When metagenomics meets stable-isotope probing: progress and perspectives

Yin Chen and J. Colin Murrell

Department of Biological Sciences, the University of Warwick, Coventry, CV4 7AL, UK

The application of metagenomics, the culture-independent capture and subsequent analysis of genomic DNA from the environment, has greatly expanded our knowledge of the diversity of microbes and microbial protein families; however, the metabolic functions of many microorganisms remain largely unknown. DNA stable-isotope probing (DNA-SIP) is a recently developed method in which the incorporation of stable isotope from a labelled substrate is used to identify the function of microorganisms in the environment. The technique has now been used in conjunction with metagenomics to establish links between microbial identity and particular metabolic functions. The combination of DNA-SIP and metagenomics not only permits the detection of rare low-abundance species from metagenomic libraries but also facilitates the detection of novel enzymes and bioactive compounds.

From 16S rRNA genes to whole community sequence analyses of environmental microorganisms

One of the major breakthroughs in environmental microbiology in the past few decades is the application of culture-independent techniques to the study of microbial ecology and the distribution of microbes in the environment [1–4]. Because previously only cultivated microorganisms could be studied, the development of culture-independent techniques has fundamentally changed our way of studying microorganisms. Early work focused on the analyses of 16S rRNA genes retrieved from the environment [5] and later on the retrieval of larger DNA fragments and whole community sequences from environmental DNA, now referred to as metagenomics [6]. Metagenomic approaches have now been applied to a variety of environments, from the human gut microbiome to soils [7–9], and from the deep-sea to the indoor atmosphere [10,11]. The aims of these studies were (i) to identify the diverse microorganisms inhabiting various environments and (ii) to reconstruct metabolic pathways, thus helping to predict environmental functions of uncultivated microorganisms. Research in metagenomics is also driven by industry in the quest to retrieve novel enzymes and bioactive compounds from the environment. Diverse enzymes and secondary metabolites have been identified as arising from uncultivated microorganisms in the environment [12]. Conventional metagenomics uses DNA directly extracted from environmental samples; more recently DNA stable-isotope probing (DNA–SIP) has been used as a filter to select DNA from relevant microorganisms [13,14]. DNA–SIP was originally developed as a method to investigate the metabolic functions of microorganisms in the environment. The technique relies on the incorporation of stable-isotope labelled compounds (e.g. 13C, 15N) into microbial DNA [15] during microcosm or in situ incubation of such compounds with environmental samples (Box 1). Metagenomics combining with DNA–SIP has developed from concept to reality, and is now attracting significant attention. We summarise recent progress in SIP-metagenomic techniques and applications and discuss future prospects for this combined approach in environmental microbiology and biotechnology.

Metagenomics: success and frustration

Key steps in metagenomic studies are summarised in Figure 1. DNA is extracted directly from environmental samples and then either is used to prepare a metagenomic library by cloning DNA fragments into an appropriate vector (plasmid, cosmid, fosmid, etc.), or the DNA is directly used for high-throughput sequencing, for example using 454 pyrosequencing technology [16]. The metagenome library can be screened for genes of interest by function-based approaches to look for key metabolic functions. This depends on the expression of heterologous genes in a host such as Escherichia coli (e.g. to screen for lipase activity by the detection of clear zones around colonies on tributyrin agar plates [17]). Metagenome libraries can also be screened for target genes by sequence-based approaches (such as PCR using degenerate primers or by colony blotting and screening with radioactive probes) (reviewed in Ref. [18]).

The power of the sequence-based approach for understanding the role of microorganisms in the environment was first demonstrated in an elegant study carried out by Tyson and colleagues using a less complex environment as a model system [19]. Through the analyses of 16S rRNA gene clones it was found that all biofilms present in acidic mine drainage (Iron Mountain, USA) were dominated by two genera of bacteria (Leptospirillum and Sulfolobus) and one archeon (Ferroplasma). The authors then carried out a shotgun-sequencing experiment, and from 100 Mb of DNA sequence information they were able to reconstruct near-complete genomes of the two dominant organisms comprising these biofilms (Leptospirillum group II and Ferroplasma type II), as well as partial genomes of three...
Review 1. Top ten things you need to know before setting up a DNA-SIP experiment

1. When should I consider using DNA-SIP? DNA-SIP is one of the many methods available to link microbial identity to function (for more detail, see Ref. [27]). It is less sensitive compared to RNA-SIP [50] and phospholipid fatty acid (PLFA)-SIP [51,52]; however, it allows one to gain access to the DNA from the microorganisms of interest.

