微生物学通报
Microbiology China
tongbao@im.ac.cn

Jan. 20, 2013, 40(1): 18−33


专论与综述

在气候变化、环境污染生物清除、农田肥力保持等重要领域，以宏基因组学为基础的系列组学技术的应用已极大地加深了人们对科学规律的认识。

杨云锋

# Omics breakthroughs for environmental microbiology
## Running title: Omics for environmental microbiology

YANG Yun-Feng*

(*State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing* 100084, *China*)

**Abstract:** To discover and characterize microbial members within their community and their roles in the environment, high-throughput omics approaches are being developed and applied. Among them, sequencing- and microarray-based metagenomics is the most mature key approach, providing information basis for most other omics technologies. Metatranscriptomics, metaproteomics and community metabolomics have had limited successes but already shown promising potentials. All of the omics approaches lie on the support of bioinformatics, which becomes a major bottleneck in omics applications. These new omics technologies are revolutionizing the field of environmental microbiology in revealing the genetic potentials and functional activities of microbial communities.

**Keywords:** Microbial community, Omics, Metagenomics, Climate change, Carbon cycling, Nitrogen cycling

**Foundation item:** National Science Foundation of China (No. 41171201); State Key Joint Laboratory of Environment Simulation and Pollution Control (No. 11Z03ESPCT)
***Corresponding author:** Tel: 86-10-62784692; ✉: yangyf@tsinghua.edu.cn
**Received:** October 25, 2012; **Accepted:** November 30, 2012

# 环境微生物学的组学技术应用和突破

## 短标题：环境微生物学的组学技术

杨云锋

(清华大学 环境学院 环境模拟与污染控制国家重点联合实验室 北京 100084)

**摘 要：**由于研究环境变化和微生物群落的需要，近年来高通量组学技术得到了迅猛开发和应用。其中，基于测序和芯片技术的宏基因组学是一个关键的、最成熟的组学技术，为大多数的其它组学技术提供了支撑。相比较而言，宏转录组学、宏蛋白质组学和宏代谢组学也取得了少数的有限成功，但已经显示出可喜的潜力。所有的组学技术都有赖于生物信息学，使得后者成为组学技术应用的一个主要的技术瓶颈。这些新的组学技术对环境微生物学领域产生了革命性的影响，极大地丰富了我们对于环境微生物基因资源和功能活性的了解。

**关键词：**微生物群落，组学，宏基因组学/元基因组学，气候变化，碳循环，氮循环

## 1 Introduction

Natural environments possess large microbial reservoirs that can be exploited for new enzymes[1−2], secondary metabolite[3] and modeling microbe-mediated nutrient and energy cycling[4] as microbes play an essential role in driving earth's biogeochemical cycles[5]. However, the study of the microbial reservoirs is limited by our inability to cultivate a majority of the microbes, known as the "great plate count anomaly". It is possibly the oldest unresolved microbiological mystery. To explain it, it has been proposed that microbes stochastically awake from dormancy to exploit transient conditions in an environment[6]. Alternatively, it can also be attributed to oxidative/substrate stress-induced death on transfer to and incubation in the laboratory or simply slow-growth strategies[7]. To overcome the barrier of "great plate count anomaly", there are intensive efforts to improve the cultivation by extended incubation time, the use of conditions that simulate the natural environment, and microbial encapsulation that physically separate microbes while allowing for molecular exchanges with the environment[8−12]. However, only limited success has been achieved to date, calling for the development and adoption of cultivation-independent approaches.

Owing to the rapid development of metagenomic and other tools, we have witnessed a rapid advancement of environmental microbiology. It is clear now that phylogenetic diversity of environmental microbes in soil, freshwater, marine water and air vastly exceeds what is informed from cultivation[13−14]. Among these natural habitats, soil appears to be the most dynamic and heterogeneous environments and thus diverse in microbial population, with estimates of $10^3$–$10^7$ species of eukaryotic, bacteria, archaea and viruses in one gram of soil[15−18], making soil the most valuable resource for discovering secondary metabolites such as antibiotics of the polyketide class[19−20]. In addition, diverse soil ecosystems allow for biological behaviors of competition, parasitism, predation and so on, thus it is of great interest to investigate the dynamic interactions among microbial group. Notably, extraction of gene, mRNA, protein or metabolite information of microbial community can help determine the physiological requirements and thus im-

prove the success in cultivating the uncultivable microbes.

Microbial activity and growth are constrained by the physical and chemical properties of natural habitats. It has been shown that soil particle size is negatively correlated to the diversity and number of bacteria[21]. Meanwhile, temperature is an important controlling factor since all of living organisms have an optimal spectrum of growth temperature. In addition, it has been shown that moisture, pH and organic matter content are also often crucial in shaping microbial communities[22].

## 2 Metagenomics-sequencing

Meta is a Greek word with the meaning of *after* or *beyond*. Metagenomics can be defined as the genome research of the collective microbial assemblage in environmental samples[23]. A major advantage of metagenomics is to identify novel functional genes. It is of little debate that the technical advancement of DNA sequencing in the past two decades is the most pivotal driver that revolutionizes the field of environmental microbiology. To date, sequencing-based metagenomics approaches have been applied to a number of different environments and served as an indispensable information basis for most other omics applications to assess microbial community and ecosystem functioning (Fig. 1).

Large-scale metagenomics projects have quickly come to depend entirely on so called the 2nd-generation DNA sequencing. Roche 454 genome sequencer was the first commercially available one among this kind, which employs a sequencing-by-synthesis pyrosequencing technology[24]. The unique features of 454 pyrosequencing include emulsion PCR thermal cycling on sepharose or Styrofoam beads to amplify DNA fragments, the use of a picotiter plate (PTP) that contains more than one million wells per plate but can hold only one bead per well, and the release of a pyrophosphate molecule when a nucleotide is incorporated during DNA synthesis (where the word *pyrosequencing* is derived from). Today, Roche 454 genome sequencers have two versions-GS FLX+ system and junior system with 1 M and 100 k reads capacity, respectively. Illumina sequencers, commonly referred to as Solexa, were commercialized in 2007. It uses sequencing-by-synthesis technology coupled with bridge amplification in a flow cell[25−26].

