Array-Based Binary Analysis for Bacterial Typing

ARTICLE *in* ANALYTICAL CHEMISTRY · FEBRUARY 2005

Impact Factor: 5.64 · DOI: 10.1021/ac0488006 · Source: PubMed

4 AUTHORS, INCLUDING:

Jason Shepard
University at Albany, The State University of …
28 PUBLICATIONS **913** CITATIONS
SEE PROFILE

Yael Danin-Poleg
Technion - Israel Institute of Technology
41 PUBLICATIONS **921** CITATIONS
SEE PROFILE

Anal. Chem. **2005,** *77,* 319−326

Array-Based Binary Analysis for Bacterial Typing

Jason R. E. Shepard, Yael Danin-Poleg,‡ Yechezkel Kashi,*,‡ and David R. Walt*,†

Department of Chemistry, Tufts University, 62 Talbot Avenue, Medford, Massachusetts 02155, and Department of Biotechnology & Food Engineering, Technion−Israel Institute of Technology, Haifa 32000, Israel

An allele-specific oligonucleotide microarray was developed for rapid typing of pathogens based on analysis of genomic variations. Using a panel of *Escherichia coli* strains as a model system, selected loci were sequenced to uncover differences, such as single- or multiple-nucleotide polymorphisms as well as insertion/deletions (indels). While typical genomic profiling experiments employ specific sequences targeted to genomic DNA unique to a single strain or virulent gene, the present array is designed to type bacteria based on a patterned signature response across multiple loci. In the signature concept, all strains are interrogated by hybridizing their amplified DNA to an array containing multiple probe sequences. Allele-specific oligonucleotide probe sequences targeting each of these variable regions were synthesized and included in a custom fiber-optic array. For each locus, a set of specific probe sequences is selected, such that hybridization gives a binary signal/no signal response to each of the probes. Using this strategy for multiple loci, many pathogens or microorganisms could be classified using a limited number of probes. Because of the advantages of the fiber-optic array platform over other array formats, including sensitivity and speed, the platform described in this paper is capable of supporting a high-throughput diagnostic strategy.

Pathogenic microorganism detection and discrimination have become important objectives recently for clinical, food, and water samples. In addition, current concerns about bioterrorism have increased the need for systems that can rapidly identify pathogens. While detection of pathogenic microorganisms and identification of biowarfare agents are logical applications for microarrays, microarray-based assays can be applied to more intricate problems such as classification of emerging pathogenic strains.[1] The threat of emerging human pathogens has been realized in the past few years; examples include the emergence of SARS (severe acute respiratory syndrome) throughout the world,[2] a new strain of the avian flu in The Netherlands,[3] the reemergence of the West Nile virus in the United States,[4] and the new virulent strain of *Vibrio cholerae* O139.[5] The impact of these threats is even more ominous considering both bacteria and viruses have demonstrated the ability to develop resistance to therapeutic treatments such as antibiotics and antiviral agents.[6−9] For bacterial typing, it is often difficult to discriminate among strains that may vary from normal, commensal microflora to virulent strains. The problem is compounded by the emergence of infectious bacterial agents caused by horizontal gene transfer and elevated mutation frequencies.[8,10] Rapid discrimination and identification of the various pathogenic variants will contribute to our ability to limit epidemics and will have a major role in decreasing the incidence of food-related outbreaks. The ability to rapidly type an organism to the specific strain, particularly one that can quickly evolve to pose a more serious threat, can lead to faster crisis response and rapid containment, thereby limiting infection, aiding in treatment, and preventing potential spread of the organism.

Classical bacterial isolation and identification is performed by selective enrichment followed by plating on selective media. Species identification is mainly accomplished by biochemical characterization, and strain identification is primarily based on serology. These methods are often labor-intensive and require days to complete; thus, they do not meet the requirement for rapid identification and typing. Advances in biotechnology have resulted in the development of several other techniques for typing microorganisms, including multiplexed PCR,[11,12] electrophoretic analyses,[13−15] and microarrays.[16,17] Another method, multilocus sequence typing (MLST), is mainly used to analyze housekeeping genes and has recently been applied as a DNA-based method for bacterial strain typing.[18,19] The MLST approach has also been used for bacterial virulence genes, providing direct information about

* Corresponding authors: (e-mail) kashi@tx.technion.ac.il; david.walt@tufts.edu.
† Tufts University.
‡ Technion−Israel Institute of Technology.

