwientek, A. Sczyrba, T. Woyke, D. Soll and M. Podar (2013). "UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota." Proc Natl Acad Sci U S A **110**(14): 5540-5545.
- Carbon, P., J. P. Ebel and C. Ehresmann (1981). "The sequence of the ribosomal 16S RNA from Proteus vulgaris. Sequence comparison with E. coli 16S RNA and its use in secondary model building." Nucleic Acids Res **9**(10): 2325-2333.
- Chandrasekharaiah, M., A. Thulasi, M. Bagath, D. P. Kumar, S. S. Santosh, C. Palanivel, V. L. Jose and K. T. Sampath (2012). "Identification of cellulase gene from the metagenome of Equus burchelli fecal samples and functional characterization of a novel bifunctional cellulolytic enzyme." Appl Biochem Biotechnol **167**(1): 132-141.
- Chang, Y. J., M. Land, L. Hauser, O. Chertkov, T. G. Del Rio, M. Nolan, A. Copeland, H. Tice, J. F. Cheng, S. Lucas, C. Han, L. Goodwin, S. Pitluck, N. Ivanova, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, K. Mavromatis, K. Liolios, T. Brettin, A. Fiebig, M. Rohde, B. Abt, M. Goker, J. C. Detter, T. Woyke, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H. P. Klenk and A. Lapidus (2011). "Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium Ktedonobacter racemifer type strain (SOSP1-21)." Stand Genomic Sci **5**(1): 97-111.
- Chang, Z., P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis and D. H. Sherman (2002). "The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit." Gene **296**(1-2): 235-247.

- Charlop-Powers, Z., J. G. Owen, B. V. Reddy, M. A. Ternei and S. F. Brady (2014). "Chemical-biogeographic survey of secondary metabolism in soil." Proc Natl Acad Sci U S A **111**(10): 3757-3762.
- Chi, K. R. (2014). "Singled out for sequencing." Nat Methods **11**(1): 13-17.
- Chistoserdova, L. (2010). "Recent progress and new challenges in metagenomics for biotechnology." Biotechnol Lett **32**(10): 1351-1359.
- Choi, J. E., M. A. Kwon, H. Y. Na, D. H. Hahm and J. K. Song (2013). "Isolation and characterization of a metagenome-derived thermoalkaliphilic esterase with high stability over a broad pH range." Extremophiles **17**(6): 1013-1021.
- Chow, J., F. Kovacic, Y. Dall Antonia, U. Krauss, F. Fersini, C. Schmeisser, B. Lauinger, P. Bongen, J. Pietruszka, M. Schmidt, I. Menyes, U. T. Bornscheuer, M. Eckstein, O. Thum, A. Liese, J. Mueller-Dieckmann, K. E. Jaeger and W. R. Streit (2012). "The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases." PLoS One **7**(10): e47665.
- Chung, E. J., H. K. Lim, J. C. Kim, G. J. Choi, E. J. Park, M. H. Lee, Y. R. Chung and S. W. Lee (2008). "Forest soil metagenome gene cluster involved in antifungal activity expression in Escherichia coli." Appl Environ Microbiol **74**(3): 723-730.
- Ciric, M., C. D. Moon, S. C. Leahy, C. J. Creevey, E. Altermann, G. T. Attwood, J. Rakonjac and D. Gagic (2014). "Metasecretome-selective phage display approach for mining the functional potential of a rumen microbial community." BMC Genomics **15**: 356.
- Connon, S. A. and S. J. Giovannoni (2002). "High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates." Appl Environ Microbiol **68**(8): 3878-3885.
- Council, N. R. (2007). The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. Washington (DC), National Academies Press.
- Courtois, S., C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerinneau, P. Jeannin, P. Simonet and J. L. Pernodet (2003). "Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products." Appl Environ Microbiol **69**(1): 49-55.
- Cowan, D., Q. Meyer, W. Stafford, S. Muyanga, R. Cameron and P. Wittwer (2005). "Metagenomic gene discovery: past, present and future." Trends Biotechnol **23**(6): 321-329.
- Craig, J. W. and S. F. Brady (2011). "Discovery of a metagenome-derived enzyme that produces branched-chain acyl-(acyl-carrier-protein)s from branched-chain alpha-keto acids." Chembiochem **12**(12): 1849-1853.
- Craig, J. W., F. Y. Chang and S. F. Brady (2009). "Natural products from environmental DNA hosted in Ralstonia metallidurans." ACS Chem Biol **4**(1): 23-28.
- 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**(5): 1633-1641.
- Cramer, R. and M. Suter (1993). "Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional

- gene products linked to the genetic information responsible for their production." *Gene* **137**(1): 69-75.
- Crawford, J. M. and J. Clardy (2011). "Bacterial symbionts and natural products." *Chem Commun (Camb)* **47**(27): 7559-7566.
- Culligan, E. P., J. R. Marchesi, C. Hill and R. D. Sleator (2014). "Combined metagenomic and phenomic approaches identify a novel salt tolerance gene from the human gut microbiome." *Front Microbiol* **5**: 189.
- Daniel, R. (2005). "The metagenomics of soil." *Nat Rev Microbiol* **3**(6): 470-478.
- Das, S., S. Paul, S. K. Bag and C. Dutta (2006). "Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation." *BMC Genomics* **7**: 186.
- de Castro, A. P., M. R. Sartori da Silva, B. F. Quirino and R. H. Kruger (2013). "Combining "omics" strategies to analyze the biotechnological potential of complex microbial environments." *Curr Protein Pept Sci* **14**(6): 447-458.
- Dean, F. B., S. Hosono, L. Fang, X. Wu, A. F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S. F. Kingsmore, M. Egholm and R. S. Lasken (2002). "Comprehensive human genome amplification using multiple displacement amplification." *Proc Natl Acad Sci U S A* **99**(8): 5261-5266.
- Dean, F. B., J. R. Nelson, T. L. Giesler and R. S. Lasken (2001). "Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification." *Genome Res* **11**(6): 1095-1099.
- Degnan, P. H. and H. Ochman (2012). "Illumina-based analysis of microbial community diversity." *ISME J* **6**(1): 183-194.
- Demidov, V. V., N. O. Bukanov and D. Frank-Kamenetskii (2000). "Duplex DNA capture." *Curr Issues Mol Biol* **2**(1): 31-35.
- Detter, J. C., J. M. Jett, S. M. Lucas, E. Dalin, A. R. Arellano, M. Wang, J. R. Nelson, J. Chapman, Y. Lou, D. Rokhsar, T. L. Hawkins and P. M. Richardson (2002). "Isothermal strand-displacement amplification applications for high-throughput genomics." *Genomics* **80**(6): 691-698.
- Di Bella, J. M., Y. Bao, G. B. Gloor, J. P. Burton and G. Reid (2013). "High throughput sequencing methods and analysis for microbiome research." *J Microbiol Methods* **95**(3): 401-414.
- Dodsworth, J. A., P. C. Blainey, S. K. Murugapiran, W. D. Swingley, C. A. Ross, S. G. Tringe, P. S. Chain, M. B. Scholz, C. C. Lo, J. Raymond, S. R. Quake and B. P. Hedlund (2013). "Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage." *Nat Commun* **4**: 1854.
- Editorial (2009). "Metagenomics versus Moore's law." *Nat Meth* **6**(9): 623-623.
- Edwards, R. A., B. Rodriguez-Brito, L. Wegley, M. Haynes, M. Breitbart, D. M. Peterson, M. O. Saar, S. Alexander, E. C. Alexander, Jr. and F. Rohwer (2006). "Using pyrosequencing to shed light on deep mine microbial ecology." *BMC Genomics* **7**: 57.
- Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A.

- Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach and S. Turner (2009). "Real-time DNA sequencing from single polymerase molecules." *Science* **323**(5910): 133-138.
- Ekkers, D. M., M. S. Cretoiu, A. M. Kielak and J. D. Elsas (2012). "The great screen anomaly--a new frontier in product discovery through functional metagenomics." *Appl Microbiol Biotechnol* **93**(3): 1005-1020.
- Elleuche, S., C. Schroder, K. Sahm and G. Antranikian (2014). "Extremozymes-biocatalysts with unique properties from extremophilic microorganisms." *Curr Opin Biotechnol* **29C**: 116-123.
- Fang, Z., W. Fang, J. Liu, Y. Hong, H. Peng, X. Zhang, B. Sun and Y. Xiao (2010). "Cloning and characterization of a beta-glucosidase from marine microbial metagenome with excellent glucose tolerance." *J Microbiol Biotechnol* **20**(9): 1351-1358.
- Feinstein, L. M., W. J. Sul and C. B. Blackwood (2009). "Assessment of bias associated with incomplete extraction of microbial DNA from soil." *Appl Environ Microbiol* **75**(16): 5428-5433.
- Felczykowska, A., A. Dydecka, M. G. Bohdanowicz, G. S. T, M. Sobo, J. Kobos, S. Bloch, B. E. Nejman-Fale Czyk and W. G. G (2014). "The use of fosmid metagenomic libraries in preliminary screening for various biological activities." *Microb Cell Fact* **13**(1): 105.
- Ferreira, A. J., R. Siam, J. C. Setubal, A. Moustafa, A. Sayed, F. S. Chambergo, A. S. Dawe, M. A. Ghazy, H. Sharaf, A. Ouf, I. Alam, A. M. Abdel-Haleem, H. Lehtvaslainen, E. Ramadan, A. Antunes, U. Stingl, J. A. Archer, B. R. Jankovic, M. Sogin, V. B. Bajic and H. El-Dorri (2014). "Core microbial functional activities in ocean environments revealed by global metagenomic profiling analyses." *PLoS One* **9**(6): e97338.
- 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**(1-2): 109-123.
- Ferrer, M., A. Beloqui, J. M. Vieites, M. E. Guazzaroni, I. Berger and A. Aharoni (2009). "Interplay of metagenomics and in vitro compartmentalization." *Microb Biotechnol* **2**(1): 31-39.
- Fieseler, L., U. Hentschel, L. Grozdanov, A. Schirmer, G. Wen, M. Platzer, S. Hrvatin, D. Butzke, K. Zimmermann and J. Piel (2007). "Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges." *Appl Environ Microbiol* **73**(7): 2144-2155.
- Fitch, W. M. and E. Margoliash (1967). "Construction of phylogenetic trees." *Science* **155**(3760): 279-284.
- Fox, G. E., K. R. Pechman and C. R. Woese (1977). "Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Prokaryotic Systematics." *International Journal of Systematic Bacteriology* **27**(1): 44-57.
- Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsens, K. N. Chen and C. R. Woese (1980). "The phylogeny of prokaryotes." *Science* **209**(4455): 457-463.
- Frebourg, T. and O. Brison (1988). "Plasmid vectors with multiple cloning sites and cat-reporter gene for promoter cloning and analysis in animal cells." *Gene* **65**(2): 315-318.

- Freeman, M. F., C. Gurgui, M. J. Helf, B. I. Morinaka, A. R. Uria, N. J. Oldham, H. G. Sahl, S. Matsunaga and J. Piel (2012). "Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides." Science **338**(6105): 387-390.
- Fuhrman, J. A., M. S. Schwalbach and U. Stingl (2008). "Proteorhodopsins: an array of physiological roles?" Nat Rev Microbiol **6**(6): 488-494.
- Gabor, E., K. Liebeton, F. Niehaus, J. Eck and P. Lorenz (2007). "Updating the metagenomics toolbox." Biotechnol J **2**(2): 201-206.
- 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**(9): 879-886.
- Gabor, E. M., E. J. de Vries and D. B. Janssen (2003). "Efficient recovery of environmental DNA for expression cloning by indirect extraction methods." FEMS Microbiol Ecol **44**(2): 153-163.
- Gaida, S. M., N. R. Sandoval, S. A. Nicolaou, Y. Chen, K. P. Venkataramanan and E. T. Papoutsakis (2015). "Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries." Nat Commun **6**: 7045.
- Galbraith, E. A., D. A. Antonopoulos and B. A. White (2004). "Suppressive subtractive hybridization as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model." Environ Microbiol **6**(9): 928-937.
- Gershenson, J. and N. Dudareva (2007). "The function of terpene natural products in the natural world." Nat Chem Biol **3**(7): 408-414.
- Gilbert, J. A. and C. L. Dupont (2011). "Microbial metagenomics: beyond the genome." Ann Rev Mar Sci **3**: 347-371.
- Giovannoni, S. and U. Stingl (2007). "The importance of culturing bacterioplankton in the 'omics' age." Nat Rev Microbiol **5**(10): 820-826.
- Giovannoni, S. J., T. B. Britschgi, C. L. Moyer and K. G. Field (1990). "Genetic diversity in Sargasso Sea bacterioplankton." Nature **345**(6270): 60-63.
- Giovannoni, S. J. and U. Stingl (2005). "Molecular diversity and ecology of microbial plankton." Nature **437**(7057): 343-348.
- Goddard, J. P. and J. L. Reymond (2004). "Enzyme assays for high-throughput screening." Curr Opin Biotechnol **15**(4): 314-322.
- Green, C. D., J. F. Simons, B. E. Taillon and D. A. Lewin (2001). "Open systems: panoramic views of gene expression." J Immunol Methods **250**(1-2): 67-79.
- Griffiths, A. D. and D. S. Tawfik (2003). "Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization." EMBO J **22**(1): 24-35.
- Griffiths, A. D. and D. S. Tawfik (2006). "Miniaturising the laboratory in emulsion droplets." Trends Biotechnol **24**(9): 395-402.
- Grotzinger, S. W., I. Alam, W. Ba Alawi, V. B. Bajic, U. Stingl and J. Eppinger (2014). "Mining a database of single amplified genomes from Red Sea brine pool extremophiles-improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA)." Front Microbiol **5**: 134.
- Guan, C., W. Cui, X. He, X. Hu, J. Xu, G. Du, J. Chen and Z. Zhou (2015). "Construction and development of a novel expression system of Streptomyces." Protein Expr Purif **113**: 17-22.

- Guo, X., H. Yin, J. Cong, Z. Dai, Y. Liang and X. Liu (2013). "RubisCO gene clusters found in a metagenome microarray from acid mine drainage." Appl Environ Microbiol **79**(6): 2019-2026.
- Haeckel, E. (1866). Generelle Morphologie der Organismen: allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie. Allgemeine Anatomie der Organismen, Reimer.
- Handelsman, J. (2004). "Metagenomics: application of genomics to uncultured microorganisms." Microbiol Mol Biol Rev **68**(4): 669-685.
- 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**(10): R245-249.
- Handyside, A. H., M. D. Robinson, R. J. Simpson, M. B. Omar, M. A. Shaw, J. G. Grudzinskas and A. Rutherford (2004). "Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease." Mol Hum Reprod **10**(10): 767-772.
- Harris, T. D., P. R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J. W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S. R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss and Z. Xie (2008). "Single-molecule DNA sequencing of a viral genome." Science **320**(5872): 106-109.
- Healy, F. G., R. M. Ray, H. C. Aldrich, A. C. Wilkie, L. O. Ingram and K. T. Shanmugam (1995). "Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose." Appl Microbiol Biotechnol **43**(4): 667-674.
- Hentschel, U., J. Piel, S. M. Degnan and M. W. Taylor (2012). "Genomic insights into the marine sponge microbiome." Nat Rev Microbiol **10**(9): 641-654.
- Hess, M., A. Sczyrba, R. Egan, T. W. Kim, H. Chokhawala, G. Schroth, S. Luo, D. S. Clark, F. Chen, T. Zhang, R. I. Mackie, L. A. Pennacchio, S. G. Tringe, A. Visel, T. Woyke, Z. Wang and E. M. Rubin (2011). "Metagenomic discovery of biomass-degrading genes and genomes from cow rumen." Science **331**(6016): 463-467.
- Hicks, M. A. and K. L. Prather (2014). "Bioprospecting in the genomic age." Adv Appl Microbiol **87**: 111-146.
- Hochmuth, T., H. Niederkruger, C. Gernert, A. Siegl, S. Taudien, M. Platzer, P. Crews, U. Hentschel and J. Piel (2010). "Linking chemical and microbial diversity in marine sponges: possible role for poribacteria as producers of methyl-branched fatty acids." Chembiochem **11**(18): 2572-2578.
- Hochmuth, T. and J. Piel (2009). "Polyketide synthases of bacterial symbionts in sponges--evolution-based applications in natural products research." Phytochemistry **70**(15-16): 1841-1849.
- Hong, K. S., H. K. Lim, E. J. Chung, E. J. Park, M. H. Lee, J. C. Kim, G. J. Choi, K. Y. Cho and S. W. Lee (2007). "Selection and characterization of forest soil metagenome genes encoding lipolytic enzymes." J Microbiol Biotechnol **17**(10): 1655-1660.
- Hopwood, D. A. (2007). "Therapeutic treasures from the deep." Nat Chem Biol **3**(8): 457-458.