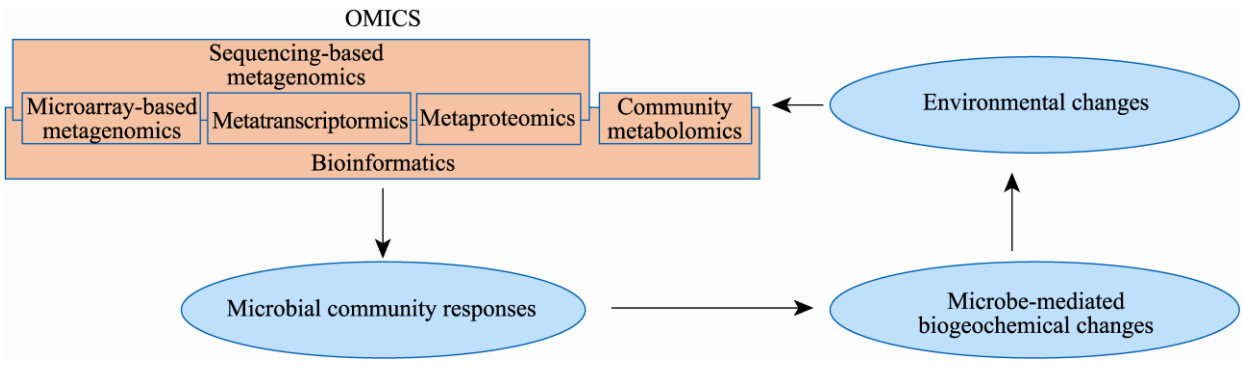


**Fig. 1 A conceptual framework of omics' roles in environmental microbiology research**

**图 1 组学技术在环境微生物学研究中发挥作用的概念框架**

Note: Among the omics approaches, sequencing-based metagenomics is usually necessary for providing informational basis of microarray-based metagenomics, metatranscriptomics and metaproteomics. Bioinformatics is considered to be essential for all of the omics approaches and a major challenge of future omics R&D. An integrated use of these omics approaches are crucial for understanding the responses of native microbes as well as their interactions with other microbes and environment, which elucidates microbe-mediated processes of soil respiration, methane and nitrous oxide emission, to name a few. Such a knowledge flow is helpful for gaining a mechanistic understanding of microbial responses to environmental changes and prediction of feedback responses at the large-scale ecosystem level using computational models. Meanwhile, it is well observed that environmental changes exert a clear impact on microbial community composition and functional activity, thereby the resulting feedback loop has dramatically complicated the research field of environmental microbiology.

The unique features include the linkage between adaptors ligated to DNA fragments and complementary oligos attached to the interior surfaces of a flow cell, which allows DNA amplification, and the incorporation and identification of fluorescent nucleotide during DNA synthesis. There are currently several versions. The HiSeq2500 platform can generate up to 120 Gb of sequencing data in roughly a day, whereas a low-end MiSeq platform can generate 1.5−2.0 Gb per run. Unlike Roche 454 pyrosequencing and Illumina technologies, the Applied Biosystems SOLiD sequencer uses a sequencing-by-oligo-ligation technology, which targets octamer DNA fragments during a DNA ligation reaction[27]. Another feature of SOLiD is the use of two-base encoding as an error-correction scheme to identify miscalls. The current SOLiD platforms have two versions: the 5500 system with 100 Gb capacity, and an improved 5500xl system with 250 Gb capacity. Life Technologies Ion Torrent uses a post-light sequencing technology by detection of hydrogen ion released during DNA synthesis. The 314, 316 and 318 series of ion chips can generate 10 Mb, 100 Mb and 1 Gb of sequencing data, respectively. Ion Proton I and II chips, which is coupled to the Ion Proton bench top sequencer, have higher capacities of 1 Gb and 10 Gb, respectively.

There are pros and cons of these 2nd-generation DNA sequencing platforms. The major advantages of Roche 454 pyrosequencing are its long read length (300–500 bp) and acceptably low rates of sequencing errors. The read length is a critical factor in matching sequences to those in GenBank or other databases[18] in order to generate accurate annotation of novel genes. To date, 454 pyrosequencing is the most commonly used platform for metagenomics. However, pyrosequencing relies on PCR-based DNA amplification, thus environmental contamination imposes a challenge in inhibiting DNA amplification. Also, the sequencing capacity of 454 pyrosequencing is low, thus the cost is high per megabase output[28]. It is roughly estimated, bearing in mind the caveat that the estimate may change rapidly, that pyrosequencing sequencing costs 0.01 cent per nucleotide[29]. In comparison, Illumina and SOLiD sequencing costs 20 times less. Thus they impose an obvious advantage over 454 pyrosequencing since it is unlikely for the latter to dramatically improve the sequencing capacity soon. Another major drawback of 454 pyrosequencing is the high errors in sequencing homopolymers, which are consecutive same nucleotides. In comparison, homopolymers are certainly less of an issue for Illumina. The major advantages of Illumina and SOLiD are the high sequencing capacity per run and the acceptably low rate of sequencing errors. The sequencing capacity is an important criterion, which improves the coverage depth of sequencing, referring to one genome equivalent, and can correct sequencing errors by additional coverage depths. However, their main drawback is the short-read length, which is inclined to generate more gaps and difficult to align, assign and annotate the sequences. Ion torrent is a relatively new technology. Similar to 454 pyrosequencing, it delivers high read-length sequences, but the capacity is low compared to Illumina and SOLiD. Notably, all of the 2nd-generation DNA sequencing technologies are prone to the errors and biases derived from DNA amplication. In this regard, the 3rd-generation DNA sequencing technologies, which are single-molecule, non-PCR sequencing technologies, are appealing in that it targets single cells and eliminates the biases introduced by template amplification. In addition, they promise remarkably high read-lengths (e.g. Pacific Biosciences). However, there are still issues such as high rates of sequencing errors, making the commercialization primitive at the time of this manuscript writing.

Metagenomics is the forefront of environmental microbiology research in recent years. It has been widely applied to analyze microbial community compositions in a variety of environmental samples from terrestrial, soil, marine, freshwater, gut microbiota, rare and extreme habitats such as deep sea floors and acid mines. Among them, 454 pyrosequencing examining 16S rDNA amplicons has been widely adopted by a large number of research groups (to list a few, [30−36]). An important point based on many of these studies is that gener-

ally the microbial community compositions appeared to be sensitive to anthropogenic and natural disturbance, unveiling a potential to use microbes as sensitive bioindicators of environmental quality and changes. In addition, selected functional gene amplicons of microbial community have also been targeted[37−40]. Meanwhile, Illumina and Ion Torrent sequencing have been successfully applied for microbial community assessment using short 16S rDNA regions[41−42].