(1) Wang, D.; Coscoy, L.; Zylberberg, M.; Avila, P. C.; Boushey, H. A.; Ganem, D.; DeRisi, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (24), 15687−15692.
(2) Holmes, K. V.; Enjuanes, L. *Science* **2003**, *300*, 1377−1378.
(3) Enserink, M. *Science* **2003**, *300*, 718.
(4) Enserink, M. *Science* **2002**, *297*, 1988−1989.
(5) Colwell, R. R. *Science* **1996**, *274*, 2025−2031.
(6) Boggs, A. F. *Expert Opin. Ther. Pat.* **2003**, *13*, 1107−1112.
(7) Gallant, J. E. *J. Clin. Virol.* **2002**, *25*, 317−333.
(8) Chopra, I.; O'Neill, A. J.; Miller, K. *Drug Resistance Updates* **2003**, *6*, 137−145.
(9) Mansky, L. M.; Bernard, L. C. *J. Virol.* **2000**, *74*, 9532−9539.
(10) Faguy, D. M. *BMC Infect. Dis.* **2003**, *3*, 13.
(11) Ibekwe, A. M.; Watt, P. M.; Grieve, C. M.; Sharma, V. K.; Lyons, S. R. *Appl. Environ. Microbiol.* **2002**, *68*, 4853−4862.
(12) Hu, Y.; Zhang, Q.; Meitzler, J. C. *J. Appl. Microbiol.* **1999**, *87*, 867−876.
(13) Schwartz, D. C.; Cantor, C. R. *Cell* **1984**, *37*, 67−75.
(14) Kourkine, I. V.; Ristic-Petrovic, M.; Davis, E.; Ruffolo, C. G.; Kapsalis, A.; Barron, A. E. *Electrophoresis* **2003**, *24*, 655−661.
(15) Koh, C. G.; Tan, W.; Zhao, M.-q.; Ricco, A. J.; Fan, Z. H. *Anal. Chem.* **2003**, *75*, 4591−4598.
(16) Chizhikov, V.; Rasooly, A.; Chumakov, K.; Levy, D. D. *Appl. Environ. Microbiol.* **2001**, *67*, 3258−3263.
(17) Dorrell, N.; Mangan, J. A.; Laing, K. G.; Hinds, J.; Linton, D.; Al-Ghusein, H.; Barrell, B G.; Stoker, N. G.; Karlyshev, A. V.; Butcher, P. D.; Wren, B. W. *Genome Res.* **2001**, *11*, 1706−1715.
(18) Clarke, S. C. *Bioessays* **2002**, *24*, 858−862.

10.1021/ac0488006 CCC: $30.25
Published on Web 12/02/2004

the potential pathogenicity of an organism.[20,21] The focus on virulence genes can be problematic due to certain strains' high mutability and rearrangements of the genes encoding toxins.[22] To counter this problem, MLST of loci harboring mononucleotide repeats from noncoding regions have been analyzed and were shown to contain much higher sequence variation (including SNPs) than housekeeping genes.[23] These variations proved to be more efficient and as reliable as MLST of housekeeping genes for strain discrimination and for inferring phylogenetic relationships in *Escherichia coli*.

A fiber-optic array platform[24−26] developed with a cross-reactive MLST response methodology, capable of characterizing large numbers of closely related *E. coli* strains is described. *E. coli* is an important organism for demonstrating pathogen detection methods due to the availability of extensive sequence data,[27,28] and its wide pathogenic spectrum from harmless, or probiotic bacterial strains, to virulent strains such as O157:H7. *E. coli* O157:H7 has contributed to the increased occurrence of food-borne illnesses, which affect tens of millions of individuals each year.[29] In this paper, we employ *E. coli* sequence information[23] to prepare a custom microarray for rapid typing of both known and emerging strains, based on a patterned "cross-reactive" response (Figure 1). Arrays were fabricated by employing six different oligonucleotide probe sequences using sequence information from five bacterial intergenic loci. With only six sequences in the array, the signal response patterns generated were capable of rapidly classifying 12 different strains.

EXPERIMENTAL PROTOCOL

Bacterial strain information, DNA extraction, PCR amplification, and sequence analysis were previously described.[23] Briefly, PCR primers for fragment amplification were originally designed from one open reading frame to the adjacent downstream open reading frame. Loci were named after the downstream open reading frame.[23,30] Oligonucleotide probe and primer sequences were selected based on the sequence data from 12 *E. coli* loci (e.g., Table 1). The specific primer and probe sequences, listed in Table 2, were synthesized by Integrated DNA Technologies (Coralville, IA). Oligonucleotide probe coupling to the microspheres (Bang's Laboratories, Fishers, IN) was performed as previously described.[24,25] Array design was performed on polymorphic loci regions with primer and probe sequence information highlighting the strain-to-strain variation, while limiting hybridization deficiencies such as hairpin structure and amplicon-dimer formation. Primer sequences were selected using the Gene Runner (Version 3.05), and probe sequences were selected by hand and scored via Oligotech software (Integrated DNA Technologies downloadable software, www.idtdna.com). Each oligonucleotide probe/microsphere position was determined with its own tracking dye embedded in the beads and confirmed via hybridization to synthetic fluorescently labeled complementary DNA.[24,31] The arrays were fabricated with multiple replicate microspheres of each probe type in the array.

PCR Amplification. A fluorescein label was included in the reverse primers via 5′-fluorescein phosphoramidite during synthesis. Forward and reverse primers, template DNA, and PCR Master Mix solutions (Promega, Madison, WI) were combined and subjected to amplification in a thermal cycler (Hybaid PCRSprint, Franklin, MA).[23] The primers were designed to minimize the amplified fragment length (Table 2) where possible. The cycling conditions for PCR consisted of denaturation at 95 °C for 5 min, followed by 5 cycles (1 min at 95 °C, 1 min at T_m, and 1 min at 72 °C), 20 cycles (1 min at 95 °C, 1 min at T_m 5 °C, and 1 min at 72 °C), and a final step of 7 min at 72 °C. Melting temperatures for the primers at each locus were as follows: *ycgW* 55 °C, *yaiN* 63 °C, *osmB* 55 °C, *galS* 55 °C, and *serW* 60 °C. PCR amplification success was verified by gel electrophoresis and sequencing where necessary (data not shown).