- Hosono, S., A. F. Faruqi, F. B. Dean, Y. Du, Z. Sun, X. Wu, J. Du, S. F. Kingsmore, M. Egholm and R. S. Lasken (2003). "Unbiased whole-genome amplification directly from clinical samples." *Genome Res* **13**(5): 954-964.
- Hu, Y., J. Chen, G. Hu, J. Yu, X. Zhu, Y. Lin, S. Chen and J. Yuan (2015). "Statistical Research on the Bioactivity of New Marine Natural Products Discovered during the 28 Years from 1985 to 2012." *Mar Drugs* **13**(1): 202-221.
- Huang, W. E., Y. Song and J. Xu (2015). "Single cell biotechnology to shed a light on biological 'dark matter' in nature." *Microb Biotechnol* **8**(1): 15-16.
- Hunter, P. (2008). "Harnessing Nature's wisdom. Turning to Nature for inspiration and avoiding her follies." *EMBO Rep* **9**(9): 838-840.
- Hwang, S., K. H. Choi, N. Yoon and J. Cha (2015). "Improvement of a Sulfolobus-E. coli shuttle vector for heterologous gene expression in Sulfolobus acidocaldarius." *J Microbiol Biotechnol* **25**(2): 196-205.
- Ilmberger, N., S. Gullert, J. Dannenberg, U. Rabausch, J. Torres, B. Wemheuer, M. Alawi, A. Poehlein, J. Chow, D. Turaev, T. Rattei, C. Schmeisser, J. Salomon, P. B. Olsen, R. Daniel, A. Grundhoff, M. S. Borchert and W. R. Streit (2014). "A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes." *PLoS One* **9**(9): e106707.
- Inceoglu, O., E. F. Hoogwout, P. Hill and J. D. van Elsas (2010). "Effect of DNA extraction method on the apparent microbial diversity of soil." *Appl Environ Microbiol* **76**(10): 3378-3382.
- Iqbal, H. A., Z. Feng and S. F. Brady (2012). "Biocatalysts and small molecule products from metagenomic studies." *Curr Opin Chem Biol* **16**(1-2): 109-116.
- Jabbour, D., A. Sorger, K. Sahm and G. Antranikian (2013). "A highly thermoactive and salt-tolerant alpha-amylase isolated from a pilot-plant biogas reactor." *Appl Microbiol Biotechnol* **97**(7): 2971-2978.
- Jacquioud, S., S. Demaneche, L. Franqueville, L. Ausec, Z. Xu, T. O. Delmont, V. Dunon, C. Cagnon, I. Mandic-Mulec, T. M. Vogel and P. Simonet (2014). "Characterization of new bacterial catabolic genes and mobile genetic elements by high throughput genetic screening of a soil metagenomic library." *J Biotechnol* **190**: 18-29.
- Jansson, J. K. and N. Tas (2014). "The microbial ecology of permafrost." *Nat Rev Microbiol* **12**(6): 414-425.
- Jimenez-Infante, F., D. K. Ngugi, I. Alam, M. Rashid, W. Baalawi, A. A. Kamau, V. B. Bajic and U. Stingl (2014). "Genomic differentiation among two strains of the PS1 clade isolated from geographically separated marine habitats." *FEMS Microbiol Ecol* **89**(1): 181-197.
- Kalyuzhnaya, M. G., A. Lapidus, N. Ivanova, A. C. Copeland, A. C. McHardy, E. Szeto, A. Salamov, I. V. Grigoriev, D. Suci, S. R. Levine, V. M. Markowitz, I. Rigoutsos, S. G. Tringe, D. C. Bruce, P. M. Richardson, M. E. Lidstrom and L. Chistoserdova (2008). "High-resolution metagenomics targets specific functional types in complex microbial communities." *Nat Biotechnol* **26**(9): 1029-1034.
- Kamanda Ngugi, D., J. Blom, I. Alam, M. Rashid, W. Ba-Alawi, G. Zhang, T. Hikmawan, Y. Guan, A. Antunes, R. Siam, H. El Dorry, V. Bajic and U. Stingl (2015). "Comparative genomics reveals adaptations of a halotolerant thaumarchaeon in the interfaces of brine pools in the Red Sea." *ISME J* **9**(2): 396-411.

- Kasting, J. F. and J. L. Siefert (2002). "Life and the evolution of Earth's atmosphere." *Science* **296**(5570): 1066-1068.
- Kennedy, J., J. R. Marchesi and A. D. Dobson (2007). "Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges." *Appl Microbiol Biotechnol* **75**(1): 11-20.
- Kim, T. K. and J. A. Fuerst (2006). "Diversity of polyketide synthase genes from bacteria associated with the marine sponge *Pseudoceratina clavata*: culture-dependent and culture-independent approaches." *Environ Microbiol* **8**(8): 1460-1470.
- Kim, U. J., H. Shizuya, P. J. de Jong, B. Birren and M. I. Simon (1992). "Stable propagation of cosmid sized human DNA inserts in an F factor based vector." *Nucleic Acids Res* **20**(5): 1083-1085.
- Knietsch, A., S. Bowien, G. Whited, G. Gottschalk and R. Daniel (2003). "Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures." *Appl Environ Microbiol* **69**(6): 3048-3060.
- Korlach, J., K. P. Bjornson, B. P. Chaudhuri, R. L. Cicero, B. A. Flusberg, J. J. Gray, D. Holden, R. Saxena, J. Wegener and S. W. Turner (2010). "Real-time DNA sequencing from single polymerase molecules." *Methods Enzymol* **472**: 431-455.
- Kumar, G., R. Rech, K. Kapolka, K. Lavrenov, E. Garnova, S. Lavasini, R. Deadman and S. Hamilton (2007). "Genomic DNA preparation using illustra GenomiPhi V2 and HY DNA Amplification kits." *Nature Methods*. pp An30-An32.
- Kumar, R., M. Sharma, R. Singh and J. Kaur (2013). "Characterization and evolution of a metagenome-derived lipase towards enhanced enzyme activity and thermostability." *Mol Cell Biochem* **373**(1-2): 149-159.
- Kvist, T., B. K. Ahring, R. S. Lasken and P. Westermann (2007). "Specific single-cell isolation and genomic amplification of uncultured microorganisms." *Appl Microbiol Biotechnol* **74**(4): 926-935.
- Kyrpides, N. C., P. Hugenholtz, J. A. Eisen, T. Woyke, M. Goker, C. T. Parker, R. Amann, B. J. Beck, P. S. Chain, J. Chun, R. R. Colwell, A. Danchin, P. Dawyndt, T. Dedeurwaerdere, E. F. DeLong, J. C. Detter, P. De Vos, T. J. Donohue, X. Z. Dong, D. S. Ehrlich, C. Fraser, R. Gibbs, J. Gilbert, P. Gilna, F. O. Glockner, J. K. Jansson, J. D. Keasling, R. Knight, D. Labeda, A. Lapidus, J. S. Lee, W. J. Li, J. Ma, V. Markowitz, E. R. Moore, M. Morrison, F. Meyer, K. E. Nelson, M. Ohkuma, C. A. Ouzounis, N. Pace, J. Parkhill, N. Qin, R. Rossello-Mora, J. Sikorski, D. Smith, M. Sogin, R. Stevens, U. Stingl, K. Suzuki, D. Taylor, J. M. Tiedje, B. Tindall, M. Wagner, G. Weinstock, J. Weissenbach, O. White, J. Wang, L. Zhang, Y. G. Zhou, D. Field, W. B. Whitman, G. M. Garrity and H. P. Klenk (2014). "Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains." *PLoS Biol* **12**(8): e1001920.
- Lage, J. M., J. H. Leamon, T. Pejovic, S. Hamann, M. Lacey, D. Dillon, R. Segraves, B. Vossbrinck, A. Gonzalez, D. Pinkel, D. G. Albertson, J. Costa and P. M. Lizardi (2003). "Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH." *Genome Res* **13**(2): 294-307.

