

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 November 2008 (20.11.2008)

PCT

(10) International Publication Number
WO 2008/140494 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(74) Agents: **CARROLL, Peter, G.** et al.; Medlen & Carroll, LLP, 101 Howard Street, Suite 350, San Francisco, CA 94105 (US).

(21) International Application Number:
PCT/US2007/024291

(22) International Filing Date:
21 November 2007 (21.11.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/860,786 22 November 2006 (22.11.2006) US

(71) Applicant (for all designated States except US): **THE BOARD OF TRUSTEES OF MICHIGAN STATE UNIVERSITY** [US/US]; 238 Administration Building, East Lansing, MI 48424-1046 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HASHSHAM, Syed, Anwar** [IN/US]; 3950 Jonquil Lane, Okemos, MI 48864 (US). **TEIDJE, James, M.** [US/US]; 1646 Lindbergh Drive, Lansing, MI 48910 (US). **GULARI, Erdogan** [US/US]; 386 Pemeberton Drive, Ann Arbor, MI 48105 (US). **TOURLOUSSE, Dieter** [BE/US]; 3839 Hunsaker Drive, Apt 923D, East Lansing, MI 48823 (US). **STEDTFELD, Robert** [US/US]; 4642 Palmer Street, Lansing, MI 48910 (US). **AHMAD, Farhan** [IN/US]; 1525 Spartan Village, Apt. 1, Michigan State University, East Lansing, MI 48823 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(54) Title: HIGH THROUGHPUT SCREENING USING MICROARRAYS

(57) Abstract: The present invention relates to the design and manufacture of microarrays for use in detecting multiple organisms in a sample in a high throughput manner. In particular, the present invention relates to the detection of waterborne pathogens in a sample of water.



WO 2008/140494 A2

HIGH THROUGHPUT SCREENING USING MICROARRAYS

This invention was made with government support from the National Institutes of Health; grant numbers 1R01RR018625-01, 5R01RR018625-02, 5R01RR018625-03, and
5 5R01RR018625-03. The United States Government has certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates to the design and manufacture of microarrays for use in detecting multiple organisms in a sample in a high throughput manner. In particular,
10 the present invention relates to the detection of waterborne pathogens in a sample of water.

BACKGROUND OF THE INVENTION

Waterborne pathogens are estimated to cause 2.2 and 5 million cases deaths per
15 year. Waterborne illnesses continue to be a problem even in developed nations where outbreaks of unknown etiology are routine and chlorine resistant pathogens are emerging as a threat to public health. Mackenzie, W.R., et al., A Massive Outbreak in Milwaukee of *Cryptosporidium* Infection Transmitted through the Public Water-Supply. New England Journal of Medicine, 1994. 331(3): p. 161-167, herein incorporated by reference.

20 Historically, people have been exposed to waterborne pathogens through accidental or inadvertent contamination of their water supply. Additionally, measures are needed for monitoring and guarding against deliberate contamination of the water supply. With this emergence of a double threat the need for continuous monitoring of the water supply has become a critical issue for public health and national security.

25 Studies on the Affymetrix platform have shown the feasibility of pathogen detection using high density microarrays. Wilson et al. (High-density microarray of small-subunit ribosomal DNA probes. Applied and Environmental Microbiology, 2002. 68(5): p. 2535-2541, herein incorporated by reference) describe development of a microarray targeting the small-subunit ribosomal DNA of 18 pathogenic bacteria (not all
30 are human pathogens). In another paper, Wilson et al. (Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Molecular and Cellular

Probes, 2002. 16(2): p. 119-127, herein incorporated by reference) developed a multi-pathogen identification microarray to detect 18 organisms targeting various genes. Both microarrays were validated with DNA obtained from air filtrate.

A common problem with such assays is the development of multiplexed assays for multiple microorganisms. Current strategies for probe design for multiplex detection of organisms often limit the number of available probes and, in turn, the robustness and accuracy of the assays. What is needed in the art are better methods of identifying probes for use in multiplex assays so that more robust and accurate assays can be designed.

10 SUMMARY OF THE INVENTION

The present invention relates to the design and manufacture of microarrays for use in detecting multiple organisms in a sample in a high throughput manner. In particular, the present invention relates to the detection of waterborne pathogens in a sample of water.

The present invention provides methods for identifying regions of DNA that will specifically identify groups of pathogenic organisms. In some embodiments, the present invention provides methods for identifying regions of DNA that will specifically identify a single pathogenic organism. In some embodiments, the present invention provides methods for providing DNA probes for identifying regions of pathogen DNA.

In further embodiments, the present invention provides methods for providing DNA probes for identifying subregions of pathogen DNA. In some embodiments, the present invention provides methods for providing nonoverlapping DNA probes.

For the purposes of the present inventions, the "maximum number of nonoverlapping probes" depends on the size of the gene sequence and the predetermined length of the probes. For example, where the gene sequence is 1000 bases and the predetermined length of the probes is 100 bases, the maximum number of nonoverlapping probes is 10. Where the gene sequence is 1000 bases and the predetermined length of the probes is 50 bases, the maximum number of nonoverlapping probes is 20 (and so on). The probes can be "generated" electronically or chemically (synthetically) or even enzymatically. A "subregion" is smaller in size than the complete gene sequence. Typical subregions are in the size range of 50-500 nucleotides. The

predetermined number of data base microorganisms depends on the degree of specificity desired. In one embodiment, the total database comprises 100 organisms and the predetermined number is 5. In one embodiment, the total database comprises 100 organisms and the predetermined number is 4. In one embodiment, the total database comprises 100 organisms and the predetermined number is 3. In one embodiment, the total database comprises 100 organisms and the predetermined number is 2. In one embodiment, the total database comprises 100 organisms and the predetermined number is 1.

In some embodiments, the present invention provides methods of providing target probes for detecting a target pathogen comprising, a) generating a maximum number of nonoverlapping probes of a predetermined length from a gene sequence derived from a target microorganism, b) comparing the sequence of each nonoverlapping probe to sequences from a plurality of database microorganisms, c) identifying at least one subregion in the gene sequence derived from a target microorganism, wherein the at least one subregion contains exact matches to less than a predetermined number of the database microorganisms; and d) generating a set of overlapping probes of a predetermined length for the at least one subregion. The present invention is not limited to any particular length of nonoverlapping probes. In some embodiments, the predetermined length of the nonoverlapping probes is from 10 to 30 bases. In some embodiments, the predetermined length of the nonoverlapping probes is from 30 to 60 bases. The present invention is not limited to subregions containing exact matches to any particular number of database microorganisms. In some embodiments, the subregion contains exact matches to less than from 5 up to 100 database microorganisms. The present invention is not limited to subregions of any particular length. In some embodiments, the subregion is from about 50 to about 500 bases in length. In some embodiments, multiple subregions containing exact matches to less than a predetermined number of database sequences are identified. In some preferred embodiments, overlapping probes of a predetermined length are generated for the multiple subregions. In further preferred embodiments, the set of overlapping probes of a predetermined length are from 15 to 30 bases in length. In other preferred embodiments, the set of overlapping probes of a predetermined length are from 40 to 60 bases in length.

In some embodiments, the methods of the present invention further comprise the step of selecting at least one probe for the at least one subregion. In some embodiments, the at least one probe for the at least one subregion has at least one base pair difference as compared to the sequences of the database microorganisms. In some embodiments, the methods of the present invention further comprise the step of arraying the at least one probe on a solid surface.

In some embodiments, the present invention provides a set of nucleic acid probes generated by the foregoing methods. In other embodiments, the present invention provides microarrays comprising the foregoing set of nucleic acid probes.

In further embodiments, the present invention provides a microarray comprising nucleic acids probes for a target organism, wherein the probes are selected by generating a maximum number of nonoverlapping probes of a predetermined length from a gene sequence derived from the target organism, comparing the sequence of each nonoverlapping probe to sequences from a plurality of database microorganisms, identifying at least one subregion in the gene sequence derived from a target microorganism, wherein the at least one subregion contains exact matches to less than a predetermined number of the database microorganisms; and generating a set of overlapping probes of a predetermined length for the at least one subregion. The present invention is not limited to microarrays formed on any particular substrate. In some embodiments, the probes are arrayed on a glass slide.

In some embodiments, the present invention provides methods for high throughput detection of multiple target organisms comprising a) providing: i) a microarray comprising sets of multiple discrete nucleic acid probes for at least one target organism, wherein the nucleic acid probes are complementary to multiple amplicons from multiple target genes of the at least one target organism; ii) primers complementary to multiple target genes of the multiple target organisms; b) using the primers to amplify the multiple target genes from the multiple target organisms to produce amplicons; c) contacting the microarray with the amplicons under conditions such that the amplicons hybridize to the discrete nucleic acid probes; d) detecting the presence of amplicon binding to the discrete nucleic acid probes. In some embodiments, the methods provide a set of nucleic acid primers. In some embodiments, the methods provide a set of nucleic

acid probes. In some embodiments, the methods provide sequences selected from the group consisting of SEQ ID Nos. 1-19206. In some embodiments, the methods provide sequences selected from the group consisting of SEQ ID Nos. 1-791. In some embodiments, the methods provide sequences selected from the group consisting of SEQ ID Nos. 792-11533. In some embodiments, the methods provide sequences selected from the group consisting of SEQ ID Nos. 11534-13225. In some embodiments, the methods provide sequences selected from the group consisting of SEQ ID Nos. 13226-19206

In some embodiments, the present invention provides a microarray comprising the set of nucleic acid probes generated by methods of the present inventions. In some embodiments, the microarray comprises at least one sequence selected from the group consisting of SEQ ID Nos. 1-19206

In some embodiments, the present invention provides a microarray comprising nucleic acids probes for a target organism, wherein said probes are selected by generating a maximum number of nonoverlapping oligonucleotide probes of a predetermined length from a gene sequence derived from said target organism, comparing the sequence of each nonoverlapping oligonucleotide probes to sequences from a plurality of database microorganisms, identifying at least one subregion in said gene sequence derived from a target microorganism, wherein said at least one subregion contains exact matches to less than a predetermined number of said database microorganisms; and generating a set of overlapping probes of a predetermined length for said at least one subregion. In some embodiments, the oligonucleotide probes are arrayed on a substrate selected from the group consisting of a glass slide and a serpentine chip. In some embodiments, the target organisms are selected from the group consisting of *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica*, *E. faecalis*, *S. enterica*, *C. parvum*, *G. intestinalis*, *K. pneumoniae*, *E. coli*. and combinations thereof. In some embodiments, the oligonucleotide probes are complementary to virulence marker genes from said target organisms. In some embodiments, the present invention provides a microarray for high throughput detection

of waterborne pathogens comprising sets of multiple discreet nucleic acid probes for at least ten target organisms, wherein said nucleic acid probes are complementary to multiple amplicons from multiple virulence marker genes from each target organism and multiple discreet nucleic acid probes are provided for each amplicon, and wherein said at least 10 target microorganism are selected from the group consisting of *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica*, *E. faecalis*, *S. enterica*, *C. parvum*, *G. intestinalis*, *K. pneumoniae*, and *E. coli*. In some embodiments, the nucleic acid probes are selected from the group consisting of oligonucleotide probes encoded by SEQ ID Nos. 1-19206 and combinations thereof.

In some embodiments, the present invention provides a method for high throughput detection of multiple target organisms a) providing: i) a microarray comprising sets of multiple discreet nucleic acid probes for multiple target organisms, wherein said nucleic acid probes are complementary to multiple amplicons from multiple target genes of said at least one target organism, and wherein multiple discreet nucleic acid probes are provided for each amplicon; ii) primers complementary to multiple target genes of said multiple target organisms; and b) using said primers to amplify said multiple target genes from said multiple target organisms to produce amplicons; c) contacting said microarray with said amplicons under conditions such that said amplicons hybridize to said discreet nucleic acid probes; and d) detecting the presence of amplicon binding to said discreet nucleic acid probes. In some embodiments, the target organisms are selected from the group consisting of *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica*, *E. faecalis*, *S. enterica*, *C. parvum*, *G. intestinalis*, *K. pneumoniae*, *E. coli*. and combinations thereof. In some embodiments, the probes are complementary to virulence marker genes from said target organisms. In some embodiments, the nucleic acid probes are selected from the group consisting of probes encoded by SEQ ID Nos. 1-19206 and combinations thereof.

The present invention provides a set of oligonucleotide probes consisting of at least 2 sequences selected from the group consisting of SEQ ID Nos. 1-19206.

The present invention provides a microarray consisting of at least 2 sequences selected from the group consisting of SEQ ID Nos. 1-19206. In some embodiments, the
5 microarray is a serpentine chip.

DESCRIPTION OF THE FIGURES

FIG. 1 shows an exemplary result of testing microbial sequences obtained by
10 compositions and methods of the present inventions for a 16s rDNA sequence. Panel A: Graph of 73 non-overlapping 20-mers. Each 20-mer sequence was then used to obtain the number of matches (exact) in The Ribosomal Database Project II (RDP-II) database and plotted on the y-axis. Panel B: For the region from 135 to 220 yielding the first and second regions with 10 matches or less (in the left panel), the 20-mer probes were tested
15 for matches in RDP-II by walking one base at a time.

FIG. 2 shows an exemplary evaluation of 791 targeted and 2,034 not-targeted probes with a composite target sample containing 47 VMG amplicons (24 labeled with Cy3 and 23 labeled with Cy5). (A) Part of the Xeotron chip after hybridization and washing up to 30
20 °C, spot features are 50 µm in diameter. (B) Distribution and boxplot of SNRs for 791 targeted (black bars and lower boxplot) and 2034 not-targeted probes (grey bars and upper boxplot). (C) Distribution of the positive fraction for 47 targeted (black bars) and 67 not-targeted probe sets (grey bars). The dashed line shows the cut-off (0.5) for presence/absence calls. (D) Success rate of probe design. The number of probes showing
25 positive signal (y-axis) vs. the number of designed probes (x-axis) for each VMG (black symbols: targeted VMGs, and grey symbols: not-targeted VMGs). The dashed lines represent varying levels of success. The slope of these lines noted is equal to the positive fraction.

30 FIG. 3 shows an exemplary targeted probe response and characterization. VMGs are sorted from bottom to top according to increasing median SNRs. (a) positive fraction,

Boxplots represent the variability in SNRs obtained for each VMG, (b) Positive fractions, and (c) G+C content of the amplified region with each symbol representing a VMG for each of 47 targeted VMG amplicons. Probe sets are sorted from bottom to top according to increasing median signal-to-noise ratio.

5

FIG. 4 shows an exemplary success rate of probe design as function of free energy of probe-target duplex formation ($\Delta G^\circ_{\text{duplex}}$). Each symbol represents 40 probes with error bars reflecting the standard deviation among three replicates. The percent positive probes is plotted at the median $\Delta G^\circ_{\text{duplex}}$ of each bin. The G+C content on the upper X-axis was derived from a linear relationship between a $\Delta G^\circ_{\text{duplex}}$ and G+C content.

10

FIG. 5 shows an exemplary performance of a probe set for a VMG biochip with DNA extracted from several types of water samples. Genomic DNA (10 pg) of each pathogen was spiked into 10 ng of background DNA, yielding a relative abundance for each pathogen of 0.1%. Spiked samples and unspiked background samples were labeled with Cy3 and Cy5, respectively, mixed and hybridized in duplicate. For spiked samples, 673 probes (46 VMG amplicons, marker gene targets chosen from validated probes as described herein) were analyzed. For unspiked samples, 2,034 probes (67 VMG amplicons) were analyzed. Boxplots represent the distribution of the positive fraction among the VMG amplicons. The dashed line represents the selected threshold for presence/absence (positive/negative) calls.

15

20

FIG. 6 shows an exemplary derivation of optimal probe design criteria demonstrating a dependence of positive fraction on Gibb's free energy of duplex formation. (a)

25

Selectivity, calculated as the difference in the percent positive probes for targeted and not-targeted probes, as function of $\Delta G^\circ_{\text{duplex}}$. Maximum selectivity (~80%) was obtained for probes with a $\Delta G^\circ_{\text{duplex}}$ of -19.3 kcal/mole and a corresponding G+C content of 47.2%. The shaded area represents the $\Delta G^\circ_{\text{duplex}}$ and G+C content range yielding a selectivity higher than 70%. (b) Percent positive probes as function of free energy of probe-target duplex formation ($\Delta G^\circ_{\text{duplex}}$) for targeted and not-targeted probes (black circles and diamonds, respectively). Each symbol represents 40 targeted and 80 not-

30

targeted probes with error bars reflecting the standard deviation among samples and replications ($n = 6$ for targeted probes, and $n = 4$ for not-targeted probes). The G+C content on the upper X-axis was derived from a linear relationship between a $\Delta G^\circ_{\text{duplex}}$ and G+C content. For spiked samples (black circles), 673 probes were analyzed. For
5 unspiked samples (black diamonds), 2034 probes were analyzed. Probes were sorted according to ΔG° , and binned (40 probes for targeted probes and 80 probes for not-targeted probes). The y-axis refers to the fraction of probes within each bin resulting in a positive signal.

10 FIG. 7A-O shows an exemplary list of 20 mer probes for waterborne pathogens.

FIG. 8 shows an exemplary list of microorganisms and virulence marker genes for use in compositions and methods of the present inventions.

15 FIG. 9 shows an exemplary description of multiplex PCR reactions and the specificity and sensitivity achievable with the probe design of the present invention.

FIG. 10 shows an exemplary Serpentine Chip, developed for use in compositions and methods of the present inventions, please note the solid (nontransparent) base.

20

FIG. 11 shows an exemplary microfluidic DNA biochip with recirculation capabilities for detecting and quantifying infectious agents by hybridizing PCR amplified products to oligonucleotide probes as provided herein (see, Figures 12 and 13): (a) the exemplary chip was approximately 1 cm², (b) a close-up view of microfluidic channels and a
25 portion of the approximately 8,000 reactors on the chip, (c) a close-up view of 6 reactors, each with a 50 micron diameter, (d) a signal to noise ratio for 5 genes belonging to one of the 20 organisms that were tested on the chip, and (e) a laser scanned signal intensities for part of the chip experimental results (a green signal represents a positive hybridization result).

30

FIG. 12 shows an exemplary list of 50 mer oligonucleotide probes for identifying waterborne pathogens developed and used as described herein.

FIG. 13 shows an exemplary list of validated 50 mer oligonucleotide probes for identifying waterborne pathogens developed and used as described herein.

FIG. 14 shows exemplary results from spot testing 10,000 50 mer probes (see, Figure 12) attached to a Serpentine Chip, oligonucleotide probes were synthesized as described herein, hybridization products were labeled with a SY3 dye, (A) results were visualized by a laser scanner, green areas were positive, dark spots were negative, (B) demonstrates exemplary low signal to noise ratios.

FIG. 15 shows an exemplary confirmation of amplification in a serpentine PCR chip demonstrating reaction products obtained from a nonleaking chip (a) microfluidic channel, (b) PCR product detectable after the 15th cycle, and (c) demonstration of success obtaining the expected size PCR product by routine gel electrophoresis.

FIG. 16 shows the stability of exemplary freeze-dried PCR reagents (A) Optimization of trehalose concentration for freeze-dried Taq Polymerase and (B) Stability of freeze-dried PCR reagents with 15% Trehalose.

FIG. 17 shows exemplary results from a helicase-dependent isothermal amplification.

DEFINITIONS

To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

The use of the article "a" or "an" is intended to include one or more. As used in this application, the singular form "a," "an," and "the" include plural references unless the context clearly dictates otherwise. For example, the term "an agent" includes a plurality of agents, including mixtures thereof.

As used herein, the term "serpentine chip" refers to a chip with serpentine shaped microchannels.

As used herein, the term "arraying" in reference to a primer or probe refers to the use of the primer or probe in a microarray, for example, by attaching the probe or primer to the microarray surface, adding the probe or primer to the microarray, and the like.

As used herein, the term "transducer device" refers to a device that is capable of converting a non-electrical phenomenon into electrical information, and transmitting the information to a device that interprets the electrical signal. Such devices can include, but are not limited to, devices that use photometry, fluorometry, and chemiluminescence; fiber optics and direct optical sensing (*e.g.*, grating coupler); surface plasmon resonance; potentiometric and amperometric electrodes; field effect transistors; piezoelectric sensing; and surface acoustic wave.

As used herein, the term "optical transparency" refers to the property of matter whereby the matter is capable of transmitting light such that the light can be observed by visual light detectors (*e.g.*, eyes and detection equipment).

As used herein, "electroluminescence" or "EL" refer to a direct conversion of electrical energy into light.

As used herein, "electroluminescent sheet" and "electroluminescent film" and "ELM" and "electroluminescent panel" and "electroluminescent wire" and "EL lamp" refer to a type of capacitor comprising a thin layer of light emitting phosphor located between two electrodes, wherein the first electrode is opaque and the second electrode is translucent to allow light to escape. For one example, AC current (400 - 1600 Hz) is applied to a phosphor resulting in the emission of light wherein phosphor chemical composition and dye pigments determine the brightness and color of the emitted light. Examples of electroluminescent compositions include ruthenium, ruthenium polypyridyl complex, tris-bipyridine, tris(bipyridine) Ru(II) complexes, tripropylamine, Ruthenium Tris(2,2')bipyridyl/Tripropylamine, and the like.

As used herein, "phosphor" refers to a substance that exhibits the phenomenon of phosphorescence, such as a natural substance, for example a transition metal compound or rare earth compounds of various types or synthetic, for example, a suitable host

material, to which an activator is added such as a copper-activated zinc sulfide and the silver-activated zinc sulfide (zinc sulfide silver).

As used herein, "ACTFEL" and "alternating current thin film electroluminescence" refer to emitted light following exposure to an electrical current.

5 As used herein, the term "film" refers to any substance capable of coating at least a portion of a substrate surface and immobilizing capture particles. Examples of materials used to make such films include, but are not limited to, agarose, acrylamide, SEPHADEX, proteins (e.g. bovine serum albumin (BSA), polylysine, collagen, etc.), hydrogels (e.g. polyethylene oxide, polyvinyl alcohol, polyhydroxyl butylate, etc.), film
10 forming latexes (e.g. methyl and ethyl acrylates, vinylidene chloride, and copolymers thereof), or mixtures thereof. In certain embodiments, films include additional material such as plasticisers (e.g. polyethylene glycol [PEG], detergents, etc.) to improve stability and/or performance of the film. In preferred embodiments, a film is a material that will react with the capture particles and present them in the same focal plane. In other
15 preferred embodiments, a film is pre-activated with cross-linking groups such as aldehydes, or groups added after the film has been formed.

As used herein, "optical signal" refers to any energy (e.g. photodetectable energy) emitted from a sample (e.g., produced from a microarray that has one or more optically excited [i.e., by electromagnetic radiation] molecules bound to its surface).

20 As used herein, "filter" refers to a device or coating that preferentially allows light of a characteristic spectra to pass through it (e.g., the selective transmission of light beams). "Polychromatic" or "broadband" as used herein, refers to a plurality of electromagnetic wavelengths emitted from a light source.

As used herein, "microarray" refers to a substrate with a plurality of molecules
25 (e.g., nucleotides) bound to its surface. Microarrays, for example, are described generally in Schena, "Microarray Biochip Technology," Eaton Publishing, Natick, Mass., 2000. Additionally, the term "patterned microarrays" refers to microarray substrates with a plurality of molecules non-randomly bound to its surface. As used herein, the term "optical detector" or "photodetector" refers to a device that generates an output signal
30 when irradiated with optical energy. Thus, in its broadest sense the term optical detector system is taken to mean a device for converting energy from one form to another for the

purpose of measurement of a physical quantity or for information transfer. Optical detectors include but are not limited to photomultipliers and photodiodes.

As used herein, the term "photomultiplier" or "photomultiplier tube" refers to optical detection components that convert incident photons into electrons via the photoelectric effect and secondary electron emission. The term photomultiplier tube is meant to include devices that contain separate dynodes for current multiplication as well as those devices which contain one or more channel electron multipliers.

As used herein, the term "photodiode" refers to a solid-state light detector type including, but not limited to PN, PIN, APD and CCD.

As used herein, the terms "plate reader" in reference to a "detection device" refer to a device to detect the transmission of light through or reflection of light (i.e., polarized light or non-polarized light of specific wavelengths) from the surface of an assay, that for the purposes of the present invention the assay is a "chip" and "PCR chip" or a "glass slide" comprising a PCR assay or a "plate" such as a 96-well plate and the like. For example, a microtiter plate reader measures transmittance, absorbance, or reflectance through, in, or from each well of a multitest device such as a microtiter testing plate (e.g., MicroPlate.TM. testing plates) or a miniaturized testing card (e.g., MicroCard.TM. miniaturized testing cards).

As used herein, "thin layer" refers to a very thin deposition of a colloidal substance (phosphor, dielectric, silver) onto the ITO coated glass plate.

As used herein, "ITO" refers to an Indium Tin Oxide ($\text{In}_{203}\text{:Sn}_{02}$) A thin layer of indium oxide that has been doped with tin; transparent, conductive coating on glass plate.

As used herein, "phosphor" refers to a powder made of a material such as zinc sulfide, doped with either copper or manganese to achieve the emission colors when exposed to an electric field.

As used herein, "dielectric layer" refers to an insulating layer that serves to even out the electric field across the phosphor layer and prevents short circuits.

As used herein, "capacitor" refers to an electrical device that can store energy in the electric field between a pair of closely spaced conductors or 'plates'. When current is passed through the capacitor, electric charges of equal magnitude, but opposite sign, build up on each plate.

As used herein, "electrode" refers to a plate of a capacitor, for example, one front electrode, such an electrode comprising as transparent ITO and one back electrode, such as an electrode comprising silver.

As used herein, "electronic power supply" refers to an electronic device that
5 produces a particular DC voltage or current from a source of electricity such as a battery or wall outlet.

As used herein, "power adapter," "transformer," or "power supply" refer to an external power supply for laptop computers or portable or semi-portable electronic device
As used herein, "AC adapter" refers to a rectifier to convert AC current to DC and a
10 transformer to convert voltage from 120 down to 9, 12, 15 or whatever is required.

As used herein, "power supply" refers to an electrical system that converts AC current from the wall outlet into the DC currents required by the computer circuitry. A computer power supply typically generates multiple voltages. For example, 12 volts is used for drives, and either 3.3 or 5 volts is used for the electronic circuitry.

15 As used herein, "external AC adaptor power brick" refers to an electronic device that produces AC current.

As used herein, "AC powered linear power supply" refers to a transformer to convert the voltage from the wall outlet to a lower voltage. An array of diodes called a diode bridge then rectifies the AC voltage to DC voltage. A low-pass filter smoothes out
20 the voltage ripple that is left after the rectification. Finally a linear regulator converts the voltage to the desired output voltage, along with other possible features such as current limiting.

As used herein, "AC current" and "Alternating Current" and "AC" refers to a type of electrical current, the direction of which is reversed at regular intervals or cycles. In
25 the United States, the standard is 120 reversals or 60 cycles per second.

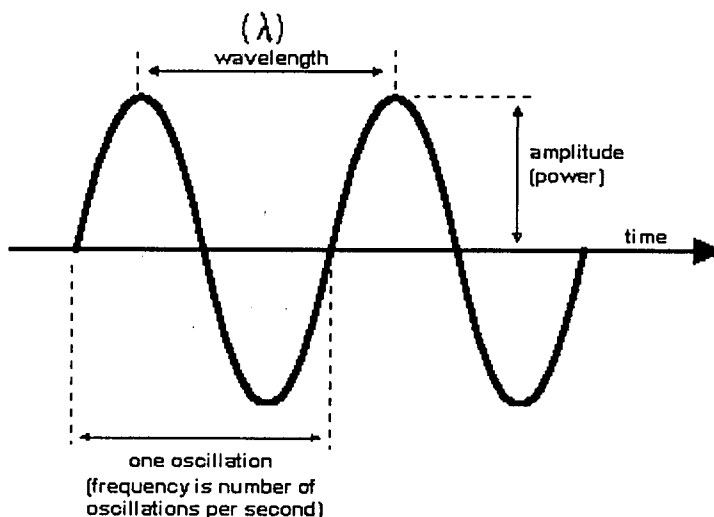
As used herein, "DC current" and "Direct Current" and "DC" refers to a type of electricity transmission and distribution by which electricity flows in one direction through the conductor, usually relatively low voltage and high current. For typical 120 volt or 220 volt devices, DC must be converted to alternating current.

30 As used herein, "battery" refers to a device that stores chemical energy and makes it available in an electrical form. Batteries comprise electrochemical devices such as one

or more galvanic cells, fuel cells or flow cell, examples include, lead acid, nickel cadmium, nickel metal hydride, lithium ion, lithium polymer, CMOS battery and the like.

As used herein, "CMOS battery" refers to a battery that maintains the time, date, hard disk and other configuration settings in the CMOS memory.

5 As used herein, "frequency" refers to a number of oscillations (vibrations) in one second. Frequency f is the reciprocal of the time T taken to complete one cycle (the period), or $1/T$. The frequency with which earth rotates is once per 24 hours. Frequency is usually expressed in units called hertz (Hz). Frequency is measured in terms "hertz" or "Hz" that refer to "oscillations per second" or "cycles per second." For example, "one
10 hertz" or "1 Hz" is equal to one cycle per second, *for example*, "one kilohertz" or "kHz" is 1,000 Hz, and "one megahertz" or "MHz" is 1,000,000 Hz. *For example*, the alternating current in a wall outlet in the U.S. and Canada is 60Hz. Electromagnetic radiation is measured in kiloHertz (kHz), megahertz (MHz) and gigahertz (GHz).



15 Frequency

As used herein, "inverter " or "rectifier" refer to a device that converts direct current electricity to alternating current either for stand-alone systems or to supply power to an electricity grid.

As used herein, "volt" and "V" refer to a unit of electrical force equal to that
20 amount of electromotive force that will cause a steady current of one ampere to flow through a resistance of one ohm.

As used herein, "voltage " refers to an amount of electromotive force, measured in volts, that exists between two points.

As used herein, "Ohm " refers to a measure of the electrical resistance of a material equal to the resistance of a circuit in which the potential difference of 1 volt
5 produces a current of 1 ampere.

As used herein, "Ampere" and "amp" refers to a unit of electrical current or rate of flow of electrons, such that one volt across one ohm of resistance causes a current flow of one ampere.

As used herein, "watt" or "W" refer to a measure of power, i.e., Volts multiplied
10 by Amps = Watts. Watt may also refer to a rate of energy transfer equivalent to one ampere under an electrical pressure of one volt, for examples, one watt equals 1/746 horsepower, or one joule per second, i.e., voltage x current =amperage.

As used herein, "Charge-Coupled Device" and "CCD" refers to an electronic memory that records the intensity of light as a variable charge.

15 As used herein, "storage CCDs" refers to either a separate array (frame transfer) or individual photosites (interline transfer) coupled to each imaging photosite.

As used herein, "CMOS" or " Complementary-symmetry/metal-oxide semiconductor" refers to a both a particular style of digital circuitry design and the family of processes used to implement that circuitry on integrated circuits (chips).

20 As used herein, "CMOS IMAGE SENSOR" refers to a "CMOS-based chip" that records intensities of light as variable charges similar to a CCD chip. In one embodiment, as CMOS chip use less power than a CCD chip.

As used herein, "optical signal" refers to any energy (*e.g.*, photodetectable energy) from a sample (*e.g.*, produced from a microarray that has one or more optically excited
25 [*i.e.*, by electromagnetic radiation] molecules bound to its surface).

As used herein, "filter" refers to a device or coating that preferentially allows light of a characteristic spectra to pass through it (*e.g.*, the selective transmission of light beams).

"Polychromatic" and "broadband" as used herein, refer to a plurality of
30 electromagnetic wavelengths emitted from a light source or sample.

As used herein, the term "micromirror array" refers to a plurality of individual light reflecting surfaces that are addressable (*e.g.*, electronically addressable in any combination), such that one or more individual micromirrors can be selectively tilted, as desired.

5 As used herein, the terms "optical detector" and "photodetector" refers to a device that generates an output signal when exposed to optical energy. Thus, in its broadest sense, the term "optical detector system" refers devices for converting energy from one form to another for the purpose of measurement of a physical quantity and/or for information transfer. Optical detectors include but are not limited to photomultipliers and
10 photodiodes, as well as fluorescence detectors.

As used herein, the term "TTL" stands for Transistor-Transistor Logic, a family of digital logic chips that comprise gates, flip/flops, counters etc. The family uses zero Volt and five Volt signals to represent logical "0" and "1" respectively.

As used herein, the term "dynamic range" refers to the range of input energy over
15 which a detector and data acquisition system is useful. This range encompasses the lowest level signal that is distinguishable from noise to the highest level that can be detected without distortion or saturation.

As used herein, the term "noise" in its broadest sense refers to any undesired disturbances (*i.e.*, signal not directly resulting from the intended detected event) within
20 the frequency band of interest. One example of noise is the summation of unwanted or disturbing energy introduced into a system from man-made and natural sources. In another example, noise may distort a signal such that the information carried by the signal becomes degraded or less reliable.

As used herein, the term "signal-to-noise ratio" refers the ability to resolve true
25 signal from the noise of a system. One example of computing a signal-to-noise ratio is by taking the ratio of levels of the desired signal to the level of noise present with the signal. In preferred embodiments of the present invention, phenomena affecting signal-to-noise ratio include, but are not limited to, detector noise, system noise, and background artifacts.

30 As used herein, the term "detector noise" refers to undesired disturbances (*i.e.*, signal not directly resulting from the intended detected energy) that originate within the

detector. Detector noise includes dark current noise and shot noise. Dark current noise in an optical detector system results from the various thermal emissions from the photodetector. Shot noise in an optical system is the product of the fundamental particle nature (*i.e.*, Poisson-distributed energy fluctuations) of incident photons as they pass
5 through the photodetector.

As used herein, the term "system noise" refers to undesired disturbances that originate within the system. System noise includes, but is not limited to noise contributions from signal amplifiers, electromagnetic noise that is inadvertently coupled into the signal path, and fluctuations in the power applied to certain components (*e.g.*, a
10 light source)

As used herein, the term "background artifacts" include signal components caused by undesired optical emissions from the microarray. These artifacts arise from a number of sources, including: non-specific hybridization, intrinsic fluorescence of the substrate and/or reagents, incompletely attenuated fluorescent excitation light, and stray ambient
15 light. In some embodiments, the noise of an optical detector system is determined by measuring the noise of the background region and noise of the signal from the microarray feature.

As used herein, the term "processor" refers to a device that performs a set of steps according to a program (*e.g.*, a digital computer). Processors, for example, include
20 Central Processing Units ("CPUs"), electronic devices, and systems for receiving, transmitting, storing and/or manipulating digital data under programmed control.

As used herein, the terms "memory device," and "computer memory" refer to any data storage device that is readable by a computer, including, but not limited to, random access memory, hard disks, magnetic (*e.g.*, floppy) disks, zip disks, compact discs,
25 DVDs, magnetic tape, and the like.