Array Fabrication. The fiber-optic arrays were fabricated as described previously.[24−26] Briefly, optical fiber bundles were polished and etched to form an array of microwells. The fiber bundles were ∼1 mm in diameter and contained tens of thousands of 3-μm-diameter microwells. Single-stranded oligonucleotide probe sequences, terminated with a 5′-amine group, were synthesized and attached to amine-functionalized microspheres via cyanuric chloride. A solution of the different probe-functionalized microspheres was added to the etched fiber face, distributing the microspheres randomly in the fiber-optic bundle. The microsphere positions were determined after array assembly as previously described with our custom-built imaging system.

Array Hybridization. The single-locus PCR products (10 μL) were combined with 7 μL of PBS buffer (0.2 M phosphate-buffered saline, 2 M NaCl, pH 7.4) and 3 μL of formamide, to a total of 20-μL volume, and hybridized directly to the array. The multilocus PCR products (7.5 μL each) were combined with 17.5 μL of PBS buffer and 20 μL of formamide, to a total volume of 75 μL. The samples were heated to 95 °C for 5 min, flash frozen, warmed to room temperature, and immediately hybridized to the array. The single-locus PCR products were hybridized for a minimum of 15 min, and the multiplexed (combined multiple loci) PCR products were hybridized for a minimum of 45 min. Posthybridization, the array was subjected to washing with 0.1 M PBS (0.1 M potassium phosphate, 1 M NaCl, pH 7.4) at 45 °C. Fluorescent images were

(19) Maiden, M. C.; Bygraves, J. A.; Feil, E.; Morelli, G.; Russell, J. E.; Urwin, R.; Zhang, Q.; Zhou, J.; Zurth, K.; Caugant, D. A.; Feavers, I. M.; Achtman, M.; Spratt, B. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3140−3145.
(20) Urwin, R.; Maiden, M. C. *Trends Microbiol.* **2003**, *10*, 479−487.
(21) Packard, E. R.; Parton, R.; Coote, J. G.; Fry, N. K. *J. Med. Microbiol.* **2004**, *53*, 355−365.
(22) Karch, H.; Meyer, T.; Russmann, H.; Heesemann, J. *Infect. Immun.* **1992**, *60*, 3464−3467.
(23) Diamant, E.; Palti, Y.; Gur-Arie, R.; Cohen, H.; Hallerman, E. M.; Kashi, Y. *Appl. Environ. Microbiol.* **2004**, *70*, 2464−2473.
(24) Epstein, J. R.; Lee, M.; Walt, D. R. *Anal. Chem.* **2002**, *74*, 1836−1840.
(25) Ferguson, J. A.; Steemers, F. J.; Walt, D. R. *Anal. Chem.* **2000**, *72*, 5618−5624.
(26) Walt, D. R. *Science* **2000**, *287*, 451−452.
(27) Blattner, F. R.; Plunkett, G., III; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. *Science* **1997**, *277*, 1453−1462.
(28) Perna, N. T.; Plunkett, G., III; Burland, V.; Mau, B.; Glasner, J. D.; Rose, D. J.; Mayhew, G. F.; Evans, P. S.; Gregor, J.; Kirkpatrick, H. A.; Postal, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E. J.; Davis, N. W.; Lim, A.; Dimalanta, E. T. Potamousis, K. D.; Apodaca, J.; Anantharaman, T. S. Lin, J. Yen, G.; Schwartz, D. C.; Welch, R. A.; Blatner, F. R. *Nature* **2001**, *409*, 529−533.
(29) Mead, P. S.; Slutsker, L.; Dietz, V.; McGaig, L. F.; Bresee, J. S.; Shapiro, C.; Griffin, P. M.; Tauxe, R. V. *Emerg. Infect. Dis.* **1999**, *5* (5), 607−625.
(30) Gur-Arie, R.; Cohen, C. J.; Eitan, Y.; Shelef, L.; Hallerman, E. M.; Kashi, Y. *Genome Res.* **2000**, *10*, 62−71.
(31) Michael, K. L.; Taylor, L. C.; Schultz, S. L.; Walt, D. R. *Anal. Chem.* **1998**, *70*, 1242−1248.

