Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Biotechnology

# Advances in monitoring environmental microbes

Terry C Hazen[1,2,3,4,5], Andrea M Rocha[1,4,5] and Stephen M Techtmann[1,4,5]

Culture-independent approaches, such as next-generation sequencing and microarray-based tools, provide insight into the identity and functional diversity of microbial communities. Although these approaches are potentially powerful tools in understanding microbial structure and function, there are a number of limitations that may bias conclusions. In order to mitigate these biases, an understanding of potential biases within each stage of the experimental process is necessary. This review focuses on the biases associated with sample collection, nucleic acid extraction, processing, sequencing analyses, and Chip technologies used in microbial ecology studies.

**Addresses**
[1] Department of Civil & Environmental Engineering, University of Tennessee, Knoxville, TN 37996, United States
[2] Department of Microbiology, University of Tennessee, Knoxville, TN 37996, United States
[3] Department of Earth & Planetarty Sciences, University of Tennessee, Knoxville, TN 37996, United States
[4] Center for Environmental Biotechnology, University of Tennessee, Knoxville, TN 37996, United States
[5] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States

Corresponding author: Hazen, Terry C (tchazen@utk.edu)

**Current Opinion in Biotechnology** 2013, **24**:526–533

This review comes from a themed issue on **Environmental biotechnology**

Edited by **Robert J Steffan** and **Juan Luis Ramos**

For a complete overview see the Issue and the Editorial

Available online 23rd November 2012

0958-1669/$ – see front matter, 

http://dx.doi.org/10.1016/j.copbio.2012.10.020

## Introduction

The primary goal of microbial ecology is to understand the structure and function of a microbial community and detail the interactions between microbes in various environments. The inability of culture-dependent methods to assess the vast majority of environmental microbes has limited an accurate understanding of these organisms. Culture-independent methods have removed this limitation and provided access to a great wealth of phylogenetic and functional diversity contained within microbial communities. This access has enabled a sharper picture of microbial communities in a variety of settings and has enhanced our ability to harness microbes for biotechnological applications.

While culture-independent approaches have their advantages, there are a number of limitations and biases that may be introduced throughout sampling and processing of environmental samples. These biases depend on a number of individual processes, starting at collection of samples and ending with bioinformatics and conclusions. Within each step bias is introduced and carried through the pipeline, thus compounding that bias in the end result (Figure 1). Therefore, to ensure the quality of the results, we must take into account potential biases at each stage in the experimental process in order to temper our conclusions as well as take steps to mitigate that bias.

Large-scale assessments of microbial communities have utilized two distinct technologies: Sequencing-based and Chip-based approaches. In this review we examine the pipeline for molecular microbial ecology and identify some of the biases associated with each of the steps. Additionally, we compare the advantages and limitations of the use of metagenomic and gene-specific molecular technologies in the characterization of microbial communities.

