Molecular and Cellular Probes (2002) **16**, 119–127
doi:10.1006/mcpr.2001.0397, available online at http://www.idealibrary.com on IDEAL®

Sequence-specific identification of 18 pathogenic microorganisms using microarray technology

W. J. Wilson, C. L. Strout, T. Z. DeSantis, J. L. Stilwell, A. V. Carrano and G. L. Andersen*

Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

(Received 13 July 2001; accepted for publication 22 October 2001)

We have developed a Multi-Pathogen Identification (MPID) microarray for high confidence identification of eighteen pathogenic prokaryotes, eukaryotes and viruses. Analysis of amplified products from pathogen genomic DNA using microarray hybridization allows for highly specific and sensitive detection, and allows the discrimination between true amplification products and false positive amplification products that might be derived from primers annealing to non-target sequences. Species-specific primer sets were used to amplify multiple diagnostic regions unique to each individual pathogen. Amplified products were washed over the surface of the microarray, and labelled with phycoerythrin-streptavidin for fluorescence detection. A series of overlapping 20-mer oligonucleotide probes hybridize to the entire diagnostic region, while parallel hybridizations on the same surface allow simultaneous screening for all organisms. Comparison to probes that differ by a single mismatch at the central position reduced the contribution of non-specific hybridization. Samples containing individual pathogens were analyzed in separate experiments and the corresponding species-specific diagnostic regions were identified by fluorescence among their highly redundant probe sets. On average, 91% of the 53 660 pathogen probes on the MPID microarray performed as predicted. The limit of detection was found to be as little as 10 fg of *B. anthracis* DNA in samples that were amplified with six diagnostic primer-pairs. In contrast, PCR products were not observed at this concentration when identical samples were prepared and visualized by agarose gel electrophoresis.

KEYWORDS: pathogen detection, PCR, GeneChip®, microarray, bioterrorist threat organisms.

INTRODUCTION

The need to adequately monitor biological threats to assure public health has led to research on advanced pathogen detection technologies.[1,2] Polymerase chain reaction (PCR) technologies have proven to be the most accurate and quick methods for identifying the presence of microbial pathogens.[3] However, false positive amplifications can occur among related organisms in the environment sharing similar genetic sequences, and false negative amplifications can result from sequence variation in the primer binding sites of the target microbe's diagnostic region or from low sensitivity of a reaction. Ideally, multiple unique diagnostic regions (both chromosomal and plasmid) should be monitored for increased confidence of correct identification of a pathogen. Many pathogens contain mobile genetic elements that can potentially interfere with proper identification using single loci detection systems.[4]

* Author to whom correspondence should be addressed: Gary L. Andersen, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, L-441, P.O. Box 808, Livermore, CA 94551, USA. Tel. +925 423 2525; Fax. +925 422 2282; E-mail: andersen2@llnl.gov

0890–8508/02/020119+09 $35.00/0

Recent development of high-density oligonucleotide arrays allows massively parallel hybridizations to occur on the same surface, permitting high levels of probe redundancy and multiple independent detections of a diagnostic DNA sequence.[5] In the Affymetrix approach, oligonucleotide probes are synthesized directly on a glass substrate coated with linkers containing a photolabile protecting group. A photolithographic mask exposes selected areas of the glass, removing the protecting group to allow synthesis of probes directly on the surface with solid-phase combinatorial chemistry.[6] Affymetrix microarrays have been successfully developed to detect both the presence of *Mycobacterium* species and their antibiotic susceptibility.[7] Polymorphisms in a 200 bp sequence region of the 16S rRNA gene were used to correctly identify any of 26 *Mycobacterium* species. A 200-bp sequence region in the *rpoB* gene containing 51 known mutations that cause rifampicin resistance was also on the microarray; it was able to identify all 15 resistant isolates.[7] A microarray described by Israel *et al.* used whole genome sequence of *Helicobacter pylori* to identify differences in gene content among pathogenic strains. Differences in several genes including a large deletion in the CAG pathogenicity island were used to distinguish an attenuated strain from a more virulent one.[8]

We have developed a Multi-Pathogen Identification (MPID) PCR-coupled high-density microarray that contains 53 660 oligonucleotide probes complementary to the internal nucleotide sequence of 142 unique diagnostic regions of 11 bacteria, 5 RNA viruses, and 2 eukaryotes identified as candidates for biological terrorism. The microarray thus allows detection based on the sequence-specific hybridization of amplified products characteristic of each pathogen. Each diagnostic region for a particular pathogen is represented on the microarray as a highly redundant set of probes (for example, see Figure 1). PCR-amplified targets obtained from pathogen containing samples are fragmented and biotinylated prior to incubation on the microarray. After hybridization and a series of stringency washes the target is stained with a streptavidin-conjugated fluorophore and the interaction of the target with specific probes is measured with epifluorescence confocal microscopy using an argon-ion laser. Integrated software measures fluorescence intensity of all probes on the array as the basis for a positive identification. This same approach can be used to identify virtually any group of organisms of interest.