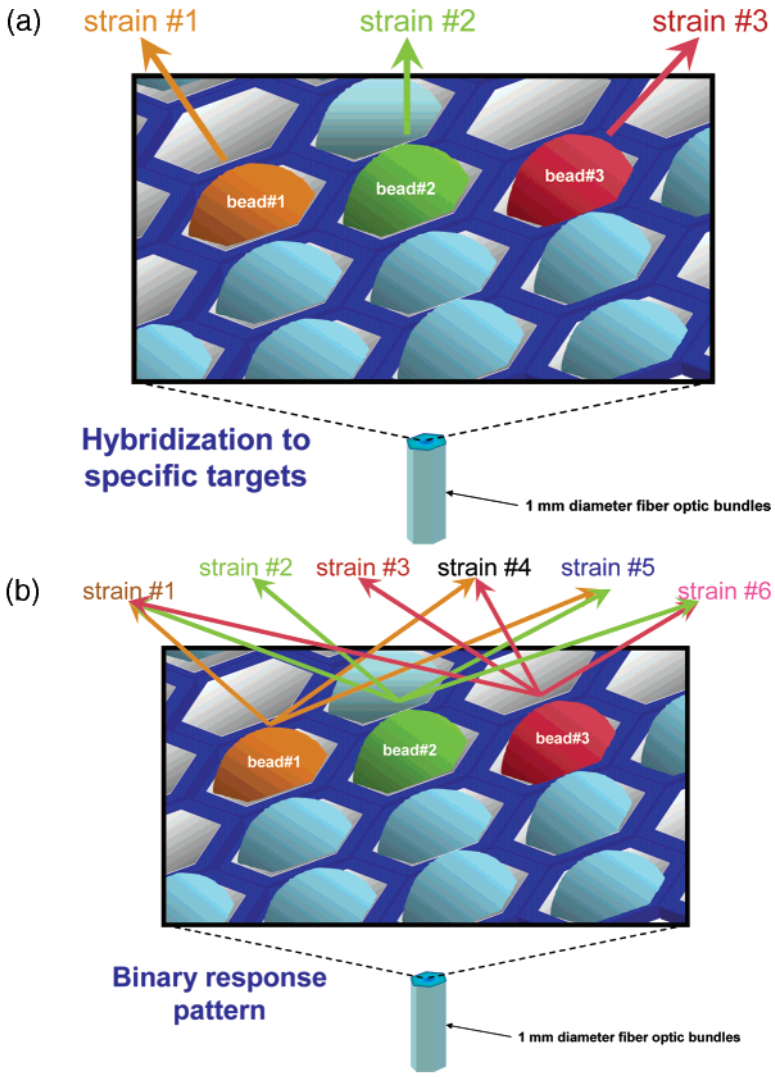

Figure 1. Array-based patterned hybridization concept. The array platform employs single-stranded oligonucleotide sequences immobilized onto microbead array features. The beads are situated on a 1-mm-diameter, hexagonally packed fiber-optic bundle. In the conventional approach (a), the individual beads respond to the presence of a single gene, typically coding for a single toxin, to identify a pathogen. In the approach described in this paper (b), the individual beads respond to the presence of bacterial strain DNA (pathogenic or nonpathogenic) in a binary ("yes" or "no") fashion. The analysis of the binary pattern generated by multiple beads contributes to a response pattern that leads to a more detailed description of the organism.

taken prior to hybridization (background) and after each rinsing stage with 5-s acquisition times. Arrays were regenerated by exposure to 90% formamide solution as previously described, allowing repetitive hybridizations on a single array.[24−26]

Instrumentation. The instrument is composed of a custom-built, modified epifluorescence imaging system[24,25] and includes a white light source, excitation and emission filter wheels, microscope objectives, and a charge-coupled device camera detector (Hamamatsu, Bridgewater, NJ). The core of the imaging system is a standard Olympus microscope (Melville, NY), controlled using IPLab software (Scananalytics, Fairfax, VA).

RESULTS

The array platform described here employs cross-reactive oligonucleotide detection elements based on comprehensive sequence information across multiple strains and multiple loci. The array-based multilocus typing format was designed to respond to multiple strains and classify each one based on a unique patterned response.

PCR Assays Targeting Polymorphisms. MLST of *E. coli* using mononucleotide repeat loci[23] was employed to develop an oligonucleotide array capable of differentiating bacterial strains based on sequence variations. A panel of 12 *E. coli* strains (Table 1) representing various pathogenic and nonpathogenic groups was selected, and their DNA was extracted from single colonies. Sequence variation at five informative loci (*serW*, *ycgW*, *yaiN*, *osmB*, *galS*) were targeted in the present study, with each locus containing numerous polymorphic regions (e.g., Table 1).[23] PCR primers were chosen to amplify the specific polymorphic regions with a size range of 100−250 bp (Table 2). The PCR products were designed to be as short as possible, to enable more efficient discrimination and incorporate a minimum number of polymorphic mismatches. The reverse PCR primers were labeled with fluorescein, producing PCR products with a fluorescently labeled strand complementary to the probe sequences. The probe sequences were selected to interrogate the *E. coli* sequence variation and were designed to hybridize to a single allele out of

Table 1. Portion of the Polymorphic Region of the *ycgW* Locus Illustrating the Sequence Variation among 12 *E. coli* Strains[a]

Strain #	Serotype	*E. coli* alleles at the *ycgW* locus
1	EHEC - O157:H⁻	A C T C A T A T G C A A A A T C A A G A A A T A A
2	K12 - DH5α	A C T C A T A T G C A A A A T C A A G A A A T A A
3	ETEC- O78:H⁻	A C T C A T A T G C A A A A T C A A G A A A T A A
4	ECOR - 08	A C T C A T A T G C A A A A T C A A G A A A T A A
5	B-SR9b	A C T C A T A T G C A A A A T A A A G A A A T G A
6	ECOR - 68	no PCR product
7	EHEC - O113:H2	A C T C A T A T G C A A A A T A A A G A A A T G A
8	ECOR - 27	A C T C A T A T G C A A A A T A A A G A A A T G A
9	EHEC - O26:H⁻	A C T C A T A T G C A A A A T A A A G A A A T G A
10	ECOR - 44	A C A C A T A T G C A A A G T C A A G A A A T A A
11	O157:H7 HER phage 1058	no PCR product
12	EPEC - O55:H7	A C T C A T A T G C A A A A T A A A G A A A T G A
	Cons.	A C * C A T A T G C A A A * T * A A G A A A T * A
probe sequence		A C T C A T A T G C A A A A T C A A G A A A T A A