The term "gene" refers to a nucleic acid (*e.g.*, DNA) sequence that comprises coding sequences necessary for the production of a polypeptide or precursor. It is intended that the term encompass polypeptides encoded by a full length coding sequence, as well as any portion of the coding sequence, so long as the desired activity and/or
30 functional properties (*e.g.*, enzymatic activity, ligand binding, etc.) of the full-length or fragmented polypeptide are retained. The term also encompasses the coding region of a

structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences that are located 5' of the coding region and which are present on the mRNA are referred to as "5' untranslated sequences." The sequences that are located 3' (*i.e.*, "downstream") of the coding region and that are present on the mRNA are referred to as "3' untranslated sequences." The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form of a genetic clone contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms, such as "polypeptide" and "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

As used herein, the terms "nucleic acid molecule encoding," "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides or polynucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide or polynucleotide, referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide or polynucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements that direct transcription of a linked gene are generally located 5' or upstream of the coding region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

As used herein, the terms "an oligonucleotide having a nucleotide sequence encoding a gene" and "polynucleotide having a nucleotide sequence encoding a gene," means a nucleic acid sequence comprising the coding region of a gene or, in other words, the nucleic acid sequence that encodes a gene product. The coding region may be present in either a cDNA, genomic DNA, or RNA form. When present in a DNA form, the oligonucleotide or polynucleotide may be single-stranded (*i.e.*, the sense strand) or double-stranded. Suitable control elements such as enhancers/promoters, splice junctions, polyadenylation signals, etc. may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript.

As used herein, the term "regulatory element" refers to a genetic element that controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element that facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements include splicing signals, polyadenylation signals, termination signals, etc.

As used herein, the terms "complementary" and "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification and hybridization reactions, as well as detection methods that depend upon binding between nucleic acids.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

A gene may produce multiple RNA species that are generated by differential splicing of the primary RNA transcript. cDNAs that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cDNAs) and regions of complete non-identity (for example, representing the presence of exon "A" on cDNA 1 wherein cDNA 2 contains exon "B" instead). Because the two cDNAs contain regions of sequence identity they will both hybridize to a probe derived from the entire gene or

portions of the gene containing sequences found on both cDNAs; the two splice variants are therefore substantially homologous to such a probe and to each other.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term " T_m " is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (*See e.g.*, Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization [1985]). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of T_m .

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Those skilled in the art will recognize that "stringency" conditions may be altered by varying the parameters just described either individually or in concert. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences (*e.g.*, hybridization under "high stringency" conditions may occur between homologs with about 85-100% identity, preferably about 70-100% identity). With medium stringency conditions, nucleic acid base pairing will occur between nucleic acids with an intermediate frequency of complementary base sequences (*e.g.*, hybridization under "medium stringency" conditions may occur between

homologs with about 50-70% identity). Thus, conditions of "weak" or "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less.

The term "selectively hybridize" means that, for particular identical sequences, a substantial portion of the particular identical sequences hybridize to a given desired sequence or sequences, and a substantial portion of the particular identical sequences do not hybridize to other undesired sequences. A "substantial portion of the particular identical sequences" in each instance refers to a portion of the total number of the particular identical sequences, and it does not refer to a portion of an individual particular identical sequence. In certain embodiments, "a substantial portion of the particular identical sequences" means at least 70% of the particular identical sequences. In certain embodiments, "a substantial portion of the particular identical sequences" means at least 80% of the particular identical sequences. In certain embodiments, "a substantial portion of the particular identical sequences" means at least 90% of the particular identical sequences. In certain embodiments, "a substantial portion of the particular identical sequences" means at least 95% of the particular identical sequences.

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (*i.e.*, replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (*i.e.*, synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q-replicase, MDV-1 RNA is the specific template for the replicase (Kacian *et al.*, Proc. Natl. Acad. Sci. USA, 69:3038 [1972], herein incorporated by reference). Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin *et al.*,

Nature, 228:227 [1970], herein incorporated by reference). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace, Genomics, 4:560 [1989], herein incorporated by reference). Finally, *Taq* and *Pfu* polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H.A. Erlich (ed.), PCR Technology, Stockton Press [1989], herein incorporated by reference).

As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

As used herein, the term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of

extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, the term "probe" refers to a molecule (*e.g.*, an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification), that is capable of hybridizing to another molecule of interest (*e.g.*, another oligonucleotide). When probes are oligonucleotides they may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular targets (*e.g.*, gene sequences). In some embodiments, it is contemplated that probes used in the present invention are labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular label. With respect to microarrays, the term probe is used to refer to any hybridizable material that is affixed to the microarray for the purpose of detecting "target" sequences in the analyte.

As used herein "probe element" and "probe site" refer to a plurality of probe molecules (*e.g.*, identical probe molecules) affixed to a microarray substrate. Probe elements containing different characteristic molecules are typically arranged in a two-dimensional array, for example, by microfluidic spotting techniques or by patterned photolithographic synthesis, etc.

As used herein, the term "target," when used in reference to hybridization assays, refers to the molecules (*e.g.*, nucleic acid) to be detected. Thus, the "target" is sought to be sorted out from other molecules (*e.g.*, nucleic acid sequences) or is to be identified as being present in a sample through its specific interaction (*e.g.*, hybridization) with another agent (*e.g.*, a probe oligonucleotide). A "segment" is defined as a region of nucleic acid within the target sequence.

A "target nucleic acid sequence" is a sequence in a sample that is not a known control gene that is added to the sample. In certain embodiments, a target nucleic acid sequence serves as a template for amplification in a PCR reaction. In certain

embodiments, a target nucleic acid sequence is a portion of a larger nucleic acid sequence. In certain embodiments, a target nucleic acid sequence is a portion of a gene.

As used herein, the term "polymerase chain reaction" ("PCR") refers to the methods described in U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, hereby
5 incorporated by reference, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a
10 DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and
15 polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable
20 parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified." In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the
25 appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by the device and systems of the present invention.

30 As used herein, the terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR

steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

As used herein, the term "amplification reagents" refers to those reagents
5 (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.).

As used herein, the terms "reverse-transcriptase" and "RT-PCR" refer to a type of
10 PCR where the starting material is mRNA. The starting mRNA is enzymatically converted to complementary DNA or "cDNA" using a reverse transcriptase enzyme. The cDNA is then used as a "template" for a "PCR" reaction.

A "multiplex amplification reaction" is an amplification reaction in which two or more target nucleic acid sequences are amplified in the same reaction. A "multiplex
15 polymerase chain reaction" or "multiplex PCR" is a polymerase chain reaction method in which two or more target nucleic acid sequences are amplified in the same reaction.

A "singleplex amplification reaction" is an amplification reaction in which only one target nucleic acid sequence is amplified in the reaction. A "singleplex polymerase chain reaction" or "singleplex PCR" is a polymerase chain reaction method in which only
20 one target nucleic acid sequence is amplified in the reaction.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

As used herein, the term "recombinant DNA molecule" as used herein refers to a
25 DNA molecule that is comprised of segments of DNA joined together by means of molecular biological techniques.

As used herein, the term "antisense" is used in reference to RNA sequences that are complementary to a specific RNA sequence (*e.g.*, mRNA). Included within this definition are antisense RNA ("asRNA") molecules involved in gene regulation by
30 bacteria. Antisense RNA may be produced by any method, including synthesis by splicing the gene(s) of interest in a reverse orientation to a viral promoter that permits the

synthesis of a coding strand. Once introduced into an embryo, this transcribed strand combines with natural mRNA produced by the embryo to form duplexes. These duplexes then block either the further transcription of the mRNA or its translation. In this manner, mutant phenotypes may be generated. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. The designation (-) (i.e., "negative") is sometimes used in reference to the antisense strand, with the designation (+) sometimes used in reference to the sense (i.e., "positive") strand.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" or "isolated polynucleotide" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell genome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be utilized to express a protein, the oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide or polynucleotide may single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide or polynucleotide may be double-stranded).

As used herein the term "coding region" when used in reference to a structural gene refers to the nucleotide sequences that encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. The coding region is bounded, in eukaryotes, on the 5' side by the nucleotide triplet "ATG" that encodes the initiator methionine and on the 3' side by one of the three triplets which specify stop codons (i.e., TA, TAG, TGA).

As used herein, the terms "purified" and "to purify" refer to the removal of contaminants from a sample.

The term "recombinant DNA molecule" as used herein refers to a DNA molecule that is comprised of segments of DNA joined together by means of molecular biological techniques.

As used herein the term "portion" when in reference to a nucleotide sequence (as in "a portion of a given nucleotide sequence") refers to fragments of that sequence. The fragments may range in size from four nucleotides to the entire nucleotide sequence minus one nucleotide.

The terms "recombinant protein" and "recombinant polypeptide" as used herein refer to a protein molecule that are expressed from a recombinant DNA molecule.

As used herein the term "biologically active polypeptide" refers to any polypeptide which maintains a desired biological activity.

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid.

As used herein, the terms "microbe" and "microbial" refer to microorganisms. In particularly preferred embodiments, the microbes identified using the present invention are bacteria (*i.e.*, eubacteria and archaea). However, it is not intended that the present invention be limited to bacteria, as other microorganisms are also encompassed within this definition, including fungi, viruses, and parasites (*e.g.*, protozoans and helminths).

As used herein, the term "reference DNA" refers to DNA that is obtained from a known organism (*i.e.*, a reference strain). In some embodiments of the invention, the reference DNA comprises random genome fragments. In particularly preferred embodiments, the genome fragments are of approximately 1 to 2 kb in size. Thus, in preferred embodiments, the reference DNA of the present invention comprises mixtures of genomes from multiple reference strains.

As used herein, the term "multiple reference strains" refers to the use of more than one reference strains in an analysis. In some embodiments, multiple reference strains within the same species are used, while in other embodiments, "multiple reference strains" refers to the use of multiple species within the same genus, and in still further embodiments, the term refers to the use of multiple species within different genera.

As used herein, the terms "test DNA" and "sample DNA" refer to the DNA to be analyzed using the method of the present invention. In preferred embodiments, this test DNA is tested in the competitive hybridization methods of the present invention, in which reference DNA(s) from multiple reference strains is/are used.

5 The terms "sample" and "specimen" in the present specification and claims are used in their broadest sense. On the one hand, they are meant to include a specimen or culture. On the other hand, they are meant to include both biological and environmental samples. These terms encompasses all types of samples obtained from humans and other animals, including but not limited to, body fluids such as urine, blood, fecal matter, 10 cerebrospinal fluid (CSF), semen, and saliva, as well as solid tissue. These terms also refers to swabs and other sampling devices that are commonly used to obtain samples for culture of microorganisms. Biological samples may be animal, including human, fluid or tissue, food products and ingredients such as dairy items, vegetables, meat and meat by-products, and waste. Environmental samples include environmental material such as 15 surface matter, soil, water, and industrial samples, as well as samples obtained from food and dairy processing instruments, apparatus, equipment, disposable, and non-disposable items. These examples are not to be construed as limiting the sample types applicable to the present invention.

20 **DETAILED DESCRIPTION OF THE INVENTION**

The present invention relates to the design and manufacture of microarrays for use in detecting multiple organisms in a sample in a high throughput manner. In particular, the present invention relates to the detection of waterborne pathogens in a sample of water.

25 Preferably, the microarrays comprise nucleic acids probes for multiple target organisms. The probes are selected by generating a maximum number of nonoverlapping probes of a predetermined length from a gene sequence derived from the target organism, comparing the sequence of each nonoverlapping probe to sequences from a plurality of database microorganisms, identifying at least one subregion in the gene sequence derived 30 from the target microorganism, wherein the subregion contains exact matches to less than

a predetermined number of the database microorganisms; and generating a set of overlapping probes of a predetermined length for said at least one subregion.

High throughput molecular tools for the detection and identification of pathogens are in increasing demand for numerous application including biodefense, food and water safety. The present invention embodies, in part, an in situ synthesized oligonucleotide probe biochip developed for the detection of 12 pathogens. A redundant set of 791 probes was selected targeting 47 virulence and marker genes (VMGs). Overall, 95% of the probes behaved as predicted based on hybridizations with a composite sample of all VMGs. After experimental evaluation, 673 probes (85% of the initial probe set) were selected for the detection of VMGs in environmental samples. A combination of multiplex PCR and the microarray allowed the detection of the pathogens at a relative abundance level between 0.1% and 0.01% for most of the VMGs. In-depth analysis of hybridization patterns showed that optimal selectivity is achieved with probes with a hybridization free energy of -19.3 kcal/mole and a G+C content of 47.2%. It was calculated that for an optimal probe set with above-described characteristics, 16 probes are needed to obtain calls with a type I and II errors of less than 0.01. This study demonstrates that microarrays with optimized and redundant probe sets, combined with statistical evaluation of presence/absence calls, are a powerful tool for environmental analysis.

As an exemplary embodiment, the present invention encompasses a microarray (VMG biochip) using an in situ synthesized microfluidic platform targeting 93 virulence and marker genes (VMGs) in the following pathogens: *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica*, *E. faecalis*, *S. enterica*, *C. parvum*, *G. intestinalis*, *K. pneumoniae*, *E. coli*. Each VMG was targeted by 17 probes on average. Many VMGs are targeted in more than one region. This design allows for redundancy in organism detection by targeting more than one VMG in an organism and redundancy in VMG detection by targeting VMGs in more than one region. In preferred embodiments, the virulence marker genes are selected from the genes listed in the table listed in Fig. 8. Fig. 8 also provides the number of probes for

each gene on an exemplary chip. Fig. 9 presents provides a description of the which targets were amplified in multiplex PCR reactions and the specificity and sensitivity of the biochip design.

The biochips of the present invention were validated by spiking DNA of 12 pathogens into environmental background DNA extracted from river water, tertiary effluent, and tap water separately. The VMG targets were enriched via multiplex PCR to improve the detection limit, labeled, and allowed to hybridize on the biochip. The chip allowed detection 11 of the 12 pathogens presented to the biochip in environmental backgrounds. The targets of the undetected organism had a relatively low G+C content. Over all the VMG biochip is robust and could be utilized for pathogen detection in environmental samples.

It is contemplated that the methods of the present invention result in highly reliable pathogen detection. This reliability is accomplished by increasing the redundancy of the detection approach. In preferred embodiments, redundancy is achieved by using detections methods using probes to multiple genes from a single organism (e.g., a pathogen) and multiple amplicons per gene as well as using multiple probes per each Amplicon or combinations thereof. In some embodiments, high sensitivity is obtained by coupling two rounds of amplification. For example, the first round can be multiplex PCR and the second round can be isothermal Klenow amplification.

A. Probe Design and Selection

Traditional methods of probe design use heuristic rules that were developed for solution-based hybridization and are not fully applicable to hybridizations performed on solid substrates. Convention probe design also fails to consider target dangling-end influence on hybridization behavior. As a result, conventional probe design can erroneously remove probes that will behave well on solid substrates, fail to remove probes that compliment problematic target fragments, and can reduce the size of probe sets. The present invention provides a new strategy for probe design and validation of probes for use on high-throughput microarray platforms. The inventive strategy takes advantage of the fact that the availability of specific probes (i.e., those probes that

hybridize only to the target of interest and do not cross-hybridize with non-targets) varies along the target sequence.

As a non-limiting example, 16S rDNA probes were designed with the following steps: 1) the number of exact matches in the Ribosomal Database Project II (RDP-II) database maintained at Michigan State University (RDP Release 9), were determined for every non-overlapping 20-mer probe possible in a 16S rDNA gene for an organism of interest (generating about 70 – 75 probes for the target sequence) and 2) generating additional probes by overlapping 19 bases in the sub-region of 16S rDNA with less than about 10 to 50 exact matches. The probes were validated using targets with specified dangling-ends for 18-mer probes. This demonstrated that the length and composition of target dangling-ends can significantly influence sensitivity and cross hybridization in reference and community assays can influence confidence in probe validation.

The specificity of the probes was examined using a single organism (*Burkholderia xenovorans* LB400), which was hybridized with a complex sample of DNA extracted from an anaerobic bioreactor. Since LB400 is not present in the bioreactor sample, this allowed evaluation of the specificity and sensitivity of LB400 probes. Results showed that specific signal increased as the conservation of probe sequences decreased (as determined by number of exact matches). Over 77% of perfect match probes displayed specificity, sensitivity and thermal melting curve characteristics in complex community samples, demonstrating that the extended probe design of the present invention is effective for retaining a larger number of high quality probes. Furthermore, portions of the sub-region yield greater specificity with as little as 1 bp difference and provide numerous probes for further evaluation. Additionally, amplified 16S rDNA provides enough target DNA to duplex with >900 perfect match probes.

Accordingly, in some preferred embodiments, the present invention provides a microarray comprising nucleic acids probes for a target organism, wherein the probes are selected by generating a maximum number of nonoverlapping probes of a predetermined length from a gene sequence derived from the target organism, comparing the sequence of each nonoverlapping probe to sequences from a plurality of database microorganisms, identifying at least one subregion in the gene sequence derived from a target microorganism, wherein the at least one subregion contains exact matches to less than a

predetermined number of the database microorganisms; and generating a set of overlapping probes of a predetermined length for the at least one subregion.

In some preferred embodiments, the number of exact matches is determined for every non-overlapping probe of a predetermined length possible in a gene sequence from an organism of interest. The predetermined length can be varied and can range from 10 bases to about 20, 30, 40, 50, 60, 70, 80, 90 or 100 bases or more. In preferred embodiments, predetermined length is about 20 bases. The number of exact matches can be obtained for each probe using Blast tools and public sequence databases (i.e., database sequences for database organisms). This data is then used to identify sub-regions of the genome of the organism of interest that have exact matches to less than a predetermined number of database organisms. The number of database organisms to which there are exact matches can vary from about 1 organism to about 50 organisms, preferably from about 5 to about 30 organisms, and more preferably from about 10 to about 20 organisms. In some preferred embodiments, the sub-regions range from 10 to 300 bases in length, and are more preferably from about 10 to about 20, 50, 75, 100, 150, 200, 250 or 300 bases in length. Panel A of Figure 1 provides a graphical depiction of this analysis for the 16s rDNA sequence.

Additional probes are then generated by overlapping sequences of a predetermined number of bases in the identified sub-regions. In preferred embodiments, the sub-regions range from 10 to 300 bases in length, and are more preferably from about 10 to about 20, 50, 75, 100, 150, 200, 250 or 300 bases in length. The sequences used for overlapping the predetermined region are preferably from about 10 to about 50 bases in length, and in some particularly preferred embodiments are from 10 to 30 bases in length, and most preferably about 19 bases in length. Since some portions of the sub-region yield greater specificity with as little as 1 base pair difference, numerous probes are identified. See, Panel A, Figure 1. In some embodiments, the probes are then validated using PCR to ensure sufficient presence of the target. In some embodiments, the optimal length and composition of target dangling-end is determined to minimize hybridization between targets, thus reducing ambiguous hybridization behavior. In some preferred embodiments, targets are fragmented to a predetermined length to eliminate cross

hybridization in a community sample and to reduce target-target dangling end hybridization.

In certain embodiments, a probe may include Watson-Crick bases or modified bases. Modified bases include, but are not limited to, the AEGIS bases (from Eragen Biosciences), which have been described, e.g., in U.S. Pat. Nos. 5,432,272; 5,965,364; and 6,001,983. In certain embodiments, bases are joined by a natural phosphodiester bond or a different chemical linkage. Different chemical linkages include, but are not limited to, a peptide bond or an LNA linkage, which is described, e.g., in published PCT applications WO 00/56748; and WO 00/66604.

In some embodiments, the means for generating the polynucleotide probes, for example, for use on a microarray, is by synthesis of synthetic polynucleotides or oligonucleotides, e.g., using N-phosphonate or phosphoramidite chemistries (Froehler et al., 1986, Nucleic Acid Res. 14:5399-5407; McBride et al., 1983, Tetrahedron Lett. 24:246-248). Other methods for synthesizing probes include those described in U.S. Pat. Nos. 6,965,040 and 6,426,184, both of which are incorporated herein by reference. Synthetic sequences are typically between about 15 and about 600 bases in length, more typically between about 20 and about 100 bases, most preferably between about 40 and about 70 bases in length.

B. Amplification Target Nucleic Acid Sequences

In some embodiments, the present invention is directed to the detection of target nucleic acid sequences from a target organism. In certain embodiments, target nucleic acid sequences include RNA and DNA. Exemplary RNA target sequences include, but are not limited to, mRNA, rRNA, tRNA, snRNA, viral RNA, and variants of RNA, such as splicing variants. Exemplary DNA target sequences include, but are not limited to, genomic DNA, plasmid DNA, phage DNA, nucleolar DNA, mitochondrial DNA, chloroplast DNA, cDNA, synthetic DNA, yeast artificial chromosomal DNA ("YAC"), bacterial artificial chromosome DNA ("BAC"), other extrachromosomal DNA, and primer extension products. Target nucleic acid sequences also include, but are not limited to, analogs of both RNA and DNA. Exemplary nucleic acid analogs include, but are not limited to, locked nucleic acids ("LNAs"), peptide nucleic acids ("PNAs"), 8-aza-7-

deazaguanine ("PPG's"), and other nucleic acid analogs. Exemplary target nucleic acid sequences include, but are not limited to, chimeras of RNA and DNA.

A variety of methods are available for obtaining a target nucleic acid sequence.

When the nucleic acid target is obtained through isolation from a biological matrix,

5 certain isolation techniques include, but are not limited to, (1) organic extraction followed by ethanol precipitation, e.g., using a phenol/chloroform organic reagent (e.g., Ausubel et al., eds., Current Protocols in Molecular Biology Volume 1, Chapter 2, Section I, John Wiley & Sons, New York (1993), herein incorporated by reference), in certain embodiments, using an automated nucleic acid extractor, e.g., the Model 341 DNA
10 Extractor available from Applied Biosystems (Foster City, Calif.); (2) stationary phase adsorption methods (e.g., Boom et al., U.S. Pat. No. 5,234,809; Walsh et al., Biotechniques 10(4): 506-513 (1991), all of which are herein incorporated by reference); and (3) salt-induced nucleic acid precipitation methods (e.g., Miller et al., Nucleic Acids Research, 16(3): 9-10 (1988), herein incorporated by reference), such precipitation
15 methods being typically referred to as "salting-out" methods. In certain embodiments, the above isolation methods may be preceded by an enzyme digestion step to help eliminate unwanted protein from the sample, e.g., digestion with proteinase K, or other like proteases. See, e.g., U.S. patent application Ser. No. 09/724,613, herein incorporated by reference.

20 In certain embodiments, a target nucleic acid sequence may be derived from any living, or once living, organism, including but not limited to, a prokaryote, a eukaryote, a plant, an animal, and a virus. In certain embodiments, a target nucleic acid sequence is derived from a human. In certain embodiments, the target nucleic acid sequence may originate from a nucleus of a cell, e.g., genomic DNA, or may be extranuclear nucleic
25 acid, e.g., originate from a plasmid, a mitochondrial nucleic acid, from various RNAs, and the like. In certain embodiments, if the sequence from the organism is RNA, it may be reverse-transcribed into a cDNA target nucleic acid sequence. In certain embodiments, the target nucleic acid sequence may be present in a double-stranded or single-stranded form.

30 In some embodiments, target nucleic acid sequences are amplified. In certain embodiments, multiple target nucleic acid sequences can be amplified in the same

reaction (e.g., in multiplex amplification reactions). In certain embodiments, more than one different multiplex amplification reaction is performed. In certain embodiments, 5 to 10 different multiplex amplification reactions are performed. In certain embodiments, 10 to 25 different multiplex amplification reactions are performed. In certain embodiments, 5 25 to 50 different multiplex amplification reactions are performed. In certain embodiments, greater than 50 different multiplex amplification reactions are performed.

In certain embodiments, a sufficient number of different amplification reactions can be performed such that all of the target nucleic acid sequences together represent all of the genes in a genome. In certain embodiments, the genome may be derived from any 10 living, or once living organism including but not limited to, a prokaryote, a eukaryote, a plant, an animal, and a virus. In certain embodiments, the genome is prokaryotic. In certain embodiments, a sufficient number of different amplification reactions can be performed such that all of the target nucleic acid sequences together represent most of the genes in a genome. In certain embodiments, a sufficient number of different amplification 15 reactions can be performed such that all of the target nucleic acid sequences together represent all of the nucleic acids in a transcriptome. In certain embodiments, a sufficient number of different amplification reactions can be performed such that all of the target nucleic acid sequences together represent most of the nucleic acids in a transcriptome. The term "transcriptome" refers to the activated genes, mRNAs, and/or transcripts found 20 in a particular tissue at a particular time.

Exemplary target nucleic acid sequences include, but are not limited to, amplification products, ligation products, transcription products, reverse transcription products, primer extension products, methylated DNA, and cleavage products. Exemplary amplification products include, but are not limited to, PCR and isothermal 25 products.

Different target nucleic acid sequences may be different portions of a single contiguous nucleic acid or may be on different nucleic acids. Different portions of a single contiguous nucleic acid may or may not overlap.

In certain embodiments, nucleic acids in a sample may be subjected to a cleavage 30 procedure. In certain embodiments, such cleavage products may be target nucleic acid sequences.

In certain embodiments, a target nucleic acid sequence is derived from a crude cell lysate. Examples of target nucleic acid sequences include, but are not limited to, nucleic acids from buccal swabs, crude bacterial lysates, blood, skin, semen, hair, bone, mucus, saliva, cell cultures, and tissue biopsies.

5 In some embodiments, one or more target nucleic acid samples are derived from communities of microorganisms. In some embodiments, the community of microorganisms is present in a public water source, such as a well, reservoir, river, stream, or lake.

10 In certain embodiments, target nucleic acid sequences are obtained from a cell, cell line, tissue, or organism that has undergone a treatment. In certain embodiments, the treatment results in the up-regulation or down-regulation of certain target nucleic acid sequences in treated cells, cell lines, tissues, or organisms.

In certain embodiments, a target nucleic acid sequence is obtained from a single cell. In certain embodiments, a target nucleic acid sequence is obtained from tens of cells.
15 In certain embodiments, a target nucleic acid sequence is extracted from hundreds of cells or more. In certain embodiments, a target nucleic acid sequence is extracted from cells of a single organism. In certain embodiments, a target nucleic acid sequence is extracted from cells of two or more different organisms. In certain embodiments, a target nucleic acid sequence concentration in a PCR reaction ranges from about 1 to about 10,000,000
20 molecules per reaction.

In certain embodiments, each primer is sufficiently long to prime the template-directed synthesis of the target nucleic acid sequence under the conditions of the amplification reaction. In certain embodiments, the lengths of the primers depends on many factors, including, but not limited to, the desired hybridization temperature between
25 the primers, the target nucleic acid sequence and the complexity of the different target nucleic acid sequences to be amplified, and other factors. In certain embodiments, a primer is about 15 to about 35 nucleotides in length. In certain embodiments, a primer is fewer than 15 nucleotides in length. In certain embodiments, a primer is greater than 35 nucleotides in length.

30 In certain embodiments, a set of primers comprises at least one set of primers which comprises at least one designed portion and at least one random portion. In certain

embodiments, the designed portion of a primer set is at the 5' end of the primers. In certain embodiments, the designed portion of a primer set is at the 3' end of the primers. In certain embodiments, the designed portion of a primer set is in the center of the primers. In certain embodiments, the designed portion of a primer set includes two or more designed portions. In certain embodiments, the designed portions of a primer set are located in two or more portions separated by random portions.

The amplification methods used in the present invention generally use polymerases. In certain embodiments, a polymerase is active at 37°C. In certain embodiments, a polymerase is active at a temperature other than 37°C. In certain embodiments, a polymerase is active at a temperature greater than 37°C. In certain embodiments, a polymerase is active at both 37°C. and other temperatures.

In certain embodiments, a thermostable polymerase remains active at a temperature greater than about 42°C. In certain embodiments, a thermostable polymerase remains active at a temperature greater than about 50°C. In certain embodiments, a thermostable polymerase remains active at a temperature greater than about 60°C. In certain embodiments, a thermostable polymerase remains active at a temperature greater than about 70°C. In certain embodiments, a thermostable polymerase remains active at a temperature greater than about 80°C. In certain embodiments, a thermostable polymerase remains active at a temperature greater than about 90°C.

Exemplary thermostable polymerases include, but are not limited to, *Thermus thermophilus* HB8 (described, e.g., in U.S. Pat. No. 5,789,224); mutant *Thermus thermophilus* HB8, including, but not limited to, *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus thermophilus* HB8 (A271; F669Y; E683W), and *Thermus thermophilus* HB8 (D18A; F669Y); *Thermus oshimai* (described, e.g., in U.S. Provisional Application No. 60/334,798, filed Nov. 30, 2001, corresponding to U.S. Application No. 20030194726, *Thermus oshimai* Nucleic Acid Polymerases, published Oct. 16, 2003); mutant *Thermus oshimai*, including, but not limited to, *Thermus oshimai* (G43D; F665Y); *Thermus scotoductus* (described, e.g., in U.S. Provisional Application No. 60/334,489, filed Nov. 30, 2001); mutant *Thermus scotoductus*, including, but not limited to, *Thermus scotoductus* (G46D; F668Y); *Thermus thermophilus* 1B21 (described, e.g., in U.S. Provisional Application No. 60/336,046, filed Nov. 30, 2001),

mutant *Thermus thermophilus* 1B21, including, but not limited to, *Thermus thermophilus* 1B21 (G46D; F669Y); *Thermus thermophilus* GK24 (described, e.g., in U.S. Provisional Application No. 60/336,046, filed Nov. 30, 2001); mutant *Thermus thermophilus* GK24, including, but not limited to, *Thermus thermophilus* GK24 (G46D; F669Y); *Thermus*
 5 *aquaticus* polymerase; mutant *Thermus aquaticus* polymerase, including, but not limited to, *Thermus aquaticus* (G46D; F667Y) (AmpliTaq.RTM. FS or Taq (G46D; F667Y), described, e.g., in U.S. Pat. No. 5,614,365), Taq (G46D; F667Y; E6811), and Taq (G46D; F667Y; T664N; R660G); *Pyrococcus furiosus* polymerase; mutant *Pyrococcus furiosus* polymerase; *Thermococcus gorgonarius* polymerase; mutant *Thermococcus*
 10 *gorgonarius* polymerase; *Pyrococcus* species GB-D polymerase; mutant *Pyrococcus* species GB-D polymerase; *Thermococcus* sp. (strain 9.degree.N-7) polymerase; mutant *Thermococcus* sp. polymerase; *Bacillus stearothermophilus* polymerase; mutant *Bacillus stearothermophilus* polymerase; *Tsp* polymerase; mutant *Tsp* polymerase; ThermalAce™ polymerase (Invitrogen); *Thermus flavus* polymerase; mutant *Thermus flavus*
 15 polymerase; *Thermus litoralis* polymerase; mutant *Thermus litoralis* polymerase. In certain embodiments, a thermostable polymerase is a mutant of a naturally-occurring polymerase.

Exemplary non-thermostable polymerases include, but are not limited to DNA polymerase I; mutant DNA polymerase I, including, but not limited to, Klenow fragment
 20 and Klenow fragment (3'>5' exonuclease minus); T4 DNA polymerase; mutant T4 DNA polymerase; T7 DNA polymerase; mutant T7 DNA polymerase; phi29 DNA polymerase; and mutant phi29 DNA polymerase.

In certain embodiments, a polymerase is a processive polymerase. In certain embodiments, a processive polymerase remains associated with the template for two or
 25 more nucleotide additions. In certain embodiments, a non-processive polymerase disassociates from the template after the addition of each nucleotide. In certain embodiments, a processive DNA polymerase has a characteristic polymerization rate. In certain embodiments, a processive DNA polymerase has a polymerization rate of between 5 to 300 nucleotides per second. In certain embodiments, a processive DNA
 30 polymerase has a higher processivity in the presence of accessory factors, such as one or more additives. In certain embodiments, the processivity of a processive DNA

polymerase may be influenced by the presence or absence of accessory single-stranded DNA-binding proteins and helicases. In certain embodiments, the net polymerization rate will depend on the enzyme concentration, because at higher concentrations there are more re-initiation events and thus the net polymerization rate is increased. In certain
5 embodiments, the processive polymerase is Bst polymerase.

"Strand displacement" as used herein refers to the phenomenon in which a chemical, physical, or biological agent causes at least partial dissociation of a nucleic acid that is hybridized to its complementary strand. In certain embodiments, a DNA polymerase is a strand displacement polymerase. In certain embodiments, a processive
10 DNA polymerase is also a strand displacement polymerase, which is capable of displacing a hybridized strand encountered during replication. In certain embodiments, a strand displacement polymerase requires a factor that facilitates strand displacement to be capable of displacing a hybridized strand encountered during replication. In certain
15 embodiments, a strand displacement polymerase is capable of displacing a hybridized strand encountered during replication in the absence of a strand displacement factor. In certain embodiments, the strand displacement polymerase lacks 5' to 3' exonuclease activity.

In certain embodiments, the dissociation of a nucleic acid that is hybridized to its complementary strand occurs in a 5' to 3' direction in conjunction with replication. In
20 certain embodiments, where a primer extension reaction forms a newly synthesized strand while displacing a second nucleic acid strand from the template nucleic acid strand, both the newly synthesized and displaced second nucleic acid strand have the same base sequence, which is complementary to the template nucleic acid strand. In certain embodiments, a molecule comprises both strand displacement activity and another
25 activity. In certain embodiments, a molecule comprises both strand displacement activity and polymerase activity. In certain embodiments, strand displacement activity is the only activity associated with a molecule. Enzymes that possess both strand displacement activity and polymerase activity include, but are not limited to, E. coli DNA polymerase 1, the Klenow fragment of DNA polymerase 1, the bacteriophage T7 DNA polymerase,
30 the bacteriophage T5 DNA polymerase, the (p29 polymerase, and the Bst polymerase. Certain methods of using enzymes possessing strand displacement activity include, but

are not limited to, those described by, e.g., Kornberg, A., DNA Replication, W.H. Freeman & Co., San Francisco, Calif., 1980.

The term "strand displacement replication" refers to nucleic acid replication which involves strand displacement. In certain embodiments, strand displacement is facilitated through the use of a strand displacement factor, such as a helicase. In certain 5 embodiments, a DNA polymerase that can perform a strand displacement replication in the presence of a strand displacement factor is used in strand displacement replication. In certain embodiments, the DNA polymerase does not perform a strand displacement replication in the absence of such a factor. Strand displacement factors useful in strand 10 displacement replication include, but are not limited to, BMRF1 polymerase accessory subunit (Tsurumi et al., J. Virology 67(12):7648-7653 (1993)), herein incorporated by reference), adenovirus DNA-binding protein (Zijderveld and van der Vliet, J. Virology 68(2):1158-1164 (1994)), herein incorporated by reference), herpes simplex viral protein ICP8 (Boehmer and Lehman, J. Virology 67(2):711-715 (1993); Skalter and Lehman, 15 Proc. Natl. Acad. Sci. USA 91(22):10665-10669 (1994), all of which are herein incorporated by reference); single-stranded DNA binding proteins (SSB; Rigler and Romano, J. Biol. Chem. 270:8910-8919 (1995), herein incorporated by reference); phage T4 gene 32 protein (Villemain and Giedroc, Biochemistry 35:14395-14404 (1996); and calf thymus helicase (Siegel et al., J. Biol. Chem. 267:13629-13635 (1992), all of which 20 are herein incorporated by reference). Strand displacement amplification (SDA) reaction methods include, but are not limited to, those described in, e.g., Fraiser et al., U.S. Pat. No. 5,648,211; Cleuziat et al., U.S. Pat. No. 5,824,517; and Walker et al., Proc. Natl. Acad. Sci. U.S.A. 89:392-396 (1992), all of which are herein incorporated by reference).