MATERIALS AND METHODS

Diagnostic region and primer selection

Two strategies were used to choose each pathogen's diagnostic regions of DNA. Sequences were derived from pathogenicity and virulence genes and also by shotgun sequencing of cloned fragments from previously uncharacterized DNA regions throughout the pathogen genomes. Primers were designed to amplify these regions in multiplex reactions: all primer pairs had similar T_m ($57° \pm 2°$), lacked significant secondary structure, contained 45–60% GC, and had a sensitivity level of at least 10 ng initial DNA concentration. Amplified products were designed to be under 500 bp, an important factor when working with environmental samples containing degraded DNA.

Primer pairs were screened against the GenBank public database for potential interference from similar nucleotide sequence.[9] Nonetheless, empirical testing from a range of environmental samples was the best predictor for the success of a primer pair in the detection of a specific pathogen. Primer pairs were tested to ensure that all strains in our collection for each pathogen amplified DNA of the predicted size. The potential diagnostic regions were screened against a panel of closely related organisms based on their phylogenetic relationship and against DNA from other sources likely to occur in clinical or environmental samples for cross-amplification. Examples of DNA sources used to check to performance of the primers included several mixtures of soil, fresh and ocean water, and human blood. Approximately 5% of the primer-pairs satisfactorily passed the selection process by meeting these criteria. A minimum of 3 and maximum of 10 diagnostic regions were chosen to represent each organism, depending on how many primer pairs were left after the selection process. Their internal diagnostic sequence was used to create probes for the microarray. Each diagnostic region for a particular pathogen was represented on the microarray as a set of probes averaging 189 probe pairs, 20-mer in length, per diagnostic region, and 1491 probe pairs for each organism.

PCR amplification of pathogens

Amplification reactions were performed on a Perkin Elmer 9600 thermocycler with 2·5 U Taq polymerase (Boehringer Mannheim, Indianapolis, IN, USA), 1 ng of genomic DNA, and 10 mm dNTP. PCR conditions of all primer pairs were optimized to be the same: 1 cycle of 4 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at 52°C, and 1 min at 72°C, and a final extension

