Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biosensors and Bioelectronics 20 (2004) 668–683
BIOSENSORS & BIOELECTRONICS
www.elsevier.com/locate/bios

# Potential of DNA microarrays for developing parallel detection tools (PDTs) for microorganisms relevant to biodefense and related research needs

Syed A. Hashsham[a,b,*], Lukas M. Wick[c], Jean-Marie Rouillard[e], Erdogan Gulari[e], James M. Tiedje[b,d]

[a] *Department of Civil and Environmental Engineering, Michigan State University, A 126 Research Complex-Engineering, East Lansing, MI 48824, USA*
[b] *The Center for Microbial Ecology, Michigan State University, 540 Plant and Soil Science Building, East Lansing, MI 48824, USA*
[c] *National Food Safety and Toxicology Center, Michigan State University, 165 Food Safety and Toxicology Building, East Lansing, MI 48824, USA*
[d] *Department of Crop and Soil Sciences, Michigan State University, East Lansing, MI 48824, USA*
[e] *Department of Chemical Engineering, The University of Michigan, 3302 G.G. Brown Building, Ann Arbor, MI 48109, USA*

Available online 2 September 2004

**Abstract**

Development of parallel detection tools using microarrays is critically reviewed in view of the need for screening multiple microorganisms in a single test. Potential research needs with respect to probe design and specificity, validation, sample concentration, selective target enrichment and amplification, and data analysis are discussed. Data illustrating selected probe design issues for detecting multiple targets in mixed microbial systems is presented. Challenges with respect to cost, time, and ease of use compared to other methods are also summarized.


*Keywords:* DNA microarrays; Parallel detection tools; Biodefense

## 1. Background

The National Academy report on countering bioterrorism (NRC, 2002) identifies intelligence, detection, surveillance, and diagnosis as key elements of biodefense. It states that "emphasis should be on defense, simply because proliferation of biological weapons is difficult to control." It further suggests that the science and technology community should be focused on defenses against biological weapons and that the means to do so "include environmental detection of biological agents together with pre-clinical, clinical, and agricultural surveillance and diagnosis." The total number of known microorganisms (bacteria, eukaryotes, and viruses) that are of interest to various federal agencies responsible for the safety of air, water, food, animals, and agricultural products runs into the hundreds. Hence, development of parallel detection tools (PDTs) capable of rapidly and economically identifying a broad spectrum, if not all, of the microorganisms relevant to a given matrix (air, water, soil, food, plant, paper, manure, tissue, body fluid, etc.) is critical.

Such PDTs will most likely employ an array of distinguishing genetic and functional signatures (e.g., virulence factors) collected from all the microorganisms of interest to identify the most likely candidate(s) present in a sample of interest. Subsequent analysis of the organism(s) by existing approaches, such as culturing and sequencing, can provide the final confirmation. There are several advantages in adapting such a two-tiered approach for microbial detection related to biodefense. Paramount among which is the elimination of the initial guesswork. There may not be enough time to carry out traditional time-consuming tests one by one to identify the potential threat(s). PDTs will also provide better resolution with respect to the strain in cases where a sufficient number of genetic sequences exist. It will have a lower cost and time of testing in comparison to a battery of individual tests. A

* Corresponding author. Tel.: +1 517 355 8241
*E-mail address:* hashsham@egr.msu.edu (S.A. Hashsham).

0956-5663/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bios.2004.06.032

high throughput and parallel approach for microbial detection using genetic and functional signatures has the additional advantage of being able to detect dangerous elements without the need to identify the microorganisms. This capability may be able to address the concerns associated with the intentional or unintentional mobility of the virulent elements. It has also been identified as a critical need for biodefense (Enserink, 2001; Nass, 2001; Talbot, 2001).

Numerous tools to detect single microorganisms utilizing antibodies, gene sequences, or functional signatures already exist for many of the problematic pathogens (see www.aoac.org). Tools to detect those microorganisms that routinely impact the safety of commercial products have received the most attention. For example, hundreds of commercial kits are available to detect *Escherichia coli* O157:H7 and more are being developed to reduce time and cost. Other microbial pathogens have only a few extremely specialized and time-consuming (up to a week) methods available for their detection. It is evident that any parallel approach developed to screen for hundreds of microorganisms together will initially be less competitive compared to some well-established kits developed for a single or a few microorganisms. PDTs will most likely not compete in terms of cost when detection of a single known microorganism is of interest and for which individual detection kits are available. At the current cost of hundreds of dollars per test, they will also be less competitive to multiplex polymerase chain reaction approaches if detection of only a dozen or so organisms is needed. The real advantages of PDTs lie in their application to instances where broad-spectrum detection is a necessity, e.g., when the suspect agent is unknown and screening must be done for a large number of microorganisms or when strain level fingerprinting is necessary for source tracking based on multiple genetic signatures. They may also be useful to characterize the type of microorganisms present as a background.

High throughput DNA microarrays or biochips provide one of the best platforms for developing PDTs. Developed in the 1990s, and initially used mainly for gene expression analysis (Schena et al., 1995; Lockhart et al., 1996), microarrays are now routinely employed for comparative genomics (McCluskey et al., 2002; Schoolnik, 2002; Wells and Bennik, 2003), microbial detection (Chizhikov et al., 2001; del Cerro et al., 2002; Loy et al., 2002; Wilson et al., 2002a, 2002b; Busch et al., 2003; Hashsham et al., 2003; Wu et al., 2003), single nucleotide polymorphisms analysis (Huber et al., 2002; Iwasaki et al., 2002; Lindroos et al., 2002; Urakawa et al., 2002), sequencing (Drmanac et al., 1998), and many other applications. An oligonucleotide DNA microarray consists of short (usually 20- to 70-mer) (Lipshutz et al., 1999; Kane et al., 2000; Hughes et al., 2001) signature sequences (called probes) on a solid substrate (usually glass or silicon wafers). These probes are either synthesized on the substrate or attached mechanically, using DNA printers after synthesis in a DNA synthesizer. There are DNA microarrays that use long (100- to 500-mer or longer) probes produced by polymerase chain reaction but they are considered less useful for parallel microbial detection due to loss in specificity (The chipping forecast, 1999; Kane et al., 2000). They may be more useful, however, for strain identification of a particular species by microarray-based comparison of gene sequences (McCluskey et al., 2002; Schoolnik, 2002; Wells and Bennik, 2003). Probe design is the critical first step (after making a list of the microorganisms to detect) in the development of PDTs. After a DNA chip is designed and fabricated, it must be validated for the targets and matrices of interest in the real world. Validation is accomplished by extraction of the DNA or RNA from the sample, labeling it with a fluorescent dye (using separate protocols for each) and hybridizing the labeled DNA or RNA with the probes on the biochip. The hybridized signals are then read, using a laser scanner to determine quantitative abundance or presence/absence of the targets. Data analysis is an important component in determining presence/abundance of the targets and spans all aspects of PDT development, from probe design to signal detection.

Many DNA microarray platforms are available now to develop PDTs. Some synthesize the probes in situ, using light and various photolithography techniques (e.g., Affymetrix (Lockhart et al., 1996); NimbleGen (Hasan et al., 1997); Xeotron (Gao et al., 2001); and Febit (Baum et al., 2003)). Others attach the probe to the substrate, using DNA printers after synthesizing it in a DNA synthesizer (e.g., Agilent (Hughes et al., 2001), gel-pad arrays (El Fantroussi et al., 2003), and glass slide arrays (Schena et al., 1995)). Examples of pilot-scale efforts exist for most of them (Urakawa et al., 2002; Wilson et al., 2002a, 2002b; Hashsham et al., 2003). As in other microbial detection methods, the most basic issues including selection of the intended targets and genes for probe design, sample concentration, platform associated advantages and limitations, validation and data analysis, and ultimate use must be addressed in applying microarray-based tools (Fig. 1). Detection limit, specificity, quantification of abundance, dynamic range, reliability, speed, cost, sample concentration, and approaches for target amplification and enrichment must all be optimized for performance. Since only a single or a few protocols must be used to detect multiple microorganisms, optimizing so many parameters is a daunting task.

Concentration of microbes (often termed sample concentration) for developing most of these tools is an equally daunting task. This is because often a small number of the target microorganisms present in large quantities of water (1–1000 L), food (1–100 g), air (10–100 s of $m^3$), or other media (soil, body fluid, manure, tissue, paper, etc.) must be concentrated into a few microliters without concentrating the inhibitory materials. Only then can the sample be applied to a PDT for hybridization and detection. Numerous optimized protocols for sample concentration already exist for protozoa (Sturbaum et al., 2002), viruses (Scott et al., 2002), and some bacteria (Safarikova and Safarik, 2001) using mainly membrane filtration (United States Environmental Protection Agency, 2001a, 2001b) and immunomagnetic separation processes (Safarikova and Safarik, 2001; Guy et al.,

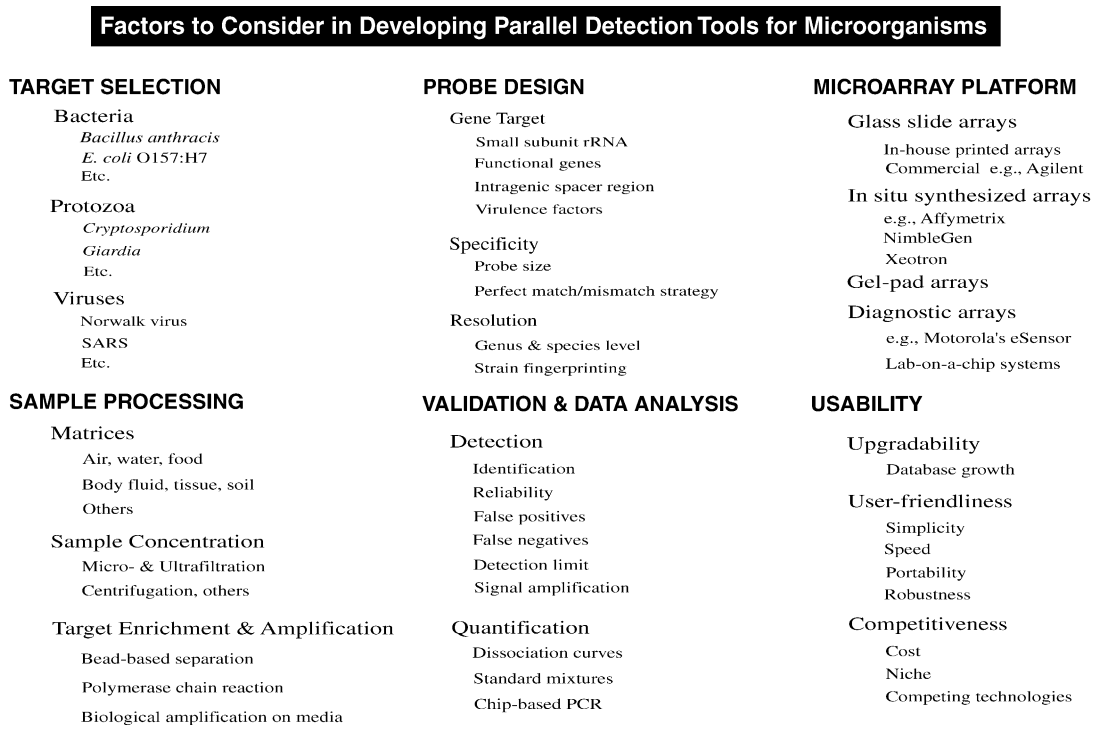


Fig. 1. Many factors must be considered in developing tools for parallel microbial detection. Sample processing and validation are the rate limiting steps.

