



Detecting unknown sequences with DNA microarrays: explorative probe design strategies.

Eric Dugat-Bony, Eric Peyretailade, Nicolas Parisot, Corinne Biderre-Petit,
Faouzi Jaziri, David R.C. Hill, Sébastien Rimour, Pierre Peyret

► To cite this version:

Eric Dugat-Bony, Eric Peyretailade, Nicolas Parisot, Corinne Biderre-Petit, Faouzi Jaziri, et al..
Detecting unknown sequences with DNA microarrays: explorative probe design strategies.. Environ-
mental Microbiology, 2012, 14 (2), pp.356-71. 10.1111/j.1462-2920.2011.02559.x . hal-00822134

HAL Id: hal-00822134

<https://hal.science/hal-00822134>

Submitted on 28 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Minireview

Detecting unknown sequences with DNA microarrays:
explorative probe design strategies

Eric Dugat-Bony,^{1,2} Eric Peyretailade,^{1,2}
Nicolas Parisot,^{1,2} Corinne Biderre-Petit,^{1,2}
Faouzi Jaziri,^{3,4} David Hill,^{3,4} Sébastien Rimour^{1,2} and
Pierre Peyret^{1,2*}

¹Clermont Université, Université Blaise Pascal,
Laboratoire Microorganismes: Génome et
Environnement, BP 10448, F-63000, Clermont-Ferrand.

²UMR CNRS 6023, Université Blaise Pascal, 63000
Clermont-Ferrand, France.

³Clermont Université, Université Blaise Pascal, LIMOS,
BP 10448, F-63000 Clermont-Ferrand.

⁴UMR CNRS 6158, LIMOS, F-63173 Aubière.

Summary

Designing environmental DNA microarrays that can be used to survey the extreme diversity of microorganisms existing in nature, represents a stimulating challenge in the field of molecular ecology. Indeed, recent efforts in metagenomics have produced a substantial amount of sequence information from various ecosystems, and will continue to accumulate large amounts of sequence data given the qualitative and quantitative improvements in the next-generation sequencing methods. It is now possible to take advantage of these data to develop comprehensive microarrays by using explorative probe design strategies. Such strategies anticipate genetic variations and thus are able to detect known and unknown sequences in environmental samples. In this review, we provide a detailed overview of the probe design strategies currently available to construct both phylogenetic and functional DNA microarrays, with emphasis on those permitting the selection of such explorative probes. Furthermore, exploration of complex environments requires particular attention on probe sensitivity and specificity criteria. Finally, these innovative probe design approaches require exploiting newly available high-density microarray formats.

Introduction

The microbial world represents the most important and diverse group of organisms living on earth (Whitman *et al.*, 1998; Curtis *et al.*, 2002), comprising most of the diversity of the three domains of life defined by Woese and colleagues (1990): *Archaea*, *Bacteria* and *Eucarya*. Furthermore, these organisms are widely distributed across many environmental habitats, even the most extreme. Their numerous enzymatic machineries have allowed them to adapt to almost every ecological niche and take advantage of any environmental condition (Øvreås, 2000; Guerrero and Berlanga, 2006). Despite our increasing knowledge of the role of microorganisms in ecosystem functioning, our current vision of the microbial world is still incomplete and several issues remain unsolved. This is partially explained (i) by the tremendous diversity of the genes and metabolisms of the existing species but also of ecological niches and (ii) by technological limits such as our inability to culture the majority of microorganisms (Amann *et al.*, 1995; Pace, 1997).

Because of this huge microbial biocomplexity, high-throughput molecular tools allowing simultaneous analyses of existing populations are greatly needed (Torsvik and Øvreås, 2002; Xu, 2006). Massive sequencing based on next-generation sequencing (NGS) technologies and microarrays are currently the most promising and complementary approaches to address these tasks (Claesson *et al.*, 2009; Roh *et al.*, 2010; van den Bogert *et al.*, 2011). Using NGS, two specific strategies can be applied: metagenomics, which refers to the study of the collective genomes in a given environmental community and the 16S rDNA amplicon sequencing approach. In principle, these methods enable: (i) access to the wide diversity of microbial communities, (ii) identification of unknown microorganisms and (iii) the potential to link structure to functions (Simon and Daniel, 2009). Some limitations of metagenomics, however, have been demonstrated: for example, the huge difficulty of managing large amounts of sequence data, or the short sequence read length (400–500 bases maximum with 454 FLX Titanium instrument from Roche), which complicates contigs assembling, or the sequencing errors caused by NGS

Received 31 March, 2011; accepted 4 July, 2011. *For correspondence. E-mail pierre.peyret@univ-bpclermont.fr; Tel. (+33) 473405139; Fax (+33) 473407670.

technologies (Roh *et al.*, 2010). Furthermore, Quince and colleagues (2008) estimated that detecting 90% of the richness in some hyperdiverse environments could require tens of thousands of times the current sequencing effort, which is inconceivable. Oligonucleotide microarray technologies have, however, been widely used for gene detection and gene expression quantification, and more recently, were adapted to profiling environmental communities in a flexible and easy-to-use manner (Zhou, 2003; Wagner *et al.*, 2007). These approaches can monitor the presence, or the expression, of thousands of genes, combining qualitative and quantitative aspects in only one experiment (Tiquia *et al.*, 2004; Marcelino *et al.*, 2006; Dugat-Bony *et al.*, 2011). Furthermore, this technology appears well adapted to multi-sample comparison. Although several whole-genome arrays have been developed in the last few years, phylogenetic oligonucleotide arrays (POAs), targeting the 16S rRNA genes, as well as functional gene arrays (FGAs), targeting key genes encoding enzymes involved in metabolic processes, are the two major approaches to assess diversity of microbial communities in the environment (Wagner *et al.*, 2007). Currently, the most comprehensive tools developed are the high-density PhyloChip, with nearly 500 000 oligonucleotide probes to almost 9000 operational taxonomic units (Brodie *et al.*, 2006), and the GeoChip 3.0 with ~28 000 probes covering approximately 57 000 gene variants from 292 functional gene families (He *et al.*, 2010). Whereas microarrays were demonstrated as being sufficiently sensitive, with detection of sequences representing genomic material from 0.05% to 5% of the total environmental community (Bodrossy *et al.*, 2003; Peplies *et al.*, 2004; Loy *et al.*, 2005; Gentry *et al.*, 2006; Marcelino *et al.*, 2006; Palmer *et al.*, 2006; Huyghe *et al.*, 2008), these methods require a sequence *a priori* to determine probes and hence allow surveys only of microorganisms with available sequences in public databases (Chandler and Jarrell, 2005; Wagner *et al.*, 2007).

The main problem that must be faced to construct oligonucleotide microarrays dedicated to microbial ecology is the probe design step. Indeed, environmental microarrays often require this step to be manually performed. Although numerous general probe design programmes are currently freely accessible for academics [for recent reviews see Lemoine and colleagues (2009)], only few may be useful for microbial ecology applications and are listed in Table 1. This review aims to show how probe design strategies can avoid the limitation of sequence availability and make possible the detection of previously uncharacterized microbial populations present in nature. We emphasize various recent methods combining the use of both degenerate and non-degenerate oligonucleotide probes to target either 16S rRNA markers, or new proteic variants. In conclusion, we highlight other procedures and

Table 1. Appropriate probe design software for microbial ecology studies.

Software	Applications in microbial ecology	Accessibility and user interface	URL	Reference
ARB	POA	Downloadable, standalone GUI (L, M)	http://www.arb-home.de/	Ludwig <i>et al.</i> (2004)
PRIMROSE	POA	Downloadable, GUI (L, W, M)	http://www.bioinformatics-toolkit.org/Primrose/index.html	Ashelford <i>et al.</i> (2002)
ORMA	POA, FGA	Matlab Script	Upon request	Severgnini <i>et al.</i> (2009)
PhylArray	POA	Web interface	http://ig2m.u-clermont1.fr/serimour/phyarray/	Milton <i>et al.</i> (2007)
HPD	FGA	Downloadable, standalone GUI (W)	Not available	Chung <i>et al.</i> (2005)
ProDesign	FGA	Web interface	http://www.uhnresearch.ca/labs/tillier/ProDesign/ProDesign.html	Feng and Tillier (2007)
HiSpOD	FGA, WGA	Web interface	http://fc.isima.fr/~g2lm/hispod/	Dugat-Bony <i>et al.</i> (2011)
Metabolic Design	FGA	Downloadable from a website, GUI (W)	ftp://195.221.123.90/	Terrat <i>et al.</i> (2010)
CommOligo (v 2.0)	FGA, WGA	Downloadable, standalone GUI (W)	http://leg.ou.edu/software.htm	Li <i>et al.</i> (2005)
OligoWiz (v 2.0)	FGA, WGA	Downloadable client programme, GUI (L, W, M)	http://www.cbs.dtu.dk/services/OligoWiz2/	Wernersson and Nielsen (2005)
ROSO	FGA, WGA	Web interface or standalone GUI (S, W, M) upon request	http://pbil.univ-lyon1.fr/roso/Home.php	Reymond <i>et al.</i> (2004)
ArrayOligoSelector	FGA, WGA	Downloadable, command line (L)	http://arrayoligosel.sourceforge.net/	Bozdech <i>et al.</i> (2003)
OligoArray (v 2.1)	FGA, WGA	Downloadable, command line (L)	http://berry.engin.umich.edu/oligoarray2_1/	Rouillard <i>et al.</i> (2003)
OligoPicker	FGA, WGA	Downloadable, command line (L)	http://pga.mgh.harvard.edu/oligopicker/	Wang and Seed (2003)
PROBEmr	POA, FGA, WGA	Web interface	Not available	Emrich <i>et al.</i> (2003)
YODA	FGA, WGA	Downloadable, standalone GUI (L, W, M)	Not available	Nordberg (2005)
ProbeSelect	WGA	Available upon request, command line (L)	Not available	Li and Stormo (2001)

FGA, functional gene array; GUI, graphical user interface; L, Linux; M, MacOS; POA, phylogenetic oligonucleotide array; S, SunOS; W, Windows; WGA, whole-genome array.

limitations that must be circumvented, to improve microarray development in terms of specificity and sensitivity.