[a] The polymorphic positions are red. A portion of the *ycgW#1* target sequence is at the bottom of the table. Alleles are color-coded. The stringency is set to allow hybridization only to the perfect match (blue), such that all sequences noncomplementary to the *ycgW* probe 1 do not hybridize. Note: strains 6 and 11 have a null allele, where no product was amplified via PCR. A detailed description of the *E. coli* strains used in the study is in Diamant et al.[23]

Table 2. Primer and Probe Sets Used for the MLST Binary Microarray Approach[a]

locus	forward primer	reverse primer (fluorescein labeled)	product size (bp)
serW	5′-TTC-ACA-ggT-AAC-ATA-CTC−CAC-3′	5′-CCC−CTC-ACC-gCC-ATA-TTT-AA-3′	116
ycgW	5′-TTg-TTA-TgT-CTT-ATC−CCA-Cgg-3′	5′-CAT-CCA-TTg-AgA-TTC−CTT-gCT-3′	156
osmB	5′-ggT-gAT-AAT-gAC-TTC−CTg-T-3′	5′-CAA-CCA-ggA-ATC-ATC-TTA-*g*-3′[b]	101
galS	5′-gCg-CTA-CAT-CAC-gAA-Tgg-Tg-3′[b]	5′-CgA-TTC−Acg-Aag-TCC-TgT-ATT-C-3′	110
yaiN	5′-AAT-TTA-TCC-ggT-gAA-TgT-ggT-3′[b]	5′-ggA-CgC−CAg-AAA-CAC-gCT-AC-3′	250

locus	probe sequence	probe length (bp)
serW	5′-CAT-gTT-CAC-TgC−CgT-ACA-gAC−AgA-TAA-AAT-gCg-AAA-AAA-AAg−CTC-3′	45
ycgW #1	5′-ggT-TCA-TTA-ggA-TgT-TTA-TTT-CTT-gAT-TTT-gCA-TAT-gAg-TAT-ATT-AC-3′	46
ycgW #2	5′-ggT-TCA-TTA-ggA-TgT-TCA-TTT-CTT-TAT-TTT-gCA-TAT-gAg-TAT-ATT-AC-3′	46
osmB	5′-TAA-TTT-TAT-ATC-TTg-AgA-gTg-TTA-ATA-ACA-ggT-AAA-TAg-TCT-3′	42
galS	5′-gTA-AAT-gAC-TgC-TTg-CTg-CCg-gCT-AAT-TTg-TCA-3′	33
yaiN	5′-TTT-CCA-ggT-CAg-ATg-ggC-TgC-AAg-TTg-CAg-ACC-gTT-ATA-ATC-AT-3′	44

[a] The strain and locus sequence information was previously published in Diamant et al.[23] [b]Sequences of these primers were published previously.

multiple alleles present at each polymorphic locus (for example, see Table 1). The probe sequences were long enough (33−46-mers) to hybridize to the PCR products directly. Shorter probe sequences (12−24-mers), encompassing only a single polymorphism, were tested but resulted in low signal and low discrimination (data not shown). The successful probe sequences included regions with multiple point mutations to allow discrimination between completely homologous and nonhomologous hybridization. In this manner, each locus produced a binary response; if a particular strain had the complementary target sequence to the specific allele probe sequence, a positive fluorescent signal was measured, whereas if the target sequence was not fully complementary, no signal was present. This binary response at each probe enabled us to interrogate specific alleles at multiple loci, resulting in a unique "bar code" for each of the 12 strains tested.

Array Hybridization. Because the array is reusable, the same array could test each of the PCR products either individually or in a multiplexed hybridization across the five loci from each strain. The combined PCR products from a single strain, when hybridized to the array, produced a patterned response that allowed each strain to be uniquely identified. A total of 3 arrays that included