In certain embodiments, the ability of a polymerase to carry out strand 25 displacement replication can be determined by using the polymerase in a strand displacement replication assay such as those described, e.g., in U.S. Pat. No. 6,642,034, herein incorporated by reference or in a primer-block assay described, e.g., in Kong et al., J. Biol. Chem. 268:1965-1975 (1993), herein incorporated by reference.

In certain embodiments, an amplification reaction comprises a blend of 30 polymerases. In certain such embodiments, at least one polymerase possesses exonuclease activity. In certain embodiments, none of the polymerases in an

amplification reaction possess exonuclease activity. Exemplary polymerases that may be used in an amplification reaction include, but are not limited to, phi.29 DNA polymerase, taq polymerase, stoffel fragment, Bst DNA polymerase. *E. coli* DNA polymerase 1, the Klenow fragment of DNA polymerase 1, the bacteriophage T7 DNA polymerase, the bacteriophage T5 DNA polymerase, and other polymerases known in the art. In certain
5 embodiments, a polymerase is inactive in the reaction composition and is subsequently activated at a given temperature.

In certain embodiments, an amplification reaction composition is formed comprising (a) two or more target nucleic acid sequences, (b) at least one set of primers,
10 and (c) at least one polymerase. In certain embodiments, an amplification reaction composition is formed comprising two or more target nucleic acid sequences, at least one primer set, dNTPs, at least one buffering agent and at least one polymerase. In certain such embodiments, the amplification reaction is incubated under conditions that allow the formation of one or more amplification products. In certain embodiments, the
15 amplification reaction further includes one or more additives. In certain embodiments, no strand displacement factors are required for strand displacement.

In certain embodiments, an amplification reaction composition comprises strand displacement factors. Exemplary strand displacement factors include, but are not limited to, helicases and single stranded DNA binding protein. In certain embodiments, the
20 temperature of the reaction affects strand displacement. In certain embodiments, a temperature of approximately 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, 65 °C, 70 °C, 75 °C, 80 °C, 85 °C, or 90 °C. facilitates strand displacement by allowing segments of double stranded DNA to separate and reanneal.

In certain embodiments, the temperature of the amplification reaction is kept at
25 isothermal reaction conditions. The term "isothermal reaction conditions" refers to conditions wherein the temperature is kept substantially constant. In certain embodiments, isothermal reaction conditions prevent the template DNA from being completely disassociated. In certain embodiments, short primers can hybridize to a double stranded template maintained at an isothermal temperature. In certain such
30 embodiments, the primers that strand invade and anneal to the template DNA can be extended by a strand-displacing DNA polymerase. In certain embodiments, an

amplification process is isothermal at 50 °C and uses Bst DNA polymerase for strand displacement and extension. In certain embodiments, an amplification reaction uses a fragment of Bst DNA polymerase with the 3'→5' exonuclease activity removed ("the large fragment of Bst DNA polymerase").

5 Certain amplification methods include, but are not limited to, Random PCR or Primer Extension Preamplification-PCR (PEP-PCR) (Zhang et al., Proc. Natl. Acad. Sci., USA 89: 5847-51 (1992), herein incorporated by reference), Linker Adapter PCR (Miyashita et al., Cytogenet. Cell Genet. 66(1): 54-57 (1994), herein incorporated by reference), Tagged-PCR (Grothues et. al., Nuc. Acids Res. 21(5) 1321-1322 (1993),
10 herein incorporated by reference), Inter-Alu-PCR (Bicknell et. al., Genomics 10:186-192 (1991), herein incorporated by reference), Degenerate Oligonucleotide Primed-PCR (DOP-PCR) (Cheung et al., Proc. Natl. Acad. Sci., USA 93:14676-14679 (1996), herein incorporated by reference), Improved-Primer Extension Preamplification PCR (I-PEP-PCR) (Dietmaier et al., Amer. J. Pathology 154(1): 83-95 (1999), herein incorporated by
15 reference and U.S. Pat. No. 6,365,375, herein incorporated by reference), LL-DOP PCR (Kittler et al., Anal. Biochem. 300:237-244 (2002) , herein incorporated by reference), Balanced PCR amplification (Makrigiorgos et. al., Nature Biotech. 20:936-939 (2002) , herein incorporated by reference), Multiple Displacement Amplification (MDA) (U.S. Pat. Nos. 6,124,120 and 6,280,949, all of which are herein incorporated by reference),
20 and Random Primer Amplification (RPA) (U.S. Pat. No. 5,043,272, herein incorporated by reference). In certain embodiments, multiplex amplification may be used (see, e.g., Published U.S. Patent Application No. 2004-0175733 A1, herein incorporated by reference).

 In certain embodiments, multiplex amplification is used to distinguish between
25 target nucleic acid sequences that have single nucleotide polymorphisms ("SNP"). In certain such embodiments, one or more multiplex amplification reactions include one or more primer sets specific for two or more target nucleic acid sequences that differ at only a single nucleotide and are present in similar abundance. In certain such embodiments, the one or more multiplex amplification reactions further include one or more probes
30 with different detectable labels specific for the presence or absence of that particular

single nucleotide. In certain such embodiments, the signal from the label in the multiplex amplification reaction is detected as an indicator of the presence of one or more SNPs.

In certain embodiments, multiplex amplification is used for melting curve analysis. In certain such embodiments, a multiplex amplification reaction includes two or more primer sets specific for two or more target nucleic acid sequences of similar abundance and also includes one or more probes that intercalate into double-stranded target nucleic acid sequences and does not bind to single-stranded target nucleic acid sequences. In certain such embodiments, a multiplex amplification reaction includes two or more primer sets specific for two or more target nucleic acid sequences and also includes one or more probes that bind to single-stranded target nucleic acid sequences but does not bind to double-stranded target nucleic acid sequences. In certain such embodiments, the one or more probes includes a detectable label, and the label is detectable only when the one or more probes interact with their target nucleic acid sequences. In certain such embodiments, the temperature of the reaction is modified gradually and the signal from the detectable label is monitored such that the shift of the one or more target nucleic acid sequences from single-stranded to double-stranded or from double-stranded to single-stranded as a function of temperature is recorded. In certain such embodiments, the signal from the detectable label is monitored using real-time PCR.

In certain embodiments, a multiplex amplification reaction includes two or more primer sets specific for two or more target nucleic acid sequences and also includes two or more probes that bind to single-stranded target nucleic acid sequences but do not bind to double-stranded target nucleic acid sequences. In certain embodiments, the two or more probes include a detectable label, and the label is detectable only when the two or more probes interact with their target nucleic acid sequence(s). In certain embodiments, the temperature of the reaction is modified gradually and the signal from the detectable label is monitored such that the shift of one or more target nucleic acid sequences from single-stranded to double stranded or from double-stranded to single-stranded as a function of temperature is recorded. In certain such embodiments, the signal from the detectable label is monitored using real-time PCR.

In certain embodiments, one or more target nucleic acid sequences undergo a treatment before being included in an amplification reaction. In certain embodiments, a target nucleic acid treatment selectively modifies a target nucleic acid according to the methylation state of the target nucleic acid sequence (see, e.g., Published U.S. Patent Application No. 2004-0101843, U.S. Pat. No. 6,265,171; and U.S. Pat. No. 6,331,393, all of which are herein incorporated by reference). In certain embodiments, the sample from which one or more target nucleic acid sequences is derived (e.g., a cell, tissue, etc.) undergoes a treatment prior to the inclusion of the target nucleic acid sequences from that sample in a multiplex amplification reaction. In certain embodiments, the amplification of one or more target nucleic acid sequences from a treated sample is compared to the amplification of one or more target nucleic acid sequences from an untreated control sample. In certain such embodiments, the expression of one or more genes in response to the treatment is determined.

In certain embodiments, the products of two or more amplification reactions are combined. In certain such embodiments, the products of one amplification reaction may have a different amplification profile than the products of the second amplification reaction.

In preferred embodiments, multiplex PCR reactions are performed with nucleic acid samples from a community of microorganisms. In particularly preferred embodiments, primers are designed to amplify target regions identified according to the probe design and selection methods described above. The amplified target regions are designated amplicons. In some embodiments, multiple target regions from a gene of interest are amplified to produce multiple amplicons. In further embodiments, multiple target regions from multiple target genes of interest are amplified. In some embodiments, it is necessary to conduct more than one multiplex PCR reaction for each nucleic acid sample so that a large number of target regions can be amplified. For example, five multiplex PCR reactions can be performed using ten different primer sets to amplify fifty different amplicons for assay on a microarray. In some particularly preferred embodiments, the primers have a length of from 18-30 bases, a GC content of from 40-60%, and annealing temperature of from about 52 to about 58 degrees Celsius. In some embodiments, the primers are designed for detection of multiple waterborne pathogens.

In some particularly preferred embodiments, the primers are designed to amplify portions of virulence-associated genes. In some embodiments, about 5, 10, 12, 15 or 20 waterborne pathogens are analyzed. In some embodiments, about 10, 20, 30, 40, 50, 60 or more genes are analyzed, for example, about 10, 20, 30, 40, 50, 60 virulence-associated genes.

As described above, in preferred embodiments, the assay design is redundant. For each organism assayed, probes are used for multiple genes from each target organism, multiple amplicons from each gene are analyzed, and multiple discrete probes are used for each amplicon. Accordingly, in some embodiments, about 5, 10, 15, 20, 30, 40, or 50 to about 100 genes are analyzed in an assay of the present invention. In further embodiments, about 1, 3, 5, 10, 15 or 20 to about 50 amplicons per gene are amplified and analyzed in an assay of the present invention. In further embodiments, about 1, 3, 5, 10, 15 or 20 to about 50 probes for each amplicon are used in assay of the present invention. In some embodiments, the probes are attached to a microarray and amplicons are hybridized to the microarray, as described in more detail below.

In certain embodiments, a label is attached to one or more amplicons during either a first amplification with a thermostable polymerase or a second amplification with a non-thermostable polymerase. In some embodiments, the labels are included in the amplification reaction. In preferred embodiments, the labels have one or more of the following properties: (i) provides a detectable signal; (ii) interacts with a second label to modify the detectable signal provided by the second label, e.g., FRET (Fluorescent Resonance Energy Transfer); (iii) stabilizes hybridization, e.g., duplex formation; and (iv) provides a member of a binding complex or affinity set, e.g., affinity, antibody/antigen, ionic complexes, hapten/ligand, e.g., biotin/avidin. In certain embodiments, use of labels can be accomplished using any one of a large number of known techniques employing known labels, linkages, linking groups, reagents, reaction conditions, and analysis and purification methods.

Labels include, but are not limited to, light-emitting, light-scattering, and light-absorbing compounds which generate or quench a detectable fluorescent, chemiluminescent, or bioluminescent signal (see, e.g., Kricka, L. in *Nonisotopic DNA Probe Techniques* (1992), Academic Press, San Diego, pp. 3-28, and *Non-Radioactive*

Labelling, A Practical Introduction, Garman, A. J. (1997) Academic Press, San Diego).

Fluorescent reporter dyes useful as labels include, but are not limited to, fluoresceins (see, e.g., U.S. Pat. Nos. 5,188,934; 6,008,379; and 6,020,481), rhodamines (see, e.g., U.S. Pat. Nos. 5,366,860; 5,847,162; 5,936,087; 6,051,719; and 6,191,278),

- 5 benzophenoxazines (see, e.g., U.S. Pat. No. 6,140,500), energy-transfer fluorescent dyes, comprising pairs of donors and acceptors (see, e.g., U.S. Pat. Nos. 5,863,727; 5,800,996; and 5,945,526), and cyanines (see, e.g., Kubista, WO 97/45539), as well as any other fluorescent moiety capable of generating a detectable signal. Examples of fluorescein dyes include, but are not limited to, 6-carboxyfluorescein; 2',4',1,4,-
10 tetrachlorofluorescein; and 2',4',5',7',1,4-hexachlorofluorescein. In certain embodiments, the fluorescent label is selected from SYBR®-green, 6-carboxyfluorescein ("FAM"), TET, ROX, VICTM, and JOE. In certain embodiments, a label is a radiolabel.

- In certain embodiments, labels are hybridization-stabilizing moieties which serve to enhance, stabilize, or influence hybridization of duplexes, e.g. intercalators and
15 intercalating dyes (including, but not limited to, ethidium bromide and SYBR.RTM. green), minor-groove binders, and cross-linking functional groups (see, e.g., Blackburn, G. and Gait, M. Eds. "DNA and RNA structure" in Nucleic Acids in Chemistry and Biology, 2.sup.nd Edition, (1996) Oxford University Press, pp. 15-81, herein incorporated by reference). In certain embodiments, labels effect the separation or
20 immobilization of a molecule by specific or non-specific capture, for example biotin, digoxigenin, and other haptens (see, e.g., Andrus, A. "Chemical methods for 5' non-isotopic labeling of PCR probes and primers" (1995) in PCR 2: A Practical Approach, Oxford University Press, Oxford, pp. 39-54, herein incorporated by reference).

- In certain embodiments, different amplicons comprise detectable and different
25 labels that are distinguishable from one another. For example, in certain embodiments, labels are different fluorophores capable of emitting light at different, spectrally-resolvable wavelengths (e.g., 4-differently colored fluorophores); certain such labeled probes are known in the art and described above, and in, e.g., U.S. Pat. No. 6,140,054, herein incorporated by reference and Saiki et al., 1986, Nature 324:163-166, herein
30 incorporated by reference. Preferably, the detectable label is a fluorescent label, e.g., by incorporation of nucleotide analogs. Other labels suitable for use in the present invention

include, but are not limited to, biotin, imminobiotin, antigens, cofactors, dinitrophenol, lipoic acid, olefinic compounds, detectable polypeptides, electron rich molecules, enzymes capable of generating a detectable signal by action upon a substrate, and radioactive isotopes. Preferred radioactive isotopes include ^{32}P , ^{35}S , ^{14}C , ^{15}N and ^{125}I .

5 Fluorescent molecules suitable for the present invention include, but are not limited to, fluorescein and its derivatives, rhodamine and its derivatives, texas red, 5'carboxy-fluorescein (FMA), 2',7'-dimethoxy-4',5'-dichloro-6-carboxy-fluorescein (JOE), N,N,N',N'-tetramethyl-6-carboxy-rhodamine (TAMRA), 6'carboxy-X-rhodamine (ROX), HEX, TET, IRD40, and IRD41. Fluorescent molecules that are suitable for the invention
10 further include: cyamine dyes, including but not limited to Cy3, Cy3.5 and Cy5; BODIPY dyes including but not limited to BODIPY-FL, BODIPY-TR, BODIPY-TMR, BODIPY-630/650, and BODIPY-650/670; and ALEXA dyes, including but not limited to ALEXA-488, ALEXA-532, ALEXA-546, ALEXA-568, and ALEXA-594; as well as other fluorescent dyes which will be known to those who are skilled in the art. Electron rich
15 indicator molecules suitable for the present invention include, but are not limited to, ferritin, hemocyanin, and colloidal gold. Alternatively, in less preferred embodiments the target polynucleotides may be labeled by specifically complexing a first group to the polynucleotide. A second group, covalently linked to an indicator molecules and which has an affinity for the first group, can be used to indirectly detect the target
20 polynucleotide. In such an embodiment, compounds suitable for use as a first group include, but are not limited to, biotin and imminobiotin. Compounds suitable for use as a second group include, but are not limited to, avidin and streptavidin.

C. Microarrays

25 In preferred embodiments, probes designed by the methods described above are used on microarrays. A microarray is an array of positionally-addressable binding (e.g., hybridization) sites on a support. Each of such binding sites comprises a plurality of polynucleotide molecules of a probe bound to the predetermined region on the support. Microarrays can be made in a number of ways, of which several are described herein
30 below (see e.g., Meltzer, 2001, Curr. Opin. Genet. Dev. 11(3):258-63; Andrews et al., 2000, Genome Res. 10(12):2030-43; Abdellatif, 2000, Circ. Res. 86(9):919-20; Lennon,

2000, *Drug Discov. Today* 5(2):59-66; Zweiger, 1999, *Trends Biotechnol.* 17(11):429-36, all of which are herein incorporated by reference). However produced, microarrays share certain characteristics. The arrays are preferably reproducible, allowing multiple copies of a given array to be produced and easily compared with each other. Preferably, the microarrays are made from materials that are stable under binding (e.g., nucleic acid hybridization) conditions. However, both larger and smaller (e.g., 0.5 cm.^{sup.2} or less) arrays are also contemplated and may be preferable, e.g., for simultaneously evaluating a very large number of different probes. Any of these methods described herein may be used for making microarrays comprising the probes designed and identified as described above.

In preferred embodiments, methods of the invention utilize polynucleotide probes synthesized directly on the support to form the array. The probes are attached to a solid support or surface, which may be made, e.g., from glass, plastic (e.g., polypropylene, nylon), polyacrylamide, nitrocellulose, gel, or other porous or nonporous material. There are a variety of techniques known for producing arrays containing thousands of oligonucleotides complementary to defined sequences, at defined locations on a surface. For example, photolithographic techniques for synthesis in situ (see, Fodor et al., 1991, *Science* 251:767-773; Pease et al., 1994, *Proc. Natl. Acad. Sci. U.S.A.* 91:5022-5026; Lockhart et al., 1996, *Nature BioTechnology* 14:1675; U.S. Pat. Nos. 5,489,678; 5,578,832; 5,556,752; 5,510,270; 6,197,506; 5,143,854, 5,424,186 and 6,346,413, all of which are herein incorporated by reference) or other methods for rapid synthesis and deposition of defined oligonucleotides (Blanchard et al., *Biosensors & Bioelectronics* 11:687-690) may be used. In some particularly preferred embodiments, microarrays are synthesized by the methods described by Gao, X.L., et al., A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Research*, 2001. 29(22): p. 4744-4750, herein incorporated by reference. See also, U.S. Pat. Publ. 20020081582 and 20030138363 and U.S. Pat. No. 6,375,903, all of which are incorporated herein by reference.

Other methods for making microarrays, e.g., by masking (Maskos and Southern, 1992, *Nucl. Acids. Res.* 20:1679-1684, herein incorporated by reference), may also be used. In principle, and as noted supra, any type of array, for example, dot blots on a nylon

hybridization membrane (see Sambrook et al., supra) could be used. However, as will be recognized by those skilled in the art, very small arrays will frequently be preferred because hybridization volumes will be smaller.

In some embodiments, microarrays of the invention are manufactured by means of an ink jet printing device for oligonucleotide synthesis, e.g., using the methods and systems described by Blanchard in International Patent Publication No. WO 98/41531, published Sep. 24, 1998; Blanchard et al., 1996, *Biosensors and Bioelectronics* 11:687-690; Blanchard, 1998, in *Synthetic DNA Arrays in Genetic Engineering*, Vol. 20, J. K. Setlow, Ed, Plenum Press, New York at pages 111-123; Hughes et al., 2001, *Nature BioTechnology* 19:342-347; and U.S. Pat. No. 6,028,189 to Blanchard, all of which are herein incorporated by reference. Specifically, the oligonucleotide probes in such microarrays are preferably synthesized in arrays, e.g., on a glass slide, by serially depositing individual nucleotide bases in microdroplets of a high surface tension solvent such as propylene carbonate. The microdroplets have small volumes (e.g., 100 pL or less, more preferably 50 pL or less) and are separated from each other on the microarray (e.g., by hydrophobic domains) to form circular surface tension wells which define the locations of the array elements (i.e., the different probes). Polynucleotide probes are attached to the surface covalently at the 3' end of the polynucleotide.

D. Assays Using Microarrays

In a preferred embodiment in the present invention, sample processing is through hybridization of the amplicons obtained through multiplex PCR on a nucleotide microarray. In a more preferred embodiment, the microarray is an oligonucleotide array. Preferably, the microarray contains in the range of 20 to 50,000 nucleic acid probes. The probes are identified by the methods described in detail above. The probes can be arranged in a variety of patterns. For example, the probes can be arranged in rows and columns, polygonal (e.g., hexagonal), or circular patterns, etc. In preferred embodiments, the microarrays comprise a redundant set of probes. For each organism assayed, the microarray comprises probes for multiple genes from each target organism. In some embodiments, the microarray further comprises probes to multiple amplicons from each

the multiple genes that are analyzed. In further embodiments, the microarray comprises multiple discrete probes for each amplicon.

Preferably, the density of probes on a microarray is about 100 different (i.e., non-identical) probes per 1 cm^2 or higher. More preferably, a microarray used in the methods of the invention will have at least 550 probes per 1 cm^2 , at least 1000 probes per 1 cm^2 , at least 1500 probes per 1 cm^2 or at least 2000 probes per 1 cm^2 . In a particularly preferred embodiment, the microarray is a high density array, preferably having a density of at least about 2500 different probes per 1 cm^2 . The microarrays used in the invention therefore preferably contain at least 2500, at least 5000, at least 10000, at least 15000, at least 20000, at least 25000, at least 50000 or at least 55000 different (i.e., non-identical) probes.

Such polynucleotide probes are preferably of the length of 15 to 200 bases, more preferably of the length of 20 to 100 bases, most preferably 40-60 bases. It will be understood that each probe sequence may also comprise a linker (e.g., spacer) in addition to the sequence that is complementary to its target sequence. As used herein, a linker refers to a chemical structure between the sequence that is complementary to its target sequence and the surface. The linker need not be a nucleotide sequence. For example, the linker can be composed of a nucleotide sequence, or peptide nucleic acids, hydrocarbon chains, etc.

It will be appreciated that when a sample of target nucleic acid molecules, e.g., cDNA complementary to the RNA of a cell is made and hybridized to a microarray under suitable hybridization conditions, the level of hybridization to the site in the array will reflect the prevalence of the corresponding complementary sequences in the sample. For example, when detectably labeled (e.g., with a fluorophore) amplicon is hybridized to a microarray, the site on the array corresponding to a nucleotide sequence that is not in the sample will have little or no signal (e.g., fluorescent signal), and a nucleotide sequence that is prevalent in the sample will have a relatively strong signal. The relative abundance of different nucleotide sequences in a sample may be determined by the signal strength pattern of probes on a microarray.

The target polynucleotides may be from any source. For example, the target polynucleotide molecules may be naturally occurring nucleic acid molecules such as

genomic or extragenomic DNA molecules isolated from an organism, or RNA molecules, such as mRNA molecules, isolated from an organism. Alternatively, the polynucleotide molecules may be synthesized, including, e.g., nucleic acid molecules synthesized enzymatically in vivo or in vitro, such as cDNA molecules, or polynucleotide molecules synthesized by PCR, RNA molecules synthesized by in vitro transcription, etc. The sample of target polynucleotides can comprise, e.g., molecules of DNA, RNA, or copolymers of DNA and RNA. In preferred embodiments, the target polynucleotides of the invention will correspond to particular genes or to particular gene transcripts (e.g., to particular mRNA sequences expressed in cells or to particular cDNA sequences derived from such mRNA sequences). However, in many embodiments, particularly those embodiments wherein the polynucleotide molecules are derived from mammalian cells, the target polynucleotides may correspond to particular fragments of a gene transcript. For example, the target polynucleotides may correspond to different exons of the same gene, e.g., so that different splice variants of that gene may be detected and/or analyzed.

In preferred embodiments, the target polynucleotides to be analyzed are prepared in vitro from nucleic acids extracted from cells. For example, in one embodiment, RNA is extracted from cells (e.g., total cellular RNA, poly(A)+ messenger RNA, fraction thereof) and messenger RNA is purified from the total extracted RNA. Methods for preparing total and poly(A)+ RNA are well known in the art, and are described generally, e.g. in Sambrook et al., *supra*. In one embodiment, RNA is extracted from cells of the various types of interest in this invention using guanidinium thiocyanate lysis followed by CsCl centrifugation and an oligo dT purification (Chirgwin et al., 1979, *Biochemistry* 18:5294-5299). In another embodiment, total RNA is extracted from cells using guanidinium thiocyanate lysis followed by purification on RNeasy columns (Qiagen). cDNA is then synthesized from the purified mRNA using, e.g., oligo-dT or random primers. In preferred embodiments, the target polynucleotides are cRNA prepared from cDNA prepared from purified mRNA or from total RNA extracted from cells. As used herein, cRNA can either be complementary to (anti-sense) or of the same sequence (sense) as the sample RNA. The extracted RNA molecules are amplified using a process in which double-stranded cDNA molecules are synthesized from the sample RNA molecules using primers linked to an RNA polymerase promoter. As a result, RNA

polymerase promoters can be incorporated into either or both strands of the cDNA. Using the RNA polymerase promoter that is on the first strand of the cDNA molecule, cRNA can be synthesized that is the same sequence as the sample RNA. To synthesize cRNA complementary to the sample RNA, transcription can be initiated from the RNA
5 polymerase promoter that is on the second strand of the double-stranded cDNA molecule using an RNA polymerase (see, e.g., U.S. Pat. Nos. 5,891,636, 5,716,785; 5,545,522 and 6,132,997; see also, U.S. Pat. No. 6,271,002 and U.S. Provisional Patent Application Ser. No. 60/253,641, filed on Nov. 28, 2000, by Ziman et al., all of which are herein incorporated by reference). Both oligo-dT primers (U.S. Pat. Nos. 5,545,522 and
10 6,132,997, herein incorporated by reference) or random primers (U.S. Provisional Patent Application Ser. No. 60/253,641, filed on Nov. 28, 2000, by Ziman et al., herein incorporated by reference) that contain an RNA polymerase promoter or complement thereof can be used. Preferably, the target polynucleotides are short and/or fragmented polynucleotide molecules which are representative of the original nucleic acid population
15 of the cell. In one embodiment, total RNA is used as input for cRNA synthesis. An oligo-dT primer containing a T7 RNA polymerase promoter sequence can be used to prime first strand cDNA synthesis. When second strand synthesis is desired, random hexamers can be used to prime second strand cDNA synthesis by a reverse transcriptase. This reaction yields a double-stranded cDNA that contains the T7 RNA polymerase promoter at the 3'
20 end. The double-stranded cDNA can then be transcribed into cRNA by T7 RNA polymerase. In preferred embodiments, the target nucleic acids are amplified as described in detail above.

The target polynucleotides to be analyzed are preferably detectably labeled. For example, cDNA can be labeled directly, e.g., with nucleotide analogs, or indirectly, e.g.,
25 by making a second, labeled cDNA strand using the first strand as a template. Alternatively, the double-stranded cDNA can be transcribed into cRNA and labeled. In some preferred embodiments, amplicons are labeled by including suitable labeled substrates during amplification with either thermostable or non-thermostable polymerases.

30 In preferred embodiments, nucleic acid hybridization and wash conditions are chosen so that the polynucleotide molecules to be analyzed (or target polynucleotide

molecules) specifically bind or specifically hybridize to the complementary polynucleotide sequences of the array, preferably to one or more specific array sites, wherein its complementary sequence is located.

Arrays containing double-stranded probe DNA situated thereon are preferably subjected to denaturing conditions to render the DNA single-stranded prior to contacting with the target polynucleotide molecules. Arrays containing single-stranded probe DNA (e.g., synthetic oligodeoxyribonucleic acids) may need to be denatured prior to contacting with the target polynucleotide molecules, e.g., to remove hairpins or dimers which form due to self complementary sequences.

Optimal hybridization conditions will depend on the length (e.g., oligomer versus polynucleotide greater than 200 bases) and type (e.g., RNA, or DNA) of probe and target nucleic acids. General parameters for specific (i.e., stringent) hybridization conditions for nucleic acids are described in Sambrook et al., (supra), and in Ausubel et al., 1987, *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York, herein incorporated by reference. For example, when cDNA microarrays are used, typical hybridization conditions are hybridization in 5X SSC plus 0.2% SDS at 65°C. for four hours, followed by washes at 25 °C. in low stringency wash buffer (1X SSC plus 0.2% SDS), followed by 10 minutes at 25 °C. in higher stringency wash buffer (0.1X SSC plus 0.2% SDS) (Hughes et al., 2001, *Nature BioTechnology* 19:342-347). Useful hybridization conditions are also provided in, e.g., Tijessen, 1993, *Hybridization With Nucleic Acid Probes*, Elsevier Science Publishers B. V. and Kricka, 1992, *Nonisotopic DNA Probe Techniques*, Academic Press, San Diego, Calif, herein incorporated by reference.

In some preferred embodiments, when fluorescently labeled probes are used, the fluorescence emissions at each site of a transcript array can be, preferably, detected by scanning confocal laser microscopy. In one embodiment, a separate scan, using the appropriate excitation line, is carried out for each of the two fluorophores used. Alternatively, a laser can be used that allows simultaneous specimen illumination at wavelengths specific to the two fluorophores and emissions from the two fluorophores can be analyzed simultaneously (see Shalon et al., 1996, *Genome Res.* 6:639-645, herein incorporated by reference). In a preferred embodiment, the arrays are scanned with a

laser fluorescence scanner with a computer controlled X-Y stage and a microscope objective. Sequential excitation of the two fluorophores is achieved with a multi-line, mixed gas laser, and the emitted light is split by wavelength and detected with two photomultiplier tubes. Such fluorescence laser scanning devices are described, e.g., in
5 Schena et al., 1996, Genome Res. 6:639-645. Alternatively, the fiber-optic bundle described by Ferguson et al., 1996, Nature BioTechnology 14:1681-1684, herein incorporated by reference, may be used to monitor mRNA abundance levels at a large number of sites simultaneously.

Signals are recorded and, in a preferred embodiment, analyzed by computer, e.g.,
10 using a 12 bit or 16 bit analog to digital board. In one embodiment, the scanned image is despeckled using a graphics program (e.g., Hijaak Graphics Suite) and then analyzed using an image gridding program that creates a spreadsheet of the average hybridization at each wavelength at each site. If necessary, an experimentally determined correction for cross talk (or overlap) between the channels for the two fluors may be made. For any
15 particular hybridization site on the transcript array, a ratio of the emission of the two fluorophores can be calculated.

In some preferred embodiments, arrays are designed for the high-throughput analysis of multiple waterborne pathogens. In some embodiments, the assays are configured to detect the presence of from about 5, 10, 15, or 20 to about 30 different
20 waterborne pathogens. In some preferred embodiments, the assays of the present invention are designed to detect the presence of at least one of *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Yersinia enterocolitica* and
25 combinations thereof. In some preferred embodiments, target regions of Virulence Maker Genes (VMGs) are amplified and then hybridized to a microarray comprising probes that hybridize to the VMGs. In some preferred embodiments, the microarray comprises about 5, 10, 15, 20, 25, or 30 to about 100 different probes for each VMG for each organism. As described above, in some embodiments, the probes to multiple VGAs
30 from a single organism are included on the microarray and multiple different probes per amplicon from a given target VGA are included on the microarray. In preferred

embodiments, the probes are identified as described in detail above. Figure 7 provides a list of exemplary probes identified by the methods described herein and that are suitable for use on microarrays of the present invention. Accordingly, in some embodiments, the microarrays of the present invention comprise any of probes encoded by SEQ ID Nos. 1-791, or combinations thereof.

In some embodiments, the arrays are analyzed using the system described in co-pending application 60/860,702, entitled, Electroluminescent-Based Fluorescence Detection Device, filed November 22, 2006, incorporated herein by reference. In particular, the device described below can be utilized in combination with the primers, probes and microarrays described herein. For example, the present invention provides fluorescence detection devices comprising an electroluminescent light (EL) source that provide static and/or real-time fluorescent read-outs in a number of formats including visual and digital. In further examples, the present invention provides fluorescence detection devices comprising an electroluminescent light (EL) source that provides PCR assay capabilities, such as thermal cycling and/or isothermal amplification assays, computational capabilities for data read-outs, and read-out capabilities in a number of formats including visual and digital.

It is not intended that the present invention be limited by the nature of the reactions carried out in the electroluminescent fluorescence detection device. Reactions include, but are not limited to, chemical and biological reactions. Biological reactions include, but are not limited to mRNA transcription, nucleic acid amplification, DNA amplification, cDNA amplification, sequencing, and the like. It is also not intended that the invention be limited by the particular purpose for carrying out the biological reactions. In one diagnostic application, it may be desirable to simply detect the presence or absence of a particular pathogen. In another diagnostic application, it may be desirable to simply detect the presence or absence of specific allelic variants of pathogens in a clinical sample. For example, different species or subspecies of bacteria may have different susceptibilities to antibiotics; rapid identification of the specific species or subspecies present aids diagnosis and allows initiation of appropriate treatment.

The present invention provides a device, comprising, a) an electroluminescent light source, b) an excitation filter, c) a biological sample holder, and d) an emission

filter, wherein said biological sample holder, is disposed between said excitation filter and said emission filter and said electroluminescent light source is adjacent to said excitation filter so that light produced by said electroluminescent light source passes through said excitation filter to illuminate said biological sample holder. The present invention is not limited to a particular electroluminescent light source. Indeed, a variety of electroluminescent light sources may be incorporated, including, but not limited to a blue, blue-green and green electroluminescent film. Indeed, a variety of emission filters and excitation filters may be incorporated, including, but not limited to Super Gel filters, in any case, the emission filter and excitation filter should be optically compatible with the electroluminescent light source and a target fluorescent molecule. The present invention is not limited to a particular biological sample holder. Indeed, a variety of biological sample holders may be used, including, but not limited to a biological sample holder of the present invention. In one embodiment, the biological sample holder is compatible with a PCR chip. In one embodiment, the biological sample holder is compatible with a microarray chip comprising the probes described above. In one embodiment, the biological sample holder is stationary. In one embodiment, the biological sample holder is mobile.

D. Other Uses of Probes and Amplicons.

In various embodiments, amplified target nucleic acid sequences can be used for any purpose for which nucleic acids are used. Certain exemplary uses for amplification products include, but are not limited to, forensic purposes, genotyping, sequencing, detecting SNPs, detecting microsatellite DNA, detecting expression of genes, quantifying expression of genes, nucleic acid library construction, melting curve analysis, and any other purpose that involves manipulating and/or detecting nucleic acids or nucleic acid sequences.

In certain embodiments, amplification products may be used in any process that uses nucleic acids. Exemplary assays in which amplification products may be used include, but are not limited to, agarose gel electrophoresis, picogreen assays, oligonucleotide ligation assays, and assays described in U.S. Pat. Nos. 5,470,705, 5,514,543, 5,580,732, 5,624,800, 5,807,682, 6,759,202, 6,756,204, 6,734,296, 6,395,486,

U.S. patent application Ser. Nos. 09/584,905 and 09/724,755, and Published U.S. Patent Application No. US 2003-0190646 A1, all of which are herein incorporated by reference.

In certain embodiments, amplification products are treated before they are used in a downstream process. Such treatments include, but are not limited to, heating or enzymatic digestion of amplification products prior to their use in a downstream process.