General criteria for probe design

In silico probe design is one of the most critical step for microarray experiments because the selected oligonucleotide probe set will have to combine: (i) sensitivity (e.g. probes should detect low abundance targets in complex mixtures), (ii) specificity (e.g. probes should not cross-hybridize with non-target sequences) and (iii) uniformity (e.g. probes should display similar hybridization behaviour) (Loy and Bodrossy, 2006; Wagner *et al.*, 2007). According to Lemoine and colleagues (2009), this process requires dealing with many parameters and currently available probe design programmes differ in the choice of criteria that are considered to select the best probe set (Table 2).

Sensitivity

The sensitivity generally increases with probe length, as the binding energy for longer probe-target hybrid complexes is typically higher and hybridization kinetics are irreversible (Hughes *et al.*, 2001; Relogio *et al.*, 2002; Letowski *et al.*, 2004). For example, probes of 60 mers can detect targets with eightfold higher sensitivity than those of 25 mers (Chou *et al.*, 2004). However, their threshold for differentiation is at 75–90% sequence similarity (Kane *et al.*, 2000; Taroncher-Oldenburg *et al.*, 2003; Tiquia *et al.*, 2004), which indicates a poor specificity (Li *et al.*, 2005). In contrast, short oligonucleotide probes are more specific, allowing discrimination of single nucleotide polymorphisms under optimal conditions, but at the cost of reduced sensitivity (Relogio *et al.*, 2002). Furthermore, the formation of stable secondary self-structures like stem-loops, hairpins and probe-to-probe dimerization by the probes or their targets is another crucial factor that must be considered to minimize loss of microarray sensitivity. However, despite a good knowledge of the thermodynamic properties of nucleic acid duplex formation and dissociation in solution (SantaLucia *et al.*, 1996) and the availability of several algorithms like Mfold (Zuker, 2003) or Hyther (Bommarito *et al.*, 2000) for their accurate prediction, these calculations should be treated cautiously in the microarray context due to the limited knowledge on the thermodynamics of hybridization at solid–liquid interfaces (Pozhitkov *et al.*, 2006; 2007).

Specificity

The specificity of microarray hybridization is one of the main effectors of the result quality (Kane *et al.*, 2000;

Evertsz *et al.*, 2001; Koltai and Weingarten-Baror, 2008). Therefore, it is crucial that oligonucleotide probes must be unique with respect to all non-target sequences. To check probe specificity, software usually use results produced by algorithms such as BLAST or suffix array method, to search for cross-hybridization against databases constructed in accordance with the microarray application. In this step, potential cross-hybridization prediction are usually based on Kane's recommendations (probe should not have a total percent identity > 75–80% with a non-target sequence, or contiguous stretches of identity > 15 nt with a non-target sequence) (Kane *et al.*, 2000) or thermodynamics calculations (duplex's stability between the probe and the non-target sequence). Moreover, low-complexity regions such as those containing long homopolymers may also contribute to affect probe specificity and must therefore be avoided for probe design (Wang and Seed, 2003; Leparac *et al.*, 2009).

Uniformity

Because microarray technology relies on the simultaneous hybridization of many probes under the same conditions (salt concentration, temperature, etc.), it is important to ensure that the selected probes have thermodynamic behaviours as uniform as possible (Loy and Bodrossy, 2006; Wagner *et al.*, 2007). The easiest way to achieve this is to select probes with homogeneous structural properties such as probe length, G + C content, melting temperature (T_m) or binding capacities (ΔG).

Characterization of environmental species with POAs

The classical way to characterize members of complex bacterial communities relies on the small subunit ribosomal RNA gene (16S rRNA) analysis. This target is particularly well adapted to phylogenetic studies as it contains highly conserved and variable moieties permitting reliable and detailed bacterial classification. Moreover, the advent of many PCR-based approaches, as well as sequencing projects, has led to the explosion of 16S rRNA gene sequences now available in major specialized sequence repositories, such as SILVA (Pruesse *et al.*, 2007), Greengenes (DeSantis *et al.*, 2006) and the Ribosomal Database Project (RDP) (Cole *et al.*, 2009).

In order to rapidly survey prokaryotic communities present in complex environments high-throughput tools have been developed, such as POAs using the SSU rRNA biomarker (Wilson *et al.*, 2002; Brodie *et al.*, 2006; Palmer *et al.*, 2006; DeSantis *et al.*, 2007). The main obstacle in designing a POA, however, is potential cross-hybridization. In many cases, the 16S rRNA genes of the type species are too conserved to allow the design of

Table 2. Comparison of probe design software features.

Software	Probe length (nt)	Secondary structure	Low-complexity	GC content	T_m	ΔG	Degenerate probes	Cross-hybridization assessment	Database for specificity test
ARB	Fixed by the user (10–100)	No	No	Yes	Yes	No	No	Local alignment and thermodynamic calculations	ARB-Silva Database
PRIMROSE	Fixed by the user (3–100)	No	No	No	No	No	Yes	ND	RDP-II Database
ORMA	Fixed by the user	No	Yes	No	Yes	No	Yes	No	No
PhylArray	Fixed by the user (20–70)	No	No	Yes	Yes	No	Yes	BLAST and Kane's specifications	Custom non-redundant SSU rRNA database (95 Mo)
HPD	Fixed by the user (20–70)	Yes	No	Yes	Yes	Yes	No	BLAST and Kane's specifications	Input sequence dataset
ProDesign	Fixed by the user (20–70)	Yes	Yes	Yes	Yes	Yes	No	Spaced seed hashing and Kane's specifications	Input sequence dataset
HiSpOD	Fixed by the user	No	Yes	Yes	Yes	No	Yes	BLAST and Kane's specifications	EnvExBase (10Go) Complete CDS Database
Metabolic Design	Fixed by the user	No	No	No	No	No	Yes	BLAST and Kane's specifications	EnvExBase (10Go) Complete CDS Database
CommOligo (v 2.0)	Fixed by the user	Yes	Yes	Yes	Yes	No	No	Global alignment, thermodynamic calculations and Kane's specifications	Input sequence dataset
OligoWiz (v 2.0)	Fixed by the user	Yes	No	No	Yes	No	No	BLAST, Kane's specifications and thermodynamic calculations	Single organism genome
ROSO	Fixed by the user	Yes	Yes	Yes	Yes	Yes	No	BLAST	External fasta file (typically single organism genome)
ArrayOligoSelector	Fixed by the user	Yes	Yes	Yes	No	No	No	BLAST and thermodynamic calculations	External fasta file (typically single organism genome)
OligoArray (v 2.1)	Fixed by the user (15–75)	Yes	Yes	Yes	Yes	No	No	BLAST and thermodynamic calculations	External fasta file (typically single organism genome)
OligoPicker	Fixed by the user (20–100)	Yes	Yes	No	Yes	No	No	BLAST	Input sequence dataset or external fasta file (typically single organism genome)
PROBEmer	Fixed by the user	Yes	No	Yes	Yes	Yes	No	Suffix array approach	RDP (v 8.1), single organism genome or external fasta file
YODA	Fixed by the user	Yes	Yes	Yes	Yes	No	No	BLAST and Kane's specifications	External fasta file (typically single organism genome)
ProbeSelect	Fixed by the user	Yes	Yes	No	No	Yes	No	Suffix array approach and thermodynamic calculation	Single organism genome

ND, not determined.

discriminatory probes (Bae and Park, 2006). To circumvent this problem, a hierarchical design allows probing for microbial taxa at different phylogenetic levels (Huyghe *et al.*, 2008; Liles *et al.*, 2010), providing information on the presence or absence of the branches and the twigs on the Tree of Life.

Probe design for POA

Both fully automated software and manual approaches have been developed to design POAs, taking into account the main criteria for efficient probe design, which are sensitivity and specificity. Currently, three programmes have been developed to work with structured data for retrieving and analysing sequences from dedicated databases and to operate a phylogenetic probe design targeting the 16S rRNA gene.