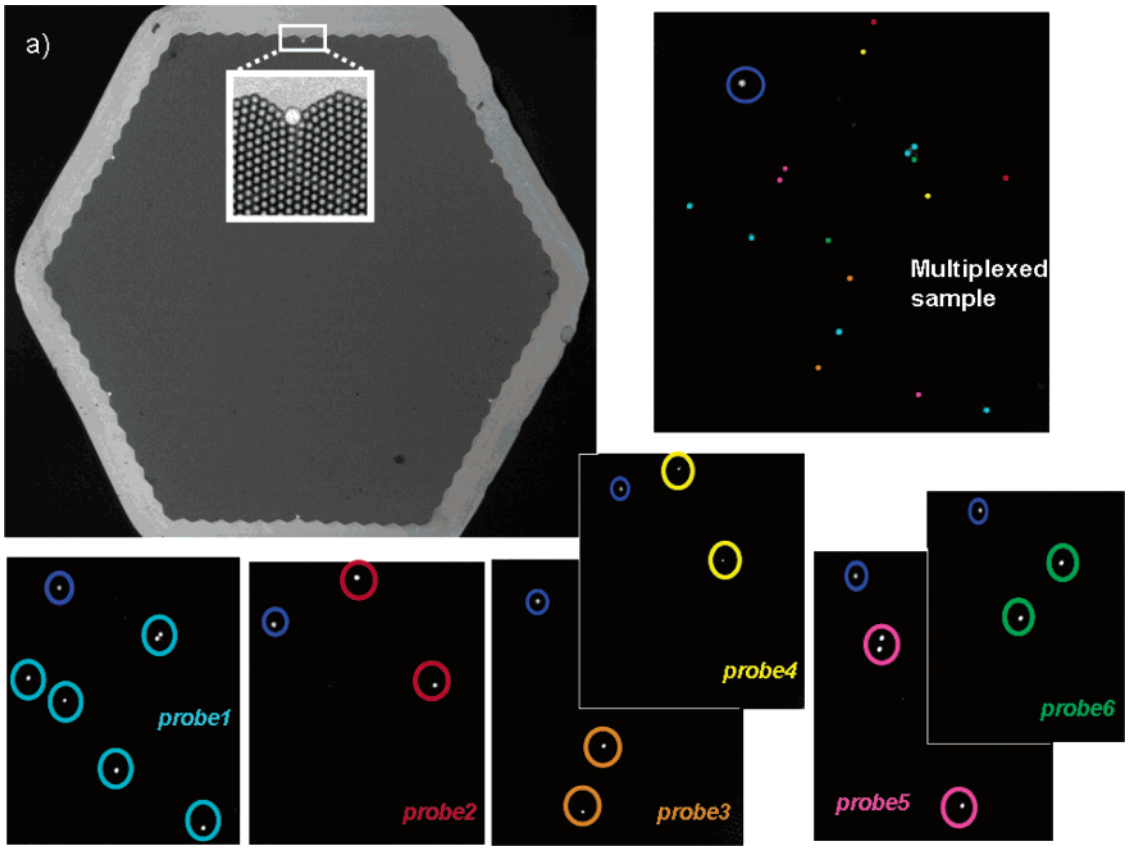

Figure 2. (a) Image of the hexagonally packed fiber-optic bundle illustrating the individual positions in the array. The 1-mm-diameter fiber bundle is composed of ~50 000 individual 3-μm optical fibers, each capable of housing a single oligonucleotide-functionalized bead. The remaining images are from a portion of a fiber bundle, illustrating the responses from each individual bead type and a multiplexed response from all of the beads simultaneously. Each bead type is marked with a different color in the array. The navy blue circled mark is a positional marker and is included in the array for correct alignment.

identical probe sequences were used to type each of the 12 bacterial strains. Each strain typing was done as a multiplexed hybridization, i.e., a single multiplexed hybridization solution containing the PCR products from five separate PCR reactions, for each of the five loci. To confirm the multiplexed hybridization responses, the individual locus response of each strain was conducted with a separate hybridization. One such array is presented in Figure 2, where the responses showing the positions of each of the six different probe sequences are shown. The results from the patterned response profiles from the different strains are shown in Table 3. The results illustrate the patterns for all the strains, color-coded with blue for a positive response and yellow for a negative response. In each of the different strains, the response pattern to the array is unique. As can be seen, the approach reduces to a simple threshold yes/no determination of hybridization to each probe, with the *minimum* average positive signal greater than 5 times the signal differences attributed to negative responses (Table 3), with the exception of *ycgW#2* probe where lower differences between positive and negative signals were observed. A typical positive detection signal is based on signal intensities greater than three times the standard deviation of the background signal (3σ). The detection thresholds reported here are determined from hybridizations to known standards. It is possible that threshold values may vary between experiments, but with this unknown determination, a 5-fold threshold value is actually more stringent than 3σ and defines a higher threshold of difference between a positive and a negative signal. This platform has the ability to perform rapid, sensitive assays, with hybridizations taking less than 1 h. Between hybridizations, the array was regenerated within 5 s, via exposure to organic denaturants, allowing subsequent hybridizations to the same array with additional samples. While the overall assay time required for this method is still dependent on the PCR step, many common array platforms use much more lengthy, or even overnight, hybridizations causing the hybridization to be the limiting factor in high-throughput analysis.[32]

Discrimination Ability of the Patterned MLST Method. Successful discrimination was achieved, differentiating among all 12 tested isolates based on the identified sequence variation. In addition, targeting specific point mutations facilitated the discrimination of closely related strains, as was expected from their sequence differences. For example, two different isolates of *E. coli* O157 serogroup were differentiated (strains 1 (O157:H$^-$) and 11 (O157:H7). The closely related strain O55:H7 (strain 12) was also separately identified as well. Separate classification of these three strains demonstrates the ability to subcategorize pathogenic serotypes in addition to their identification as pathogen strains. From sequence information encompassing dozens of *E. coli* strains,[23] it is clear that if a number of other O157:H7 isolates were hybridized to the array, they would similarly group together

(32) Yang, L.; Tran, D. K.; Wang, X. *Genome Res.* **2001**, *11*, 1888−1898.