## Sample collection

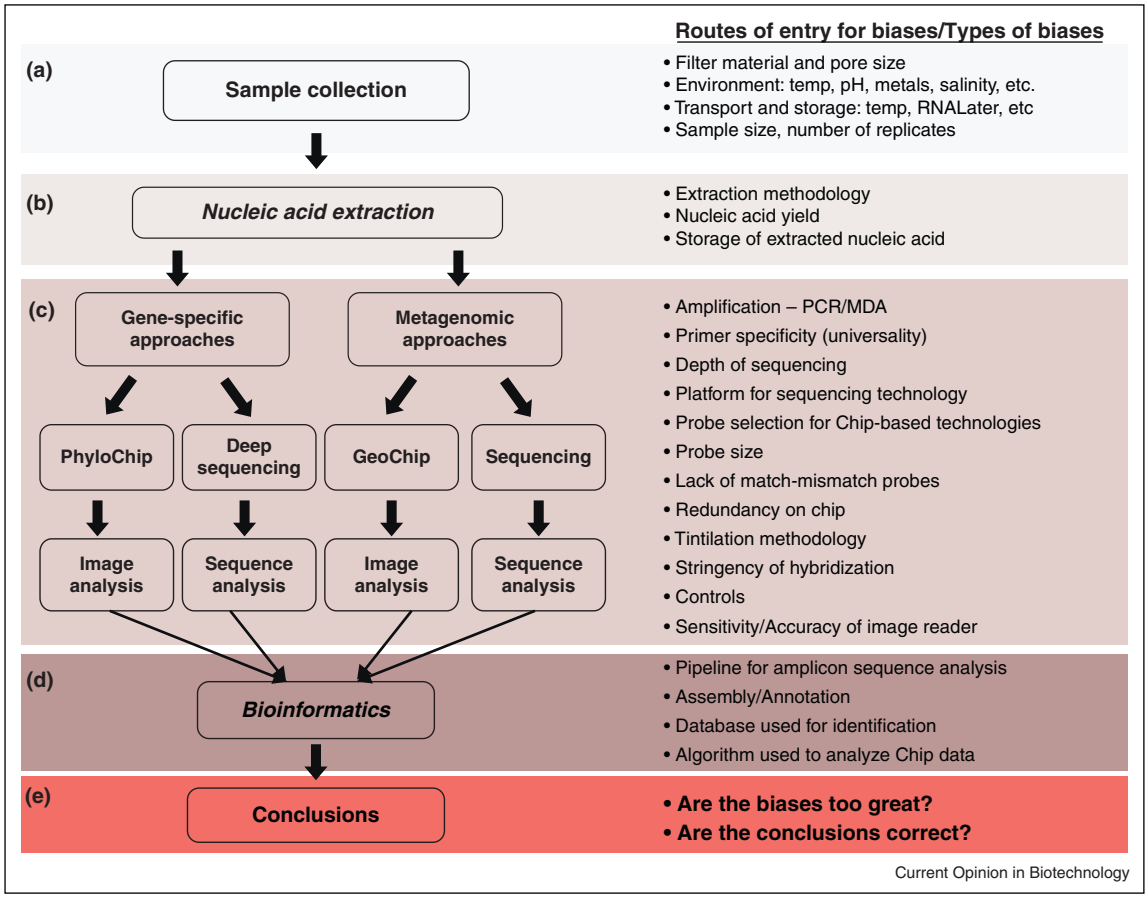
The most fundamental step in monitoring microbial communities is obtaining a representative sample of the community. From this sample, DNA will be extracted and further used to examine the community structure and function. Within this intial step, biases that can limit the overall analysis of the results may be introduced. For example, in environments where microbial biomass is low, it may be difficult to obtain enough environmental sample to characterize the entire microbial community. Therefore, the sampling approach must be carefully considered to limit the introduction of potential bias that will affect the analysis of a microbial community.

A thorough review of prior work performed on the community of interest or related communities can provide important information (i.e. estimates of biomass, known community members, chemical parameters, etc.) to aid in designing a sampling strategy that limits the introduction of biases. Factors that must be considered in order to limit biases related to sampling methodologies include: sample type (e.g. water versus sediment), amount of sample collected, methodology for collection of samples (e.g. filter versus centrifugation), and materials used (e.g. filter type).

### Sediment samples

For sediment samples, core or grab samples are often collected. During handling of the sediment samples the

**Figure 1**




Pipeline of individual processes associated with culture-independent approaches. Within steps A–E, biases are introduced and carried through, resulting in compounding bias. Biases introduced at earlier stages are further amplified by the end of the pipeline. Light color represents little bias and darker color represents increase in the number of bias present. **(a)** Sample collection is the initial step in culture-independent approaches; **(b)** extraction of nucleic acids; **(c)** molecular techniques and analyses associated with culture-independent approaches; **(d)** bioinfomatics; and **(e)** conclusions.

microbial community may be exposed to conditions different from their natural environment. Ideally, samples should be immediately stored at −80 °C until nucleic acids are extracted. This is especially important if RNA is the desired product. Another concern with processing sediment samples is the presence of substances such as humic acids. If humics are extracted along with DNA or present with the sample, they can potentially interfere with downstream applications like PCR [1,2,3•,4•,5,6]. Additionally, cleanup methods may result in loss of DNA within the sample [7,8].

## Water samples

Water samples are either collected using filtration or centrifugation. Selection of collection methods is generally dependent on the sample volume required. For example, if large volumes of water are required, filtration is preferable due the relative ease of filtering large volumes of water and the difficulty of transporting large volumes of water to be centrifuged.