The first programme is the Probe Design tool included in the ARB programme package (Ludwig *et al.*, 2004) which is commonly used to select 10–100 mer oligonucleotides. The first step in the programme consists of the target group selection. Second, the algorithm identifies unique sequence stretches that could serve as target sites, and subsequently returns a sorted list of potential oligonucleotides. Third, the suggested probes can be matched against all sequences in the database using the Probe Match software programme. ARB also proposes different sets of predefined probes, each targeting distinct phylogenetic groups. It has been widely used to develop low-density custom-made POAs, containing up to a few hundred oligonucleotide probes. These probes usually target either restricted microorganism groups known to perform a specific metabolism (Loy *et al.*, 2002; Kelly *et al.*, 2005; Franke-Whittle *et al.*, 2009), or belonging to a specific taxon (Castiglioni *et al.*, 2004; Lehner *et al.*, 2005; Loy *et al.*, 2005; Kyselkova *et al.*, 2008; Schonmann *et al.*, 2009; Liles *et al.*, 2010), or living in a habitat/ecosystem of particular interest (Neufeld *et al.*, 2006; Sanguin *et al.*, 2009). To illustrate this purpose, Sanguin and colleagues (2009) identified multiple changes in rhizobacterial community composition associated with the decline of take-all disease of wheat caused by the soil-borne fungus *Gaeumannomyces graminis* by using a taxonomic 16S rRNA-based microarray targeting both *Bacteria*, *Archaea* and the OP11 and OP2 candidate divisions. ARB has also been used to construct phylogenetic microarrays based on other biomarkers such as protein coding genes (Bodrossy *et al.*, 2003; Duc *et al.*, 2009).

The second programme is the PRIMROSE programme (Ashelford *et al.*, 2002), which uses standard or custom databases, and allows the design of degenerate probes. Initially, a multiple alignment is produced using all the different sequences representing a given taxon. Every probe is subsequently tested against all the sequences of

the initial database, to characterize potential cross-hybridizations and to verify good coverage of the targeted taxon. Although this tool was developed to identify both phylogenetic probes and primers, it has been mainly applied to PCR-based and FISH (fluorescent *in situ* hybridization) approaches (Rusch and Amend, 2004; Yu *et al.*, 2005; Feldhaar *et al.*, 2007; Boeckeaert *et al.*, 2008; Klitgaard *et al.*, 2008; Muhling *et al.*, 2008; Gittel *et al.*, 2009; Fraune *et al.*, 2010; Bers *et al.*, 2011). Few applications of POAs using PRIMROSE have been reported. Blaskovic and Barak (2005) reported the development of a user-friendly chip to specifically detect tick-borne bacteria responsible of human and animal diseases.

Nevertheless, neither of these two applications is built specifically for the determination of discriminating positions within a set of very similar sequences. The third programme is ORMA (Oligonucleotide Retrieving for Molecular Applications), which represents a good alternative solution (Severgnini *et al.*, 2009). This programme designs and selects oligonucleotide probes for molecular application experiments on sets of highly similar sequences. Although it was first applied to the design of probes targeting 16S rRNA genes, this software can be used on any set of highly correlated sequences, such as new potential phylogenetic biomarkers. Using this programme, Candela and colleagues (2010) designed the HTF-Microbi.Array allowing high taxonomic level fingerprinting of the human intestinal microbial community.

In parallel, other computational approaches not implemented under fully automated software were developed to design high-density POAs (thousands of oligonucleotide probes) allowing a comprehensive screening for all known bacterial or archaeal taxa with a single microarray (Wilson *et al.*, 2002; DeSantis *et al.*, 2007). These approaches rely on sophisticated algorithms for the design of a multitude of probes and for the analyses of highly complex hybridization patterns. The best example is the PhyloChip developed by Brodie and colleagues (2006), which contains 500 000 probes based on the Affymetrix GeneChip platform. This tool is able to simultaneously identify thousands of taxa present in an environmental sample and has been applied to characterize prokaryotic communities from ecosystems such as urban atmosphere (Brodie *et al.*, 2007), grassland soils (Cruz-Martinez *et al.*, 2009; DeAngelis *et al.*, 2009), Antarctic soils (Yergeau *et al.*, 2009), mining-impacted soils (Rastogi *et al.*, 2010a,b), metal-contaminated river sediments (Rastogi *et al.*, 2011), terrestrial volcanic glasses (Kelly *et al.*, 2010), rhizosphere of potato (Weinert *et al.*, 2011), citrus leaf (Sagaram *et al.*, 2009), endotracheal aspirates from patients colonized by *Pseudomonas aeruginosa* (Flanagan *et al.*, 2007), and pearly eyed thrasher eggs (Shawkey *et al.*, 2009). Recently, due to increased interest in microbes of human and animal

gastrointestinal tracts, a number of high-density microarrays were also developed to study the composition and activity of intestinal microbiota (Palmer *et al.*, 2006; Paliy *et al.*, 2009; Rajilic-Stojanovic *et al.*, 2009).

The main limitation of all the strategies proposed for 16S rRNA probe design is that they only ensure the survey of known microorganisms with available sequences in public databases. Unfortunately, the vast majority of microbial species is still unidentified and, therefore, is not represented by sequences in public ribosomal rRNA databases. A major challenge for the future is improvement of microarray technology to, in part, rely on new strategies for the design of explorative probes targeting sequences, which have not yet been described.

Explorative probe design strategies for POA

The 'multiple probe concept' consists of using several probes targeting an organism at similar and different phylogenetic/taxonomic levels. Designing probes using this concept significantly reduces the risk of misidentification, and often allows discrimination of bacteria down to the species level (Ludwig *et al.*, 1998; Loy and Bodrossy, 2006; Schliep and Rahmann, 2006; Huyghe *et al.*, 2008; Schonmann *et al.*, 2009; Liles *et al.*, 2010). Arrays constructed using this concept may ensure the detection of unknown taxa using probes defined from known higher phylogenetic levels. Because such probes are strictly complementary to available sequences, however, they do not harbour the explorative power to detect microorganisms with uncharacterized phylogenetic signatures.

Currently, the only software dedicated to POAs offering the possibility of designing explorative probes, is the PhylArray programme (Milton *et al.*, 2007). This algorithm generates 16S rRNA probes to globally monitor known and unknown bacterial communities in complex environments. The first step in the design is the extraction of all available sequences corresponding to a given taxon from a custom 16S rRNA curated database. Second, a multiple sequence alignment is performed using the ClustalW algorithm (Thompson *et al.*, 1994). Third, a degenerate consensus sequence is produced taking into account sequence variability at each position, which allows the selection of degenerate probes. Finally, all combinations from each degenerate probe are checked for cross-hybridization against the 16S rRNA database. Among the combinations derived from each degenerate probe, some correspond to sequences not previously included in public databases (Fig. 1). They should, therefore, allow for the exploration of the as yet undescribed fraction of environmental microbial communities. Moreover, comparative experimental evaluations indicate that probes designed with PhylArray yield a higher sensitivity and specificity than those designed with the PRIMROSE and ARB strat-

egies (Milton *et al.*, 2007). Recently, a microarray designed with the PhylArray strategy has been employed to evaluate the bacterial diversity in two different soils (Delmont *et al.*, 2011). The authors highlighted the significant influence of several parameters like sampling depth or DNA extraction protocols on the biodiversity estimation.

Detection of functional signatures for FGA design

Assessing the metabolic potential of microorganisms in natural ecosystems is an interesting goal in microbial ecology. In fact, some authors estimate that individual environmental samples, like soil, may contain between 10^3 and 10^7 different bacterial genomes (Curtis *et al.*, 2002; Gans *et al.*, 2005), each of them harbouring thousands of genes. In this context, high-density oligonucleotide FGAs provide the best high-throughput tools to access this tremendous genetic content (He *et al.*, 2008). GeoChips, composed of 50 mer probes designed with CommOligo (Li *et al.*, 2005), are currently the most comprehensive FGAs. Indeed, these microarrays have evolved over several generations and now target key genes involved in most microbial functional processes such as carbon, nitrogen, phosphorus and sulfur cycles, energy metabolism, antibiotic resistance, metal resistance and organic contaminant degradation (Rhee *et al.*, 2004; He *et al.*, 2007; 2010; 2011). However, being able to encompass the full diversity of gene family sequences encountered in nature, described in databanks or not, is still one of the most difficult challenges for the future. Most FGAs described to date only monitor sequences available in databases and, therefore, cannot appraise the unknown part of the microbial gene diversity present in complex environments. A more extensive coverage of the probe set is, therefore, crucial and designing explorative probes represents a pertinent and essential approach.

Characterization of new functional signatures from nucleic sequence alignment

Many probe design programmes are currently freely accessible for academics [for recent reviews see Lemoine and colleagues (2009)]. Most of them were developed for use on single-genome datasets, and hence, are limited to the determination of probes targeting specific gene sequences (Table 2). In contrast, few strategies offer the opportunity to design probes allowing a broad coverage of multiple sequence variants for a given gene family.

