US 20140155283A1

(19) **United States**

(12) **Patent Application Publication**

Venkateswaran et al.

(10) **Pub. No.: US 2014/0155283 A1**

(43) **Pub. Date: Jun. 5, 2014**

(54) **MICROARRAY FOR DETECTING VIABLE ORGANISMS**

(75) Inventors: **Kasthuri Venkateswaran**, Arcadia, CA (US); **Christina Dock**, Pasadena, CA (US); **Gary Aandersen**, Berkeley, CA (US); **Todd DeSantis**, Livermore, CA (US)

(73) Assignee: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA**, Oakland, CA (US)

(21) Appl. No.: **13/984,257**

(22) PCT Filed: **Feb. 10, 2012**

(86) PCT No.: **PCT/US12/24724**

§ 371 (c)(1),
(2), (4) Date: **Jan. 21, 2014**

Related U.S. Application Data

(60) Provisional application No. 61/441,820, filed on Feb. 11, 2011.

Publication Classification

(51) **Int. Cl.**
C12Q 1/68 (2006.01)

(52) **U.S. Cl.**
CPC ***C12Q 1/689*** (2013.01)
USPC **506/9**; 506/16

(57) **ABSTRACT**

A methodology of microarray using the fluorescent DNA intercalating agent propidium monoazide (PMA) to selectively block DNA of dead cells from amplification and its application in detecting and enumerating viable microbes in complex microbial communities is described. A phylogenetic array is used in the preferred embodiment to enhance the sensitivity of the method. The PMA-Microarray assay is particularly applicable for monitoring samples from environments with extremely low microbial burden such as spacecraft surfaces.

Figure 1.

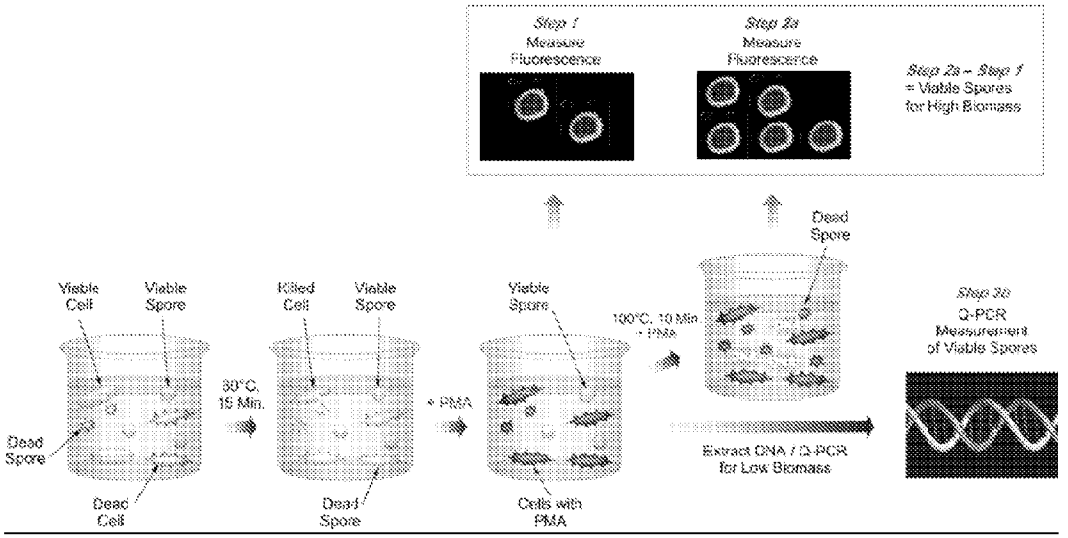

Step 1
Measure
Fluorescence
Step 2a
Measure
Fluorescence
Step 2a – Step 1
= Viable Spores
for High Biomass
Viable
Cell
Viable
Spore
Dead
Spore
Dead
Cell
80°C,
15 Min.
Killed
Cell
Viable
Spore
Dead
Spore
+ PMA
Viable
Spore
Cells with
PMA
100°C, 10 Min.
+ PMA
Dead
Spore
Extract DNA / Q-PCR
for Low Biomass
Step 2b
Q-PCR
Measurement
of Viable Spores

Figure 2.

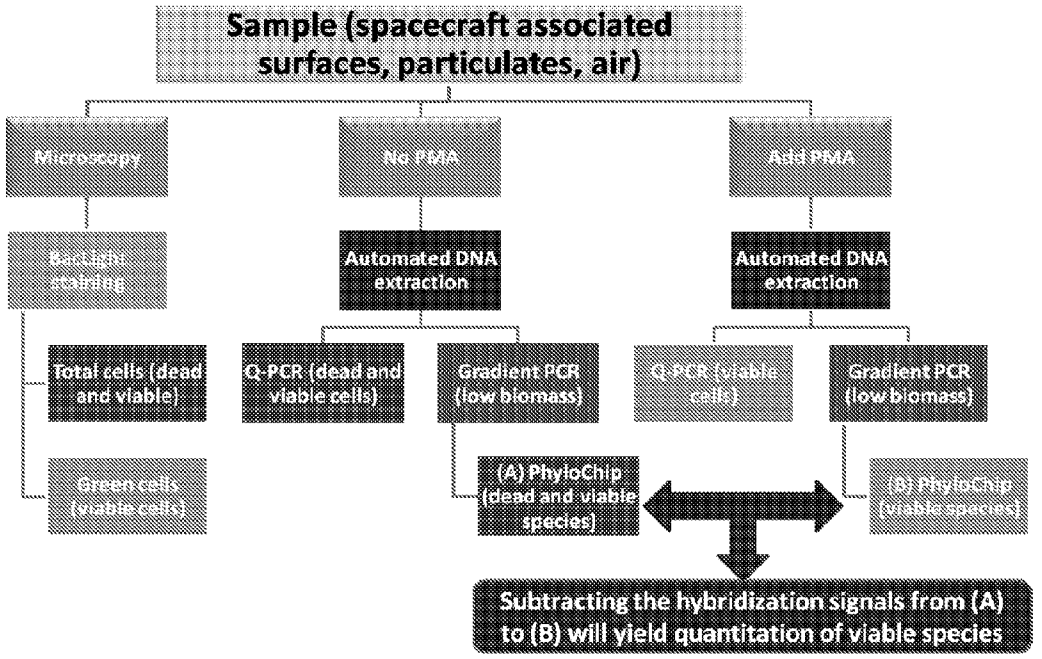

Sample (spacecraft associated surfaces, particulates, air)
Microscopy
No PMA
Add PMA
Backlight staining
Automated DNA extraction
Automated DNA extraction
Total cells (dead and viable)
Q-PCR (dead and viable cells)
Gradient PCR (low biomass)
Q-PCR (viable cells)
Gradient PCR (low biomass)
Green cells (viable cells)
(A) PhyloChip (dead and viable species)
(B) PhyloChip (viable species)
Subtracting the hybridization signals from (A) to (B) will yield quantitation of viable species

Figure 3

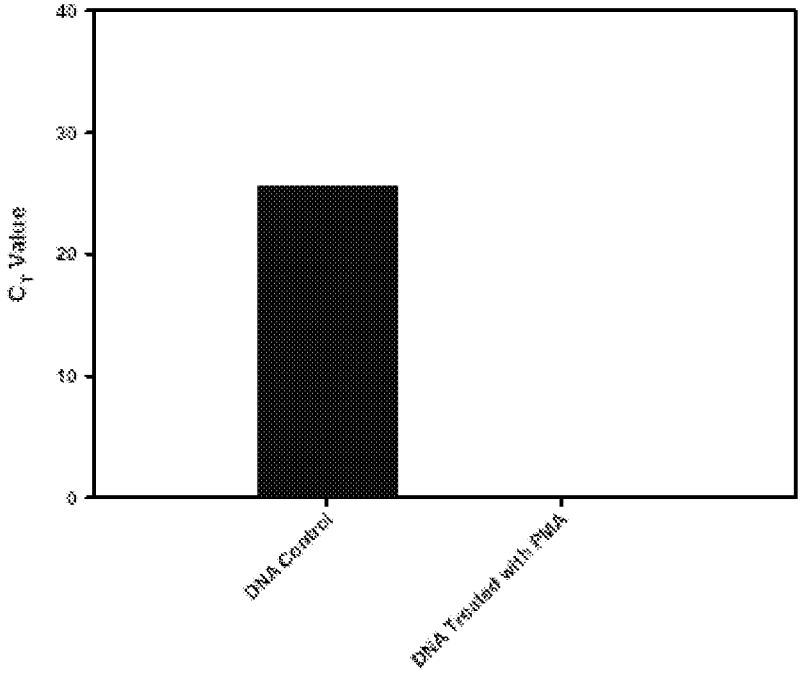


Figure 4

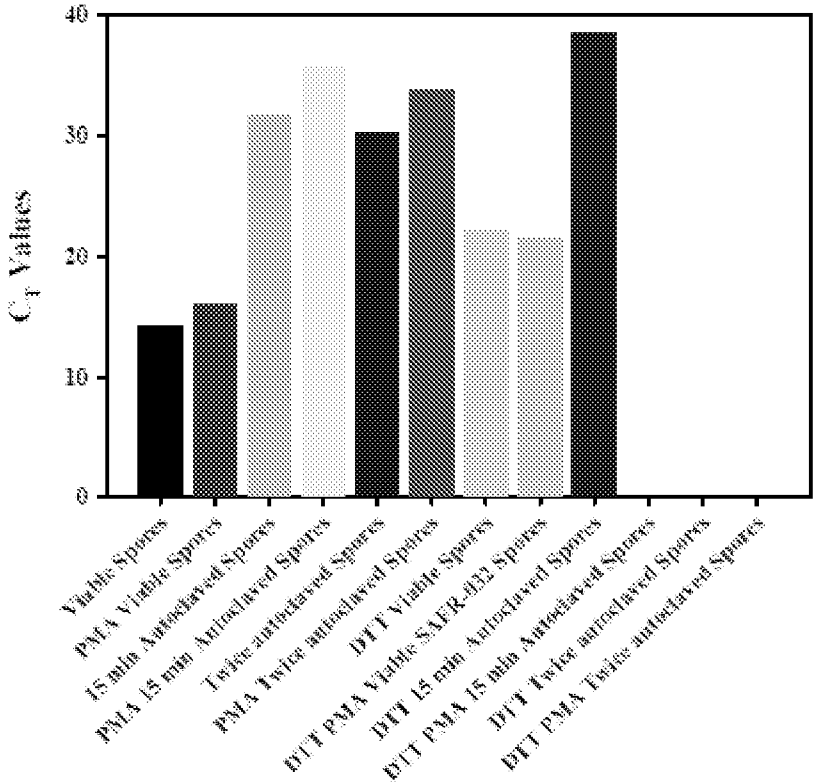

40
30
20
10
0
C_T Values
Viable Spores
PMA Viable Spores
15 min Autoclaved Spores
PMA 15 min Autoclaved Spores
Twice autoclaved spores
PMA Twice autoclaved Spores
DFT Viable Spores
DFT PMA Viable SAFR-032 Spores
DFT 15 min Autoclaved Spores
DFT PMA 15 min Autoclaved Spores
DFT Twice autoclaved Spores
DFT PMA Twice autoclaved Spores