A potential source of bias associated with filtration-based sampling methods is the selection of the filter membrane for collection of cells and extraction of DNA [9–11]. There are a number of membranes available, each with its own advantages and disadvantages. No matter what filter is used, one of the greatest concerns associated with DNA extraction from filters is the ability to recover the cells and nucleic acids from the filter material. One concern is the presence of non-biological contaminants, such as metals, in the water. If these materials are sorbed to the filter, there is a possibility that nucleic acids and lipids may be degraded or destroyed. Additionally, sorption of nucleic acids to the filter material could potentially bias results. For some filters, cells may adhere to the material thus limiting the overall yields of DNA as well as

biasing the extracted nucleic aids in favor non-adhesive cells. One approach to control for adhesion to the filter is to run controls using cultured organisms and/or pure DNA to verify that most, if not all, of the DNA is recovered from your filter. Although these quality controls are often not performed or reported, they are a key step in limiting biases associated with sample collection.

Pre-filters are often used to allow for filtration of greater volumes of water, particularly in turbid samples. The presence (or absence) of microbial cells collected on pre-filters is often not considered. Frequently these filters are disposed of and are not preserved for extraction of nucleic acids. In these cases, the particle-associated or aggregate or floc microbial community collected on the pre-filter is lost. Therefore the sample is biased in favor of the single cells or small particles with cells.

### Sample size

The number of cells collected in a sample must be sufficient to generate enough nucleic acids for downstream applications. Many extreme communities have very low biomass. As such, the volume of filtered water required for culture-independent studies, may range from a few to several liters. In extreme cases, volumes of up to >1000 L of filtered water have been reported [9,12]. Depending on the sample site (e.g. ocean versus groundwater well), quantification of microbial cell numbers at the time of sampling is not always possible. In these instances the total yield of DNA extracted from the sample may be low if microbial biomass is low.

## DNA extraction methods

Currently, there are a number of widely accepted and commercially available protocols for extraction of nucleic acids from various environments [3•,13,14,15•,16–18]. Since there are a number of papers comparing methods for extracting nucleic acids in current literature [3•,8,19,20], we will focus on biases associated with the general extraction processes.

A major concern with any extraction method is its ability to lyse and recover all of the nucleic acids from every microbe in that community. For example, Gram-positive bacteria are notorious for being difficult to lyse. It is therefore important to confirm that the extraction protocol used is tested on cultured strains before using that method on environmental samples.

Other methods include commercially available kits (e.g. Powerlyzer™, PowerSoil® DNA isolation kit, PowerWater® DNA Isolation Kit) or bead milling approaches [8,16]. These approaches rely upon a combination of physical and chemical processes to lyse the cells. One concern with bead beating is the possibility of shearing nucleic acids, thus lowering the overall quantity of intact DNA and RNA in the sample [18,21]. This in turn can bias gene-specific or metagenomic analysis. Additionally, bead-beating processes may result in increased recovery of humic acids, thus contaminating the sample [3•].

Other extraction methodologies utilize freeze-thaw grinding to lyse cells from environmental samples [18,21]. While this method has been shown to be effective for extracting high molecular weight DNA from both Gram positive and Gram negative cells, in addition to being time consuming, this approach may not be as effective for recovering cells directly from filters. Some filters, such as glass microfiber filters (Whatman GF/F) dissolve easily in lysis solution under freeze-thaw grinding conditions. However, these filters are often not desirable owing to the lack of durability during sample collection. If the filter material does not pulverize easily, a rinsing step to desorb cells from the filter is required before the sample can be ground. Washing the filters can introduce bias if removal of the cells from the filter is not complete. The wash-solution (e.g. phosphate buffer) and method for washing the filters may also bias results if microbial cells are lysed during the wash and if nucleic acids are lost during the rinse.

Since there is not one method that fits all environments, performing controls before nucleic acid extraction of field samples is necessary to identify potential biases. These controls might include gel electrophoresis to estimate the quality and amounts of release DNA, adsorption of cells and DNA to particles, spiked standards of known cell density and known amounts of DNA [11,17]. Since the quality of extracted nucleic acids greatly influences the quality of downstream data, it is essential to perform quality assurance or control checks before use. Quick assays, such as gel electrophoresis, to verify the effectiveness of the methodology may help identify potential problems early.

To prevent further bias, proper storage of extracted nucleic acids is essential. RNA should be extracted on ice and stored immediately at −80 °C. DNA, owing to its chemical structure can be extracted at room temperature and will not degrade as easily as RNA, but care should still be taken in handling and storing DNA as to preserve its integrity [22•,23•]. Since nucleic acids may be shipped to other laboratories for molecular analyses, proper storage during shipment is necessary.