In certain embodiments, high-throughput assay systems are used. In certain embodiments, a high-throughput assay system includes a plurality of multiplex amplification reactions. In certain such embodiments, the plurality of multiplex amplification reactions is contained on one or more plates or cards, in separate reaction spaces (including, but not limited to, wells or spots). In certain such embodiments, each of the plurality of multiplex amplification reactions amplifies two to five target nucleic acid sequences of similar abundance. In certain such embodiments, each of the plurality of multiplex amplification reactions amplifies more than five target nucleic acid sequences of similar abundance. In certain such embodiments, each of the plurality of multiplex amplification reactions includes a sufficient number of differently-labeled probes such that the amplification product of each target nucleic acid sequence can be separately identified. In certain such embodiments, the amplification reaction proceeds using real-time PCR.

Exemplary high-throughput assay systems include, but are not limited to, an Applied Biosystems plate-reader system (using a plate with any number of wells, including, but not limited to, a 96-well plate, a 384-well plate, a 768-well plate, a 1,536-well plate, a 3,456-well plate, a 6,144-well plate, and a plate with 30,000 or more wells), the ABI 7900 Micro Fluidic Card system (using a card with any number of wells, including, but not limited to, a 384-well card), other microfluidic systems that exploit the use of TaqMan probes (including, but not limited to, systems described in WO 04083443 A1, and published U.S. Patent Application Nos. 2003-0138829 A1 and 2003-0008308 A1, all of which are herein incorporated by reference), other micro card systems (including, but not limited to, WO04067175 A1, and published U.S. Patent Application Nos. 2004-083443 A1, 2004-0110275 A1, and 2004-0121364 A1, all of which are herein incorporated by reference), the Invader® system (Third Wave Technologies), the OpenArray™ system (Biotrove), systems including integrated fluidic circuits (Fluidigm),

and other assay systems known in the art. In certain embodiments, multiple different labels are used in each multiplex amplification reaction in a high-throughput multiplex amplification assay system such that a large number of different target nucleic acid sequences can be analyzed on a single plate or card. In certain embodiments, a high-throughput multiplex amplification assay system is capable of analyzing most of the genes in a genome on a single plate or card. In certain embodiments, a high-throughput multiplex amplification assay system is capable of analyzing all genes in an entire genome on a single plate or card. In certain embodiments, a high-throughput multiplex amplification assay system is capable of analyzing most of the nucleic acids in a transcriptome on a single plate or card. In certain embodiments, a high-throughput multiplex amplification assay system is capable of analyzing all of the nucleic acids in a transcriptome on a single plate or card.

When referring to analyzing most of the genes in a genome by performing one or more amplification reactions, for each gene analyzed, either an entire gene may be amplified or a portion of an entire gene may be amplified. When referring to analyzing all of the genes in a genome by performing one or more amplification reactions, for each gene analyzed, either an entire gene may be amplified or a portion of an entire gene may be amplified. When referring to analyzing most of the nucleic acids in a transcriptome by performing one or more amplification reactions, for each nucleic acid analyzed, either an entire nucleic acid or a portion of an entire nucleic acid may be amplified. When referring to analyzing all of the nucleic acids in a transcriptome by performing one or more amplification reactions, for each nucleic acid analyzed, either an entire nucleic acid or a portion of an entire nucleic acid may be amplified.

EXPERIMENTAL

The following examples serve to illustrate certain embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosures which follow, the following abbreviations apply: N (normal); M (molar); mM (millimolar); μ M (micromolar); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); pmol (picomoles); g (grams); mg (milligrams); μ g (micrograms); ng (nanograms); pg (picograms); L and l (liters); ml

(milliliters); μ l (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); U (units); min (minute); s and sec (second); deg (degree); °C (degrees Centigrade/Celsius), polymerase chain reaction (PCR).

5 Example 1

This Example describes the development of PCT primers and oligonucleotide probes used in a microarray assay for waterborne bacterial pathogens.

10 I. Development of primers and probes and materials and methods for providing one embodiment of a microarray assay for a waterborne bacterial pathogen.

Oligonucleotide probe design and chip synthesis. Two sets of probes were designed and included on a test microarray.

15 A first set of probes ($n = 791$, SEQ ID NOs: 1-791) was designed to detect 35 VMGs covering 12 pathogens, listed herein. A total of 47 PCR primer sets were designed for these genes (Table 1), with each amplicon being targeted by 7 to 35 probes, with an average of 17 probes per amplicon.

20 A second set of probes ($n = 2,034$) targeted 67 VMGs for the 12 pathogens and VMGs for 5 additional pathogens. For these genes, no primers were designed and the probes were analyzed as not-targeted probes to assess the specificity of the probe design.

VMG primers and probes designed and selected in this study were described to detect the selected pathogens in PCR-based assays. Thus, based on 671 sequences retrieved from Genbank (May-April 2004). Probes were selected to satisfy the following criteria i) the probes were complementary to all retrieved sequences of a given gene with species level specificity, and ii) the probes contained at least two mismatches to all other non-target sequences within the database. At least 2,825 probes were selected satisfying these criteria. The Gibbs free energy of probe-target duplex formation assuming complementary targets ($\Delta G^\circ_{\text{duplex}}$) ranged from -14.1 to -23.9 kcal/mole, with a mean of -18.4 and a variability (standard deviation) of 2.1 kcal/mole.

Further, oligonucleotide 18-mer probes were designed for 93 virulence and marker genes (VMGs) of 17 pathogens (Tables 1 and 2) and even further, 50 mer probes were designed, Figure 12.

Each of the 47 VMG regions obtained in this study was targeted by, on average, 17 probes with a range of 7 to 35 probes per region (Table 1).

The designed probes were synthesized *in situ* on microfluidic chips using a proprietary light-directed synthesis technology developed by the University of Michigan (Gao, X.L., et al., Nucleic Acids Research, 2001. 29(22):p. 4744-4750, herein incorporated by reference) and commercialized by Xeotron (Houston, TX, now part of Invitrogen, Carlsbad, CA).

The microfluidic chip contained 8,000 microreactors with a diameter of 50 μ m and were interconnected with flow channels (Fig. 2a). The *in situ* synthesis technology employed conventional phosphoramidite chemistry with photogenerated acid for deprotection of the 5'-hydroxyl groups of nucleotide monomers for the synthesis of oligonucleotide probes. The probes were attached to the substrate via a spacer consisting of Ts and C₁₈ with an effective length of 12 nucleotides. The microarray also contained at least 22 randomly spaced control spots containing solely linker chemistry to assess background signal intensity

DNA extraction from bacterial strains and environmental samples. Genomic DNA from at least 12 reference bacteria was used to validate the VMG biochip. Cultures of *A. hydrophila* (ATCC7966), *Clostridium perfringens* (ATTC12916), *Salmonella* (ATCC13311), *L. monocytogenes* (ATCC15313), *P. aeruginosa* (ATCC10145), *V. parahaemolyticus* (ATCC43996), and *Y. enterocolitica* (ATCC55075) were obtained from the American Type Culture Collection (ATCC, Manassas, VA). Cells were grown overnight according to the provided instructions. Cells were grown overnight according to the instructions provided by ATCC for the respective organisms. Genomic DNA was extracted from 1 to 2 ml of the cultures using the DNeasy Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. For *V. cholerae* (ATCC39315), *C. jejuni* (ATCC700819), *H. pylori* (ATCC700392), *S. aureus* (ATCC700699), and *L. pneumophila* (ATCC33152), purified genomic DNA was obtained from ATCC.

Tap water, river water, and tertiary effluent from a wastewater treatment plant

were collected as described herein. The samples were filtered through 0.45 µm nitrocellulose filters (Millipore, Billerica, MA) upon arrival in the lab. DNA was extracted from the microorganisms accumulated on the filters by cutting the filters into at least 8 pieces and processing the filter pieces using the MegaPrep UltraClean™ Soil DNA Kit (Mo Bio Laboratories, Carlsbad, CA) according to the provided instructions. Extracts were further purified through ethanol precipitation, followed by final purification with the QIAquick PCR purification kit (Qiagen). DNA quantification was performed with NanoDrop® ND-1000 spectrophotometer using a wavelength of 260 nm (NanoDrop Technologies, Wilmington, DE).

Polymerase chain reaction amplification of target genes. Forty seven primer sets were designed to flank regions of high probe density in 35 VMGs (Table 1). Primers were selected to have similar annealing temperature (two sets; one with 53 °C and another with 58 °C) and for covering the majority of alleles of a given VMG available in the GenBank database. Primers were synthesized and obtained from Integrated DNA Technologies (Corallville, IA).

Uniqueness of the primers was confirmed by Blast search against the GenBank database. To ensure primer specificity, mismatches to related non-target sequences were located near the 3' end of the primer. For multiplex PCR, primers were segregated into five primer sets with each set containing nine to ten primer pairs. Combinations of primer sets were selected so that each pathogen was targeted in at least two different multiplex PCR reactions (except for *P. aeruginosa*).

PCR reaction mixtures (25 µl) consisted of 1x PCR buffer, 2 mM MgCl₂, 1.5 (monoplex PCR) or 3 units (multiplex PCR) of AmpliTaq Gold (Roche Molecular Systems, Switzerland), 200 µM deoxynucleoside triphosphates (Invitrogen), 500 nM each primer, 200 ng BSA (New England Biolaboratories, Beverly, MA), and a 1 µl of DNA solution. After initial enzyme activation at 94 °C for 10 min, 35 cycles of the following program was used for amplification: denaturation at 94 °C for 60 s, annealing at 53 °C or 58 °C for 60 s, elongation at 72 °C for 60 s; followed by a final elongation at 72 °C for 7 min. PCR mixtures were purified using a QIAquick PCR purification kit (Qiagen), and amplification products quantified with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies).

Fluorescent labeling of DNA targets (PCR products). PCR products pooled from different PCR reactions were labeled according to an amino-allyl dUTP (aa-dUTP) incorporation and cyanine-dye coupling protocol as described (Stedtfeld, et al., 2007. Appl. Environ. Microbiol. 73:380-389, 58; Wick, et al., Nucleic Acids Res. 34:e26, herein incorporated by reference) with minor modifications. Klenow reactions were incubated for 2 hr instead of 90 min, and dye coupling reactions were incubated for 80 instead of 60 minutes. Briefly, PCR products were amplified and aa-dUTP was incorporated in the amplification products with the Bioprime DNA labeling kit (Invitrogen, San Diego, CA) using a 5:1 aa-dUTP:dTTP ratio and an incubation time of 120 min. Purified products were then coupled with cyanine dye by incubation for 80 min in a 1:1 mixture of 0.1 M sodium carbonate buffer (pH 9.3) and *N*-hydroxysuccinimide ester cyanine dye (Amersham Biosciences).

Microarray hybridizations and melting curve profiles. Hybridizations were performed and melting profiles obtained as described by Wick, L.M., et al., Nucleic Acids Research, 2006. 34(3), herein incorporated by reference. Replicates on one chip were run on consecutive days to minimize storage of used chips. Briefly, after priming of the chips, target DNA (200 pmol of Cy dye) was hybridized overnight at 20 °C using the M-2 hybridization station (Invitrogen, formerly Xeotron Corporation, Houston, TX) with the following hybridization buffer: 35% deionized formamide (Ambion), 6× SSPE (pH 6.6, Invitrogen) and 0.4% Triton X-100 (Sigma). After washing of the chips, the initial point of the melting curve profile was obtained by washing the chip with high stringency wash buffer (20 mM NaCl, 10 mM Na₂PO₄, 5 mM Na₂EDTA, pH adjusted to 6.6 with HCl) for 1.4 min at 25 °C, and imaging of the chip. High stringency wash buffer was degassed under vacuum to prevent formation of air bubbles during the wash steps. Subsequent development of the melting profile was performed by manual cycles of washing and scanning of the chip, repeated at 1 degree intervals until 60 °C was reached. At the end of this series, the chip was additionally stripped for 2.5 min each at 60, 45 and 30 °C with nuclease-free water (Sigma). The tubing of the hybridization station was washed for 20 min before and after each experiment to prevent carry-over between experiments.

Data acquisition. Hybridization signal intensities were obtained from scanned images using GenePix 5.0 (Axon Instruments, Union City, CA). Fluorescence signal intensities were extracted from the scanned images using GenePix5.0 (Axon Instruments, Union City, CA), yielding values between 0 and 65,535 arbitrary units (a.u.). The median of all pixel intensities within a spot was used as raw spot intensity. Subsequent data analysis was done with Microsoft Excel (Microsoft, Redmond, WA), and plotting with SigmaPlot 9.0 (Systat Software, Point Richmond, CA). Raw spot intensities were divided by the mean intensity of 22 empty control spots to yield signal-to-noise ratios (SNRs) (Stedtfeld., et al., Appl. Environ. Microbiol. 73:380-389, herein incorporated by reference). The SNR was computed for each wash step between 30 and 45 °C. An overall SNR was subsequently obtained by averaging all SNR within this interval to include variation in both signal strength and probe-target dissociation behavior. This estimation is equivalent to calculating an overall SNR based on the area under the melting curve within this temperature interval. The SNR for all probes was normalized by dividing by the median SNR of the 2,034 not-targeted probes. The SNR obtained after this normalization step served as the basis for positive/negative calls, with probes yielding a normalized SNR equal or greater than 3 considered positive. In analysis where averaged SNR among replicates were required, SNR were additionally normalized to the replicate with the median SNR (for triplicate experiments), or the replicate with the lowest SNR (for duplicate experiments) using the slope of the linear regression line between replicates as normalization factor. The PF was computed for each probe set by dividing the number of probes with $\text{SNR} \geq 3$ within a set by the number of probes within that set (Wilson et al., Mol. Cell. Probes 16:119-127, herein incorporated by reference).

Preparation of environmental samples, and set-up for hybridization. The power of the VMG chip to detect pathogens was tested three different background samples, tap water, river water, and tertiary effluent from a waste water treatment plant. A mixture of genomic DNA consisting of all 12 pathogens was spiked in DNA extracted from each background. The spike consisted of 0.1% each pathogen (10 pg pathogen DNA, approximately 3,000 genome copies assuming a genome size of 3 MB, in 10 ng background DNA). Spiked samples and unspiked samples were enriched using multiplex PCR assays. The spiked samples were labeled with Cy3 and the unspiked samples were

labeled with Cy5. Both samples were mixed, and hybridized on the same VMG chip in duplicate without a dye swap. The above experimental design was also used to assess DNA from the 12 pathogens at an abundance of 0.01% in DNA extracted from river water and an abundance of 0.001% in DNA extracted from tertiary effluent.

5 **Experimental design.** The specificity and sensitivity of the assays was evaluated using pathogen genomic DNA spiked into DNA obtained from different water sources (tap water, river water, and tertiary effluent from a wastewater treatment plant). A first set of samples was prepared by spiking 10 pg of genomic DNA of each pathogen (equivalent to approximately 1,400 to 5,500 genome copies depending on the pathogen)
10 into 10 ng of DNA from the different water samples, yielding a relative abundance of 0.1% for each of the pathogens. Both spiked and unspiked samples were enriched using the multiplex PCR assays. The amplicons prepared from spiked water samples were labeled with Cy3 and the amplicons from unspiked water samples were labeled with Cy5. Equal amounts of both samples (100 pmol of each dye or ~2.5 µg of DNA) were mixed
15 and hybridized in duplicate. For the unspiked water samples, 2,034 not-targeted probes were evaluated to address assay specificity. For the spiked samples, 673 probes (see Results and Discussion for information on selection of probes) were examined to address assay sensitivity. A second set of samples was prepared to assess the detection limit of the assay. This set contained samples with pathogens spiked at a relative abundance level
20 of 0.01% in river water and 0.001% in tertiary effluent.

Data analysis. All data analysis was performed using Excel (Microsoft, Redmond, WA), Systat 11 and SigmaPlot 9.0 (Systat Software, Point Richmond, CA). Signal-to-noise ratio (SNR) was calculated for each probe using the mean signal intensity of 22 control spots (empty wells) as noise (approximately equal to XXX). For each scan,
25 temperature-dependent SNR were calculated as the ratio of the spot intensity and background signal. An overall SNR was calculated using the area under the melting curve (AUC) between 30 and 45 °C by averaging the SNR measurements in this interval. Each SNR was normalized by dividing individuals SNR by the median SNR of all 2,034 non-targeted probes. A signal was considered positive when its intensity was at least three
30 times the background signal intensity (SNR>3). In cases where averages of SNR among replicates were needed, the SNR were additionally normalized to the replicate with the

median SNR (for triplicate experiments), or the replicate with the lowest SNR (for duplicate experiments).

The metric for presence/absence calls at the VMG level was the positive fraction.

See Wilson, W.J., et al., *Sequence-specific identification of 18 pathogenic*

5 *microorganisms using microarray technology*. Molecular and Cellular Probes, 2002.

16(2): p. 119-127, herein incorporated by reference. The positive fraction was calculated by dividing the number of probes designed for the VMG by the number of probes with positive signal for the VMG. VMGs with positive fractions of 0.5 or higher were scored as present.

10 **Gibbs free energy calculation.** Gibbs free energy of probe-target duplex formation ($\Delta G^{\circ}_{\text{duplex}}$) was calculated with the DINAMelt web server (Markham, et al, DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Research, 2005. 33: p. W577-W581, herein incorporated by reference). This server simulates hybridization behavior of nucleic acids in solution using a general statistical mechanics
15 approach developed by Dimitrov, R.A. and M. Zuker, Prediction of hybridization and melting for double-stranded nucleic acids. Biophysical Journal, 2004. 87(1): p. 215-226, herein incorporated by reference. A temperature of 43°C was used instead of the actual 20°C hybridization temperature to account for the presence of 35% formamide in the hybridization solution. See Blake, R.D. and S.G. Delcourt, Thermodynamic effects of
20 formamide on DNA stability. Nucleic Acids Research, 1996. 24(11): p. 2095-2103; Urakawa, H., et al., Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. Applied and Environmental Microbiology, 2002. 68(1): p. 235-244, all of which are herein incorporated by
reference). The Na⁺ concentration used was 1 M.

25 II. Evaluation of the VMG chip with amplified targets.

Probe success rate was evaluated through hybridization of a target mixture containing all 47 VMG amplicons obtained by individual PCR amplifications. Of the 47 amplicons, 24 were labeled with Cy3 and 23 were labeled with Cy5, Miller et al., An *In-*
30 *situ* Synthesized Virulence and Marker Gene (VMG) Biochip for the Detection of Bacterial Pathogens in Water, in preparation for submission, herein incorporated by

reference).

A SNR of 3 was used as a threshold for positive signals. The total number of probes (2,825) was divided into targeted (720) and non targeted probes (2,105). Twenty four VMG targets were labeled with Cy3 and twenty three VMG targets were labeled with Cy5 (Table 1). Both samples were mixed and hybridized to the VMG chip. Using this experimental set-up, 5,650 possible hybridization events could be analyzed. Evaluations consider probes presented with target (targeted probes) and probes not presented with target (non-targeted probes) separately.

10 **Table 1.** Gene-specific primer sets were designed for 35 targeted VMGs and tested on water samples.

I.D. No.	Pathogen	VMG amplicon	Environmental Samples Present/absent calls and PF in spiked water samples						
			No. of designed probes	No. of probes with S/B ≥ 3 with targets obtained by multiplex PCR	No. of probes with S/B ≥ 3 with targets obtained by multiplex PCR	0.1% tap water	0.1% tertiary effluent	0.1% river water	0.01% river water
33	<i>A. hydrophila</i>	alt	8	6	5	+ (0.60)	+	+	- (0.20)
8		ast(1)	15	13	12	+ (0.71)	+	+	+
1		ast(2)	10	9	7	+ (0.86)	+	+	- (0.07)
36	<i>C. jejuni</i>	cadF	25	21	21	+ (0.67)	+	+	- (0.26)
22		cdtA	14	13	13	+ (0.92)	+	+	+
46		cdtC	31	22	21	+ (0.74)	+	+	+
35		hipO	14	12	12	+ (0.75)	+	+	- (0.38)
39	<i>C. perfringens</i>	cpe(1)	11	9	9	- (0.00)	-	- (0.00)	- (0.11)

						(0.06)			
32		cpe(2)	9	8	7	- (0.29)	- (0.07)	- (0.07)	- (0.00)
37		plc(1)	17	14	14	- (0.04)	- (0.07)	- (0.07)	- (0.14)
47		plc(2)	16	11	11	- (0.09)	- (0.27)	- (0.36)	- (0.18)
23	<i>H. pylori</i>	flaA(1)	18	18	18	+ (0.89)	+ (0.92)	+ (1.00)	+ (0.75)
27		flaA(2)	8	8	7	+ (1.00)	+ (1.00)	+ (1.00)	- (0.21)
26		ureB(1)	11	10	10	+ (0.70)	+ (0.85)	+ (0.75)	+ (0.5)
11		ureB(2)	23	23	23	+ (0.91)	+ (0.87)	+ (0.87)	- (0.28)
31	<i>L. monocytogenes</i>	inlA	19	19	19	+ (0.97)	+ (0.89)	+ (1.00)	- (0.45)
29		inlB	10	9	9	+ (0.89)	+ (0.78)	+ (0.67)	- (0.33)
41		lisO	17	13	11	+ (0.55)	+ (0.73)	+ (0.55)	- (0.27)
45		plcA	32	21	16	+ (0.84)	+ (0.63)	+ (0.56)	- (0.13)
28		plcB	18	15	15	+ (0.87)	+ (0.87)	+ (0.83)	- (0.43)
24	<i>L. pneumophila</i>	dnaJ(1)	17	16	16	+ (0.78)	+ (0.75)	+ (0.78)	- (0.28)
18		dnaJ(2)	17	17	16	+ (0.81)	+ (0.75)	+ (0.75)	+ (0.53)
14		Mip	8	7	7	+ (0.71)	+ (0.57)	+ (0.57)	- (0.43)
13	<i>P. aeruginosa</i>	oprL	18	18	16	+ (1.00)	+ (0.94)	+ (1.00)	- (0.28)
4	<i>Salmonella</i>	fimA	21	21	20	+ (1.00)	+ (0.98)	+ (1.00)	+ (0.93)
2		hilA(1)	12	12	12	+ (1.00)	+ (1.00)	+ (1.00)	+ (0.75)
17		hilA(2)	23	21	21	+ (0.90)	+ (0.95)	+ (0.90)	- (0.45)
6		invA	15	15	12	+ (0.92)	+ (0.92)	+ (0.58)	- (0.33)
9		invE(1)	18	18	17	+ (0.97)	+ (0.91)	+ (0.65)	- (0.38)

30		invE(2)	9	9	9	+	+	+	-
						(0.56)	(0.72)	(0.61)	(0.22)
43	<i>S. aureus</i>	nuc(1)	10	10	9	-	+	+	-
						(0.44)	(0.50)	(0.56)	(0.0)
38		nuc(2)	11	10	10	+	+	+	-
						(0.60)	(0.60)	(0.65)	(0.3)
42		seC	12	9	8	-	+	-	-
						(0.38)	(0.63)	(0.38)	(0.19)
34		tsst(1)	17	17	17	-	-	-	-
						(0.35)	(0.38)	(0.29)	(0.12)
44		tsst(2)	13	10	10	+	+	+	-
						(0.50)	(0.65)	(0.70)	(0.15)
40	<i>V. cholerae</i>	ctxB	20	20	2	x	x	x	x
15		hlyA	15	15	15	+	+	+	-
						(0.94)	(0.97)	(0.94)	(0.47)
5		ompU	18	18	18	+	+	+	-
						(0.79)	(0.82)	(0.53)	(0.21)
16		toxR	35	31	31	+	+	+	-
						(0.92)	(0.90)	(0.85)	(0.40)
3		zot	31	31	31	+	+	+	-
						(0.79)	(0.95)	(0.79)	(0.45)
	<i>V.</i>								
19	<i>parahaemolyticus</i>	tdh	20	20	19	+	+	+	+
						(0.92)	(0.97)	(0.95)	(0.71)
7		tlh(1)	19	19	19	+	+	+	-
						(0.92)	(0.92)	(0.73)	(0.38)
20		tlh(2)	13	13	13	+	+	+	-
						(0.92)	(0.92)	(0.95)	(0.45)
25		toxR(1)	20	20	19	+	+	+	+
						(1.00)	(1.00)	(1.00)	(0.50)
12		toxR(2)	12	12	12	+	+	+	-
						(0.62)	(0.87)	(0.70)	(0.30)
10	<i>Y. enterocolitica</i>	ail	32	30	30	+	+	+	-
						(0.67)	(0.67)	(0.75)	(0.33)
21		ystA	7	7	6	+	+	+	-
						(0.60)	(1.00)	(1.00)	(0.20)
Total			791	720	673				

The experiment described above was performed using 24 amplified regions labeled with Cy3 and 23 amplified regions labeled with Cy5 to better evaluate the specificity and behavior of the targeted probes (Table 1). Of the 393 probes presented with their Cy3 labeled targets, 337 (85.5%) displayed positive signals in at least three replicates. Furthermore 10 out of the 393 (2.5%) probes displayed positive signals in at least one of the replicates with the Cy5 labeled amplicons with SNR always less than 4.4 (median 3.4). Of the 398 probes presented with their Cy5 labeled targets, 383 (96.2%) yielded positive signals in at least three replicates. Furthermore, only 18 of these 383 (4.5%) probes displayed positive signals with the Cy3 labeled non-target sample. Of the 2,034 non-target probes, 27 (1.3%) and 34 (1.7%) probes displayed positive signals with the Cy3 and Cy5 labeled non-target samples, respectively. The SNR of these non-specific hybridization events were considerably smaller than the signals observed for probes presented with targets. SNR for the latter ranged from 3 to 1000 while of the non-target probes only 8 probes yielded SNR higher than 5. The good separation between the signal distributions for targeted and non-targeted probes demonstrates the specificity of the probes (Fig. 2B). Overall, of the 791 targeted probes, 720 (91.0%) produced positive signals with SNR up to 1,000. Of the 2,034 not-targeted probes, 61 probes (3.0%) yielded positive signals with either the Cy3 or Cy5 labeled amplicons. The SNR for targeted and not-targeted probes were well separated, except for a small fraction of the probes (Fig. 2b), indicating the good discriminatory power of the probes. Taken as a whole, out of the 5,650 interrogated individual hybridization reactions, 95% behaved as expected in terms of positive/negative hybridization signal.

A positive fraction (PF) was calculated each probe set (and corresponding amplicon) by dividing the number of positive probes by the number of interrogated probes within that set. A PF obtained based on the number of initially designed probes is equivalent to the success level of probe selection for individual probe sets. Of the 47 probe sets presented with targets, 22 sets displayed a positive fraction of 1, indicating that 100% of the probes in that set yielded positive signal. Eighteen sets displayed a positive fraction between 1 and 0.8, and 7 sets displayed a positive fraction between 0.8 and 0.6 (Fig. 2C). Wilson et al. (Wilson, W.J., et al., Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Molecular and Cellular Probes*,

2002. 16(2): p. 119-127) also observed positive fractions between 0.8 and 1 for 142 probe sets when hybridized with amplified targets. Of the 67 non-targeted probe sets, 34 sets displayed a positive fraction of 0, and 31 sets displayed a positive fraction between 0 and 0.1. The two highest positive fractions for non-targeted probe sets were 0.125 and 0.3.

5 Targeted VMGs displayed a positive fraction above 0.5 while the non-targeted VMGs displayed a positive fraction much lower than 0.5. Accordingly, a positive fraction of 0.5 was adopted as a threshold for presence/absence calls at the gene level. A plot of the number of positive probes vs. the number of designed probes also demonstrates the excellent separation between the positive fractions for targeted and not-targeted probe
10 sets (Fig. 2D). It is also evident that the positive fraction was not related to the number of designed probes. High success rates (>90%) for the design of oligonucleotide probes have been described in numerous recent studies for 16S rRNA genes and sets of functional genes, and demonstrates that current probe design approaches are adequate and successful.

15 Substantial differences in SNR were observed among probes targeting the same VMG (Fig. 3B). For probe sets, differences up to a 100- to 500-fold were observed. When sorting the probe sets according to their median SNR, an overall trend of increasing positive fraction with higher median SNRs becomes apparent (Fig. 3A). The G+C content of the amplified VMG region displayed a similar trend (Fig. 3C). This trend
20 also explained, at least in part, the overall lower success level for the probe sets hybridized with the Cy3 labeled amplicons (85.5%) in comparison with probe sets targeted by the Cy5 labeled amplicons (96.2%). The G+C content for Cy3 labeled amplicons ($36.5 \pm 5.4\%$) was, on average, lower than the G+C content of the Cy5 labeled amplicons ($48.6 \pm 7.9\%$). This also implies that probe design for genes or genomes with a
25 considerably low overall G+C content may yield lower success rates if continuous regions with higher G+C content are not present.

Multiplex amplification of VMG targets. The use of multiplex PCR can greatly increase the efficiency of sample analysis, but it can also introduce bias. This bias was examined by hybridizing targets produced in multiplex PCR and comparing the VMG
30 chip response to that obtained with individually amplified gene regions. Out of 720 probes giving positive signal in response to individually amplified targets, 673 yielded

positive signals in response to multiplex amplified targets (Table 1). Twenty eight of the probe sets responded the same to both individually amplified and multiplex amplified targets. Thirteen probe sets displayed a one probe difference, three sets displayed a two probe difference, one set displayed a three probe difference, and one set displayed a five probe difference between individually and multiplex amplified targets. The probes that did not respond to multiplex targets displayed relatively low SNR in response to individually amplified targets (data not shown). Eighteen of 20 probes for the *ctxB* gene of *V. cholera ctxB* displayed a differential response to multiplex amplified targets. This gene regions was not amplified to an abundance detectable via the VMG chip in multiplex PCR. Only *ctxB* of *V. cholerae* was not amplified to a level detectable by the microarray, and displayed a positive fraction of 0.2 (2 positive out of 20 probes).

Success level of probes as function of $\Delta G^\circ_{\text{duplex}}$. The use of hybridization free energy (ΔG°) was evaluated as a parameter to facilitate *in silico* probe selection. 791 targeted probes were sorted according to $\Delta G^\circ_{\text{duplex}}$, binned (40 probes per bin), and the percent positive probes in each bin calculated. The latter is equivalent to the success level for probes with $\Delta G^\circ_{\text{duplex}}$ within the bin range. A drastic decrease in the success rate was observed for probes with a $\Delta G^\circ_{\text{duplex}}$ less negative than -17 kcal/mole (Fig. 4), and a corresponding G+C content lower than of 34.4%. Similar analysis with variable SNR thresholds displayed analogous trends, but were shifted towards more negative $\Delta G^\circ_{\text{duplex}}$ for increasing SNR thresholds (not shown). The factors contributing the unevenness in probe signals for a given target may explain the lack of hybridization signal for probes with a highly negative $\Delta G^\circ_{\text{duplex}}$. Therefore, -17 kcal/mole might be a useful threshold to increase the success rate of probe selection. Similarly, Loy et al. (Loy, A., et al., 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". Applied and Environmental Microbiology, 2005. 71(3): p. 1373-1386.) proposed a threshold of -16 kcal/mole for 18-mer oligonucleotide probes targeting 16S rRNA genes. Other studies also demonstrated the utility of ΔG° as a probe selection parameter. See Matveeva, O.V., et al., Thermodynamic calculations and statistical correlations for oligo-probes design. Nucleic Acids Research, 2003. 31(14): p. 4211-4217; Luebke, K.J., R.P. Balog, and H.R. Garner, Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to

RNA transcripts. Nucleic Acids Research, 2003. 31(2): p. 750-758. The selection criteria mentioned above were based on hybridization patterns observed with high abundance and low complexity samples. Rules derived from such experiments are not necessarily stringent enough to be applied to work with environmental samples. Rules for such applications are discussed below. Furthermore, small deviations from the proposed optima might occur for other hybridization platforms and experimental conditions.

The results of inventors' experiments display a weak albeit significant correlation between probe hybridization ΔG_0 and SNR (data not shown), as do the studies of Matveeva et al. and Luebke et al. This weak correlation and the large variability within each ΔG_0 bin described above caused some to reject probe hybridization ΔG_0 as a probe selection parameter. See, Pozhitkov, A., et al., Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. Nucleic Acids Research, 2006. 34(9). The inventors' results show that probe hybridization ΔG_0 does have value as a selection parameter, it can be used to predict the behavior of groups of probes with accuracy (Fig. 5). Furthermore, strong hybridization signals were mainly observed by the inventors for probes with a highly negative $\Delta G^\circ_{\text{duplex}}$.

Based on the analysis of responsive and non-responsive probes, a $\Delta G^\circ_{\text{duplex}}$ threshold of -17 kcal/mole serves as a valuable design rule for future oligonucleotide probe (18-mers) screening exercises. This recommendation is in agreement with Loy et al. who proposed a $\Delta G^\circ_{\text{duplex}}$ threshold of -16 kcal/mole for 18-mer probes targeting the 16S rRNA gene. The latter was suggested based on the observation that the majority of probes displaying positive signal (including cross-hybridization signals) were attributed to duplexes with a $\Delta G^\circ_{\text{duplex}}$ more negative than -16 kcal/mole. It should be noted that small deviations from the above $\Delta G^\circ_{\text{duplex}}$ thresholds may be observed for hybridizations performed under different experimental conditions and/or for theoretical $\Delta G^\circ_{\text{duplex}}$ estimates using different nearest neighbor model parameters. Nevertheless, interpreting the -17 kcal/mole $\Delta G^\circ_{\text{duplex}}$ threshold for 18-mers in terms of their inferred G+C content (34.3%) suggested that the derived values might be applicable to most studies. It should be recognized, however, that the above $\Delta G^\circ_{\text{duplex}}$ criterion was derived from hybridization patterns observed with high abundance and low complexity target mixtures. Hence, probe

design rules derived from such hybridizations may not be stringent enough to be applied to nucleic acid target mixtures obtained from environmental samples, containing low abundance target sequences in the presence of excess of non-target sequences.

Performance of the VMG chip with environmental samples. The

5 performance of the VMG biochip probe sets was further validated by hybridization the biochip with PCR amplification products generated from pathogen genomic DNA spiked in total DNA extracted from three different water samples: tap water, tertiary effluent from a wastewater treatment plant, and river water. In the following analysis, PFs for targeted amplicons were calculated based on the 673 probes selected after hybridization
10 of amplicon mixtures from multiplex PCR.

Genomic DNA of 12 pathogens was spiked at an abundance of 0.1% into DNA obtained from three background samples: tap water (drinking water), river water and waste water treatment plant tertiary effluent. Eleven pathogens were detected in all three backgrounds, with all of the gene regions for these organisms indicating presence (Table
15 1). The response of *S. aureus* probes indicated presence in all three backgrounds, although the gene regions were in disagreement between background samples. The gene regions of *C. perfringens* were not detected in any of the background samples. This is likely due to the low G+C content of the *C. perfringens* probe sets.