An important but often ignored issue of metagenomics is to extract DNA representative of total microbial diversity, which is particularly important in light of recent recognition of the importance to assess the rare biosphere for its functional roles[43]. In consideration of environmental heterogeneity, it is advisable to pool multiple samples prior to analysis. The environmental matrix and contaminants are also factors to be considered since DNA recovery is highly dependent on the environmental matrix. In soil, the strong adherence of microbial cells and DNA to soil matrix makes DNA recovery inefficient[44−45]. Meanwhile, contaminants such as humic acids, polyphenols, polysaccharides and nucleases can interfere with efficient DNA amplification or labeling[45−48]. In general, DNA extraction should be harsh enough to lyse a variety of microbes, but with the permissible degree of DNA fragmentation. A soil grinding method with the use of liquid nitrogen has been widely adopted[47]. Nevertheless, commercial kits (e.g., MoBio Laboratories, Qiagen) can provide acceptable DNA yields under many cases. Regardless of DNA extraction methods, subsequent DNA purification is usually required. A combination of phenol and chloroform is typical, and treatment with hexadecyltrimethylammonium bromide (CTAB)[49−53] is often useful. Unfortunately, purification steps inevitably result in a significantly decreased DNA yield and increased processing time. Occasionally, the issue of DNA purification can be alleviated by diluting DNA along with contaminants[54], or by addition of bovine serum albumin to prevent inhibition of humic acids[55].

Based on the discussion, the choice of sequencing technologies depends not only on the cost, but on whether data analysis and DNA sampling methods will allow the completion of the project. The cost, labor and time required for sequencing has been reduced dramatically in the past two decades, as a result of the rapid technology development, and this trend is likely to continue. Despite serious drawbacks for de novo sequencing, 454 pyrosequencing has made significant progress in providing useful information about microbial community[56]. However, as read length is gradually extended, Illumina technology is likely to prove more cost-effective at current pricing and output and thus outweigh 454 pyrosequencing in the near future. We predict that mass de novo sequencing of microbial community desires more advanced and low-cost technologies than the current ones on the market. However, such technologies will soon be available, given the current rate of technology advancement. When it occurs, the burden of difficulty will probably be shifted from the laboratory-based research to the interpretation of sequence information, calling for significant advances in bioinformatics and computational biology to solve algorithmic challenges in dealing with the issues of read-length, raw accuracy and throughput. Notably, the sequencing cost is not necessarily the major cost of metagenomics. With the astonishing rate to produce sequencing data, storage and analysis costs have become a significant part of the budget. For all second-generation technologies, the availability of bioinformatics software is not yet optimized to the level that makes researchers satisfactory.

## 3 Metagenomics-microarray

The use of microarrays is an excellent alternative to high-throughput DNA sequencing to target functional genes and phylogenetic markers. Microarrays have a number of advantages in that they are high-throughput, specific, quantitative, free of biases of PCR amplification and random sampling[57], and allow for comparison between multiple samples. However, the probe design limits to existing sequence database, thus microarrays can only detect what is already known. Therefore, microarrays are commonly used for generally characterized ecosys-

tems instead of novel gene discovery.

PhyloChip and GeoChip are two most popular microarrays[58−61]. PhyloChip is based on diagnostic markers of phylogenetic diversity such as the 16S rRNA gene, whereas GeoChip contains probes to target a wide range of functional genes for their detection in a given environment. The most advanced version of GeoChip to date, GeoChip 4.0, is designed to detect a staggering number of ~152 000 functional genes involved in biogeochemical cycles of carbon, nitrogen, phosphorus and sulfur, and stress responses such as antibiotics and metal resistance. Recently, several studies have used a combination of PhyloChip and GeoChip to explore information of phylogenetic and functional genes simultaneously[62−64].

PhyloChip is in direct competition of high-throughput sequencing of 16S rDNA genes. With the rapid development of sequencing technology, the application of PhyloChip is limited now. In contrast, GeoChip is complementary to sequencing by targeting a wide array of functional genes. Thus, GeoChip has been widely used for environmental studies. In a US Department of Energy's Field Research Center at Oak Ridge, TN, USA, GeoChip was used to profile microbial community of uranium-contaminated groundwater under bioremediation treatment with ethanol injection[65]. The results showed that the functional potentials of microbial community was very responsive to bioremediation and provided mechanistic explanation on ethanol, as an electron donor, for uranium reduction and immobilization. In marine water, Lu et al. used GeoChip to examine the effect of a massive deep sea oil spill in the Gulf of Mexico[66]. The results unveiled a huge functional potential for intrinsic bioremediation or natural attenuation since metabolic genes required for degradation were highly enriched in the oil plume. In a newly grown deep sea protochimney of hydrothermal vent at the Juan de Fuca Ridge, GeoChip detected vast, dramatic changes of microbial communities along the growth of chimney, suggesting that hydrothermal microbial communities are highly dynamic in metabolism and physiology. Meanwhile, analyzing soil microbial communities by GeoChip appeared to be a major focus. Soil is estimated to contain the most sophiscated microbial communities[15−16] and cannot be fully explored by the capacities of current technologies of DNA sequencing[57]. Consequently, GeoChip appeared to be most suitable as an alternative. In a long-term experimentally warming grassland site of Oklahoma, USA, the results of GeoChip showed that microbial communities played crucial roles in regulating soil carbon dynamics[67]. In an alpine grassland ecosystem of Tibetan plateau, China, livestock grazing stimulated functional potentials of nitrogen mineralization and nitrification, but suppressed those of carbon cycling of carbon fixation, degradation and methane, which could be linked to the changes of nitrogen and carbon inputs[68]. In contrast, functional potentials of virulence, stress and antibiotics resistance genes were increased due to the presence of livestock. Worthy of mention is that GeoChip is a high-resolution metagenomics tool suitable to tackle microbial ecology theories. It has been shown by GeoChip that a relatively flat gene-area relationship was present for forest soil microbial community, suggesting that the turnover in space of microbial communities was very low[69]. GeoChip is also applicable for engineered systems. It revealed that the functional diversity of microbial electrolysis cells were high, with the highest hydrogen yields linked to the highest microbial diversity[70]. Further analysis indicated that reactor operation, rather than startup conditions of microbial electrolysis cells were crucial in shaping microbial communities.

Despite wide applications, GeoChip has several inherent challenges just as any of the other omics technologies[71]. The probe coverage is crucial in comprehensive exploration of microbial communities. However, the fact that continuous addition of a tremendous volume of novel sequences in public databases suggests that our knowledge of microbial communities is very limited. Consequently, the probe coverage of GeoChip is far behind what is sufficient. Specificity is another crucial issue since many gene variants exist in an environmental sample. Therefore, a challenge of GeoChip is to take

into account of gene variants with high homologies, even though many of them are unknown sequences in the environment.