## Overview of molecular techniques

High quality nucleic acids can be used with a number of techniques in order to monitor microbial communities. In many cases, molecular methods are used to address the identity and functional repertoire of community members. Recent advances in technology have enabled these questions to be addressed using both next-generation sequencing and microarray-based tools. Each of these approaches has its unique advantages and disadvantages

**Table 1**

**Comparision of molecular techniques for assessing community structure (gene-specific) and community function (Metagenomic) approaches**

| Pipeline | Advantages | Disadvantages | Ref. |
|---|---|---|---|
| Gene-Specific | | | |
| Sequence-based | 1) Ability to use sequence data to more finely define phylogeny | 1) Biases introduced by PCR amplification | [17–22,23•] |
| | | 2) Biases introduced with selection of primer pair | |
| | 2) Ability to identify novel species | 3) Will be skewed toward dominant organisms | |
| Phylochip | 1) Able to detect low abundance organisms<br>2) It is possible to be performed without PCR amplification | 1) Biases introduced by PCR amplification<br>2) Biases introduced with selection of primer pair<br>3) Will only see organisms that have probes on the array | [24,25,32] |
| Metagenomic | | | |
| Sequence-based | 1) Sequence data could be used for bioprospecting | 1) Low abundance organisms not represented – highly dependent upon depth of sequencing | [37–39,40•,41,42] |
| | 2) Able to identify novel gene families<br>3) Able to be assembled into larger contigs that could be used to link taxonomy to function. | 2) Limited by the accuracy of the annotation | |
| Geochip | 1) Confidence in assigned function | 1) Limited by the functions probed for on the chip [33,45,46] | |
| | 2) Able to detect low abundance genes from low abundance organisms | 2) No indication of similarity to other known genes<br>3) Unable to identify novel classes of genes | |

(Table 1), which must be considered when utilizing these techniques.

## Gene specific approaches

Sequencing of a single type of gene from an environment enables a focused analysis of the species diversity or the diversity of a particular functional group in that environment. Most commonly, the 16s rRNA gene is sequenced to determine the taxonomy [24–28]. Both microarray and sequencing-based 16s rRNA profiling approaches require an amplification step to increase the signal to observable levels [29–32]. However, well-known biases are introduced by any PCR amplification step [33,34]. PCR protocols have been modified to minimize this bias [35,36], yet these biases cannot be eliminated.

16s rRNA primer pairs are designed to target the broadest range of organisms. However, they have been shown to enrich the sample for some sequences while underrepresenting if not missing others [37,38]. Thus primer selection has great potential to introduce bias into the analysis. Recently, in an effort to eliminate PCR bias, a microarray approach was used in a PCR-independent manner [39]. While, this technique was successfully employed in this study, more work needs to be done to make the PCR-independent approach universally applicable.

Sequence-based approaches unlike microarray-based techniques are able to identify novel taxa as well as provide phylogenetic information about the species present. Phylochip-based approaches only detect the sequences with probes on the chip. This major caveat of phylochip-based protocols is overcome through the large number of probes present on the chip (representing over 50 000 strains [22•]). However, if there are no probes related to an environmental species, that species would not be detected.

In sequenced based-techniques, every sequence represents one sample from the pool of DNA present. Therefore, the dataset becomes more complete with greater sampling events [40•]. A complete data set is a prerequisite of accurately assessing the microbial diversity of a community. In mid to high diversity communities, the number of times that pool of DNA needs to be sampled, in order to generate a complete dataset via sequencing, must approach infinity [41]. To attempt to overcome this limitation, massively parallel sequencing has been employed to greatly increase the sample size. However it is still likely that many of the low abundance organisms will be underrepresented or absent. Also, systematic biases introduced in earlier steps (i.e. biases introduced in sample collection, DNA extraction, etc.) will only be amplified by deeper sequencing. Microarray based methods sample the entire pool of DNA at once allowing low abundance sequences to be detected. Thus the microarray data is not biased in favor of the most abundant organisms.

### Metagenomic approaches

The use of techniques aimed at elucidating the functional potential of a community has greatly advanced our understanding of the metabolic diversity of microbes and has led to a clearer picture of a community's function. Metagenomic techniques also have their advantages and biases. While metagenomic approaches have similar limitations to the gene-specific techniques, the biases of metagenomic approaches are further compounded by the more diverse set of sequences to be sampled (i.e. all of the genes versus only one type of gene). Similarly to PCR-based approaches many metagenomic techniques utilize multiple displacement amplification (MDA) to amplify whole genomes for down stream analysis such as sequencing, functional gene arrays and single cell genomics. While MDA has enabled the study of low biomass environments, this amplification step is known to introduce many biases, which may alter the conclusions [4•].