- Lammle, K., H. Zipper, M. Breuer, B. Hauer, C. Buta, H. Brunner and S. Rupp (2007). "Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning." *J Biotechnol* **127**(4): 575-592.
- Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin and N. R. Pace (1985). "Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses." *Proc Natl Acad Sci U S A* **82**(20): 6955-6959.
- Lasken, R. S. (2012). "Genomic sequencing of uncultured microorganisms from single cells." *Nat Rev Microbiol* **10**(9): 631-640.
- Lasken, R. S. and M. Egholm (2003). "Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens." *Trends Biotechnol* **21**(12): 531-535.
- Lasken, R. S. and J. S. McLean (2014). "Recent advances in genomic DNA sequencing of microbial species from single cells." *Nat Rev Genet* **15**(9): 577-584.
- Lasken, R. S. and T. B. Stockwell (2007). "Mechanism of chimera formation during the Multiple Displacement Amplification reaction." *BMC Biotechnol* **7**: 19.
- Lawley, B. and G. W. Tannock (2012). "Nucleic acid-based methods to assess the composition and function of the bowel microbiota." *Gastroenterol Clin North Am* **41**(4): 855-868.
- Lee, M. H. and S. W. Lee (2013). "Bioprospecting potential of the soil metagenome: novel enzymes and bioactivities." *Genomics Inform* **11**(3): 114-120.
- Leis, B., A. Angelov and W. Liebl (2013). "Screening and expression of genes from metagenomes." *Adv Appl Microbiol* **83**: 1-68.
- Lewin, A., A. Wentzel and S. Valla (2013). "Metagenomics of microbial life in extreme temperature environments." *Curr Opin Biotechnol* **24**(3): 516-525.
- Liebl, W., A. Angelov, J. Juergensen, J. Chow, A. Loeschcke, T. Drepper, T. Classen, J. Pietruszka, A. Ehrenreich, W. R. Streit and K. E. Jaeger (2014). "Alternative hosts for functional (meta)genome analysis." *Appl Microbiol Biotechnol* **98**(19): 8099-8109.
- Liles, M. R., B. F. Manske, S. B. Bintrim, J. Handelsman and R. M. Goodman (2003). "A census of rRNA genes and linked genomic sequences within a soil metagenomic library." *Appl Environ Microbiol* **69**(5): 2684-2691.
- Liles, M. R., L. L. Williamson, J. Rodbumrer, V. Torsvik, L. C. Parsley, R. M. Goodman and J. Handelsman (2009). "Isolation and cloning of high-molecular-weight metagenomic DNA from soil microorganisms." *Cold Spring Harb Protoc* **2009**(8): pdb prot5271.
- Ling, L. L., T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, T. F. Schaberle, D. E. Hughes, S. Epstein, M. Jones, L. Lazarides, V. A. Steadman, D. R. Cohen, C. R. Felix, K. A. Fetterman, W. P. Millett, A. G. Nitti, A. M. Zullo, C. Chen and K. Lewis (2015). "A new antibiotic kills pathogens without detectable resistance." *Nature* **517**(7535): 455-459.
- Lopanik, N. B., J. A. Shields, T. J. Buchholz, C. M. Rath, J. Hothersall, M. G. Haygood, K. Hakansson, C. M. Thomas and D. H. Sherman (2008). "In vivo and in vitro transacylation by BryP, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont." *Chem Biol* **15**(11): 1175-1186.
- Lopez-Lopez, O., M. E. Cerdan and M. I. Gonzalez-Siso (2013). "Hot spring metagenomics." *Life (Basel)* **3**(2): 308-320.

- Lorenz, P. and J. Eck (2005). "Metagenomics and industrial applications." *Nat Rev Microbiol* **3**(6): 510-516.
- Lorenz, P., K. Liebeton, F. Niehaus and J. Eck (2002). "Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space." *Curr Opin Biotechnol* **13**(6): 572-577.
- Marcy, Y., T. Ishoey, R. S. Lasken, T. B. Stockwell, B. P. Walenz, A. L. Halpern, K. Y. Beeson, S. M. Goldberg and S. R. Quake (2007). "Nanoliter reactors improve multiple displacement amplification of genomes from single cells." *PLoS Genet* **3**(9): 1702-1708.
- Marcy, Y., C. Ouverney, E. M. Bik, T. Losekann, N. Ivanova, H. G. Martin, E. Szeto, D. Platt, P. Hugenholtz, D. A. Relman and S. R. Quake (2007). "Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth." *Proc Natl Acad Sci U S A* **104**(29): 11889-11894.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley and J. M. Rothberg (2005). "Genome sequencing in microfabricated high-density picolitre reactors." *Nature* **437**(7057): 376-380.
- Marris, E. (2006). "Marine natural products: drugs from the deep." *Nature* **443**(7114): 904-905.
- Marshall, I. P., P. C. Blainey, A. M. Spormann and S. R. Quake (2012). "A Single-cell genome for *Thiovulum* sp." *Appl Environ Microbiol* **78**(24): 8555-8563.
- Martinez, A., S. J. Kolvek, C. L. Yip, J. Hopke, K. A. Brown, I. A. MacNeil and M. S. Osburne (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." *Appl Environ Microbiol* **70**(4): 2452-2463.
- Martinez-Garcia, M., D. M. Brazel, B. K. Swan, C. Arnosti, P. S. Chain, K. G. Reitenga, G. Xie, N. J. Poulton, M. Lluesma Gomez, D. E. Masland, B. Thompson, W. K. Bellows, K. Ziervogel, C. C. Lo, S. Ahmed, C. D. Gleasner, C. J. Detter and R. Stepanauskas (2012). "Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of Verrucomicrobia." *PLoS One* **7**(4): e35314.
- Martinez-Martinez, M., I. Lores, C. Pena-Garcia, R. Bargiela, D. Reyes-Duarte, M. E. Guazzaroni, A. I. Pelaez, J. Sanchez and M. Ferrer (2014). "Biochemical studies on a versatile esterase that is most catalytically active with polyaromatic esters." *Microb Biotechnol* **7**(2): 184-191.
- Maurer, K. H. (2004). "Detergent proteases." *Curr Opin Biotechnol* **15**(4): 330-334.
- McLean, J. S., M. J. Lombardo, J. H. Badger, A. Edlund, M. Novotny, J. Yee-Greenbaum, N. Vyahhi, A. P. Hall, Y. Yang, C. L. Dupont, M. G. Ziegler, H. Chitsaz, A. E. Allen, S. Yooseph, G. Tesler, P. A. Pevzner, R. M. Friedman, K. H. Nealson, J. C. Venter and R. S. Lasken (2013). "Candidate phylum TM6 genome recovered from a hospital sink

- biofilm provides genomic insights into this uncultivated phylum." Proc Natl Acad Sci U S A **110**(26): E2390-2399.
- McMahon, M. D., C. Guan, J. Handelsman and M. G. Thomas (2012). "Metagenomic analysis of *Streptomyces lividans* reveals host-dependent functional expression." Appl Environ Microbiol **78**(10): 3622-3629.
- McNally, B., A. Singer, Z. Yu, Y. Sun, Z. Weng and A. Meller (2010). "Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays." Nano Lett **10**(6): 2237-2244.
- Miao, V., M. F. Coeffet-LeGal, D. Brown, S. Sinnemann, G. Donaldson and J. Davies (2001). "Genetic approaches to harvesting lichen products." Trends Biotechnol **19**(9): 349-355.
- Miyake, S. and U. Stengl (2011). *Proteorhodopsin*. eLS, John Wiley & Sons, Ltd.
- Montaser, R. and H. Luesch (2011). "Marine natural products: a new wave of drugs?" Future Med Chem **3**(12): 1475-1489.
- Moore, R. L. and B. J. McCarthy (1967). "Comparative study of ribosomal ribonucleic acid cistrons in enterobacteria and myxobacteria." J Bacteriol **94**(4): 1066-1074.
- Moser, M. J., R. A. DiFrancesco, K. Gowda, A. J. Klingele, D. R. Sugar, S. Stocki, D. A. Mead and T. W. Schoenfeld (2012). "Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme." PLoS One **7**(6): e38371.
- Mullis, K. B. and F. A. Faloona (1987). "Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction." Methods Enzymol **155**: 335-350.
- Nacke, H., M. Engelhaupt, S. Brady, C. Fischer, J. Tautzt and R. Daniel (2012). "Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes." Biotechnol Lett **34**(4): 663-675.
- Newman, D. J., G. M. Cragg and K. M. Snader (2000). "The influence of natural products upon drug discovery." Natural Product Reports **17**(3): 215-234.
- Newman, D. J., G. M. Cragg and K. M. Snader (2003). "Natural products as sources of new drugs over the period 1981-2002." J Nat Prod **66**(7): 1022-1037.
- Nichols, D., N. Cahoon, E. M. Trakhtenberg, L. Pham, A. Mehta, A. Belanger, T. Kanigan, K. Lewis and S. S. Epstein (2010). "Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species." Appl Environ Microbiol **76**(8): 2445-2450.
- Noro, J. C., J. A. Kalaitzis and B. A. Neilan (2012). "Bioactive natural products from Papua New Guinea marine sponges." Chem Biodivers **9**(10): 2077-2095.
- Nyyssonen, M., H. M. Tran, U. Karaoz, C. Weihe, M. Z. Hadi, J. B. Martiny, A. C. Martiny and E. L. Brodie (2013). "Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries." Front Microbiol **4**: 282.
- O'Hanlon, L. H. (2006). "Scientists are searching the seas for cancer drugs." J Natl Cancer Inst **98**(10): 662-663.
- Ogram, A., G. S. Sayler and T. Barkay (1987). "The extraction and purification of microbial DNA from sediments." Journal of Microbiological Methods **7**(2-3): 57-66.
- Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace and D. A. Stahl (1986). "Microbial ecology and evolution: a ribosomal RNA approach." Annu Rev Microbiol **40**: 337-365.