## 4 Metatranscriptomics

While mass de novo whole-genome sequencing of microbial community is not yet feasible due to the cost or technical difficulty, an alternative is to focus on protein-coding sequence by sequencing mRNA[72], which is called metatranscriptomics. Metatranscriptomics randomly sequences mRNA fragments and requires computational assembly, thus serves as readily accessible 'entry point' for studying microbial community[73]. It represents the next logical step to reach beyond genomic potentials of the microbial community toward *in situ* activity, since gene transcription is required for functioning. Conversely, metatranscriptomics can be utilized for gene discovery[74]. To date, metatranscriptomics projects use 454 pyrosequencing for its relative long read-lengths[73,75−76] because very short-read technologies (Illumina or Solid) prevent efficient de novo assembly of transcripts that do not have a closely related template as efficient. Also, the resolution of genes with similar sequences is more challenging for Illumina or Solid technologies. However, given the advantage in sequence output, very short read sequencing technologies (Illumina or Solid) will provide sufficient depth for effective assembly in many genes, which is important since it allows better sequence assembly as well as coverage of rare mRNA sequences. In this regard, it is also likely that there is a shift from 454 pyrosequencing to Illumina, since its read-length continues to be improved. Regardless, the cost of metatranscriptomics is likely to fall quickly in the near future, thanks to the advancement of sequencing technologies to analyze large numbers of microbial cDNAs quickly and cheaply.

Recovering mRNA from natural environments is challenging because of mRNA instability and the need to separate mRNAs from more abundant rRNAs and tRNAs. In addition, mRNAs of housekeeping genes are of little interest for many studies, thus removing them is also useful. These challenges have prevented the wide application of metatranscriptomics, despite recent efforts to address them. The first study of this kind successfully recovered mRNAs encoding proteins functioning in carbon and nitrogen cycling from freshwater and marine bacterioplanktons[77]. Using metatranscriptomics, immense metabolic diversity was unveiled in the ocean since a large fraction of environmental mRNAs was novel[78]. Similarly, a large fraction of soil mRNAs were indicative of novel genes, signifying the importance of exploring the unknown[79]. A milestone publication utilizing metatranscriptomics came out in 2008, in which it was shown that key metabolic genes such as photosynthesis, carbon and nitrogen cycling were highly represented in RNA pools of marine surface water[78]. Further analysis showed that a large fraction of ocean water RNA pools is comprised of microbial small RNAs[80]. Many of small RNAs were mapped to intergenic regions of microbial genomes generated from similar habitats, suggestive of potentially new regulatory elements. Although metatranscriptomics is not yet a quantitative technology, there has been effort to use transcriptomics profiling for comparative study. More transcripts of photosynthesis, carbon metabolism and oxidative phosphorylation were present in the surface ocean microbial community for the day samples than the night ones, and *Cyanobacteria* transcripts were abundant throughout the day/night cycles[81]. In soil, metatranscriptomics was used to profile the community diversity of forest soil eukaryotic microbes[79]. An interesting result was that the taxonomic distribution from metatranscriptomics and metagenomics did not coincide. This suggested that the omics profiling was highly incomplete, signifying the importance to develop and utilize sequencing technologies with higher capacity. Similarly, a more recent report showed that more than half of metatranscriptomics sequences from beech or spruce forest soil were previously unknown sequences[82]. Among them, the researchers have identified a novel family of oligopeptide transporters of fungi and provided experimental evidence that it is functional[83]. This success indicated the potential of meta-

transcriptomics in gene discovery and biotechnological applications.

## 5 Metaproteomics

It is well understood that metagenomics and metatranscriptomics are not sufficient to understand the activity and physiology of microbial community, calling for the analysis of environmental samples at the protein level. Metaproteomics, which analyzes protein profiles of microbial communities, can address the function more directly than metagenomics and metatranscriptomics since proteins are responsible for performing function and activities. To date, there are a variety of applications such as protein cataloging, comparative and quantitative proteomics, post-translational modifications, protein-protein interactions and genotyping, allowing for the establishment of linkages between gene and function in microbial communities.

A state-of-arts, standard metaproteomics experiment usually includes protein extraction and purification from an environmental sample, followed by enzymatic digestion of proteins, MS and MS/MS analysis, and protein identification based on reference database searches. The reference protein databases are typically based on genomics or metagenomics project. Thus, proteins cannot be resolved if their nucleotide sequences are not available or annotated. Conversely, proteomic data can contribute to the refinement of sequencing annotations[84]. Thus, the combination of metaproteomics and metagenomics is very helpful in the analysis of environmental samples. Indeed, a term proteogenomics has been coined to indicate the simultaneous performance of genomics and proteomics[85−86].

Although metaproteomics is still in its infancy, there are excellent studies using large-scale MS/MS-based metaproteomics, or proteogenomics in some cases. A prominent example is the protein profiling of natural biofilms in an acid mine drainage[87], which are low complexity models due to the extreme conditions. More than 2 000 proteins from the top five abundant species were sampled to saturation (up to 48% protein coverage), whereas as low as 1% of the total population can also be detected. A subsequent study unveiled interpopulation recombination since strain-specific genome information is available for the dominant species[88], which was confirmed by extensive semi-quantitative protein profiling[89]. The virtue of proteogenomics was also demonstrated by the identification of 45%−65% of the predicted proteins of the dominant species in acid mines[90]. Also, proteogenomics was carried out in study microbial community of plant leaves[91], leading to the identification of protein profiling o predominant species of microbial community (*Sphingomonas, Methylobacterium* and *Pseudomonas*). Similarly, microbial community in the termite hindgut or sheep rumen was recently analyzed by proteogenomics[74,92], which revealed the presence of a large number of cellulolytic and xylanolytic enzymes. Other proteomics applications include a comparative metaproteome study to examine the mechanisms of enhanced biological phosphorus removal (EBPR), a process of interest to a wastewater treatment plant[93−94]. The results indicated distinct differences in protein profiles and revealed metabolic pathways essential for EBPR in uncultured dominant microbial species of the EBPR reactor. In highly complex environments, microbial communities in lake water, soil or aquifier sediments were analyzed, entailing the broad applicability of proteomics[95−97].

However, several needs should be fulfilled for further development of proteomics. The protein extraction methods need to be developed to deal with organic and metal contaminants in environmental samples. The sensitivity and accuracy of mass spectrometer is also crucial for improvement in protein coverage. Meanwhile, quantitative metaproteomics is urgently needed for comparisons between samples, in addition to the need of software tools to overcome current challenges in data analysis. Successes in these issues will be important to advance our understanding of earth's biogeochemical cycles in which microbial communities play important roles and harness proteins with biological activities desirable for biotechnological applications.

## 6 Community metabolomics

Community metabolomics is the large-scale

metabolite profiling in a given environmental sample. Metabolites, which are small chemical molecules, are required for the building blocks, energy consumption and signal communication of microbial community, providing a direct description of the physiological and communication status[98]. Therefore, metabolite profiling is important for the comprehensive investigation of microbial community. However, since metabolites vary in molecular weight, size, solubility and other physical and chemical properties, it imposes a challenge on comprehensive detection and measurement of metabolites. In addition, as other omics approaches, another challenge is to extract and interpret the vast amount of information from community metabolomics, entailing the need for bioinformatics tool development and validation.