With the availability of more and more sequences corresponding to functional genes (complete genome sequencing and environmental studies from specific functional markers), new programmes have been developed in the last decade taking into account this wide diversity. Hierarchical Probe Design (HPD) was the first programme

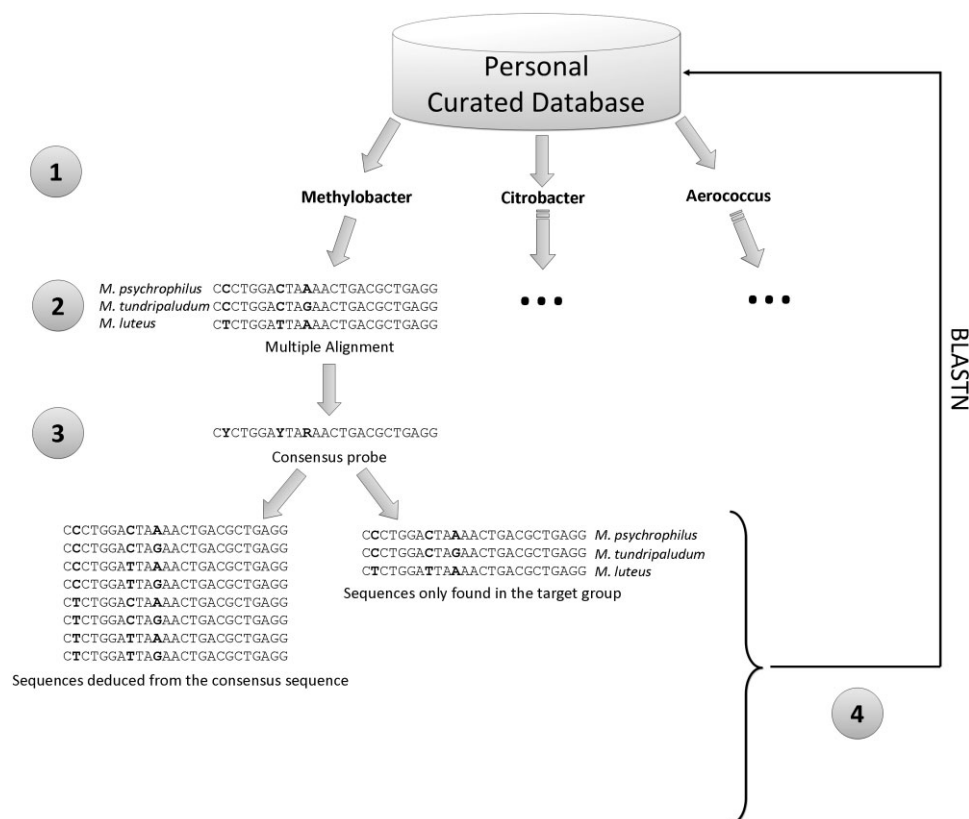


Fig. 1. PhylArray programme workflow. PhylArray programme is composed of four steps: (i) sequence extraction for each taxon, (ii) multiple sequence alignment, (iii) degenerate consensus sequence production and probe selection and (iv) specificity tests against the 16S rRNA database.

dedicated to functional oligonucleotide determination based on the concept of cluster-specific probes (Chung *et al.*, 2005). The first step of the programme consists of the alignment and hierarchical clustering of input sequences in order to generate all possible candidate probes. The optimal probe set is subsequently determined according to probe quality criteria, including cluster coverage, specificity, GC content and hairpin energy. Although this tool is not explorative, it automatically produces probes against all nodes of the clustering tree, providing an extensive coverage of known variants from a conserved functional gene. Using this programme, Rinta-Kanto and colleagues (2011) developed a taxon-specific microarray targeting sulfur-related gene transcription in members of *Roseobacter* clade, using data from 13 genome sequences. This FGA consisted of 1578 probes to 431 genes and was applied to the study of diverse natural *Roseobacter* communities. The results revealed that dimethylsulfoniopropionate was not preferred over other organic carbon and sulfur substrates by these populations.

ProDesign, developed by Feng and Tillier (2007), uses similar clustering methods with the aim of detecting all members of a same gene family in environmental

samples. But, unlike HPD, this software uses spaced seed hashing, rather than a suffix tree algorithm, in order to benefit from permitted mismatches between a probe and its targets, and ensures the re-clustering of groups for which no probe was found. This results in a significant improvement in sequence coverage. As with HPD, however, this tool does not provide probes targeting uncharacterized nucleic acid sequences. In addition, to the best of our knowledge, no application using this design strategy has been reported in literature.

Although both of these strategies allow a wider range of sequence variants to be covered, and, therefore, appear best suited to describe microbial communities from complex environments, their main drawbacks are their inability to generate explorative probes and the absence of specificity tests (i.e. searching for potential cross-hybridizations) against large databases representative of microbial diversity. Recently, an efficient functional microarray probe design algorithm, called HiSpOD (High Specific Oligo Design), was proposed to overcome this problem (Dugat-Bony *et al.*, 2011). It is particularly useful for studying microbial communities in their environmental context. HiSpOD takes into account classical parameters for the design of effective probes (probe

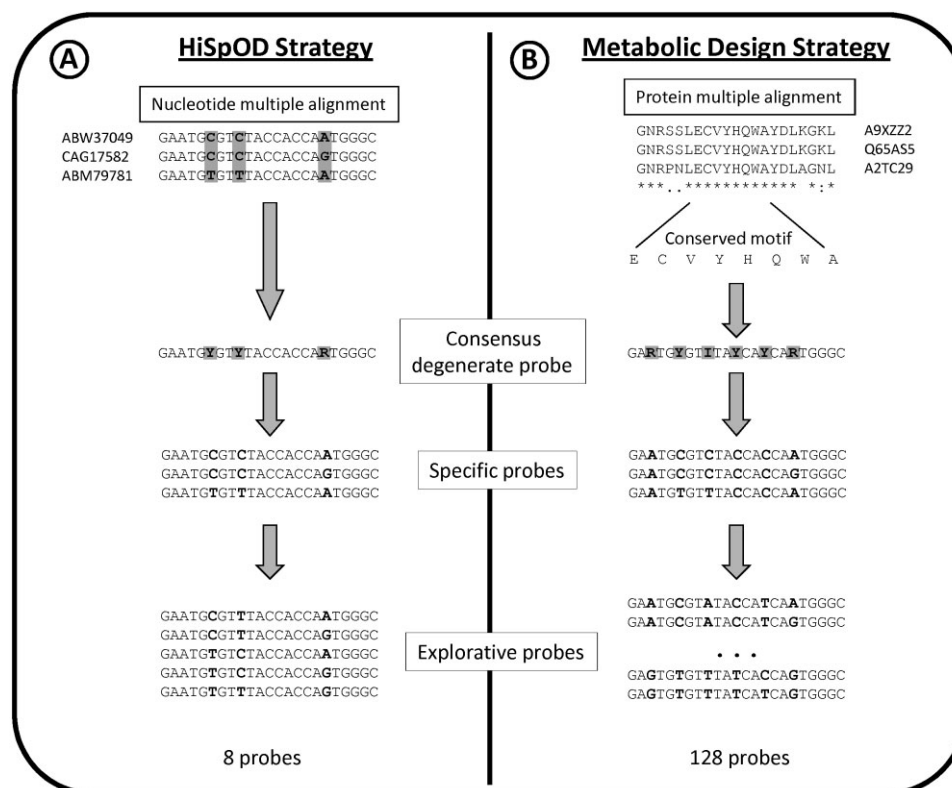


Fig. 2. Explorative probe design strategies implemented in (A) HiSpOD and (B) Metabolic Design software. The example shows probe design for the *bphA1c* gene encoding the Salicylate 1-hydroxylase alpha subunit involved in PAH degradation from three distinct *Sphingomonas* or *Sphingobium* species with both strategies.

length, Tm, GC%, complexity) and combines supplemental properties not considered by previous programmes. First, it can allow for the design of degenerate probes for gene families after multiple alignments of nucleic sequences belonging to the same gene family, and the production of consensus sequences. All combinations deduced from these degenerate probes are then divided into two groups. The first corresponds to specific probes for sequences available in databanks, and the second to explorative probes, which represent potential new signatures not corresponding to any previously described microorganisms (Fig. 2A). Both the probe sets covering the most likely gene sequence variants and those covering new combinations not yet deposited in databanks are created based on multiple mutation events already identified. Second, the specificity of all selected probes is checked against a large formatted database dedicated to microbial communities, the EnvExBase (Environmental Expressed sequences data-Base) composed of all coding DNA sequences (CDSs) from Prokaryotes (PRO), Fungi (FUN) and Environmental (ENV) taxonomic divisions of the EMBL databank, in order to limit cross-hybridizations. To validate this strategy, a microarray focusing on the genes involved in

chloroethene solvent biodegradation was developed as a model system and enabled the identification of active cooperation between *Sulfurospirillum* and *Dehalococcoides* populations in the decontamination of a polluted groundwater (Dugat-Bony *et al.*, 2011).

Use of protein sequence signatures for probe design

Unlike the strategies outlined above, a number of new strategies have been proposed to initiate probe design not from nucleic acid sequences, but from conserved peptidic regions, in order to survey all potential nucleic acid variants.

The first strategy based on this principle was described by Bontemps and colleagues (2005) and called CODEHMOP (for Consensus DEgenerate Hybrid Motif Oligonucleotide Probe). It comes from an adaptation of the CODEHOP (for Consensus DEgenerate Hybrid Oligonucleotide Primer) PCR primer design strategy, originally developed to identify distantly related genes encoding proteins that belong to known families (Rose *et al.*, 1998; 2003; Boyce *et al.*, 2009). In the CODEHMOP strategy, conserved amino acid motifs are identified from multiple alignments of protein sequences. Then, all possible

nucleic combinations (15–21 nucleotides) from the most highly conserved region (5–7 amino acids) of each protein motif are recreated and flanked by 5' and 3' fixed ends (12–15 nucleotides each), derived from the most frequent nucleotide at each position. The final probes are called 'hybrids', as they consist of a variable central core, to target a larger diversity, with some nucleic combinations not corresponding to any yet described sequences, and two fixed end sequences (available in databanks) added to increase probe length. The authors used this approach to design a prototype DNA array covering all described and undescribed *nodC* (nodulation gene) sequences in bacteria, and applied it to legume nodules (Bontemps *et al.*, 2005). This strategy allowed the authors to detect new *nodC* sequences exhibiting less than 74% identity with known sequences.