2003; Kuczynska et al., 2003). Inconsistent and poor recovery of the target organisms and the need to process larger volumes of high turbidity waters are the leading research problems for sample concentration (Robertson and Gjerde, 2000). A comprehensive approach to enrich and/or concentrate all types of microorganisms in a given matrix would be ideal. Efforts to concentrate protozoa and viruses together have been made but recovery of all the three groups of microorganisms is less common because concentration methods to recover bacteria rely heavily upon species-specific antibodies and media enrichment. For sequence-based detection, amplification and/or further magnetic bead based enrichment of the target sequence is also an option (Chen and Griffiths, 2001). The success of this step depends upon the availability of suitable primers/probes and the absence of inhibitory materials in the sample matrix. Overcoming technical difficulties as well as improving speed, size, simplicity, and economics of sample concentration and target enrichment are key areas that will benefit from renewed research interest.

Detection limit and quantification of abundance are also important issues in developing PDTs. Sequence-based detection of a few pathogens in the presence of a large number of harmless microorganisms in an environmental matrix is a challenging task due to the use of miniaturized platforms with limited quantity of probes. At present the lower detection limit for a target organism using DNA microarrays and short oligonucleotides probes is approximately 1% by DNA abundance (Denef et al., 2003). This obviously is inferior to many technologies that provide much better sensitivity but are not as high throughput as microarrays. Sample concentration and target enrichment do serve as key mechanisms to improve the detection limit (Chen and Griffiths, 2001; Safarikova and Safarik, 2001; Sturbaum et al., 2002; Bopp et al., 2003; Kuczynska et al., 2003). However, concentration may also reduce signal intensity due to loss of target material. Signal amplification by improvements in methods for target labeling (Yguerabide and Yguerabide, 2001; Bao et al., 2002) and in the synthesis of microarray probe density and fidelity are some of the other areas that may improve detection limit. The overall improvement achievable by an appropriate combination of the existing protocols is expected to be insufficient for the current needs.

Quantification of the detected targets on microarrays has received the least attention because most of the traditional applications of this technology depend upon the ratio of the signals obtained for two samples and absolute quantification was seldom needed. For parallel microbial detection, absolute quantification of the target microorganisms is a necessary objective in many cases. Two approaches to accomplish this, one based on the dissociation curve of the probe and target duplex (El Fantroussi et al., 2003) and the other using standard mixtures to obtain ratios, are currently possible and have received limited attention for developing PDTs.

Some additional complexities that may be considered specific to microarray-based methods for parallel microbial detection include but are not limited to the following: (i) strain-level resolution in probe design, (ii) the ability to incorporate sequence information for new threats of interest as soon as it becomes available, (iii) data analysis requiring statistical knowledge about false positives and false negatives,

and (iv) the cost of development for new chip designs when warranted.

In the following section, a list of selected microorganisms relevant to biodefense as indicated by various federal agencies is presented along with the information about why they are important. The sections that follow include a description of gene sequence databases useful in their detection, some of the microarray technologies that are available to develop parallel microbial detection tools, and the research needs to develop PDTs for biodefense.

## 2. Microorganisms relevant to biodefense

Table 1 lists selected examples of microorganisms that are of concern to human health along with the associated concerns and matrices. Various federal agencies including the Centers for Disease Control and Prevention (CDC), United States Environmental Protection Agency (US EPA), United States Department of Agriculture (USDA), United State Department of Health and Human Services (HHS), and Animal and Plant Health Inspection Service (APHIS) have prepared an extensive list of such microorganisms (available through their respective websites) due to their impact on human and animal health and agriculture. None of these lists should be considered exhaustive because information about harmful microorganisms is continuously growing. The approach adopted by various agencies to manage the threats from these microorganisms depends upon their mandate, health risk, and economic considerations. The CDC groups all microorganisms and the associated toxins into three priority areas named A, B, and C, according to their threat level and uses comprehensive species-specific monitoring during outbreaks as a management tool. The US EPA relies on indicator organisms as the most economical tool indicative of human fecal pollution in drinking water. In the food and agriculture industry, management of microbiological threats is accomplished mainly through the detection of individual microorganisms most relevant to a given category of food. Some of these organisms or groups are ubiquitous and exist in the environment due to human and animal fecal pollution and in reservoirs that respond positively and negatively to different environmental conditions (e.g., *E. coli*, enterococci, *Cryptosporidium*, *Vibrio cholerae*, and enteric viruses). Many of them are constant health threats and have been traditionally controlled in drinking water by monitoring indicator organisms or by using appropriate treatment technologies. Other microorganisms do not ordinarily exist in the environment, except in localized endemic cases, and are more relevant to intentional release scenarios (e.g., *Yersinia pestis*, *Bacillus anthracis*). It is obvious that areas experiencing endemic presence of a pathogen will require extra attention to detect intentional release.

Table 1
Selected examples of microorganisms important to human health

| Microorganism | Concern/disease | Transmission matrix | Carriers | Cases per year in the US |
|---|---|---|---|---|
| Bacteria | | | | |
| *Bacillus anthracis* | Anthrax: cutenaceous, gastrointestinal and inhalation | Infected items, animal products, food and water | Animals | 20 in 2001 |
| *Clostridium botulinum* (toxin) | Toxin causes muscle paralysis | Food | Infant intestinal tract, wounds | 110 |
| *Francisella tularensis* | Tularemia | Air, ticks and deerflies | Rodents, rabbits, and hares | 200 |
| *Brucella* species | Brucellosis (flu-like symptoms in humans) | Infected food | Animals | 100–200 |
| *Salmonella* species | Diarrhea, fever, and abdominal cramps | Raw food | Humans, contaminated food | 40,000 (600)[a] |
| *Escherichia coli* O157:H7 | Bloody diarrhea, kidney failure | Food (ground beef), water | Cattles | 73,000 (61) |
| *Shigella* species | Diarrhea, fever, and stomach cramps | Food, water, sewage | Humans | 18,000 |
| *Burkholderia pseudomallei* | Melioidosis (similar to glanders) | Contaminated water, soil, air | Soil and water of the tropics | 0–5 |
| *Listeria monocytogenes* | Fever, fatigue, Nausea, vomiting and diarrhea | Food, water | Soil and water | 1850 (425) |
| *Campylobacter jejuni* | Dysentery, bloody diarrhea | Food, water | Wildlife, birds | $2.1–2.4 \times 10^6$ (500) |
| *Helicobacter pylori* | Duodenal and gastric ulcer | Contaminated water | Humans | 500,000–850,000 |
| Eukaryotes | | | | |
| *Cryptosporidium parvum* | Gastrointestinal illness (e.g., diarrhea, vomiting, cramps) | Water | Human and fecal animal waste | |
| *Giardia lamblia* | Gastrointestinal illness (e.g., diarrhea, vomiting, cramps) | Water | Human and fecal animal waste | |
| Viruses | | | | |
| Norwalk virus | Intestinal illness | Water | Food, water | |
| Enteric viruses (indicator virus) | Gastrointestinal illness (e.g., diarrhea, vomiting, cramps) | Water | Human and animal fecal waste | |

[a] Numbers in parenthesis represents the number of deaths.

In addition to detection of microorganisms at the species level, it is often critical to distinguish the strain(s) present for source tracking and attribution (e.g., in the case of *Bacillus anthracis*, *E. coli*, and *Bacteroides)*. Similarly, determining the genotypes of *Cryptosporidium* (Straub et al., 2002) and their viability may be more relevant than knowledge of its presence in a water source using signature gene sequences. PDTs are especially useful for such applications. For example, a better strain level resolution may be achieved by using genes related to virulence and other functions. Viability can be determined by targeting the ribosomes or the mRNA. However, these steps do add to the chip complexity, validation, and use protocol.

As mentioned before, the microbiological quality of water is regulated by the U.S. Environmental Protection Agency. It mandates monitoring of indicator microorganisms, mainly *E. coli* and enterococci, and the use of appropriate treatment technologies. These groups are commonly used as indicators of human fecal pollution and some of the members are also pathogens. Over time it has been shown that none of these indicator species, as tested by using many commercially available kits, specifically indicate human fecal contamination. These indicators can also emanate from domestic animals and wildlife, and some may even grow naturally in certain environments. The need to improve the tests for detecting existing indicator organisms as well as to explore new species that are better indicators of human fecal pollution (e.g., *Clostridium* sp., *Eubacterium*, *Bifidobacterium* sp., *Bacteroides* sp., etc.) is evident.

## 3. Genetic sequence databases for probe design

Once a list of microorganisms relevant to a niche is finalized, the next step in developing PDTs is the selection of gene targets to design probes. The objective is to detect all of the organisms included in the list at the stated resolution and none outside the list. This later requirement of high specificity is of course as good as the sequence databases itself. Many genes have been used as targets for detecting closely related microorganisms including the small subunit ribosomal RNA (16S rRNA for the prokaryotic microorganisms and 18S rRNA for the eukaryotic microorganisms), intragenic spacer region (between the 16S and 23S genes), and genes for virulence factors and functions specific to a given guild or group of microorganisms. Examples are: shiga-like toxin genes in *E. coli*, nitrite reductase (*nirS*) in denitrifiers, genes contributing to antibiotic resistance in pathogens and commensals, *rpoB* gene responsible for encoding β-subunit of the RNA polymerase, and *hsp*70 for *Cryptosporidium* genotyping (Straub et al., 2002).