Figure 5

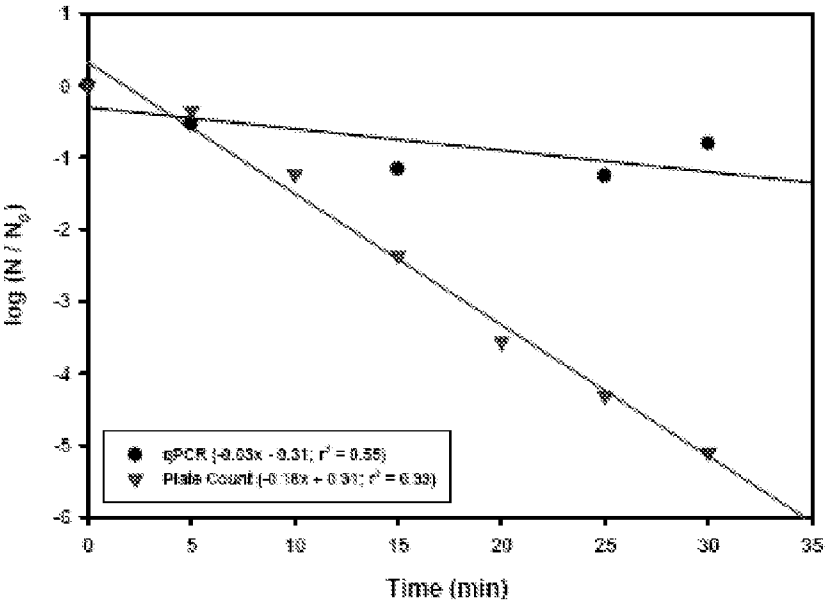

log (N / N0)
Time (min)
qPCR (-0.03x - 0.31; r² = 0.55)
Plate Count (-0.18x + 0.31; r² = 0.93)

Figure 6

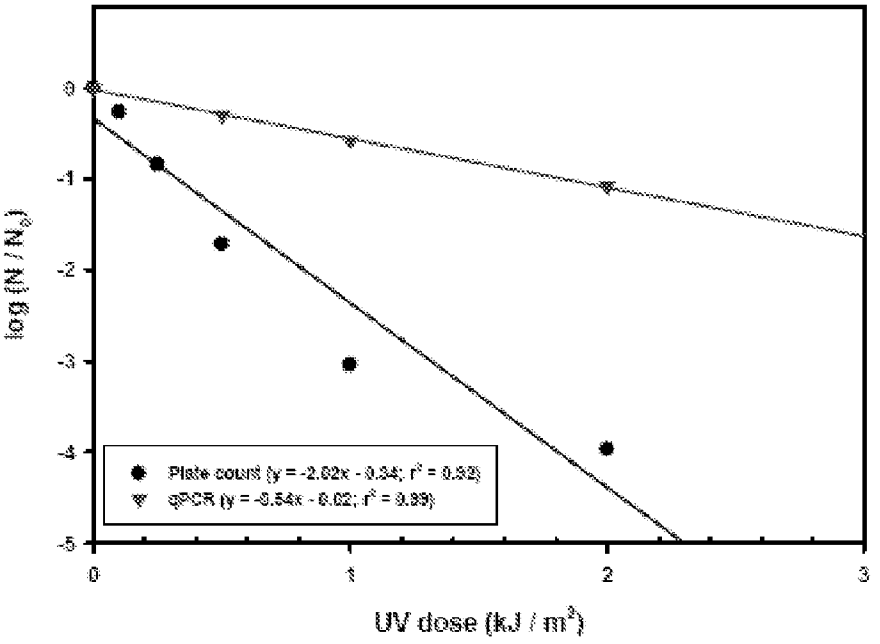
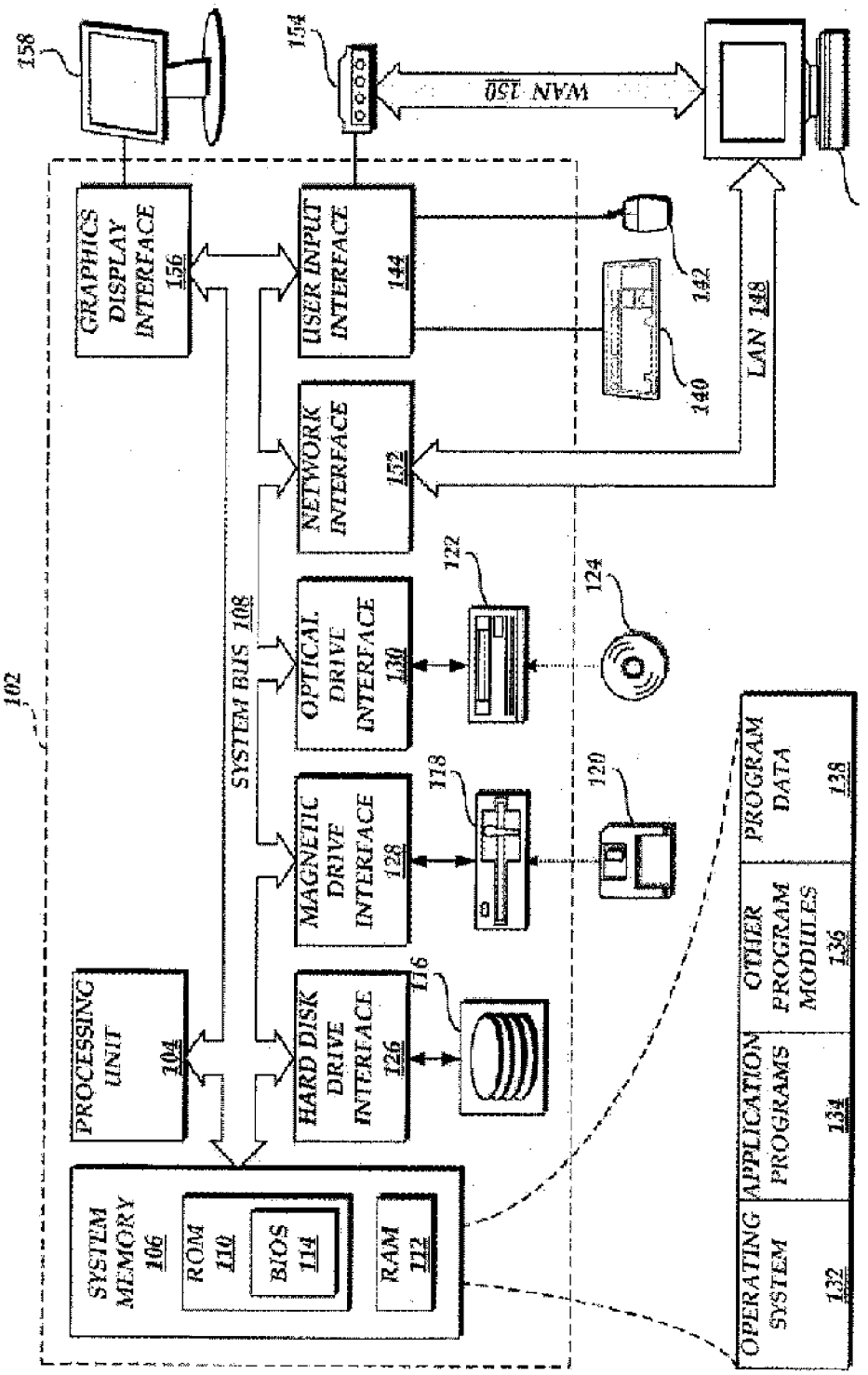

log (N / N0)
0
-1
-2
-3
-4
-5
0
1
2
3
UV dose (kJ / m2)
Plate count (y = -2.02x - 0.34; r2 = 0.92)
qPCR (y = -0.54x - 0.02; r2 = 0.89)

Figure 7


158
154
WAN 150
GRAPHICS DISPLAY INTERFACE 156
USER INPUT INTERFACE 144
142
140
LAN 148
NETWORK INTERFACE 172
122
124
102
SYSTEM BUS 108
OPTICAL DRIVE INTERFACE 130
118
120
MAGNETIC DRIVE INTERFACE 128
116
PROCESSING UNIT 104
HARD DISK DRIVE INTERFACE 126
SYSTEM MEMORY 106
ROM 110
BIOS 114
RAM 112
OPERATING SYSTEM 132
APPLICATION PROGRAMS 134
OTHER PROGRAM MODULES 136
PROGRAM DATA 138

MICROARRAY FOR DETECTING VIABLE ORGANISMS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. §119(e) to U.S. Provisional Application No. 61/441,820 filed Feb. 11, 2011, the content of which is herein expressly incorporated by references in their entireties.

[0002] This application is related to U.S. Application Ser. No. 61/259,565 [Attorney Docket No. IB-2733P1], filed on Nov. 9, 2009; U.S. Application Ser. No. 61/317,644 [Attorney Docket No. IB-2733P2], filed on Mar. 25, 2010; U.S. Application Ser. No. 61/347,817 [Attorney Docket No. IB-2733P3], filed on May 24, 2010 and co-pending U.S. patent application Ser. No. 12/474,204, filed on May 28, 2009, which are hereby incorporated by reference in their entirety.

[0003] This application is related to the co-pending international application having application number PCT/US2010/040106 [Attorney Docket No. IB-2733PCT], filed on Jun. 25, 2010, which is incorporated herein by reference.

[0004] This application is also related to the co-pending U.S. patent application Ser. No. 13/152,213, filed on Jun. 2, 2011, which is hereby incorporated by reference in its entirety.

STATEMENT OF GOVERNMENTAL SUPPORT

[0005] The invention claimed herein was made in the performance of work under DOE and NASA contracts, and is subject to the provisions of Public Law 96-517 (35 USC 202) in which the Contractor has elected to retain title. This invention was made in part under Contract DE-AC02-05CH11231 awarded by the Department of Energy and under Contract Nos. NNH09ZDA001N and NAS7-1407 awarded by NASA. The government has certain rights in this invention.