The application of the CODEHMOP strategy is limited by the fact that it is not implemented into a fully automated programme and no probe specificity test is incorporated. Nevertheless, this approach appears to be the most comprehensive way of encompassing the larger diversity of gene sequence variants potentially found for enzymes mediating a given function. Furthermore, Terrat and colleagues (2010) developed a new software programme called Metabolic Design, which ensures *in silico* reconstruction of metabolic pathways, the identification of conserved motifs from protein multiple alignments and the generation of efficient explorative probes through a simple convenient graphical interface. In this case, before the probe design stage, the user reconstructs the chosen metabolic pathway *in silico* with all substrates and products from each metabolic step. One reference enzyme for each of these steps is selected and its protein sequence extracted from a curated database (by default, Swiss-Prot), which is then used to retrieve all homologous proteins from complete databases (Swiss-Prot and TrEMBL). After selecting the most pertinent homologous sequences, they are aligned to begin the probe design stage. The amino acids are back-translated for each molecular site identified, taking into account all genetic code redundancy, to produce a degenerate nucleic consensus sequence. All degenerate probes that meet the criteria defined by the user are retained (probe length and maximal degeneracy). All the specific possible combinations for each degenerate probe are subsequently checked for potential cross-hybridizations against a representative database (i.e. EnvExBase as in the HiSpOD programme). Finally, an output file, listing all degenerate probes selected by the user, permits the deduction of all possible combinations and organizes them into specific probes and exploratory probes (Fig. 2B). The approach was validated by studying enzymes involved in the degradation of polycyclic aromatic hydrocarbons (Terrat *et al.*, 2010).

Towards circumventing microarray limitations

Despite the emergence of new design strategies, such as those presented above, the determination of a high-quality probe set appears to be crucial, especially in an environmental ecology context (Liebich *et al.*, 2006; Leparç *et al.*, 2009). Although explorative potential represents a major criterion for fingerprint determination, other parameters also impact considerably on probe sensitivity and specificity, and, therefore, require particular attention (Zhou, 2003; Wagner *et al.*, 2007).

Optimization of probe size criterion

Generally, POAs employed for microbial community analysis contain short probes (typically 24–25 mers) (Brodie *et al.*, 2006; Paliy *et al.*, 2009; Rajilic-Stojanovic *et al.*, 2009), whereas FGAs are built either with short (15–30 mers) (Bodrossy *et al.*, 2003; Stralis-Pavese *et al.*, 2004) or long oligonucleotides (40–70 mers) (Kane *et al.*, 2000; Religio *et al.*, 2002; He *et al.*, 2007). The main limitation of microarrays based on short oligonucleotide probes, therefore, is the need to use, in most cases, PCR-amplified targets to ensure enrichment and thereby increase sensitivity, but this also introduces an inherent PCR bias (Suzuki and Giovannoni, 1996; Peplies *et al.*, 2004; Vora *et al.*, 2004).

An alternative approach to design oligonucleotide probes, which combines excellent specificity with a potentially high sensitivity, is the use of the GoArrays strategy developed by Rimour and colleagues (2005) (software available at <http://g2im.u-clermont1.fr/serimour/goarrays.html>). In this approach, the oligonucleotide probe consists of the concatenation of two short subsequences that are complementary to disjointed regions of the target, with an insertion of a short random linker (e.g. 3–6 mer) (Fig. 3). This strategy has been shown to improve microarray efficiency for a wide range of applications (Rimour *et al.*, 2005; Zhou *et al.*, 2007; Pariset *et al.*, 2009; Kang *et al.*, 2010).

Specificity improvement using large databases

Because only a small portion of the natural microbial diversity has been identified, it is a major challenge to design appropriate probes specific to unique markers that do not cross-hybridize with similar unknown sequences (Chandler and Jarrell, 2005). Most of the currently available probe design software have been developed for non-environmental applications and performs specificity tests only against a reduced set of sequences, such as whole-genome data or specific sets of genes (Lemoine *et al.*, 2009). The study of microbial communities, however, requires dedicated databases that are as representative

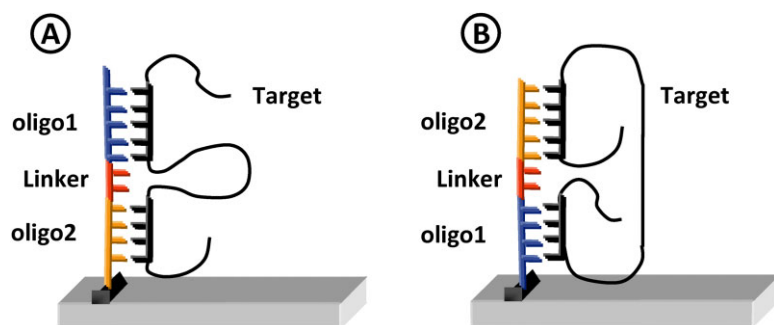


Fig. 3. Representation of the GoArrays strategy. In this strategy, two short oligonucleotide probes are concatenated with a random linker. Depending on the probes' positions, the target can form two kinds of stable loops during hybridization (A and B). (For the color version of this figure, please refer to the Web version of this article.)

as possible of all non-target sequences potentially present in environmental samples. GenBank (Benson *et al.*, 2011), European Nucleotide Archive (ENA) (Leinonen *et al.*, 2011) and the DNA Data Bank of Japan (DDBJ) (Kaminuma *et al.*, 2011) are the most complete nucleic sequence databases publicly available to perform specificity tests. Dealing with such databases, however, is too time-consuming for probe design task, and, in this instance not really appropriate as some subsets of these databases correspond to sequences from organisms such as *Metazoa*, which are typically not considered in microbial ecology. Furthermore, for studies focusing on particular biomarkers, other sequence information need not to be considered.

For example, within POAs, each probe must be specific with respect to all small subunit (SSU) rRNA sequences, which may be present in the sample during hybridization. Curated and dedicated secondary databases have been already constructed [RDP (Cole *et al.*, 2009), Greengenes (DeSantis *et al.*, 2006) and SILVA (Pruesse *et al.*, 2007)], assembling all SSU rRNA sequences described on public databases. The differences between these databases come from the construction and update pipelines that lead to distinct sizes: SILVA (Release 104) contains 1 304 069 16S rRNA sequences, RDP (Release 10) 1 545 680 and Greengenes (03/22/2011) 855 446. These large databases, therefore, are well adapted to phylogenetic probe design. PhylArray software (Milton *et al.*, 2007) was developed before these databases were publicly available, and, therefore, uses its own highly curated (full length and quality filtered) and automatically updated prokaryotic SSU rRNA database (122 337 sequences for the last release).

Because environmental FGAs target coding sequences (CDS), the database used for specificity tests must include all known CDSs that may be encountered in natural environments. To the best of our knowledge, EnvExBase (integrated in both HiSpOD and Metabolic Design programmes) is the first CDSs database dedicated to microbial ecology (Terrat *et al.*, 2010; Dugat-Bony *et al.*, 2011). For its construction, all annotated transcript sequences and their associated 5' and 3'

untranslated regions in all classes of EMBL Prokaryotes (PRO), Fungi (FUN) and Environmental (ENV) taxonomic divisions, were extracted and curated to remove bad-quality sequences. It represents a 9 129 323 sequence database.

The rapid growth of datasets, particularly environmental datasets, has led to an important increase in computational requirements coupled with a fundamental change in the way algorithms are conceived and designed [e.g. mpi-BLAST (Darling *et al.*, 2003)]. Consequently, parallel computing is essential, and algorithms must be deployed on large cluster infrastructures or computing grids, if specificity tests and alignments are to be performed with reasonable data processing times (Gardner *et al.*, 2006; Thorsen *et al.*, 2007).

Adaptation of the microarray format to the design strategy

Explorative design strategies targeting unknown sequences involve the use of degenerate probes (Bontemps *et al.*, 2005; Milton *et al.*, 2007; Terrat *et al.*, 2010; Dugat-Bony *et al.*, 2011). Consequently, the selected strategy will greatly influence the choice between the two major DNA microarray types (*ex situ* or *in situ*), the platform and the density (Dufva, 2005; Ehrenreich, 2006; Kawasaki, 2006). When using *in situ* synthesis microarrays, such as the Agilent, Affymetrix and NimbleGen platforms, all combinations resulting from a degenerate probe must be independently synthesized. This will exponentially increase the final number of probes for the array production (density). For instance, concerning the CODE-HMOP (Bontemps *et al.*, 2005) and Metabolic Design strategies (Terrat *et al.*, 2010), because the genetic code often involves degeneracy at the third position of each codon, a 24 mer probe (targeting a seven amino acid conserved motif) will generate at least 128 combinations (assuming a minimal degeneracy rate of two for each codon). This value will reach at least 131 072 for a 51 mer probe containing 17 degenerate positions. Conversely, *ex situ* platforms allow the degenerate probes (all combinations mixed together) to be spotted in the same location

Table 3. Characteristics of the main commercially distributed high-density microarrays.

Type of array	Technology	Probe length	Max features	Max plex
Spotted Arrays	Robot spotting Pre-made DNA	Any	~100 000	1
Affymetrix	Photolithography <i>in situ</i>	<100 mer (generally 25 mer)	~6 000 000	1
NimbleGen	Micro-mirrors <i>in situ</i>	<100 mer (generally 50–75 mer)	4 200 000	12
Agilent	Inkjet <i>in situ</i>	<100 mer (generally 25–60 mer)	1 000 000	8

on the array and consequently reduce the total amount of features.

