Minireview

Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities

Hikaru Suenaga*
Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, 1-1-1 Higashi, Tsukuba 305-8566, Japan.

Summary
A major research goal in microbial ecology is to understand the relationship between gene organization and function involved in environmental processes of potential interest. Given that more than an estimated 99% of microorganisms in most environments are not amenable to culturing, methods for culture-independent studies of genes of interest have been developed. The wealth of metagenomic approaches allows environmental microbiologists to directly explore the enormous genetic diversity of microbial communities. However, it is extremely difficult to obtain the appropriate sequencing depth of any particular gene that can entirely represent the complexity of microbial metagenomes and be able to draw meaningful conclusions about these communities. This review presents a summary of the metagenomic approaches that have been useful for collecting more information about specific genes. Specific subsets of metagenomes that focus on sequence analysis were selected in each metagenomic studies. This ‘targeted metagenomics’ approach will provide extensive insight into the functional, ecological and evolutionary patterns of important genes found in microorganisms from various ecosystems.

Introduction
An understanding of how microbes and/or their genes interact in various environmental niches represents a major challenge for the environmental microbiologist. However, in many environments, more than 99% of the microorganisms cannot be cultured with readily available technologies. The list of microorganisms in culture collections may skew our view of microbial diversity in nature. Even with the recent success of novel culture methods (Sait et al., 2002; Zengler et al., 2002; Stevenson et al., 2004; Binnerup et al., 2008; Nichols et al., 2010), it is still difficult to provide the appropriate conditions to induce growth of most relevant microbes.

The development of culture-independent techniques, which bypass the need for isolation and laboratory cultivation of individual species, has fundamentally changed studies in environmental microbiology. This discipline has its roots in the analysis of 16S rRNA genes collected from the environment (Olsen et al., 1986). Instead of cataloguing only rRNAs, these techniques are now capable of handling genomic DNA retrieved from naturally occurring microbial communities and is referred to as metagenomics (Handelsman et al., 1998).

Metagenomics technology, which allows us to obtain various gene or pathway components from the community as a whole, has led to the accumulating number of DNA data sets. These extended genome sequences are currently being exploited for novel biotechnological applications (Streit and Schmitz, 2004) and for increasing our knowledge of microbial ecology (Tyson et al., 2004). Metagenomics approaches have been applied to understand the structure (gene/species richness and distribution) and the functional (metabolic) potential of environmental microbial communities.

Based on different screening methods, metagenomic studies are grouped into three classes: (i) shotgun analysis using mass genome sequencing, (ii) activity-driven studies that are designed to search for specific microbial functions and (iii) sequence-driven studies that link genome information with phylogenetic or functional marker genes of interest (Riesenfeld et al., 2004). More recently, next-generation sequencing technologies (Shendure and Ji, 2008; Harismendy et al., 2009) have been...
developed to directly determine the whole collection of genes within an environmental sample without constructing a metagenomic library (Edwards et al., 2006; Tumbaugh et al., 2006; Cox-Foster et al., 2007; Biddle et al., 2008; Brulc et al., 2009; Palenik et al., 2009).

The four metagenomics approaches described above can be characterized as unselective (shotgun analysis and next-generation sequencing) and targeted (activity-driven and sequence-driven studies) metagenomics based on their random and directed sequencing strategies respectively. Unselective metagenomics is becoming an increasingly common strategy due to its simplicity and cost-effectiveness in DNA sequencing. Many projects of random sequencing of microbial communities, such as bacteria, archaea and viruses, have been reported (Chen and Pachter, 2005).

Here, we focus on ‘targeted metagenomics’ studies, which combine metagenomic library screening and subsequent sequencing analysis. The approach is a more effective means to understanding the content and composition of genes for key ecological processes in microbial communities.

‘Excessive’ advances in DNA sequencing technology

The use of shotgun sequencing technology for understanding microbial species and their role in ecosystems was first reported by Tyson and colleagues (2004) using a low-complexity environment. The authors reported that two bacterial genera (Leptospirillum, Sulfobacillus) and one archaeon (Ferroplasma) dominated the biofilms present in an acidic mine drainage habitat based on 16S rDNA sequence analyses. A shotgun sequence experiment was then performed. Nearly complete genomes of the two dominant species and the partial genomes of three less dominant species were reconstructed. As a result, microbial metabolic pathways were discovered in this particular ecosystem. Thus, the detection of metabolic routes via metagenomics opened the door to post-metagenomic studies that gain further insight into genetic networks and metabolic pathways in environmental microbes.