The median positive fractions of the 46 targeted probe sets were 0.8, 0.82, and
20 0.79 in tap water, tertiary effluent, and river water, respectively (Fig. 6). This precision in three different background samples illustrates the robustness of the chip. The median positive fraction for the 67 non-targeted gene regions were 0.03, 0.09, and 0.18 for the three backgrounds, with tap water resulting in the narrowest distribution of non-targeted fractions and river water resulting in the widest. This could be due to the relative
25 complexity of the background samples, with a greater amount of cross hybridization expected in samples containing greater diversity of target sequence. Using a positive fraction threshold of 0.5, the number of detectable gene regions was 39, 40, and 37, again illustrating the robustness of the chip. Positive fractions presented in the preceding analysis were based on the 673 probes that produced positive signals in response to
30 multiplex amplified targets.

Six pathogens were detected in river water background DNA when spiked at an

abundance of 0.01%. Even though these six organisms were detected, no more than half of the probe sets targeting each organism indicated presence. This illustrates the utility in targeting multiple gene regions in the detection of organisms present at low abundance. These detection limits are comparable to those obtained in studies by Maynard et. al. [15] and Wilson et. al. Maynard et al. (Maynard, C., et al., Waterborne pathogen detection by use of oligonucleotide-based microarrays. Applied and Environmental Microbiology, 2005. 71(12): p. 8548-8557.) were able to detect Salmonella at a relative abundance of 0.1% by targeting. In a study similar to one described here, Wilson et al. was able to detect in a background of DNA obtained from air samples based on a positive fraction threshold of 0.8. Other studies described lower detection limits, albeit in different matrices. Pathogen templates spiked into tertiary effluent background DNA at an abundance of 0.001% resulted in no detection.

Hybridization free energy range associated with best probe selectivity.

Probe selectivity, defined as the difference between the percent for the targeted and non-targeted probes, is presented in Fig. 6 as a function of ΔG_0 . The highest selectivity of 80% was obtained for probes with a ΔG_0 of approximately -19.3 kcal/mole. Probes with a ΔG_0 higher or lower than this value displayed lower selectivity. More than 70% selectivity was obtained for probes with a ΔG_0 between approximately -18.6 and -21.1 kcal/mole, which correlates to a G+C content of $47.2 \pm 5\%$.

Two distinct ΔG_0 regions were evident: region 1 with a ΔG_0 lower than -19 kcal/mole and region 2 with a ΔG_0 higher than -19 kcal/mole. Moving from the lowest ΔG_0 to -19 kcal/mole, the positive bin fraction of the targeted probes is consistent at 0.9 in region 1 while the positive bin fraction of the non-targeted probes decreases with a slope of -0.07. Moving from -19 kcal/mole to higher ΔG_0 , the positive bin fraction of the targeted probes decreases with a slope of -0.17 in region 2 while the positive bin fraction of the non-targeted probes remains relatively constant.

A probe having a more negative ΔG_0 is more likely to give a positive signal, but it is also more prone to cross hybridization with non-targets (Held, G.A., G. Grinstein, and Y. Tu, Modeling of DNA microarray data by using physical properties of hybridization. Proceedings of the National Academy of Sciences of the United States of America, 2003. 100(13): p. 7575-7580 and Fig. 7 top panel). In contrast, a probe having

a more positive ΔG_0 has a low tendency to hybridize with non-targets as well as a lower likelihood to give a positive signal. Starting from the optimal ΔG_0 of -19.3 kcal/mole and moving in the more negative direction, the increase in cross hybridization positive bin fraction is less drastic than the decrease in target hybridization positive bin fraction as you move in the more positive direction. This means that if there is a choice between selecting a probe that has a ΔG_0 of -20 or -18 kcal/mole, the former is more likely to give a higher selectivity. Consequently, it is more difficult to obtain quality probes for and detect lower G+C content regions than higher G+C content regions.

For this analysis, probes were sorted according to their ΔG_0 values, and the fraction of probes with positive signal was calculated for bins of 40 targeted probes and bins of 80 non-targeted probes (Fig. 6 top panel). The hybridizations using tap water as a background resulted in very little non-targeted probe signal (cross-hybridization), so the non-targeted probes from that background were not included in this analysis. The targeted probes from all three background experiments were assessed.

Probe selectivity, defined as the capability of a probe to detect intended targets with the exclusion of non-targets, is the parameter that needs to be optimized during probe selection. Probe selectivity was accordingly calculated as the percent of targeted probes giving positive signal minus the percent of non-targeted probes yielding positive signals. A peak shaped profile was observed when plotting the selectivity as function of ΔG_0 (Fig. 5B). The highest selectivity (approximately 80%) was observed for probes with a ΔG_0 of -19.3 kcal/mole. Probes with a ΔG_0 different from this optimum displayed lower selectivities. Higher than 70% selectivity was observed for probes with a ΔG_0 between -18.6 and -21.1 kcal/mole. Using a relationship between a probe's G+C content and its ΔG_0 , this amounts to a G+C content range between 43.4 and 56.3%, with an optimum at 47.2%. For probes with more negative ΔG_0 values, the decreased selectivity was due to the increased tendency of such probes to cross-hybridize with non-targets (Fig. 4A). For probes with less negative ΔG_0 values, the observed decrease in selectivity was explained by the decrease in percent positive probes for targeted probes (Fig. 4A). Finally, it is interesting to observe that the decrease for probes with less negative ΔG_0 values was more pronounced than that for probes with more negative ΔG_0 values. Two aspects need to be recognized when interpreting the results in Fig. 5. First, the amount of

cross-hybridization depends on the sample that is hybridized. For this analysis, results from river water and tertiary effluent were combined with exclusion of the results from tap water since only minimal cross-hybridization was observed with this background. Second, the drastic decrease in percent positive probes for targeted probes will be less pronounced for more abundant targets.

Hybridization free energy was shown to be a valuable parameter for the selection of promising probes. Probe selectivity was highly depended on ΔG_0 , and a maximum was observed for probes with a ΔG_0 of -19.3 kcal/mole. The capability of theoretical (solution-based) models to predict hybridization signals on microarrays, and their applicability in facilitating probe selection remains in debate. See Loy, A., et al., 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". Applied and Environmental Microbiology, 2005. 71(3): p. 1373-1386; Pozhitkov, A., et al., Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. Nucleic Acids Research, 2006. 34(9); Pozhitkov, A., et al., Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. Nucleic Acids Research, 2006. 34(9), herein incorporated by reference. Although it is recognized that other factors, besides $\Delta G^\circ_{\text{duplex}}$, determine the signal strength of hybridization events, the inventors propose that $\Delta G^\circ_{\text{duplex}}$ is a valuable parameter that can be used to increase the success rate of in silico probe selection.

In conclusion, the inventors developed and validated a coupled format of multiplex PCR and an *in situ* synthesized DNA biochip for the detection of 12 bacterial pathogens in various water samples. By using redundant probe sets targeting various VMGs and inference of presence/absence calls based on the PF, false-positive calls could be eliminated. Pathogens could be detected at a relative abundance level of 0.1% to 0.01%, depending on the pathogen and VMG. Analysis of the hybridization patterns also yielded optimal probe selection criteria for short oligonucleotides. Finally, the relatively high overall success rate of primer and probe design observed in this work, along with implementation of the optimized probe design criteria, indicates that unvalidated VMGs are contemplated for use to detect untested pathogens.

Example 2

This example describes on-chip PCR using primers identified and designed according to the present invention.

5 **DNA Targets.** Genomic DNA from 17 bacterial pathogens (Table 1) was used as the source of virulence and marker genes. *A. hydrophila*, *C. perfringens*, *E. faecalis*, *L. monocytogenes*, *L. pneumophila*, *P. aeruginosa*, *S. enterica*, *V. parahaemolyticus*, *Y. enterocolitica* type strains were obtained from the American Type Culture Collection (ATCC; Manassas, VA) and grown as per the protocol provided. For *H. pylori*, *C. jejuni*,
10 *C. parvum*, *G. intestinalis*, *K. pneumoniae*, *S. aureus*, *V. cholerae*, only genomic DNA was obtained from the ATCC. *E. coli* was kindly provided by Dr. Thomas Whittam at Michigan State University. DNA from pure cultures was extracted using Promega Wizard DNA Extraction Kit (Promega, Madison, WI).

Genomic DNA Background Samples. Tap water (drinking water) (40 L), river
15 water (10 L) from the Red Cedar River in East Lansing (MI), and tertiary effluent (20 L) from a local wastewater treatment plant (East Lansing, MI). Samples were filtered through 0.45 µm nitrocellulose filters (Millex, Billerica, MA) immediately after collection. Genomic DNA was extracted from the filters as instructed with the MegaPrep UltraClean™ Soil DNA Kit (Mo Bio Laboratories, Carlsbad, CA). Genomic DNA was
20 purified further through ethanol precipitation, followed by final purification with the QIAquick PCR purification kit (Qiagen). DNA quantification was performed with the NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). All DNA quantification was performed with the NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE).

25 **PCR Primer Design.** Ninety-six virulence and marker genes (VMG) from 20 human pathogens were targeted. Primer design consisted of the following: i. Kodon (Applied Maths, Austin, TX) was used to generate 96 consensus sequences for each VMG (>4 genes for each organism). ii. Primer Express (Applied Biosystems, Foster City, CA) was used to design primers from the consensus sequence. A majority of
30 designed amplicons had a maximum length of 150 bases and T_m of 59 °C. Some genes required longer amplicons (<250 bases) for generating acceptable primer pairs. iii.

Designed primers were BLASTed against Genbank database using a script that automatically highlighted unspecific primers. Specificity was based on the extent of 3' end perfect matches to other bacterial sequences. Sequences were selected manually based on results of BLAST output. iv. When available, primers described in the literature for successful quantitative PCR were also used. Overall, 110 primers pairs were designed and extracted from the literature.

PCR on BioTrove OpenArray. Primers were synthesized by Sigma-Aldrich (St. Louis, MO) and pre-loaded (128 nm) into BioTrove OpenArray plates (Woburn, MA). Two subarrays (each having 64 wells for 56 assays) were used for each PCR reaction sample. PCR reaction mixtures (10 µl per sample or 5 µl for each sample array) consisted of 1 x LightCycler Fast Start DNA Master SYBR Green I Kit (Roche Applied Sciences, Indianapolis, IN), 1.6 x SYBR Green I, 0.5% Glycerol, 0.2% Pluronic F-68, 1 mg per ml BSA (New England Biolaboratories, Beverly, MA), 2.5 mM MgCl₂, 8% formamide, and a DNA mixture.

Experimental Design. Experimental samples consisted of the following: 1) Genomic DNA from 17 pathogenic organisms (pure cultures) was tested individually (6 ng in total sample or 20 pg per reaction well). 2) Genomic DNA from 4, 8, and 17 pathogenic organisms was mixed and tested simultaneously (20 pg per reaction well per organism). 3) Genomic DNA from 17 pathogens was mixed together and spiked at various concentrations into background genomic DNA from waste water tertiary effluent and river water samples (20 pg, 2 pg, 200 fg, 20 fg, 2 fg of each of 17 pathogens mixed together spiked into 66.6 pg of background genomic DNA per reaction well). Samples were tested in triplicate.

Table 2: List of Pathogens Selected for On-Chip PCR Methods Development.

Genus	Species	Marker genes Targeted and No. of sequences used to make consensus sequence for primer design	Designed Primer pairs	Primer pairs from literature	ATCC Organisms tested with Biotrove
<i>Aeromonas</i>	<i>hydrophila</i> ,	alt, exeF, tapA, arcV	5	-	7966
<i>Burkholderia</i>	<i>mallei</i> , <i>pseudomallei</i>	pilA, pilD, orf13, orf11, SL1631OMP	4	2	-
<i>Campylobacter</i>	<i>jejuni</i> ,	hipO, racR, mapA, gyrA, cdtC, cdtB, cdtA	5	2	700819
<i>Clostridium</i>	<i>perfringens</i>	cpe, plc, pfo, cpbs1, etx	5	-	12916
<i>Enterococcus</i>	<i>faecalis</i> , <i>faecium</i>	cylA, ace, cylLS, esp, gelE	5	-	19433
<i>Escherichia</i>	<i>coli</i>	uidA, stx1, stx2, eae, papG	3	3	from lukas

(including <i>Shigella</i>)					O157:H7
<i>Helicobacter</i>	<i>pylori</i>	cagA, cagE, ureA, virD4	3	2	700392
<i>Klebsiella</i>	<i>pneumoniae</i>	nuc, magA, kvgS, kci, kca	5	-	700721
<i>Legionella</i>	<i>pneumophila</i>	lepB, icmQ, lepA, imp	4	1	33152
<i>Leptospira</i>	<i>interrogans</i>	hlyB, lipA, chpl, lip21, ompL1	4	1	-
<i>Listeria</i>	<i>monocytogenes</i>	actA, hlyA, inlA, mpl, plcB	4	1	15313
<i>Mycobacterium</i>	MAC <i>paratuberculosis</i> , MTB <i>tuberculosis</i> , <i>leprae</i>	erp, glnA1, mmpA, mmpB, IS900	4	1	-
<i>Pseudomonas</i>	<i>aeruginosa</i>	exoS, iasA, pcrV, pilDXcpA, popD	5	-	10145
<i>Salmonella</i>	<i>salmonella</i>	fimA, flicC, fljB, invA	3	1	13311
<i>Staphylococcus</i>	<i>aureus</i>	alphahly, seA, seC, tsst1	3	1	700699
<i>Vibrio</i>	<i>cholerae</i> , <i>mimicus</i> , <i>parahaemolyticus</i> , <i>vulnificus</i>	ace, ompU, tdh, tlh, toxR, ctxA, ctxB, mshA, tcpA, zot,	9	2	39315 43996
<i>Yersinia</i>	<i>enterocolitica</i> , <i>pestis</i> , <i>pseudotuberculosis</i>	crp, yopD, yscD, ystA, bipA	4	2	55075
<i>Cryptosporidium</i>	<i>parvum</i> , <i>hominis</i>	gp40, hsp70, COWP, cp23, cgd7	8	1	PRA-67 genotype II
<i>Giardia</i>	<i>intestinalis</i> , <i>lamblia</i>	B-giardin, VSPH71, VSP4177, VSP4173	5	2	30888 A-1
Total: 20		96	(88 + 22)	=110	17

Example 3.

This example describes developmental stages of microfluidics systems for use in detecting pathogens using PCR primers, 20 mer and 50 mer PCR oligonucleotide probes designed and used according to methods of the present invention. Further, this example demonstrates the use of these oligonucleotide probes in combination with microfluidic and serpentine chips (for example, see, Figures 10 and 11) for PCR reactions, (Hashsham, et al., *Microbe*, Volume 2, Number 11, 2007, herein incorporated by reference) (Figure 14).

Microfluidics-based assays were used for detecting and quantifying infectious agents by hybridizing PCR amplified products onto oligonucleotide probes. For example, the inventors developed and validated a chip (Fig. 11) containing 8,000 microreactors, each with a diameter of 50 microns. Each reactor had oligonucleotide probes synthesized in situ using a low-cost, light-directed DNA synthesis technology. The chip was used to screen 20 different pathogens per run, based on their respective virulence and marker genes.

Presence of each pathogen was confirmed by targeting 3-6 genes per organism. Positive signals were confirmed by hybridizing amplicons for each gene to 5 - 20

standardized probes. (FIG. 11d). This method was found to lower false positives rather than standard approaches that rely on a single marker. (FIG. 14B).

More specifically, the inventors used methods of providing probe sequences, as described herein, for generating 50 mer probes (SEQ ID Nos. 792-11533, Figure 12.

5 These sequences underwent initial evaluations, wherein at least SEQ ID Nos. 11534-13225 and SEQ ID Nos. 13226-19206 were validated (Figure 13) for hybridizing to test pathogen sequences (Figure 14A and 14B).

One of the most challenging tasks of using microfluidics based chips with oligonucleotide probes of the present inventions was sealing of the chip after primer and
10 sample placement inside of the chip because of the small reagent volume which evaporates even after one cycle if leaks are present. The inventors demonstrate a leakproof amplification reaction 15(a) with real time monitoring 15(b). In this experiment, the products were diffused throughout the chip with a relatively low SNR. Presence of the right size of product was confirmed by standard gel electrophoresis 15(c).

15 A key point noted by the inventors was the appearance of the product after the 15th cycle.

Example 4.

This example describes stability of freeze dried Taq polymerase and optimization of Trehalose concentrations for use in compositions and methods of the present
20 inventions.

For field applications of a microarray (PCR) chip comprising primers and probes of the present inventions, the inventors contemplate chips with primers and reagents already dispensed in them. However, this implies that the primers/polymerase/reagents must be made stable at room temperature or even under hot climates. A common practice
25 to obtain freeze-dried reagents is to add sugar (e.g., Trehalose) at the time of freeze-drying. Optimization of the trehalose concentration and stability of the freeze-dried reagents for long periods (6 to 12 months) are two key aspects. A trehalose concentration of 15% has generally been reported as optimal in literature and confirmed in the inventors lab (Figure 4), although lower concentrations seem to work as well. The reagents were
30 stable for at least one month (FIG.16).

Example 5.

This example describes isothermal amplification using a helicase enzyme and primers of the present inventions for use in compositions and methods of the present inventions.

5 Helicase-dependent amplification is isothermal (at around 60 °C) and does not require temperature cycling. The inventors assessed the performance of this enzyme under 21 different conditions that indicated that less than 10 min. was needed for the signals to cross the background threshold. This experiment was conducted at high target concentration (~10,000 copies). Further test are needed to evaluate the detection limit, replication, and primer design. Helicase (BioHelix Corporation, Beverly, Massachusetts, 10 www.biohelix.com/). (FIG.17)

All publications and patents mentioned in the above specification are herein incorporated by reference in their entirety. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art 15 without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in biochemistry, chemistry, molecular biology, 20 microbiology, immunology and water pathogens or related fields are intended to be within the scope of the following claims.

CLAIMS

What is claimed is:

- 5 1. A method providing target oligonucleotide probes for detecting a target pathogen comprising,
 - a) generating a maximum number of nonoverlapping probes of a predetermined length from a gene sequence derived from a target microorganism,
 - b) comparing the sequence of each nonoverlapping probe to sequences from
10 a plurality of database microorganisms,
 - c) identifying at least one subregion in said gene sequence derived from a target microorganism, wherein said at least one subregion contains exact matches to less than a predetermined number of said database microorganisms;
 - d) generating a set of overlapping oligonucleotide probes of a predetermined
15 length for said at least one subregion.
2. The method of claim 1, wherein said predetermined length of said nonoverlapping oligonucleotide probes is from 10 to 60 bases.
- 20 3. The method of claim 1, wherein said subregion is from about 50 to 500 bases in length.
4. The method of claim 1, wherein multiple subregions containing exact matches to less than a predetermined number of database sequences are identified.
25
5. The method of claim 4, wherein overlapping oligonucleotide probes of a predetermined length are generated for said multiple subregions.
6. The method of claim 1, further comprising selecting at least one oligonucleotide
30 probe for said at least one subregion.

7. The method of claim 6, wherein said at least one oligonucleotide probe for said at least one subregion has at least one base pair difference as compared to said sequences of said database microorganisms.

8. The method of claim 7, further comprising the step of arraying said at least one probe on a solid surface.

9. A set of nucleic acid probes generated by the method of claim 7.

10. A microarray comprising the set of nucleic acid probes of claim 9.

11. A microarray comprising nucleic acids probes for a target organism, wherein said probes are selected by generating a maximum number of nonoverlapping oligonucleotide probes of a predetermined length from a gene sequence derived from said target organism, comparing the sequence of each nonoverlapping oligonucleotide probes to sequences from a plurality of database microorganisms, identifying at least one subregion in said gene sequence derived from a target microorganism, wherein said at least one subregion contains exact matches to less than a predetermined number of said database microorganisms; and generating a set of overlapping probes of a predetermined length for said at least one subregion.

12. The microarray of claim 11, wherein said oligonucleotide probes are arrayed on a substrate selected from the group consisting of a glass slide and a serpentine chip.

13. The microarray of Claim 11, wherein said target organisms are selected from the group consisting of *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica*, *E. faecalis*, *S. enterica*, *C. parvum*, *G. intestinalis*, *K. pneumoniae*, *E. coli*. and combinations thereof.

14. The microarray of Claim 13, wherein said oligonucleotide probes are complementary to virulence marker genes from said target organisms.

15. A microarray for high throughput detection of waterborne pathogens comprising sets of multiple discrete nucleic acid probes for at least ten target organisms, wherein said nucleic acid probes are complementary to multiple amplicons from multiple virulence marker genes from each target organism and multiple discrete nucleic acid probes are provided for each amplicon, and wherein said at least 10 target microorganism are selected from the group consisting of *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica*, *E. faecalis*, *S. enterica*, *C. parvum*, *G. intestinalis*, *K. pneumoniae*, and *E. coli*.

16. The microarray of Claim 15, wherein said nucleic acid probes are selected from the group consisting of oligonucleotide probes encoded by SEQ ID Nos. 1-19206 and combinations thereof.

17. A method for high throughput detection of multiple target organisms

a) providing:

i) a microarray comprising sets of multiple discrete nucleic acid probes for multiple target organisms, wherein said nucleic acid probes are complementary to multiple amplicons from multiple target genes of said at least one target organism, and wherein multiple discrete nucleic acid probes are provided for each amplicon;

ii) primers complementary to multiple target genes of said multiple target organisms; and

b) using said primers to amplify said multiple target genes from said multiple target organisms to produce amplicons;

c) contacting said microarray with said amplicons under conditions such that said amplicons hybridize to said discrete nucleic acid probes; and

d) detecting the presence of amplicon binding to said discrete nucleic acid probes.

18. The method of Claim 19, wherein said target organisms are selected from the group consisting of *Aeromonas hydrophila*, *Campylobacter jejuni*, *Clostridium perfringens*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica*, *E. faecalis*, *S. enterica*, *C. parvum*, *G. intestinalis*, *K. pneumoniae*, *E. coli*. and combinations thereof.

19. The method of Claim 19, wherein said probes are complementary to virulence marker genes from said target organisms.

20. The method of Claim 19, wherein said nucleic acid probes are selected from the group consisting of probes encoded by SEQ ID Nos. 1-19206 and combinations thereof.

21. A set of oligonucleotide probes consisting of at least 2 sequences selected from the group consisting of SEQ ID Nos. 1-19206.

22. A microarray consisting of at least 2 sequences selected from the group consisting of SEQ ID Nos. 1-19206.

23. The microarray of claim 22, wherein said microarray is a serpentine chip.

Figure 1

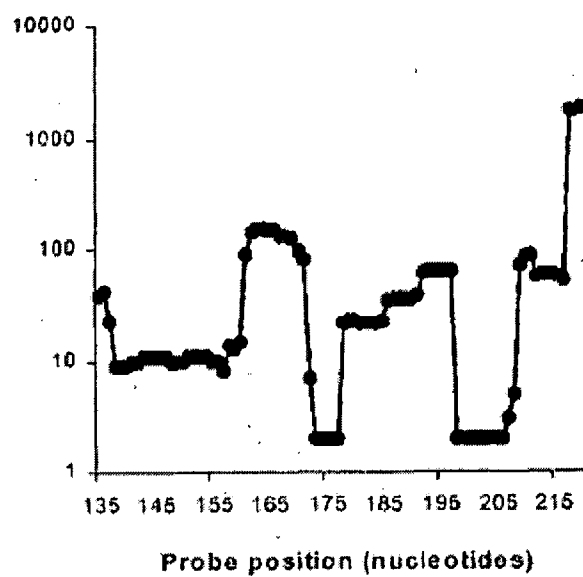
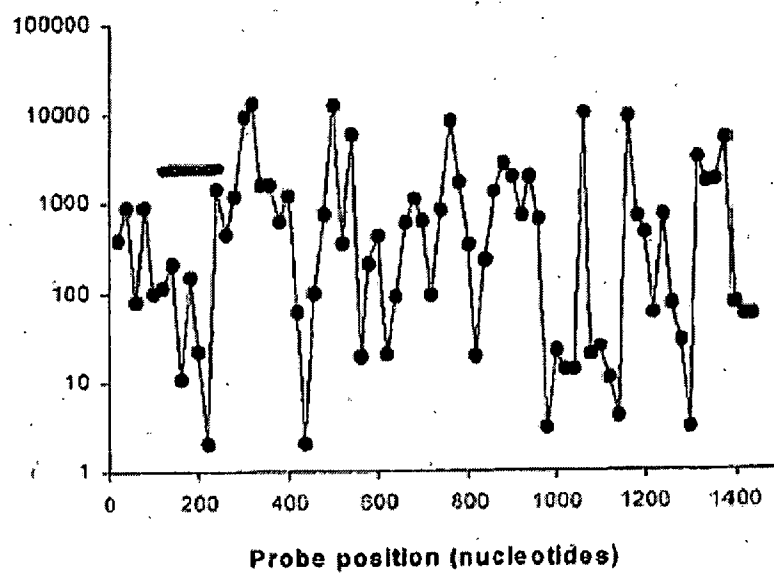