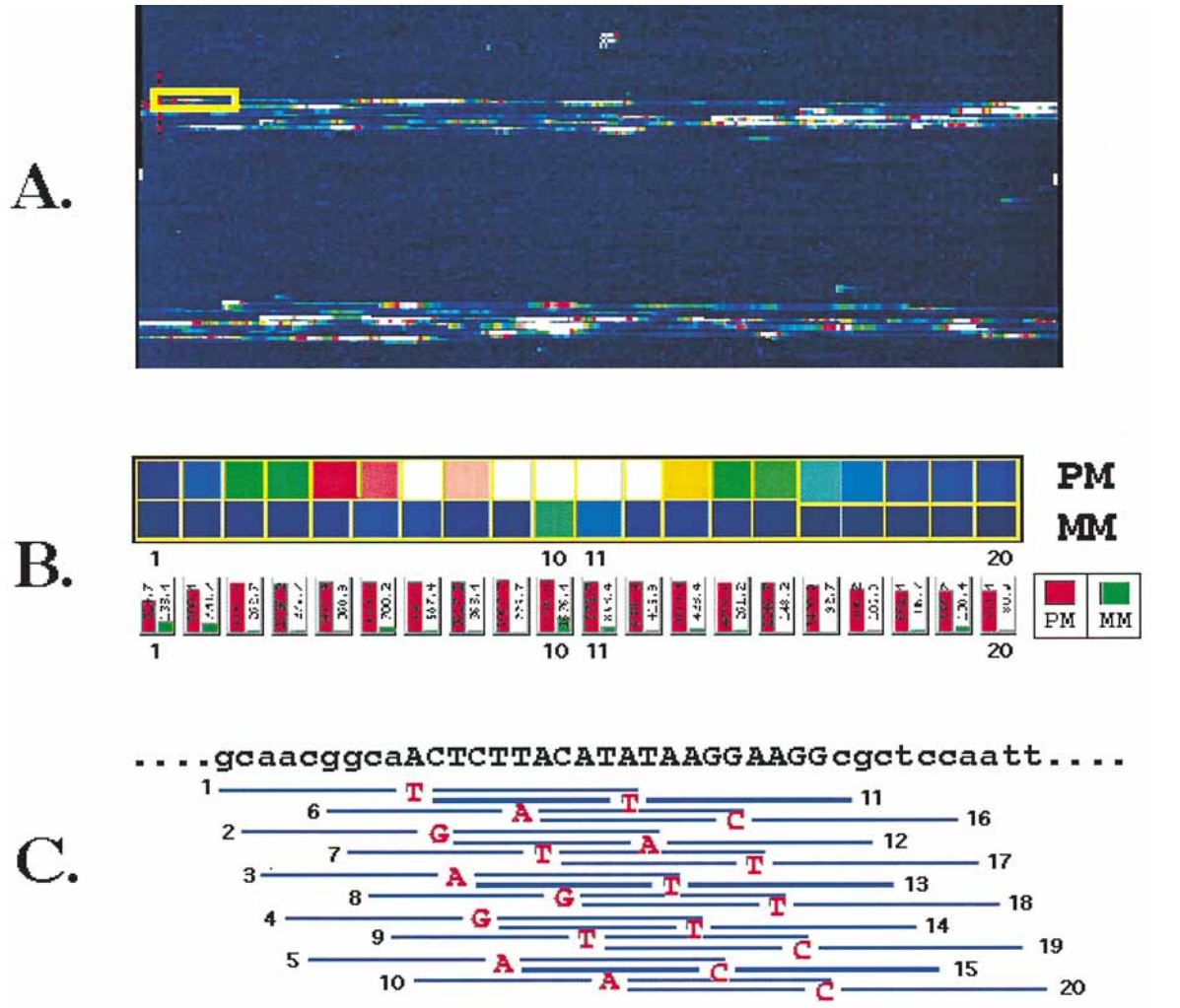

Fig. 1. Pathogen screening with oligonucleotide arrays. (A) Scanned image of a portion of the MPID microarray showing relative fluorescence intensity (in pseudocolor scale, increasing from blue to white) of probes corresponding to 8 *Yersinia pestis*-diagnostic regions from which amplicons and probes are derived. (B) Magnification of the first 20 probe pairs of probe set Yp5 (from *Y. pestis caf1* virulence gene) hybridizing to *Y. pestis* target DNA (yellow box in A above); each 20-base perfect match (PM) probe square distinctly matching the pathogen's diagnostic sequence is paired with mismatch (MM) partner, physically located in the adjacent, lower position on the microarray. The relative fluorescence intensity results for each probe pair is graphically displayed with red bars for PM intensity and green bars for MM intensity. This probe set continues with additional 328 probe pairs. (C) Microarray probe design. Overlapping sequences advance by one nucleotide over the entire diagnostic sequence, essentially sequencing the entire region by hybridization. The beginning of the jcaf1 gene sequence (capitalized) is shown, with the first 20 corresponding MM probes underneath, identical to the PM probes except for a mismatch at position 11 of its partners 20-meroligonucleotide (substitution highlighted in red).

of 7 min at 72°C. The PCR amplicon products were observed on 3% NuSieve 3 : 1 (FMC Corporation, Rockland, ME, USA) gels after electrophoresis and ethidium bromide staining.

A standardized RT-PCR was used for amplification of viral RNA. RT reactions were performed on a Perkin-Elmer 9600 with 25 U M-MuLV reverse transcriptase (Gibco-BRL, Bethesda, MD), 20 U RNase inhibitor (PE Biosystems, Foster City, CA, USA), RT buffer containing 10 mM Tris-HCl (pH 7·6), 72 mM KCl, and 1·5 mM $MgCl_2$, 10 mM DTT, 800 µm each dNTP, 100 pmol reverse primer and 1 ng Genomic RNA in DEPC-treated RNase free water (Research Genetics). RT Reaction conditions were: 1 cycle of 15 min at 42°C, 1 cycle of 5 min at 99°C, and 1 cycle of 5 min at 5°C. A 20 µl product from the RT reaction was used in the PCR reaction containing 50 pmol forward primer, 2·5 U Advantage2 Polymerase and its buffer system (Clon-Tech, Palo Alto, CA, USA). PCR reaction conditions were: 1 cycle of 2 min at 95°C, 35 cycles of 15 s at 95°C and 30 s at 60°C, 1 cycle of 7 min at 72°C. Products were observed as described above.

For sensitivity studies, Advantage2 Polymerase Mix (Clontech, Palo Alto, CA, USA) was employed. PCR conditions used were those for genomic DNA amplification. Sensitivity was assessed using six *Bacillus anthracis* primer sets, *B. anthracis* DNA starting concentrations of 1 pg, 100 fg, 10 fg, and 0, and amplification cycles of 35, 45, and 55.

PCR product preparation and array hybridization

PCR amplified diagnostic fragments were pooled and a 45 µl aliquot was fragmented to an average length of approximately 50 bp with a 5 µl fragmentation solution containing 0·25 units of DNaseI (Affymetrix, Santa Clara, CA, USA). In our experience, fragments under 100 bp had a stronger hybridization signal than full-length amplicons on the microarray. Fragmentation was performed on a PE 9600 thermocycler for 12 min at 25°C followed by enzyme inactivation by heating at 95°C for 10 min. A 21·5 µl aliquot of the fragmented DNA was then 3′ end-labelled with 50 units of terminal transferase (Boehringer Mannheim) in 1 × TdT buffer (0·2 M potassium cacodylate, 25 mm tris-HCl [pH 6·6], 0·25 mg/ml BSA; Boehringer Mannheim) and 2·5 mm cobalt chloride for a 2 h incubation at 37°C.