Other user choices may also affect the final number of probes per array. Replication is crucial to achieve reliable data for microarrays (Spruill *et al.*, 2002). Multiple replicates of the same probe provide some back-up in case a feature cannot be evaluated due to technical artefacts, such as dye precipitations or dust particles. A statistical estimation has deduced that at least three replicates should be made (Lee *et al.*, 2000). Second, multiple probes per gene could be designed in order to increase confidence in the results (Loy *et al.*, 2002; Chou *et al.*, 2004) and to mask misleading signal variations whose causes (e.g. target secondary structure, probe folding, etc.) are not yet fully understood (Pozhitkov *et al.*, 2007). Third, some platforms, such as Affymetrix GeneChips, determine probe pairs where each probe ('match') is accompanied by a negative control with a single differing base in the middle of the probe ('mismatch probe') in order to discriminate between real signals and those due to non-specific hybridizations (Lipshutz *et al.*, 1999).

To address this problem of probe number, several commercial companies have proposed two major types of high-density microarrays whose main characteristics are described in Table 3: (i) *in situ* synthesized microarrays, distributed by Agilent (<http://www.chem.agilent.com>), NimbleGen (<http://www.nimblegen.com>) and Affymetrix (<http://www.affymetrix.com>), which can attain billions of probes and be physically divided into multi-arrays per slide (up to 12) to perform simultaneous analyses of several samples on a single experiment; and (ii) spotted microarrays [e.g. Arrayit (<http://www.arrayit.com>)] with a current printing capacity close to 100 000 features per microarray.

Concluding remarks and future directions

Assessing the extreme microbial diversity encountered in environmental samples represents an exciting challenge that could create a better understanding of microbial community functioning. Environmental DNA microarrays, with the opportunity to survey both known and unknown microorganisms through explorative probe design, are one of the most powerful approaches for achieving this goal. Future perspectives in this domain will be to systematically integrate this innovative concept into probe design

workflows, especially by offering the possibility to design degenerate probes targeting sequence clusters. Furthermore, to efficiently recognize signals due to unknown targets, it will be particularly useful to develop automatic procedures to analyse microarray data. In addition, using explorative probe design in sequence capture approaches that couple with NGS, such as those originally developed for direct selection of human genomic loci (Albert *et al.*, 2007), could also improve this gene characterization. Indeed, sequence capture elution products should allow the full identification and characterization of new taxa when using POAs or new protein coding genes with FGAs.

The constant increase in available sequences (Cochrane *et al.*, 2009) means that databases for specificity tests must be regularly updated. As a result, probe datasets must be re-computed as frequently as possible in order to take into account all deposited data. Nevertheless, assessing probe specificity against large databases is a time-consuming task. To overcome this problem, two complementary strategies could be employed: (i) Creation of databases specific to each ecological compartment. Usually, specificity tests are not performed against a suitable subset of sequences mainly due to lack of databases for microbial ecology. Depending on the environment studied it would be more relevant to perform these tests against reduced databanks dedicated to specific ecosystems (soil, marine, freshwater, gut, etc.).

(ii) Parallelization of probe design algorithms. Perspectives to limit computation time are based on exploiting the computational resources available using specialized frameworks such as Message Passing Interface (MPI) or heterogeneous systems including General-purpose Processing on Graphics Processing Units (GPGPU). With the recent development of extremely fast broadband networks, it has become possible to distribute the calculations at larger and larger scales over different geographical locations (Schadt *et al.*, 2010). Cluster, grid or emerging cloud computing are all examples of shared computing resources where probe design algorithms can be deployed. Being able to improve the bioinformatics tools applied to environmental microbiology through algorithm deployment on such shared computational resources, and combining them with automatic update pipelines, are two important challenges and strategies for the future of the field of molecular ecology.

Acknowledgments

This work was supported by the grant ID 2598 from the 'Agence De l'Environnement et de la Maîtrise de l'Energie' (ADEME, France); the Grant ANR-07-ECOT-005-05 for the programme PRECODD Evasol from 'Agence Nationale de la Recherche' (ANR, France); the Grant ANR-08-BIOENERGIES-0 for the programme BIOENERGIES AnaBio-H2 from 'Agence Nationale de la Recherche' (ANR, France); and the INSU-EC2CO programme from 'Centre National de la Recherche Scientifique' (CNRS, France). We thank David Tottey for reviewing the English version of the manuscript.

References

- Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., *et al.* (2007) Direct selection of human genomic loci by microarray hybridization. *Nat Methods* **4**: 903–905.
- Amann, R., Ludwig, W., and Schleifer, K. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.
- Ashelford, K.E., Weightman, A.J., and Fry, J.C. (2002) PRIM-ROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res* **30**: 3481–3489.
- Bae, J.W., and Park, Y.H. (2006) Homogeneous versus heterogeneous probes for microbial ecological microarrays. *Trends Biotechnol* **24**: 318–323.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2011) GenBank. *Nucleic Acids Res* **39**: D32–D37.
- Bers, K., Sniegowski, K., Albers, P., Breugelmans, P., Hendrickx, L., De Mot, R., and Springael, D. (2011) A molecular toolbox to estimate the number and diversity of *Variovorax* in the environment: application in soils treated with the phenylurea herbicide linuron. *FEMS Microbiol Ecol* **76**: 14–25.
- Blaskovic, D., and Barak, I. (2005) Oligo-chip based detection of tick-borne bacteria. *FEMS Microbiol Lett* **243**: 473–478.
- Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A., and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* **5**: 566–582.
- Boeckeaert, C., Vlaeminck, B., Fievez, V., Maignien, L., Dijkstra, J., and Boon, N. (2008) Accumulation of trans C18:1 fatty acids in the rumen after dietary algal supplementation is associated with changes in the *Butyrivibrio* community. *Appl Environ Microbiol* **74**: 6923–6930.
- van den Bogert, B., de Vos, W.M., Zoetendal, E.G., and Kleerebezem, M. (2011) Microarray Analysis and Barcoded Pyrosequencing Provide Consistent Microbial Profiles Depending on the Source of Human Intestinal Samples. *Appl Environ Microbiol* **77**: 2071–2080.
- Bommarito, S., Peyret, N., and SantaLucia, J., Jr (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res* **28**: 1929–1934.
- Bontemps, C., Golfier, G., Gris-Liebe, C., Carrere, S., Talini, L., and Boivin-Masson, C. (2005) Microarray-based detection and typing of the *Rhizobium* nodulation gene *nodC*: potential of DNA arrays to diagnose biological functions of interest. *Appl Environ Microbiol* **71**: 8042–8048.
- Boyce, R., Chilana, P., and Rose, T.M. (2009) iCODEHOP: a new interactive program for designing CONsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. *Nucleic Acids Res* **37**: W222–W228.
- Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B., and DeRisi, J.L. (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* **4**: R9.
- Brodie, E.L., Desantis, T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L., *et al.* (2006) Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* **72**: 6288–6298.
- Brodie, E.L., DeSantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M., and Andersen, G.L. (2007) Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci USA* **104**: 299–304.
- Candela, M., Consolandi, C., Severgnini, M., Biagi, E., Castiglioni, B., Vitali, B., *et al.* (2010) High taxonomic level fingerprint of the human intestinal microbiota by ligase detection reaction – universal array approach. *BMC Microbiol* **10**: 116.
- Castiglioni, B., Rizzi, E., Frosini, A., Sivonen, K., Rajaniemi, P., Rantala, A., *et al.* (2004) Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria. *Appl Environ Microbiol* **70**: 7161–7172.
- Chandler, D.P., and Jarrell, A.E. (2005) Taking arrays from the lab to the field: trying to make sense of the unknown. *Biotechniques* **38**: 591–600.
- Chou, C.C., Chen, C.H., Lee, T.T., and Peck, K. (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res* **32**: e99.
- Chung, W.-H., Rhee, S.-K., Wan, X.-F., Bae, J.-W., Quan, Z.-X., and Park, Y.-H. (2005) Design of long oligonucleotide probes for functional gene detection in a microbial community. *Bioinformatics* **21**: 4092–4100.
- Claesson, M.J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J.R., Smidt, H., *et al.* (2009) Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. *PLoS ONE* **4**: e6669.
- Cochrane, G., Akhtar, R., Bonfield, J., Bower, L., Demiralp, F., Faruque, N., *et al.* (2009) Petabyte-scale innovations at the European Nucleotide Archive. *Nucleic Acids Res* **37**: D19–D25.
- Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.
- Cruz-Martinez, K., Suttle, K.B., Brodie, E.L., Power, M.E.,