Subsequently, several whole-community shotgun sequencing studies were performed (DeLong et al., 2006; Gill et al., 2006; Wegley et al., 2007; Gilbert et al., 2008). These studies provided much useful information about potential metabolic pathways of uncultivated environmental microorganisms. At the same time, the ability to reconstruct metagenomic DNA fragments and determine metabolic routes decreased dramatically as the genomic complexity increased. In the microbial genomic diversity assay of the Sargasso Sea, Venter and colleagues (2004) determined the nucleotide sequences of over 1 billion base pairs. Their results showed that large fragments could only be assembled for the most dominant community members and the majority of the fragments obtained through shotgun sequencing could not be assigned to any specific microorganism. In predicting the function of more diverse ecosystems, the power of the environmental shotgun sequencing approach is rather limited because of inherent problems associated with limited accessibility to the genomes of less abundant members of the community. A ‘rare biosphere’, such as Methylphaga-like methyloths, reported to play a significant role in methanol metabolism in marine environments, could not be captured by such an approach (Neufeld et al., 2008).

The recent development of second-generation sequence technologies has enabled us to obtain much more DNA information from highly complex microbial communities (Mardis, 2008). The analysis of longer contiguous sequences allows us to identify not only open reading frames, but also operons. Longer gene units, such as mobile genetic elements, are evident only when large genomic fractions can be assembled. However, short sequence reads (varying between 20 and 700 bp depending on the sequencing technology used) that are dissociated from their original species can be joined to lengths usually not exceeding 5000 bp due to the large phylogenetic complexity. Consequently, the reconstruction of a whole genome is not possible (Wooley et al., 2010). Even if reconstruction was successful, identifying an entire transcriptional unit can be problematic (Wommack et al., 2008; Hoff, 2009). Thus, these fragmented and lower-quality sequences lack sufficient information to understand their function and ecological relevance. Furthermore, another growing problem is the handling and management of the massive amount of metagenomic sequence data generated by these technologies, which requires computational infrastructure and accessible tools for storage and further analysis.

Targeted metagenomics strategies

Two different metagenomics strategies are possible: unselective and targeted metagenomics (Fig. 1). Random sequencing of an environmental DNA pool using high-throughput sequencing technology is becoming increasingly common. However, as described above, methods for generating and interpreting metagenomics data remain in the early stages of development. In a targeted metagenomics approach, a deliberately selected DNA pool is subjected to sequencing to reduce genetic complexity. The selection process is usually based on (i) sequence-driven screening or (ii) function-driven screening. By focusing efforts on sequence analysis, targeted metagenomics can provide broad coverage and extensive redundancy of sequences for targeted genes and reveal

specific genome areas directly linked to an ecological function, even at low abundances within a metagenome (Table 1). Better sequence coverage of the obtained target metagenomics can be beneficial for genome assembly and subsequent data analysis. Classified studies on targeted metagenomics are summarized below.

**Targeted metagenomics based on sequence-driven screening**

Using the 16S ribosomal RNA gene (16S rDNA) as a phylogenetic marker gene is one of the most common methods of identifying genome fragments derived from specific groups of microorganisms that have not yet been cultured or that play an important role in the environment (Acinas *et al.*, 2004).

Many Antarctic bacteria have adapted to cold conditions (Fuhrman and Azam, 1980; Saunders *et al.*, 2003; Methe *et al.*, 2005). The hypothesis of cold adaptation derived from diverse bacteria was tested at the amino acid level by comparative genomics of metagenomic DNA fragments (Grzymski *et al.*, 2006). A library generated from DNA obtained from Antarctic coastal water bacterioplankton contained at least 105 bacterial rRNA genes. Among these clones, six fosmid clones carrying various 16S rDNAs, which were ecologically relevant and/or phylogenetically unique, were selected and completely sequenced. Amino acid usage patterns and their relevance to cold adaptation were analysed in the six environmental metagenome fragments affiliated with currently uncultivated Antarctic marine bacterioplankton species. The most significant Antarctic bacterial protein sequences included reduction of salt bridge-forming residues and reduction of stabilizing hydrophobic clusters. These characteristics were not specific to any one phylum, COG role category or G + C content and indicate underlying genomic and biochemical adaptations to cold conditions.