Probe arrays were prehybridized with 200 µl of Rapid Hyb buffer (Amersham, Arlington Heights, III, USA) containing 0·8 M tetramethylammonium chloride (to improve specific annealing of the fragmented target to the probes;[10]) in a custom hybridization rotisserie oven (Stovall Life Science, Greensboro, NC, USA) for 10 min at 45°C and 60 rev/min. A target cocktail containing 17 µl of the fragmented and end-labelled DNA in the hybridization buffer was heat denatured at 95°C for 5 min followed by 45°C for 5 min before being placed in the microarray. Hybridization was performed in the rotisserie oven for 3 h at 45°C and 60 rev/min. The microarray was washed and stained in a GeneChip Fluidics Station 400 (Affymetrix, Inc.) according to the following protocol: 4 washes with 6 × SSPE/0·005% Triton X-100, 2 washes with 1 × SSPE/0·005% Triton X-100, stain with 2 µg ml^{-1} streptavidin phycoerythrin (SAPE) conjugate (Molecular Probes, Eugene, OR, USA) and 0·95 mg/ml acetylated BSA (Sigma, St Louis, MO, USA) in 6 × SSPE/0·005% Triton X-100) for 15 min at 40°C. Excess SAPE was removed with 3 washes with 1 × SSPE/0·005% Triton X-100 at 25°C. The stained probe array was scanned using an argon ion laser confocal microscope (GeneChip Scanner 50; Molecular Dynamics) and a photomultiplier tube detected the emission through a 560 nM longpass filter.[11]

Data analysis

The image obtained by the scan was analyzed using Affymetrix GeneChip Analysis Suite, version 3.3.[12] Using the default algorithms, background fluorescence was subtracted from all probe cells and an average noise value (Q) was determined from the variations in intensity among pixels within the same probe cell. Each probe pair is comprised of a 20-mer perfectly matching (PM) probe and a mismatch control probe (MM) that differs by a single base difference at a central position. The MM probe acts as a specificity control to allow discrimination between random cross-hybridization and true signals. Each probe pair was independently evaluated to determine if the intensity of fluorescence from the PM cell was greater than 1.5 times the intensity from the MM cell and the PM–MM intensity difference was four times greater than the noise (Q). Probe pairs meeting these criteria were called positive. Inversely, probe pairs were deemed negative when the MM fluorescence intensity was considerably greater than that of the PM as judged by same two criteria. Pairs with intensities that did not exceed these thresholds in either direction were regarded as neutral.

The likelihood of the presence of the potential PCR products amplified from the diagnostic genome sequences was determined by two metrics, the Positive Fraction and the Log Average Ratio. The Positive Fraction was the ratio of probe pairs scored as positive divided by the total number of probes in a set. The Log Average Ratio described the extent of hybridization across a probe set by determining the ratio of the PM to MM intensities for each probe pair, taking the Log of the resulting values, then averaging them across the probe set.

RESULTS

Hybridization specificity

The performance of the MPID microarray for sequence-specific detection of PCR amplified products was assessed. Samples containing individual pathogens were amplified with pathogen-specific primers and placed on separate microarrays. After a 3-h hybridization, the sample was drained from the array and a series of washes removed weakly hybridizing amplification products. The strength of hybridization with sample DNA (target) labeled with streptavidin-phycoerythrin conjugate was measured for all 53 660 probes using a confocal fluorescence microscope (provided by Affymetrix, Santa Clara, CA). A sequence-specific interaction with the target DNA

Table 1. The specificity of probe pairs for single organism detection in separate experiments

Organism	Strain	Probe Sets[a]	Total Probe Pairs[b]	Positive Fraction[c]
Prokaryotes				
Bacillus anthracis	Ames	8	2333	0·87
Brucella abortus, (Biotype 1)	ATTC-23448	10	1414	0·97
Clostridium botulinum, (Type A)	ATTC-25763	9	1470	0·84
Clostridium perfringens	ATTC-13124	10	1754	0·86
Coxiella burnetii	Nine Mile	9	1022	0·89
Francisella tularensi	83A5698	8	1786	0·97
Rickettsia prowazekii	ATTC-VR-142	7	828	0·90
Staphylococcus aureus	ATTC-14458	8	1483	0·87
Vibrio cholerae	ATTC-14035	9	1844	0·96
Vibrio alginolyticus	ATTC-17749	9	1173	0·92
Yersinia pestis	ATTC-19428	8	2376	0·95
Viruses				
Western equine encephalitis	Tr25717	7	1342	0·89
Eastern equine encephalitis	Ten Broeck	6	1495	0·92
Marberg	Musoke	3	822	1·00
Ebola	Mayinga	4	690	0·89
Venezuelan equine encephalitis	VR-1250CAF	7	1053	0·89
Eukaryotes				
Alexandrium cantenella	Bgt1	10	1591	0·89
Fusarium sporotrichoides	ATCC-24631	8	2354	0·94

[a] Total number of organism specific probe sets, one probe set per amplicon.
[b] Sum of all probe pairs from all probe sets that are unique to each organism.
[c] Fraction of organism-specific probes scoring positive when amplified target DNA from the same organism is hybridized to the microarray.