2. What can DNA-SIP tell me? DNA-SIP will identify active microorganisms that metabolise and incorporate the labelled substrate into their DNA during a given SIP incubation.

3. What substrate concentration should be used and how long should the incubation be? Unfortunately, there is no standard answer. However, typically, we have found that incorporation of ~50 μmol 13C per g (for soil or sediment samples) and ~5 μmol 13C per ml (for water samples) is sufficient to detect active microorganisms above background. It is highly recommended that target substrate utilisation activity for a given environmental sample is measured before setting up SIP incubations. The overall solution will be to consider the balance of the concentration of labelled substrate (i.e. close to in situ concentrations), activity of the samples and incubation time.

4. Do I need a 13C- or 14N-incubation control? This is highly recommended. Fingerprints (see below) between labelled and non-labelled DNA (both of which could be distributed over several gradient fractions) can therefore be compared. This is extremely useful in identifying ‘heavy’ DNA from SIP incubations.

5. Can I use partially labelled substrate? The use of partially-labelled substrate for DNA-SIP is not recommended. This could cause insufficient stable-isotope incorporation into DNA and further complicate subsequent ‘heavy’ DNA identification and isolation; however, recent studies have successfully used partially labelled substrate for DNA-SIP (e.g. [53,54]).

6. Do I need to worry about cross-feeding? Although it could be difficult to completely remove the problem of cross-feeding, SIP incubations need to be optimised to label only primary consumers while still yielding sufficient isotope-labelled DNA. However, in some circumstances, cross-feeding of stable-isotope labeling is desirable to track carbon flow through a food web (e.g. [55-57]).

7. What method should I use to extract DNA from SIP-incubated samples? This depends on later applications using the ‘heavy’ DNA. For example, avoid methods that shear DNA if large DNA fragments are needed. If high concentrations of humic substances are present, extracted DNA could need to be purified before isopycnic ultracentrifugation.

8. How do I know where the ‘heavy’ DNA is after isopycnic centrifugation? There are six methods for identifying ‘heavy’ DNA. The more these are used, the more reliable the data are. One should be aware that a second round of ultracentrifugation with bis-benzimide can be required in some cases because unlabelled DNA with a high G+C content can migrate at the same position as labelled DNA with a relatively low G+C content [58].

9. What if the ‘heavy’ DNA retrieved is not sufficient for analysis? The quantity of ‘heavy’ DNA retrieved ranges from ng to μg. One way to obtain more ‘heavy’ DNA is, of course, to combine ‘heavy’ DNA from several CsCl gradients. However, if a substantial amount of DNA is needed, such as for making a metagenome library, multiple displacement amplification can be performed from the ‘heavy’ DNA using Phi 29 DNA polymerase [25,35].

10. What are the minimum instruments required for carrying out a SIP experiment? Instruments for measuring substrate utilisation activity, an ultracentrifuge for isopycnic centrifugation, and an analytical balance.

other less-dominant microorganisms, thereby piecing together the metabolic routes of this rather simple ecosytem. Inspired by this successful study, a number of whole-community shotgun-sequencing projects were carried out; these have yielded a wealth of information on potential metabolic pathways of uncultivated microorganisms in the environment [7,20,21,22,23]. However, at the same time it was realised that our ability to piece together metagenomic DNA fragments decreases dramatically as the complexity of the environment sampled increases. Aiming at assessing the microbial diversity in Sargasso Sea, Venter and colleagues found it was only possible to assemble large scaffolds of genomic DNA from the most dominant species in the samples, and the majority of sequences obtained through shotgun sequencing could not be assigned to any specific microorganism [24]. Inevitably this community shotgun-sequencing approach is of rather limited use in predicting the function of community members of low abundance, and encounters the following two problems. First, in order to gain insights into the functions of less-abundant community members (often referred to as the ‘rare biosphere’), greater in-depth sequencing needs to be carried out, and this is generally not cost-effective. Second, the so-called ‘rare biosphere’ can be significant players in a particular function (for example, methylophy in the marine water column) that cannot be captured by such conventional approaches [25].