BACKGROUND

[0006] There is a continuing need for the development of a sensitive, rapid method that will detect and enumerate total viable microorganisms. Culture-based plating methods, including the NASA standard assay, have remained the "gold standard," even though such methods are both laborious and time consuming. Nucleic-acid-based amplification such as the polymerase chain reaction (PCR), particularly real-time quantitative-PCR (qPCR), provides a promising approach that allows rapid detection in a wide range of applications. Nonetheless, PCR-based methods are rarely used for process validation, because such methods cannot discriminate between live and inactive microorganisms due to the high stability of the DNA from dead bacterial cells. Detecting the number or kinds of viable microorganisms is useful in a variety of applications.

SUMMARY OF THE INVENTION

[0007] The present application provides a method of detecting and enumerating low levels of microorganisms. The method combines propidium monoazide (PMA) assay with PhyloChip microarray analysis, allowing rapid, selective detection and enumeration of viable microbes in microbial communities containing both live and dead microorganisms.

[0008] In one aspect, the present application provides a method for detecting live cells in a sample. In some embodiments, the method comprises selectively amplifying a nucleic acid from live cells from a sample comprising live and dead cells; and detecting the presence, absence, relative abundance, and/or quantity of one or more operational taxon units (OTUs) in the sample based on hybridization of amplified nucleic acid to a plurality of probes complementary to 16s rRNA sequences, wherein said one or more OTUs consist essentially of viable organisms from said sample. In some embodiments, the nucleic acid is DNA. In some embodiments, selectively amplifying a nucleic acid comprises pre-treating the sample with an agent that selectively modifies a nucleic acid of dead cells. In some embodiments, the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent, such as propidium monoazide. In some embodiments, the probes are used to detect the presence, absence, relative abundance, and/or quantity of at least 10,000 different OTUs in a single assay. In some embodiments, the presence, absence, relative abundance, and/or quantity is detected with a confidence level greater than 95%. In some embodiments, the method further comprises quantifying the number of live cells in the sample.

[0009] In one aspect, the present application provides a method for detecting viable organisms in a sample. In some embodiments, the method comprises selectively amplifying a nucleic acid from live cells from a sample comprising live and dead cells; and determining the presence, absence, relative abundance, and/or quantity of at least 1,000 different OTUs in a single assay, wherein said OTUs consist essentially of viable organisms from said sample. In some embodiments, selectively amplifying comprises pre-treating the sample with an agent that selectively modifies a nucleic acid of dead cells. In some embodiments, the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent, such as propidium monoazide. In some embodiments, contacting a sample with an agent that selectively modifies a nucleic acid of dead cells in the sample is followed by exposure to visible light. In some embodiments, the presence, absence, relative abundance, and/or quantity is detected with a confidence level greater than 95%. In some embodiments, the method further comprises quantifying the number of live cells in the sample.

[0010] In one aspect, the present application provides a method for detecting and optionally quantifying viable organisms in a sample. In some embodiments, the method comprises (a) selectively amplifying a nucleic acid from live cells from a sample comprising live and dead cells; (b) hybridizing amplified nucleic acid to a plurality of probes; (c) determining hybridization signal strength distributions for a plurality of different interrogation probes, each of which is complementary to a section within one or more highly conserved polynucleotides in one or more target OTUs; (d) determining hybridization signal strength distributions for a plurality of mismatch probes, wherein for each interrogation probe, one or more different corresponding mismatch probes comprising one or more nucleotide mismatches with said section within said one or more highly conserved polynucleotides are included in the plurality of mismatch probes; and, (e) using the hybridization signal strengths of the interrogation probes and mismatch probes to determine the probability that the hybridization signal for the different interrogation probes represents the presence, absence, relative abundance, and/or quantity of said one or more OTUs, wherein said one or more OTUs consist essentially of viable organisms from said sample. In some embodiments, selectively amplifying

comprises pre-treating the sample with an agent that selectively modifies a nucleic acid of dead cells. In some embodiments, the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent, such as propidium monoazide. In some embodiments, contacting a sample with an agent that selectively modifies a nucleic acid of dead cells in the sample is followed by exposure to visible light. In some embodiments, the highly conserved polynucleotides are selected from the group consisting of 16S rRNA gene, 23S rRNA gene, 5S rRNA gene, 5.8S rRNA gene, 12S rRNA gene, 18S rRNA gene, 28S rRNA gene, gyrB gene, rpoB gene, fusA gene, recA gene, coxl gene, nif13 gene, RNA molecules derived therefrom, and a combination thereof. In some embodiments, each interrogation probe has 4 or more corresponding mismatch probes in the plurality of mismatch probes. In some embodiments, the probes are used to detect the presence, absence, relative abundance, and/or quantity of at least 10,000 different OTUs in a single assay. In some embodiments, the probes are attached to a substrate, such as a bead or a microsphere. In some embodiments, the probes are attached to a substrate comprising glass, plastic, or silicon. In some embodiments, the presence, absence, relative abundance, and/or quantity is detected with a confidence level greater than 95%.

[0011] In one aspect, the present application provides a kit for detecting and optionally quantifying live cells in a sample. In some embodiments, the kit comprises (a) an agent that selectively modifies a nucleic acid of dead cells; (b) a plurality of different interrogation probes, each of which is complementary to a section within one or more highly conserved polynucleotides in one or more target operational taxon units (OTUs); and, (c) a plurality of mismatch probes, wherein for each interrogation probe, one or more different corresponding mismatch probes comprising one or more nucleotide mismatches with said section within said one or more highly conserved polynucleotides are included in the plurality of mismatch probes. In some embodiments, the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent, such as propidium monoazide. In some embodiments, each interrogation probe has 4 or more corresponding mismatch probes in the plurality of mismatch probes. In some embodiments, the highly conserved polynucleotides are selected from the group consisting of 16S rRNA gene, 23S rRNA gene, 5S rRNA gene, 5.8S rRNA gene, 12S rRNA gene, 18S rRNA gene, 28S rRNA gene, gyrB gene, rpoB gene, fusA gene, recA gene, coxl gene, nif13 gene, RNA molecules derived therefrom, and a combination thereof.

INCORPORATION BY REFERENCE

[0012] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0014] FIG. **1** shows a schematic of the rapid measurement of bacterial spores using PMA system by fluorescence method.

[0015] FIG. **2** shows a comprehensive flow chart depicting a PMA-Microarray experimental process.

[0016] FIG. **3** shows a graph of real-time PCR using an *E. coli* DNA control before and after treatment with PMA.

[0017] FIG. **4** shows a graph of *B. pumilus* SAFR-032 spores with different treatments. Gene target for qPCR is gyrB.

[0018] FIG. **5** shows a graph of heat resistance of *B. pumilus* spores at 90° C. The D-values for plate and qPCR are 5 and 33 min, respectively.

[0019] FIG. **6** shows a graph of UV inactivation curve of *B. pumilus* spores. Curves are based on the averages of all results in triplicate.

[0020] FIG. **7** illustrates an example of a suitable computer system environment.

DETAILED DESCRIPTION

Definitions

[0021] As used herein, the term "oligonucleotide" refers to a polynucleotide, usually single stranded, that is either a synthetic polynucleotide or a naturally occurring polynucleotide. The length of an oligonucleotide is generally governed by the particular role thereof, such as, for example, probe, primer and the like. Various techniques can be employed for preparing an oligonucleotide, for instance, biological synthesis or chemical synthesis. A nucleic acid disclosed herein will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramide (Beaucage, et al., Tetrahedron, 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem., 35:3800 (1970); Sprinzl, et al., Eur. J. Biochem., 81:579 (1977); Letsinger, et al., Nucl. Acids Res., 14:3487 (1986); Sawai, et al., Chem. Lett., 805 (1984), Letsinger, et al., J. Am. Chem. Soc., 110:4470 (1988); and Pauwels, et al., Chemica Scripta, 26:141 (1986)); phosphorothioate (Mag, et al, Nucleic Acids Res., 19:1437 (1991); and U.S. Pat. No. 5,644,048); phosphorodithioate (Briu, et al., J. Am. Chem. Soc., 111:2321 (1989)); 0-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press); and peptide nucleic acid backbones and linkages (see Egholm, J. Am. Chem. Soc., 114:1895 (1992); Meier, et al., Chem. Int. Ed. Engl., 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson, et al., Nature, 380:207 (1996), all of which are incorporated by reference)). Other analog nucleic acids include those with positive backbones (Denpcy, et al., Proc. Natl. Acad. Sci. USA, 92:6097 (1995)); non-ionic backbones (U.S. Pat. Nos. 5,386,023; 5,637,684; 5,602,240; 5,216,141; and 4,469,863; Kiedrowshi, et al., Angew. Chem. Intl. Ed. English, 30:423 (1991); Letsinger, et al., J. Am. Chem. Soc., 110:4470 (1988); Letsinger, et al., Nucleosides & Nucleotides, 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker, et al., Bioorganic & Medicinal Chem. Lett., 4:395 (1994); Jeffs, et al., J. Biomolecular NMR, 34:17 (1994); Tetrahedron Lett., 37:743 (1996)); and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Anti-

sense Research", Ed. Y. S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins, et al., Chem. Soc. Rev., (1995) pp. 169-176). Several nucleic acid analogs are described in Rawls, C & E News, Jun. 2, 1997, page 35. All of these references are hereby expressly incorporated by reference.

[0022] The nucleic acid may be DNA, RNA, or a hybrid and may contain any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthanine, hypoxanthanine, isocytosine, isoguanine, and base analogs such as nitropyrrole and nitroindole, etc. Oligonucleotides can be synthesized by standard methods such as those used in commercial automated nucleic acid synthesizers and later attached to an array, bead or other suitable surface. Alternatively, the oligonucleotides can be synthesized directly on the assay surface using photolithographic or other techniques. In some embodiments, linkers are used to attach the oligonucleotides to an array surface or to beads.