for each probe pair was calculated by measuring the ratio of fluorescence intensity of the sequence-specific probe (PM) and the mismatch control probe (MM). When the PM probe has significantly higher fluorescence intensity than the MM probe (arbitrarily defined as PM/MM $\geq 1{\cdot}5$) the probe pair was called positive. A series of probe pairs were designed to overlap the sequence for each diagnostic region, advancing by one nucleotide over the entire amplicon, creating a highly redundant probe set (Fig. 1C). Probe sets defining 142 distinct diagnostic regions were designed to allow identification based on the internal sequence of multiple, independent amplicons for 11 bacteria, 5 RNA viruses, and 2 eukaryotic microorganisms on the microarray. Results of independent experiments for 18 organisms for identification of each pathogen represented on the microarray are presented in Table 1. For each pathogen tested, between 3 and 10 probe sets were used to identify species-specific diagnostic regions. A probe set was scored as present when a majority of its component probe pairs were called positive by analysis with Affymetrix software.

We found an average of 91% of all probe pairs were positive when samples with complementary sequences exactly matching a set of PM probes were placed on a microarray. Each pathogen was correctly identified by the sum of the interactions of multiple independent probes generated by the hybridization of amplicons of predicted sequence. For example, a sample containing *Clostridium botulinum* amplified 9 diagnostic regions characteristic of this pathogen. Probe sets on the microarray for detection of these amplicons had between 75% and 94% of the individual probe pairs scored as positive with the remainder not scored (PM and MM were essentially the same) or scored negative. Combining all 1470-probe pairs of the 9 *Clostridium botulinum* specific probe sets resulted in a positive fraction of 0·84, the lowest value of all tested pathogens (Table 1). The absolute fluorescence intensity of the PM probes were in all cases substantially higher in probe sets in which corresponding target DNA was present. However, we found that the pattern of hybridization to the hundreds of PM/MM pairs in a probe set was a more accurate predictor of target diagnostic region presence, especially at low concentrations, than strength of signal alone.

The MPID microarray's reproducibility was assessed twofold; by comparing different target preparations of *Brucella abortus* DNA on two chips and by comparing the same *B. abortus* DNA target preparation applied to two separate chips (Fig. 2: exp 1 and 2, respectively). The sample-to-sample variation

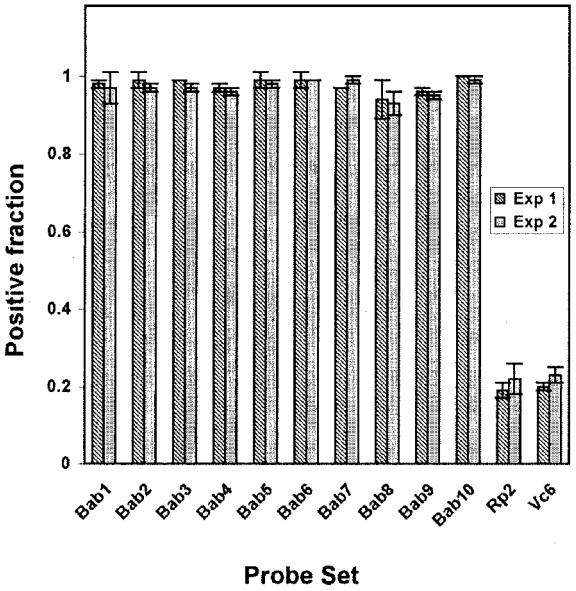

Fig. 2. Experimental variability within *Brucella abortus* probe sets. Exp 1, PCR variation: two separate amplifications of *B. abortus* PCR products. Exp 2, chip-to-chip variation: same sample preparation processed on two different MPID microarrays. Means (SEM) for duplicate samples of both Exp 1 and Exp 2. All *B. abortus* probe sets reported positive; negative probe sets from *Vibrio cholerae* (Vc6) and *Rickettsia prowazekii* (Rp2) are shown for comparison.

of the 10 *B. abortus* probe sets (comprising 1414 probe pairs) was very low; indicating that target DNA consistently finds its unique probes on the microarrays.

Hybridization sensitivity

The sensitivity of the MPID microarray for the detection for *Bacillus anthracis* (anthrax) was assessed using 6 of the species-specific primer pairs for PCR amplification from sample extracts containing 1 pg, 100 fg and 10 fg of genomic DNA as well as a no DNA control. Detection of the 6 diagnostic regions was achieved by hybridization to the array with even the lowest concentration, equivalent to 2 genome copies, using 35 cycles of amplification (Fig. 3); visible bands were not observed at this concentration by standard gel electrophoresis. Increasing the number of PCR cycles to 45 or 55 resulted in a poorly resolved set of amplified products on a gel, including spurious bands and a background smear suggestive of non-specific amplification. However, the background Positive Fraction and Log Average values increased, which most likely correlates to the same non-specific hybridization seen in the gels (data not shown). We concluded that the 35-cycle PCR was best for GeneChip target production, and used 35 cycles for all further tests. Analyzing the pattern of multiple PM/MM probes had the advantage of a much higher level of sensitivity for identification of the diagnostic sequences when compared to the analysis of hybridization strength of PM probes alone.