Figure 2

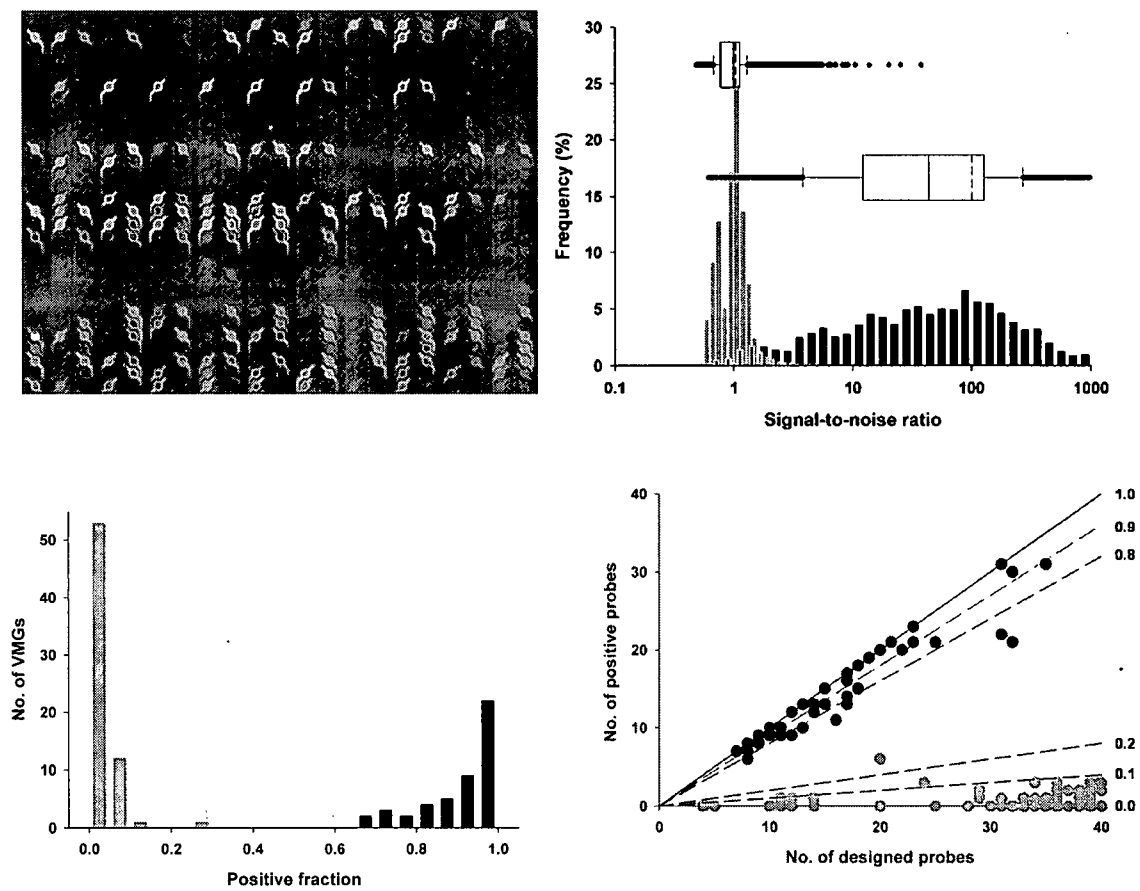


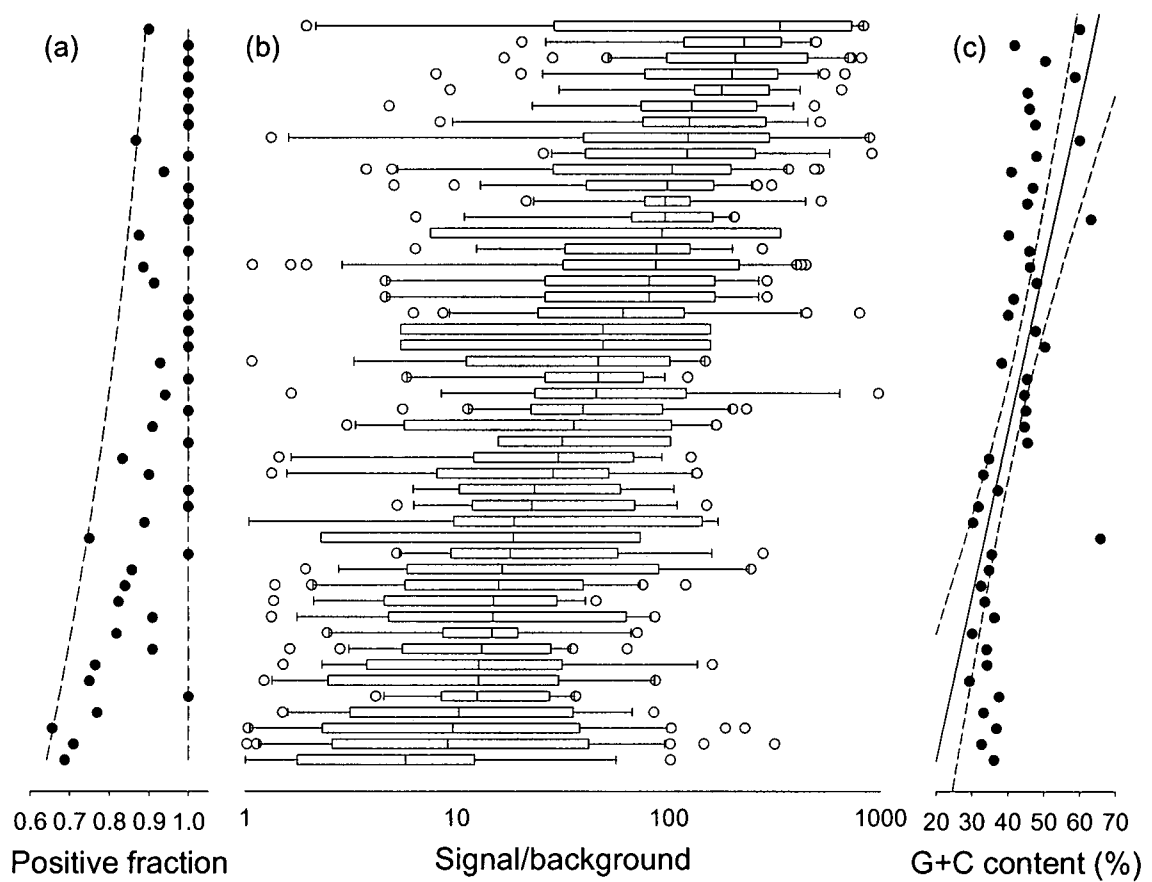
Figure 3

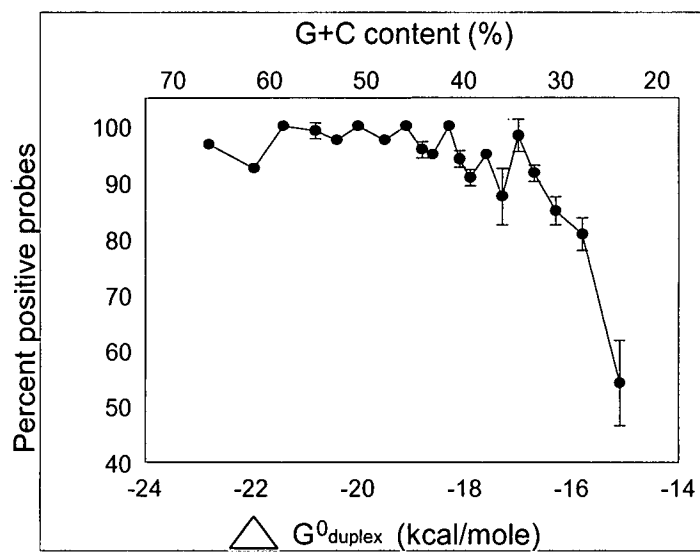
Figure 4

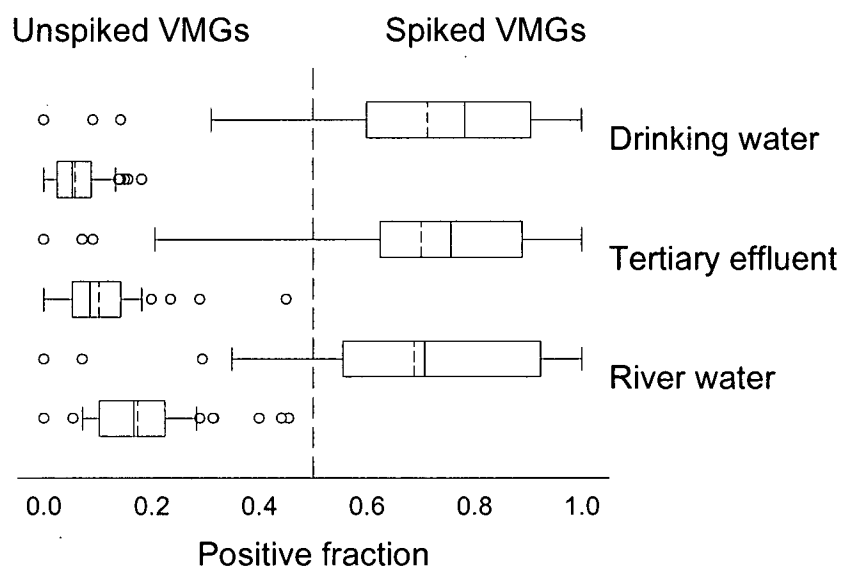
Figure 5

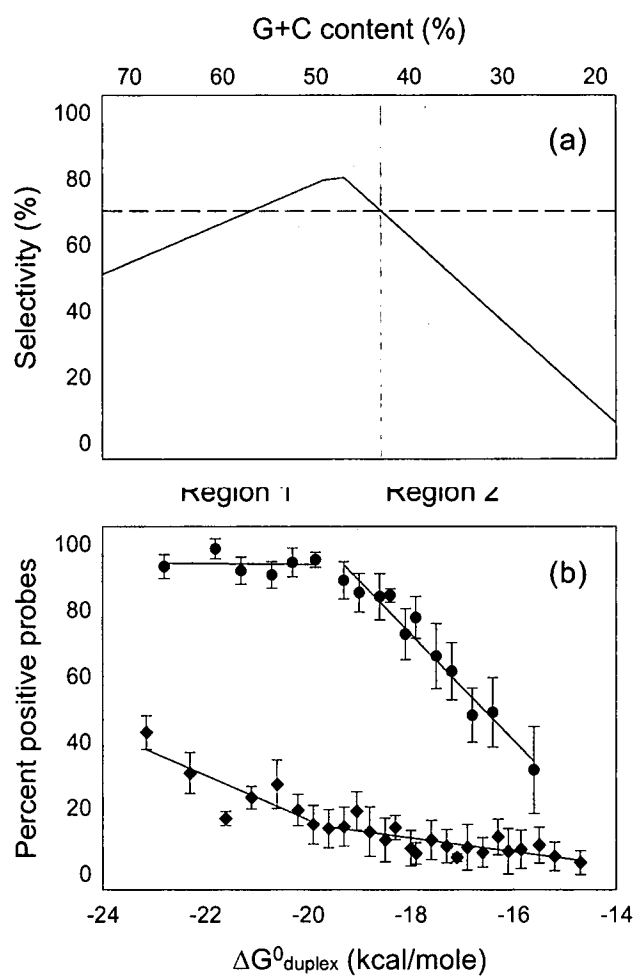
Figure 6

FIGURE 7

Name	Sequence	Length		
AH_alt_PM_22	GGTCGGTCTGGCCAATGG	18	1 SEQ. ID NO	1
AH_alt_PM_23	CGGTCTGGCCAATGGCAA	18	2 SEQ. ID NO	2
AH_alt_PM_26	CTCCGGCACCACCTTGCT	18	3 SEQ. ID NO	3
AH_alt_PM_27	CGGCACCAACTTGCTGGC	18	4 SEQ. ID NO	4
AH_alt_PM_28	GCCAACCAGTCCATCGGC	18	5 SEQ. ID NO	5
AH_alt_PM_29	CCAGTCCATCGGCAATGC	18	6 SEQ. ID NO	6
AH_alt_PM_33	CCAGTGCAAGACTGCGG	18	7 SEQ. ID NO	7
AH_alt_PM_34	GCTGCAAGACTGCGGTGC	18	8 SEQ. ID NO	8
AH_ast_PM_11	GGGCAAGCCGCTCAGCTA	18	9 SEQ. ID NO	9
AH_ast_PM_12	TTTGGCGAGGGTCAGCTC	18	10 SEQ. ID NO	10
AH_ast_PM_14	CAACTCCTGGCTCGACAT	18	11 SEQ. ID NO	11
AH_ast_PM_15	ACTCAAGCGCATCGATCA	18	12 SEQ. ID NO	12
AH_ast_PM_17	GGACGATCGACTGGGCAA	18	13 SEQ. ID NO	13
AH_ast_PM_18	TCAGCAAACCTCAAGAGTG	18	14 SEQ. ID NO	14
AH_ast_PM_19	TGCCAGCTCGGGATCTG	18	15 SEQ. ID NO	15
AH_ast_PM_2	GACCAGTACCGTATCAAG	18	16 SEQ. ID NO	16
AH_ast_PM_20	AAGGGATGAGCGGGTCTC	18	17 SEQ. ID NO	17
AH_ast_PM_22	ACTTCCACCAGAAGGTGA	18	18 SEQ. ID NO	18
AH_ast_PM_24	GAGCCGCTTCACCAAGTGA	18	19 SEQ. ID NO	19
AH_ast_PM_25	GAGCCGCAATCCGAAGG	18	20 SEQ. ID NO	20
AH_ast_PM_26	TGGCTGTTCTTTGCCGAC	18	21 SEQ. ID NO	21
AH_ast_PM_27	GTGATGACATCAACTTCT	18	22 SEQ. ID NO	22
AH_ast_PM_28	GTGCAGCAGTGGATCACC	18	23 SEQ. ID NO	23
AH_ast_PM_29	CCTATCGCTCCATCGTCA	18	24 SEQ. ID NO	24
AH_ast_PM_3	GTTCCGGGAAGTACCTT	18	25 SEQ. ID NO	25
AH_ast_PM_30	TGATGCTGCACGGAAGTGG	18	26 SEQ. ID NO	26
AH_ast_PM_31	CAAGCATCTGGATCGGCA	18	27 SEQ. ID NO	27
AH_ast_PM_34	GCACTCTGGTCACAGCAG	18	28 SEQ. ID NO	28
AH_ast_PM_4	CGCTCAGACAACCTGTTC	18	29 SEQ. ID NO	29
AH_ast_PM_5	GCCGACTTCAAGGTCACC	18	30 SEQ. ID NO	30
AH_ast_PM_6	TCGAGATGGCGTTGGAAA	18	31 SEQ. ID NO	31
AH_ast_PM_8	TCAACGGCTTCTCCATGC	18	32 SEQ. ID NO	32
AH_ast_PM_9	CCATCAACAGCTCGCCCA	18	33 SEQ. ID NO	33
CJ_cadF_PM_10	CTTGTTATCATTTTGAC	18	34 SEQ. ID NO	34
CJ_cadF_PM_11	TTGGCTTGATCAATTAGA	18	35 SEQ. ID NO	35
CJ_cadF_PM_12	ATCAATTAGAATTTGGGT	18	36 SEQ. ID NO	36
CJ_cadF_PM_13	TTAGAGCATTATTCTGAT	18	37 SEQ. ID NO	37
CJ_cadF_PM_15	ACAAGAACTTATTTGAGT	18	38 SEQ. ID NO	38
CJ_cadF_PM_16	TGATGTAGGTGAGAAATT	18	39 SEQ. ID NO	39
CJ_cadF_PM_18	CTATGGTTTAGCAGGTGG	18	40 SEQ. ID NO	40
CJ_cadF_PM_19	TTAGCAGGTGGAGGATAT	18	41 SEQ. ID NO	41
CJ_cadF_PM_20	GTGGAGGATATGAGGATT	18	42 SEQ. ID NO	42
CJ_cadF_PM_21	TATGATAATAAAAGCGGT	18	43 SEQ. ID NO	43
CJ_cadF_PM_22	GTCTTAGTGATTCTTTGG	18	44 SEQ. ID NO	44
CJ_cadF_PM_24	ATGCAAACCATAATTGGG	18	45 SEQ. ID NO	45
CJ_cadF_PM_25	GGGTTTCAACTTTAGGTA	18	46 SEQ. ID NO	46
CJ_cadF_PM_27	GGTATTAGTTTTGGTTTT	18	47 SEQ. ID NO	47
CJ_cadF_PM_28	CTGTAGAAGAAGTTGCTG	18	48 SEQ. ID NO	48
CJ_cadF_PM_29	CAAGCAAATGTCCTGTT	18	49 SEQ. ID NO	49
CJ_cadF_PM_30	GGTGCTTTGTTAGATGAA	18	50 SEQ. ID NO	50
CJ_cadF_PM_31	TATTTCTTTGGAAGGTCA	18	51 SEQ. ID NO	51
CJ_cadF_PM_32	TTGGAAGGTCAATTTGGT	18	52 SEQ. ID NO	52
CJ_cadF_PM_34	AATCCAACTTTTCAAGAA	18	53 SEQ. ID NO	53

FIGURE 7

CJ_cadF_PM_35	TCAAAGAAATTGCAAAAAG	18	54 SEQ. ID NO	54
CJ_cadF_PM_37	ATACTATTCTTGAAGGAC	18	55 SEQ. ID NO	55
CJ_cadF_PM_38	CAAGAGCTTATAATCAAA	18	56 SEQ. ID NO	56
CJ_cadF_PM_39	ATAATCAAAAGCTTTCTG	18	57 SEQ. ID NO	57
CJ_cadF_PM_40	TTTCTGAAAGACGTGCTA	18	58 SEQ. ID NO	58
CJ_cdtA_PM_10	GTGCTGATTTAGTACCTA	18	59 SEQ. ID NO	59
CJ_cdtA_PM_11	GTACCTATTACTCCTATT	18	60 SEQ. ID NO	60
CJ_cdtA_PM_12	TCCTATTACCCACCTTT	18	61 SEQ. ID NO	61
CJ_cdtA_PM_13	ACAAGCAATAGTGCCAAC	18	62 SEQ. ID NO	62
CJ_cdtA_PM_14	AACAATAATGCAGCAAAT	18	63 SEQ. ID NO	63
CJ_cdtA_PM_15	AATGGGATCAATCCTCGC	18	64 SEQ. ID NO	64
CJ_cdtA_PM_16	TCCTCGCTTTAAAGACGA	18	65 SEQ. ID NO	65
CJ_cdtA_PM_17	TGAAAATCGCCCTGCGGT	18	66 SEQ. ID NO	66
CJ_cdtA_PM_18	TTAACCATTTTAGGCCCT	18	67 SEQ. ID NO	67
CJ_cdtA_PM_20	ACGGTTTGGGCTTTAGCA	18	68 SEQ. ID NO	68
CJ_cdtA_PM_5	CGTTCATTTGGAGAATTT	18	69 SEQ. ID NO	69
CJ_cdtA_PM_6	ATACTGATCCTTTAAAC	18	70 SEQ. ID NO	70
CJ_cdtA_PM_7	TTAAAACTAGGACTTGAA	18	71 SEQ. ID NO	71
CJ_cdtA_PM_8	ACTTGAACCTACTTTTCC	18	72 SEQ. ID NO	72
CJ_cdtC_PM_10	GGGATACCTCAAAAGATC	18	73 SEQ. ID NO	73
CJ_cdtC_PM_11	GATCCTATTGATCAAAAT	18	74 SEQ. ID NO	74
CJ_cdtC_PM_15	AGCTGATGAACTTCCTTT	18	75 SEQ. ID NO	75
CJ_cdtC_PM_16	GAACTTCCTTTTGTTAT	18	76 SEQ. ID NO	76
CJ_cdtC_PM_17	TGTGCAATTTACAAATCC	18	77 SEQ. ID NO	77
CJ_cdtC_PM_18	AAAGAAAGCGATCTTTGT	18	78 SEQ. ID NO	78
CJ_cdtC_PM_19	TTTGTTTAGCCATCTTAG	18	79 SEQ. ID NO	79
CJ_cdtC_PM_20	AGCCATCTTAGAAGATGG	18	80 SEQ. ID NO	80
CJ_cdtC_PM_21	TTAGAAGATGGAACCTTT	18	81 SEQ. ID NO	81
CJ_cdtC_PM_22	ATGGAACCTTTGGAGCAA	18	82 SEQ. ID NO	82
CJ_cdtC_PM_23	TGGAGCAAAATCTTGTCA	18	83 SEQ. ID NO	83
CJ_cdtC_PM_24	AAATCTTGTCAAGATGAT	18	84 SEQ. ID NO	84
CJ_cdtC_PM_25	GTCAAGATGATCTAAAAG	18	85 SEQ. ID NO	85
CJ_cdtC_PM_26	GATGGTAAATTAGAAACT	18	86 SEQ. ID NO	86
CJ_cdtC_PM_28	TTTCTATAATGCCAACAA	18	87 SEQ. ID NO	87
CJ_cdtC_PM_29	AACAACAACCTCAGCTGT	18	88 SEQ. ID NO	88
CJ_cdtC_PM_30	ACTTCAGCTGTGCAAATT	18	89 SEQ. ID NO	89
CJ_cdtC_PM_31	CTGTGCAAATTCGTTCTT	18	90 SEQ. ID NO	90
CJ_cdtC_PM_32	CTTTAGTTTTGGAATCTG	18	91 SEQ. ID NO	91
CJ_cdtC_PM_33	TTTGGAATCTGATGAATG	18	92 SEQ. ID NO	92
CJ_cdtC_PM_34	TCTGATGAATGTATAGTA	18	93 SEQ. ID NO	93
CJ_cdtC_PM_36	AAAACGCTTTGGAATAGC	18	94 SEQ. ID NO	94
CJ_cdtC_PM_37	TTTGGAATAGCCCCTTGC	18	95 SEQ. ID NO	95
CJ_cdtC_PM_38	TAGCCCCTTGCAACCCTAG	18	96 SEQ. ID NO	96
CJ_cdtC_PM_39	TTGCACCCTAGATCCTAT	18	97 SEQ. ID NO	97
CJ_cdtC_PM_4	ACTCCTACTGGAGATTTG	18	98 SEQ. ID NO	98
CJ_cdtC_PM_5	CTGGAGATTTGAAAGATT	18	99 SEQ. ID NO	99
CJ_cdtC_PM_6	TTACCGAAATGGTTTCTA	18	100 SEQ. ID NO	100
CJ_cdtC_PM_7	TCTATAAGAAGCTTAGAA	18	101 SEQ. ID NO	101
CJ_cdtC_PM_8	TTTAAGCGCCTTTAGGGA	18	102 SEQ. ID NO	102
CJ_cdtC_PM_9	GCCTTTAGGGATACCTCA	18	103 SEQ. ID NO	103
CJ_hip_PM_11	TCATGATGGACATACTAC	18	104 SEQ. ID NO	104
CJ_hip_PM_12	GACATACTACTTCTTTAT	18	105 SEQ. ID NO	105
CJ_hip_PM_13	CTTTATTTTCAACCTGCT	18	106 SEQ. ID NO	106
CJ_hip_PM_14	AACCTGCTGAAGAGGGTT	18	107 SEQ. ID NO	107

FIGURE 7

CJ_hip_PM_15	TGATAGAAGATGGATTGT	18 108 SEQ. ID NO	108
CJ_hip_PM_17	GGTGCGATGATGGCTTCT	18 109 SEQ. ID NO	109
CJ_hip_PM_18	GATGGCTTCTTCGGATAG	18 110 SEQ. ID NO	110
CJ_hip_PM_3	TAGACTTACAAGGCGAAT	18 111 SEQ. ID NO	111
CJ_hip_PM_4	CGTCATCAAATTCATGAA	18 112 SEQ. ID NO	112
CJ_hip_PM_5	TGAAAATCCTGAGCTTGG	18 113 SEQ. ID NO	113
CJ_hip_PM_6	CTGAGCTTGGTTTGTATG	18 114 SEQ. ID NO	114
CJ_hip_PM_7	AATTTGGTTATGAGGTTT	18 115 SEQ. ID NO	115
CJ_hip_PM_8	TTCGTGCAGATATGGATG	18 116 SEQ. ID NO	116
CJ_hip_PM_9	CAAGAATGCACAAATTTG	18 117 SEQ. ID NO	117
CP_cpe_PM_10	TTAGCCAATCATTAAC	18 118 SEQ. ID NO	118
CP_cpe_PM_13	AATTTATACAAGCATCTG	18 119 SEQ. ID NO	119
CP_cpe_PM_14	CATCTGTAGAATATGGAT	18 120 SEQ. ID NO	120
CP_cpe_PM_15	AATAACTATAGGAGAACA	18 121 SEQ. ID NO	121
CP_cpe_PM_16	ATCTGTATCTACAACTGC	18 122 SEQ. ID NO	122
CP_cpe_PM_26	GGTTCATTAATTGAACT	18 123 SEQ. ID NO	123
CP_cpe_PM_27	CATCTACAGATATAGAAA	18 124 SEQ. ID NO	124
CP_cpe_PM_28	GCTGCTACAGAAAGATTA	18 125 SEQ. ID NO	125
CP_cpe_PM_3	CTTATTTCTGAGGATTTA	18 126 SEQ. ID NO	126
CP_cpe_PM_30	AAACTCAAATCCAGCTGG	18 127 SEQ. ID NO	127
CP_cpe_PM_31	GTAATTTATATGATTGGC	18 128 SEQ. ID NO	128
CP_cpe_PM_32	TTGGCGTTCTTCTAACTC	18 129 SEQ. ID NO	129
CP_cpe_PM_33	AACTCATACCCTTGGACT	18 130 SEQ. ID NO	130
CP_cpe_PM_35	ATAGAATCTTAGCTAGCA	18 131 SEQ. ID NO	131
CP_cpe_PM_38	AATTAGAACAGTCCTTAG	18 132 SEQ. ID NO	132
CP_cpe_PM_5	TCAAATTTAAGTGATGGA	18 133 SEQ. ID NO	133
CP_cpe_PM_6	GGAGATGGTTGGATATTA	18 134 SEQ. ID NO	134
CP_cpe_PM_7	ATATTAGGGGAACCCTCA	18 135 SEQ. ID NO	135
CP_cpe_PM_8	CCCTCAGTAGTTTCAAGT	18 136 SEQ. ID NO	136
CP_cpe_PM_9	TCCTAATGAAACAGGTAC	18 137 SEQ. ID NO	137
CP_plc_PM_1	TGATGGAACAGGAACCTCA	18 138 SEQ. ID NO	138
CP_plc_PM_11	GGGGAATCACAAATAAGA	18 139 SEQ. ID NO	139
CP_plc_PM_13	CAGCATTAGCTAGATATG	18 140 SEQ. ID NO	140
CP_plc_PM_14	AGCTAGATATGAATGGCA	18 141 SEQ. ID NO	141
CP_plc_PM_15	TATGAATGGCAAAGAGGA	18 142 SEQ. ID NO	142
CP_plc_PM_16	CAAGCTACATTCTATCTT	18 143 SEQ. ID NO	143
CP_plc_PM_17	CTATGCACTATTTTGGAG	18 144 SEQ. ID NO	144
CP_plc_PM_18	TTTTGGAGATATAGATAC	18 145 SEQ. ID NO	145
CP_plc_PM_19	GATACTCCATATCATCCT	18 146 SEQ. ID NO	146
CP_plc_PM_2	ACAGGAACTCATGCTATG	18 147 SEQ. ID NO	147
CP_plc_PM_20	ATCCTGCTAATGTTACTG	18 148 SEQ. ID NO	148
CP_plc_PM_24	AGGTTTTGCTAAAACAGG	18 149 SEQ. ID NO	149
CP_plc_PM_25	ATATACTATAGTCATGCT	18 150 SEQ. ID NO	150
CP_plc_PM_26	ATAGTCATGCTAGCATGA	18 151 SEQ. ID NO	151
CP_plc_PM_27	TGCTAGCATGAGTCATAG	18 152 SEQ. ID NO	152
CP_plc_PM_28	ATGAGTCATAGTTGGGAT	18 153 SEQ. ID NO	153
CP_plc_PM_29	ATAGTTGGGATGATTGGG	18 154 SEQ. ID NO	154
CP_plc_PM_3	AATGAACCAGAAAGTGTA	18 155 SEQ. ID NO	155
CP_plc_PM_31	AGATTCTTACACGATGTA	18 156 SEQ. ID NO	156
CP_plc_PM_32	TACACGATGTATCAGAGG	18 157 SEQ. ID NO	157
CP_plc_PM_33	AGAGGGTAATGATCCATC	18 158 SEQ. ID NO	158
CP_plc_PM_34	AATGATCCATCAGTTGGA	18 159 SEQ. ID NO	159
CP_plc_PM_35	AACTAGTAGCTTACATAT	18 160 SEQ. ID NO	160
CP_plc_PM_36	AAGATGCTGGAACAGATG	18 161 SEQ. ID NO	161

FIGURE 7

CP_plc_PM_37	TGGAACAGATGACTACAT	18 162 SEQ. ID NO	162
CP_plc_PM_38	GGAATCAAAACAAAGGAT	18 163 SEQ. ID NO	163
CP_plc_PM_39	AGAATGGGAAATGGACAA	18 164 SEQ. ID NO	164
CP_plc_PM_40	GAAATGGACAACCCAGGA	18 165 SEQ. ID NO	165
CP_plc_PM_5	ATTAGGTTCTACTTATCC	18 166 SEQ. ID NO	166
CP_plc_PM_6	GATCATTTCTGGGATCCT	18 167 SEQ. ID NO	167
CP_plc_PM_7	TTCTCAAAGGATAATAGT	18 168 SEQ. ID NO	168
CP_plc_PM_8	AGGATAATAGTTGGTATT	18 169 SEQ. ID NO	169
CP_plc_PM_9	TAGTTGGTATTTAGCTTA	18 170 SEQ. ID NO	170
HP_flmA_PM_1	TTGGGTCAAGCGATTGCC	18 171 SEQ. ID NO	171
HP_flmA_PM_10	ACCGTTAAGGTTAAAGCG	18 172 SEQ. ID NO	172
HP_flmA_PM_11	GTTAAGGTTAAAGCGACT	18 173 SEQ. ID NO	173
HP_flmA_PM_12	GAATCTCGTAAAGCGATT	18 174 SEQ. ID NO	174
HP_flmA_PM_13	CGATTCAATCTGACATCG	18 175 SEQ. ID NO	175
HP_flmA_PM_14	TTCAATCTGACATCGTTC	18 176 SEQ. ID NO	176
HP_flmA_PM_15	AATCTGACATCGTTCGTT	18 177 SEQ. ID NO	177
HP_flmA_PM_16	CTGACATCGTTCGTTTGA	18 178 SEQ. ID NO	178
HP_flmA_PM_17	ACATCGTTCGTTTGATTG	18 179 SEQ. ID NO	179
HP_flmA_PM_18	TCGTTGTTTGATTCAAG	18 180 SEQ. ID NO	180
HP_flmA_PM_2	GTCAAGCGATTGCCAACA	18 181 SEQ. ID NO	181
HP_flmA_PM_3	AAGCGATTGCCAACACGA	18 182 SEQ. ID NO	182
HP_flmA_PM_33	GGCGCTTGAATTTGCGCA	18 183 SEQ. ID NO	183
HP_flmA_PM_34	TGGGATTGAAATCAAAAC	18 184 SEQ. ID NO	184
HP_flmA_PM_35	GCGATTGGTTTTGGGGAA	18 185 SEQ. ID NO	185
HP_flmA_PM_36	ATTGGTTTTGGGGAATCT	18 186 SEQ. ID NO	186
HP_flmA_PM_37	GGTTTTGGGGAATCTCAA	18 187 SEQ. ID NO	187
HP_flmA_PM_38	TTTGGGGAATCTCAAGTG	18 188 SEQ. ID NO	188
HP_flmA_PM_39	GGGGAATCTCAAGTGGCA	18 189 SEQ. ID NO	189
HP_flmA_PM_4	CGATTGCCAACACGAATG	18 190 SEQ. ID NO	190
HP_flmA_PM_40	GAATCTCAAGTGGCAGAA	18 191 SEQ. ID NO	191
HP_flmA_PM_5	TTGCCAACACGAATGACG	18 192 SEQ. ID NO	192
HP_flmA_PM_6	CCAACACGAATGACGGCA	18 193 SEQ. ID NO	193
HP_flmA_PM_7	ACACGAATGACGGCATGG	18 194 SEQ. ID NO	194
HP_flmA_PM_8	CGAATGACGGCATGGGGA	18 195 SEQ. ID NO	195
HP_flmA_PM_9	GTTGCGGATAAGGCTATG	18 196 SEQ. ID NO	196
HP_ureB_PM_10	AGCTTTTGCAAGCGGTGT	18 197 SEQ. ID NO	197
HP_ureB_PM_11	AATGGATGCTCAGAGCGG	18 198 SEQ. ID NO	198
HP_ureB_PM_12	GCTGAAGAATATTCTATG	18 199 SEQ. ID NO	199
HP_ureB_PM_13	TAAAGGTAACGCTTCTAA	18 200 SEQ. ID NO	200
HP_ureB_PM_14	AAATACGATGTGCAAGTC	18 201 SEQ. ID NO	201
HP_ureB_PM_15	GCAAGTCGCTATCCACAC	18 202 SEQ. ID NO	202
HP_ureB_PM_16	TCCACACAGACACTTTGA	18 203 SEQ. ID NO	203
HP_ureB_PM_17	GGACGCACTATGCACACT	18 204 SEQ. ID NO	204
HP_ureB_PM_18	GCACACTTCCACACTGA	18 205 SEQ. ID NO	205
HP_ureB_PM_19	CCACTAACCCCACTATCC	18 206 SEQ. ID NO	206
HP_ureB_PM_20	ACTATCCCTTTCCTGTG	18 207 SEQ. ID NO	207
HP_ureB_PM_21	CACTGTGAATACAGAAGC	18 208 SEQ. ID NO	208
HP_ureB_PM_22	ACCACTTGATAAAAAGCA	18 209 SEQ. ID NO	209
HP_ureB_PM_23	CGCTGATTCAAGGATCCG	18 210 SEQ. ID NO	210
HP_ureB_PM_24	GGATCCGCCCTCAAACCA	18 211 SEQ. ID NO	211
HP_ureB_PM_25	ATTGCGGCTGAAGACACT	18 212 SEQ. ID NO	212
HP_ureB_PM_26	AGACACTTTGCATGACAT	18 213 SEQ. ID NO	213
HP_ureB_PM_27	ATGGGTCGTGTGGGTGAA	18 214 SEQ. ID NO	214
HP_ureB_PM_28	TTGGCAAACAGCTGACAA	18 215 SEQ. ID NO	215

FIGURE 7

HP_ureB_PM_29	TTGGCCGCTTGAAAGAAG	18 216 SEQ. ID NO	216
HP_ureB_PM_3	TTGATCGCTGAAGTAGAA	18 217 SEQ. ID NO	217
HP_ureB_PM_30	CAAACGCTACTTGTCTAA	18 218 SEQ. ID NO	218
HP_ureB_PM_31	TGTCTAAATACACCATTA	18 219 SEQ. ID NO	219
HP_ureB_PM_32	ACCATTAACCCAGCGATC	18 220 SEQ. ID NO	220
HP_ureB_PM_33	CGCTCATGGGATTAGCGA	18 221 SEQ. ID NO	221
HP_ureB_PM_34	AAGTGGGCAAAGTGGCTG	18 222 SEQ. ID NO	222
HP_ureB_PM_35	GAACGCTTCTATCCCTAC	18 223 SEQ. ID NO	223
HP_ureB_PM_36	AATACGATGCAAACATCA	18 224 SEQ. ID NO	224
HP_ureB_PM_4	TAGAACATGACTACACCA	18 225 SEQ. ID NO	225
HP_ureB_PM_5	AACCCTAGCAAAGAAGAA	18 226 SEQ. ID NO	226
HP_ureB_PM_6	ACCGGTATTTATAAAGCG	18 227 SEQ. ID NO	227
HP_ureB_PM_7	AAGACATGCAAGATGGCG	18 228 SEQ. ID NO	228
HP_ureB_PM_8	TATTGACACACACATCCA	18 229 SEQ. ID NO	229
HP_ureB_PM_9	AACAAATCCCTACAGCTT	18 230 SEQ. ID NO	230
LM_hly_PM_10	ATGTGCCGCCAAGAAAAG	18 231 SEQ. ID NO	231
LM_hly_PM_12	TGAATGCAATTTTCGAGCC	18 232 SEQ. ID NO	232
LM_hly_PM_13	GGAATTAGTAGAAAATCA	18 233 SEQ. ID NO	233
LM_hly_PM_14	GCAGTAAATACATTAGTG	18 234 SEQ. ID NO	234
LM_hly_PM_15	AGTGGAAAGATGGAATGA	18 235 SEQ. ID NO	235
LM_hly_PM_16	CAAGCTTATCCAAATGTA	18 236 SEQ. ID NO	236
LM_hly_PM_17	ATTATGATGACGAAATGG	18 237 SEQ. ID NO	237
LM_hly_PM_18	AAAGCTGTAAATAATAGC	18 238 SEQ. ID NO	238
LM_hly_PM_19	AGCTTGAATGTAAACTTC	18 239 SEQ. ID NO	239
LM_hly_PM_20	CGCAATCAGTGAAGGGAA	18 240 SEQ. ID NO	240
LM_hly_PM_21	GAAGAAGTCATTAGTTTT	18 241 SEQ. ID NO	241
LM_hly_PM_23	ATGTTAATGAACCTACAA	18 242 SEQ. ID NO	242
LM_hly_PM_24	ACAAGACCTTCCAGATTT	18 243 SEQ. ID NO	243
LM_hly_PM_5	AAGTCCTAAGACGCCAAT	18 244 SEQ. ID NO	244
LM_hly_PM_6	GAAACACGCGGATGAAAT	18 245 SEQ. ID NO	245
LM_hly_PM_8	GGATTGGATTACAATAAA	18 246 SEQ. ID NO	246
LM_hly_PM_9	TAGTATACCACGGAGATG	18 247 SEQ. ID NO	247
LM_inlA_PM_10	ATATTAGTGCGCTTTCAG	18 248 SEQ. ID NO	248
LM_inlA_PM_11	GCGCTTTCAGGTTTAACT	18 249 SEQ. ID NO	249
LM_inlA_PM_12	TTTGGTAATCAAGTGACA	18 250 SEQ. ID NO	250
LM_inlA_PM_13	TCAAGTGACAGATTTAAA	18 251 SEQ. ID NO	251
LM_inlA_PM_14	GATTTAAAACCATTAGCT	18 252 SEQ. ID NO	252
LM_inlA_PM_15	AACGACTAGATATTTCAA	18 253 SEQ. ID NO	253
LM_inlA_PM_16	TATTAGTGTTCTGGCTAA	18 254 SEQ. ID NO	254
LM_inlA_PM_17	TTCTGGCTAAATTAACCA	18 255 SEQ. ID NO	255
LM_inlA_PM_19	GAAAGTCTTATCGCTACT	18 256 SEQ. ID NO	256
LM_inlA_PM_20	TCGCTACTAACAACCAAA	18 257 SEQ. ID NO	257
LM_inlA_PM_21	TTTTAACAAATTTGGACG	18 258 SEQ. ID NO	258
LM_inlA_PM_22	AATTTGGACGAATTATCC	18 259 SEQ. ID NO	259
LM_inlA_PM_23	CGAATTATCCTTAAATGG	18 260 SEQ. ID NO	260
LM_inlA_PM_24	CCTTAAATGGTAACCACT	18 261 SEQ. ID NO	261
LM_inlA_PM_25	TTGGCGAGTTTAACAAAC	18 262 SEQ. ID NO	262
LM_inlA_PM_6	AACTGGTTTGACTTTGTT	18 263 SEQ. ID NO	263
LM_inlA_PM_7	TGACTTTGTTCAACAATC	18 264 SEQ. ID NO	264
LM_inlA_PM_8	CTAACAAATTTAAATCGG	18 265 SEQ. ID NO	265
LM_inlA_PM_9	GATTAGTGATATTAGTGC	18 266 SEQ. ID NO	266
LM_inlB_PM_10	ACGGATATAACGGTTCTT	18 267 SEQ. ID NO	267
LM_inlB_PM_11	TTGTCTCTCGAAGATAAC	18 268 SEQ. ID NO	268
LM_inlB_PM_12	CCACTTGCAGGTTTAACT	18 269 SEQ. ID NO	269

FIGURE 7

LM_inlB_PM_15	CCAGAAATAATAAGTGAT	18 270 SEQ. ID NO	270
LM_inlB_PM_16	AAATGGCATTACCAGAA	18 271 SEQ. ID NO	271
LM_inlB_PM_17	TCTATCAGCCAGTCACTA	18 272 SEQ. ID NO	272
LM_inlB_PM_18	AGATTTTCATGGGAGAGTA	18 273 SEQ. ID NO	273
LM_inlB_PM_19	ACTGAAAGAGGTTTACAC	18 274 SEQ. ID NO	274
LM_inlB_PM_8	TATAAACGGACTTGTTCA	18 275 SEQ. ID NO	275
LM_inlB_PM_9	AGCTGGAAAGTTTGTATT	18 276 SEQ. ID NO	276
LM_plcA_PM_10	CCAGGTACACATGATACG	18 277 SEQ. ID NO	277
LM_plcA_PM_11	TACGATGAGCTATAACGG	18 278 SEQ. ID NO	278
LM_plcA_PM_12	ACGGAGACATGACGTGGA	18 279 SEQ. ID NO	279
LM_plcA_PM_13	ACCACTGGCTCAAACACA	18 280 SEQ. ID NO	280
LM_plcA_PM_14	CACAAACGATGTCATTGT	18 281 SEQ. ID NO	281
LM_plcA_PM_15	CAACAACTAGAAGCAGGA	18 282 SEQ. ID NO	282
LM_plcA_PM_16	AGGAATACGGTACATCGA	18 283 SEQ. ID NO	283
LM_plcA_PM_17	TCGATATTAGAGCAAAAG	18 284 SEQ. ID NO	284
LM_plcA_PM_19	CTTTCAGGCGTATTAGAA	18 285 SEQ. ID NO	285
LM_plcA_PM_20	AGAAACGATTACTCAATT	18 286 SEQ. ID NO	286
LM_plcA_PM_21	TATGCGTTTAAAAGACGA	18 287 SEQ. ID NO	287
LM_plcA_PM_22	CAAAACAGCAACGATAGT	18 288 SEQ. ID NO	288
LM_plcA_PM_23	CCAACCACTAATCAACAT	18 289 SEQ. ID NO	289
LM_plcA_PM_24	TTACACTACTCCCAGAAC	18 290 SEQ. ID NO	290
LM_plcA_PM_25	GAAGTACACGAGCAATA	18 291 SEQ. ID NO	291
LM_plcA_PM_26	ATCCCTACATTAAGAT	18 292 SEQ. ID NO	292
LM_plcA_PM_27	TATTACTTTCAGAGAACC	18 293 SEQ. ID NO	293
LM_plcA_PM_28	CCATTAGTCATTAATTCA	18 294 SEQ. ID NO	294
LM_plcA_PM_29	GGCGCATCCAACCAAGTA	18 295 SEQ. ID NO	295
LM_plcA_PM_3	GTTTTAGTACTCTGTTTT	18 296 SEQ. ID NO	296
LM_plcA_PM_30	AAGATGACTACAATGGTC	18 297 SEQ. ID NO	297
LM_plcA_PM_31	GGTCCGAGTGTGAAAACA	18 298 SEQ. ID NO	298
LM_plcA_PM_33	TGTCCAGACTGCTTATCA	18 299 SEQ. ID NO	299
LM_plcA_PM_34	AACAACTGTTTCTTAAC	18 300 SEQ. ID NO	300
LM_plcA_PM_35	CATATTAGTGCCACTTCA	18 301 SEQ. ID NO	301
LM_plcA_PM_36	CATTACACCTCGTCAGT	18 302 SEQ. ID NO	302
LM_plcA_PM_4	GCATATTCGCTTAATAAC	18 303 SEQ. ID NO	303
LM_plcA_PM_5	TAAGCCAATAAAGAAGTC	18 304 SEQ. ID NO	304
LM_plcA_PM_6	ACTCTGTAACACAAAGC	18 305 SEQ. ID NO	305
LM_plcA_PM_7	TGGATGTCCGCTCTACCT	18 306 SEQ. ID NO	306
LM_plcA_PM_8	ACCTGACACAACGAGCCT	18 307 SEQ. ID NO	307
LM_plcA_PM_9	GCCTAGCAGCACTCTCTA	18 308 SEQ. ID NO	308
LM_plcB_PM_10	CCGCGGATAACCCGACAA	18 309 SEQ. ID NO	309
LM_plcB_PM_11	ACAAATACTGACGTAAAT	18 310 SEQ. ID NO	310
LM_plcB_PM_12	TAAATACGCACTATTGCG	18 311 SEQ. ID NO	311
LM_plcB_PM_13	GTAAATCATATGCGAGCT	18 312 SEQ. ID NO	312
LM_plcB_PM_14	GAGCTAATTTAATGAATG	18 313 SEQ. ID NO	313
LM_plcB_PM_15	TCGATAAACAAATAGCTC	18 314 SEQ. ID NO	314
LM_plcB_PM_16	AGCTCAAGGAATATATGA	18 315 SEQ. ID NO	315
LM_plcB_PM_17	TATGATGCGGATCATAAA	18 316 SEQ. ID NO	316
LM_plcB_PM_19	TTTTGCTAATGCGAAAT	18 317 SEQ. ID NO	317
LM_plcB_PM_2	TAGTCTTTCCGGTAACGA	18 318 SEQ. ID NO	318
LM_plcB_PM_20	AAAATAACAGGAGCAAAG	18 319 SEQ. ID NO	319
LM_plcB_PM_21	TTTCAATCAATCGGTGAC	18 320 SEQ. ID NO	320
LM_plcB_PM_3	CGATAAAAGCAAATGCCT	18 321 SEQ. ID NO	321
LM_plcB_PM_4	TGCCTGTTGTGATGAATA	18 322 SEQ. ID NO	322
LM_plcB_PM_5	GAATACTTACAAACACCC	18 323 SEQ. ID NO	323