The number of sequences of a specific gene available in a database depends upon the type of gene. RDP-II©—the database for ribosomal RNA genes, maintained by the Center for Microbial Ecology (CME) at Michigan State University, currently contains more than 97,000 sequences compared to only 16,277 sequences in the year 2000. Because the 16S ribosomal RNA gene is used to classify the living world into a phylogenetic framework and has become a cornerstone for bacterial identification (Cole et al., 2003), it is the most common target gene for developing detection tools (Loy et al., 2003). Besides phylogeny, there are many other advantages in using the 16S rRNA gene for detection. Its phylogenetic group-based amplification is possible due to the extensive database and conserved gene sequence regions. Because a single cell contains between 10,000 and 60,000 ribosomes under different growth conditions (Commission on Physical Sciences, 1999), its detection at the RNA level is also possible avoiding the PCR and its biases.

There are some disadvantages also with the use of the 16S rRNA gene as a target for detection. Because the molecule is only 1500 bp long and contains highly conserved regions, it cannot yield signature sequences for closely related strains. Therefore, microorganisms cannot be differentiated at the strain level using the sequence variations in the 16S rRNA gene. Some further resolution can be gained by using the 23S rRNA gene but the database for this molecule is much smaller and experience in using it as a detection target is limited. The copy number of the 16S rRNA genes may also vary from one copy to as many as 15 copies for various organisms, complicating the quantification of abundance (Klappenbach et al., 2001). Large variations in signals strengths with very small changes in probe sequences are also reported for 16S rRNA detection and have been attributed to the secondary structure of RNAs.

Alternative gene targets (e.g., *gyrB*, *rpoB*, and genes for virulence or a specific function) do not have some of the above disadvantages. Hence, they are increasingly being used for detection and especially for differentiation at the strain level. Because the isolation of these genes from the environment is more complicated than for the 16S rRNA, and the driving force to collect the sequence (phylogenetic positioning) is not as strong as for the 16S rRNA, the number of sequences available for these is limited. For most of the genes related to virulence or a specific function, there is no dedicated public database (except for the comprehensive gene sequence databases, e.g., GenBank, DDBJ, and EMBL). Whenever it exists due to the efforts of individual researchers (e.g., RISSC: database for 16S–23S RNA gene spacer regions (Garcia-Martinez et al., 2001), *gyrB* (Watanabe et al., 2001), the number of sequences is limited and its maintenance, unless funded from large centralized source, is poor.

With the extraordinary growth in sequence availability and ease in sequencing, the existence of specialized databases for genes related to a specific function and the number of sequences in them is expected to grow. For example, scientists at the Center for Microbial Ecology are developing an antibiotic resistance (AR) gene database. This information is critical if we consider that the most common treatment for anthrax is Ciprofloxacin™ for which AR genes are known to exist with no technical barriers for their manipulation. Furthermore, the growth and spread of antibi-

otic resistance genes among pathogens and commensals is considered a major new challenge in controlling infectious disease (Isaacson and Torrence, 2002). Similarly under a project sponsored by the US Environmental Protection Agency (EPA), researchers at Michigan State University are developing a virulence factor activity relationships (VFAR) database to provide a means of predicting the degree of health hazard for a microorganism from its distinguishable and virulent gene sequences (National Research Council, 2001). Such dedicated databases are expected to be helpful in developing classes of virulent genes that may be tracked irrespective of their host. PDTs using such sequences may also be helpful in identifying new pathogens that are present in the environment but have not yet been identified or isolated. At present, databases for genes related to a specific function, when available, are in their infancy. They are being developed at the pace permitted by resources available to individual researchers. It is well known that probes developed only a few years ago and considered specific to a given microorganism have become less specific mainly due to the availability of more sequences in a database related to the target organism. Hence, probes included on PDTs may need to be updated periodically.

## 4. Microarray platforms available for the development of PDTs

Major biochip synthesis platforms available today to develop PDTs include in-house printed glass slide arrays (Wang et al., 2002a, 2002b; Relogio et al., 2002), immobilized gel-pad system (Liu et al., 2001; Barsky et al., 2002; Urakawa et al., 2002; El Fantroussi et al., 2003), Affymetrix (Wilson et al., 2002a, 2002b), Agilent, NimbleGen, Illumina, and Xeotron (Hashsham et al., 2003). This list is not exclusive and represents only platform types that either were used in published research related to products resembling PDTs or have a known interest in developing PDTs. Other microarray platforms may be equally useful for this purpose. Each may have some advantages and disadvantages associated with them. Excellent reviews focusing on the main features of most of these technologies are plentiful. Hence, the main features relevant to the development of PDTs are listed in Table 2.

Probe length, array density, flexibility (i.e., ability to change the probe design during its development and use), stepwise yield of probe synthesis, development cost, and cost per chip are some of the key parameters that determine the suitability of a microarray platform to develop PDTs. All platforms can synthesize or print probes suitable for PDTs, which is generally 18- to 30-mer. If needed, some are capable of synthesizing longer probes (e.g., Agilent, NimbleGen, Xeotron). All have the required density to address the needs for a given niche. Affymetrix has higher probe densities but it is more expensive and less flexible due to the use of masks. NimbleGen is also a high-density array (393,000 probes) but is only available as a service combined with synthesis, hybridization, and data analysis. Agilent provides glass slide technology commercially at a price that is comparable to other commercial platforms. In-house printed glass slide arrays, although more expensive and cumbersome to develop, are the most cost effective during use. Revision of probes is possible but requires careful tracking of the synthesized probe stock. Febit and PamGene are more recent systems. Febit combines in situ array synthesis, hybridization, scanning, and data analysis in the same equipment. PamGene arrays are known for their fast hybridization protocols, completing the process in less than one hour. Xeotron's in situ synthesis chip platform is also a recent invention (Gao et al., 2001). It has a probe density of 7963 features, is one of the least expensive for developmental work, and comparable in cost to other systems during use. Fig. 2 shows an in situ synthesized biochip from Xeotron that we are using to develop PDTs for pathogen detection, source tracking, whole genome DNA and expression analysis, and community fingerprinting. The key feature of this platform is its flexibility. Probe design can change from one chip to another without incurring much cost. The chip, a microfluidic device using modified generic hybridization chemistry and detection protocols, can also be integrated with other automated sample processing devices currently under development.

## 5. Issues related to the development of PDTs

Preparing a list of microorganisms, selection of gene targets, and choice of a suitable microarray platform provides

Table 2
Key aspects of oligonucleotides biochip synthesis technologies for parallel microbial detection

| | Parameter | Affymetrix | NimbleGen[a] | Agilent | Gel pad | Glass slides | Febit | PamGene | Xeotron |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Probe length | 20 (30)[b] | 70 (85) | 60 (85) | 20 (85) | 70 (85) | 25 | 20 | 45 (100) |
| 2. | Array density | 500000 | 393000 | 22000 | 2000 | 40000 | 6000 | 1600 | 7963 |
| 3. | Stepwise yield | 90+% | 97.5% | 93+% | 93+% | 93+% | –[c] | –[c] | 97–99+% |
| 4. | Flexibility | + | +++++ | ++ | ++ | ++ | +++++ | ++++ | +++++ |
| 5. | Development cost | ++++++ | + | ++ | + | + | + | + | + |
| 6. | Cost per chip | ++++ | ++++ | +++ | + | + | ++ | ++ | +++ |

[a] Available as a service only.
[b] Most common probe length (current maximum possible length).
[c] Not available at present.

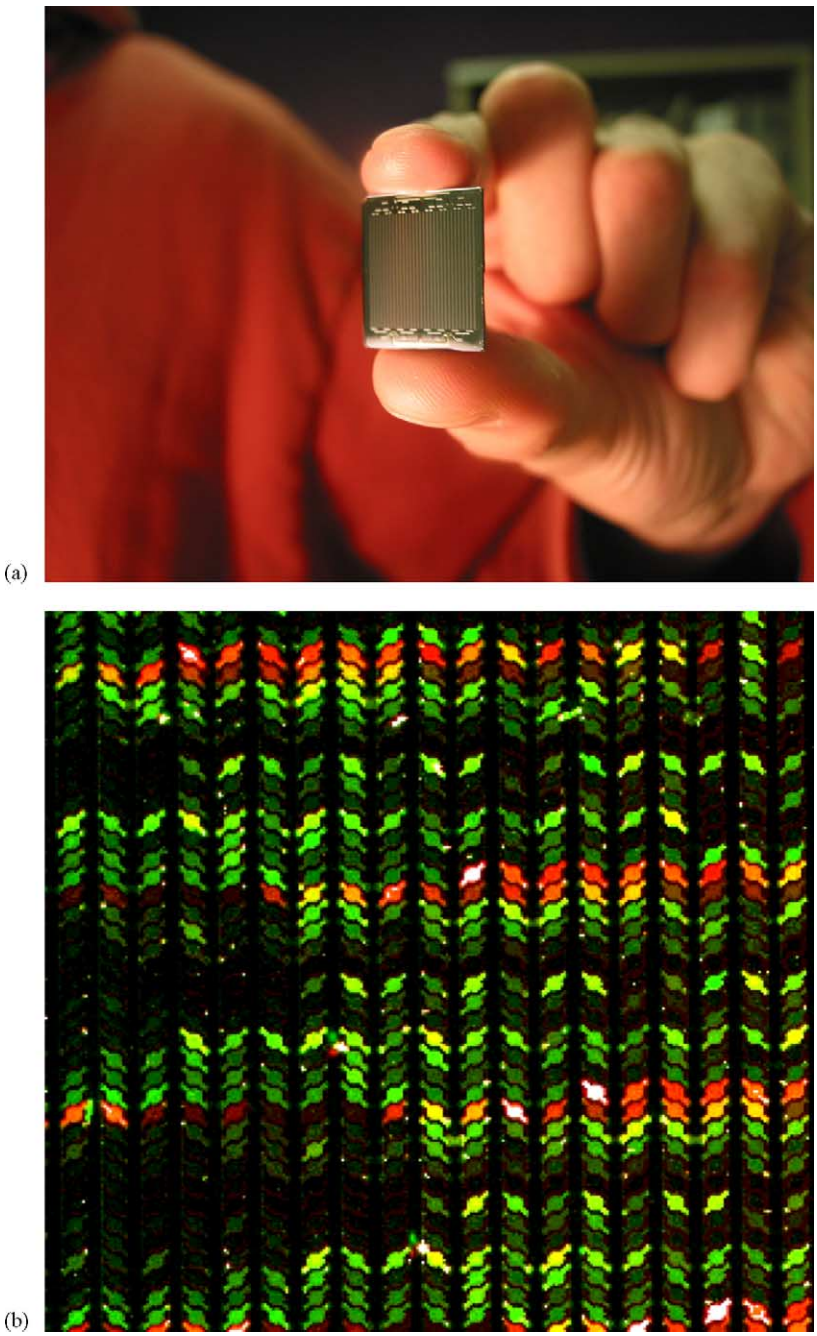


Fig. 2. A XeoChip$^{TM}$ containing 7963 probes synthesized in situ in microfluidic reactors (a), signals obtained for gene targets using a laser scanner (b). Only 10% of the chip area is shown for clarity. (*Photo:* courtesy of Vincent Denef, Wouter Donckerwolcke, and Yoshinobu Matsamura).

the starting point to address the research issues related to the development of PDTs. Many of these issues are highlighted in Fig. 1. Some that are relevant to all platforms are described below in more detail including: (i) probe design, specificity, and false signals, (ii) validation, quantification, and detection limit, (iii) sample concentration, (iv) selective target amplification and enrichment, and (v) data analysis tools for PDTs.