Detection in a DNA mixture

To assess the effectiveness of the MPID microarray for pathogen detection in complex environmental samples, we spiked a DNA preparation from an air sample with different concentrations of genomic DNA from *Francisella tularensis* and *Yersinia pestis*. The air sample was collected in Livermore, California on a polyester filter with a 1-micron pore size using a customized air-sampling device. Two ng of purified DNA from air (equivalent to 4×10^3 l air) was mixed with a combination of 50 pg, or 0·5 pg of genomic DNA from both *F. tularensis* and *Y. pestis*. The DNA mixtures and an air DNA only control were used for amplification of the diagnostic regions from both pathogens. All probe sets to both *F. tularensis* and *Y. pestis* were detected on a microarray when the sample containing 50 pg of each pathogen and 2 ng of air DNA was hybridized to an array (Table 2). The highest two background hybridizations observed were to a probe set for *Brucella abortus* and a probe set for *Vibrio cholerae* with 48% of the probes having a higher signal for the PM than the MM (Positive Fraction 0·48) in both cases. To reduce the contribution of false positive signals, we set a limit of greater than 80% of the probes having a significantly higher signal for the PM than the MM probe, which we consider highly unlikely to occur by chance. At the lower concentration of 500 fg (100 genome equivalents) detection of specific pathogen probe sets was variable with 5 of 8 positive for *F. tularensis* and 3 of 8 positive for *Y. pestis* (Table 2).

DISCUSSION

Coupling PCR technology with the hybridization of the resulting amplification products on microarrays yields an extremely powerful detection system. The reliability of the detection system is further enhanced by the simultaneous identification of multiple diagnostic regions to obtain a highly accurate fingerprint of each pathogen. We have used this technology to design a custom microarray that contains 53 660 oligonucleotide probes to unique diagnostic regions of 18 pathogens identified as candidates for biological terrorism. We have demonstrated the sensitivity of detection with this system with as low as 10 fg of

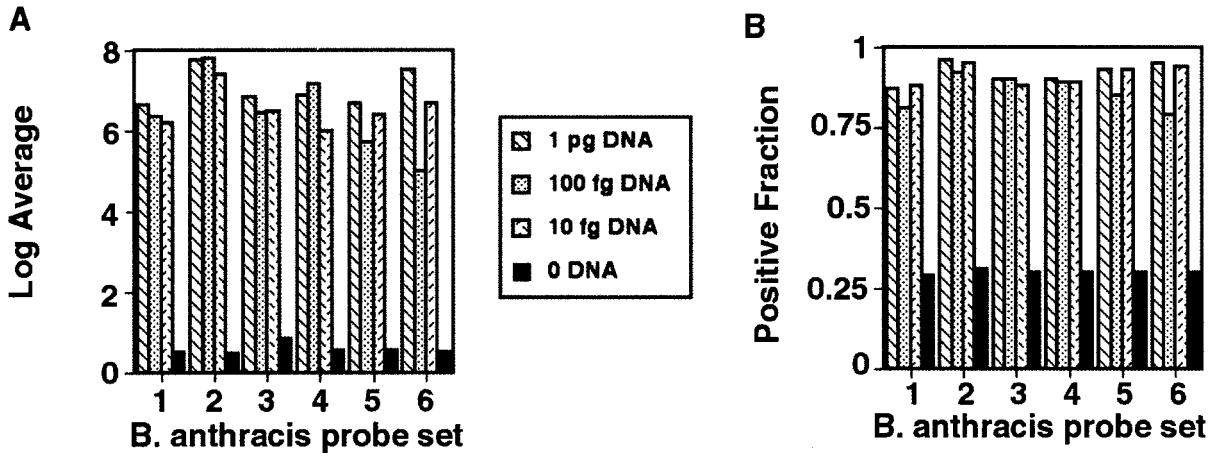

Fig. 3. Sensitivity of the MPID microarray assay using 1 pg, 100 fg, 10 fg *B. anthracis* DNA starting concentrations and a no DNA control with 35 cycles of amplification. Positive identification was observed with a DNA equivalent of only 2 genome copies (~10 fg). (A) A statistical measure of confidence that a DNA fragment was present using a Log Average Ratio. The value is calculated taking a log of the average of the intensity ratio for a probe set [log (PM/MM)] and multiplying by 10. Values below 1 suggest random cross hybridization and values above 4 indicate a high confidence that the DNA region is present. (B) Fraction of probe pairs positive (positive fraction) for each of 6 *B. anthracis* specific probe sets.

purified genomic DNA or 500 fg of pathogen DNA in a complex environmental sample containing potential PCR inhibitors and competing target DNA from other microorganisms. This microarray has potential applications in the public health and law enforcement arenas, where accurate pathogen monitoring and identification is important for the early treatment of individuals in natural infectious outbreaks or in threats of bioterrorism.