Meanwhile, the application of metagenomics has led to the discovery and characterisation of a wide range of novel biocatalysts and pharmacologically important compounds (reviewed in Ref. [12,26]). A number of small- to medium-scale companies (e.g. Diversa, now known as Verenium;
Metagenomics, together with in vitro evolution and high-throughput screening technologies, provides industry with an unprecedented chance to bring biomolecules into industrial application [26]. However, so far only relatively few enzymes discovered through metagenomics are used in established biotechnological processes [12]. One of the reasons is probably ‘the low frequency of clones of a desired nature’ [18] that can be retrieved from metagenome libraries (Table 1).

Combining DNA–SIP with metagenomics: from concept to reality

The problems that conventional metagenomics have encountered imply that preselection of DNA from the environment would be of considerable benefit for improving the performance of current metagenomic approaches (i.e. to enhance gene detection frequency and to reduce the complexity of the target microbial community). SIP is one of a number of culture-independent methods developed to link the identity of microorganisms in the environment to particular functions (reviewed in Ref. [27]); however, SIP also allows the simultaneous recovery of DNA from active target microorganisms in the environment. Stable-isotope labelled DNA (i.e. ‘heavy’ DNA) can be separated from unlabelled DNA (background community) by density-gradient (isopycnic) ultracentrifugation (Figure 1). Traditionally, the ‘heavy’ DNA is subjected to PCR amplification of ribosomal RNA and functional genes (e.g. pmoa, encoding a key subunit of particulate methane monoxygenase – a functional gene marker for aerobic methane-oxidising bacteria [methanotrophs]) to identify the taxonomic identity of these microorganisms that feed on the labelled substrate (reviewed in Ref. [28]).

The application of metagenomics based on shotgun sequencing in determining microbial diversity is frustrated by the inability to completely resolve more complex communities. It was proposed by Schloss and Handelsman that combining DNA–SIP with metagenomics could help to reduce sample complexity [18], whereby SIP is used as a filter to enrich for the DNA of microorganisms that carry out a particular function. A proof-of-concept study was first carried out by Dumont et al. who were investigating methane-utilising bacteria in a forest soil [13]. Using ‘heavy’ DNA from a 13CH4-labelled forest soil sample, the authors constructed a bacterial artificial chromosome (BAC) library. After screening 2,300 clones (average size ~25 kb), one BAC clone was found to contain a pmocAB operon encoding the three subunits of particulate methane monoxygenase, a key enzyme involved in the methane-utilisation pathway in aerobic methanotrophs. The advantage of using DNA–SIP is highlighted by comparison to a similar non-SIP study [29] where a metagenome library was constructed using DNA directly extracted from a forest soil sample and 250,000 fosmid clones (average size ~40 kb) had to be screened to find one pmocAB operon. Although methanotrophs were detected in this study, they are not usually a dominant group of microorganisms in many soils and would have been missed in other conventional metagenomic studies. The study by Dumont and colleagues demonstrated the feasibility of combining DNA–SIP with metagenomics to determine the function

![Figure 1. Key steps in conventional and DNA–SIP-enabled metagenomics.](http://www.verenium.com)
of less abundant members of the microbial community in this environment [13].