The overall ecological functions of *Planctomyces* have not been studied in depth despite their wide distribution in marine environments (Llobet-Brossa *et al.*, 1998; Rusch *et al.*, 2003). Given that the group is diverse and its members are often difficult to culture, a deep understanding of the ecological function of this group would only be accessible by metagenomic approaches (Woebken *et al.*, 2007). A metagenomic library from marine upwelling systems was screened for fosmids containing planctomycete 16S rRNA genes by polymerase chain reaction (PCR). Six clones thus obtained were sequenced and compared with all available planctomycete genome sequences. The results confirmed that sulfatases were of particular importance for the environmental niche of many marine planctomycetes and are likely to be of general importance for carbon recycling from complex sulfated heteropolysaccharides in marine habitats. Furthermore, single-carbon (C1) metabolism genes were found on some of the fosmids and in all planctomycete genomes except the anammox (anaerobic ammonia oxidation) bacterium *Candidatus Kuenenia stuttgartiensis*, suggesting a general relevance of these genes for *Planctomyces*. The notable lack of these genes only within the anammox *Planctomyces* suggests as yet uncharacterized distinctiveness of these organisms from all other reported planctomycete genera.

Nesbø and colleagues (2005) constructed an rDNA-gene-enriched metagenomic library using a vector con-
### Table 1. Comparison of unselective and targeted metagenomics.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Marker gene/phenotype</th>
<th>Screening method</th>
<th>Sequence size (Mb)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unselective metagenomics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acid mine drainage</td>
<td></td>
<td>–</td>
<td>76</td>
<td>Tyson et al. (2004)</td>
</tr>
<tr>
<td>Sargasso sea</td>
<td></td>
<td>–</td>
<td>1625</td>
<td>Venter et al. (2004)</td>
</tr>
<tr>
<td>North Pacific Subtropical Gyre</td>
<td></td>
<td>–</td>
<td>64</td>
<td>DeLong et al. (2006)</td>
</tr>
<tr>
<td>Human distal gut</td>
<td></td>
<td>–</td>
<td>78</td>
<td>Gill et al. (2006)</td>
</tr>
<tr>
<td>Coral</td>
<td></td>
<td>–</td>
<td>32</td>
<td>Wegley et al. (2007)</td>
</tr>
<tr>
<td>Controlled coastal ocean mesocosm</td>
<td></td>
<td>–</td>
<td>323</td>
<td>Gilbert et al. (2008)</td>
</tr>
<tr>
<td><strong>Targeted metagenomics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sequence-driven screening</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antarctic coastal water</td>
<td>16S rDNA</td>
<td>PCR</td>
<td>0.260</td>
<td>Grymski et al. (2006)</td>
</tr>
<tr>
<td>Coast and shelf water</td>
<td>16S rDNA of planctomycete</td>
<td>PCR</td>
<td>0.233</td>
<td>Woebken et al. (2007)</td>
</tr>
<tr>
<td>Harbour sediment</td>
<td>I-Ceul site (23S rRNA gene)</td>
<td>–</td>
<td>0.438</td>
<td>Nesbo et al. (2005)</td>
</tr>
<tr>
<td>Marine sediment</td>
<td>dsrAB, aprAB</td>
<td>PCR</td>
<td>0.111</td>
<td>Mussmann et al. (2005)</td>
</tr>
<tr>
<td>Grassland soil</td>
<td>nirS, nirK, nirZ</td>
<td>Colony hybridization</td>
<td>0.218</td>
<td>Demanèche et al. (2009)</td>
</tr>
<tr>
<td><strong>Function-driven screening</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faecal from organically reared pig</td>
<td>Tetracycline resistance</td>
<td>Growth assay</td>
<td>0.051</td>
<td>Kazimierczak et al. (2009)</td>
</tr>
<tr>
<td>Activated sludge</td>
<td>Bleomycin resistance</td>
<td>Growth assay</td>
<td>0.099</td>
<td>Mori et al. (2008)</td>
</tr>
<tr>
<td>Cow rumen</td>
<td>Hydrolytic activity</td>
<td>Enzyme activity assay</td>
<td>0.772</td>
<td>Ferrer et al. (2005)</td>
</tr>
<tr>
<td>Activated sludge</td>
<td>Extradiol dioxygenase</td>
<td>Enzyme activity assay</td>
<td>1.37</td>
<td>Suenaga et al. (2009a)</td>
</tr>
<tr>
<td><strong>Other methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest soil</td>
<td>Incorporation of the $^{13}$CH&lt;sub&gt;4&lt;/sub&gt; pmoA, nmoX, mxaF</td>
<td>SIP</td>
<td>0.015</td>
<td>Dumont et al. (2006)</td>
</tr>
<tr>
<td>Lake sediment</td>
<td>Incorporation of labelled C&lt;sub&gt;12&lt;/sub&gt; compounds</td>
<td>SIP</td>
<td>255</td>
<td>Kalyuzhnaya et al. (2008)</td>
</tr>
<tr>
<td>Freshwater and marine sediment</td>
<td>Magnetotactic bacteria</td>
<td>Magnetic collection</td>
<td>0.521</td>
<td>Jogler et al. (2009)</td>
</tr>
</tbody>
</table>
taining an I-Ceu site (Marshall and Lemieux, 1992), which exists at position 1923 in the 23S rRNA gene of most bacteria, and specifically cloned DNA fragments containing the 23S rRNA gene. Phylogenetic analysis of the protein-coding sequences of 12 fosmids indicated that a significant fraction of the genes was acquired by lateral gene transfer. In several cases, co-transfer of functionally related genes can be inferred.