[0023] As used herein, the terms "nucleic acid," "nucleic acid molecule," and "polynucleotide" refer to a compound or composition that is a polymeric nucleotide or nucleic acid polymer. The nucleic acid molecule may be a natural compound or a synthetic compound. The nucleic acid molecule can have from about 2 to 5,000,000 or more nucleotides. The larger nucleic acid molecules are generally found in the natural state. In an isolated state, the nucleic acid molecule can have about 10 to 50,000 or more nucleotides, usually about 100 to 20,000 nucleotides. It is thus obvious that isolation of a nucleic acid molecule from the natural state often results in fragmentation. It may be useful to fragment longer target nucleic acid molecules, particularly RNA, prior to hybridization to reduce competing intramolecular structures. Fragmentation can be achieved chemically, enzymatically, or mechanically. Typically, when the sample contains DNA, a nuclease such as deoxyribonuclease (DNase) is employed to cleave the phosphodiester linkages. Nucleic acid molecules, and fragments thereof, include, but are not limited to, purified or unpurified forms of DNA (dsDNA and ssDNA) and RNA, including tRNA, mRNA, rRNA, mitochondrial DNA and RNA, chloroplast DNA and RNA, DNA/RNA hybrids, biological material or mixtures thereof, genes, chromosomes, plasmids, cosmids, the genomes of microorganisms, e.g., bacteria, yeasts, phage, chromosomes, viruses, viroids, molds, fungi, or other higher organisms such as plants, fish, birds, animals, humans, and the like. The polynucleotide can be only a minor fraction of a complex mixture such as a biological sample.

[0024] As used herein, the term "hybridize" refers to the process by which single strands of polynucleotides form a double-stranded structure through hydrogen bonding between the constituent bases. The ability of two polynucleotides to hybridize with each other is based on the degree of complementarity of the two polynucleotides, which in turn is based on the fraction of matched complementary nucleotide pairs. The more nucleotides in a given polynucleotide that are complementary to another polynucleotide, the more stringent the conditions can be for hybridization and the more specific will be the binding between the two polynucleotides. Increased stringency may be achieved by elevating the temperature, increasing the ratio of co-solvents, lowering the salt concentration, and combinations thereof.

[0025] As used herein, the terms "complementary," "complement," and "complementary nucleic acid sequence" refer to the nucleic acid strand that is related to the base sequence in another nucleic acid strand by the Watson-Crick base-pairing rules. In general, two polynucleotides are complementary when one polynucleotide can bind another polynucleotide in an anti-parallel sense wherein the 3'-end of each polynucleotide binds to the 5'-end of the other polynucleotide and each A, T(U), G, and C of one polynucleotide is then aligned with a T(U), A, C, and G, respectively, of the other polynucleotide. Polynucleotides that comprise RNA bases can also include complementary G/U or U/G basepairs. Two complementary strands may comprise complementary regions comprising all or one or more portions of one or both strands.

[0026] As used herein, the term "clustering tree" refers to a hierarchical tree structure in which observations, such as organisms, genes, and polynucleotides, are separated into one or more clusters. The root node of a clustering tree consists of a single cluster containing all observations, and the leaf nodes correspond to individual observations. A clustering tree can be constructed on the basis of a variety of characteristics of the observations, such as sequences of the genes and morphological traits of the organisms. Many techniques known in the art, e.g. hierarchical clustering analysis, can be used to construct a clustering tree. A non-limiting example of the clustering tree is a phylogenetic, taxonomic or evolutionary tree.

[0027] As used herein, the terms "operational taxon unit," "OTU," "taxon," "hierarchical cluster," and "cluster" are used interchangeably. An operational taxon unit (OTU) refers to a group of one or more organisms that comprises a node in a clustering tree. The level of a cluster is determined by its hierarchical order. In some embodiments, an OTU is a group tentatively assumed to be a valid taxon for purposes of phylogenetic analysis. In some embodiments, an OTU is any of the extant taxonomic units under study. In yet other embodiments, an OTU is given a name and a rank. For example, an OTU can represent a domain, a sub-domain, a kingdom, a sub-kingdom, a phylum, a sub-phylum, a class, a sub-lass, an order, a sub-order, a family, a subfamily, a genus, a subgenus, or a species. In some embodiments, OTUs can represent one or more organisms from the kingdoms eubacteria, protista, or fungi at any level of a hierarchal order. In some embodiments, an OTU represents a prokaryotic or fungal order.

[0028] As used herein, the term "kmer" refers to a polynucleotide of length k. In some embodiments, k is an integer from 1 to 1000. In some embodiments, k is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, or 1000.

[0029] As used herein, the term "perfect match probe" (PM probe) refers to a kmer which is 100% complementary to at least a portion of a highly conserved target gene or polynucleotide. The perfect complementarity usually exists throughout the length of the probe. Perfect probes, however, may have a segment or segments of perfect complementarity that is/are flanked by leading or trailing sequences lacking complementarity to the target gene or polynucleotide.

[0030] As used herein, the term "mismatch probe" (MM probe) refers a control probe that is identical to a corresponding PM probe at all positions except for one, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides of the PM probe. Typically, the non-

identical position or positions are located at or near the center of the PM probe. In some embodiments, the mismatch probes are universal mismatch probes, e.g., a collection of mismatch probes that have no more than a set number of nucleotide variations or substitutions compared to positive probes. For example, the universal mismatch probes may differ in nucleotide sequence by no more than five nucleotides compared to any one PM probe in the PM probe set. In some embodiments, a MM probe is used adjacent to each test probe, e.g., a PM probe targeting a bacterial 16S rRNA sequence, in the array.

[0031] As used herein, the term "probe pair" refers to a PM probe and a corresponding MM probe. In some embodiments, the PM probes and the MM probes are scored in relation to each other during data processing and statistic analysis. As used herein, the term "a probe pair associated with an OTU" is defined as a pair of probes consisting of an OTU-specific PM probe and its corresponding MM probe.

[0032] As used herein, a "sample" is from any source, including, but not limited to a biological sample, a gas sample, a fluid sample, a solid sample, or any mixture thereof.

[0033] As used herein, a "microorganism" or "organism" includes, but is not limited to, a virus, viroids, bacteria, archaea, fungi, protozoa and the like.

[0034] The term "sensitivity" refers to a measure of the proportion of actual positives which are correctly identified as such.

[0035] The term "specificity" refers to a measure of the proportion of actual negatives which are correctly identified as such.

[0036] The term "confidence level" refers to the likelihood, expressed as a percentage, that the results of a test are real and repeatable, and not random. Confidence levels are used to indicate the reliability of an estimate and can be calculated by a variety of methods.

Live Cell Detection

[0037] In some embodiments, the methods comprise selectively amplifying a nucleic acid from live cells from a sample comprising live and dead cells. In some embodiments, the nucleic acid is DNA and/or RNA. In some embodiments, selectively amplifying comprises pre-treating the sample with an agent that selectively modifies a nucleic acid of dead cells. Selective modification can be the result of differences between the accessibility of DNA in live cells and the DNA in and/or derived from dead cells to the selective agent, such as an agent that is a cell membrane- and/or cell wall-impermeant agent. In some embodiments, the DNA of dead cells is more accessible the selective agent than the DNA of live cells. Selective modification can comprise intercalation in the nucleic acid. Exposure to an agent that selectively modifies a nucleic acid of dead cells may be followed by exposure to a second agent that further modifies the nucleic acid modified by the selective agent. In some embodiments, the second agent is visible light. In some embodiments, modified nucleic acid is selectively removed prior to amplification. In some embodiments, modified nucleic acid is present during amplification but is not efficiently amplified. For example, unmodified DNA can be amplified at least about 5-fold, 10-fold, 25-fold, 50-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 600-fold, 700-fold, 800-fold, 900-fold, 1000-fold, 2000-fold, 5000-fold, 10000-fold, 20000-fold, 50000-fold, 100000-fold, or more than modified DNA. In some embodiments, the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent. Examples of DNA intercalating agents include, but are not limited to ethidium compounds, such as ethidium monoazide, ethidium bromide, and ethidium diazide; propidium compounds, such as propidium monozaide and propidium iodide; and 7-Amino-actinomycin D.

[0038] DNA intercalating agents such as ethidium monoazide (EMA) and propidium monoazide (PMA) can be used to selectively distinguish between viable and dead bacterial cells, because they selectively penetrate the membranes of dead cells. Once penetrated, they intercalate the DNA and, upon photolysis using visible light, produce stable DNA monoadducts, which are not efficiently amplified by PCR, either through removal during or subsequent to the DNA extraction process, or by serving as poor templates in the amplification process. Therefore, once PMA is used to treat bacterial cells prior to DNA extraction, only the DNA of viable cells will be available for PCR amplification. In some embodiments, excess amounts of the selective agent are removed before nucleic acid extraction. Following selective amplification of nucleic acid from live cells, amplified nucleic acid can be subjected to a detection process, such as quantitative PCR or a probe-based analysis. In some embodiments, amplified nucleic acids are hybridized with a plurality of probes.

[0039] In some embodiments, detection of amplified nucleic acid comprises detection of rRNA or DNA encoding rRNA, such as 16S rRNA. Molecular methods based on detecting and analyzing 16S rRNA have been developed to enable the identification of specific microorganism in a mixed bacterial population. Among such methods, DNA microarray analysis is significantly more sensitive and robust than traditional rRNA gene sequencing methods. A 16S rRNA-based phylogenetic microarray ("PhyloChip") allows detailed measurements of microbial community composition in a high-throughput and reproducible manner. Phylogenetic microarrays and methods of making and using the same are described in U.S. Application Ser. No. 61/259,565 [Attorney Docket No. IB-2733P1], filed on Nov. 9, 2009; U.S. Application Ser. No. 61/317,644 [Attorney Docket No. IB-2733P2], filed on Mar. 25, 2010; U.S. Application Ser. No. 61/347,817 [Attorney Docket No. IB-2733P3], filed on May 24, 2010; and co-pending U.S. patent application Ser. No. 12/474,204, filed on May 28, 2009, and co-pending international application having application number PCT/US2010/040106 [Attorney Docket No. IB-2733PCT], filed on Jun. 25, 2010, all of which are hereby incorporated by reference in their entirety for all purposes.

[0040] Features of phylogenetic microarrays disclosed herein include the use of multiple oligonucleotide probes for every known category of prokaryotic organisms for high-confidence detection, and the pairing of at least one mismatch probe for every perfectly matched probe to minimize the effect of nonspecific hybridization. In some embodiments, each perfect match probe corresponds to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more mismatch probes. These and other features, alone or in combination as described herein, make arrays of the methods and systems disclosed herein extremely sensitive, allowing identification of very low levels of microorganisms.

Biosignatures

[0041] In one aspect, the present application utilizes a biosignature of OTUs. As used herein, the term "biosignature" refers to an association of the level of one or more members of

one or more OTUs with a particular condition or state, such as a classification, diagnosis, prognosis, and/or predicted outcome of a disease condition in a subject. In some embodiments, the biosignature comprises a determination of the presence, absence, relative abundance, and/or quantity of at least 1, 2, 3, 4, 5, 10, 20, 50, 100, 250, 500, 1000, 5000, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 250,000, 500,000 or 1,000,000 OTUs in a sample using a single assay. In some embodiments, the biosignature comprises the presence of or changes in the level of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 75, 100, 125, 150, 175, 200, 250, 300, or more OTUs. In some embodiments, the OTUs consist essentially of viable organisms from a sample.