One of the major criticisms of DNA–SIP is that usually relatively high concentrations (compared to in situ concentrations) of $^{13}$C compounds need to be added to SIP incubations in order for the desired amount of stable isotope to be incorporated into DNA over a reasonable timescale, a prerequisite for making a metagenome library from stable-isotope-labelled DNA (i.e. ‘heavy’ DNA). The problem is that isotope label can flow from primary utilisers to secondary consumers by ‘cross-feeding’ [13,30]. In addition, it is documented that some microorganisms can be inhibited by higher substrate concentrations (e.g. Candidatus Nitroso-sphaera gargensis, an ammonia-oxidising archaeon, can be partially inhibited by ammonium at a concentration of 3.08 mM [31]) whereas others can be inactive if substrate concentrations are too low (e.g. a Pseudomonas strain isolated from a polycyclic aromatic hydrocarbon-contaminated groundwater only utilises naphthalene at concentrations higher than 30 μM [32]). Successful labeling with near in situ concentrations of substrate was achieved in a study with the goal of finding methanol-utilising microorganisms in the ocean [25]. Using 1 μM $^{13}$C-methanol it was found that Methylophaga-related microorganisms were metabolising methanol in sea water. However, only nanograms of ‘heavy’ DNA were obtained from the DNA–SIP experiment and multiple displacement amplification (MDA) was necessary to generate sufficient amounts of DNA to make a fosmid library. After screening fosmids by PCR, a novel methanol dehydrogenase cluster was found in the metagenomic library; this was clearly from a bacterium closely related to Methylophaga spp., thus indicating that these bacteria are actively involved in the metabolism of methanol in the marine environment. The MDA method has only recently begun to be used by environmental microbiologists (reviewed in Ref. [33]) and it was noted that chimeras could be introduced by mechanisms that are not fully understood [34]. Another study [35] showed that it was possible to minimise chimera formation by treating MDA-generated DNA with a series of enzymes before making a fosmid library. This technique did not significantly affect the size of MDA-generated DNA. These two studies demonstrated that it is possible to combine DNA–SIP, using near in situ substrate conditions, with metagenomics to fully characterise the function of microorganisms which may well be present in low numbers. Taking the advantage of the high sequencing capacity at the Joint Genome Institute (USA), an elegant study was carried out by Kalyuzhnaya et al. [36], where the authors were interested in the one-carbon (C₁) cycling in sediment from Lake Washington. DNA–SIP experiments were carried out using a number of $^{13}$C₁ compounds and metagenome libraries were made from purified ‘heavy’ DNA. Rather than screening for particular functions, the libraries were subjected to a shotgun-sequencing approach and a near-complete genome of the methylo troph Methylophaca mobilis was reconstructed from the sequences retrieved from the $^{13}$C-DNA metagenome library. This bacterium is clearly a key player in C₁ cycling in the Lake Washington sediment [37]. Furthermore, this study proved the notion that whole-genome shotgun metagenomics can be used to reconstruct genomes and metabolic pathways of ‘rare’ microorganisms if combined with the DNA–SIP approach.

Having seen the inherent problems associated with conventional function-based metagenomics, a number of methods have been evaluated to improve gene detection frequency. One of these is to establish enrichment cultures for microorganisms harbouring the desired functions before DNA extraction is carried out (e.g. Refs. [38,39]). However, this inevitably results in the loss of microbial diversity because a few fast-growing microorganisms can outcompete the rest of the microbial community. Another method that can potentially be applied to enrich for DNA

---

Table 1. Comparison of conventional metagenomics and DNA-SIP-enabled metagenomics

<table>
<thead>
<tr>
<th>Key features</th>
<th>Limitations</th>
<th>Research needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional metagenomics</td>
<td>Heavily biased towards the most abundant species in the environment</td>
<td>Improved bioinformatics for the reconstruction of metabolic pathways from massive gene-sequence datasets and to assist binning of these sequences for improving sequence assembly</td>
</tr>
<tr>
<td></td>
<td>- Allows culture-independent analyses of microbial communities</td>
<td>- Technology development for large-scale sequencing; reducing cost</td>
</tr>
<tr>
<td></td>
<td>- Allows the identification of novel enzymes and bioactive compounds from uncultivated microorganisms in the environment</td>
<td>- Automated screening methods and technologies for function-based screening for novel enzymes and pharmacologically relevant compounds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Better understanding of genes of unknown function in current databases</td>
</tr>
<tr>
<td>DNA–SIP-enabled metagenomics</td>
<td>Isolating sufficient DNA for metagenomic library construction is an issue</td>
<td>Prevent chimaera formation during MDA of $^{13}$C-DNA</td>
</tr>
<tr>
<td></td>
<td>- Can establish a direct link between identity and function</td>
<td>Development of high-throughput methods for DNA–SIP analysis of multiple samples</td>
</tr>
<tr>
<td></td>
<td>- Less abundant ‘rare species’ can be targeted with the same sequencing effort</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Reconstruction of individual genomes is feasible due to reduced complexity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Gene-mining ‘hit-rate’ is improved</td>
<td></td>
</tr>
</tbody>
</table>