### *5.1. Probe design, specificity, and false signals*

The accuracy of detection of PDTs will obviously depend for a part on the quality of the oligonucleotide sequences on the microarray. Three major criteria are considered during oligonucleotide design. First, the sequences used on the chips have to be specific to their respective targets to avoid any cross-hybridization. Second, these oligonucleotides should

not fold to form stable secondary structures that may compete with the probe during hybridization. This is especially true when naturally folded sequences (rRNA) are targeted. Finally, these oligonucleotides should have consistent thermodynamic properties (i.e., melting temperatures) across the chips to behave similarly during hybridization.

Probes targeting the 16S rRNA to detect a single microorganism or a group of microorganisms are mostly short, 18–25 nucleotides in length. Generally, the longer the probe the less specific it is due to chances of cross-hybridization with other closely related target sequences and decreased destabilizing effect of a single mismatch in a long sequence during hybridization (Kane et al., 2000; Kaderali and Schliep, 2002). Probe lengths for genes related to a specific function may be somewhat longer, especially if the gene sequence is unique and detection of single nucleotide polymorphisms that may be present in different strains is not an objective (Bekal et al., 2003). Probe design for PDTs can be accomplished by one or more of the following four approaches: (i) by designing probes for individual microorganisms or a phylogenetically related group using software commonly used for 16S rRNA probe design, (ii) by tiling the target into the maximum number of possible n-mers, (iii) by finding all the unique 18- to 30-mers in a set of target gene sequences, using bioinformatics approaches (Zhang et al., 2002), and (iv) by first identifying unique genes among all the microorganisms of interest and then finding the unique signatures for each organism.

The first approach has been mostly used for designing phylogenetic probes targeting the 16S rRNA gene. However, its application to other types of genes has been shown successfully. The software most commonly used for this purpose is ARB, as evident by a large database of individually developed probes (Loy et al., 2003). Primrose, a user-friendly package for PC users has also been recently developed (Ashelford et al., 2002). The tiling approach simply involves breaking down the target gene into a number of n-mers without consideration for conserved or unique regions. It does not depend on preliminary screening of non-unique regions in the gene. Alternatively, the 16S rRNA gene database can also be analyzed for all unique short oligonucleotides probes (Zhang et al., 2002). Such a set of unique probes that is continuously updated may be quite useful in keeping PDTs up to date because it could be used to theoretically and experimentally determine the impact of new sequence information. This approach is extendable to other gene sequences related to a specific function. A limitation of the existing software tools is their inability to design probes simultaneously for a large number of distantly related targets. Probe design must be attempted one by one for each phylogenetic member or group.

Bioinformatics approaches for identifying unique n-mers in a set of genes, differ from the above probe design approach in the sense that they do not require identification of the target organisms prior to probe design. Instead, suitable probes for the identified microorganisms can be chosen from the unique signature set. Since most of these approaches are of recent origin (Li and Stormo, 2001; Kaderali and Schliep, 2002; Pozhitkov and Tautz, 2002; Rahman, 2002; Rouillard et al., 2002; Zhang et al., 2002; Matveeva et al., 2003), no user-friendly software tools exist to accomplish the task. The algorithms, however, are generally open source and available from the respective authors. This approach is most useful when the number of organisms to be targeted is large or a large number of probes are needed for a small set of microorganisms. This approach is more likely to be used for probes targeting genes of a functional category. This is due to the insufficient number of sequences for genes related to a specific function and their unavailability in a single database similar to RDP-II. Tools for exploring primers for their universal amplification and for detection are not well developed. Current approaches use mostly manual collection of the related sequences for a gene target related to a specific function and use various alignment tools to identify regions of gene sequences. This yields probes specific to a single or group of microorganisms. For the hybridization data shown in Fig. 2, the design of probes for the detection of antibiotic resistance genes involved identifying and collecting all the sequences of AR genes from various databases, and using ARB, GeneDoc (Nicholas et al., 1997), and Primrose for various steps in the process of probe design.

The fourth probe design approach is conceptual at present and should be treated as such. Every pathogenic microorganism is expected to have a set of genes that are unique compared to all other microorganisms. Identification of these unique genes, using established tools, such as BLAST is trivial. Whether these unique genes also contain signature sequences useful for the detection of the corresponding microorganisms is not yet explored. Also, the presence or absence of these genes in unknown genotypes is not known. The existence of signature sequences, however, is expected. Similar approaches have been proposed to develop virulence factor activity relationships for microorganisms (National Research Council, 2001). Sequencing of important new pathogens is not considered a bottleneck within a reasonable timeframe. Thus, it is theoretically possible to identify a set of unique genes that encompasses all the sequenced microorganisms, design gene specific probes using well established software tools (Li and Stormo, 2001; Rahman, 2002; Rouillard et al., 2002), and use them as target genes for detection. When DNA or RNA from isolates or enrichments is available, identification of the pathogen will be similar to the results obtained by Wang et al. with a virus chip (Wang et al., 2002a, 2002b). However, if the suspected target is present in a background of large amounts of community DNA, the problem of detection limit needs to be addressed. This is because, any single gene will generally constitute much less than 1% by weight of the total community DNA (Cho and Tiedje, 2002) and therefore, will not be easily detectable. If many such genes with varying sequences and functions are selected as targets for microarrays, amplification with a universal primer will not be an option. It may be possible, however, to selectively enrich these targets using mul-

tiple capture probes on magnetic beads (Chen and Griffiths, 2001).

The primary requirement of probe design specific to a single organism is that it must be able to specifically detect the target gene in a mixture of other genes without cross reactivity to other sequences from closely related strains. Specificity is generally achieved through a combination of theoretical analysis and experimental verification. All the above probe design approaches, except the tiling strategy, carry out some theoretical analysis to yield a list of signature sequences specific to the intended target. However, theoretical analysis alone does not reliably predict the probe behavior during use. This is due to gaps in knowledge related to the thermodynamics of probe hybridization with its target on various substrates and under varying buffer conditions. The theoretical analysis assumes a certain concentration of targets, determination of which is one of the objectives of a PDT. Because the target concentrations and their relative abundance are unknown and numerous, it is often treated as incalculable. Hence, experimental validation of probe behavior and optimization of hybridization conditions is necessary. Optimization is a difficult and cumbersome task even when one or a few microorganisms are targeted. With PDTs, the problem gets amplified significantly, since a single hybridization protocol must now be suitable for all probes. One approach, that we have adopted for high throughput experimental screening of probes, is the use of heat maps, displayed in Fig. 3. The theoretically designed good probes and their mismatch counterparts can be clustered in order to distinguish promising probes from bad probes. Wash temperatures that show the highest discrimination between perfect match and mismatch are also readily visible for each probe. Ideally, the theoretical tools should be able to predict this behavior under various environmental and target mixture scenarios.

In spite of careful probe design and experimental validation, there is always a chance for observing positive signals when the intended target is absent (false positive) or missing detection when the target is present (false negative). Possible reasons for this error include: (i) more than 97% of the total microbial diversity is still unknown, (ii) knowledge about the thermodynamics of probe hybridization on the support is poor, and (iii) protocol optimization is economically feasible for only a few types of matrices and microorganism mixtures. High false positives and negatives are extremely undesirable and must be accounted for in validation. Analysis of false positives will require more care for PDTs with

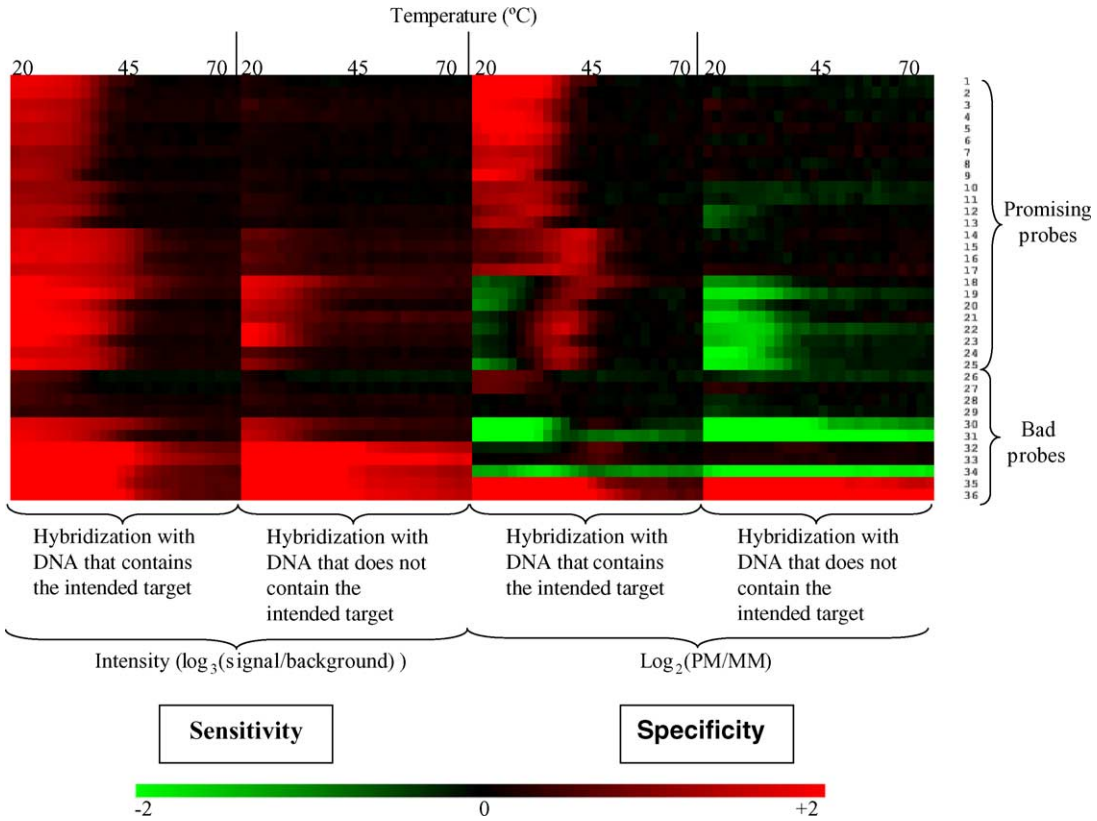


Fig. 3. Heat map showing sensitivity and specificity of a set of probes. Promising probes have a relatively high signal intensity (red color) and a high PM/MM ratio (red color) when hybridized with the intended target DNA, but a low signal intensity (black or only low red color) and a low PM/MM ratio (black or green color) when hybridized with non-target DNA (probes 1–25). Some probes only have a small temperature range where they show a good PM/MM ratio (probe 18–25). Bad probes show a low specificity and/or a low signal intensity (probes 26–29). Some probes with high intensity have a very low specificity, no specificity (probes 30–34) or only apparently a high specificity (probes 35–36). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