FIGURE 7

LM_plcB_PM_7	GCATGATATTGACAGCAA	18 324 SEQ. ID NO	324
LM_plcB_PM_8	CAAATTACCACATAAACT	18 325 SEQ. ID NO	325
LM_plcB_PM_9	AAACTTAGTTGGTCCGCG	18 326 SEQ. ID NO	326
LP_dnaJ_PM_10	TGGATCCTTCAATGGGTG	18 327 SEQ. ID NO	327
LP_dnaJ_PM_11	GAGGCCCGAGGTGGTTTTG	18 328 SEQ. ID NO	328
LP_dnaJ_PM_12	TGGGCGTCAATCACGTGG	18 329 SEQ. ID NO	329
LP_dnaJ_PM_13	GCGCTGATTTGCAATTTA	18 330 SEQ. ID NO	330
LP_dnaJ_PM_14	GCAATTGACCCCTTGAAGA	18 331 SEQ. ID NO	331
LP_dnaJ_PM_15	ATAGGAAAAGAAGTTGAA	18 332 SEQ. ID NO	332
LP_dnaJ_PM_16	GAAATTACCGTTCCAAGA	18 333 SEQ. ID NO	333
LP_dnaJ_PM_17	AGACATGGTACTTGTACT	18 334 SEQ. ID NO	334
LP_dnaJ_PM_18	TTGTGAAACATGCCAGGG	18 335 SEQ. ID NO	335
LP_dnaJ_PM_19	GGGCATGGGGCAAGTCAG	18 336 SEQ. ID NO	336
LP_dnaJ_PM_20	CAACAAACCTGCCCCACT	18 337 SEQ. ID NO	337
LP_dnaJ_PM_23	TGTCATGGTCAAGGAAGA	18 338 SEQ. ID NO	338
LP_dnaJ_PM_24	AAATGTCAAAATCCCAGGC	18 339 SEQ. ID NO	339
LP_dnaJ_PM_25	GGCTGGAGTCGATAATGG	18 340 SEQ. ID NO	340
LP_dnaJ_PM_26	TGGTGATCGAGTCCGTTT	18 341 SEQ. ID NO	341
LP_dnaJ_PM_27	CGGAGAAGGTGAAGCAGG	18 342 SEQ. ID NO	342
LP_dnaJ_PM_28	AGGGGTGCATGGTGGTGG	18 343 SEQ. ID NO	343
LP_dnaJ_PM_3	CGAAATGCGAGTGATGCA	18 344 SEQ. ID NO	344
LP_dnaJ_PM_30	GCATTGTGAAGTACCTAT	18 345 SEQ. ID NO	345
LP_dnaJ_PM_31	TTTGCAACTGCGGCGTTA	18 346 SEQ. ID NO	346
LP_dnaJ_PM_32	TTAGGAGGCTCGATCGAA	18 347 SEQ. ID NO	347
LP_dnaJ_PM_33	CACTTTGGAAGGCCGAGT	18 348 SEQ. ID NO	348
LP_dnaJ_PM_34	CGGCGGAAACTCAAACAG	18 349 SEQ. ID NO	349
LP_dnaJ_PM_35	TTTCGCTTGAGAAGTAAA	18 350 SEQ. ID NO	350
LP_dnaJ_PM_36	AAAGGGATGAAATCTGTC	18 351 SEQ. ID NO	351
LP_dnaJ_PM_37	GTCAGAGGATATGGACAA	18 352 SEQ. ID NO	352
LP_dnaJ_PM_38	GATTTGCTATGCAAGGTA	18 353 SEQ. ID NO	353
LP_dnaJ_PM_39	AAACACCGGTCAATTTAT	18 354 SEQ. ID NO	354
LP_dnaJ_PM_4	CATATCGCAGGCTGGCAA	18 355 SEQ. ID NO	355
LP_dnaJ_PM_40	GTTACAGGATTCATTAGA	18 356 SEQ. ID NO	356
LP_dnaJ_PM_5	CAATGAAATACCATCCCCG	18 357 SEQ. ID NO	357
LP_dnaJ_PM_6	CCGATCGGAATCCAGGTG	18 358 SEQ. ID NO	358
LP_dnaJ_PM_7	AAGCCTATAACATCTTGT	18 359 SEQ. ID NO	359
LP_dnaJ_PM_8	AACGCGCCGCTTATGATC	18 360 SEQ. ID NO	360
LP_mip_PM_1	GCAATGTCAACAGCAATG	18 361 SEQ. ID NO	361
LP_mip_PM_2	TGTCAACAGCAATGGCTG	18 362 SEQ. ID NO	362
LP_mip_PM_3	AGCAATGGCTGCAACCGA	18 363 SEQ. ID NO	363
LP_mip_PM_4	GACAAGGATAAGTTGTCT	18 364 SEQ. ID NO	364
LP_mip_PM_5	TTGTCTTATAGCATTGGT	18 365 SEQ. ID NO	365
LP_mip_PM_6	AAGGCATGCAAGACGCTA	18 366 SEQ. ID NO	366
LP_mip_PM_7	CAAATGAAAGACGTTCTT	18 367 SEQ. ID NO	367
LP_mip_PM_8	ACTGGTCGTCTGATTGAT	18 368 SEQ. ID NO	368
PA_oprL_PM_14	TCTGCGTGCGATCACCAC	18 369 SEQ. ID NO	369
PA_oprL_PM_15	CGATCACCACCTTCTACT	18 370 SEQ. ID NO	370
PA_oprL_PM_16	ACCTTCTACTTCGAGTAC	18 371 SEQ. ID NO	371
PA_oprL_PM_17	CTTCGAGTACGACAGCTC	18 372 SEQ. ID NO	372
PA_oprL_PM_18	ACGACAGCTCCGACCTGA	18 373 SEQ. ID NO	373
PA_oprL_PM_19	TCCGACCTGAAGCCGGAA	18 374 SEQ. ID NO	374
PA_oprL_PM_20	GAAGCCGGAAGCCATGCG	18 375 SEQ. ID NO	375
PA_oprL_PM_22	CTGGACGTACACGCGAAA	18 376 SEQ. ID NO	376
PA_oprL_PM_23	ACACGCGAAAGACCTGAA	18 377 SEQ. ID NO	377

FIGURE 7

PA_oprL_PM_24	AAGACCTGAAAGGCAGCG	18 378 SEQ. ID NO	378
PA_oprL_PM_26	CGCGTAGTGCTGGAAGGC	18 379 SEQ. ID NO	379
PA_oprL_PM_27	TGGAAGGCCACACCGACG	18 380 SEQ. ID NO	380
PA_oprL_PM_30	GCGAGTACAATATGGCTC	18 381 SEQ. ID NO	381
PA_oprL_PM_33	GCCGTTCCAGCGCTACCTG	18 382 SEQ. ID NO	382
PA_oprL_PM_34	GCGCTACCTGGTGCTGCA	18 383 SEQ. ID NO	383
PA_oprL_PM_36	GGTTTCCTATGGTAAAGA	18 384 SEQ. ID NO	384
PA_oprL_PM_39	GACGAGCAGTCCTGGGCT	18 385 SEQ. ID NO	385
PA_oprL_PM_40	GTCCTGGGCTCAGAACCG	18 386 SEQ. ID NO	386
S_fimA_PM_14	GACGCTGGGTCAATACCG	18 387 SEQ. ID NO	387
S_fimA_PM_15	TGGGTCAATACCGTACCG	18 388 SEQ. ID NO	388
S_fimA_PM_16	CAATACCGTACCGCCAGC	18 389 SEQ. ID NO	389
S_fimA_PM_17	CCGTACCGCCAGCTTTAC	18 390 SEQ. ID NO	390
S_fimA_PM_18	CAGGTGCCTTTCTCCATC	18 391 SEQ. ID NO	391
S_fimA_PM_19	GCCTTTCTCCATCGTCCT	18 392 SEQ. ID NO	392
S_fimA_PM_20	TCTCCATCGTCCTGAATG	18 393 SEQ. ID NO	393
S_fimA_PM_21	ATCGTCCTGAATGACTGC	18 394 SEQ. ID NO	394
S_fimA_PM_22	CCTGAATGACTGCGATCC	18 395 SEQ. ID NO	395
S_fimA_PM_23	CCGTGGCTTTCTCTGGTC	18 396 SEQ. ID NO	396
S_fimA_PM_24	TGGTCAGGCAGATAACAC	18 397 SEQ. ID NO	397
S_fimA_PM_25	CTCTGCGGACAATAGCAC	18 398 SEQ. ID NO	398
S_fimA_PM_27	GTGCGGATTGAGATTCTT	18 399 SEQ. ID NO	399
S_fimA_PM_32	CTGGTTGAAGGCACCAAT	18 400 SEQ. ID NO	400
S_fimA_PM_33	AAGGCACCAATACGCTGC	18 401 SEQ. ID NO	401
S_fimA_PM_34	ACCAATACGCTGCGTTTT	18 402 SEQ. ID NO	402
S_fimA_PM_35	TACGCTGCGTTTTACCGC	18 403 SEQ. ID NO	403
S_fimA_PM_36	TGCGTTTTACCGCACGCT	18 404 SEQ. ID NO	404
S_fimA_PM_37	TTTACCGCACGCTATAAG	18 405 SEQ. ID NO	405
S_fimA_PM_38	CGCACGCTATAAGGCAAC	18 406 SEQ. ID NO	406
S_fimA_PM_40	CCAGGCTAATGCCGACGC	18 407 SEQ. ID NO	407
S_hilA_PM_10	CATTACTCTATCGTGAAG	18 408 SEQ. ID NO	408
S_hilA_PM_11	TATGCGCCCTTTGGCCTG	18 409 SEQ. ID NO	409
S_hilA_PM_12	CCGGTGACCATTACGAAG	18 410 SEQ. ID NO	410
S_hilA_PM_13	AGTGTTAAGGATATTCTT	18 411 SEQ. ID NO	411
S_hilA_PM_14	TTACGCCCCGATTATTAT	18 412 SEQ. ID NO	412
S_hilA_PM_15	CAGATGATACCCGATGGT	18 413 SEQ. ID NO	413
S_hilA_PM_16	CACCAGGAAAGCATTAAAG	18 414 SEQ. ID NO	414
S_hilA_PM_18	ATTCCCGGACTTCGCTGG	18 415 SEQ. ID NO	415
S_hilA_PM_19	GAGTTAAATCAATACACC	18 416 SEQ. ID NO	416
S_hilA_PM_20	GCTGACTCAATGCGTTAA	18 417 SEQ. ID NO	417
S_hilA_PM_21	GCTGGCAGAATGCTACCT	18 418 SEQ. ID NO	418
S_hilA_PM_22	AGAACATGCGATTAAGGC	18 419 SEQ. ID NO	419
S_hilA_PM_23	GGACCACAATAATCCACA	18 420 SEQ. ID NO	420
S_hilA_PM_24	ATTACTGGGGCTAATTAA	18 421 SEQ. ID NO	421
S_hilA_PM_25	CTCAGAATACATCGTCGG	18 422 SEQ. ID NO	422
S_hilA_PM_26	ACTTACTTTTCGCCCATT	18 423 SEQ. ID NO	423
S_hilA_PM_27	GCTGGAATCTTTTCATGG	18 424 SEQ. ID NO	424
S_hilA_PM_28	TGGAGGAGGCCTTACAAA	18 425 SEQ. ID NO	425
S_hilA_PM_29	TAAAATTGGACCCAACGC	18 426 SEQ. ID NO	426
S_hilA_PM_30	AGCTGTGGATTACCTATT	18 427 SEQ. ID NO	427
S_hilA_PM_31	GCTATACGTTTAGGCGAT	18 428 SEQ. ID NO	428
S_hilA_PM_32	ACACCTGCAGGATAATCC	18 429 SEQ. ID NO	429
S_hilA_PM_33	AAGTATGCAGGTTATGTT	18 430 SEQ. ID NO	430
S_hilA_PM_34	CATGAACTGGCACGAAAA	18 431 SEQ. ID NO	431

FIGURE 7

S_hilA_PM_35	GCAGGAAATAACAGGACT	18 432 SEQ. ID NO	432
S_hilA_PM_36	ACGCTGAATATTGTCAGA	18 433 SEQ. ID NO	433
S_hilA_PM_37	GTGCCTTACCGACGATAA	18 434 SEQ. ID NO	434
S_hilA_PM_38	CAGCGTATAGATAATAAT	18 435 SEQ. ID NO	435
S_hilA_PM_39	TTACCGTTAGTGCTGGTT	18 436 SEQ. ID NO	436
S_hilA_PM_40	GCTATTGCCGAGAAAATG	18 437 SEQ. ID NO	437
S_hilA_PM_5	AATCTCTTACCCGCTGTA	18 438 SEQ. ID NO	438
S_hilA_PM_6	AGAGCATCGTTACATTGA	18 439 SEQ. ID NO	439
S_hilA_PM_7	TAATCGTCCGGTCGTAGT	18 440 SEQ. ID NO	440
S_hilA_PM_8	CCTACGACTCATACATTG	18 441 SEQ. ID NO	441
S_hilA_PM_9	CAGGATCAGGTTCAATCC	18 442 SEQ. ID NO	442
S_invA_PM_24	CGGTATTCGGTGGGGAT	18 443 SEQ. ID NO	443
S_invA_PM_25	GATGGTCTTGTGCGCCAG	18 444 SEQ. ID NO	444
S_invA_PM_26	GTCTTGTCGCCCAGATCC	18 445 SEQ. ID NO	445
S_invA_PM_27	CTATTTTAAATTCCGTGA	18 446 SEQ. ID NO	446
S_invA_PM_28	TTTAAATTCCGTGAAGCA	18 447 SEQ. ID NO	447
S_invA_PM_29	AATTCGTGAAGCAAAAC	18 448 SEQ. ID NO	448
S_invA_PM_30	CCGTGAAGCAAAACGTAG	18 449 SEQ. ID NO	449
S_invA_PM_31	GAGCAGCCGCTCAGTATT	18 450 SEQ. ID NO	450
S_invA_PM_32	AGCCGCTCAGTATTGAGG	18 451 SEQ. ID NO	451
S_invA_PM_33	GCTCAGTATTGAGGAAAA	18 452 SEQ. ID NO	452
S_invA_PM_34	TGTTGATTAATGAGATCC	18 453 SEQ. ID NO	453
S_invA_PM_35	GATTAATGAGATCCGTGT	18 454 SEQ. ID NO	454
S_invA_PM_37	AATATTTTCGGTATTCAGG	18 455 SEQ. ID NO	455
S_invA_PM_38	TTTCGGTATTCAGGAAAC	18 456 SEQ. ID NO	456
S_invA_PM_40	AACGTGTTTCCGTGCGTA	18 457 SEQ. ID NO	457
S_invE_PM_10	CGCGTAGCTTATTTCTTG	18 458 SEQ. ID NO	458
S_invE_PM_11	CTTCGTCGTAAAGACCTG	18 459 SEQ. ID NO	459
S_invE_PM_12	TCGTAAAGACCTGGAAGA	18 460 SEQ. ID NO	460
S_invE_PM_13	ACCTGGAAGAGATCGTGC	18 461 SEQ. ID NO	461
S_invE_PM_14	GAAGAGATCGTGCGGAAA	18 462 SEQ. ID NO	462
S_invE_PM_16	CAGCTATCGGCAATTTAT	18 463 SEQ. ID NO	463
S_invE_PM_17	TATTAACCGATATTGACG	18 464 SEQ. ID NO	464
S_invE_PM_18	CTGGAGTTTGGCCAGCTT	18 465 SEQ. ID NO	465
S_invE_PM_19	GTTTGGCCAGCTTTTACG	18 466 SEQ. ID NO	466
S_invE_PM_2	TGATGCGCAACAGGCGGA	18 467 SEQ. ID NO	467
S_invE_PM_20	GCCAGCTTTTACGACGCC	18 468 SEQ. ID NO	468
S_invE_PM_21	ACTGTTTGTGAGTACATT	18 469 SEQ. ID NO	469
S_invE_PM_22	TTGTGAGTACATTGTTGT	18 470 SEQ. ID NO	470
S_invE_PM_28	CAGCCACATGAAGTGGAT	18 471 SEQ. ID NO	471
S_invE_PM_29	GTTAGCCGATATTATAGG	18 472 SEQ. ID NO	472
S_invE_PM_3	AGGCGGAAATACAACAGG	18 473 SEQ. ID NO	473
S_invE_PM_30	CCGATATTATAGGTTTGA	18 474 SEQ. ID NO	474
S_invE_PM_31	TATAGGTTTGAATGCGTT	18 475 SEQ. ID NO	475
S_invE_PM_36	AAGAATATTGGCAGGAAG	18 476 SEQ. ID NO	476
S_invE_PM_37	TATTGGCAGGAAGAATTG	18 477 SEQ. ID NO	477
S_invE_PM_38	GCAGGAAGAATTGTTAAT	18 478 SEQ. ID NO	478
S_invE_PM_39	AAGAATTGTTAATGGCGT	18 479 SEQ. ID NO	479
S_invE_PM_4	AAATTTGTCCAGTCGACG	18 480 SEQ. ID NO	480
S_invE_PM_40	AATGGCGTTACGTAGTAT	18 481 SEQ. ID NO	481
S_invE_PM_6	CCAGTAATTTATCTAACA	18 482 SEQ. ID NO	482
S_invE_PM_8	TTACGTCAGGCGCGTAGC	18 483 SEQ. ID NO	483
S_invE_PM_9	TCAGGCGCGTAGCTTATT	18 484 SEQ. ID NO	484
SA_nuc_PM_10	CGCTACTAGTTGCTTAGT	18 485 SEQ. ID NO	485

FIGURE 7

SA_nuc_PM_12	GTTTCAAGTCTAAGTAGC	18 486 SEQ. ID NO	486
SA_nuc_PM_13	GTCTAAGTAGCTCAGCAA	18 487 SEQ. ID NO	487
SA_nuc_PM_14	TAGCTCAGCAAATGCATC	18 488 SEQ. ID NO	488
SA_nuc_PM_15	GCAAATGCATCACAAACA	18 489 SEQ. ID NO	489
SA_nuc_PM_16	ACAGATAACGGCGTAAAT	18 490 SEQ. ID NO	490
SA_nuc_PM_17	ACGGCGTAAATAGAAGTG	18 491 SEQ. ID NO	491
SA_nuc_PM_18	AAATAGAAGTGGTTCTGA	18 492 SEQ. ID NO	492
SA_nuc_PM_19	ACAGTATATAGTGCAACT	18 493 SEQ. ID NO	493
SA_nuc_PM_29	CACCTGAAACAAAGCATC	18 494 SEQ. ID NO	494
SA_nuc_PM_30	AGGTGTAGAGAAATATGG	18 495 SEQ. ID NO	495
SA_nuc_PM_31	GCAAGTGCATTTACGAAA	18 496 SEQ. ID NO	496
SA_nuc_PM_32	TGAAGTCGAGTTTGACAA	18 497 SEQ. ID NO	497
SA_nuc_PM_33	TAAATATGGACGTGGCTT	18 498 SEQ. ID NO	498
SA_nuc_PM_34	GGACGTGGCTTAGCGTAT	18 499 SEQ. ID NO	499
SA_nuc_PM_35	GCTTAGCGTATATTTATG	18 500 SEQ. ID NO	500
SA_nuc_PM_36	TATATTTATGCTGATGGA	18 501 SEQ. ID NO	501
SA_nuc_PM_37	TAAACGAAGCTTTAGTTC	18 502 SEQ. ID NO	502
SA_nuc_PM_38	AGCTTTAGTTCGTCAAGG	18 503 SEQ. ID NO	503
SA_nuc_PM_39	AAGGCTTGGCTAAAGTTG	18 504 SEQ. ID NO	504
SA_nuc_PM_9	TGTTTCGAAAGGGCAATA	18 505 SEQ. ID NO	505
SA_seC_PM_25	TTTGATAATGGGAACCTTA	18 506 SEQ. ID NO	506
SA_seC_PM_26	AATGGGAACCTTACAAAAT	18 507 SEQ. ID NO	507
SA_seC_PM_30	AAGTGCAAACCTGATAAGA	18 508 SEQ. ID NO	508
SA_seC_PM_31	AGCTCAAGAACTAGACAT	18 509 SEQ. ID NO	509
SA_seC_PM_32	AGAACTAGACATAAAAGC	18 510 SEQ. ID NO	510
SA_seC_PM_33	AGACATAAAAGCTAGGAA	18 511 SEQ. ID NO	511
SA_seC_PM_35	AGTTTAAACAGTTCACCAT	18 512 SEQ. ID NO	512
SA_seC_PM_36	ACAGTTCACCATATGAAA	18 513 SEQ. ID NO	513
SA_seC_PM_37	TTATTGAAAATAACGGCA	18 514 SEQ. ID NO	514
SA_seC_PM_38	AAAATAACGGCAATACTT	18 515 SEQ. ID NO	515
SA_seC_PM_39	AGGCGATAAGTTTGACCA	18 516 SEQ. ID NO	516
SA_seC_PM_40	GATAGAAGTCCACCTTAC	18 517 SEQ. ID NO	517
SA_tsst1_PM_10	GTAGTGGGTCTGACACTT	18 518 SEQ. ID NO	518
SA_tsst1_PM_11	TCTGACACTTTTACAAAT	18 519 SEQ. ID NO.E520	519
SA_tsst1_PM_12	CAAATAGTGAAGTTTATG	18 520 SEQ. ID NO	520
SA_tsst1_PM_13	GAAGTTTATAGATAATTCC	18 521 SEQ. ID NO	521
SA_tsst1_PM_14	ATTCCTTAGGATCTATGC	18 522 SEQ. ID NO	522
SA_tsst1_PM_15	GGATCTATGCGTATAAAA	18 523 SEQ. ID NO	523
SA_tsst1_PM_16	TTCCGAGTCCTTATTAT	18 524 SEQ. ID NO	524
SA_tsst1_PM_17	TTATAGCCCTGCTTTTAC	18 525 SEQ. ID NO	525
SA_tsst1_PM_18	CTGCTTTTACAAAAGGGG	18 526 SEQ. ID NO	526
SA_tsst1_PM_19	AGCCAACATACTAGCGAA	18 527 SEQ. ID NO	527
SA_tsst1_PM_2	AGCCCTTTGTTGCTTGCG	18 528 SEQ. ID NO	528
SA_tsst1_PM_22	CATTTCCAAATAAGTGGC	18 529 SEQ. ID NO	529
SA_tsst1_PM_23	AAGTGGCGTTACAAATAC	18 530 SEQ. ID NO	530
SA_tsst1_PM_24	CCTACTCCAATAGAACTA	18 531 SEQ. ID NO	531
SA_tsst1_PM_25	GAACTACCTTTAAAGTT	18 532 SEQ. ID NO	532
SA_tsst1_PM_26	GTAAAGGTTTCATGGTAAA	18 533 SEQ. ID NO	533
SA_tsst1_PM_27	TCATGGTAAAGATAGCCC	18 534 SEQ. ID NO	534
SA_tsst1_PM_28	AAGATAGCCCCTTAAAGT	18 535 SEQ. ID NO	535
SA_tsst1_PM_3	AGATTTTACCCCTGTTCC	18 536 SEQ. ID NO	536
SA_tsst1_PM_30	TTTAGACTTTGAAATTGCG	18 537 SEQ. ID NO	537
SA_tsst1_PM_31	TTCGTCATCAGCTAACTC	18 538 SEQ. ID NO	538
SA_tsst1_PM_32	AACTCAAATACATGGATT	18 539 SEQ. ID NO	539

FIGURE 7

SA_tsst1_PM_33	ACATGGATTATATCGTTC	18 540 SEQ. ID NO	540
SA_tsst1_PM_34	TATATCGTTCAAGCGATA	18 541 SEQ. ID NO	541
SA_tsst1_PM_35	TCAAGCGATAAAACGGGT	18 542 SEQ. ID NO	542
SA_tsst1_PM_4	CAAATAATCAAAACTGCA	18 543 SEQ. ID NO	543
SA_tsst1_PM_5	AAAGCATCTACAAACGAT	18 544 SEQ. ID NO	544
SA_tsst1_PM_7	TAAAGGATTGCTAGACT	18 545 SEQ. ID NO	545
SA_tsst1_PM_8	TTGCTAGACTGGTATAGT	18 546 SEQ. ID NO	546
SA_tsst1_PM_9	CTGGTATAGTAGTGGGTC	18 547 SEQ. ID NO	547
VC_ctxB_PM_10	TACCACAACACACAAATA	18 548 SEQ. ID NO	548
VC_ctxB_PM_11	CGCTAAATGATAAGATAT	18 549 SEQ. ID NO	549
VC_ctxB_PM_12	GTATACAGAATCTCTAGC	18 550 SEQ. ID NO	550
VC_ctxB_PM_13	ACAGAATCTCTAGCTGGA	18 551 SEQ. ID NO	551
VC_ctxB_PM_14	GAGAGATGGCTATCATTA	18 552 SEQ. ID NO	552
VC_ctxB_PM_15	GATGGCTATCATTACTTT	18 553 SEQ. ID NO	553
VC_ctxB_PM_16	GCTATCATTACTTTTAAG	18 554 SEQ. ID NO	554
VC_ctxB_PM_18	TACTTTTAAGAATGGTGC	18 555 SEQ. ID NO	555
VC_ctxB_PM_19	TTTCAAGTAGAAGTACCA	18 556 SEQ. ID NO	556
VC_ctxB_PM_20	AAGTAGAAGTACCAGGTA	18 557 SEQ. ID NO	557
VC_ctxB_PM_21	AGAAGTACCAGGTAGTCA	18 558 SEQ. ID NO	558
VC_ctxB_PM_22	GTACCAGGTAGTCAACAT	18 559 SEQ. ID NO	559
VC_ctxB_PM_23	CAGGTAGTCAACATATAG	18 560 SEQ. ID NO	560
VC_ctxB_PM_24	AGCGATTGAAAGGATGAA	18 561 SEQ. ID NO	561
VC_ctxB_PM_25	ATTGAAAGGATGAAGGAT	18 562 SEQ. ID NO	562
VC_ctxB_PM_26	AAAGGATGAAGGATACCC	18 563 SEQ. ID NO	563
VC_ctxB_PM_27	GATGAAGGATACCCTGAG	18 564 SEQ. ID NO	564
VC_ctxB_PM_28	AAGGATACCCTGAGGATT	18 565 SEQ. ID NO	565
VC_ctxB_PM_6	TATGCACATGGAACACCT	18 566 SEQ. ID NO	566
VC_ctxB_PM_7	TTGTGTGCAGAATACCAC	18 567 SEQ. ID NO	567
VC_ctxB_PM_8	GTGCAGAATACCACAACA	18 568 SEQ. ID NO	568
VC_ctxB_PM_9	AGAATACCACAACACACA	18 569 SEQ. ID NO	569
VC_hly_PM_10	TCAGAAAGGCTTATGGGG	18 570 SEQ. ID NO	570
VC_hly_PM_12	AAAGGGCAGTCGTCCGTT	18 571 SEQ. ID NO	571
VC_hly_PM_13	GTAAAGATGCCAATATCT	18 572 SEQ. ID NO	572
VC_hly_PM_14	CGTATCAGCCTAGATGAT	18 573 SEQ. ID NO	573
VC_hly_PM_16	TGATCAACTCGGTTATCG	18 574 SEQ. ID NO	574
VC_hly_PM_17	AGTTTGGAGCCAGTTATA	18 575 SEQ. ID NO	575
VC_hly_PM_18	ACGTTAGATGCCTATTTT	18 576 SEQ. ID NO	576
VC_hly_PM_2	GCAATATCCAGTCCAACC	18 577 SEQ. ID NO	577
VC_hly_PM_3	GTTGGCAAATATCAATGA	18 578 SEQ. ID NO	578
VC_hly_PM_4	TGGTGAAGCGGCGGATAT	18 579 SEQ. ID NO	579
VC_hly_PM_5	TTAGTCAAGTCGCTGATA	18 580 SEQ. ID NO	580
VC_hly_PM_6	ACGCATTACCGAGCTTAG	18 581 SEQ. ID NO	581
VC_hly_PM_7	CGCGATTTGGTGATTAAC	18 582 SEQ. ID NO	582
VC_hly_PM_8	GATTTCAAGTCAGATCAGT	18 583 SEQ. ID NO	583
VC_hly_PM_9	AGCAGAGATGCAAGCCCA	18 584 SEQ. ID NO	584
VC_ompU_PM_23	CAAGCTACCGTTTTGCTG	18 585 SEQ. ID NO	585
VC_ompU_PM_24	GACCGTAACGCTGTTGAC	18 586 SEQ. ID NO	586
VC_ompU_PM_25	TAATGTTGTAAGTAAAC	18 587 SEQ. ID NO	587
VC_ompU_PM_26	CTGCCAAGTACTCTGACA	18 588 SEQ. ID NO	588
VC_ompU_PM_27	AACGGTGAAGATGGTTAC	18 589 SEQ. ID NO	589
VC_ompU_PM_28	CTCTCTGTCTGCTATCTA	18 590 SEQ. ID NO	590
VC_ompU_PM_29	ACACTTTCGGTGACACTG	18 591 SEQ. ID NO	591
VC_ompU_PM_30	AGGTGCTGGCTACGCAGA	18 592 SEQ. ID NO	592
VC_ompU_PM_31	TCAAGACGATCAAAACGA	18 593 SEQ. ID NO	593

FIGURE 7

VC_ompU_PM_32	CCGCTTCTTACCGCATGG	18 594 SEQ. ID NO	594
VC_ompU_PM_33	AACCTGTACTTCGCAGGA	18 595 SEQ. ID NO	595
VC_ompU_PM_34	ACTGTTCACCTGACGGTGA	18 596 SEQ. ID NO	596
VC_ompU_PM_35	AGCTAGCGAAAGACGTTG	18 597 SEQ. ID NO	597
VC_ompU_PM_36	GACTACACTGGCTACGAG	18 598 SEQ. ID NO	598
VC_ompU_PM_37	CGACATACAACAATGCAG	18 599 SEQ. ID NO	599
VC_ompU_PM_38	TTTGCTATCGACGCAACT	18 600 SEQ. ID NO	600
VC_ompU_PM_39	TACTACTTCAAGCCAAAC	18 601 SEQ. ID NO	601
VC_ompU_PM_40	CGCTCTTACATCTCTTAC	18 602 SEQ. ID NO	602
VC_toxR_PM_10	AAGTCGATGATTCCAGCT	18 603 SEQ. ID NO	603
VC_toxR_PM_11	GCTTAACCCAAGCCATTT	18 604 SEQ. ID NO	604
VC_toxR_PM_12	TTCGACTCTGCGCAAAAT	18 605 SEQ. ID NO	605
VC_toxR_PM_13	AATGCTCAAAGATTGAC	18 606 SEQ. ID NO	606
VC_toxR_PM_14	AGTCCCCACAATACGTCA	18 607 SEQ. ID NO	607
VC_toxR_PM_15	TCAAAACGGTTCCGAAGC	18 608 SEQ. ID NO	608
VC_toxR_PM_16	AGCGCGGTTACCAATTGA	18 609 SEQ. ID NO	609
VC_toxR_PM_17	TGATCGCCCGAGTGGAAA	18 610 SEQ. ID NO	610
VC_toxR_PM_18	GGTTGAAGAAGAGATGGC	18 611 SEQ. ID NO	611
VC_toxR_PM_19	GGCTCGCGAAAACGAAGC	18 612 SEQ. ID NO	612
VC_toxR_PM_20	CTGCTCATGACATCTCTC	18 613 SEQ. ID NO	613
VC_toxR_PM_21	TCAGCCAGAATCTGTCAA	18 614 SEQ. ID NO	614
VC_toxR_PM_22	ATACGCAGAATCAAGCAG	18 615 SEQ. ID NO	615
VC_toxR_PM_23	CAGTGTGCCTTCATCAGC	18 616 SEQ. ID NO	616
VC_toxR_PM_24	AGCCACTGTAGTGAACAC	18 617 SEQ. ID NO	617
VC_toxR_PM_25	GCAGCCAGCCAATGTCGT	18 618 SEQ. ID NO	618
VC_toxR_PM_26	CGTGGCGAATAAATCGGC	18 619 SEQ. ID NO	619
VC_toxR_PM_27	AAACTTGGGGAATCGACT	18 620 SEQ. ID NO	620
VC_toxR_PM_28	ACTGTTTATTCTGATAGC	18 621 SEQ. ID NO	621
VC_toxR_PM_29	AGCGGTCTTACTTCCCCT	18 622 SEQ. ID NO	622
VC_toxR_PM_30	CCTCGCAGTATTACTGCT	18 623 SEQ. ID NO	623
VC_toxR_PM_31	AACCCAAGCCAATCCAGC	18 624 SEQ. ID NO	624
VC_toxR_PM_32	GCTTTAAACCCCTAACGG	18 625 SEQ. ID NO	625
VC_toxR_PM_33	CGGTTGTGCGATGGCGTAG	18 626 SEQ. ID NO	626
VC_toxR_PM_34	CGTCAATATGCCGAATAA	18 627 SEQ. ID NO	627
VC_toxR_PM_35	TAACCACCCTGATCTTTC	18 628 SEQ. ID NO	628
VC_toxR_PM_36	GGCTACCGTCAATCGAAC	18 629 SEQ. ID NO	629
VC_toxR_PM_37	AACATACTGGTGGACTCA	18 630 SEQ. ID NO	630
VC_toxR_PM_38	ATAACCAGTTAACGCTGA	18 631 SEQ. ID NO	631
VC_toxR_PM_39	AATTACATTCACAGCCCT	18 632 SEQ. ID NO	632
VC_toxR_PM_5	CTCTGATTGACAAAGAAG	18 633 SEQ. ID NO	633
VC_toxR_PM_6	GAGATCATTGATTAGGC	18 634 SEQ. ID NO	634
VC_toxR_PM_7	CCGAATTCTTTGGCTGCT	18 635 SEQ. ID NO	635
VC_toxR_PM_8	AACGTCCAAACGAGGTAA	18 636 SEQ. ID NO	636
VC_toxR_PM_9	CGCAATGATTTGCATGAC	18 637 SEQ. ID NO	637
VC_zot_PM_12	TAAAGGCGCTCGACACGC	18 638 SEQ. ID NO	638
VC_zot_PM_13	GCTTTTGACATGCATCGT	18 639 SEQ. ID NO	639
VC_zot_PM_14	GCATCGTCACCACGGCTG	18 640 SEQ. ID NO	640
VC_zot_PM_15	ACGGCTGGGATATCTGCC	18 641 SEQ. ID NO	641
VC_zot_PM_16	ATCTGCCTAACACGCCT	18 642 SEQ. ID NO	642
VC_zot_PM_17	CACGCCTAACATTGCCAA	18 643 SEQ. ID NO	643
VC_zot_PM_18	TTGCCAAAGTGACAACA	18 644 SEQ. ID NO	644
VC_zot_PM_19	AACATGATAAGAGAGGCG	18 645 SEQ. ID NO	645
VC_zot_PM_20	AGAGGCGGCGGAGATAGG	18 646 SEQ. ID NO	646
VC_zot_PM_21	AGATAGGGTATCGCCAAT	18 647 SEQ. ID NO	647