Conventional metagenomics

One of the major criticisms of DNA–SIP is that usually relatively high concentrations (compared to in situ concentrations) of $^{13}$C compounds need to be added to SIP incubations in order for the desired amount of stable isotope to be incorporated into DNA over a reasonable timescale, a prerequisite for making a metagenome library from stable-isotope-labelled DNA (i.e. ‘heavy’ DNA). The problem is that isotope label can flow from primary utilisers to secondary consumers by ‘cross-feeding’ [13,30]. In addition, it is documented that some microorganisms can be inhibited by higher substrate concentrations (e.g. Candidatus Nitroso-sphaera gargensis, an ammonia-oxidising archaeon, can be partially inhibited by ammonium at a concentration of 3.08 mM [31]) whereas others can be inactive if substrate concentrations are too low (e.g. a Pseudomonas strain isolated from a polycyclic aromatic hydrocarbon-contaminated groundwater only utilises naphthalene at concentrations higher than 30 μM [32]). Successful labeling with near in situ concentrations of substrate was achieved in a study with the goal of finding methanol-utilising microorganisms in the ocean [25]. Using 1 μM $^{13}$C-methanol it was found that Methylophaga-related microorganisms were metabolising methanol in sea water. However, only nanograms of ‘heavy’ DNA were obtained from the DNA–SIP experiment and multiple displacement amplification (MDA) was necessary to generate sufficient amounts of DNA to make a fosmid library. After screening fosmids by PCR, a novel methanol dehydrogenase cluster was found in the metagenomic library; this was clearly from a bacterium closely related to Methylophaga spp., thus indicating that these bacteria are actively involved in the metabolism of methanol in the marine environment. The MDA method has only recently begun to be used by environmental microbiologists (reviewed in Ref. [33]) and it was noted that chimeras could be introduced by mechanisms that are not fully understood [34]. Another study [35] showed that it was possible to minimise chimera formation by treating MDA-generated DNA with a series of enzymes before making a fosmid library. This technique did not significantly affect the size of MDA-generated DNA. These two studies demonstrated that it is possible to combine DNA–SIP, using near in situ substrate conditions, with metagenomics to fully characterise the function of microorganisms which may well be present in low numbers. Taking the advantage of the high sequencing capacity at the Joint Genome Institute (USA), an elegant study was carried out by Kalyuzhnaya et al. [36], where the authors were interested in the one-carbon (C₁) cycling in sediment from Lake Washington. DNA–SIP experiments were carried out using a number of $^{13}$C₁ compounds and metagenome libraries were made from purified ‘heavy’ DNA. Rather than screening for particular functions, the libraries were subjected to a shotgun-sequencing approach and a near-complete genome of the methylo troph Methylophaca mobilis was reconstructed from the sequences retrieved from the $^{13}$C-DNA metagenome library. This bacterium is clearly a key player in C₁ cycling in the Lake Washington sediment [37]. Furthermore, this study proved the notion that whole-genome shotgun metagenomics can be used to reconstruct genomes and metabolic pathways of ‘rare’ microorganisms if combined with the DNA–SIP approach.