- Orellana, L. H., R. L. Rodriguez, S. Higgins, J. C. Chee-Sanford, R. A. Sanford, K. M. Ritalahti, F. E. Löffler and K. T. Konstantinidis (2014). "Detecting nitrous oxide reductase (NosZ) genes in soil metagenomes: method development and implications for the nitrogen cycle." *MBio* **5**(3): e01193-01114.
- Osborn, A. M. and C. J. Smith (2005). *Molecular microbial ecology*. New York ; Abingdon England, Taylor & Francis.
- Pace, B. and L. L. Campbell (1971). "Homology of ribosomal ribonucleic acid diverse bacterial species with *Escherichia coli* and *Bacillus stearothermophilus*." *J Bacteriol* **107**(2): 543-547.
- Pace, N., D. Stahl, D. Lane and G. Olsen (1986). The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences. *Advances in Microbial Ecology*. K. C. Marshall, Springer US. **9**: 1-55.
- Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." *Science* **276**(5313): 734-740.
- Pace, N. R., G. J. Olsen and C. R. Woese (1986). "Ribosomal RNA phylogeny and the primary lines of evolutionary descent." *Cell* **45**(3): 325-326.
- Park, S. J., C. H. Kang, J. C. Chae and S. K. Rhee (2008). "Metagenome microarray for screening of fosmid clones containing specific genes." *FEMS Microbiol Lett* **284**(1): 28-34.
- Park, S. J., D. H. Kim, M. Y. Jung, S. J. Kim, H. Kim, Y. H. Kim, J. C. Chae and S. K. Rhee (2012). "Evaluation of a fosmid-clone-based microarray for comparative analysis of swine fecal metagenomes." *J Microbiol* **50**(4): 684-688.
- Patel, D. D., A. K. Patel, N. R. Parmar, T. M. Shah, J. B. Patel, P. R. Pandya and C. G. Joshi (2014). "Microbial and Carbohydrate Active Enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet." *Gene* **545**(1): 88-94.
- Piel, J. (2002). "A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles." *Proc Natl Acad Sci U S A* **99**(22): 14002-14007.
- Piel, J. (2004). "Metabolites from symbiotic bacteria." *Nat Prod Rep* **21**(4): 519-538.
- Piel, J. (2006). "Bacterial symbionts: prospects for the sustainable production of invertebrate-derived pharmaceuticals." *Curr Med Chem* **13**(1): 39-50.
- Piel, J. (2009). "Metabolites from symbiotic bacteria." *Nat Prod Rep* **26**(3): 338-362.
- Piel, J. (2011). "Approaches to capturing and designing biologically active small molecules produced by uncultured microbes." *Annu Rev Microbiol* **65**: 431-453.
- Piel, J., D. Butzke, N. Fusetani, D. Hui, M. Platzer, G. Wen and S. Matsunaga (2005). "Exploring the chemistry of uncultivated bacterial symbionts: antitumor polyketides of the pederin family." *J Nat Prod* **68**(3): 472-479.
- Piel, J., I. Hofer and D. Hui (2004). "Evidence for a symbiosis island involved in horizontal acquisition of pederin biosynthetic capabilities by the bacterial symbiont of *Paederus fuscipes* beetles." *J Bacteriol* **186**(5): 1280-1286.
- Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani and S. Matsunaga (2004). "Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*." *Proc Natl Acad Sci U S A* **101**(46): 16222-16227.

- Pinard, R., A. de Winter, G. J. Sarkis, M. B. Gerstein, K. R. Tartaro, R. N. Plant, M. Egholm, J. M. Rothberg and J. H. Leamon (2006). "Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing." BMC Genomics **7**: 216.
- Podar, M., C. B. Abulencia, M. Walcher, D. Hutchison, K. Zengler, J. A. Garcia, T. Holland, D. Cotton, L. Hauser and M. Keller (2007). "Targeted access to the genomes of low-abundance organisms in complex microbial communities." Appl Environ Microbiol **73**(10): 3205-3214.
- Pope, P. B., S. E. Denman, M. Jones, S. G. Tringe, K. Barry, S. A. Malfatti, A. C. McHardy, J. F. Cheng, P. Hugenholtz, C. S. McSweeney and M. Morrison (2010). "Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores." Proc Natl Acad Sci U S A **107**(33): 14793-14798.
- Procopio, R. E., I. R. Silva, M. K. Martins, J. L. Azevedo and J. M. Araujo (2012). "Antibiotics produced by Streptomyces." Braz J Infect Dis **16**(5): 466-471.
- Putz, A. and P. Proksch (2010). Chemical Defence in Marine Ecosystems. Annual Plant Reviews Volume 39: Functions and Biotechnology of Plant Secondary Metabolites, Wiley-Blackwell: 162-213.
- Quaiser, A., Y. Zivanovic, D. Moreira and P. Lopez-Garcia (2011). "Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara." ISME J **5**(2): 285-304.
- Rabausch, U., N. Ilmberger and W. R. Streit (2014). "The metagenome-derived enzyme RhaB opens a new subclass of bacterial B type alpha-l-rhamnosidases." J Biotechnol.
- Rabausch, U., J. Juergensen, N. Ilmberger, S. Bohnke, S. Fischer, B. Schubach, M. Schulte and W. R. Streit (2013). "Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes." Appl Environ Microbiol **79**(15): 4551-4563.
- Radajewski, S., I. R. McDonald and J. C. Murrell (2003). "Stable-isotope probing of nucleic acids: a window to the function of uncultured microorganisms." Curr Opin Biotechnol **14**(3): 296-302.
- Raghunathan, A., H. R. Ferguson, Jr., C. J. Bornarth, W. Song, M. Driscoll and R. S. Lasken (2005). "Genomic DNA amplification from a single bacterium." Appl Environ Microbiol **71**(6): 3342-3347.
- Rappe, M. S. and S. J. Giovannoni (2003). "The uncultured microbial majority." Annu Rev Microbiol **57**: 369-394.
- Rashamuse, K. J., D. F. Visser, F. Hennessy, J. Kemp, M. P. Roux-van der Merwe, J. Badenhorst, T. Ronneburg, R. Francis-Pope and D. Brady (2013). "Characterisation of two bifunctional cellulase-xylanase enzymes isolated from a bovine rumen metagenome library." Curr Microbiol **66**(2): 145-151.
- Rath, C. M., B. Janto, J. Earl, A. Ahmed, F. Z. Hu, L. Hiller, M. Dahlgren, R. Kreft, F. Yu, J. J. Wolff, H. K. Kweon, M. A. Christiansen, K. Hakansson, R. M. Williams, G. D. Ehrlich and D. H. Sherman (2011). "Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743." ACS Chem Biol **6**(11): 1244-1256.

- Reddy, T. B., A. D. Thomas, D. Stamatis, J. Bertsch, M. Isbandi, J. Jansson, J. Mallajosyula, I. Pagani, E. A. Lobos and N. C. Kyrpides (2015). "The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification." *Nucleic Acids Res* **43**(Database issue): D1099-1106.
- Reid, A. (2014). "From Woese to Wired: the unexpected payoffs of basic research." *RNA Biol* **11**(3): 205-206.
- Reyes-Duarte, D., M. Ferrer and H. Garcia-Arellano (2012). "Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries." *Methods Mol Biol* **861**: 101-113.
- Ridley, C. P., D. John Faulkner and M. G. Haygood (2005). "Investigation of *Oscillatoria spongelliae*-dominated bacterial communities in four dictyoceratid sponges." *Appl Environ Microbiol* **71**(11): 7366-7375.
- Rinke, C., J. Lee, N. Nath, D. Goudeau, B. Thompson, N. Poulton, E. Dmitrieff, R. Malmstrom, R. Stepanauskas and T. Woyke (2014). "Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics." *Nat Protoc* **9**(5): 1038-1048.
- Rinke, C., P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J. F. Cheng, A. Darling, S. Malfatti, B. K. Swan, E. A. Gies, J. A. Dodsworth, B. P. Hedlund, G. Tsiamis, S. M. Sievert, W. T. Liu, J. A. Eisen, S. J. Hallam, N. C. Kyrpides, R. Stepanauskas, E. M. Rubin, P. Hugenholtz and T. Woyke (2013). "Insights into the phylogeny and coding potential of microbial dark matter." *Nature* **499**(7459): 431-437.
- Rittmann, S. and C. Herwig (2012). "A comprehensive and quantitative review of dark fermentative biohydrogen production." *Microb Cell Fact* **11**: 115.
- Rodrigue, S., R. R. Malmstrom, A. M. Berlin, B. W. Birren, M. R. Henn and S. W. Chisholm (2009). "Whole genome amplification and de novo assembly of single bacterial cells." *PLoS One* **4**(9): e6864.
- Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman and R. M. Goodman (2000). "Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms." *Appl Environ Microbiol* **66**(6): 2541-2547.
- Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcon, V. Souza, G. Bonilla-Rosso, L. E. Eguarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Neilson, R. Friedman, M. Frazier and J. C. Venter (2007). "The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific." *PLoS Biol* **5**(3): e77.
- Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich (1988). "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." *Science* **239**(4839): 487-491.
- Salipante, S. J., D. J. Roach, J. O. Kitzman, M. W. Snyder, B. Stackhouse, S. M. Butler-Wu, C. Lee, B. T. Cookson and J. Shendure (2015). "Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains." *Genome Res* **25**(1): 119-128.

- Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.
- Schloss, P. D. and J. Handelsman (2003). "Biotechnological prospects from metagenomics." Curr Opin Biotechnol **14**(3): 303-310.
- Schmidt, E. W. and M. S. Donia (2010). "Life in cellulose houses: symbiotic bacterial biosynthesis of ascidian drugs and drug leads." Curr Opin Biotechnol **21**(6): 827-833.
- Schmidt, E. W., M. S. Donia, J. A. McIntosh, W. F. Fricke and J. Ravel (2012). "Origin and variation of tunicate secondary metabolites." J Nat Prod **75**(2): 295-304.
- 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**(14): 4371-4378.
- Schmidt, T. M. and D. A. Relman (1994). "Phylogenetic identification of uncultured pathogens using ribosomal RNA sequences." Methods Enzymol **235**: 205-222.
- Schmieder, R. and R. Edwards (2011). "Fast identification and removal of sequence contamination from genomic and metagenomic datasets." PLoS One **6**(3): e17288.
- Schroder, C., S. Elleuche, S. Blank and G. Antranikian (2014). "Characterization of a heat-active archaeal beta-glucosidase from a hydrothermal spring metagenome." Enzyme Microb Technol **57**: 48-54.
- Sebat, J. L., F. S. Colwell and R. L. Crawford (2003). "Metagenomic profiling: microarray analysis of an environmental genomic library." Appl Environ Microbiol **69**(8): 4927-4934.
- Shapiro, E., T. Biezuner and S. Linnarsson (2013). "Single-cell sequencing-based technologies will revolutionize whole-organism science." Nat Rev Genet **14**(9): 618-630.
- Sharpton, T. J. (2014). "An introduction to the analysis of shotgun metagenomic data." Front Plant Sci **5**: 209.
- Shendure, J., G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra and G. M. Church (2005). "Accurate multiplex polony sequencing of an evolved bacterial genome." Science **309**(5741): 1728-1732.
- Shi, W., S. Xie, X. Chen, S. Sun, X. Zhou, L. Liu, P. Gao, N. C. Kyrpides, E. G. No and J. S. Yuan (2013). "Comparative genomic analysis of the microbiome [corrected] of herbivorous insects reveals eco-environmental adaptations: biotechnology applications." PLoS Genet **9**(1): e1003131.
- Shizuya, H., B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri and M. Simon (1992). "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector." Proc Natl Acad Sci U S A **89**(18): 8794-8797.
- Siegl, A., J. Kamke, T. Hochmuth, J. Piel, M. Richter, C. Liang, T. Dandekar and U. Hentschel (2011). "Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges." ISME J **5**(1): 61-70.
- Simon, C. and R. Daniel (2009). "Achievements and new knowledge unraveled by metagenomic approaches." Appl Microbiol Biotechnol **85**(2): 265-276.
- Simon, C. and R. Daniel (2011). "Metagenomic analyses: past and future trends." Appl Environ Microbiol **77**(4): 1153-1161.

- Sims, D., I. Sudbery, N. E. Iltott, A. Heger and C. P. Ponting (2014). "Sequencing depth and coverage: key considerations in genomic analyses." Nat Rev Genet **15**(2): 121-132.
- Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent and L. E. Hood (1986). "Fluorescence detection in automated DNA sequence analysis." Nature **321**(6071): 674-679.
- Sogin, S. J., M. L. Sogin and C. R. Woese (1972). "Phylogenetic measurement in procaryotes by primary structural characterization." J Mol Evol **1**(2): 173-184.
- Sorek, R., Y. Zhu, C. J. Creevey, M. P. Francino, P. Bork and E. M. Rubin (2007). "Genome-wide experimental determination of barriers to horizontal gene transfer." Science **318**(5855): 1449-1452.
- Staley, J. T. and A. Konopka (1985). "Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats." Annu Rev Microbiol **39**: 321-346.
- 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**(1-2): 25-37.
- Steele, H. L. and W. R. Streit (2005). "Metagenomics: advances in ecology and biotechnology." FEMS Microbiol Lett **247**(2): 105-111.
- 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**(3): 591-599.
- Stevens, D. C., K. R. Conway, N. Pearce, L. R. Villegas-Penaranda, A. G. Garza and C. N. Boddy (2013). "Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in Escherichia coli." PLoS One **8**(5): e64858.
- Steward, G. F. and M. S. Rappé (2007). "What's the 'meta' with metagenomics?" ISME JOURNAL **1**(2): 100.
- Stokke, R., I. Roalkvam, A. Lanzen, H. Haflidason and I. H. Steen (2012). "Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments." Environ Microbiol **14**(5): 1333-1346.
- Su, C., L. Lei, Y. Duan, K. Q. Zhang and J. Yang (2012). "Culture-independent methods for studying environmental microorganisms: methods, application, and perspective." Appl Microbiol Biotechnol **93**(3): 993-1003.
- Suen, G., J. J. Scott, F. O. Aylward, S. M. Adams, S. G. Tringe, A. A. Pinto-Tomas, C. E. Foster, M. Pauly, P. J. Weimer, K. W. Barry, L. A. Goodwin, P. Bouffard, L. Li, J. Osterberger, T. T. Harkins, S. C. Slater, T. J. Donohue and C. R. Currie (2010). "An insect herbivore microbiome with high plant biomass-degrading capacity." PLoS Genet **6**(9): e1001129.
- Taupp, M., K. Mewis and S. J. Hallam (2011). "The art and design of functional metagenomic screens." Curr Opin Biotechnol **22**(3): 465-472.
- Tawfik, D. S. and A. D. Griffiths (1998). "Man-made cell-like compartments for molecular evolution." Nat Biotechnol **16**(7): 652-656.
- Tchigvintsev, A., H. Tran, A. Popovic, F. Kovacic, G. Brown, R. Flick, M. Hajighasemi, O. Egorova, J. C. Somody, D. Tchigvintsev, A. Khusnutdinova, T. N. Chernikova, O. V. Golyshina, M. M. Yakimov, A. Savchenko, P. N. Golyshin, K. E. Jaeger and A. F. Yakunin

- (2014). "The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes." Appl Microbiol Biotechnol.
- Tebbe, C. C. and W. Vahjen (1993). "Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast." Appl Environ Microbiol **59**(8): 2657-2665.
- Thomas, T., J. Gilbert and F. Meyer (2012). "Metagenomics - a guide from sampling to data analysis." Microb Inform Exp **2**(1): 3.
- Thompson, L. R., C. Field, T. Romanuk, D. Ngugi, R. Siam, H. El Dorry and U. Stingl (2013). "Patterns of ecological specialization among microbial populations in the Red Sea and diverse oligotrophic marine environments." Ecol Evol **3**(6): 1780-1797.
- Torsvik, V. L. and J. Goksoyr (1978). "Determination of bacterial DNA in soil." Soil Biology and Biochemistry **10**(1): 7-12.
- Tourlousse, D. M., F. Kurisu, T. Tobino and H. Furumai (2013). "Sensitive and substrate-specific detection of metabolically active microorganisms in natural microbial consortia using community isotope arrays." FEMS Microbiol Lett **342**(1): 70-75.
- Tsai, Y. L. and B. H. Olson (1991). "Rapid method for direct extraction of DNA from soil and sediments." Appl Environ Microbiol **57**(4): 1070-1074.
- Tseng, C. H. and S. L. Tang (2014). "Marine microbial metagenomics: from individual to the environment." Int J Mol Sci **15**(5): 8878-8892.
- Tuffin, M., D. Anderson, C. Heath and D. A. Cowan (2009). "Metagenomic gene discovery: how far have we moved into novel sequence space?" Biotechnol J **4**(12): 1671-1683.
- Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar and J. F. Banfield (2004). "Community structure and metabolism through reconstruction of microbial genomes from the environment." Nature **428**(6978): 37-43.
- 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**(1): 88-93.
- Uchiyama, T. and K. Miyazaki (2009). "Functional metagenomics for enzyme discovery: challenges to efficient screening." Curr Opin Biotechnol **20**(6): 616-622.
- Uchiyama, T. and K. Miyazaki (2010). "Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes." Appl Environ Microbiol **76**(21): 7029-7035.
- Uchiyama, T. and K. Watanabe (2008). "Substrate-induced gene expression (SIGEX) screening of metagenome libraries." Nat Protoc **3**(7): 1202-1212.
- Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J. A. Malek, G. Costa, K. McKernan, A. Sidow, A. Fire and S. M. Johnson (2008). "A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning." Genome Res **18**(7): 1051-1063.
- Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers and H. O. Smith (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." Science **304**(5667): 66-74.