[0042] In some embodiments, the biosignature is associated with a single condition. In some embodiments, the biosignature is associated with a combination of conditions. A biosignature can be obtained for any sample, including but not limited to: tissue samples; cell culture samples; bacterial culture samples; samples obtained from a subject, including biopsies, body fluids and other excreted material; pulmonary samples; environmental samples; other samples as described herein; materials derived therefrom; and combinations thereof. In some embodiments, a biosignature of a test sample is compared to a known biosignature, and a determination is made as to likelihood that the biosignatures are the same. The biosignature to which the biosignature of the test sample is compared can be determined before, after, or at substantially the same time as that of the test sample. Biosignatures can be the result of one or more analyses of one or more samples from a particular source. In some embodiments, a biosignature is indicative of a response to treatment. In some embodiments, a biosignature is used as a basis for the selection of a mode of treatment.

[0043] In some embodiments, the biosignature of a test sample is a combination of two or more independent biosignatures, such as 2, 3, 4, 5, 6, 7, 8, 9, 10 or more independent biosignatures. In some embodiments, each of the two or more biosignatures contained in a sample are assayed simultaneously. In a further embodiment, a subset of biosignatures can be evaluated through the use of low-density detection systems, comprising the determination of the presence, absence, relative abundance, and/or quantity of no more than 10, 25, 50, 100, 250, 500, 1000, 2000, or 5000 OTUs.

[0044] In some embodiments, a biosignature comprises a measure of the number of members, such as the number of live members, in one or more bacterial families or OTUs. The number of members can range from 0 to 10000 or more, such as 0 to 5000, 0 to 2500, 0 to 1000, 0 to 2000, 0 to 1000, 0 to 900, 0 to 800, 0 to 700, 0 to 600, 0 to 500, 0 to 400, 0 to 300, 0 to 200, 0 to 100, 0 to 50, 0 to 25, 0 to 20, 0 to 10, or 0 to 5. In some embodiments, a biosignature comprises the presence of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 2500, 5000, 10000, or more members of one or more bacterial families or OTUs, or the presence of a range that includes any two of these values as end points. In some embodiments, a biosignature comprises a ratio between numbers of members in two or more bacterial families or OTUs. The numerator and denominator of such ratios can include overlapping sets of bacterial families or OTUs. Ratios of the numbers of members in two or more bacterial families can compare a first set of one or more bacterial families or OTUs to a second set of one or more bacterial families or OTUs, where there is at least one bacterial family or OTU difference between the first and second set. A set of bacterial families or OTUs can include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or more OTUs.

[0045] In one aspect, the present application provides methods, systems, and compositions for detecting and identifying a plurality of biomolecules and/or organisms in a sample. The present application utilizes the ability to differentiate between individual organisms or OTUs, as well as, in some embodiments, between live and dead members of OTUs. In one aspect, the individual organisms or OTUs are identified using organism-specific and/or OTU-specific probes, e.g., oligonucleotide probes. More specifically, some embodiments relate to selecting organism-specific and/or OTU-specific oligonucleotide probes useful in detecting and identifying biomolecules and organisms in a sample. In some embodiments, an oligonucleotide probe is selected on the basis of the cross-hybridization pattern of the oligonucleotide probe to regions within a target oligonucleotide and its homologs in a plurality of organisms. The homologs can have nucleotide sequences that are at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.5% identical. Such oligonucleotides can be gene, or intergenetic sequences, in whole or a portion thereof. The oligonucleotides can range from 10 to over 10,000 nucleotides in length. In some other embodiments, a method is provided for detecting the presence of an OTU in a sample based at least partly on the cross-hybridization of the OTU-specific oligonucleotide probes to probes specific for other organisms or OTUs. In some embodiments, the biosignature to which a sample biosignature is compared comprises a positive result for the presence of the targets for one or more probes.

[0046] In one aspect, the present application provides a diagnostic system for the determination or evaluation of a biosignature of a sample. In some embodiments, the diagnostic system comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 75, 100, 125, 150, 175, 200, 250, 300, or more probes. In some embodiments, the diagnostic system comprises up to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 75, 100, 125, 150, 175, 200, 250, 300, or more probes.

High Capacity Systems

[0047] In one aspect of the present application, a high capacity system is provided for determining a biosignature of a sample by assessing the total microorganism population of a sample in terms of the microorganisms present and optionally their percent composition of the total population. In some embodiments, the microorganisms so determined consist essentially of viable organisms from a sample. In some embodiments, the system comprises a plurality of probes that are capable of determining the presence or quantity of at least 1000, 5000, 10000, 20000, 30000, 40000, 50000, 60000, or more different OTUs in a single assay. Typically, the probes selectively hybridize to a highly conserved polynucleotide. Usually, the probes hybridize to the same highly conserved polynucleotide or within a portion thereof. Generally, the highly conserved polynucleotide or fragment thereof comprises a gene or fragment thereof. Non-limiting examples of highly conserved polynucleotides comprise nucleotide sequences found in the 16S rRNA gene, 23S rRNA gene, 5S rRNA gene, 5.8S rRNA gene, 12S rRNA gene, 18S rRNA gene, 28S rRNA gene, gyrB gene, rpoB gene, fusA gene, recA gene, coxl gene and/or nifD gene. In other embodi-

ments, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, 15 or more, 20 or more, 25 or more, or 50 or more collections of probes are employed, each of which specifically hybridizes to a different highly conserved polynucleotide or portion thereof. For example, a first collection of probes binds to the same region of the 16S rRNA gene, a second collection of probes binds to the same region of the 16S rRNA gene that is different from the region bound by probes in the first collection, and a third collection of probes binds to the same region of the 23S rRNA gene. The use of two or more collections of probes where each collection recognizes distinct and separate highly conserved polynucleotides or portions thereof allows for the generation and testing of more probes the use of which can provide greater discrimination between species or OTUs.

[0048] Highly conserved polynucleotides usually show at least 80%, 85%, 90%, 92%, 94%, 95%, or 97% homology across a domain, kingdom, phylum, class, order, family or genus, respectively. The sequences of these polynucleotides can be used for determining evolutionary lineage or making a phylogenetic determination and are also known as phylogenetic markers. In some embodiments, a biosignature comprises the presence, absence, and/or abundance of a combination of phylogenetice markers. The OTUs detected by the probes disclosed herein can be bacterial, archeal, fungal, or eukaryotic in origin. Additionally, the methodologies disclosed herein can be used to quantify OTUs that are bacterial, archaeal, fungal, or eukaryotic. By combining the various probe sets, a system for the detection of bacteria, archaea, fungi, eukaryotes, or combinations thereof can be designed. Such a universal microorganism test that is conducted as a single assay can provide great benefit for assessing and understanding the composition and ecology of numerous environments, including characterization of biosignatures for various samples, environments, conditions, and contaminants.

[0049] In another aspect of the present application, a system is provided that is capable of determining the probability of presence and optionally quantity of at least 10000, 20000, 30000, 40000, 50000, 60000, or more different OTUs of a single domain in a single assay. In some embodiments, such a system makes a probability determination with a confidence level greater than 90%, 91%, 92%, 93%, 94%, 95%, 99% or 99.5%. In some embodiments, a biosignature can comprise the combined result of each probability determination.

[0050] Some embodiments provide a method of selecting an oligonucleotide probe that is specific for a node in a clustering tree. In some embodiments, the method comprises selecting a highly conserved target polynucleotide and its homologs for a plurality of organisms; clustering the polynucleotides and homologs of the plurality of organisms into a clustering tree; and determining a cross-hybridization pattern of a candidate oligonucleotide probe that hybridizes to a first polynucleotide to each node on the clustering tree. This determination is performed (e.g., in silico) to determine the likelihood that the probe would cross hybridize with homologs of its target complementary sequence. The candidate oligonucleotide probe can be complementary to a highly conserved target polynucleotide, a fragment of the highly conserved target or one of its homologs in one of the plurality of organisms. In some embodiments, a method is provided for the determination of the cross-hybridization pattern of a variant of the candidate oligonucleotide probe to each node on the clustering tree, wherein the variant corresponds to the candidate oligonucleotide probe but comprises at least 1 nucleotide mismatch; and selecting or rejecting the candidate oligonucleotide probe on the basis of the cross-hybridization pattern of the candidate oligonucleotide probe and the cross-hybridization pattern of the variant. In some embodiments, the node is an operational taxon unit (OTU). In some embodiments, the node is a single organism.

[0051] Some embodiments provide a method of selecting an OTU-specific oligonucleotide probe for use in detecting a plurality of organisms in a sample. In some embodiments, the method comprises: selecting a highly conserved target polynucleotide and its homologs from the plurality of organisms; clustering the polynucleotides of the target gene and its homologs from the plurality of organisms into one or more operational taxonomic units (OTUs), wherein each OTU comprises one or more groups of similar nucleotide sequence; determining the cross-hybridization pattern of a candidate OTU-specific oligonucleotide probe to the OTUs, wherein the candidate OTU-specific oligonucleotide probe corresponds to a fragment of the target gene or its homolog from one of the plurality of organisms; determining the cross-hybridization pattern of a variant of the candidate OTU-specific oligonucleotide probe to the OTUs, wherein the variant comprises at least 1 nucleotide mismatch from the candidate OTU-specific oligonucleotide probe; and selecting or rejecting the candidate OTU-specific oligonucleotide probe on the basis of the cross-hybridization pattern of the candidate OTU-specific oligonucleotide probe and the cross-hybridization pattern of the variant. In some embodiments, the candidate OTU-specific oligonucleotide probe is selected if the candidate OTU-specific oligonucleotide probe does not cross-hybridize with any polynucleotide that is complementary to probes from other OTUs. In further embodiments, the candidate OTU-specific oligonucleotide probe is selected if the candidate OTU-specific oligonucleotide probe cross-hybridizes with the polynucleotide in no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50, 100, 200, 500, or 1000 other OTU groups.