FIGURE 7

VC_zot_PM_23	GGGGCAAAGTTTACCCTG	18 648 SEQ. ID NO	648
VC_zot_PM_24	TGCAGCCAACTCTGGACA	18 649 SEQ. ID NO	649
VC_zot_PM_25	TGACACGCCAAGTCAAAA	18 650 SEQ. ID NO	650
VC_zot_PM_27	GCTGTGGAAAGACAGAAA	18 651 SEQ. ID NO	651
VC_zot_PM_28	ACAGAAAGATCCTTTTCT	18 652 SEQ. ID NO	652
VC_zot_PM_29	TTTTCTTGTTCCGGCATGG	18 653 SEQ. ID NO	653
VC_zot_PM_30	GGGGAAATGATGCAACTA	18 654 SEQ. ID NO	654
VC_zot_PM_31	GCAACTATCGAGTCAGAG	18 655 SEQ. ID NO	655
VC_zot_PM_32	GTCAGAGCAATCCGAGCC	18 656 SEQ. ID NO	656
VC_zot_PM_33	CCGAGCCTCAGTCAAAGG	18 657 SEQ. ID NO	657
VC_zot_PM_34	AATGCTGTCGGGAGCAAG	18 658 SEQ. ID NO	658
VC_zot_PM_35	TTTGTATTGGTCGGCTTT	18 659 SEQ. ID NO	659
VC_zot_PM_36	CGGCTTTGTGTCCAAGAT	18 660 SEQ. ID NO	660
VC_zot_PM_37	TGATGAGCGTTATCGCCT	18 661 SEQ. ID NO	661
VC_zot_PM_38	ATCGCCTCGTAGACAATT	18 662 SEQ. ID NO	662
VC_zot_PM_4	GGCCGTCACATCATCACG	18 663 SEQ. ID NO	663
VC_zot_PM_5	CATCACGAATGTGCGAGG	18 664 SEQ. ID NO	664
VC_zot_PM_6	GCTTAAACCTTGAACGCA	18 665 SEQ. ID NO	665
VC_zot_PM_7	GACGGTCGCTTAACGATG	18 666 SEQ. ID NO	666
VC_zot_PM_8	GCACTGGGCGAGAAAGGA	18 667 SEQ. ID NO	667
VC_zot_PM_9	GAAAGGACGCGTTTCTCT	18 668 SEQ. ID NO	668
VP_tdh_PM_10	ATGTTTATGGTCAATCAG	18 669 SEQ. ID NO	669
VP_tdh_PM_11	TTTATGGTCAATCAGTAT	18 670 SEQ. ID NO	670
VP_tdh_PM_12	ATGGTCAATCAGTATTCA	18 671 SEQ. ID NO	671
VP_tdh_PM_13	ATCAGTATTCACAACGTC	18 672 SEQ. ID NO	672
VP_tdh_PM_14	AGTATTCACAACGTCAGG	18 673 SEQ. ID NO	673
VP_tdh_PM_15	ATTCACAACGTCAGGTAC	18 674 SEQ. ID NO	674
VP_tdh_PM_16	CACAACGTCAGGTACTAA	18 675 SEQ. ID NO	675
VP_tdh_PM_17	AACGTCAGGTACTAAATG	18 676 SEQ. ID NO	676
VP_tdh_PM_18	GACATCCTACATGACTGT	18 677 SEQ. ID NO	677
VP_tdh_PM_19	ATCCTACATGACTGTGAA	18 678 SEQ. ID NO	678
VP_tdh_PM_20	TACATGACTGTGAACATT	18 679 SEQ. ID NO	679
VP_tdh_PM_21	TGACTGTGAACATTAATG	18 680 SEQ. ID NO	680
VP_tdh_PM_26	ATAAAGACTATACAATGG	18 681 SEQ. ID NO	681
VP_tdh_PM_3	CCATCTGTCCCTTTTCT	18 682 SEQ. ID NO	682
VP_tdh_PM_4	TCTGTCCCTTTTCTGCC	18 683 SEQ. ID NO	683
VP_tdh_PM_5	CCCCGGTTCTGATGAGAT	18 684 SEQ. ID NO	684
VP_tdh_PM_6	CGGTTCTGATGAGATATT	18 685 SEQ. ID NO	685
VP_tdh_PM_7	TCTCTGACTTTTGGACAA	18 686 SEQ. ID NO	686
VP_tdh_PM_8	TTTTGGACAAACCGTAAT	18 687 SEQ. ID NO	687
VP_tdh_PM_9	TGGACAAACCGTAATGTA	18 688 SEQ. ID NO	688
VP_tlh_PM_10	CCGTCAGCGTTGTGAAGC	18 689 SEQ. ID NO	689
VP_tlh_PM_11	AACGCAGACATTACGTTC	18 690 SEQ. ID NO	690
VP_tlh_PM_12	GACAATCGCTTCTCATAC	18 691 SEQ. ID NO	691
VP_tlh_PM_13	CACACGATCTGGAGCAAC	18 692 SEQ. ID NO	692
VP_tlh_PM_14	GCAGCAATGCAGCCAGAT	18 693 SEQ. ID NO	693
VP_tlh_PM_15	GTGGTTGCACTCGGTGAC	18 694 SEQ. ID NO	694
VP_tlh_PM_16	TTGTCTGATACAGGCAAC	18 695 SEQ. ID NO	695
VP_tlh_PM_17	TTTAACGCATCACAAATGG	18 696 SEQ. ID NO	696
VP_tlh_PM_18	TTCCCTAACCCGAACAGC	18 697 SEQ. ID NO	697
VP_tlh_PM_19	TTCTTAGGTCACTTCTCC	18 698 SEQ. ID NO	698
VP_tlh_PM_2	TACTCCCGCTTGCTTCT	18 699 SEQ. ID NO	699
VP_tlh_PM_20	GGTTTTGTGTGGACAGAA	18 700 SEQ. ID NO	700
VP_tlh_PM_25	GCAAACTGGCGAAGAAC	18 701 SEQ. ID NO	701

FIGURE 7

VP_tlh_PM_26	CAGCAAACACCTTGTTTA	18 702 SEQ. ID NO	702
VP_tlh_PM_27	ACTTCATGAACTACAACC	18 703 SEQ. ID NO	703
VP_tlh_PM_28	TGCAGAAGCACTGATTCG	18 704 SEQ. ID NO	704
VP_tlh_PM_3	AAGAGCCAACCTTATCAC	18 705 SEQ. ID NO	705
VP_tlh_PM_30	CTTCATGTTGATGACACT	18 706 SEQ. ID NO	706
VP_tlh_PM_31	GTTTAAGTACTCAACACA	18 707 SEQ. ID NO	707
VP_tlh_PM_32	AGAGATCGACAAAATTCG	18 708 SEQ. ID NO	708
VP_tlh_PM_33	GAAAGTGCTTGAGATGAA	18 709 SEQ. ID NO	709
VP_tlh_PM_34	GATGTA CTACAAAGCGCA	18 710 SEQ. ID NO	710
VP_tlh_PM_35	ACTCACGCCTTGTTTCGAG	18 711 SEQ. ID NO	711
VP_tlh_PM_36	GAGCACGGTTTCGTGAAC	18 712 SEQ. ID NO	712
VP_tlh_PM_37	CCTTGTTTGGACATCAAC	18 713 SEQ. ID NO	713
VP_tlh_PM_38	ATGTACACCCACGCATTG	18 714 SEQ. ID NO	714
VP_tlh_PM_4	TCTGAAGTGATCAGCACG	18 715 SEQ. ID NO	715
VP_tlh_PM_5	CAAACCTATACCTATGTT	18 716 SEQ. ID NO	716
VP_tlh_PM_6	GCACCAGCTACTCGAAAG	18 717 SEQ. ID NO	717
VP_tlh_PM_7	ATGGTAGCTACTTCACCA	18 718 SEQ. ID NO	718
VP_tlh_PM_8	TACTGGTGGAGCTCCGTT	18 719 SEQ. ID NO	719
VP_tlh_PM_9	CTACACCAACACGTCGCA	18 720 SEQ. ID NO	720
VP_toxR_PM_10	GAGAGACCAAACGAAGTT	18 721 SEQ. ID NO	721
VP_toxR_PM_11	AAGTTTTAACCCGTAACG	18 722 SEQ. ID NO	722
VP_toxR_PM_12	ACCCGTAACGAGCTTCAC	18 723 SEQ. ID NO	723
VP_toxR_PM_13	CGAGCTTCACGAGTTTGT	18 724 SEQ. ID NO	724
VP_toxR_PM_14	ACGAGTTTGTTTGGCGTG	18 725 SEQ. ID NO	725
VP_toxR_PM_15	TTGAGGTGGATGACTCAA	18 726 SEQ. ID NO	726
VP_toxR_PM_16	GATGACTCAAGCCTGACT	18 727 SEQ. ID NO	727
VP_toxR_PM_17	AAGCCTGACTCAAGCGAT	18 728 SEQ. ID NO	728
VP_toxR_PM_18	CTCAAGCGATTTCTACTC	18 729 SEQ. ID NO	729
VP_toxR_PM_19	CTACTCTGCGTAAGATGT	18 730 SEQ. ID NO	730
VP_toxR_PM_20	TTCAACCAAATCTCCAGA	18 731 SEQ. ID NO	731
VP_toxR_PM_21	AATCTCCAGAGTTTGTTA	18 732 SEQ. ID NO	732
VP_toxR_PM_22	GAGTTTGTTAAAACCGTT	18 733 SEQ. ID NO	733
VP_toxR_PM_23	TAAAACCGTTCCAAAACG	18 734 SEQ. ID NO	734
VP_toxR_PM_24	TTCCAAAACGAGGCTATC	18 735 SEQ. ID NO	735
VP_toxR_PM_25	CGAGGCTATCAACTCATT	18 736 SEQ. ID NO	736
VP_toxR_PM_26	CATTTGTA CTGTTGAACG	18 737 SEQ. ID NO	737
VP_toxR_PM_27	CTGTTGAACGCCTAAGCC	18 738 SEQ. ID NO	738
VP_toxR_PM_28	GCAATCGTTGAACCAGAA	18 739 SEQ. ID NO	739
VP_toxR_PM_29	AAGCGCCAGTAGTACCTG	18 740 SEQ. ID NO	740
VP_toxR_PM_30	TTTGGCACTATTACTACC	18 741 SEQ. ID NO	741
VP_toxR_PM_31	TACCGATTTGCGTACTGC	18 742 SEQ. ID NO	742
VP_toxR_PM_32	GCGGAATCTCAGTTCCGT	18 743 SEQ. ID NO	743
VP_toxR_PM_33	TCAGTTCCGTCAGATTGG	18 744 SEQ. ID NO	744
VP_toxR_PM_34	GTCAGATTGGTGAGTATC	18 745 SEQ. ID NO	745
VP_toxR_PM_35	CCAGTGATGACACCTGTA	18 746 SEQ. ID NO	746
VP_toxR_PM_36	CCTTCTATTGAGCAGTGC	18 747 SEQ. ID NO	747
VP_toxR_PM_37	TGAGCAGTGCATTGAACG	18 748 SEQ. ID NO	748
VP_toxR_PM_38	CATTGAACGCTACGTTAA	18 749 SEQ. ID NO	749
VP_toxR_PM_39	GCTACGTTAAGCACCATG	18 750 SEQ. ID NO	750
VP_toxR_PM_8	CGAAAGCCGTATACTCCT	18 751 SEQ. ID NO	751
VP_toxR_PM_9	GTATACTCCTGATGTTGG	18 752 SEQ. ID NO	752
YE_ail_PM_10	ATGTAAGAAATGGGT	18 753 SEQ. ID NO	753
YE_ail_PM_11	ATGGGTATACATTGGATA	18 754 SEQ. ID NO	754
YE_ail_PM_12	ATTGGATAATGACCCTAA	18 755 SEQ. ID NO	755

FIGURE 7

YE_ail_PM_13	ACCCTAAAGGTTTTAACC	18 756 SEQ. ID NO	756
YE_ail_PM_14	TTTTAACCTGAAGTACCG	18 757 SEQ. ID NO	757
YE_ail_PM_15	ACCTGAAGTACCGTTATG	18 758 SEQ. ID NO	758
YE_ail_PM_16	AAGTACCGTTATGAACTC	18 759 SEQ. ID NO	759
YE_ail_PM_17	CCGTTATGAACTCGATGA	18 760 SEQ. ID NO	760
YE_ail_PM_18	ATGAACTCGATGATAACT	18 761 SEQ. ID NO	761
YE_ail_PM_19	CTCGATGATAACTGGGGA	18 762 SEQ. ID NO	762
YE_ail_PM_20	TGATAACTGGGGAGTAAT	18 763 SEQ. ID NO	763
YE_ail_PM_21	CTGGGGAGTAATAGGTTT	18 764 SEQ. ID NO	764
YE_ail_PM_22	GAGTAATAGGTTTCGTTTG	18 765 SEQ. ID NO	765
YE_ail_PM_23	AGGTTTCGTTTGCTTATAC	18 766 SEQ. ID NO	766
YE_ail_PM_24	ATGGCAGTAATAAGTTTG	18 767 SEQ. ID NO	767
YE_ail_PM_26	AATGGGGCCATCTTCCG	18 768 SEQ. ID NO	768
YE_ail_PM_27	ACGAATATGTTAGCCTTT	18 769 SEQ. ID NO	769
YE_ail_PM_28	TATGTTAGCCTTTATGGA	18 770 SEQ. ID NO	770
YE_ail_PM_29	AAGTAAGACGTCAATGGC	18 771 SEQ. ID NO	771
YE_ail_PM_30	AGACGTCAATGGCATAACG	18 772 SEQ. ID NO	772
YE_ail_PM_31	TCAATGGCATAACGGGGCA	18 773 SEQ. ID NO	773
YE_ail_PM_33	ACGGGGCAGGGGTGCAAT	18 774 SEQ. ID NO	774
YE_ail_PM_34	GCAGGGGTGCAATTC AAC	18 775 SEQ. ID NO	775
YE_ail_PM_35	TCCAAATTTTGTCATTGA	18 776 SEQ. ID NO	776
YE_ail_PM_36	ATTTTGTCATTGACGCTT	18 777 SEQ. ID NO	777
YE_ail_PM_37	GTCATTGACGCTTCATAT	18 778 SEQ. ID NO	778
YE_ail_PM_38	TGACGCTTCATATGAATA	18 779 SEQ. ID NO	779
YE_ail_PM_39	CTTCATATGAATACTCCA	18 780 SEQ. ID NO	780
YE_ail_PM_40	TATGAATACTCCAACTC	18 781 SEQ. ID NO	781
YE_ail_PM_6	TGTTAATGTGTACGCTGC	18 782 SEQ. ID NO	782
YE_ail_PM_7	ATGTGTACGCTGCGAGTG	18 783 SEQ. ID NO	783
YE_ail_PM_9	AAGCCATGTAAAAGAAAA	18 784 SEQ. ID NO	784
YE_ysts1_PM_10	GCCACCAGCCGAAGTCAG	18 785 SEQ. ID NO	785
YE_ysts1_PM_4	CAAGAAACAGTTTCAGGG	18 786 SEQ. ID NO	786
YE_ysts1_PM_5	TCAGTGATGCATTATCGA	18 787 SEQ. ID NO	787
YE_ysts1_PM_6	GATGCATTATCGACACCA	18 788 SEQ. ID NO	788
YE_ysts1_PM_7	TATCGACACCAATAACCG	18 789 SEQ. ID NO	789
YE_ysts1_PM_8	TCGACACCAATAACCGCT	18 790 SEQ. ID NO	790
YE_ysts1_PM_9	AATAACCGCTGAGGTATA	18 791 SEQ. ID NO	791

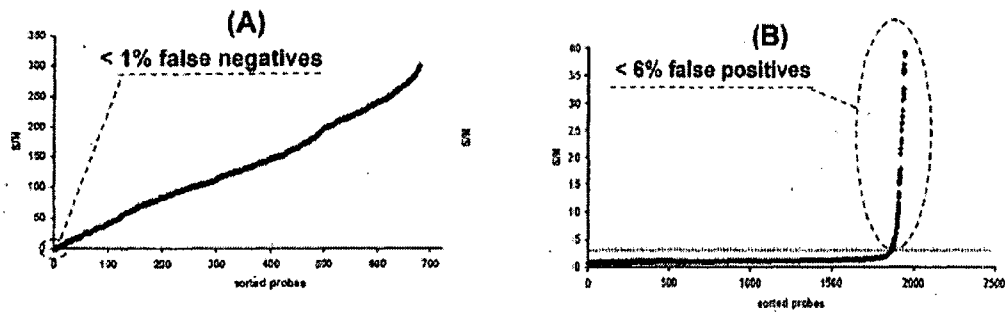
Figure 8A

<i>Pseudomonas aeruginosa</i>		
	exotoxin (etA)	40
	GDP-mannose dehydrogenase (algD)	40
	outer membrane proteins (oprI and oprL)	80
<i>Salmonella</i>		
	type I fimbriae (fimA)	40
	phase 1 flagellin (fljC)	120
	phase 2 flagellin (fljB)	5
	regulator (hilA)	40
	type III secretion system (invA and invE)	40
	enterotoxin (stn)	
<i>Staphylococcus aureus</i>		
	exfoliative toxins (etA and etB)	80
	enterotoxins (seA, seB, seC, seD, seE, seI, seG, seH)	338
	toxic shock syndrome toxin (tsst1)	40
	nuclease (nuc)	40
<i>Vibrio cholerae</i>		
	cholera toxin (ctxA and ctxB)	80
	hemolysin (hlyA)	40
	accessory cholera toxin (ace)	40
	zona occludens toxin (zot)	40
	outer membrane protein (ompU)	40
	response regulator (toxR)	40
	toxin-coregulated pilus (tcpA)	160
<i>Vibrio parahaemolyticus</i>		
	thermostable direct hemolysin (tdh)	40
	thermolabile hemolysin (tl)	40
	response regulator (toxR)	40
<i>Yersinia enterocolitica</i>		
	attachment-invasion locus (ail)	40
	adhesin (yadA)	40
	type III secretion system (virF)	40
	enterotoxin (ystA)	10
	enterotoxin (ystB)	13

Figure 8B

Pathogens	Genes	No. of probes on chip
<i>Campylobacter jejuni</i>		
	cytolethal distending toxin (cdtA, cdtB and cdtC)	120
	fibronectin binding protein (cadF)	40
	Invasion antigen B (ciaB)	40
	response regulator (racR)	40
	hippuricase (hipO)	40
	type IV secretion system (virB11)	40
<i>Clostridium perfringens</i>		
	alpha toxin (plc)	40
	enterotoxin (cpe)	40
	beta-toxin (cpb)	80
	epsilon toxin (etx)	40
	lota toxin (IA)	80
<i>Enterococcus faecalis</i>		
	collagen adhesion (ace)	40
	cytolysin (cylA)	40
	endocarditis specific antigen (efaA)	40
	surface protein (esp)	40
	gelatinase (gelE)	40
<i>Helicobacter pylori</i>		
	vacuolating cytotoxin (vacA)	21
	cytotoxin-associated antigen (cagA)	12
	urease (ureA and ureB)	80
	flagellin (flaA)	40
<i>Listeria monocytogenes</i>		
	internalins (inlA and inlB)	80
	listeriolysin (lisA)	40
	Invasion-associated protein (iap)	40
	phospholipase (plcA and plcB)	80
	metalloprotease (mpl)	40
	actin binding protein (actA)	40
	protease (clpE)	40
	regulator protein (prfA)	
<i>Legionella pneumophila</i>		
	macrophage infectivity potentiator (mip)	9
	heat-shock protein (dnaJ)	40
	type IV secretion system (dotA)	40

Figure 9



C

Pathogen	PCR Batch				
	1	2	3	4	5
<i>C. jejuni</i>	cdtA, hipO	cadF	cdtB	-	-
<i>H. pylori</i>	-	ureB	flaA	flaA	ureB
<i>C. perfringens</i>	cpe	cpe, plc	plc	-	-
<i>L. monocytogenes</i>	inlA	-	hlyA, plcB	-	-
<i>E. faecalis</i>	gelE	efaA	ace	gelE	ace
<i>S. aureus</i>	nuc	nuc, tsstI	tsstI	-	-
<i>V. cholerae</i>	-	ctxA	ctxB, toxR	zot	-
<i>V. parahaemolyticus</i>	tdh	-	-	tlh	toxR, tlh
<i>Y. enterocolitica</i>	-	ystB	-	-	ail
<i>Salmonella</i>	invA	-	-	fimA	hilA
<i>P. aeruginosa</i>	-	-	-	oprL	-
<i>L. pneumophila</i>	-	-	-	dnaJ, mip	dnaJ
Primer Pairs/Genes	8	9	9	8	7
Probes	154	153	129	167	145

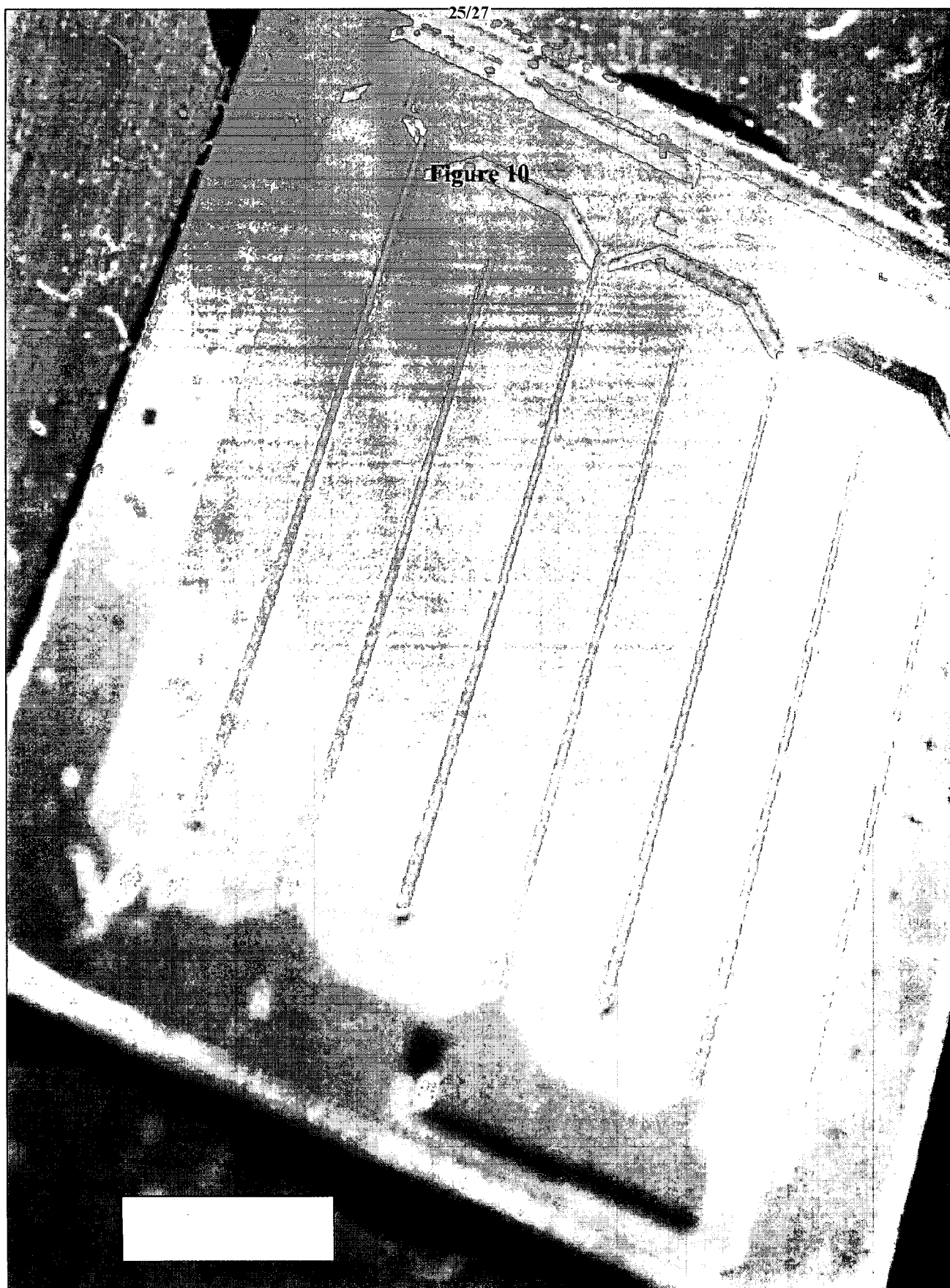


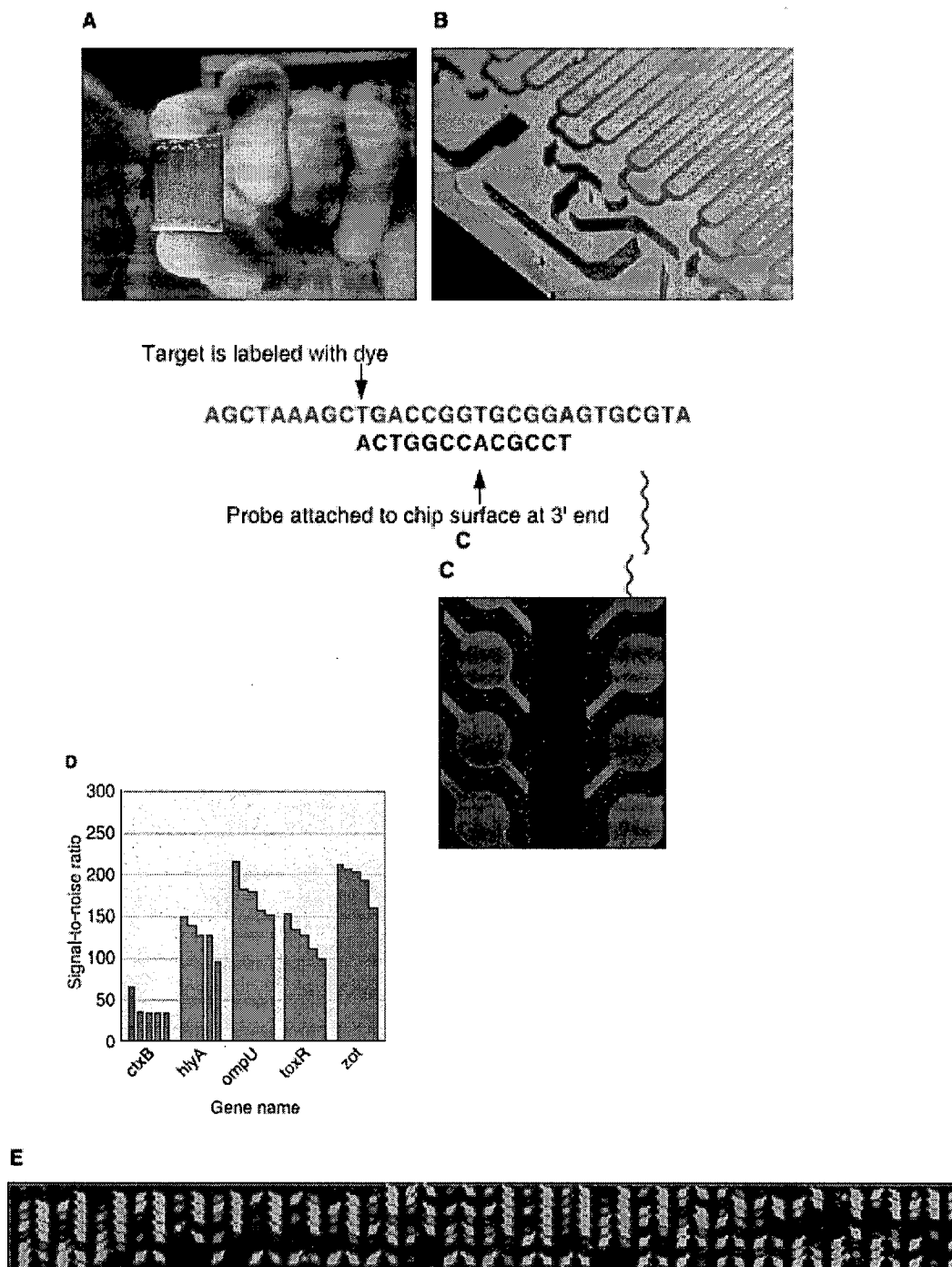
Figure 11

FIGURE 12

SEQ ID NO: 792 Mycobacterium_leprae_erp_AL583917
GAGCTTGTAATGACCTGCTAAAAGTTGCGAACCAGTTGGGCGCCAGCCA
SEQ ID NO: 793 Mycobacterium_leprae_erp_AL583917
TTGTCGGGCTCAGTAACGCCGCGCGATATCACTGATTCCTGTCACGTA
SEQ ID NO: 794 Mycobacterium_leprae_erp_AL583917
AATGACCTGCTAAAAGTTGCGAACCAGTTGGGCGCCAGCCAGGTCATGGA
SEQ ID NO: 795 Mycobacterium_leprae_erp_AL583917
CAAACCCGCGGCCAAACACTATGAATTCAAACAAGCAGCATCGATAGCCG
SEQ ID NO: 796 Mycobacterium_leprae_erp_AL583917
AAAGTTGCGAACCAGTTGGGCGCCAGCCAGGTCATGGACCTAATCAAGGG
SEQ ID NO: 797 Mycobacterium_leprae_erp_AL583917
CACGTTGGGCGGTACATACCGATCAGTACCAGCAGCGGTGAGCTTGTA
SEQ ID NO: 798 Mycobacterium_leprae_erp_AL583917
GTACATCACCGATCAGTACCAGCAGCGGTGAGCTTGTAATGACCTGCTA
SEQ ID NO: 799 Mycobacterium_leprae_erp_AL583917
TAACTGGCACCGATGATCCAGGTAATGGCCTGAGAACTCCCGGTTTGACC
SEQ ID NO: 800 Mycobacterium_leprae_erp_AL583917
GACCGCGAGCGCCAAACCCGCGGCCAAACACTATGAATTCAAACAAGCAG
SEQ ID NO: 801 Mycobacterium_leprae_erp_AL583917
TGCCTTCGCTAACTGGCACCGATGATCCAGGTAATGGCCTGAGAACTCCC
SEQ ID NO: 802 Mycobacterium_leprae_erp_AL583917
TTTACGAACCGACCGCGAGCGCCAAACCCGCGGCCAAACACTATGAATTC
SEQ ID NO: 803 Mycobacterium_leprae_erp_AL583917
ATCGCCAGTCCATGCGCATATTTCTTGTGTTACGAACCGACCGCGAGCGC
SEQ ID NO: 804 Mycobacterium_leprae_erp_AL583917
GACCAGCCCCGATCTGACAAATCAGGAGCTAGGGACACCTGTGCTACCG
SEQ ID NO: 805 Mycobacterium_leprae_erp_AL583917
CACCTAGCAATCGCCAGTCCATGCGCATATTTCTTGTGTTACGAACCGA
SEQ ID NO: 806 Mycobacterium_leprae_erp_AL583917
CCAGCCAGGTCATGGACCTAATCAAGGGTGTGGTGATGCCAGCGGTCATG
SEQ ID NO: 807 Mycobacterium_leprae_erp_AL583917
TCAACCTACCGCCGGTGCCTTCGCTAACTGGCACCGATGATCCAGGTAAT
SEQ ID NO: 808 Mycobacterium_leprae_erp_AL583917
ATTCAAACAAGCAGCATCGATAGCCGACCTGCCCCGAGAAAGTGCTGGACG
SEQ ID NO: 809 Mycobacterium_leprae_erp_AL583917
ATAAGCACGGTCGCCACCCTAGCAATCGCCAGTCCATGCGCATATTTCTT
SEQ ID NO: 810 Mycobacterium_leprae_erp_AL583917
CAAACACTATGAATTCAAACAAGCAGCATCGATAGCCGACCTGCCCCGAG
SEQ ID NO: 811 Mycobacterium_leprae_erp_AL583917
TCCATGCGCATATTTCTTGTGTTACGAACCGACCGCGAGCGCCAAACCCG
SEQ ID NO: 812 Mycobacterium_leprae_glnA1_AL583920
TGAGTTTCGACTCACGCTCGAACGGTTCGTTTTACCAAGTCGATGCCATT
SEQ ID NO: 813 Mycobacterium_leprae_glnA1_AL583920
TCACGCTCGAACGGTTCGTTTTACCAAGTCGATGCCATTTCTGGGTGGTG
SEQ ID NO: 814 Mycobacterium_leprae_glnA1_AL583920
TACAAGCGTTTGGTTCCCGGCTATGAGGCCCGATTAACTAGTTTACAG
SEQ ID NO: 815 Mycobacterium_leprae_glnA1_AL583920
TCGAGGAAGACCACGACTACCTCACTGAGGGAAGCGTGTTACACCCCGAC
SEQ ID NO: 816 Mycobacterium_leprae_glnA1_AL583920
GAGAACGAAATCGAACCGGTCAACATTCGACCGCATCCGTACGAGTTTGC
SEQ ID NO: 817 Mycobacterium_leprae_glnA1_AL583920
CTGCATGCGGCCGATGACATGCAACTGTACAAATACATTGTCAAAAACAC
SEQ ID NO: 818 Mycobacterium_leprae_glnA1_AL583920
TGGATCAACTTCAAACGCGAGAACGAAATCGAACCGGTCAACATTTCGACC
SEQ ID NO: 819 Mycobacterium_leprae_glnA1_AL583920
TTTGGTTCCCGGCTATGAGGCCCGATTAACTAGTTTACAGCCAGCGCA
SEQ ID NO: 820 Mycobacterium_leprae_glnA1_AL583920
GATCGAGCCTCAAGCCCCGGTTGACAAAGACCTCTACGAACTGCCTCCGG
SEQ ID NO: 821 Mycobacterium_leprae_glnA1_AL583920
TCGACTCGGTGAGTTTCGACTCACGCTCGAACGGTTCGTTTTACCAAGTC
SEQ ID NO: 822 Mycobacterium_leprae_glnA1_AL583920
TTGACAAAGACCTCTACGAACTGCCTCCGGAGGAGGCCGAGACATTCCG
SEQ ID NO: 823 Mycobacterium_leprae_glnA1_AL583920
CCGATGACATGCAACTGTACAAATACATTGTCAAAAACACCGCGTGCGACG