- Verastegui, Y., J. Cheng, K. Engel, D. Kolczynski, S. Mortimer, J. Lavigne, J. Montalibet, T. Romantsov, M. Hall, B. J. McConkey, D. R. Rose, J. J. Tomashek, B. R. Scott, T. C. Charles and J. D. Neufeld (2014). "Multisubstrate isotope labeling and metagenomic analysis of active soil bacterial communities." *MBio* **5**(4): e01157-01114.
- Videnska, P., M. M. Rahman, M. Faldynova, V. Babak, M. E. Matulova, E. Prukner-Radovic, I. Krizek, S. Smole-Mozina, J. Kovac, A. Szmolka, B. Nagy, K. Sedlar, D. Cejkova and I. Rychlik (2014). "Characterization of egg laying hen and broiler fecal microbiota in poultry farms in croatia, czech republic, hungary and slovenia." *PLoS One* **9**(10): e110076.
- Wang, C., D. J. Meek, P. Panchal, N. Boruvka, F. S. Archibald, B. T. Driscoll and T. C. Charles (2006). "Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants." *Appl Environ Microbiol* **72**(1): 384-391.
- Wang, G. Y., E. Graziani, B. Waters, W. Pan, X. Li, J. McDermott, G. Meurer, G. Saxena, R. J. Andersen and J. Davies (2000). "Novel natural products from soil DNA libraries in a streptomycete host." *Org Lett* **2**(16): 2401-2404.
- Ward, D. M., R. Weller and M. M. Bateson (1990). "16S rRNA sequences reveal numerous uncultured microorganisms in a natural community." *Nature* **345**(6270): 63-65.
- Warnecke, F., P. Luginbuhl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernandez, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz and J. R. Leadbetter (2007). "Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite." *Nature* **450**(7169): 560-565.
- Warren, R. L., J. D. Freeman, R. C. Levesque, D. E. Smailus, S. Flibotte and R. A. Holt (2008). "Transcription of foreign DNA in Escherichia coli." *Genome Res* **18**(11): 1798-1805.
- Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Lin, E. Mathur, J. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Soll, K. O. Stetter, J. M. Short and M. Noordewier (2003). "The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism." *Proc Natl Acad Sci U S A* **100**(22): 12984-12988.
- Webster, N. S. and R. T. Hill (2001). "The culturable microbial community of the Great Barrier Reef sponge Rhopaloeides odorabile is dominated by an α -Proteobacterium." *Marine Biology* **138**(4): 843-851.
- Webster, N. S., K. J. Wilson, L. L. Blackall and R. T. Hill (2001). "Phylogenetic diversity of bacteria associated with the marine sponge Rhopaloeides odorabile." *Appl Environ Microbiol* **67**(1): 434-444.
- Weller, R. and D. M. Ward (1989). "Selective Recovery of 16S rRNA Sequences from Natural Microbial Communities in the Form of cDNA." *Appl Environ Microbiol* **55**(7): 1818-1822.

- Wexler, M., P. L. Bond, D. J. Richardson and A. W. Johnston (2005). "A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase." *Environ Microbiol* **7**(12): 1917-1926.
- Wild, J., Z. Hradecna and W. Szybalski (2002). "Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones." *Genome Res* **12**(9): 1434-1444.
- Wilson, M. C., T. Mori, C. Ruckert, A. R. Uria, M. J. Helf, K. Takada, C. Gernert, U. A. Steffens, N. Heycke, S. Schmitt, C. Rinke, E. J. Helfrich, A. O. Brachmann, C. Gurgui, T. Wakimoto, M. Kracht, M. Crusemann, U. Hentschel, I. Abe, S. Matsunaga, J. Kalinowski, H. Takeyama and J. Piel (2014). "An environmental bacterial taxon with a large and distinct metabolic repertoire." *Nature* **506**(7486): 58-62.
- Winogradsky, S. (1887). "Ber Schwefelbakterien." *Botanische Zeitung* **45**.
- Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms." *Proc Natl Acad Sci U S A* **74**(11): 5088-5090.
- Woese, C. R., G. E. Fox, L. Zablen, T. Uchida, L. Bonen, K. Pechman, B. J. Lewis and D. Stahl (1975). "Conservation of primary structure in 16S ribosomal RNA." *Nature* **254**(5495): 83-86.
- Wong, D. W., V. J. Chan, H. Liao and M. J. Zidwick (2013). "Cloning of a novel feruloyl esterase gene from rumen microbial metagenome and enzyme characterization in synergism with endoxylanases." *J Ind Microbiol Biotechnol* **40**(3-4): 287-295.
- Woyke, T. and J. Jarett (2015). "Function-driven single-cell genomics." *Microb Biotechnol* **8**(1): 38-39.
- Woyke, T., G. Xie, A. Copeland, J. M. Gonzalez, C. Han, H. Kiss, J. H. Saw, P. Senin, C. Yang, S. Chatterji, J. F. Cheng, J. A. Eisen, M. E. Sieracki and R. Stepanauskas (2009). "Assembling the marine metagenome, one cell at a time." *PLoS One* **4**(4): e5299.
- Wu, L., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje and J. Zhou (2001). "Development and evaluation of functional gene arrays for detection of selected genes in the environment." *Appl Environ Microbiol* **67**(12): 5780-5790.
- Xia, Y., F. Ju, H. H. Fang and T. Zhang (2013). "Mining of novel thermo-stable cellulolytic genes from a thermophilic cellulose-degrading consortium by metagenomics." *PLoS One* **8**(1): e53779.
- Xu, J. (2006). "Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances." *Mol Ecol* **15**(7): 1713-1731.
- Yooseph, S., K. H. Nealson, D. B. Rusch, J. P. McCrow, C. L. Dupont, M. Kim, J. Johnson, R. Montgomery, S. Ferriera, K. Beeson, S. J. Williamson, A. Tovchigrechko, A. E. Allen, L. A. Zeigler, G. Sutton, E. Eisenstadt, Y. H. Rogers, R. Friedman, M. Frazier and J. C. Venter (2010). "Genomic and functional adaptation in surface ocean planktonic prokaryotes." *Nature* **468**(7320): 60-66.
- Yooseph, S., G. Sutton, D. B. Rusch, A. L. Halpern, S. J. Williamson, K. Remington, J. A. Eisen, K. B. Heidelberg, G. Manning, W. Li, L. Jaroszewski, P. Cieplak, C. S. Miller, H. Li, S. T. Mashiyama, M. P. Joachimiak, C. van Belle, J. M. Chandonia, D. A. Soergel, Y. Zhai, K. Natarajan, S. Lee, B. J. Raphael, V. Bafna, R. Friedman, S. E. Brenner, A. Godzik, D. Eisenberg, J. E. Dixon, S. S. Taylor, R. L. Strausberg, M. Frazier and J. C. Venter (2007). "The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families." *PLoS Biol* **5**(3): e16.

- Youssef, N. H., P. C. Blainey, S. R. Quake and M. S. Elshahed (2011). "Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone Spring, Oklahoma)." Appl Environ Microbiol **77**(21): 7804-7814.
- Zhang, K., A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm and G. M. Church (2006). "Sequencing genomes from single cells by polymerase cloning." Nat Biotechnol **24**(6): 680-686.
- Zimmermann, K., M. Engeser, J. W. Blunt, M. H. Munro and J. Piel (2009). "Pederin-type pathways of uncultivated bacterial symbionts: analysis of o-methyltransferases and generation of a biosynthetic hybrid." J Am Chem Soc **131**(8): 2780-2781.
- Zuckerkandl, E. and L. Pauling (1965). "Molecules as documents of evolutionary history." J Theor Biol **8**(2): 357-366.

Figure legends

Figure 1. Timeline plot of important events in microbial ecology.

Events are labeled in vertical text grouped by color: blue for 16S rRNA methods, red for metagenomics, green for next-generation sequencing, and indigo for single cell genomics. The numbers in the parentheses refer to corresponding references shown below.

(1) (Brosius, Palmer et al. 1978); (2) (Carbon, Ebel et al. 1981); (3) (Lane, Pace et al. 1985); (4) (Weller and Ward 1989); (5) (Schmidt, DeLong et al. 1991); (6) (Healy, Ray et al. 1995); (7) (Handelsman, Rondon et al. 1998); (8) (Beja, Suzuki et al. 2000, Rondon, August et al. 2000); (9) (Breitbart, Salamon et al. 2002); (10) (Tyson, Chapman et al. 2004); (11) (Margulies, Egholm et al. 2005); (12) (Kvist, Ahring et al. 2007, Marcy, Ouverney et al. 2007); (13) (Bentley, Balasubramanian et al. 2008, Harris, Buzby et al. 2008, Valouev, Ichikawa et al. 2008); (14) (Rodrigue, Malmstrom et al. 2009); (15) (Yooseph, Nealson et al. 2010); (16) (Siegl, Kamke et al. 2011, Youssef, Blainey et al. 2011); (17) (Freeman, Gurgui et al. 2012); (18) (Rinke, Schwientek et al. 2013); (19) (Chi 2014).