Community metabolomics utilizes analytic chemistry tools, such as nuclear magnetic resonance spectroscopy and chromatographic separation techniques coupled to mass spectrometry[99]. Gas chromatography coupled to mass spectrometry (GC-MS), which targets volatile metabolites either naturally or via chemical derivatisation, has been widely used for community metabolomics. It enables the profiling of hundreds of low-molecular-weight metabolites of up to 1 000 Da, such as amino acids, fatty acids and sugars, in a single run[100]. The coupled mass analyzers include low-resolution quadrupole instruments, or fast-scanning time of flight (TOF) MS. In contrast, liquid chromatography coupled to mass spectrometry (LC-MS) can detect and quantify higher molecular mass metabolites and do not require derivatization. Another metabolomics technique, nuclear magnetic resonance (NMR), is also popular since it detects a wide range of metabolites[101] and is fast, cheap, noninvasive and quantitative. However, the sensitivity of NMR is relatively low. As any of these techniques has limitations in detecting the huge metabolic diversity and inherent analytical errors, a multifaceted approach with combinations of GC-MS, LC-MS and NMR should be considered.

There have been several studies of community metabolomics[102−104]. For example, a strong effect of gut microflora was observed on mammalian blood metabolites, suggesting that microbes directly affected the metabolism capacity of human body[103]. This finding was of great interest for exploring microbiome toward improving human health. However, the application of metabolomics for studying microbial community is still in an early stage. It is thereby important to develop more quantitative tools, and integrate metabolomics with data sets from the other "omics" to improve data quality and provide novel insights than a single "omics" technology alone can offer.

## 7 Bioinformatics

Omics typically produce very large datasets covering a large number of previously undescribed genomes. Therefore, a major challenge in omics applications in environmental microbiology lies in handling and interpretation of high-density data. Bioinformatics is the application of informatics tools to biological data, including high-throughput sequence quality scoring, alignment, assembly, computation of relative abundance, data access and comparison across various platforms.

Raw data of high-throughput Omics experiments typically contain a high level of inherent noise, which has a variety of context-dependent error distributions. Bioinformatics will be important to resolve this issue. However, to date there is no appropriate software that can assemble de novo genomes from short reads of less than 50 bp[18]. Therefore, reconstruction of microbial genomes from highly diverse microbial community is not available. Given the non-exhaustive nature of omics dataset, sequence assembly will be of limited benefit and prone to high errors of chimeric assembly. Regardless, there has been recent effort to develop new assembly tools[105−107].

There appears to be a huge need of sequence alignment, annotation and function prediction by comparing DNA reads against previously known gene products and exploring the taxonomical content[108−110]. Notably, the predictive power of the sequence annotation is limited by known gene functions in public databases, thus the utility of se-

quence annotation is affected by bias caused by inaccurate annotation in the databases. BLAST (Basic Local Alignment Search Tool)[111] is an essential program to compare sequence to NCBI DNA and protein databases. Until recently, CLUSTALW has been the most widely used program for sequence alignment[112], but several improved alternatives, such as T-COFFEE and PROBCONS, have been developed[113]. However, BLAST is not adequate for handling short-read sequences, calling for new algorithms tailed for short-reads. An increasing number of alignment algorithms have thus been developed to fulfill the need[114]. MAQ37 and SHRiMP take sequence quality in account during sequence alignment. Other algorithms use accelerated processing to make gapped alignment more efficient[114−115].

An emerging, exciting development of bioinformatics is to generate co-occurrence patterns in association networks based on high density omics data, which is robust to the typically high level of inherent noise in omics data, and help unveil interaction within microbial community and topological details of microbial community structure[116]. Furthermore, environmental parameters can also be incorporated to establish linkages between microbial community and natural environments. Using a random matrix theory-based algorithm, association networks were reconstructed from both sequencing- and microarray-based data to examine long-time effects of elevated carbon dioxide[117−119]. The results showed that the networks possessed general topological features of complex systems and elevated carbon dioxide clearly altered the network interactions of microbial phylogenetic groups or functional genes. Similarly, association networks were reconstructed from sequencing-based metagenomics data collected from ocean time-series site of southern California, revealing the dynamics of microbial communities over time and possible 'keystone' species[120−121].

As most environmental microbiologists are not professional bioinformaticians, it is necessary to develop user-friendly automated omics "pipelines" that allow for automated annotation, phylogenetic classification and metabolic pathway reconstruction of metagenomics sequences. mg-RAST, the abbreviation of the metagenomics RAST server[122], has been used to reconstruct metabolic pathways of several different microbial communities[123]. Meanwhile, the stand alone program MEGAN can rapidly provide phylogenetic classification, functional assessments and visual outputs[109].

The importance of setting up consensus guidelines of metadata MIGS (Minimum Information about a Genome Sequence) has been accepted in the bioinformatics community, which might simulate the most successful MIAME (Minimum Information About a Microarray Experiment) guidelines for microarray experiments[124]. MIGS should include environmental data of climate, land management and environmental measurements and consistent sequence reads format styles for multiple samples, which are useful for comparison of different metagenomics projects and large-scale re-analysis with novel bioinformatics ideas.

## 8 Conclusions

In summary, the vast numbers and great diversity of microbial community, tethered with the heterogeneity of the environment, impose major challenges. Omics breakthroughs have greatly facilitated the understanding of the interactions within microbial communities and with their environment, albeit inherent biases for any of the approaches. Metagenomics is the most mature and widely used omics tool. Representing functional potentials, metagenomics reveals the range of potential activities in microbial communities. Meanwhile, metatranscriptomics has been used to indicate the gene activity since mRNA is abundant only in live cells. In contrast, the development and application of metaproteomics and community metabolomics are more relevant in reflecting microbial physiology and interaction within microbial communities and with environment. Although there are many difficulties in extracting and quantifying protein and metabolite components, there have been efforts in addressing these issues and the coupling of genomic and proteomic approaches[87,95].

The use of omics for microbial communities is still in the infancy but likely to become more prominent when the cost falls and the necessary bioinformatics tools are in place. In addition, parallel applications of metagenomics, metatranscriptomics, metaproteomics and community metabolomics will allow for simultaneous study of microbial community at different layers. These methods hold great promise in revealing the overall picture of microbial diversity and activity, as well as rare spheres of previously unknown and uncharacterized microbes but possibly with functions that are important for the community. Now it is only an exposition for an exciting era of environmental microbiology.

## 参 考 文 献


[1] Hjort K, Bergström M, Adesina MF, et al. Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil[J]. FEMS Microbiology Ecology, 2010, 71(2): 197−207.

[2] Liu N, Yan X, Zhang ML, et al. Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes[J]. Applied and Environmental Microbiology, 2011, 77(1): 48−56.

[3] Lombard N, Prestat E, Van Elsas JD, et al. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics[J]. FEMS Microbiology Ecology, 2011, 78(1): 31−49.