Having seen the inherent problems associated with conventional function-based metagenomics, a number of methods have been evaluated to improve gene detection frequency. One of these is to establish enrichment cultures for microorganisms harbouring the desired functions before DNA extraction is carried out (e.g. Refs. [38,39]). However, this inevitably results in the loss of microbial diversity because a few fast-growing microorganisms can outcompete the rest of the microbial community. Another method that can potentially be applied to enrich for DNA
from organisms of interest involves the use of 5-Bromo-2-deoxyuridine (BrdU) [40,41]. BrdU can be incorporated into newly synthesised DNA by metabolically active community members in response to stimuli (such as substrates) and BrdU-labelled DNA can then be separated from community DNA using anti-BrdU antibodies [40,41]. However, the application of this method in capturing functionally-relevant organisms responding to stimuli is probably heavily biased because two of the four capturing strains tested did not incorporate BrdU into their DNA [40]. The use of ‘heavy’ DNA from SIP experiments in order to improve the detection frequency of genes of interest for biotechnological applications has also been investigated [14]. The authors were screening environmental samples for novel glycerol dehydratases, key enzymes in the production of 1,3-propanediol, an important building block for biopolymer synthesis. It was realised that the chance of finding a positive clone from metagenome libraries using function-based screening was rather low (one positive clone per 1.14 Gbp metagenomic DNA screened) [38]. Therefore, the authors compared two approaches, and used either DNA from an enrichment culture (40 mM of 13C-glycerol) or ‘heavy’ DNA from a stable-isotope labeling experiment using 13C3-glycerol (40 mM) to make a metagenomic library. The gene detection frequency using the ‘heavy’ DNA from SIP was 2—3 times higher than with DNA from an enrichment culture. A similar study was published recently [42]. The authors retrieved biphenyl-dioxygenase sequences from the ‘heavy’ DNA extracted from a 13C-biphenyl-amended microcosm using a sediment sample from a contaminated river. The organisation of biphenyl-dioxygenase genes in the retrieved clone differed from that of the upper bph operons from known biphenyl-degrading microorganisms, thus indicating the novelty of this enzyme. These two studies clearly demonstrate the feasibility and great potential of SIP-enabled metagenomics in the discovery of novel industry-relevant biocatalysts from the environment.

Future perspectives for DNA–SIP-based metagenomic studies

Until now, only a few studies have been published that combine DNA–SIP techniques with metagenomics (Table 2). It is clear from these studies that this approach can have advantages over conventional metagenomics. Although combining DNA–SIP with metagenomics can improve gene detection frequency for enzyme discovery and significantly facilitate the reduction of community complexity for whole-genome shotgun sequencing, these two factors are not the only bottlenecks in current metagenomic studies (Table 1). For example, data analysis of massive metagenome sequences is not trivial [43]. One of the most serious challenges is population heterogeneity. In fact, the assembled genome of Methylotenera mobilis from SIP ‘heavy’ DNA is thought to be from a few closely related strains rather than from a clonal strain [36]. Genome assembly can benefit from better sequence coverage, and therefore more sequencing effort can be helpful for later genome assembly. However, although next-generation sequence technologies have significantly reduced the cost of DNA sequencing, cost is still an issue for many researchers involved in metagenomic research projects, and access to sophisticated bioinformatics expertise is also required. This is reflected in the literature – large-scale shotgun metagenomic studies have so far only been carried out by relatively few research groups worldwide.

Another major issue for current metagenomic studies is that it is at best a guess to infer the metabolism of microorganisms from their DNA sequences alone, and subsequent experimental validation is required. Combining DNA–SIP with metagenomics can help to resolve the functions of target microorganisms at the population level (i.e. depending on the uptake of particular substrates). In many cases it is also desirable to pinpoint the function of microorganisms from the environment at the single-cell level. There is now tremendous technological development in this regard. This includes combining cell-sorting and