Metagenomic sequencing has identified new classes of enzymes and has been used as a bioprospecting tool for enzyme discovery and biotechnological advancement [6,42–45]. The sequences obtained from metagenomic sequencing can be used to further characterize these novel proteins. Furthermore, sequence reads can be assembled into larger contigs, which have the potential to elucidate the genomic context, further clarifying the function of particular genes. If a contig also contains a phylogenetic anchor, it is possible to begin to link taxonomy and function for some unculturable species within the environment [46,47]. However, as with gene-specific approaches, the ability to assemble a genome is highly dependent upon the diversity of the community and the depth of sequencing.

The presence of a certain functional gene in a metagenomic dataset provides direct evidence regarding the functional complement of the dominant species. However, many of the low abundance organisms, whose genes are often missed by sequencing, may be an integral part of that community. The absence of a function within a metagenomic dataset may not be conclusive evidence that a particular function truly is absent. But is more likely an artifact of insufficient coverage.

Another caveat of metagenomic sequencing is that interpretation of sequence data rests upon accurate annotation. It is known that annotation is a point where biases may be introduced [40•,48,49]. Therefore, stringent annotation pipelines need to be utilized when analyzing metagenomic data to guard against incorrect annotations and in turn spurious conclusions about the functional complement of an ecosystem.

Functional gene microarray approaches (e.g. Geochip) are able to address many of the weakness of sequence-based approaches. Since Geochip probes are designed based on characterized functional classes, a microarray-approach circumvents the dependence upon annotation, removing annotation bias. Geochip also provides an opportunity to evaluate the functional capacity of low abundance members of the community [22•,50•]. Like phylochip, Geochip is able to sample the whole pool of DNA at once, not biasing the sample toward dominant organisms.

Conversely, Geochip has weaknesses of its own. Geochip is limited in its inability to detect novel gene families. The latest version of the Geochip includes about 84 000 probes targeting 410 gene families [22•,50•,51]. However the technology only detects known sequences whose probes are on the chip, which limits the utility of Geochip in identifying novel biotechnologically relevant gene families. Furthermore, while functional gene array probes are based on 'well conserved' region of genes, there can be great divergence among genes from the same family. Therefore, Geochip may miss members of the same families probed for on the chip.

The performance of Chip-based technology is dependent upon the design of the array and the quality assessments built into the array. Some of the design issues that may affect the fidelity of an array include, match–mismatch checks, the length of the probes used and the redundancy of spots for a particular gene. Match–mismatch probe sets allow for a determination of the amount of cross hybridization. This measurement can be used to provide a more accurate and specific readout of gene presence and abundance. Match–mismatch probes are included on the phylochip. However, this technology is not built into the functional gene arrays. Therefore, the amount of cross-hybridization and the similarity of an environmental gene to that on the chip cannot be measured using Geochip. Probe length also factors into the quality of the data from Chip-based technologies. Phylochip uses 25 mer oligos, whereas the Geochip probes are approximately 50 mer oligos [22•]. The longer probes provide more specificity to the target of interest and have the potential to eliminate some mismatch errors. The redundancy of probes on a chip is also an additional array design feature, which can control for issues regarding how the DNA sample is spread on the chip. These controls would provide confidence that the measured intensities are correct and not an artifact of insufficient spreading of the sample throughout the Chip.

Additional biases can be introduced into Chip-based technologies in the design of the experiment. For array-based techniques, the DNA must be labeled with a fluorochrome. The process of labeling could introduce bias if the DNA is not labeled uniformly, thus skewing the apparent relative abundance of the DNAs. The stringency of the hybridization is another step that has great potential to impact the results. A less stringent

hybridization could result in false positives, whereas a highly stringent hybridization will result in false negatives. Finally, the sensitivity of the image reader can affect the results of the experiment. The sensitivity and accuracy of the image reader will determine the detection limit as well as overall quantification of the genes of interest.