specialized probe designs to accommodate genetic variation in the target populations so that the range of natural and artificially manipulated variants can be detected. Understanding the background indigenous populations, is also critical to the reliable detection of the infectious agent. This implies that flexibility of PDTs to allow revisions and upgrades both in probe design and data analysis based on the availability of new sequences is as critical as the initial development itself.

### 5.2. Validation, quantification, and detection limit

It is conceivable that probes can be designed for most microorganisms in a given database. Synthesis of a PDT containing these probes is also trivial, given the technological advances. Validation, however, to determine the usefulness of PDTs, is the most expensive and the rate-limiting step (Wilson et al., 2002a, 2002b; Urakawa et al., 2002; Wang et al., 2002a). Ideally, a PDT must be validated for each of the microorganisms individually and in mixtures of various targets and background microorganisms expected to be present in a given sample. It must also address the issue of closely related strains that are not harmful but present in the background matrix. The protocol(s) must work for all types of matrices. Validation, accompanied by careful choice of probes, may also enable PDTs that are useful for discovering new threats. Consider, for example, the recently developed glass slide arrays containing 70-mer probes for more than 140 viruses (Wang et al., 2002a). Although, it was individually validated for only a handful of viruses, new types of viruses not represented on the chip were also detectable. Such an approach may be extremely useful in cataloguing the signature of background microbial communities or viruses present in the environment on a global scale (Anderson et al., 2003). The more difficult need to address is that validation must also be revised with new information and reflect the level of uncertainty in the positive signals due to the lack of knowledge about the unknown microorganisms.

Hybridization signals for probes are never obtained as a 0 or 1 event (indicating an absence or presence) for the corresponding targets. A continuum of signals is obtained between a minimum that is close to zero, given by the background fluorescence of the array substrate, and a maximum determined by the dynamic range of the laser scanner, probe behavior, and the amount of target. A higher signal for a given probe compared to another probe giving a lower signal does not necessarily mean a higher abundance for the first target (*Note:* e.g., the signal difference of Probe A and B in Fig. 3, targeting the same gene). This was one of the reasons for using a ratio approach in applications of microarrays to gene expression studies. There, an increase in the ratio of cDNA present in two samples (hybridized on the same array) was indicative of an increase in the relative amount of cDNA in one sample compared to another. Absolute quantification was not needed. Individual samples to be hybridized on PDTs are not expected to use ratios because the samples are separated in time and space. It is possible to imagine the use of a standard mixture of microorganisms (at the cellular, DNA, RNA, labeled DNA, or labeled RNA level) against which all unknown samples must be compared, but the utility and ease of use of such mixture(s) as standards is yet to be demonstrated. Until then, it will be assumed that the use of single sample hybridization will be more common for PDTs.

Considering quantification is important, validation must be considered at two levels: validation of the presence or absence and validation of the abundance. The first level only indicates whether the target is present above or below an arbitrarily specified “detection limit”. All signal intensities below the detection limit, including those obtained with no target present, are labeled as absent or below detection limit. All signal intensities above the detection limit are termed as present. The second level of validation provides abundance in addition to the presence and absence. Quantification of abundance is important because it is not sufficient to know which organism among the known threats is present. It is much more informative to know how many (or how much) of the known threats are present. Relative abundance, obtainable using mixtures of standard targets and two-dye experiments similar to relative gene expression studies, are obviously less useful for biodefense-related applications. Whether microarrays will be able to provide quantification of abundance of a given target in single dye experiments is still being explored.

If validation is carried out at a single temperature, quantification of relative abundance is only possible by using the ratio approach and standard mixtures as mentioned above. An alternative to this approach is to determine the hybridization behavior of the probe and target as a function of hybridization temperature, i.e., obtain a dissociation curve for each probe on the platform of interest. For protocols using membrane hybridization, obtaining such a curve was always a part of the detection process (Zheng et al., 1996). A dissociation curve indicates what fraction of the total target DNA or RNA of a given microorganism remains bound to the probe at various temperatures (Liu et al., 2001; El Fantroussi et al., 2003). Unless this fraction is known for each probe, determination of abundance for the corresponding targets is not likely. The temperature at which 50% of the total signal for a given probe is lost in solution is called the dissociation temperature ($T_d$).

Dissociation temperature is loosely related to the melting temperature ($T_m$), calculated theoretically using the Nearest-Neighbor model (Santalucia, 1998; Rouillard et al., 2002). Breslauer et al. (Breslauer et al., 1986) and Santalucia (Santalucia, 1998) have extensively studied the thermodynamic parameters that are needed for the nearest neighbor model. One of the parameters is the total molar concentration of the annealing oligonucleotides. Because this parameter is an unknown to be determined by the PDT, all theoretical calculations use an assumed constant value for comparative purposes. Hence, determination of a dissociation curve on a chip at least during the developmental phase is important. Such a dissociation curve combined with the dissociation curve of a mismatch probe could enhance the reliability of detection and help minimize false positive and false negative signals.

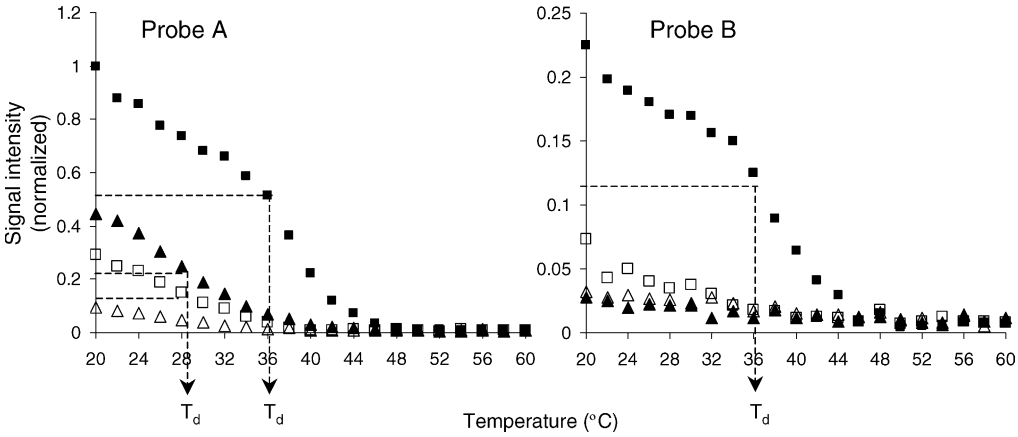


Fig. 4. Non-equilibrium melting curves for two 20-mer probes (A, B) specific to Shiga toxin 1A (perfect match, PM) and their mismatch probes (MM, single mismatch at position 10). Genomic DNA (gDNA) of *E. coli* Sakai and *E. coli* K12 was labeled with Cy3 and Cy5 respectively and hybridized to a Xeotron chip at 20 °C for 16 h. Melting curves were then generated by washing the chip for 2.5 min in 2 °C steps and scanning after each wash. (■) Sakai gDNA bound to PM; (□) Sakai gDNA bound to MM; (▲) K12 gDNA bound to PM; (△) K12 gDNA bound to MM. $T_d$ values were inferred from melting curves. Note the different scale of the y-axis in A and B.

Non-equilibrium melting curves are a variation of the equilibrium dissociation curves discussed above and are obtained by washing an array at increasing temperatures after complete hybridization at a low temperature. Examples of non-equilibrium melting curves obtained by hybridizing labeled genomic DNA of two *E. coli* strains (Sakai and K12) to a microarray with 20-mer probes targeting the Shiga toxin 1A (*stx*1A), are shown in Fig. 4. The Sakai strain contains the *stx*1A gene, whereas the K12 strain does not. As expected, both probes A and B yield the strongest signal for the Sakai strain bound to the perfect match (PM) probe. Although, targeting the same probe and having similar $T_m$ values, the two probes vary in signal strength (note the different scale on *y*-axis). With the Sakai strain, probe A shows about five times higher signal than probe B. At lower temperatures, even the non-specific signal of probe A obtained with the K12 strain is higher than the specific signal of probe B obtained with Sakai strain. In this case, known concentrations of known genomic DNA were hybridized, making the interpretation simple. However, if an unknown sample of unknown concentration yields a signal like the one obtained here from K12 with probe A, interpretation will be more difficult. If only one temperature point is taken into account, e.g., 32 °C, it is difficult to determine if the signal is from a specific target at low concentration or from a non-specific target, since signals for the PM probe are considerably higher than for the MM probe. The $T_d$ values obtained from melting curves are able to determine the specificity of the signals. Whereas, $T_d$ for the MM probe for the Sakai strain is clearly lower than for the PM probe, both $T_d$ for PM and MM probe have the same low value for the K12 strain.