---

**Table 2. Studies combining DNA–SIP with metagenomics**

<table>
<thead>
<tr>
<th>Environmental samples</th>
<th>Procedures and results</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acidic soil from a forest</td>
<td>Forest soil (5 g) was exposed to 50 ml of 13CH4; ‘heavy’ DNA was digested with BamHl and ligated to a BAC vector pCC1BACTM (average insert size ~25 kb); 2300 colonies were screened by colony blot and two clones containing a pmoCAB operon were found.</td>
<td>[13]</td>
</tr>
<tr>
<td>Marine sediment</td>
<td>Enrichment culture was exposed to 13C-glycerol; ‘heavy’ DNA was ligated to plasmid pBluescript (insert size ~2 kb); a library (87.8 Mbp) was screened by colony blot, 27 positive clones containing B12-dependent glycerol- and diol-dehydratases were isolated.</td>
<td>[14]</td>
</tr>
<tr>
<td>Marine surface water</td>
<td>Surface seawater was exposed to 1 μM 13C-methanol; ‘heavy’ DNA was subjected to MDA amplification; MDA-generated DNA was size-selected (25—60 kb) and ligated to fosmid vector pCC1FOS; 1500 colonies were screened by colony PCR with degenerate primers for mxaF; one colony containing a Methylphilhaga-related methanol-dehydrogenase gene cluster was obtained.</td>
<td>[25]</td>
</tr>
<tr>
<td>Acidic peat soil</td>
<td>Peat soil (5 g) was exposed to 10,000 ppm 13CH4; ‘heavy’ DNA was subjected to MDA and subsequent enzyme treatments to minimise potential chimeras formation, and then size-selected (30—50 kb) and ligated to fosmid vector pCC1FOS; 2300 colonies were screened by PCR and colony blot; three clones containing 16S rRNA genes from methanotrophs and one clone containing genes involved in methanol utilisation were found.</td>
<td>[35]</td>
</tr>
<tr>
<td>Sediment and water from a lake</td>
<td>10 g of sediment and 90 ml of water were exposed to 1—10 mM 13C substrates (methane, methanol, methylamine, formaldehyde and formate); ‘heavy’ DNA was used to construct a plasmid library in pUC18 (insert size 1—3 kb); 225 Mbp of DNA sequence was obtained using a shotgun-sequencing approach; a nearly complete genome of Methyltenera mobilis was obtained.</td>
<td>[36]</td>
</tr>
<tr>
<td>Contaminated river sediment</td>
<td>5 g of sediment were exposed to 10 mg of 13C-biphenyl; ‘heavy’ DNA was size-selected (25—40 kb) and ligated to a cosmid vector pWEBTM; the cosmid library (1568 colonies, average size 30—40 kb) was screened by PCR. This yielded one positive clone containing an aromatic-ring-hydroxylation dioxygenase within a 31 kb cosmid insert.</td>
<td>[42]</td>
</tr>
</tbody>
</table>
microfluidics with metagenomics, combining Raman microscopy with single-cell genomics [44], combining secondary ion mass spectrometry with fluorescence in situ hybridisation [45], and combining microautoradiography with fluorescence in situ hybridisation [46]. Readers are referred to other recent reviews and references therein for further coverage of these exciting new technologies [27,47].

The other difficulty in inferring function based on metagenome sequence lies in the fact that the majority of the genes in public databases lack a definitive functional assignment. 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 [16]. In fact, even for the best studied model organism, E. coli, the functions of approximately 50% of the genes in its genome have not yet been experimentally confirmed [48]. Current functional assignment for genes from metagenomes as well as from individual genomes is based on homology searches using BLAST tools that heavily depend on the quality as well as the completeness of current databases. On the one hand, the fact that two genes share high sequence-similarity does not guarantee that they perform a similar biological function; on the other hand, functional redundancy implies that one particular function could be performed by several proteins with no significant homology (e.g. Ref. [49]). It is therefore possible that large errors (or at least imprecisions) are present in current metagenome function assignments, and therefore also in metabolic pathway reconstruction and prediction. Improvement in the standard of gene annotation in public databases is urgently needed, not only for environmental microbiologists but for all users of these databases.

The biotechnological potential of combining DNA–SIP with metagenomics could offer considerable advances in the near future, particularly in view of the urgent need for novel enzymes in industry. Significant improvements in gene-detection frequency can reduce the cost of finding a novel enzyme. However, a number of issues need to be resolved before this technique can be widely used. The major difficulty is probably the development of a high-throughput production line for analysing multiple DNA–SIP incubations and 13C-DNA isolation. DNA–SIP was originally designed to analyse just a few samples simultaneously and is time-consuming. Bioindustry uses high-throughput methods for screening multiple samples, and conventional metagenomics has already been implemented in this process. Similar high-throughput methodology will need to be developed in order that SIP-metagenomics can be used for large-scale enzyme discovery. This is certainly achievable in the future, but will need close collaboration between environmental microbiologists and bioindustries.

SIP-metagenomics, as are many other approaches developed in the past few years, is aimed at improving the usefulness of metagenomics in understanding the unseen microbial majority. The combination of metagenomics with other contemporary and complementary technologies will guarantee a productive future for metagenomics.

Acknowledgements
Y.C. and J.C.M. acknowledge the Natural Environment Research Council (U.K.) for funding. We thank Dr. K. Purdy for his comments on the manuscript.

References
Review


Have your say

Would you like to respond to any of the issues raised in this month’s *T*IM? Please contact the Editor (etj.tim@elsevier.com) with a summary outlining what will be discussed in your letter and why the suggested topic would be timely. You can find author guidelines at our new website:

http://www.cell.com/trends/microbiology