## Bioinformatics

The data generated in these molecular methods must be processed through various bioinformatics pipelines. This processing allows the data, either in the form of an image (microarray) or sequence data (sequencing-based), to be converted into a form that can be analyzed in order to draw conclusions about the microbial community. These pipelines have the potential to introduce many biases into the final conclusions. The systematic biases introduced in earlier steps of the workflow will be further amplified in these bioinformatics techniques. The techniques used for the sequencing and array technologies are distinct and have their own biases. For example, the pipeline for analysis of amplicon-based or gene-specific techniques (e.g. QIIME, UniFrac, etc.) can greatly bias the final results. Further, analysis of metagenomic sequencing data is highly dependent upon gene calling and annotation. Gene calling is particularly difficult with metagenomic sequencing owing to the short reads and the inability of many sequences to be assembled. Assembly is also another major difficulty with sequence-based approaches. Many next-generation sequencing technology provides short reads, which then must be assembled [30]. If improperly assembled or short sequences are fed further down the pipeline into the gene calling and annotation steps those reads may be assigned errant functions, thus skewing the data [48,49]. Additional biases can be introduced when the database for gene identification is chosen. Different databases have different standards for classifying and curating gene identities and functional categories. High quality databases must be used when identifying genes from similarity to a known database.

Bioinformatic analysis of microarray data presents its own difficulties [31]. Much of this data is processed through various algorithms to convert the spot signal on the image to a response score. There exists some evidence that the choice of algorithm can greatly skew the results. Therefore it is important to consider the potential biases introduced in the choice of algorithm to process microarray-based data.

## Conclusion

It is important that as a community of researchers we identify biases throughout our experimental pipelines. Biases introduced into these workflows at earlier stages such as in the sample collection or DNA extraction steps could be amplified greatly throughout the process and be reflected in the quality of the data. Every technique discussed in this review is far from perfect and has the potential to introduce biases into the analysis. However, these techniques have great potential to advance our understanding of the structure and function of microbial communities. These tools have enabled unprecedented access to microbial communities and have allowed scientists to ask questions previously thought impossible to answer. However, we must balance the power of these techniques with a realistic understanding of their strengths and weaknesses. Comprehensive testing of sampling and extraction techniques before sample collection can mitigate some of these biases. Additionally the use of several techniques or pipelines to answer the same question can provide multiple lines of evidence for a particular community structure or the presence of a functional group. When sequence-based and array-based techniques are combined a more full and accurate picture of a microbial community can be gained [23•,50•]. As these technologies advance and many of these biases are mitigated, there may be one unifying technology that will unambiguously assess both community structure and function. Until then, the combination of these elegant culture-independent techniques to provide multiple lines of evidence will enable a more accurate assessment of microbial communities.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest


1. Picard C, Cecile P, Paget E, Nesme X, Simonet P: **Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction**. *Appl Environ Microb* 1992, **58**:2717-2722.

2. Wilson IG: **Inhibition and facilitation of nucleic acid amplification**. *Appl Environ Microb* 1997, **63**:3741-3751.

3. • Miller DN, Bryant JE, Madsen EL, Ghiorse WC: **Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples**. *Appl Environ Microbiol* 1999, **65**:4715-4724.

This research describes potential biases introduced within individual steps involved in extraction of nucleic acids. Additionally, this paper introduces modifications to current technologies for higher DNA yields representative of whole microbial communities.

4. • Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie EL, Hazen TC *et al.*: **Environmental whole-genome amplification to access microbial populations in contaminated sediments**. *Appl Environ Microbiol* 2006, **72**:3291-3301.

This paper describes multiple displacement amplification (MDA) and describes its limitations and bias. This paper has been heavily cited for the basic use of MDA for rare biosphere amplification.

5. Picard C, Ponsonnet C, Paget E, Nesme X, Simonet P: **Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain-reaction**. *Appl Envrion Microbiol* 1992, **58**:2717-2722.

6. Wilson IG: **Inhibition and facilitation of nucleic acid amplification**. *Appl Envrion Microbiol* 1997, **63**:3741-3751.

7. Dong DX, Yan A, Liu HM, Zhang XH, Xu YQ: **Removal of humic substances from soil DNA using aluminium sulfate**. *J Microbiol Meth* 2006, **66**:217-222.

8. Zhou JZ, Bruns MA, Tiedje JM: **DNA recovery from soils of diverse composition**. *Appl Envrion Microbiol* 1996, **62**:316-322.

9. Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, DeSantis TZ, Gihring TM, Lapidus A, Lin L-H, Lowry SR *et al.*: **Environmental genomics reveals a single-species ecosystem deep within earth**. *Science* 2008, **322**:275-278.