The 16S rDNA sequences reflect only the phylogenetic classification of host bacteria and not necessarily the metabolic function of these organisms (Jaspers and Overmann, 2004). For example, *Escherichia coli* K-12 and O157 : H7 (Perna et al., 2001) or strains of *Bacillus anthracis* and *Bacillus cereus* (Ash et al., 1991), which clearly exhibited different ecotypes based on their virulence properties, showed identical 16S rRNA gene sequences. Therefore, other genetic markers that are associated with metabolic function are needed.

Dissimilatory sulfate reduction is a key process in the mineralization of organic matter in marine sediments (Shen et al., 2001). Up to 50% of organic carbon in coastal sediments is mineralized anaerobically by sulfate-reducing prokaryotes (Jørgensen, 1982). From marine sediments, Mussmann and colleagues (2005) screened three fosmid DNA fragments harbouring a core set of essential genes for dissimilatory sulfate reduction, including genes associated with the reduction of sulfur intermediates (*dsrAB* gene) and the synthesis of the prosthetic group of the dissimilatory sulfate reductase (*aprA* gene). Sequence analysis of all fosmid inserts revealed the genomic context of the key enzymes of dissimilatory sulfate reduction as well as novel genes functionally involved in sulfate respiration in their franking regions. Based on the results of the comparative genome analyses, a more comprehensive model of dissimilatory sulfate reduction was presented. The results support the hypothesis that the set of genes responsible for dissimilatory sulfate reduction were concomitantly transferred in a single event among prokaryotes.

A metagenomics approach combined with molecular screening (colony hybridization) and pyrosequencing was performed to gain insight into the genetic organization and diversity of gene clusters or operons for denitrification (Ginolhac et al., 2004; Demanèche et al., 2009). Denitrification is a microbial respiratory process within the nitrogen cycle responsible for the return of fixed nitrogen to the atmosphere. This targeted metagenomics study indicated that the gene clusters involved in denitrification were probably subject to shuffling by endogenous gene displacement or by horizontal gene transfer between bacteria.

The advantage of using sequence-driven screening is that it is not dependent on expression of cloned genes in foreign hosts. Additionally, well-established techniques, such as PCR or hybridization, can be employed for different targets when using these metagenomic approaches. On the other hand, the sequence-driven approach involves designing DNA probes and primers derived from conserved regions of known gene or protein families. Various software (e.g. Edgar, 2004; Ludwig et al., 2004; Zhang et al., 2007) and databases (e.g. FunGene, http://fungene.cme.msu.edu/index.spr) can help to develop effective screening strategies.

**Targeted metagenomics based on functional-driven screening**

As an alternative to phylogenetic markers, the use of known gene functions of interest is a more direct route to the discovery of gene clusters with related metabolic roles in microbial communities. Furthermore, function-driven screening strategies can potentially provide a means to reveal undiscovered genes or gene families that cannot be detected by sequence-driven approaches, although specific screening systems are often required.

Simple and certain activity-based strategies, such as colony growth in the presence of antibiotics on agar media, have often been used. Ten clones resistant to tetracycline were identified from a bacterial artificial chromosome (BAC) library (Kazimierczak et al., 2009) and three clones resistant to bleomycin were identified from a fosmid library (Mori et al., 2008). Both metagenomic libraries were constructed using DNA collected from antibiotic-free environments (i.e. faecal samples of organically reared adult pigs and activated sludge treating coke-plant wastewater respectively). Most of the genes resided on putative mobile genetic elements, which presumably contribute to the maintenance and dissemination of antibiotic resistance in even antibiotic-free environments.