The issue of absence, low abundance, or below detection limit for biological agents is different from that faced for chemicals. Due to the potential for growth under the appropriate environmental conditions, a microorganism that is below the detection limit can easily become the dominant organism in a system. Most of the natural pathogen outbreaks are indicative of this major difference. Moreover, for many pathogens the minimum infectious dose is quite low. Hence, continuously improving the detection limit for all pathogens is critical. At the DNA level, the problem of detection limit for a single gene extracted from a mixed microbial community is qualitatively similar to the problem of detecting rare messenger RNAs in human brain cells. The Human genome has some 30,000 genes. In a typical cell, expression of only 300 or so genes is dominant. The expression of the remaining genes is rare but presumably many of them do influence the overall functioning of the cell, making their expression also important to measure. The current detection limit of microarray based approaches for gene targets present in mixed communities is much less than desirable (Denef et al., 2003). Using nitrite reductase (*nirS*), naphthalene dioxygenase (*nahA*), and *Escherichia coli* O-antigen biosynthesis gene, Cho and Tiedje (Cho and Tiedje, 2002) demonstrated that the lower detection limit on glass slide arrays using large PCR amplified products is 1–10 pg of the model gene (Cho and Tiedje, 2002). Assuming a microbial genome of 5 Mb for *Pseudomonas stutzeri*, one copy of *nirS* gene in *P. stutzeri*, and 10 pg of lower detection limit, they highlighted that 50 ng of *P. stutzeri* DNA must be present in the 1 μg of total DNA hybridized to the glass slide. This implies that the DNA of the target microorganism must constitute at least 5% of the total DNA. Obviously, this limit must be improved by at least two to three orders of magnitude.

As a desirable goal, microarrays must be able to detect 1 cell of the target microorganism in a background of 2000–20,000 other types of microorganisms as background. In addition, the total amount of sample DNA needed for hybridization must also be reduced by 1–2 orders of magnitude. Thus, in addition to detecting many microorganisms in parallel, PDTs need to attain detection limits that are

comparable to existing methods for the detected pathogens. The key areas that have the greatest potential are preferential enrichment using magnetic beads and indirect and direct signal amplification followed by enhancement in quantity and quality of probes. Recently, Denef et al. (Denef et al., 2003) have improved this detection limit to approximately 1% of the total population based on DNA abundance for glass slide arrays using short oligonucleotides probes by better labeling strategies. While, selected examples exist that demonstrate ways to achieve better detection limits (Yguerabide and Yguerabide, 2001; Bao et al., 2002), the need for further improvement is universally recognized.

### 5.3. Sample concentration

All microbial detection methods need approaches to bring the target microorganisms to the testing platform. It is apparent that all the promises of a PDT depend upon successful transfer of the targets from a large volume of matrix to the chip in a manner that does not inhibit the hybridization process. Microorganisms suspected to be present in large quantities of matrices (1–100 $m^3$ of air, 1–100 L of water, 1–100 g of food, and variable quantities of other media) must be transferred to μL volume of liquid before their detection. Because results from PDTs will only be as good as the concentrated sample that is put upon them, it is critical to ensure that methods for sample concentration exclude material that is inhibitory to hybridization or plugs the membrane filters before a requisite amount of water has passed. For air, which is least problematic, membrane filtration followed by transfer of the retained material to a liquid medium has been the preferred approach for developing a number of microbial detection devices. For drinking water requiring filtration of up to 1000 L of finished water, concentration by membrane filtration followed by recovery is also relatively simple. Membrane filtration technologies (microfiltration, ultrafiltration, nanofiltration, and reverse osmosis) are now mature for water purification. Microfiltration and ultrafiltration membranes efficiently concentrate larger microorganisms like *Cryptosporidium* and *Giardia* but for viruses, charged membranes are more advantageous (Scott et al., 2002). For surface waters, generally a lower volume of 1–100 L is sufficient because of the higher concentrations of microorganisms present. Filters and protocols are constantly being optimized for this purpose, yielding 40–80% recoveries (e.g., Envirochek HV filtration system with 1 μm pore size from Pall Corporation and Filta-Max system from IDEXX Laboratories Inc.). Some of these approaches have also found acceptance by the EPA for routine monitoring of *Cryptosporidium* (EPA Method 1622 (US EPA, 2001a)) and/or *Giardia* (EPA Method 1623 (US EPA, 2001b)), and viruses. However, large variations in performance still exist depending upon the sample volume, filtration flow rate, and the number of microorganisms and particles present. For viruses, the nature of the virus particle, water characteristics, and membrane characteristics are all important parameters determining their recovery. Many of these existing membrane filtration approaches can be evaluated for their ability to concentrate all types of microorganisms in parallel. Portability of the instrument, performance, cost and reusability of filter, and recoverability of the concentrated organisms are the main issues that must be addressed. The concentration step for water may be followed by enrichment but for other matrices (food, soil, body fluid, tissue, etc.), direct enrichment by selective media is more common.

### 5.4. Selective target amplification and enrichment

If traditional means of sample concentration fail to attain a desired minimum for detection limit, alternative approaches to concentrate the target organisms or the target genes are needed. These include amplification of the target gene using universal primers, biological enrichment, i.e., growth of the target microorganism on selective media, and enrichment using capture probes on magnetic beads. The first approach is most useful for the 16S rRNA genes because of the availability of universal primers. The second approach is time consuming and dependent on the recoverability and selectivity of the enrichment media. The third approach is quite promising but still under development. Most of the enrichment and amplification strategies carry associated biases but they are also indispensable tools to enhance the detection limit. Selective media and PCR amplification have been in use for decades. However, as mentioned before, amplification of gene families related to a specific function may be problematic due to the lack of suitable universal primers.

The use of capture probes coated on magnetic beads is an emerging technology used to enrich genes related to a specific function or genes that do not possess known conserved regions for universal amplification. Magnetic separation itself is not new but its application to molecular biology has grown considerably in the last decade. Super-paramagnetic particles of 1–5 μm and colloidal beads of 50–200 nm, coated with capture probes can be used to selectively enrich the target molecules and separate them using a magnetic field. Super-paramagnetic particles are attracted to a magnetic field but do not exhibit any residual magnetism once the field is removed. Several companies (e.g., Dynabeads from Dynal Biotech, Avibeads from Aviva Biosceinces, and ClinProt from Bruker Daltonics) now produce beads for enrichment of many types of biological molecules, including DNA, RNA, and specific proteins. The thermodynamics of capture probe hybridization on beads is also being addressed (Stevens et al., 1999). Integration of the steps involving concentration, amplification, and enrichment with detection to obtain a simple and reasonable protocol and device is the goal of many current research projects (Straub and Chandler, 2003). The integration of many of these strategies to PDTs in a manner that delivers consistent performance for the matrices of interest is yet to be demonstrated.

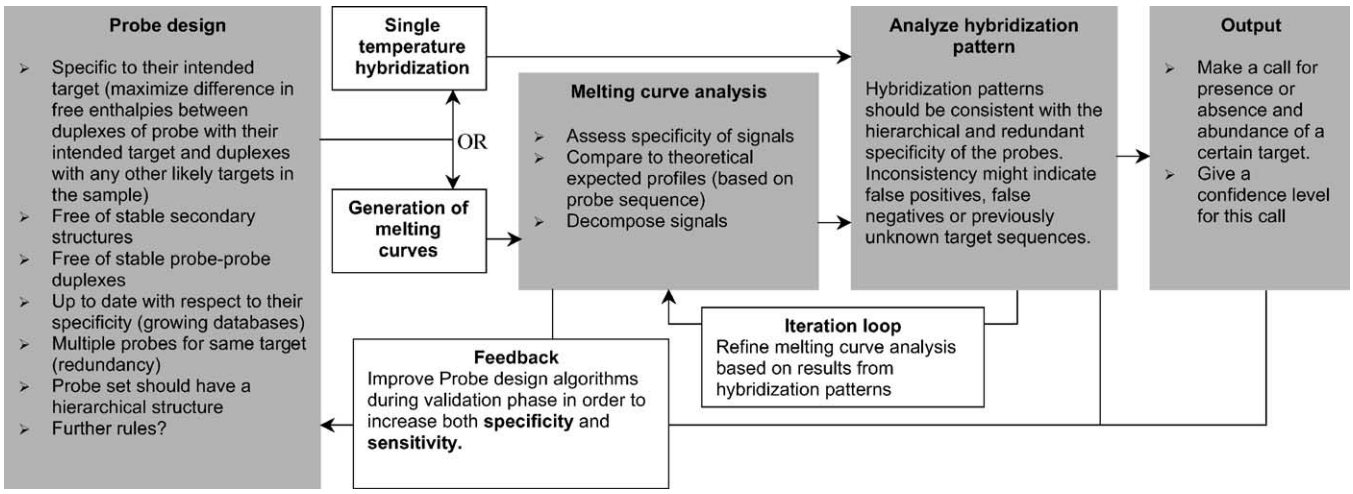


Fig. 5. Major steps and challenges in probe design and data analysis for PDTs.

### 5.5. Data analysis tools for PDTs

Numerous tools for data processing from microarrays exist but were originally developed for gene expression analysis, not microbial detection. New modules making the call for the presence/absence or abundance have to be developed and added to or used in conjunction with the current data analysis tools. Statistical and uncertainty analysis tools must also be incorporated to reflect the areas where knowledge gaps exist. Besides the traditional data analysis of microarrays, a number of additional data analysis issues need to be addressed for PDTs. They are related to one or more of the following: probe design approach, performance of probes during validation, limited and constantly growing database, characteristics and abundance of the known microorganisms present as background, hybridization behavior of the PDTs, sample processing and enrichment effects, and analysis of dissociation curves (if used). Fig. 5 depicts these issues in the form of a flowchart, starting from probe design to making a call for the presence or absence of a given target. Gray boxes indicate fields where further fundamental research and development of powerful bioinformatics tools are needed. Data analysis for PDTs is closely related to probe design. High sensitivity and specificity are fundamental characteristics of good probes. Several factors influencing sensitivity and specificity are well established and are generally incorporated in the probe design strategy. However, hybridization behavior of probes can only partly be predicted by these factors and more research is needed (Matveeva et al., 2003). Redundant and hierarchical design of probes is an essential component to increase reliability. Consistently high or low signals for all probes will increase the confidence level for a call about the presence or absence of a certain target. Inconsistent results indicate false results or binding of a previously unknown target (Liu et al., 2001; Loy et al., 2002).

Dissociation curves are considered key in parallel microbial detection (Urakawa et al., 2002). User-friendly tools to analyze all types of temperature dependent probe behavior will be extremely useful for developing PDTs. Ideally, melting curves should be predictable based on the probe sequence and then compared to the experimentally obtained curves. Moreover, if one or several unspecific targets are bound to a probe in addition to the specific target, it would be desirable to deconvolute the observed combined melting curve into the underling melting curves from the different targets. These results could then be used in the interpretation of hybridization patterns and making the positive or negative calls with an associated uncertainty. In an iterative approach, these results might be used to refine the melting curve analysis, since certain targets may cross-hybridize with closely related targets.