10. Haugland RA, Siefring S, Lavender J, Varma M: **Influences of sample interference and interference controls on quantification of enterococci fecal indicator bacteria in surface water samples by the qPCR method**. *Water Res* 2012, **46**:5989-6001.

11. Urakawa H, Martens-Habbena W, Stahl DA: **High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method**. *Appl Envrion Microbiol* 2010, **76**:2129-2135.

12. Hemme CL, Deng Y, Gentry TJ, Fields MW, Wu LY, Barua S, Barry K, Tringe SG, Watson DB, He ZL *et al.*: **Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community**. *ISME J* 2010, **4**:660-672.

13. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ: **Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition**. *Appl Envrion Microbiol* 2000, **66**:5488-5491.

14. Lovell CR, Piceno Y: **Purivication of DNA from Estuarine sediments**. *J Microbiol Meth* 1994, **20**:161-174.

15. • MacLean LCW, Pray TJ, Onstott TC, Brodie EL, Hazen TC, Southam G: **Mineralogical, chemical and biological characterization of an anaerobic biofilm collected from a borehole in a deep gold mine in South Africa**. *Geomicrobiol J* 2007, **24**:491-504.
This work describes the complexity of geochemical processes within biofilm communities, including precipation of metals, which could potentially introduce bias during sample collection.

16. Ogram A, Sayler GS, Barkay T: **The extraction and purification of microbial DNA from Sediments**. *J Microbiol Meth* 1987, **7**:57-66.

17. Steffan RJ, Goksoyr J, Bej AK, Atlas RM: **Recovery of DNA from soils and sediments**. *Appl Envrion Microbiol* 1988, **54**:2908-2915.

18. Tsai YL, Olson BH: **Rapid method for direct extraction of DNA from soil and sediments**. *Appl Envrion Microbiol* 1991, **57**:1070-1074.

19. Leff LG, Dana JR, McArthur JV, Shimkets LJ: **Comparison of methods of DNA extraction from stream sediments**. *Appl Envrion Microbiol* 1995, **61**:1141-1143.

20. More MI, Herrick JB, Silva MC, Ghiorse WC, Madsen EL: **Quantitative cell-lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment**. *Appl Envrion Microbiol* 1994, **60**:1572-1580.

21. Hurt RA, Qiu XY, Wu LY, Roh Y, Palumbo AV, Tiedje JM, Zhou JH: **Simultaneous recovery of RNA and DNA from soils and sediments**. *Appl Envrion Microbiol* 2001, **67**:4495-4503.

22. • Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N, Jansson JK, Probst A, Borglin SE, Fortney JL *et al.*: **Deep-sea oil plume enriches indigenous oil-degrading bacteria**. *Science* 2010, **330**:204-208.
This paper provides the proof of principle for the Phylochip3 and GeoChip4 and is a good example of multiple lines of evidence necessary for molecular microbial ecology pipelines.

23. • Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL, Han J, Holman HYN, Hultman J, Lamendella R *et al.*: **Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill**. *ISME J* 2012, **6**:1715-1727.
This paper combines many sequence-based techniques to characterize the functional capacity of an environment and compares the sequence-based data with the Chip-based data from Lu *et al.* to obtain a full picture of the functional capacity of the environment.

24. Armougom F, Raoult D: **Exploring microbial diversity using 16S rRNA high-throughput methods**. *J Comput Sci Syst Biol* 2009, **2**:74-92.

25. Mardis ER: **The impact of next-generation sequencing technology on genetics**. *Trends Genet* 2008, **24**:133-141.

26. Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R: **Microbiology in the post-genomic era**. *Nat Rev Microbiol* 2008, **6**:419-430.

27. Nossa CW, Oberdorf WE, Yang LY, Aas JA, Paster BJ, DeSantis TZ, Brodie EL, Malamud D, Poles MA, Pei ZH: **Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome**. *World J Gastroenterol* 2010, **16**:4135-4144.

28. Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA: **Microbial ecology and evolution – a ribosmomal-RNA approach**. *Ann Rev Microbiol* 1986, **40**:337-365.

29. Brodie EL, DeSantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL, Hazen TC, Richardson PM, Herman DJ, Tokunaga TK *et al.*: **Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation**. *Appl Environ Microbiol* 2006, **72**:6288-6298.