[0052] Some embodiments provide a method of selecting a set of organism-specific oligonucleotide probes for use in detecting a plurality of organisms in a sample. In some embodiments, the method comprises: identifying a highly conserved target polynucleotide and its homologs in the plurality of organisms; determining the cross-hybridization pattern of a candidate organism-specific oligonucleotide probe to the sequences of the highly conserved target polynucleotide and its homologs in the plurality of organisms, wherein the candidate oligonucleotide probe corresponds to a fragment of the target sequence or its homolog from one of the plurality of organisms; determining the cross-hybridization pattern of a variant of the candidate organism-specific oligonucleotide probe to the sequences of the highly conserved target sequence and its homologs in the plurality of organisms, wherein the variant comprises at least 1 nucleotide mismatch from the candidate organism-specific oligonucleotide probe; and selecting or rejecting the candidate organism-specific oligonucleotide probe on the basis of the cross-hybridization pattern of the candidate organism-specific oligonucleotide probe and the cross-hybridization pattern of the variant of the candidate organism-specific oligonucleotide probe.

[0053] In some embodiments, an OTU-specific oligonucleotide probe does not cross-hybridize with any polynucleotide that is complementary to probes from other OTUs. In other

embodiments, an OTU-specific oligonucleotide probe cross-hybridizes with the polynucleotide in no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50, 100, 200, 500, or 1000 other OTU groups. Some embodiments utilize a set of organism-specific oligonucleotide probes for use in detecting a plurality of organisms in a sample. In further embodiments, the candidate organism-specific oligonucleotide probe is selected if the candidate organism-specific oligonucleotide probe only hybridizes with the target nucleic acid molecule of no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50 unique organisms in the plurality of organisms. In other embodiments, the process is iterative with multiple candidate specific-specific oligonucleotide probes selected. Frequently, the selected organism-specific oligonucleotide probes are clustered and aligned into groups of similar sequences that allow for the detection of an organism with high confidence based on no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 50, or 60 organism-specific oligonucleotide probe matches per OTU. Generally, the candidate organism that the organism-specific oligonucleotide probes detect corresponds to a leaf or node of at least one phylogenetic, genealogic, evolutionary, or taxonomic tree. Knowledge of the position that a candidate organism detected by the organism-specific oligonucleotide probe occupies on a tree provides relational information of the organism to other members of its domain, phylum, class, subclass, order, family, subfamily, or genus.

[0054] In some embodiments, the method disclosed herein selects and/or utilizes a set of organism-specific oligonucleotide probes that are a hierarchical set of oligonucleotide probes that can be used to detect and differentiate a plurality of organisms. In some embodiments, the method selects and/or utilizes organism-specific or OTU-specific oligonucleotide probes that allow a comprehensive screen for at least 80%, 85%, 90%, 95%, 99% or 100% of all known bacterial or archaeal taxa in a single analysis, and thus provides an enhanced detection of different desired taxonomic groups. In some embodiments, the identity of all known bacterial or archaeal taxa comprises taxa that were previously identified by the use of oligonucleotide specific probes, PCR cloning, and sequencing methods. Some embodiments provide methods of selecting and/or utilizing a set of oligonucleotide probes capable of correctly categorizing mixed target nucleic acid molecules into their proper operational taxonomic unit (OTU) designations. Such methods can provide comprehensive prokaryotic or eukaryotic identification, and thus comprehensive biosignature characterization.

[0055] In some embodiments, the selected OTU-specific oligonucleotide probe is used to calculate the relative abundance of one or more organisms that belong to a specific OTU at differing levels of taxonomic identification. In some embodiments, an array or collection of microparticles comprising at least one organism-specific or OTU-specific oligonucleotide probe selected by the method disclosed herein is provided to infer specific microbial community activities. For example, the identity of individual taxa in a microbial consortium from an anaerobic environment for instance, a marsh, can be determined along with their relative abundance. If the consortium is suspected of harboring microorganisms capable of butanol fermentation, then after providing a suitable feedstock in an anaerobic environment if the production of butanol is noted, then those taxa responsible for butanol fermentation can be inferred by the microorganisms that have abundant quantities of 16S rRNA. The present application provides methods to measure taxa abundance based on the detection of directly labeled 16S rRNA.

[0056] In some embodiments, multiple probes are selected for increasing the confidence level and/or sensitivity level of identification of a particular organism or OTU. The use of multiple probes can greatly increase the confidence level of a match to a particular organism. In some embodiments, the selected organism-specific oligonucleotide probes are clustered and aligned into groups of similar sequence such that detection of an organism is based on 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35 or more oligonucleotide probe matches. In some embodiments, the oligonucleotide probes are specific for a species. In other embodiments, the oligonucleotide probe recognizes related organisms such as organisms in the same subgenus, genus, subfamily, family, sub-order, order, sub-class, class, sub-phylum, phylum, sub-kingdom, or kingdom.

[0057] Perfect match (PM) probes are perfectly complementary to the target polynucleotide, e.g., a sequence that identifies a particular organism or OTU. In some embodiments, a system comprises mismatch (MM) control probes. Usually, MM probes are otherwise identical to PM probes, but differ by one or more nucleotides. Probes with one or more mismatch can be used to indicate non-specific binding and a possible non-match to the target sequence. In some embodiments, the MM probes have one mismatch located in the center of the probe, e.g., in position 13 for a 25mer probe. The MM probe is scored in relation to its corresponding PM probe as a "probe pair." MM probes can be used to estimate the background hybridization, thereby reducing the occurrence of false positive results due to non-specific hybridization, a significant problem with many current detection systems. If an array is used, such as an Affymetrix high density probe array or Illumina bead array, ideally, the MM probe is positioned adjacent or close to its corresponding PM probe on the array. Sample PM and MM probes are provided as SEQ ID NOs: 1-50.

[0058] In general, each MM probe differs from its corresponding PM probe by at least one nucleotide. In some embodiments, the MM probe differs from its corresponding PM probe by 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides. Within a MM probe, the mismatched nucleotide or nucleotides can include any of the 3 central bases that are not found in the same position or positions in the PM probe. For example, with a 25mer PM probe that has a guanine at the 13^{th} position, i.e., the central nucleotide, the MM probes comprise probes with adenine, thymine, uracil or cytosine at the 13^{th} position. Similarly, with a 25mer PM probe with an adenine at the 12^{th} nucleotide position and a guanine at the 13^{th} nucleotide position when read from the 3' direction, the possible MM probes comprise probes with guanine at the 12^{th} nucleotide and adenine, thymine or cytosine at the 13^{th} nucleotide position; cytosine at the 12^{th} nucleotide position and adenine, thymine or cytosine at the 13^{th} nucleotide position; and thymine at the 12^{th} nucleotide position and adenine, thymine or cytosine at the 13^{th} nucleotide position. In some embodiments, the mismatched nucleotide or nucleotides include any one or more of the nucleotides in a corresponding PM probe. Increasing the number of MM probes and/or the mis-match positions represented can be used to enhance quantification, accuracy, and confidence.

[0059] Some embodiments relate to a method of selecting and/or utilizing a set of oligonucleotide probes that enable simultaneous identification of multiple prokaryotic taxa with a relatively high confidence level. Typically, the confidence level of identification is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.5%. In general, an OTU refers to an individual species or group of highly related species that share an average of at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 99.5%, or more sequence homology in a highly conserved region. Multiple MM probes can be utilized to enhance the quantification and confidence of the measure. In some embodiments, each interrogation probe (e.g. PM probe) of a plurality of interrogation probes has from about 1 to about 20 or more corresponding mismatch control probes, such as from 2 to 20, 3 to 20, 4 to 20, 5 to 20, 6 to 20, 7 to 20, 8 to 20, 9 to 20, 10 to 20, 11 to 20, 12 to 20, 13 to 20, 14 to 20, 15 to 20, 16 to 20, 17 to 20, 18 to 20, or 19 to 20 corresponding mismatch control probes. In further embodiments, each interrogation probe has from about 1 to about 10, about 1 to about 5, about 1 to 4, 1 to 3, 2 or 1 corresponding mismatch probes. In some embodiments, each interrogation probe has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more corresponding mismatch probes. These interrogation probes target unique regions within a target nucleic acid sequence, e.g., a 16S rRNA gene, and provide the means for identifying at least about 10, 20, 50, 100, 500, 1000, 2000, 5000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 250000, 500000, or 1000000 taxa. In some embodiments, multiple targets can be simultaneously assayed or detected in a single assay through a high-density oligonucleotide probe system. The sum of all target hybridizations is used to identify specific prokaryotic taxa. The result is a more efficient and less time consuming method of identifying unculturable or unknown organisms. The present application can also provide results that could not previously be achieved, e.g., providing results in hours where other methods would require days. In some embodiments, a microbiome (i.e., sample) can be assayed to determine the identity and optionally the abundance of its constituent microorganisms in less than 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 hour.

[0060] In some embodiments, the set of OTU-specific oligonucleotide probes comprises from about 1 to about 500 probes for each taxonomic group. In some embodiments, the probes are proteins including antibodies, or nucleic acid molecules including oligonucleotides or fragments thereof. In some embodiments, an oligonucleotide probe corresponds to a nucleotide fragment of the target nucleic acid molecule. In some embodiments, from about 1 to about 500, about 2 to about 200, about 5 to about 150, about 8 to about 100, about 10 to about 35, or about 12 to about 30 oligonucleotide probes can be designed for each taxonomic grouping. In other embodiments, a taxonomic group can have at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, or more probes. In some embodiments, various taxonomic groups can have different numbers of probes, while in other embodiments, all taxonomic groups have a fixed number of probes per group. Multiple probes in a taxonomic group can provide additional data that can be used to make a determination, also known as "making a call" as to whether an OTU is present or not. Multiple probes also allow for the removal of one or more probes from the analysis based on insufficient signal strength, cross hybridization or other anomalies. Removing probes can increase the confidence level of results and further allow for the detection of low abundant microorganisms. The oligonucleotide probes can each be from about 5 to about 100 nucleotides, from about 10 to about 50 nucleotides, from about 15 to about 35 nucleotides, or from about 20 to about 30 nucleotides. In some embodiments, the probes are at least 5-mers, 6-mers, 7-mers, 8-mers, 9-mers, 10-mers, 11-mers, 12-mers, 13-mers, 14-mers, 15-mers, 16-mers, 17-mers, 18-mers, 19-mers, 20-mers, 21-mers, 22-mers, 23-mers, 24-mers, 25-mers, 26-mers, 27-mers, 28-mers, 29-mers, 30-mers, 31-mers, 32-mers, 33-mers, 34-mers, 35-mers, 36-mers, 37-mers, 38-mers, 39-mers, 40-mers, 41-mers, 42-mers, 43-mers, 44-mers, 45-mers, 46-mers, 47-mers, 48-mers, 49-mers, 50-mers, 51-mers, 52-mers, 53-mers, 54-mers, 55-mers, 56-mers, 57-mers, 58-mers, 59-mers, 60-mers, 61-mers, 62-mers, 63-mers, 64-mers, 65-mers, 66-mers, 67-mers, 68-mers, 69-mers, 70-mers, 71-mers, 72-mers, 73-mers, 74-mers, 75-mers, 76-mers, 77-mers, 78-mers, 79-mers, 80-mers, 81-mers, 82-mers, 83-mers, 84-mers, 85-mers, 86-mers, 87-mers, 88-mers, 89-mers, 90-mers, 91-mers, 92-mers, 93-mers, 94-mers, 95-mers, 96-mers, 97-mers, 98-mers, 99-mers, 100-mers or combinations thereof.