The diversity of the major rumen enzymes has been characterized by an activity-based metagenomics approach (Ferrer et al., 2005). The rumen habitat consists mostly of obligate anaerobic microorganisms, including fungi, protozoa, bacteria and archaea. Bacterial and fungal components of this ecosystem are responsible for the rapid degradation of plant polymers (Tajima et al., 1999; Ramsak et al., 2000). A metagenomic phage library from the rumen content of a dairy cow was screened for hydrolase activity directly on agar medium. In total, 22 clones with distinct hydrolytic activities were identified and characterized. Among them, four hydrolases exhibited no sequence similarity to enzymes from public databases and four clones did not match any putative catalytic residues involved in the catalytic function of esterase and cellulose-like enzymes. These results suggest that function-based screening of metagenomic libraries facilitates the functional
assignment of many properties that have been classified as hypothetical proteins.

In many cases, prokaryotic phenotypes are the result of the concerted expression of many genes that are often arranged into adjacent operons or super-operonic clusters. Thus, the entire gene set of a pathway of interest can be captured by targeting the central or essential gene in that metabolic pathway.

The ability to use various aromatic compounds as sources of carbon and energy is widespread in bacteria. In the natural environment, these bacteria contribute greatly to the breakdown of aromatic compounds and to the global carbon cycle (Esteve-Núñez et al., 2001; Furukawa et al., 2004). Among the enzymes that degrade aromatic compounds, extradiol dioxygenases (EDOs) catalyse ring cleavage of catecholic compounds to produce yellow meta-cleavage compounds (Elis and Bolin, 1996). Metagenomic DNA extracted from activated sludge for industrial wastewater was cloned into fosmids (Suñana et al., 2007). The resulting E. coli library was screened for EDO activity using catechol as a substrate and 38 clones were subjected to sequence analysis (Suñana et al., 2009a,b). As a result, various types of gene subsets were identified that were not similar to previously reported pathways that complete degradation. The distribution of these genes among the different genome segments was reported in some isolated sphingomonads (Miller et al., 2010) or rhodococcal (McLeod et al., 2006) strains. The metagenomic data revealed that the complete degradation pathways identified in many isolated bacteria were, in fact, rare. Additionally, through in silico assembly of the inserts from several fosmid clones, a meta-plasmid designated as pSKYE1 was reconstructed. Plasmid pSKYE1 seems to act as a ‘detoxification apparatus’ that may confer survival capabilities to its bacterial host in the microbial community. Thus, targeted metagenomic approaches based on functional-driven screening can be used to obtain novel findings of targeted biological functions.

Other targeted metagenomics studies

The pre-treatment of a collected environmental sample before DNA extraction for the enrichment of target genes is also effective in targeted metagenomics. Stable-isotope probing (SIP) allows the simultaneous recovery of labelled DNA from active target microorganisms that play a particular function in the environment (Radajewski et al., 2000; Wellington et al., 2003). These collected DNAs have often been directly subjected to PCR amplification of rRNA genes to taxonomically identify microorganisms that feed on the labelled substrate. Recently, the combination of SIP with shotgun sequencing was used to determine a complete pmoCAB operon encoding the three subunits of a particular methane monooxygenase (Stolyar et al., 1999). PmoCAB is a key enzyme involved in the methane utilization pathway of aerobic methanotrophs, a subgroup of the methylo trophs (Dumont et al., 2006). A metagenomic BAC clone library was constructed with 13C-DNA from a 13CH4-labelled forest soil sample. After colony hybridization and subsequent sequence determination, one clone was found to contain the pmoCAB operon.

Although methanotrophs were detected in this study, they usually are not the dominant group of microorganisms in soils and would have been missed by a random metagenome shotgun method. More recently, the nearly complete genome of a novel methylothroph Methylotenera mobilis, which comprises less than 0.4% of the total bacterial population in the sampled environment, was determined by using SIP followed by shotgun sequencing (Kalyuzhnaya et al., 2008). Methylotrophy, the metabolism of organic compounds containing no carbon–carbon bonds (i.e. C1 compounds) such as methane, methanol and methylated amines, is an important component of the global carbon cycle (Hanson and Hanson, 1996). Genome analysis of methylotrophs will expand our current knowledge of the microbial metabolisms of the C1 compound in environment.