30. Degnan PH, Ochman H: **Illumina-based analysis of microbial community diversity**. *ISME J* 2012, **6**:183-194.

31. DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL: **High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment**. *Microb Ecol* 2007, **53**:371-383.

32. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ: **Microbial diversity in the deep sea and the underexplored "rare biosphere"**. *Proc Nat Acad Sci USA* 2006, **103**:12115-12120.

33. Sipos R, Székely A, Révész S, Márialigeti K: In *Addressing PCR Biases in Environmental Microbiology Studies Bioremediation.* Edited by Cummings SP. Humana Press; 2010:37-58.

34. Sipos R, Szekely AJ, Palatinszky M, Revesz S, Marialigeti K, Nikolausz M: **Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis**. *FEMS Microbiol Ecol* 2007, **60**:341-350.

35. Hori M, Fukano H, Suzuki Y: **Uniform amplification of multiple DNAs by emulsion PCR**. *Biochem Biophys Res Commun* 2007, **352**:323-328.

36. Williams R, Peisajovich SG, Miller OJ, Magdassi S, Tawfik DS, Griffiths AD: **Amplification of complex gene libraries by emulsion PCR**. *Nat Meth* 2006, **3**:545-550.

37. Hong SH, Bunge J, Leslin C, Jeon S, Epstein SS: **Polymerase chain reaction primers miss half of rRNA microbial diversity**. *ISME J* 2009, **3**:1365-1373.

38. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO: **Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies**. *Nucleic Acids Res* 2012, **1–11** http://dx.doi.org/10.1093/nar/gks808.

39. DeAngelis KM, Wu CH, Beller HR, Brodie EL, Chakraborty R, DeSantis TZ, Fortney JL, Hazen TC, Osman SR, Singer ME *et al.*: **PCR amplification-independent methods for detection of microbial communities by the high-density microarray PhyloChip**. *Appl Environ Microbiol* 2011, **77**:6313-6322.

40. • Delmont TO, Simonet P, Vogel TM: **Describing microbial communities and performing global comparisons in the 'omic era**. *ISME J* 2012, **6**:1625-1628.
This paper summarizes some of the unique challenges posed by advances in sequence-based technologies.

41. Mende DR, Waller AS, Sunagawa S, Jarvelin AI, Chan MM, Arumugam M, Raes J, Bork P: **Assessment of metagenomic assembly using simulated next generation sequencing data**. *PLoS ONE* 2012, **7**:11.

42. Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S, Sapra R, Hazen TC, Simmons BA, VanderGheynst JS *et al.*: **Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community**. *PLoS ONE* 2010, **5**:9.

43. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo SJ, Clark DS, Chen F, Zhang T *et al.*: **Metagenomic discovery of biomass-degrading genes and genomes from cow rumen**. *Science* 2011, **331**:463-467.

44. Li LL, McCorkle SR, Monchy S, Taghavi S, van der Lelie D: **Bioprospecting metagenomes: glycosyl hydrolases for converting biomass**. *Biotech Biofuels* 2009, **2**:11.

45. Lorenz P, Eck J: **Metagenomics and industrial applications**. *Nat Rev Microbiol* 2005, **3**:510-516.

46. Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD, Land ML, VerBerkmoes NC, Hettich RL, Banfield JF: **Enigmatic, ultrasmall, uncultivated Archaea**. *Proc Nat Acad Sci USA* 2010, **107**:8806-8811.

47. Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE: **De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities**. *ISME J* 2012, **6**:81-93.

48. Wooley JC, Godzik A, Friedberg I: **A primer on metagenomics**. *PLoS Comput Biol* 2010, **6**:13.

49. Wooley JC, Ye YZ: **Metagenomics: facts and artifacts, and computational challenges**. *J Comput Sci Technol* 2010, **25**:71-81.

50. • Lu ZM, Deng Y, Van Nostrand JD, He ZL, Voordeckers J, Zhou AF, Lee YJ, Mason OU, Dubinsky EA, Chavarria KL *et al.*: **Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume**. *ISME J* 2012, **6**:451-460.
This paper is an excellent example of the use of GeoChip technology to study the functional capacity of a particular environment.

51. Carter MQ, Xue K, Brandl MT, Liu F, Wu L, Louie JW, Mandrell RE, Zhou J: **Functional metagenomics of *Escherichia coli* O157:H7 interactions with spinach indigenous microorganisms during biofilm formation**. *PLoS ONE* 2012, **7**:e44186.