[0061] Some embodiments provide methods of selecting multiple, confirmatory, organism-specific or OTU-specific probes to increase the confidence of detection. In some embodiments, the methods also select one or more mismatch (MM) probes for every perfect match (PM) probe to minimize the effect of cross-hybridization by non-target regions. The organism-specific and OTU-specific oligonucleotide probes selected by the methods disclosed herein can simultaneously identify thousands of taxa present in an environmental sample and allow accurate identification of microorganisms and their phylogenetic relationships in a community of interest. Systems that use the organism-specific and OTU-specific oligonucleotide probes selected by the methods disclosed herein and the computational analysis disclosed herein have numerous advantages over rRNA gene sequencing techniques. Such advantages include reduced cost per microbiome analysis, and increased processing speed per sample or microbiome from both the physical analysis and the computational analysis point of view. In general, the analysis procedures are not adversely affected by chimeras, are not subject to creating artificial phylotypes, and are not subject to barcode PCR bias. Additionally, quantitative standards can be run with a microbiome sample disclosed herein.

[0062] Some embodiments provide a method for selecting and/or utilizing a set of OTU- or organism-specific oligonucleotide probes for use in an analysis system or bead multiplex system for simultaneously detecting a plurality of organisms in a sample. The method targets known diversity within target nucleic acid molecules to determine microbial community composition and establish a biosignature. The target nucleic acid molecule is typically a highly conserved polynucleotide. In some embodiments, the highly conserved polynucleotide is from a highly conserved gene, whereas in other embodiments the polynucleotide is from a highly conserved region of a gene with moderate or large sequence variation. In further embodiments, the highly conserved region can be an intron, exon, or a linking section of nucleic acid that separates two genes. In some embodiments, the highly conserved polynucleotide is from a "phylogenetic" gene. Phylogenetic genes include, but are not limited to, the 5.8S rRNA gene, 12S rRNA gene, 16S rRNA gene-prokary-

otic, 16S rRNA gene-mitochondrial, 18S rRNA gene, 23S rRNA gene, 28S rRNA gene, gyrB gene, rpoB gene, fusA gene, recA gene, coxl gene, and the nifD gene. With eukaryotes, the rRNA gene can be nuclear, mitochondrial, or both. In some embodiments, the 16S-23S rRNA gene internal transcribed spacer (ITS) can be used for differentiation of closely related taxa with or without the use of other rRNA genes. For example, rRNA, e.g., 16S or 23S rRNA, acts directly in the protein assembly machinery as a functional molecule rather than having its genetic code translated into protein. Due to structural constraints of 16S rRNA, specific regions throughout the gene have a highly conserved polynucleotide sequence; although, non-structural segments can have a high degree of variability. Probing the regions of high variability can be used to identify OTUs that represent a single species level, while regions of less variability can be used to identify OTUs that represent a subgenus, a genus, a subfamily, a family, a sub-order, an order, a sub-class, a class, a subphylum, a phylum, a sub-kingdom, or a kingdom. The methods disclosed herein can be used to select organism-specific and OTU-specific oligonucleotide probes that offer a high level of specificity for the identification of specific organisms, OTUs representing specific organisms, or OTUs representing specific taxonomic group of organisms. The systems and methods disclosed herein are particularly useful in identifying closely related microorganisms and OTUs from a background or pool of closely related organisms.

[0063] The probes selected and/or utilized by the methodologies disclosed herein can be organized into OTUs that provide an assay with a sensitivity and/or specificity of more than 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%. In some embodiments, sensitivity and specificity depends on the hybridization signal strength, number of probes in the OTU, the number of potential cross hybridization reactions, the signal strength of the mismatch probes, if present, background noise, or combinations thereof. In some embodiments, an OTU containing one probe can provide an assay with a sensitivity and specificity of at least 90%, while another OTU can require at least 20 probes to provide an assay with sensitivity and specificity of at least 90%.

[0064] Some embodiments relate to methods for phylogenetic analysis system design and signal processing and interpretation for use in detecting and identifying a plurality of biomolecules and organisms in a sample. More specifically, some embodiments relate to a method of selecting a set of organism-specific oligonucleotide probes for use in detecting a plurality of organisms in a sample with a high confidence level. Some embodiments relate to a method of selecting a set of OTU-specific oligonucleotide probes for use in detecting a plurality of organisms in a sample with a high confidence level.

[0065] In the case of highly conserved polynucleotides like 16S rRNA that can have only one to a few nucleotides of sequence variability over any 15- to 30-bp region targeted by probes for discrimination between related microbial species, it is advantageous to maximize the probe-target sequence specificity in an assay system. Some embodiments provide methods of selecting organism-specific oligonucleotide probes that effectively minimize the influence of cross-hybridization. In some embodiments, the method comprises: (a) identifying sequences of a target nucleic acid molecule corresponding to the plurality of organisms; (b) determining the cross-hybridization pattern of a candidate organism-specific oligonucleotide probe to the target nucleic acid molecule from the plurality of organisms, wherein the candidate oligonucleotide probe corresponds to a sequence fragment of the target nucleic acid molecule from the plurality of organisms; (c) determining the cross-hybridization pattern of a variant of the candidate organism-specific oligonucleotide probe to the target nucleic acid molecule from the plurality of organisms, wherein the variant of the candidate organism-specific oligonucleotide probe comprises at least 1 nucleotide mismatch compared to the candidate organism-specific oligonucleotide probe; and (d) selecting or rejecting the candidate organism-specific oligonucleotide probe on the basis of the cross-hybridization pattern of the candidate organism-specific oligonucleotide probe and the cross-hybridization pattern of the variant of the candidate organism-specific oligonucleotide probe. In some embodiments, a method of selecting a set of OTU-specific oligonucleotide probes for use in detecting a plurality of organisms in a sample is provided. In some embodiments, the method comprises: (a) identifying sequences of a target nucleic acid molecule corresponding to the plurality of organisms; (b) clustering the sequences of the target nucleic acid molecule from the plurality of organisms into one or more Operational Taxonomic Units (OTUs), wherein each OTU comprises one or more groups of similar sequences; (c) determining the cross-hybridization pattern of a candidate OTU-specific oligonucleotide probe to the OTUs, wherein the candidate OTU-specific oligonucleotide probe corresponds to a sequence fragment of the target nucleic acid molecule from one of the plurality of organisms; (d) determining the cross-hybridization pattern of a variant of the candidate OTU-specific oligonucleotide probe to the OTUs, wherein the variant of the candidate OTU-specific oligonucleotide probe comprises at least 1 nucleotide mismatch compared to the candidate OTU-specific oligonucleotide probe; and (e) selecting or rejecting the candidate OTU-specific oligonucleotide probe on the basis of the cross-hybridization pattern of the candidate OTU-specific oligonucleotide probe to the OTUs and the cross-hybridization pattern of the variant of the candidate OTU-specific oligonucleotide probe to the OTUs. In some embodiments, candidate OTU-specific oligonucleotide probe are rejected when the candidate OTU-specific oligonucleotide probe or its variant are predicted to cross-hybridize with other target sequences. In some embodiments, a predetermined amount of predicted cross-hybridization is allowed.

[0066] In some embodiments, selected oligonucleotide probes are synthesized by any relevant method known in the art. Some examples of suitable methods include printing with fine-pointed pins onto glass slides, photolithography using pre-made masks, photolithography using dynamic micromirror devices, ink jet printing, or electrochemistry. In one example, a photolithographic method can be used to directly synthesize the chosen oligonucleotide probes onto a surface. Suitable examples for the surface include glass, plastic, silicon and any other surface available in the art. In certain examples, the oligonucleotide probes can be synthesized on a glass surface at an approximate density from about 1,000 probes per μm^2 to about 100,000 probes per μm^2, preferably from about 2000 probes per μm^2 to about 50,000 probes per μm^2, more preferably from about 5000 probes per μm^2 to about 20,000 probes per μm^2. In one example, the density of the probes is about 10,000 probes per μm^2. The number of probes on the array can be quite large e.g., at least 10^5, 10^6, 10^7, 10^8 or 10^9 probes per array. Usually, for large arrays only

a relatively small proportion (i.e., less than about 1%, 0.1% 0.01%, 0.001%, 0.00001%, 0.000001% or 0.0000001%) of the total number of probes of a given length target an individual OTU. Frequently, lower limit arrays have no more than 10, 25, 50, 100, 500, 1000, 5000, or 10000, 25000, 50000, 100000 or 250000 probes.

[0067] Typically, the arrays or microparticles have probes to one or more highly conserved polynucleotides. The arrays or microparticles can have further probes (e.g. confirmatory probes) that hybridize to functionally expressed genes, thereby providing an alternate or confirmatory signal upon which to base the identification of a taxon. For example, an array can contain probes to 16S rRNA gene sequences from *Yersinia pestis* and *Vibrio cholerae* and also confirmatory probes to *Y. pestis* caf1 virulence gene or *V. cholerae* zonula occludens toxin (zot) gene. The detection of hybridization signals based on probes binding to 16S rRNA polynucleotides associated with a particular OTU coupled with the detection of a hybridization signal based on a confirmatory probe can provide a higher level of confidence that the OTU is present. For instance, if hybridization signals are detected for the probes associated *Y. pestis* OTU and the confirmatory probe also displays a hybridization signal for the expression of *Y. pestis* caf1 then the confidence level subscribed to the presence or quantity of *Y. pestis* will be higher than the confidence level obtained from the use of OTU probes alone.

[0068] A range of lengths of probes can be employed on the arrays or microparticles. As noted above, a probe can consist exclusively of a complementary segments, or can have one or more complementary segments juxtaposed by flanking, trailing and/or intervening segments. In the latter situation, the total length of complementary segment(s) can be more important that the length of the probe. In functional terms, the complementary segment(s) of the PM probes should be sufficiently long to allow the PM probes to hybridize more strongly to a target polynucleotide e.g., 16S rRNA, compared with a MM probe. A PM probe usually has a single complementary segment having a length of at least 15 nucleotides, and more usually at least 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 30 bases exhibiting perfect complementarity.