[4] Stockdale EA, Brookes PC. Detection and quantification of the soil microbial biomass-impacts on the management of agricultural soils[J]. The Journal of Agricultural Science, 2006, 144(4): 285−302.

[5] Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles[J]. Science, 2008, 320(5879): 1034−1039.

[6] Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation[J]. Microbiological Review, 1995, 59(1): 143−169.

[7] Janssen PH. Dormant microbes: scouting ahead or plodding along?[J]. Nature, 2009, 458(7240): 831−831.

[8] Zengler K, Toledo G, Rappé M, et al. Cultivating the uncultured[J]. Proceedings of the National Academy of Sciences of the United States of American, 2002, 99(24): 15681−15686.

[9] Ingham CJ, Sprenkels A, Bomer J, et al. The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms[J]. Proceedings of the National Academy of Sciences of the United States of American, 2007, 104(46): 18217−18222.

[10] Ben-Dov E, Kramarsky-Winter E, Kushmaro A. An *in situ* method for cultivating microorganisms using a double encapsulation technique[J]. FEMS Microbiology Ecology, 2009, 68(3): 363−371.

[11] Nunes da Rocha U, Van Overbeek L, Van Elsas JD. Exploration of hitherto-uncultured bacteria from the rhizosphere[J]. FEMS Microbiology Ecology, 2009, 69(3): 313−328.

[12] Nichols D, Cahoon N, Trakhtenberg EM, et al. Use of ichip for high-throughput *in situ* cultivation of "uncultivable" microbial species[J]. Applied and Environmental Microbiology, 2010, 76(8): 2445−2450.

[13] Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity[J]. Journal of Bacteriology, 1998, 180(18): 4765−4774.

[14] Curtis TP, Sloan WT. Exploring microbial diversity—A vast below[J]. Science, 2005, 309(5739): 1331−1333.

[15] Gans J, Wolinsky M, Dunbar J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil[J]. Science, 2005, 309(5739): 1387−1390.

[16] Schloss PD, Handelsman J. Toward a census of bacteria in soil[J]. PLoS Computational Biology, 2006, 2(7): e92.

[17] Fierer N, Breitbart M, Nulton J, et al. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil[J]. Applied and Environmental Microbiology, 2007, 73(21): 7059−7066.

[18] Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters[J]. Applied and Environmental Microbiology, 2008, 74(5): 1453−1463.

[19] Burgess JG, Jordan EM, Bregu M, et al. Microbial antagonism: a neglected avenue of natural products research[J]. Progress in Industrial Microbiology, 1999, 35: 27−32.

[20] Garbeva P, de Boer W. Inter-specific interactions between carbon-limited soil bacteria affect behavior and gene expression[J]. Microbial Ecology, 2009, 58(1): 36−46.

[21] Sessitsch A, Weilharter A, Gerzabek MH, et al. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment[J]. Applied and Environmental Microbiology, 2001, 67(9): 4215−4224.

[22] Hassink J, Bouwman L, Zwart KB, et al. Relationships between soil texture, physical protection of organic matter, soil biota, and C and N mineralization in grassland soils[J]. Geoderma, 1993, 57(1/2): 105−128.

[23] Handelsman J, Rondon MR, Brady SF, et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products[J]. Chemistry and Biology, 1998, 5(10): R245−R249.

[24] Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors[J]. Nature, 2005, 437(7057): 376−380.

[25] Adessi C, Matton G, Ayala G, et al. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms[J]. Nucleic Acids Research, 2000, 28(20): e87.

[26] Fedurco M, Romieu A, Williams S, et al. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies[J]. Nucleic Acids Research, 2006, 34(3): e22.

[27] Mardis ER. The impact of next-generation sequencing technology on genetics[J]. Trends in Genetics, 2008, 24(3): 133−141.

[28] Claesson MJ, Wang Q, O'Sullivan O, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions[J]. Nucleic Acids Research, 2010, 38(22): e200.

[29] Hudson ME. Sequencing breakthroughs for genomic ecology and evolutionary biology[J]. Molecular Ecology Resources, 2008, 8(1): 3−17.

[30] Sogin ML, Morrison HG, Huber JA, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere"[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(32): 12115−12120.

[31] Roesch LFW, Fulthorpe RR, Riva A, et al. Pyrosequencing enumerates and contrasts soil microbial diversity[J]. The ISME Journal, 2007, 1(4): 283−290.

[32] Breitbart M, Hoare A, Nitti A, et al. Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Cienegas, Mexico[J]. Environmental Microbiology, 2008, 11(1): 16−34.

[33] Murphy EF, Cotter PD, Healy S, et al. Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models[J]. Gut, 2010, 59(12): 1635−1642.

[34] Rousk J, Bååth E, Brookes PC, et al. Soil bacterial and fungal communities across a pH gradient in an arable soil[J]. The ISME Journal, 2010, 4(10): 1340−1351.

[35] Nacke H, Thürmer A, Wollherr A, W et al. Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils[J]. PLoS One, 2011, 6(2): e17000.

[36] Serino M, Luche E, Gres S, et al. Metabolic adaptation to a high-fat diet is associated with a change in the gut microbiota[J]. Gut, 2012, 61(4): 543−553.

[37] Leininger S, Urich T, Schloter M, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils[J]. Nature, 2006, 442(7104): 806−809.

[38] Mou XZ, Sun SL, Edwards RA, et al. Bacterial carbon processing by generalist species in the coastal ocean[J]. Nature, 2008, 451(7179): 708−711.

[39] Pegard A, Miquel C, Valentini A, et al. Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces[J]. Journal of Agricultural and Food Chemistry, 2009,

57(13): 5700−5706.

[40] Kowalczyk R, Taberlet P, Coissac E, et al. Influence of management practices on large herbivore diet—Case of European bison in Białowieża Primeval Forest (Poland)[J]. Forest Ecology and Management, 2011, 261(4): 821−828.

[41] Lazarevic V, Whiteson K, Huse S, et al. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing[J]. Journal of microbiological methods, 2009, 79(3): 266−271.

[42] Miller W, Hayes VM, Ratan A, et al. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil)[J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(30): 12348−12353.

[43] Elshahed MS, Youssef NH, Spain AM, et al. Novelty and uniqueness patterns of rare members of the soil biosphere[J]. Applied and Environmental Microbiology, 2008, 74(17): 5422−5428.

[44] Bakken LR. Separation and purification of bacteria from soil[J]. Applied and Environmental Microbiology, 1985, 49(6): 1482−1487.

[45] Frostegård Å, Courtois S, Ramisse V, et al. Quantification of bias related to the extraction of DNA directly from soils[J]. Applied and Environmental Microbiology, 1999, 65(12): 5409−5420.