Table 3. Hybridization Patterns of the 12 *E. coli* Strains to the Array[a]

#	*E. coli* strain	*ycgW#1*	*serW*	*yaiN*	*galS*	*osmB*	*ycgW#2*
1	EHEC - O157:H$^-$	234	141	34.3	7.9	95.6	26.7
2	K12 - DH5α	113	119	5.2	6.7	249	27.0
3	ETEC - O78:H$^-$	350	10.3	34.4	4.0	128.5	27.6
4	ECOR - 08	101	7.5	9.8	1.5	97.3	28.3
5	B-SR9b	4.8	72	30.3	7.9	64	53.3
6	ECOR - 68	5.3	63	40.9	9.6	57	13.0
7	EHEC - O113:H2	4.6	16.7	33.9	23.0	105	59.5
8	ECOR - 27	19.9	15.5	31.6	5.5	70	49.4
9	EHEC - O26:H$^-$	19.4	9.3	31.7	1.7	11.0	58.0
10	ECOR - 44	15.0	15.2	35.9	2.4	120	8.9
11	O157:H7 HER phage 1058	8.0	13.6	3.6	29.1	69	26.6
12	EPEC - O55:H7	19.2	18.7	8.5	27.1	102	44.1
	Ave. positive	200	98.8	34.2	26.4	105	52.9
	Ave. negative	12.0	13.3	6.8	5.2	11.0	22.6

[a] A positive response is shown in blue, and a negative response is shown in yellow. For each strain, the combination of responses results in a unique pattern. The pattern can be thought of as a "bar code" response to the specific signature of each strain hybridized to the array. Actual array response values are shown in the table.

Table 4. Determination of an Unknown Isolate as a Wild-Type Strain via Comparison to a Series of Known Hybridization Responses (1, 7, 9, and 11)[a]

#	*E. coli strain*	*ycgW#1*	*serW*	*yaiN*	*galS*	*osmB*	*ycgW#2*
1	O157:H$^-$	**29.1**	**123.7**	**77.2**	**9.6**	**11.8**	**1.1**
7	O113:H2	**15.7**	**10.9**	**63.8**	**35.3**	**12.8**	**28.6**
9	O26:H$^-$	**14.3**	**10**	**213.7**	**14.8**	**0.2**	**38.1**
11	O157:H7	**4.7**	**10.7**	**10.5**	**35.6**	**15.9**	**7.6**
A	unknown	**5.9**	**13.3**	**83.1**	**15.7**	**34.5**	**5.9**
A	unknown	**5.9**	**13.3**	**83.1**	**15.7**	**34.5**	**5.9**
10	ECOR-44 (from Table 3)						

[a] Green is considered a positive response and pink is a negative response.

into one of the above two patterns, resulting in pathogen subcategories.

The patterned response method was also used to classify an "unknown" strain from our panel (Table 4). The unknown sample (A in Table 4) was subjected to PCR in each of the five loci separately, and then PCR products were pooled and hybridized to the array. Four of the previously identified strains were run as controls (1, 7, 9, and 11; Table 1). The four controls were employed to provide known positive (blue) and negative (yellow) responses for each probe sequence, essentially bracketing the unknown responses within the known responses, and providing positive and negative "reference" responses for each probe. While it is not essential to perform such reference hybridizations each time, confidence in making the binary calls is improved as each PCR

reaction and the subsequent hybridizations to the array vary between the different loci. From these standards, the pattern from the unknown strain was determined, providing objective classification of the unknown. The unknown (A) strain gave a response pattern identical to a wild-type nonpathogenic strain ECOR-44 (10; Table 1) included in the original 12 classified strains and was determined to be nonpathogenic.

DISCUSSION

While many assays have targeted and detected the presence of genes coding for specific toxins,[33−35] no efficient, rapid method exists that can fully differentiate large numbers of strains to the strain or serotypic level with a single platform. Array platforms are more capable of carrying out the level of throughput necessary for this type of analysis, as microarray analysis is optimally intended for large-scale, simultaneous analysis of complex mixtures.[36] Alternative assays that target specific genes or toxins can pose problems for pathogen classification. Emerging pathogenic microorganisms that omit these target genes could preclude detection, or strain differences resulting from mutations or gene transfer could provide false positives or false negatives.[37,38] Pathogenic *E. coli*, under certain conditions, have been shown to lose genes coding for toxins and would provide a negative result to many of these specific tests.[22] While these problems would also affect the system described here, the strain would be identified as genetically related to a known pathogen. The diversity of strains within a species coupled to the plethora of different microorganisms that exist would overrun the capabilities of most detection methods except that of a microarray. Classifying multiple strains with this MLST array format avoids the need for extensive sequence analysis of the tested bacterial strain.

Array Design. The MLST fiber-optic array was designed to highlight the variations across different closely related microbial strains of a single species. The sequence design for this array exploits the intrinsic polymorphisms existing among *E. coli* strains, including insertions, deletions, and single-base differences using a binary readout (positive or negative for each probe). Employing a binary response simplifies the analysis of a locus that encompasses multiple alleles and provides an efficient way to perform strain analysis. The amount of information that can be garnered from this type of assay is high, as a binary response from n different probe sequences provides 2^n total possible classifications, where n equals the number of probes interrogated. In the present case, the array should be able to identify 2^6 or 64 possible classifications. In principle, an array containing 20 probe sequences would enable more than 1.0×10^6 possible variation classifications (2^{20}), underscoring the power of the technique.

The concept of binary discrimination is illustrated in Table 1, which shows partial sequence information of the *ycgW* locus for each of the 12 strains tested. As seen in the table, there are three distinct hybridization possibilities of the 12 PCR products to the *ycgW*#1 probe sequence. The first possibility is that the PCR products will be perfectly matched to the probe sequence, providing a positive signal. In the second case, the PCR products have polymorphic differences such that hybridization to the probe sequence would be precluded under the stringency conditions employed. The probe sequences were designed to keep the base mismatches in the middle of the probe where possible, as internal sequence mismatches are easier to differentiate compared to terminal mismatches.[39] The third instance is when no PCR product is formed; thus, no hybridization response is possible, indicating a null allele. Both the second and third possibilities would be expected to give no hybridization signal. In this manner, only the strains with a perfect match (highlighted in blue in Table 1) provided a positive response to the specific probe sequence.