An enrichment procedure was also performed for the mass collection of environmental magnetotactic bacteria (MTB) virtually free of non-magnetic contaminants (Jogler et al., 2009). MTB synthesize magnetosomes, which are membrane-enclosed organelles consisting of magnetite (Fe3O4) crystals or, less commonly, greigite (Fe3S4) (Bazylinski and Frankel, 2004). MTB do not form a coherent phylogenetic group, but the trait of magnetotaxis is found in species within different phylogenetic clades, which are mostly unavailable in pure culture (Faivre and Schuler, 2008). The selective cloning of larger DNA fragments from magnetotactic metagenomes was achieved by a two-step magnetic enrichment. Sequence analysis of fosmid clones revealed that uncultivated MTB exhibited similar, yet different, organizations of the magnetosome island, which may account for the diversity in biomineralization and magnetotaxis observed in MTB from various environments.

Targeted metagenomics analyses based on pre-treatment for enrichment of specific microorganisms or genes can reveal relevant genome areas directly linked to an ecological function even in low abundance within a metagenome.

Future aspects

The accumulation of data sets of microbial genes and genomes in metagenomics is growing rapidly, but our understanding of their functional and ecological relevance is proceeding more slowly. Sequencing technology is
accelerating and third-generation sequencers that will enable reading of the DNA sequence of a single chromosome in a single pass with few or no fragments will likely be established in the near future (Clarke et al., 2009; Eid et al., 2009). One of the most serious problems with metagenomics methods using second-generation sequencing technology related to fragment assembly may be solved by using these long-read sequencing technologies.

However, the other serious problem related to gene prediction and annotation remains unsolved. Current functional assignment for genes from metagenomes is based on homology searches using BLAST tools (Altschul et al., 1990) that depend heavily on the quality and completeness of current databases. Homology-based methods are effective only when the information of the reference sequences is accurate. Actually, a number of genomes in the current databases contain misannotations (Schnoes et al., 2009). Furthermore, this method is not sensitive, particularly for genomes that lack sequence relatives, and can miss novel genes that may potentially be the most interesting. A review of prokaryotic protein diversity in different shotgun metagenome studies indicated that 30–60% of the proteins cannot be assigned to known functions using current public databases (Vieites et al., 2009). Many hypothetical proteins that have been identified may be ecologically important. Targeted metagenomics, based on function-driven screening, can reveal specific genome sequences directly linking an ecological function although the ORF lacks sequence homology to known genes and is identified as a hypothetical protein in current databases. Furthermore, the functions of genes located in the neighbourhood of the target genes would also be revealed by complete sequence analysis of the inserts. This was supported by the observation that intergenic distances tend to be shorter between genes of the same operon than between operons (Salgado et al., 2000) and neighbouring ORFs are more likely to be functionally associated (Overbeek et al., 1999; Korbel et al. 2004). Time series studies at one site are also being used to investigate how microorganisms and their activities co-vary with environmental changes. The transition of metadata, which contains various environmental measurements as well as metagenomic sequence data, will help to predict the key genes for crucial environmental processes and to understand the variability of microbial genomes in the process of adaptive evolution.

Sequence data of metagenomics tell us ‘what is there’ and ‘what it is capable of doing’. ‘What it is actually doing’ is revealed by evaluating transcription (mRNA) and translation (protein) data. The correlation between specific gene expression patterns and changes in community composition or the functional property of the microbial ecosystem can be addressed more directly by applying these omics approaches (Benndorf et al., 2007; Maron et al., 2007; Shrestha et al., 2009). The combined analysis of targeted metagenomics with targeted metatranscriptomics or metaproteomics has the potential to provide a novel perspective on microbial community dynamics.

In this review, the targeted metagenomics approaches based on the screening step are mainly presented. Selective DNA collection from the environment also seems to be efficient for targeted metagenomics approaches. For example, plasmids represent a comparatively uncharacterized metagenomic space and they frequently carry useful genes for their host such as those involved in biodegradation and metal or antibiotic resistance. Total plasmid DNA can be targeted as a ‘metamobilome’ study. In the near future, it will be possible to develop targeted metagenomics methodologies with improved experimental techniques such as environmental sampling, DNA extraction, cloning as well as library screening.

References