Figure 2. An illustration of major steps in metagenomics and single cell genomics for biotechnology.

Schematic diagram showing the major steps in metagenomic discovery pipeline (left column) and single cell genomics (right column). The coloring scheme indicates that in metagenomics the taxonomic identity of the DNA is lost. In contrast, in single cell genomics, the taxonomic information is preserved and one can trace the taxonomic source of each MDA amplified genome and thus the clones resulting from these genomes.

Figure 1

Figure 2

Tables

Table 1: Comparison of sequence- and function-based screening of metagenomic libraries.

Sequence-based	Function-based
Homology-based method that relies on sequence matches	Functional metagenomics or bioprospecting metagenomics exploits gene(s) expression
Prior knowledge of probe or primer sequence is required	No probe or primer required
Sequence-based analysis helps reconstructing metabolic potential and quantifying ecological diversity by comparison with existing databases	Function-based screenings might discover novel genes/proteins or processes by selecting the positive clone libraries which may have entirely different sequences compared to existing databases
No prior knowledge of substrate is required	Substrate used in screening should be equal to the target substrate in order to select for specific function
No prior knowledge of product is required	Knowledge of products of the enzymatic reaction helps in detection of the positive clone(s)
No phenotypic signals while screening	Phenotype is must for selection of clone such as color production, zone of inhibition, fluorescence etc.
Expression of the cloned fragments is not required	Expression is must to detect the phenotype
Methods: PCR and Fluorescent in situ hybridization, Metagenomic microarray, Shotgun metagenomics	Methods: Functional complementation, HPTLC, Phenotypic microarray, Community isotope array, IVC-FACS, SIGEX, PIGEX

Table 2: List of enzymes identified by sequence-based metagenomics methods

Enzymes	Source	Approach	Reference
Nitrous oxide reductase, <i>nosZ</i>	Soil	High-throughput sequencing	(Orellana, Rodriguez et al. 2014)
Lipase	Soil (hot spring)	PCR, directed evolution	(Kumar, Sharma et al. 2013)
Glycosyl hydrolases	Cellulose- degrading Sludge	High-throughput sequencing	(Xia, Ju et al. 2013)
Sulfatases	Bathypelagic and sea bottom sediment	High-throughput sequencing	(Quaiser, Zivanovic et al. 2011)
Heterodisulfide reductase, F420H ₂ :quinone oxidoreductase	Methane-enriched marine sediments	High-throughput sequencing, metaproteomics	(Stokke, Roalkvam et al. 2012)
Glycosyl hydrolases	Buffalo rumen	High-throughput sequencing	(Patel, Patel et al. 2014)
Glycosyl hydrolases	Elephant feces	High-throughput sequencing	(Ilmberger, Gullert et al. 2014)
Carbohydrate active enzymes	Leaf-cutter ant fungus garden	High-throughput sequencing	(Suen, Scott et al. 2010)
Glycosyl hydrolases	Wallaby	High-throughput sequencing, Sanger sequencing	(Pope, Denman et al. 2010)
Antibiotic resistance genes	Hen feces	High-throughput sequencing	(Videnska, Rahman et al. 2014)
Glycosyl hydrolases	Termite hindgut	Sanger sequencing, PCR	(Warnecke, Luginbuhl et al. 2007)

Table 3: List of enzymes identified by functional-metagenomics approach

Enzymes	Source	Approach	Reference
Antibacterial enzymes, lipases, amylases, nucleases	Soil (Agriculture)	BAC libraries, Biological screening on solid media	(Rondon, August et al. 2000)
Esterase	Soil (Island)	Fosmid libraries, 96-well plate-based assay	(Choi, Kwon et al. 2013)
Glycosyl hydrolase	Soil (grassland)	Fosmid libraries, Biological screening on solid media	(Nacke, Engelhaupt et al. 2012)
N-acyl amino acid synthases	Soil	Plasmid libraries, HPLC-MS	(Craig and Brady 2011)
Lipases (cold-adaptive)	Soil (Peat swamp forest)	Fosmid, Biological screening on solid media	(Bunterngsook, Kanokratana et al. 2010)
Lipases, amylases, phosphatases and dioxygenases	Compost	Plasmid libraries, Biological screening on solid media	(Lammle, Zipper et al. 2007)
Lipases	Soil (forest)	Fosmid libraries, Biological screening on solid media	(Hong, Lim et al. 2007)
Carboxylesterases (cold-active and salt-resistant)	Oil contaminated area in sea	Phage libraries, Biological screening on solid media	(Tchigvintsev, Tran et al. 2014)
Beta-glucosidase	Hydrothermal spring (water, mud, sediment)	Plasmid libraries, Biological screening on solid media	(Schroder, Elleuche et al. 2014)
Ribulose-1,5-bisphosphate carboxylase/oxygenase	Hydrothermal vent	Fosmid libraries, Biological screening in liquid media, HPLC	(Bohnke and Perner 2014)
Beta-glucosidase	Marine	BAC libraries, Biological screening on solid media	(Fang, Fang et al. 2010)
Esterase	Cow rumen	Phage libraries, Biological screening on solid media	(Wong, Chan et al. 2013)
Bifunctional cellulase-xylanase	Bovine rumen	Fosmid libraries, Biological screening on solid media	(Rashamuse, Visser et al. 2013)
Glycosyltransferases	Elephant feces	Fosmid libraries, Metagenome extract thin-layer chromatography	(Rabausch, Juergensen et al. 2013)
Bifunctional glucosidase/xylosidase	Yak rumen	Cosmid libraries, Biological screening in liquid media	(Bao, Huang et al. 2012)
Alpha-amylase (high thermal- and salt-tolerance)	Biogas reactor	Plasmid and phage libraries, Biological screening on solid media	(Jabbour, Sorger et al. 2013)
Thermostable lipase	Enrichment culture at 65-75 °C	Cosmid libraries, Biological screening on solid media	(Chow, Kovacic et al. 2012)

Table 4: Single-cell amplified genomes (SAGs) for novel uncultivated microbial lineages with potentially interesting bioactivity pathways.

Name (taxonomy)	Habitat (source)	Genome size	Salient features	metabolic	Reference
Clade (Crenarchaeota)	C1b Free living archaea (soil)	130 kbp	NA		(Kvist, Ahring et al. 2007)
TM7 phylum	(candidate Free living bacteria (soil)	1.83 Mbp	Transporters for multidrug resistance, type IV secretion system		(Podar, Abulencia et al. 2007)
TM7 phylum	(candidate Symbiotic bacteria (human mouth)	2.86 Mbp	Glycosyl hydrolases, genes for type IV pilus biosynthesis		(Marcy, Ouverney et al. 2007)
<i>Poribacteria</i> phylum	(candidate Symbiotic bacteria (in marine sponges)	1.6 Mbp	Polyketide synthases (PKSs), isopropyl steroids cluster		(Siegl, Kamke et al. 2011)
OP11 division	(candidate Free living bacteria (from Zodletone spring)	270 kbp	Endoglucanase, amylopullulanase, laccase, antibiotic resistances and antibiotic production		(Youssef, Blainey et al. 2011)
SR1 phylum	(candidate Symbiotic bacteria (human mouth and skin)	460 kbp	Peptidase, pectinase, glycosyl hydrolase, in frame UGA codon for glycine		(Campbell, O'Donoghue et al. 2013)
OP9 (candidate phylum Atribacteria)	Free living bacteria (hot spring sediment)	110-872 kbp, composite Mbp	Glycohydrolases, endoglucanases		(Dodsworth, Blainey et al. 2013)
<i>Thiovulum</i> genus) (second fastest bacterium ever observed with speed of 615 $\mu\text{m/s}$)	(candidate Free living bacteria (phototrophic mats of Elkhorn Slough)	Composite Mbp	Oxidoreductases, dehydrogenases,		(Marshall, Blainey et al. 2012)
TM6 phylum	(candidate Symbiotic bacteria of an unknown host (hospital sink biofilm)	1.07 Mbp	Proteolytic peptidases		(McLean, Lombardo et al. 2013)
<i>Entotheonella</i> (candidate genus) belong to candidate phylum 'Tectomicrobia'	Symbiotic bacteria (marine sponges)	9 Mbp	28 biosynthetic gene clusters for polyketides and non-ribosomal peptides		(Wilson, Mori et al. 2014)