Another key aspect of data analysis for PDTs is the uncertainty associated with signal intensities that are closer to the background. Obviously, the signal intensities below an arbitrary cutoff (most commonly the average background plus three times the standard deviation for the background) will be termed as background. To minimize false positive, this arbitrary cutoff may be set at a higher level. The signals much above the cutoff value are positive signals for the corresponding targets. However, signals that are barely above the cutoff value pose a problem. These signals may be a result of poor background, poor probe, or small abundance of good probe. In a mixed community, if the signals emanating from the target organism happen to be in this region, it will be necessary to find alternative approaches to enhance the certainty associated with the detection. It is obvious that probability distribution curves for various targets will be an essential component of data analysis.

## 6. Strengths, weaknesses, and rate-limiting steps in the development and use PDTs

The major strengths of PDTs are: the ability (i) to screen for hundreds of pathogens together, (ii) to provide a sequence-

based resolution not feasible by using other methods, (iii) to enable high throughput source tracking, (iv) to serve as a tool for exploring unknown bacterial pathogens and viruses, and (v) to track virulence factors irrespective of their host. It is particularly advantageous when the suspect agent is unknown and screening must be done for a large number of microorganisms. It is also extremely useful when the suspect agent is new and traditional methods are not available to detect it (Wang et al., 2002a, 2002b).

Weaknesses of PDTs are, at present, its high cost, long time of testing, complexity of protocols, and lack of available systems that are validated for more than a few dozen microorganisms. The current cost of one chip, irrespective of the platform, runs into hundreds of dollars. It is not surprising that any parallel approach developed to screen for hundreds of microorganisms together may be more expansive compared to some well-established kits developed for a single or a few microorganisms. However, the choice of whether PDTs should serve as the high priced consultant in specialized cases or as the security screen against most pathogens for all needs related to air, water, food, and agriculture around the nation is an important one. It affects the direction of PDT development, PDT cost, and the hopes associated with PDTs. Although, most of the traditional tests for many of the microorganisms may also cost hundreds of dollars if used in parallel and may take days to complete, there are competing technologies (e.g., real-time PCR (RT-PCR)) that may fulfill part of the services provided by current PDTs at a comparable cost.

At present, most microarray-based hybridization protocols take several hours followed by a significant amount of time for data analysis that is not as standardized as the analysis of RT-PCR curves. Only one microarray platform (PamGene) using a unique 5D-Pulse$^{TM}$ mixing technique is able to complete the hybridization process in less than 1 h (van Beuningen et al., 2001). Although, this platform is low density, the fast kinetics demonstrate that modification in protocols and mixing approaches may reduce the hybridization time considerably. The 7-min detection of bacteria (Belgrader et al., 1999), using RT-PCR is a perfect example of what must be set as a target for speed of detection.

Complexity of protocols refers to the required skill of the personnel carrying out the test. At present, the overall process from start to finish (e.g., for water) may involve the majority of the following steps: sample concentration, DNA (or RNA) extraction, target enrichment or amplification, labeling, hybridization, signal amplification, scanning, and data analysis. Most of the initial steps, which impact the hybridization results profoundly, are difficult to automate for complex matrices. Air is an exception and considerable success has been achieved in developing fully automated and multiplexed systems for air. Although, PDTs are also amenable to automation and integration with other miniaturized devices, most of the current focus is on the development of the chip itself.

As stressed before, validation is the rate-limiting step for developing PDTs. Hence, alternative high throughput approaches for validation will be extremely useful. Other factors limiting the pace of PDT development and use include: (i) availability of gene databases related to a specific function and probe design tools for genes related to a specific function supported by sequencing of the genes useful in detection, (ii) lack of better predictive rules for probe design based on thermodynamic considerations, (iii) lack of methods for sample concentration and enrichment of seemingly unrelated genes without the need to use universal primers, (iv) poor detection limits for a single gene in a background of many genomes in various matrices, (v) lack of information about the background microbial populations complicating the detection of pathogens, (vi) lack of approaches to decrease the testing time and cost and enhance the ease of application of DNA arrays in a diagnostic setting, and (vii) lack of methods of data analysis that include statistical rules and uncertainties associated with PDTs.

To address some of the above issues, more recent methods are using multiplex PCR followed by microarrays to take advantage of both techniques. One amplifies the signals from various genetic elements of interest, and the other provides the sequence resolution not achievable by PCR. Still others are developing RT-PCR capabilities on the microarray itself. The combined strength of RT-PCR and microarrays may indeed address issues related to the abundance and detection limits faced by PDTs. In the long run, PDTs themselves may be developed similar to computer chips, as parts to be used by other devices integrating sample processing, chip synthesis, target amplification, hybridization, scanning, and data analysis. If all the known pathogens of interest, could be quantitatively detected in parallel within a short time of say less than 1 h at a total reagent cost of less than $100 (plus the one time cost of the equipment), then it sets the standard against which the performance of all the PDTs must be measured.

## References

Anderson, N.G., Gerin, J.L., Anderson, N.L., 2003. Global screening for human viral pathogens. Emerg. Infect. Dis. 9, 768–774.

Ashelford, K., Weightman, A., Fry, J., 2002. PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. Nucl. Acids Res. 30, 3481–3489.

Bao, P., Frutos, A.G., Greef, C., Lahiri, J., Muller, U., Peterson, T.C., Warden, L., Xie, X., 2002. High-sensitivity detection of DNA hybridization on microarrays using resonance light scattering. Anal. Chem. 74, 1792–1797.

Barsky, V., Kolchinsky, A., Lysov, Y., Mirzabekov, A., 2002. Biological microchips with hydrogel-immobilized nucleic acids, proteins, and other compounds: Properties and applications in genomics. Mol. Biol. 36, 437–455.

Baum, M., Bielau, S., Rittner, N., Schmid, K., Eggelbusch, K., Dahms, M., Schlauersbach, A., Tahedl, H., Beier, M., Guimil, R., Scheffler, M., Hermann, C., Funk, J.M., Wixmerten, A., Rebscher, H., Honig, M., Andreae, C., Buchner, D., Moschel, E., Glathe, A., Jager, E., Thom, M., Greil, A., Bestvater, F., Obermeier, F., Burgmaier, J., Thome, K., Weichert, S., Hein, S., Binnewies, T., Foitzik, V., Muller, M., Stahler, C.F., Stahler, P.F., 2003. Validation of a novel, fully in-

tegrated and flexible microarray benchtop facility for gene expression profiling. Nucl. Acids Res. 31, Art. No. e.151.

Bekal, S., Brousseau, R., Masson, L., Prefontaine, G., Fairbrother, J., Harel, J., 2003. Rapid identification of *Escherichia coli* pathotypes by virulence gene detection with DNA microarrays. J. Clin. Microbiol. 41, 2113–2125.

Belgrader, P., Benett, W., Hadley, D., Richards, J., Stratton Jr., P., Milanovich, R.M.F., 1999. Infectious disease: PCR detection of bacteria in seven minutes. Science 284, 449–450.

Bopp, D.J., Sauders, B.D., Waring, A.L., Ackelsberg, J., Dumas, N., Braun-Howland, E., Dziewulski, D., Wallace, B.J., Kelly, M., Halse, T., Musser, K.A., Smith, P.F., Morse, D.L., Limberger, R.J., 2003. Detection, isolation, and molecular subtyping of *Escherichia coli* O157: H7 and *Campylobacter jejuni* associated with a large waterborne outbreak. J. Clin. Microbiol. 41, 174–180.

Breslauer, K., Frank, R., Blocker, H., Markey, L., 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. U.S.A. 83, 3746–3750.

Busch, U., Knoll-Sauer, M., Muhlbauer, B., Zucker, R., Huber, I., Beck, H., 2003. Detection of pathogens in food by biochip analysis. Fleischwirtschaft 83, 111–114.

Chen, J., Griffiths, M.W., 2001. Detection of *Salmonella* and simultaneous detection of *Salmonella* and Shiga-like toxin-producing *Escherichia coli* using the magnetic capture hybridization polymerase chain reaction. Lett. Appl. Microbiol. 32, 7–11.

Chizhikov, V., Rasooly, A., Chumakov, K., Levy, D.D., 2001. Microarray analysis of microbial virulence factors. Appl. Environ. Microbiol. 67, 3258–3263.

Cho, J.C., Tiedje, J.M., 2002. Quantitative detection of microbial genes by using DNA microarrays. Appl. Environ. Microbiol. 68, 1425–1430.

Cole, J., Chai, B., Marsh, T., Farris, R., Wang, Q., Kulam, S., Chandra, S., McGarrell, D., Schmidt, T., Garrity, G., Tiedje, J., 2003. The ribosomal database project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucl. Acids Res. 31, 442–443.

1999. Mathematics, and Applications and Space Studies Board. Size Limits of Very Small Microorganisms. National Academy Press, Washington, D.C.

del Cerro, A., Marquez, I., Guijarro, J.A., 2002. Simultaneous detection of *Aeromonas salmonicida*, *Flavobacterium psychrophilum*, and *Yersinia ruckeri*, three major fish pathogens, by multiplex PCR. Appl. Environ. Microbiol. 68, 5177–5180.

Denef, V.J., Park, J., Rodrigues, J.L.M., Tsoi, T.V., Hashsham, S.A., Tiedje, J.M., 2003. Validation of a more sensitive method for using spotted oligonucleotide DNA microarrays for functional genomics studies on bacterial communities. Environ. Microbiol. 5, 933–943.

Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J.D., Drmanac, R., 1998. Accurate sequencing by hybridization for DNA diagnostics and individual genomics. Nat. Biotechnol. 16, 54–58.

El Fantroussi, S., Urakawa, H., Bernhard, A., Kelly, J., Noble, P., Smidt, H., Yershov, G., Stahl, D., 2003. Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. Appl. Environ. Microbiol. 69, 2377–2382.

Enserink, M., 2001. Biodefense hampered by inadequate tests. Science 294, 1266–1267.

Gao, X., LeProust, E., Hua, Z., Srivannavit, O., Gulari, E., Yu, P., Nishiguchi, C., Xiang, Q., Zhou, X., 2001. A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. Nucl. Acids Res. 29, 4744–4750.

Garcia-Martinez, J., Bescos, I., Rodriguez-Sala, J.J., Rodriguez-Valera, F., 2001. RISSC: a novel database for ribosomal 16S–23S RNA genes spacer regions. Nucl. Acids Res. 29, 178–180.