Typical samples from epidemiological surveys consist of a large and diverse group of background organisms, which may also contain regions of similarity to the target sequence used for diagnostic PCR amplification. The genetic complexity of this non-target DNA may also generate the amplification of a 'false positive' product when a primer anneals to a mismatched sequence, especially as the number of cycles increases.[13] Detection of amplification products by gel electrophoresis can be misleading or subject to interpretation under these circumstances. The occasional 'failure' of even highly tested primer pairs has been observed, usually attributed to the primers annealing to background organisms amplifying from complex environmental samples. For example, an IS900 insertion element was used to develop specific PCR primers for *Mycobacterium avium* subsp. *paratuberculosis*.[14,15] It was then found to cross react with environmental *Mycobacteria* sp. in ruminant feces; sequence analysis of 3 different 'false positive' products revealed 71% and 79% similarity to the region of IS900 amplified.[16] Verification of the exact sequence information between the primers of the diagnostic amplified products is the most accurate way of discriminating between true and false amplification products; the MPID microarray employs high-throughput hybridization of product sequences for accurate identification.

Equally important for establishing a reliable pathogen identification method is to have multiple confirmatory diagnostic primer pairs amplify regions of unique sequence. Bacterial genomes are extremely dynamic, and the ability of organisms to acquire genetic information (antibiotic resistance, pathogenicity and virulence genes, for example) from one another in the environment is now well known.[17] Lateral gene transfer allows a microorganism to expand its ecological niche, but can make it difficult to identify an organism based on one diagnostic region. Increasing the number of genetic regions examined increases the confidence of correct identification. Assuming diagnostic regions are pair-wise uncorrelated, an organism's false positive rate of identification becomes the product of each diagnostic region's false positive rates and the false negative rate is the sum of the negative rates.[18]

Although the microarray can detect low levels of DNA products, amplification of the pathogen's characteristic diagnostic sequences is still required to produce sufficient signal for organism identification. All PCR and RT-PCR (for amplification of virus RNA) reactions were universally standardized to the same reaction conditions. This allows PCR or RT-PCR amplifications to be performed concurrently in a high-throughput manner. After amplification, the set of

Table 2. Sensitivity of the MPID microarray assay for the detection of a pathogen mixture in a background of genomic DNA from an air sample.

	Francisella tularensis								*Yersinia pestis*							*Bab*[a]	*Vc*[b]
Pathogen DNA spiked into (Air DNA)[c]	Ft1	Ft2	Ft3	Ft4	Ft5	Ft6	Ft7	Ft8	Yp1	Yp2	Yp3	Yp4	Yp5	Yp6	Yp7	Yp8	Bab5	Vc7
50 pg (2 ng)	1·00[d]	1·00	1·00	0·99	1·00	1·00	0·99	1·00	0·98	0·98	1·00	1·00	0·99	0·97	0·96	0·95	0·48	0·48
0·5 pg (2 ng)	0·36	0·96	1·00	0·63	0·89	0·51	0·82	0·84	0·93	0·68	0·48	0·50	0·90	0·81	0·30	0·25	0·46	0·47
0 pg-(2 ng)	0·31	0·22	0·26	0·09	0·14	0·18	0·28	0·11	0·43	0·34	0·26	0·43	0·20	0·24	0·26	0·33	0·44	0·51

[a] *Brucella abortus.*
[b] *Vibrio cholerae.*
[c] Concentration each of *F. tularensis* and *Y. pestis* DNA spiked into an aerosol sample from Livermore, Californa. Within the parentheses is the concentration of environmental DNA from air.
[d] Positive Fraction value for each probe set. Values greater than 0·8 indicate presence of the diagnostic region in the sample.

diagnostic amplicons for each pathogen were pooled, fragmented and biotin end-labeled. The biotin-labeled fragment mixture was hybridized to an array at high stringency with TMAC to reduce GC bias during hybridization reactions for all sequence combinations.[19,20] The probe array was then washed, stained and scanned using an argon-ion laser confocal microscope. Preparation and analysis of an amplified sample currently requires 6 h. Recently, a microfabricated sample preparation system was described for automated control of all liquid handling steps in the GeneChip sample processing that would further reduce sample analysis time.[21]

Nucleic acid amplification (PCR) technologies have proven to be among the most rapid and accurate methods for identifying the presence of microbial pathogens.[3] One major advancement in PCR technologies has been quantitative, real time PCR; this has been accomplished using fluorescence energy transfer (FRET) probes that eliminate the need for gel electrophoresis to view products.[22] This technology has been successfully employed to detect many different pathogens, including fungi.[23] Two FRET probes have been used successfully together in a multiplex reaction to detect *Erwinia herbicola*.[24] A limitation of this system is its multiplex PCR capabilities; detection of multiple PCR products is limited to the spectral overlap of available FRET probes. By contrast, high-density arrays simultaneously detect many thousands of bases that are used to distinguish the multiple DNA fragments from each other in a single hybridization step. This allows for high throughput screening of many microorganisms and less interpretation of the data for making a positive identification.