- Andersen, G.L., and Banfield, J.F. (2009) Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *ISME J* **3**: 738–744.
- Curtis, T.P., Sloan, W.T., and Scannell, J.W. (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99**: 10494–10499.
- Darling, A., Carey, L., and Feng, W. (2003) The Design, Implementation, and Evaluation of mpiBLAST. In: *4th International Conference on Linux Clusters: The HPC Revolution 2003*. San Jose, California.
- DeAngelis, K.M., Brodie, E.L., DeSantis, T.Z., Andersen, G.L., Lindow, S.E., and Firestone, M.K. (2009) Selective progressive response of soil microbial community to wild oat roots. *ISME J* **3**: 168–178.
- Delmont, T.O., Robe, P., Cecillon, S., Clark, I.M., Constanças, F., Simonet, P., et al. (2011) Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* **77**: 1315–1324.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al. (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.
- DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubieta, I.X., Piceno, Y.M., and Andersen, G.L. (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* **53**: 371–383.
- Duc, L., Neuenschwander, S., Rehrauer, H., Wagner, U., Sobek, J., Schlappbach, R., and Zeyer, J. (2009) Development and experimental validation of a nifH oligonucleotide microarray to study diazotrophic communities in a glacier forefield. *Environ Microbiol* **11**: 2179–2189.
- Dufva, M. (2005) Fabrication of high quality microarrays. *Biomol Eng* **22**: 173–184.
- Dugat-Bony, E., Missaoui, M., Peyretailade, E., Biderre-Petit, C., Bouzid, O., Gouinaud, C., et al. (2011) HiSpOD: probe design for functional DNA microarrays. *Bioinformatics* **27**: 641–648.
- Ehrenreich, A. (2006) DNA microarray technology for the microbiologist: an overview. *Appl Microbiol Biotechnol* **73**: 255–273.
- Emrich, S.J., Lowe, M., and Delcher, A.L. (2003) PROBER: A web-based software tool for selecting optimal DNA oligos. *Nucleic Acids Res* **31**: 3746–3750.
- Evertsz, E.M., Au-Young, J., Ruvolo, M.V., Lim, A.C., and Reynolds, M.A. (2001) Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. *Biotechniques* **31**: 1182, 1184, 1186 passim.
- Feldhaar, H., Straka, J., Krischke, M., Berthold, K., Stoll, S., Mueller, M.J., and Gross, R. (2007) Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biol* **5**: 48.
- Feng, S., and Tillier, E.R.M. (2007) A fast and flexible approach to oligonucleotide probe design for genomes and gene families. *Bioinformatics* **23**: 1195–1202.
- Flanagan, J.L., Brodie, E.L., Weng, L., Lynch, S.V., Garcia, O., Brown, R., et al. (2007) Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with *Pseudomonas aeruginosa*. *J Clin Microbiol* **45**: 1954–1962.
- Franke-Whittle, I.H., Goberna, M., Pfister, V., and Insam, H. (2009) Design and development of the ANAEROCHIP microarray for investigation of methanogenic communities. *J Microbiol Methods* **79**: 279–288.
- Fraune, S., Augustin, R., Anton-Erxleben, F., Wittlieb, J., Gelhaus, C., Klimovich, V.B., et al. (2010) In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proc Natl Acad Sci USA* **107**: 18067–18072.
- Gans, J., Wolinsky, M., and Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**: 1387–1390.
- Gardner, M.K., Feng, W.-C., Archuleta, J., Lin, H., and Ma, X. (2006) Parallel genomic sequence-searching on an ad-hoc grid: experiences, lessons learned, and implications. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. SC Conference (ed.). Tampa, Florida: ACM, pp. 22.
- Gentry, T., Wickham, G., Schadt, C., He, Z., and Zhou, J. (2006) Microarray Applications in Microbial Ecology Research. *Microb Ecol* **52**: 159–175.
- Gittel, A., Sorensen, K.B., Skovhus, T.L., Ingvorsen, K., and Schramm, A. (2009) Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil reservoirs with and without nitrate treatment. *Appl Environ Microbiol* **75**: 7086–7096.
- Guerrero, R., and Berlanga, M. (2006) Life's unity and flexibility: the ecological link. *Int Microbiol* **9**: 225–235.
- He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., et al. (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67–77.
- He, Z., Deng, Y., Van Nostrand, J.D., Tu, Q., Xu, M., Hemme, C.L., et al. (2010) GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4**: 1167–1179.
- He, Z., Van Nostrand, J.D., Deng, Y., and Zhou, J.Z. (2011) Development and applications of functional gene microarrays in the analysis of the functional diversity, composition, and structure of microbial communities. *Front Environ Sci Engin China* **5**: 1–20.
- He, Z.L., Van Nostrand, J.D., Wu, L.Y., and Zhou, J.Z. (2008) Development and application of functional gene arrays for microbial community analysis. *Trans Nonferrous Met Soc China* **18**: 1319–1327.
- Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., et al. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**: 342–347.
- Huyghe, A., Francois, P., Charbonnier, Y., Tangomo-Bento, M., Bonetti, E.-J., Paster, B.J., et al. (2008) Novel Microarray Design Strategy To Study Complex Bacterial Communities. *Appl Environ Microbiol* **74**: 1876–1885.
- Kaminuma, E., Kosuge, T., Kodama, Y., Aono, H., Mashima, J., Gojobori, T., et al. (2011) DDBJ progress report. *Nucleic Acids Res* **39**: D22–D27.
- Kane, M.D., Jatke, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., and Madore, S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* **28**: 4552–4557.
- Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and McSweeney, C.S. (2010) Dysbiosis of

- fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* **16**: 2034–2042.
- Kawasaki, E.S. (2006) The end of the microarray Tower of Babel: will universal standards lead the way? *J Biomol Tech* **17**: 200–206.
- Kelly, J.J., Siripong, S., McCormack, J., Janus, L.R., Urakawa, H., El Fantroussi, S., *et al.* (2005) DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. *Water Res* **39**: 3229–3238.
- Kelly, L.C., Cockell, C.S., Piceno, Y.M., Andersen, G.L., Thorsteinsson, T., and Marteinson, V. (2010) Bacterial diversity of weathered terrestrial Icelandic volcanic glasses. *Microb Ecol* **60**: 740–752.
- Klitgaard, K., Boye, M., Capion, N., and Jensen, T.K. (2008) Evidence of multiple *Treponema* phylotypes involved in bovine digital dermatitis as shown by 16S rRNA gene analysis and fluorescence in situ hybridization. *J Clin Microbiol* **46**: 3012–3020.
- Koltai, H., and Weingarten-Baror, C. (2008) Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction. *Nucleic Acids Res* **36**: 2395–2405.
- Kyselkova, M., Kopecky, J., Felfoldi, T., Cermak, L., Omelka, M., Grundmann, G.L., *et al.* (2008) Development of a 16S rRNA gene-based prototype microarray for the detection of selected actinomycetes genera. *Antonie Van Leeuwenhoek* **94**: 439–453.
- Lee, M.L., Kuo, F.C., Whitmore, G.A., and Sklar, J. (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* **97**: 9834–9839.
- Lehner, A., Loy, A., Behr, T., Gaenge, H., Ludwig, W., Wagner, M., and Schleifer, K.H. (2005) Oligonucleotide microarray for identification of *Enterococcus* species. *FEMS Microbiol Lett* **246**: 133–142.
- Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdono-Tarraga, A., Cheng, Y., *et al.* (2011) The European Nucleotide Archive. *Nucleic Acids Res* **39**: D28–D31.
- Lemoine, S., Combes, F., and Le Crom, S. (2009) An evaluation of custom microarray applications: the oligonucleotide design challenge. *Nucleic Acids Res* **37**: 1726–1739.
- Leparc, G.G., Tuchler, T., Striedner, G., Bayer, K., Sykacek, P., Hofacker, I.L., and Kreil, D.P. (2009) Model-based probe set optimization for high-performance microarrays. *Nucleic Acids Res* **37**: e18.
- Letowski, J., Brousseau, R., and Masson, L. (2004) Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *J Microbiol Methods* **57**: 269–278.
- Li, F., and Stormo, G.D. (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* **17**: 1067–1076.
- Li, X., He, Z., and Zhou, J. (2005) Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucleic Acids Res* **33**: 6114–6123.
- Liebich, J., Schadt, C.W., Chong, S.C., He, Z., Rhee, S.K., and Zhou, J. (2006) Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. *Appl Environ Microbiol* **72**: 1688–1691.
- Liles, M.R., Turkmen, O., Manske, B.F., Zhang, M., Rouillard, J.-M., George, I., *et al.* (2010) A phylogenetic microarray targeting 16S rRNA genes from the bacterial division Acidobacteria reveals a lineage-specific distribution in a soil clay fraction. *Soil Biol Biochem* **42**: 739–747.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nat Genet* **21**: 20–24.
- Loy, A., and Bodrossy, L. (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays. *Clin Chim Acta* **363**: 106–119.
- Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J., *et al.* (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* **68**: 5064–5081.
- Loy, A., Schulz, C., Lucker, S., Schopfer-Wendels, A., Stoeker, K., Baranyi, C., *et al.* (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order 'Rhodocyclales'. *Appl Environ Microbiol* **71**: 1373–1386.
- Ludwig, W., Amann, R., Martinez-Romero, E., Schönhuber, W., Bauer, S., Neef, A., and Schleifer, K.-H. (1998) rRNA based identification and detection systems for rhizobia and other bacteria. *Plant Soil* **204**: 1–19.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Kumar, Y., *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Marcelino, L.A., Backman, V., Donaldson, A., Steadman, C., Thompson, J.R., Preheim, S.P., *et al.* (2006) Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. *Proc Natl Acad Sci USA* **103**: 13629–13634.
- Milton, C., Rimour, S., Missaoui, M., Biderre, C., Barra, V., Hill, D., *et al.* (2007) PhylArray: phylogenetic probe design algorithm for microarray. *Bioinformatics* **23**: 2550–2557.
- Muhling, M., Woolven-Allen, J., Murrell, J.C., and Joint, I. (2008) Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J* **2**: 379–392.
- Neufeld, J.D., Mohn, W.W., and de Lorenzo, V. (2006) Composition of microbial communities in hexachlorocyclohexane (HCH) contaminated soils from Spain revealed with a habitat-specific microarray. *Environ Microbiol* **8**: 126–140.
- Nordberg, E.K. (2005) YODA: selecting signature oligonucleotides. *Bioinformatics* **21**: 1365–1370.
- Øvreås, L. (2000) Population and community level approaches for analysing microbial diversity in natural environments. *Ecol Lett* **3**: 236–251.
- Pace, N.R. (1997) A Molecular View of Microbial Diversity and the Biosphere. *Science* **276**: 734–740.
- Paliy, O., Kenche, H., Abernathy, F., and Michail, S. (2009) High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. *Appl Environ Microbiol* **75**: 3572–3579.
- Palmer, C., Bik, E.M., Eisen, M.B., Eckburg, P.B., Sana, T.R., Wolber, P.K., *et al.* (2006) Rapid quantitative profiling