[46] Tebbe CC, Vahjen W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast[J]. Applied and Environmental Microbiology, 1993, 59(8): 2657−2665

[47] Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse composition[J]. Applied and Environmental Microbiology, 1996, 62(2): 316−322.

[48] Sylvia DM, Fuhrmann JJ, Hartel P, et al. Principles and Applications of Soil Microbiology[M]. 2nd ed. New Jersey: Pearson Prentice Hall, 2005.

[49] Holben WE, Jansson JK, Chelm BK, et al. DNA probe method for the detection of specific microorganisms in the soil bacterial community[J]. Applied and Environmental Microbiology, 1988, 54(3): 703−711.

[50] Selenska S, Klingmüller W. Direct detection of nif-gene sequences of *Enterobacter agglomerans* in soil[J]. FEMS Microbiology Letters, 1991, 80(2/3): 243−245.

[51] Roose-Amsaleg CL, Garnier-Sillam E, Harry M. Extraction and purification of microbial DNA from soil and sediment samples[J]. Applied Soil Ecology, 2001, 18(1): 47−60.

[52] Lee SW, Won K, Lim HK, et al. Screening for novel lipolytic enzymes from uncultured soil microorganisms[J]. Applied Microbiology and Biotechnology, 2004, 65(6): 720−726.

[53] Knaebel DB, Crawford RL. Extraction and purification of microbial DNA from petroleum-contaminated soils and detection of low numbers of toluene, octane and pesticide degraders by multiplex polymerase chain reaction and Southern analysis[J]. Molecular Ecology, 2008, 4(5): 579−591.

[54] Altshuler ML. PCR Troubleshooting: The Essential Guide[M]. Hethersett Norwich: Horizon Scientific Press, 2006.

[55] Kreader CA. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein[J]. Applied and Environmental Microbiology, 1996, 62(3): 1102−1106.

[56] Smith MG, Gianoulis TA, Pukatzki S, et al. New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis[J]. Genes and Development, 2007, 21(5): 601−614.

[57] Zhou JZ, Wu LY, Deng Y, et al. Reproducibility and quantitation of amplicon sequencing-based detection[J]. The ISME Journal, 2011b 5(8): 1303−1313.

[58] Call DR, Borucki MK, Loge FJ. Detection of bacterial pathogens in environmental samples using DNA microarrays[J]. Journal of Microbiological Methods, 2003, 53(2): 235−243.

[59] He ZL, Gentry TJ, Schadt CW, et al. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes[J]. The ISME Journal, 2007, 1(1): 67−77.

[60] Metfies K, Medlin LK. Feasibility of transferring fluorescent in situ hybridization probes to an 18S rRNA gene phylochip and mapping of signal intensities[J]. Applied and Environmental Microbiology, 2008, 74(9): 2814−2821.

[61] He ZL, Deng Y, Van Nostrand JD, et al. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity[J]. The ISME Journal, 2010, 4(9): 1167−1179.

[62] Hazen TC, Dubinsky EA, DeSantis TZ, et al. Deep-sea oil plume enriches indigenous oil-degrading bacteria[J]. Science, 2010, 330(6001): 204−208.

[63] Beazley MJ, Martinez RJ, Rajan S, et al. Microbial community analysis of a coastal salt marsh affected by the Deepwater Horizon Oil Spill[J]. PLoS One, 2012, 7(7): e41305.

[64] Wakelin SA, Barratt BIP, Gerard E, et al. Shifts in the phylogenetic structure and functional capacity of soil microbial communities follow alteration of native tussock grassland ecosystems[J]. Soil Biology and Biochemistry, 2012, doi: 10.1016/j.soilbio.2012.07.003.

[65] Van Nostrand JD, Wu WM, Wu LY, et al. GeoChip-based analysis of functional microbial communities during the reoxidation of a bioreduced uranium-contaminated aquifer[J]. Environmental Microbiology, 2009, 11(10): 2611−2626.

[66] Lu Z, Deng Y, Van Nostrand JD, et al. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume[J]. The ISME Journal, 2012, 6(2): 451−460.

[67] Zhou JZ, Xue K, Xie JP, et al. Microbial mediation of carbon-cycle feedbacks to climate warming[J]. Nature Climate Change, 2012, 2(2): 106−110.

[68] Yang YF, Wu LW, Lin QY, et al. Responses of the functional structure of soil microbial community to livestock grazing in the Tibetan alpine grassland[J]. Global Change Biology, Accepted, doi: 10.1111/gcb.12065.

[69] Zhou JZ, Kang S, Schadt CW, et al. Spatial scaling of functional gene diversity across various microbial taxa[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(22): 7768−7773.

[70] Liu WZ, Wang AJ, Cheng SA, et al. Geochip-based functional gene analysis of anodophilic communities in microbial electrolysis cells under different operational modes[J]. Environmental Science and Technology, 2010, 44(19): 7729−7735.

[71] He ZL, Van Nostrand JD, Deng Y, et al. Development and applications of functional gene microarrays in the analysis of the functional diversity, composition, and structure of microbial communities[J]. Frontiers of Environmental Science and Engineering in China, 2011, 5(1): 1−20.

[72] Adams MD, Kelley JM, Gocayne JD, et al. Complementary DNA sequencing: expressed sequence tags and human genome project[J]. Science, 1991, 252(5013): 1651−1656.

[73] Toth AL, Varala K, Newman TC, et al. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality[J]. Science, 2007, 318(5849): 441−444.

[74] Warnecke F, Hess M. A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts[J]. Journal of Biotechnology, 2009, 142(1): 91−95.

[75] Cheung F, Haas BJ, Goldberg SMD, et al. Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology[J]. BMC Genomics, 2006, 7(1): 272.

[76] Emrich SJ, Barbazuk WB, Li L, et al. Gene discovery and annotation using LCM-454 transcriptome sequencing[J]. Genome Research, 2007, 17(1): 69−73.

[77] Poretsky RS, Bano N, Buchan A, et al. Analysis of microbial gene transcripts in environmental samples[J]. Applied and Environmental Microbiology, 2005, 71(7): 4121−4126.

[78] Frias-Lopez J, Shi YM, Tyson GW, et al. Microbial community gene expression in ocean surface waters[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(10): 3805−3810.

[79] Bailly J, Fraissinet-Tachet L, Verner MC, et al. Soil eukaryotic functional diversity, a metatranscriptomic approach[J]. The ISME Journal, 2007, 1(7): 632−642.

[80] Shi YM, Tyson GW, DeLong EF. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column[J]. Nature, 2009, 459(7244): 266−269.

[81] Poretsky RS, Hewson I, Sun SL, et al. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre[J]. Environmental Microbiology, 2009, 11(6): 1358−1375.

[82] Damon C, Lehembre F, Oger-Desfeux C, et al. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils[J]. PLoS One, 2012, 7(1): e28967.