In summary, we have designed a microarray for the identification of 18 different pathogens. This detection system has the ability to screen for the presence of pathogens by the examination of specific diagnostic regions and by creating a fingerprint specific to each microbe. Using a highly redundant set of probes that allows for sequence-specific identification of each amplified product, this method offers increased confidence of accurate detection over existing technology. Amplified products hybridize to the diagnostic probes in predicted patterns that are recognized by the analysis software for identification at levels below the limit of detection by agarose gel visualization. The actual limits of detection for pathogen specific diagnostic regions were as low as 500 fg using spiked environmental samples. At this level of sensitivity, the MPID microarray offers an alternative to current methods for high throughput pathogen screening.

ACKNOWLEDGEMENTS

This work was performed under the auspices of the US Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract #W-7405-ENG-48 and was funded by the Central Intelligence Agency, Office of Research and Development.

We appreciate the technical assistance of Marge Seagraves, Jennifer Kidd, and Audrina Lee.

REFERENCES

1. Abelson, P. H. (1999). Biological warfare. *Science* **286**, 1677.
2. Henderson, D. A. (1998). Bioterrorism as a public health threat. *Emerging Infectious Diseases* **4**, 488–92.09.
3. Belgrader, P., Benett, W., Hadley, D. *et al.* (1999). PCR detection of bacteria in seven minutes. *Science* **284**, 449–50.
4. Davison, J. (1999). Genetic exchange between bacteria in the environment. *Plasmid* **42**, 73–91.
5. Lipshutz, R. J., Fodor, S. P. A., Gingeras, T. R. & Lockhart, D. J. (1999). High Density synthetic oligonucleotide arrays. *Nature Genetics* **21**, 20–4.
6. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–73.
7. Troesch, A., Nguyen, H., Miyada, C. G. *et al.* (1999). Mycobacterium species identification and rifampicin resistance testing with high-density DNA probe arrays. *Journal of Clinical Microbiology* **37**, 49–55.
8. Israel, D. A., Salama, N., Arnold, C. N. *et al.* (2001). Helicobacter pylori strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. *Journal of Clinical Investigation* **107**, 611–20.
9. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2000). GenBank. *Nucleic Acids Research* **28**, 15–18.
10. Chevet, E., Lemaitre, G. & Katinka, M. D. (1999). Low concentrations of tetramethylammonium chloride increase yield and specificity of PCR. *Nucleic Acids Research* **23**, 3343–4.
11. (1997) *GeneChip® p53 Assay manual*, Santa Clara, CA: Affymetrix, Inc.
12. (1999). *GeneChip® Analysis Suite User Guide*, Version 3.3. Pp. 383–431. Santa Clara, CA: Affymetrix, Inc.
13. Innis, M. A. & Gelfand, D. H. (1990). Optimization of PCRs. In *PCR Protocols a Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sinsky, J. J. & White, T. J., eds) Pp 3–11. San Diego: Academic Press.
14. Green, E. P., Tizard, M. L., Moss, M. T. *et al.* (1989). Sequence and characteristics of IS900, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. *Nucleic Acids Research* **17**, 9063–73.
15. Collins, D. M., Gabric, D. M. & De Lisle, G. W. (1989). Identification of a repetitive DNA sequence specific to *Mycobacterium paratuberculosis*. *FEMS Microbiology Letters* **60**, 175–8.
16. Cousins, D. V., Whittington, R., Marsh, I., Masters, A., Evans, R. J. & Kluver, P. (1999). Mycobacteria distinct from *Mycobacterium avium* subsp. *paratuberculosis* isolated from the faeces of ruminants possess IS900-like sequences detectable by IS900 polymerase chain reaction: implications for diagnosis. *Molecular and Cellular Probes* **14**, 431–442, doi: mcpr.1999.0275
17. Ochman, H., Lawrence, J. G. & Croisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
18. Mendenhall, W. (1979). *Introduction to Probability and Statistics*, 5th edn. North Scituate, MA: Duxbury Press.
19. Honore, B., Madsen, P. & Leffers, H. (1993). The tetramethylammonium chloride method for screening of cDNA libraries using highly degenerate oligonucleotides obtained by backtranslation of amino acid sequences. *Journal of Biochemical and Biophysical Methods* **27**, 39.
20. Grf, E., Tatzel, R. & Wallnoefer, P. R. (1998). Reverse hybridization on a microplate: a rapid identification of dairy microorganisms. *Milchwissenschaft* **53**, 494–8.
21. Anderson, R. C., Su, X., Bogdan, G. J. & Fenton, J. (2000). A miniature integrated device for automated multistep genetic assays. *Nucleic Acids Research* **28**, E60–e60.
22. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. (1996). Real time quantitative PCR. *Genome Research* **6**, 986–94.
23. Cruz-Perez, P., Buttner, M. P. & Stetzenbach, L. D. (2001). Detection and quantitation of *Aspergillus fumigatis* in pure culture using polymerase chain reaction. *Molecular and Cellular Probes* **15**, 81–88, doi: 10.1006/mcpr.2000.0343
24. Nasarabadi, S., Milanovich, F., Richards, J. & Belgrader, P. (1999). Simultaneous detection of TaqMan probes containing fam and tamra reporter fluorophores. *Biotechniques* **27**, 1116–8.