Guy, R.A., Payment, P., Krull, U.J., Horgen, P.A., 2003. Real-time PCR for quantification of *Giardia* and *Cryptosporidium* in environmental water samples and sewage. Appl. Environ. Microbiol. 69, 5178–5185.

Hasan, A., Stengele, K.P., Giegrich, H., Cornwell, P., Isham, K.R., Sachleben, R.A., Pfleiderer, W., Foote, R.S., 1997. Photolabile protecting groups for nucleosides: synthesis and photodeprotection rates. Tetrahedron 53, 4247–4264.

Hashsham, S.A., Rouillard, J.M., Gao, L., Callister, S.J., Denef, V.J., Tsoi, T.V., Cole, J.R., Wibbenmeyer, J., Gualri, E., Tiedje, J.M., 2003. Highly parallel microbial detection using in situ synthesized flexible biochips. In: Proceedings of the 103rd General Meeting of the American Society for Microbiology, Washington, DC.

Huber, M., Mundlein, A., Dornstauder, E., Schneeberger, C., Tempfer, C.B., Mueller, M.W., Schmidt, W.M., 2002. Accessing single nucleotide polymorphisms in genomic DNA by direct multiplex polymerase chain reaction amplification on oligonucleotide microarrays. Anal. Biochem. 303, 25–33.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., Linsley, P.S., 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. 19, 342–347.

Isaacson, R.E., Torrence, M.E., 2002. The role of antibiotics in agriculture. A report by the American Academy of Microbiology. Washington, DC.

NRC (National Research Council): Panel on Biological Issues, Committee on Science and Technology for Countering Terrorism, 2002. Countering Bioterrorism: The Role of Science and Technology. The National Academy Press, Washington, DC.

Iwasaki, H., Ezura, Y., Ishida, R., Kajita, M., Kodaira, M., Knight, J., Daniel, S., Shi, M., Emi, M., 2002. Accuracy of genotyping for single nucleotide polymorphisms by a microarray-based single nucleotide polymorphism typing method involving hybridization of short allele-specific oligonucleotides. DNA Res. 9, 59–62.

Kaderali, L., Schliep, A., 2002. Selecting signature oligonucleotides to identify organisms using DNA arrays. Bioinformatics 18, 1340–1349.

Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., Madore, S.J., 2000. Assessment of the sensitivity and specificity of oligonucleotide (50-mer) microarrays. Nucl. Acids Res. 28, 4552–4557.

Klappenbach, J.A., Saxman, P.R., Cole, J.R., Schmidt, T.M., 2001. rrndb: the Ribosomal RNA Operon Copy Number Database. Nucl. Acids Res. 29, 181–184.

Kuczynska, E., Boyer, D.G., Shelton, D.R., 2003. Comparison of immunofluorescence assay and immunomagnetic electrochemiluminescence in detection of *Cryptosporidium parvum* oocysts in karst water samples. J. Microbiol. Methods 53, 17–26.

Li, F., Stormo, G., 2001. Selection of optimal DNA oligos for gene expression arrays. Bioinformatics 17, 1067–1076.

Lindroos, K., Sigurdsson, S., Johansson, K., Ronnblom, L., Syvanen, A.C., 2002. Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. Nucl. Acids Res. 30, Art. No. e.70.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., Lockhart, D.J., 1999. High density synthetic oligonucleotide arrays. Nat. Genet. 21, 20–24.

Liu, W., Mirzabekov, A., Stahl, D., 2001. Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. Environ. Microbiol. 3, 619–629.

Lockhart, D.J., Dong, H., Byrne, M., Follettie, M., Gallo, M., Chee, M., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., Brown, E., 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 14, 1675–1680.

Loy, A., Horn, M., Wagner, M., 2003. probeBase: an online resource for rRNA-targeted oligonucleotide probes. Nucl. Acids Res. 31, 514–516.

Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J., Schleifer, K., Wagner, M., 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.

Matveeva, O.V., Mathews, D.H., Tsodikov, A.D., Shabalina, S.A., Gesteland, R.F., Atkins, J.F., Freier, S.M., 2003. Thermodynamic criteria for high hit rate antisense oligonucleotide design. Nucl. Acids Res. 31, 4989–4994.

McCluskey, J., Dowson, C.G., Mitchell, T.J., 2002. The use of microarray technology for the analysis of *Streptococcus pneumoniae*. Comp. Funct. Genomics 3, 366–368.

Nass, M., 2001. Preparing a Medical Response to Bioterrorism. Washington, DC. Invited submission to the Committee on Government Reform for its November 14, 2001 hearing: Comprehensive Medical Care for Bioterrorism Exposure.

National Research Council, 2001. Classifying Drinking Water Contaminants for Regulatory Consideration. National Academy Press, Washington, DC.

Nicholas, K., Nicholas, H.J., Deerfield, D.I., 1997. GeneDoc: Analysis and visualization of genetic variation. EMBNEW.NEWS 4, 14.

Pozhitkov, A.E., Tautz, D., 2002. An algorithm and program for finding sequence specific oligo-nucleotide probes for species identification. BMC Bioinform. 3, Art. No. 9.

Rahman, S., 2002. Rapid large-scale oligonucleotide selection for microarrays. In: IEEE Computer Society Bioinformatics Conference (CSB'02). Stanford, CA, p. 54.

Relogio, A., Schwager, C., Richter, A., Ansorge, W., Valcarcel, J., 2002. Optimization of oligonucleotide-based DNA microarrays. Nucl. Acids Res. 30, Art. No. e.51.

Robertson, l., Gjerde, B., 2000. Effect of sample holding time on recovery of *Cryptosporidium* oocysts and *Giardia* cysts from water samples. Appl. Environ. Microbiol 66, 1724–1725.

Rouillard, J., Christopher, J., Zuker, M., 2002. OligoArray: genome-scale oligonucleotide design for microarrays. Bioinformatics 18, 486–487.

Safarikova, M., Safarik, I., 2001. Immunomagnetic seperation of *Escherichia coli* O26, O111 and O157 from vegetables. Lett. Appl. Microbiol. 33, 36–39.

Santalucia, J.J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. U.S.A. 95, 1460–1465.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.

Schoolnik, G.K., 2002. Functional and comparative genomics of pathogenic bacteria. Curr. Opin. Microbiol. 5, 20–26.

Scott, T.M., Lukasik, J., Farrah, S.R., 2002. Improved method for recovery of bacteriophage from large volumes of water using negatively charged microporous filters. Can. J. Microbiol. 48, 305–310.

Stevens, P., Henry, M., Kelso, D., 1999. DNA hybridization on microparticles: determining capture-probe density and equilibrium dissociation constants. Nucl. Acids Res. 27, 1719–1727.

Straub, T., Daly, D., Wunshel, S., Rochelle, P., DeLeon, R., Chandler, D., 2002. Genotyping *Cyptospoildium parvum* with an hsp70 single-nucleotide polymorphism microarray. Appl. Environ. Microbiol. 68, 1817–1826.

Straub, T.M., Chandler, D.P., 2003. Towards a unified system for detecting waterborne pathogens. J. Microbiol. Methods 53, 185–197.

Sturbaum, G.D., Klonicki, P.T., Marshall, M.M., Jost, B.H., Clay, B.L., Sterling, C.R., 2002. Immunomagnetic separation (IMS)-fluorescent antibody detection and IMS-PCR detection of seeded *Cryptosporidium parvum* oocysts in natural waters and their limitations. Appl. Environ. Microbiol. 68, 2991–2996.

Talbot, D., 2001. Detecting bioterrorism. Technol. Rev., 35–37.

The chipping forecast, 1999. Special Suppl. Nat. Genet., 21 (1), 1–60.

US EPA, 2001a. Method 1622: Cryptosporidium in water by filtration/IMS/FA. Office of Water, United States Environmental Protection Agency, Washington, DC.

US EPA, 2001b. Method 1623: Cryptosporidium and Giardia in water by filtration/IMS/MFA. Office of Water, United States Environmental Protection Agency, Washington, DC.

Urakawa, H., Noble, P., El Fantroussi, S., Kelly, J., Stahl, D., 2002. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. Appl. Environ. Microbiol. 68, 235–244.

van Beuningen, R., van Damme, H., Boender, P., Bastiaensen, N., Chan, A., Kievits, T., 2001. Fast and specific hybridization using flow-through microarrays on porous metal oxide. Clin. Chem. 47.

Wang, D., Coscoy, L., Zylberberg, M., Avila, P., Boushey, H., Ganem, G., DeRisi, J., 2002a. Microarray based detection and genotyping of viral pathogens. Proc. Natl. Acad. Sci. 99, 15687–15692.

Wang, R., Beggs, M., Robertson, L., Cerniglia, C., 2002b. Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. FEMS Microbiol. Lett. 213, 175–182.

Watanabe, K., Nelson, J.S., Harayama, S., Kasai, H., 2001. ICB database: the gyrB database for identification and classification of bacteria. Nucl. Acids Res. 29, 344–345.

Wells, J.M., Bennik, M.H.J., 2003. Genomics of food-borne bacterial pathogens. Nutr. Res. Rev. 16, 21–35.

Wilson, K., Wilson, W., Radosevich, J., DeSantis, T., Viswanathan, V., Kuczmarski, T., Andersen, G., 2002a. High-density microarray of small-subunit ribosomal DNA probes. Appl. Environ. Microbiol. 68, 2535–2541.

Wilson, W.J., Strout, C.L., DeSantis, T.Z., Stilwell, J.L., Carrano, A.V., Andersen, G.L., 2002b. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Mol. Cell. Probes 16, 119–127.

Wu, C.F., Valdes, J.J., Bentley, W.E., Sekowski, J.W., 2003. DNA microarray for discrimination between pathogenic O157:H7 EDL933 and non-pathogenic *Escherichia coli* strains. Biosens. Bioelectron. 19, 1–8.

Yguerabide, J., Yguerabide, E.E., 2001. Resonance light scattering particles as ultrasensitive labels for detection of analytes in a wide range of applications. J. Cell Biochem. Suppl. 37, 71–81.

Zhang, Z., Wilson, R., Fox, G., 2002. Identification of characteristic oligonucleotides in the bacterial 16S ribosomal RNA sequence dataset. Bioinformatics 18, 244–250.

Zheng, D.D., Alm, E.W., Stahl, D.A., Raskin, L., 1996. Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. Appl. Environ. Microbiol. 62, 4504–4513.