[83] Damon C, Vallon L, Zimmermann S, et al. A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes[J]. The ISME Journal, 2011, 5(12): 1871−1880.

[84] Armengaud J. A perfect genome annotation is within reach with the proteomics and genomics alliance[J]. Current Opinion in Microbiology, 2009, 12(3): 292−300.

[85] Wilkins MJ, VerBerkmoes NC, Williams KH, et al. Proteogenomic monitoring of *Geobacter* physiology during stimulated uranium bioremediation[J]. Applied and Environmental Microbiology, 2009, 75(20): 6591−6599.

[86] Denef VJ, Kalnejais LH, Mueller RS, et al. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(6): 2383−2390.

[87] Ram RJ, VerBerkmoes NC, Thelen MP, et al. Community proteomics of a natural microbial biofilm[J]. Science, 2005, 308(5730): 1915−1920.

[88] Lo I, Denef VJ, VerBerkmoes NC, et al. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria[J]. Nature, 2007, 446(7135): 537−541.

[89] Denef VJ, VerBerkmoes NC, Shah MB, et al. Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation[J]. Environmental Microbiology, 2008, 11(2): 313−325.

[90] Goltsman DSA, Denef VJ, Singer SW, et al. Community genomic and proteomic analyses of chemoautotrophic iron-oxidizing "*Leptospirillum rubarum*" (group II) and "Leptospirillum ferrodiazotrophum"(group III) bacteria in acid mine drainage biofilms[J]. Applied and Environmental Microbiology, 2009, 75(3): 4599−4615.

[91] Delmotte N, Knief C, Chaffron S, et al. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria[J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(38): 16428−16433.

[92] Toyoda A, Iio W, Mitsumori M, et al. Isolation and identification of cellulose-binding proteins from sheep rumen contents[J]. Applied and Environmental Microbiology, 2009, 75(6): 1667−1673.

[93] Wilmes P, Bond PL. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms[J]. Environmental Microbiology, 2004, 6(9): 911−920.

[94] Martín HG, Ivanova N, Kunin V, et al. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities[J]. Nature Biotechnology, 2006, 24(10): 1263−1269.

[95] Schulze WX, Gleixner G, Kaiser K, et al. A proteomic fingerprint of dissolved organic carbon and of soil particles[J]. Oecologia, 2005, 142(3): 335−343.

[96] Benndorf D, Balcke GU, Harms H, et al. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater[J]. The ISME Journal, 2007, 1(3): 224−234.

[97] Benndorf D, Vogt C, Jehmlich N, et al. Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments[J]. Biodegradation, 2009, 20(6): 737−750.

[98] Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum[J]. PLoS Genetics, 2008, 4(11): e1000282.

[99] Roessner U, Beckles DM. Metabolite measurements[A] // Schwender J. Plant Metabolic Networks[M]. New York: Springer, 2009: 39−69.

[100] Jacobs A, Lunde C, Bacic A, et al. The impact of

constitutive heterologous expression of a moss $Na^+$ transporter on the metabolomes of rice and barley[J]. Metabolomics, 2007, 3(3): 307−317.

[101] Dunn WB, Ellis DI. Metabolomics: current analytical platforms and methodologies[J]. TrAC Trends in Analytical Chemistry, 2005, 24(4): 285−294.

[102] Jansson J, Willing B, Lucio M, et al. Metabolomics reveals metabolic biomarkers of Crohn's disease[J]. PLoS One, 2009, 4(7): e6386.

[103] Wikoff WR, Anfora AT, Liu J, et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites[J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(10): 3698−3703.

[104] Ponnusamy K, Choi JN, Kim J, et al. Microbial community and metabolomic comparison of irritable bowel syndrome faeces[J]. Journal of Medical Microbiology, 2011, 60(6): 817−827.

[105] Sundquist A, Ronaghi M, Tang HX, et al. Whole-genome sequencing and assembly with high-throughput, short-read technologies[J]. PLoS One, 2007, 2(5): e484.

[106] Warren RL, Sutton GG, Jones SJM, et al. Assembling millions of short DNA sequences using SSAKE[J]. Bioinformatics, 2007, 23(4): 500−501.

[107] Butler J, MacCallum I, Kleber M, et al. ALLPATHS: De novo assembly of whole-genome shotgun microreads[J]. Genome Research, 2008, 18(5): 810−820.

[108] Overbeek R, Begley T, Butler RM, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes[J]. Nucleic Acids Research, 2005, 33(17): 5691−5702.

[109] Huson DH, Auch AF, Qi J, et al. MEGAN analysis of metagenomic data[J]. Genome Research, 2007, 17(3): 377−386.

[110] Schreiber F, Gumrich P, Daniel R, et al. Treephyler: fast taxonomic profiling of metagenomes[J]. Bioinformatics, 2010, 26(7): 960−961.

[111] Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool[J]. Journal of Molecular Biology, 1990, 215(3): 403−410.

[112] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice[J]. Nucleic Acids Research, 1994, 22(22): 4673−4680.

[113] Edgar RC, Batzoglou S. Multiple sequence alignment[J]. Current Opinion in Structural Biology, 2006, 16(3): 368−373.

[114] Li RQ, Li YR, Kristiansen K, et al. SOAP: short oligonucleotide alignment program[J]. Bioinformatics, 2008, 24(5): 713−714.

[115] Ning ZM, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases[J]. Genome Research, 2001, 11(10): 1725−1729.

[116] Fuhrman JA. Microbial community structure and its functional implications[J]. Nature, 2009, 459(7244): 193−199.

[117] Zhou JZ, Deng Y, Luo F, et al. Functional molecular ecological networks[J]. MBio, 2010, 1(4): e00169-10.

[118] Zhou JZ, Deng Y, Luo F, et al. Phylogenetic molecular ecological network of soil microbial communities in response to elevated $CO_2$[J]. MBio, 2011a, 2(4): e00122-11.

[119] Deng Y, Jiang YH, Yang YF, et al. Molecular ecological network analyses[J]. BMC Bioinformatics, 2012, 13(1): 113.

[120] Gilbert JA, Steele JA, Caporaso JG, et al. Defining seasonal marine microbial community dynamics[J]. The ISME Journal, 2011, 6(2): 298−308.

[121] Steele JA, Countway PD, Xia L, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages[J]. The ISME Journal, 2011, 5(9): 1414−1425.

[122] Meyer F, Paarmann D, D'souza M, et al. The metagenomics RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes[J]. BMC Bioinformatics, 2008, 9(1): 386.

[123] Dinsdale EA, Edwards RA, Hall D, et al. Functional metagenomic profiling of nine biomes[J]. Nature, 2008, 452(7187): 629−632.

[124] Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data[J]. Nature Genetics, 2001, 29(4): 365−372.