Array-based analyses provide a secondary level of specificity above that of simple multiplexed PCR gel-based analysis.[1,36] The possibility exists that the primers would amplify another portion of the genome or a different organism altogether. Hybridization to the array with nonspecific amplification products would not result in a positive hybridization signal unless the product of the PCR reaction happened to be complementary to the probe sequence.[16] Hybridization provides an additional level of specificity over assays in which nonspecific amplified material could still appear, such as in a gel, possibly leading to a false positive result. While such as array could easily be adapted for a "presence or absence" assay that is solely used for determining pathogenicity, the method reported here is designed for rapid, specialized subclassification of a single bacterial strain.

Inherent Redundancy. Many of the strains yielded response patterns that could be identified with a few probe sequences. The remaining responses to other probes in the array provide redundant information that allowed strain identification to be confirmed. Redundant information is important for classification of emerging strains, but there remains the possibility that an emerging strain would incorporate a polymorphism into one of the probe sequences, thereby changing its patterned response. In this case, the polymorphic strain would either be determined by redundant sequences in the array or follow a "partial patterned response". Even a partial response would enable the array to identify the organism as a closely related strain. With specific or targeted arrays, discrimination may not be possible and the organism may elude identification.

MLST Hybridization. The gene expression analysis that is normally performed with oligonucleotide microarrays is aimed at observing hybridization response across the entire genome. Rather than include thousands of sequences on an array to identify a single bacterial strain, the methodology presented here takes a more efficient approach to genomic typing, employing a minimal number of array elements to differentiate multiple strains. With this binary MLST discrimination methodology, members of a pathogenic family would be expected to have closer sequence relationships, and similar response patterns, than nonpathogenic strains.[23] This same methodology could be developed for a number of different pathogenic organisms, such as salmonella or campylobacter.

(33) Call, D. R.; Brockman, F. J.; Chandler, D. P. *Int. J. Food Microbiol.* **2001**, *67*, 71−80.
(34) Louie, M.; Read, S.; Simor, A. E.; Holland, J.; Louie, L.; Ziebell, K.; Brunton, J.; Hii, J. *J. Clin. Microbiol.* **1998**, *36*, 3375−3377.
(35) Paton, A. W.; Paton, J. C. *J. Clin. Microbiol.* **1999**, *37*, 3362−3365.
(36) Vora, G. J.; Meador, C. E.; Stenger, D. A.; Andreadis, J. D. *Appl. Environ. Microbiol.* **2004**, *70* (5), 3047−3054.
(37) LeClerc, J. E.; Li, B.; Payne, W. L.; Cebula, T. A. *Science* **1996**, *274*, 1208−1211.
(38) Matic, I.; Radman, M.; Taddei, F.; Picard, B.; Doit, C.; Bingen, E.; Denamur, E.; Elion, J.; LeClerc, J. E.; Cebula, T. A. *Science* **1997**, *277*, 1833−1834.
(39) Letowski, J.; Brousseau, R.; Masson, L. *J. Microbiol. Methods* **2004**, *57* (2), 269−278.

This MLST hybridization methodology could also be applied to other related DNA-based analysis problems. An important area where this method could be useful is for cataloging specific cancer lines. Different cancer types, such as lung or breast cancer, have been subclassified with analysis of differential transcription responses, indicative of unique biological characteristics.[40−42] By providing a rapid method that can subcategorize a single cancer type based on unique polymorphisms, treatment could be tailored, thereby increasing therapeutic success.

The *E. coli* strains selected for this assay were selected to show the concept of the typing technology using a wide spectrum of strains. The PCR products from isolated *E. coli* DNA were hybridized to an array, and the response patterns were then used to classify each strain. A binary multilocus sequence typing format was devised to take advantage of polymorphisms in five different loci. Each strain's response to the 6 probe sequences in the array was used to classify 12 bacterial strains. The array successfully discriminated pathogenic from nonpathogenic strains and was able to classify different O157 serotypes. This methodology could be applied to an entire family of pathogens or for any other problem where minimal DNA-based differences exist between members.

(40) Reddy, A.; Kaelin, W. G., Jr. *Curr. Opin. Pharmacol.* **2002**, *4*, 366−373.
(41) Lotze, M. T. *Curr. Opin. Invest. Drugs* **2003**, *4* (6), 649−651.
(42) van de Vijver, M. J.; He, Y. D.; van't Veer, L. J.; Dai, H.; Hart, A. A.; Voskuil, D. W.; Scheriber, G. J.; Peterse, J. L.; Roberts, C.; Marton, M. J.; Parrish, M.; Atsma, D.; Witteveen, A.; Glas, A.; Delahaye, L.; van der Velde, T.; Bartelink, H.; Rodenhuis, S.; Rutgers, E. T.; Friend, S. H.; Bernards, R. *N. Engl. J. Med.* **2002**, *347* (25).

ACKNOWLEDGMENT

The authors thank Dr. Israel Biran for his helpful discussions with this research and for reviewing the manuscript. This research was supported by the Grand Water Research Institute, Technicon, and the Israeli Water Commission.

SUPPORTING INFORMATION AVAILABLE

Primer and probe sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review August 12, 2004. Accepted October 13, 2004.

AC0488006