- of complex microbial populations. *Nucleic Acids Res* **34**: e5.
- Pariset, L., Chillemi, G., Bongiorno, S., Romano Spica, V., and Valentini, A. (2009) Microarrays and high-throughput transcriptomic analysis in species with incomplete availability of genomic sequences. *Nat Biotechnol* **25**: 272–279.
- Peplies, J., Lau, S.C.K., Pernthaler, J., Amann, R., and Glöckner, F.O. (2004) Application and validation of DNA microarrays for the 16S rRNA-based analysis of marine bacterioplankton. *Environ Microbiol* **6**: 638–645.
- Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Staehler, P., et al. (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res* **34**: e66.
- Pozhitkov, A.E., Tautz, D., and Noble, P.A. (2007) Oligonucleotide microarrays: widely applied – poorly understood. *Brief Funct Genomic Proteomic* **6**: 141–148.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.
- Quince, C., Curtis, T.P., and Sloan, W.T. (2008) The rational exploration of microbial diversity. *ISME J* **2**: 997–1006.
- Rajilic-Stojanovic, M., Heilig, H.G., Molenaar, D., Kajander, K., Surakka, A., Smidt, H., and de Vos, W.M. (2009) Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* **11**: 1736–1751.
- Rastogi, G., Osman, S., Vaishampayan, P.A., Andersen, G.L., Stetler, L.D., and Sani, R.K. (2010a) Microbial diversity in uranium mining-impacted soils as revealed by high-density 16S microarray and clone library. *Microb Ecol* **59**: 94–108.
- Rastogi, G., Osman, S., Kukkadapu, R., Engelhard, M., Vaishampayan, P.A., Andersen, G.L., and Sani, R.K. (2010b) Microbial and mineralogical characterizations of soils collected from the deep biosphere of the former Homestake gold mine, South Dakota. *Microb Ecol* **60**: 539–550.
- Rastogi, G., Barua, S., Sani, R.K., and Peyton, B.M. (2011) Investigation of Microbial Populations in the Extremely Metal-Contaminated Coeur d'Alene River Sediments. *Microb Ecol* **62**: 1–13.
- Religio, A., Schwager, C., Richter, A., Ansorge, W., and Valcarcel, J. (2002) Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res* **30**: e51.
- Reymond, N., Charles, H., Duret, L., Calevro, F., Beslon, G., and Fayard, J.M. (2004) ROSO: optimizing oligonucleotide probes for microarrays. *Bioinformatics* **20**: 271–273.
- Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X., and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70**: 4303–4317.
- Rimour, S., Hill, D., Milton, C., and Peyret, P. (2005) GoArrays: highly dynamic and efficient microarray probe design. *Bioinformatics* **21**: 1094–1103.
- Rinta-Kanto, J.M., Burgmann, H., Gifford, S.M., Sun, S., Sharma, S., del Valle, D.A., et al. (2011) Analysis of sulfur-related transcription by *Roseobacter* communities using a taxon-specific functional gene microarray. *Environ Microbiol* **13**: 453–467.
- Roh, S.W., Abell, G.C.J., Kim, K.-H., Nam, Y.-D., and Bae, J.-W. (2010) Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends Biotechnol* **28**: 291–299.
- Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrovski, S., McCallum, C.M., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**: 1628–1635.
- Rose, T.M., Henikoff, J.G., and Henikoff, S. (2003) CODEHOP (CONsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res* **31**: 3763–3766.
- Rouillard, J.M., Zuker, M., and Gulari, E. (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* **31**: 3057–3062.
- Rusch, A., and Amend, J.P. (2004) Order-specific 16S rRNA-targeted oligonucleotide probes for (hyper)thermophilic archaea and bacteria. *Extremophiles* **8**: 357–366.
- Sagaram, U.S., DeAngelis, K.M., Trivedi, P., Andersen, G.L., Lu, S.E., and Wang, N. (2009) Bacterial diversity analysis of Huanglongbing pathogen-infected citrus, using Phylo-Chip arrays and 16S rRNA gene clone library sequencing. *Appl Environ Microbiol* **75**: 1566–1574.
- Sanguin, H., Sarniguet, A., Gazengel, K., Moenne-Loccoz, Y., and Grundmann, G.L. (2009) Rhizosphere bacterial communities associated with disease suppressiveness stages of take-all decline in wheat monoculture. *New Phytol* **184**: 694–707.
- SantaLucia, J., Jr, Allawi, H.T., and Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**: 3555–3562.
- Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L., and Nolan, G.P. (2010) Computational solutions to large-scale data management and analysis. *Nat Rev Genet* **11**: 647–657.
- Schliep, A., and Rahmann, S. (2006) Decoding non-unique oligonucleotide hybridization experiments of targets related by a phylogenetic tree. *Bioinformatics* **22**: e424–e430.
- Schönmann, S., Loy, A., Wimmersberger, C., Sobek, J., Aquino, C., Vandamme, P., et al. (2009) 16S rRNA gene-based phylogenetic microarray for simultaneous identification of members of the genus *Burkholderia*. *Environ Microbiol* **11**: 779–800.
- Severgnini, M., Cremonesi, P., Consolandi, C., Caredda, G., De Bellis, G., and Castiglioni, B. (2009) ORMA: a tool for identification of species-specific variations in 16S rRNA gene and oligonucleotides design. *Nucleic Acids Res* **37**: e109.
- Shawkey, M.D., Firestone, M.K., Brodie, E.L., and Beissinger, S.R. (2009) Avian incubation inhibits growth and diversification of bacterial assemblages on eggs. *PLoS ONE* **4**: e4522.

- Simon, C., and Daniel, R. (2009) Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol* **85**: 265–276.
- Spruill, S.E., Lu, J., Hardy, S., and Weir, B. (2002) Assessing sources of variability in microarray gene expression data. *Biotechniques* **33**: 916–920. 922–913.
- Stralis-Pavese, N., Sessitsch, A., Weilharter, A., Reichenauer, T., Riesing, J., Csontos, J., *et al.* (2004) Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. *Environ Microbiol* **6**: 347–363.
- Suzuki, M.T., and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625–630.
- Taroncher-Oldenburg, G., Griner, E.M., Francis, C.A., and Ward, B.B. (2003) Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl Environ Microbiol* **69**: 1159–1171.
- Terrat, S., Peyretailade, E., Goncalves, O., Dugat-Bony, E., Gravelat, F., Mone, A., *et al.* (2010) Detecting variants with Metabolic Design, a new software tool to design probes for explorative functional DNA microarray development. *BMC Bioinformatics* **11**: 478.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Thorsen, O., Smith, B., Sosa, C.P., Jiang, K., Lin, H., Peters, A., and Feng, W.-C. (2007) Parallel genomic sequence-search on a massively parallel system. In *Proceedings of the 4th International Conference on Computing Frontiers*. ACM (ed.). Ischia, Italy: ACM, pp. 59–68.
- Tiquia, S.M., Wu, L., Chong, S.C., Passovets, S., Xu, D., Xu, Y., and Zhou, J. (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* **36**: 664–670.
- Torsvik, V., and Øvreås, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240–245.
- Vora, G.J., Meador, C.E., Stenger, D.A., and Andreadis, J.D. (2004) Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl Environ Microbiol* **70**: 3047–3054.
- Wagner, M., Smidt, H., Loy, A., and Zhou, J. (2007) Unravelling Microbial Communities with DNA-Microarrays: Challenges and Future Directions. *Microb Ecol* **53**: 498–506.
- Wang, X., and Seed, B. (2003) Selection of oligonucleotide probes for protein coding sequences. *Bioinformatics* **19**: 796–802.
- Weinert, N., Piceno, Y., Ding, G.C., Meincke, R., Heuer, H., Berg, G., *et al.* (2011) PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS Microbiol Ecol* **75**: 497–506.
- Wernersson, R., and Nielsen, H.B. (2005) OligoWiz 2.0 – integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res* **33**: W611–W615.
- Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* **95**: 6578–6583.
- Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmarski, T.A., and Andersen, G.L. (2002) High-Density Microarray of Small-Subunit Ribosomal DNA Probes. *Appl Environ Microbiol* **68**: 2535–2541.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87**: 4576–4579.
- Xu, J. (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* **15**: 1713–1731.
- Yergeau, E., Schoondermark-Stolk, S.A., Brodie, E.L., Dejean, S., DeSantis, T.Z., Goncalves, O., *et al.* (2009) Environmental microarray analyses of Antarctic soil microbial communities. *ISME J* **3**: 340–351.
- Yu, Y., Lee, C., Kim, J., and Hwang, S. (2005) Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng* **89**: 670–679.
- Zhou, J. (2003) Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* **6**: 288–294.
- Zhou, Z., Dou, Z.-X., Zhang, C., Yu, H.-Q., Liu, Y.-J., Zhang, C.-Z., and Cao, Y.-J. (2007) A strategy to optimize the oligo-probes for microarray-based detection of viruses. *Virol Sin* **22**: 326–335.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.