[0069] In some arrays of microparticles, all probes are the same length. In other arrays of microparticles, probe length varies between quantification standard (QS) probes, negative control (NC) probes, probe pairs, probe sets (OTUs) and combinations thereof. For example, some arrays can have groups of OTUs that comprise probe pairs that are all 23 mers, together with other groups of OTUs or probe sets that comprise probe pairs that are all 25 mers. Additional groups of probes pairs of other lengths can be added. Thus, some arrays can contain probe pairs having sizes of 15 mers, 16mers, 17mers, 18mers, 19mers, 20mers, 21mers, 22mers, 23mers, 24mers, 25 mers, 26mers, 27 mers, 28mers, 29 mers, 30mers, 31mers, 32mers, 33mers, 34mers, 35mers, 36mers, 37mers, 38mers, 39mers, 40mers or combinations thereof. Other arrays can have different size probes within the same group, OTU, or probe set. In these arrays, the probes in a given OTU or probe set can vary in length independently of each other. Having different length probes can be used to equalize hybridization signals from probes depending on the hybridization stability of the oligonucleotide probe at the pH, temperature, and ionic conditions of the reaction.

[0070] In another aspect of the present application, a system is provided for determining the presence or quantity of a plurality of different OTUs in a single assay where the system comprises a plurality of polynucleotide interrogation probes, a plurality of polynucleotide positive control probes, and a plurality of polynucleotide negative control probes. In some embodiments, the system is capable of detecting the presence, absence, relative abundance, and/or quantity of at least about 5, 10, 20, 50, 100, 250, 500, 1000, 5000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 250000, 500000 or 1000000 OTUs in a sample using a single assay. In some embodiments, the polynucleotide positive control probes include 1) probes that target sequences of prokaryotic or eukaryotic metabolic genes spiked into the target nucleic acid sequences in defined quantities prior to fragmentation, or 2) probes complimentary to a pre-labeled oligonucleotide added into the hybridization mix after fragmentation and labeling. The control added prior to fragmentation collectively tests the fragmentation, biotinylation, hybridization, staining and scanning efficiency of the system. It also allows the overall fluorescent intensity to be normalized across multiple analysis components used in a single or combined experiment, such as when two or more arrays are used in a single experiment or when data from two separate experiments is combined. The second control directly assays the hybridization, staining and scanning of the system. Both types of control can be used in a single experiment.

[0071] In some embodiments, the QS standards (positive controls) are PM probes. In other embodiments, the QS standards are PM and MM probe pairs. In further embodiments, the QS standards comprise a combination of PM and MM probe pairs and PM probes without corresponding MM probes. In some embodiments, the QS standards comprise at least one, two, three, four, five, six, seven, eight, nine, ten or more MM probes for each corresponding PM probe. In a further embodiment, the QS standards comprise at least one, two, three, four, five, six, seven, eight, nine, ten or more PM probes for each corresponding MM probe. A system can comprise at least 1 positive control probe for each 1, 10, 100, or 1000 different interrogation probes.

[0072] In some cases, the spiked-in oligonucleotides that are complementary to the positive control probes vary in G+C content, uracil content, concentration, or combinations thereof. In some embodiments, the G+C % ranges from about 30% to about 70%, about 35% to about 65% or about 40% to about 60%. QS standards can also be chosen based on the uracil incorporation frequency. The QS standards can incorporate uracil in a range from about 1 in 100 to about 60 in 100, about 4 in 100 to about 50 in 100, or about 10 in 100 to about 50 in 100. In some cases, the concentration of these added oligonucleotides will range over 1, 2, 3, 4, 5, 6, or 7 orders of magnitude. Concentration ranges of about 10^5 to 10^{14}, 10^6 to 10^{13}, 10^7 to 10^{12}, 10^7 to 10^{11}, 10^8 to 10^{11}, and 10^8 to 10^{10} can be employed and generally feature a linear hybridization signal response across the range. In some embodiments, positive control probes for the conduction of the methods disclosed herein comprise polynucleotides that are complementary to the positive control sequences shown in Table 1. Other genes that can be used as targets for positive controls include genes encoding structural proteins, proteins that control growth, cell cycle or reproductive regulation, and housekeeping genes. Additionally, synthetic genes based on highly conserved genes or other highly conserved polynucleotides can be added to the sample. Useful highly conserved genes from which synthetic genes can be designed include 16S rRNA genes, 18S rRNA genes, 23SrRNA genes. Exemplary control probes are provided as SEQ ID NOs:51-100.

TABLE 1

Positive Control Sequences

Positive Control ID	Description
AFFX-BioB-5_at	*E. coli* biotin synthetase
AFFX-BioB-M_at	*E. coli* biotin synthetase
AFFX-BioC-5_at	*E. coli* bioC protein
AFFX-BioC-3_at	*E. coli* bioC protein
AFFX-BioDn-3_at	*E. coli* dethiobiotin synthetase
AFFX-CreX-5_at	Bacteriophage P1 cre recombinase protein
AFFX-DapX-5_at	*B. subtilis* dapB, dihydrodipicolinate reductase
AFFX-DapX-M_at	*B. subtilis* dapB, dihydrodipicolinate reductase
YFL039C	*Saccharomyces*, Gene for actin (Act 1p) protein
YER022W	*Saccharomyces*, RNA polymerase II mediator complex subunit (SRB4p)
YER 148 W	*Saccharomyces*, TATA-binding protein, general transcription factor (SPT15)
YEL002C	*Saccharomyces*, Beta subunit of the oligosaccharyl transferase (OST) glycoprotein complex (WBP1)
YEL024W	*Saccharomyces*, Ubiquinol-cytochrome-c reductase (RIP1)
Synthetic 16S rRNA controls	
SYNM neurolyt_st	Synthetic derivative of *Mycoplasma neurolyticum* 16S rRNA gene
SYNLc.oenos_st	Synthetic derivative of *Leuconostoc oenos* 16S rRNA gene
SYNCau.cres8_st	Synthetic derivative of *Caulobacter crescenius* 16S rRNA gene
SYNFer.nodosm_st	Synthetic derivative of *Fervidobacterium nodosum* 16S rRNA gene
SYNSap.grandi_st	Synthetic derivative of *Saprospira grandis* 16S rRNA gene

[0073] In some embodiments, the negative controls comprise PM and MM probe pairs. In further embodiments, the negative controls comprise a combination of PM and MM probe pairs and PM probes without corresponding MM probes. In other embodiments, the negative control probes comprise at least one, two, three, four, five, six, seven, eight, nine, ten or more MM probes for each corresponding negative control PM probe. A system can comprise at least 1 negative control probe for each 1, 10, 100, or 1000 different interrogation probes (PMs).

[0074] In some embodiments, the negative control probes hybridize weakly, if at all, to 16S rRNA gene or other highly conserved gene targets. The negative control probes can be complementary to metabolic genes of prokaryotic or eukaryotic origin. Generally, with negative control probes, no target material is spiked into the sample. In some embodiments, negative control probes are from the same collection of probes that are also used for positive controls, but no material complementary to the negative control probes are spiked into the sample, in contrast to the positive control probe methodology. In essence, the control probes can be universal control probes and play the role of positive or negative control probes depending on the system's design. One of skill in the art will appreciate that the universal control probes are not limited to highly conserved sequence analysis systems and have applications beyond the present embodiments disclosed herein.

[0075] In a further embodiment, probes to non-highly conserved polynucleotides are added to a system to provide species-specific identification or confirmation of results achieved with the probes to the highly conserved polynucleotides. Usually, these "confirmatory" probes cross hybridize very weakly, if at all, to highly conserved polynucleotides recognized by the perfect match probes. Useful species-specific genes include metabolic genes, genes encoding structural proteins, proteins that control growth, cell cycle or reproductive regulation, housekeeping genes or genes that encode virulence, toxins, or other pathogenic factors. In some embodiments, the system comprises at least 1, 5, 10, 20, 30, 40, 50 60, 70, 80, 90 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 5000 or 10000 species-specific probes.

[0076] In some embodiments, a system comprises an array. Non-limiting examples of arrays include microarrays, bead arrays, through-hole arrays, well arrays, and other arrays known in the art suitable for use in hybridizing probes to targets. Arrays can be arranged in any appropriate configuration, such as, for example, a grid of rows and columns. Some areas of an array comprise the OTU detection probes whereas other areas can be used for image orientation, normalization controls, signal scaling, noise reduction processing, or other analyses. Control probes can be placed in any location in the array, including along the perimeter of the array, diagonally across the array, in alternating sections or randomly. In some embodiments, the control probes on the array comprise probe pairs of PM and MM probes. The number of control probes can vary, but typically the number of control probes on the array range from 1 to about 500,000. In some embodiments, at least 10, 100, 500, 1000, 5000, 10000, 25000, 50000, 100000, 250000 or 500000 control probes are present. When control probe pairs are used, the probe pairs will range from 1 to about 250000 pairs. In some embodiments, at least 5, 50, 250, 500, 2,500, 5,000, 12500, 25000, 50000, 125000 or 250000 control probe pairs are present. The arrays can have other components besides the probes, such as linkers attaching the probes to a support. In some embodiments, materials for fabricating the array can be obtained from Affymetrix (Santa Clara, Calif.), GE Healthcare (Little Chalfont, Buckinghamshire, United Kingdom) or Agilent Technologies (Palo Alto, Calif.)

[0077] Besides arrays where probes are attached to the array substrate, numerous other technologies can be employed in the disclosed system. In some embodiments, the probes are attached to beads that are then placed on an array as disclosed by Ng et al. (Ng et al. A spatially addressable bead-based biosensor for simple and rapid DNA detection. Biosensors & Bioelectronics, 23:803-810, 2008).

[0078] In some embodiments, probes are attached to beads or microspheres, the hybridization reactions are performed in solution, and then the beads are analyzed by flow cytometry, as exemplified by the Luminex multiplexed assay system. In this analysis system, homogeneous bead subsets, each with beads that are tagged or labeled with a plurality of identical probes, are combined to produce a pooled bead set that is hybridized with a sample and then analyzed in real time with flow cytometry, as disclosed in U.S. Pat. No. 6,524,793. Bead subsets can be distinguished from each other by variations in the tags or labels, e.g., using variability in laser excitable dye content.

[0079] In some further embodiments, probes are attached to cylindrical glass microbeads as exemplified by the Illumina Veracode multiplexed assay system. Here, subsets of microbeads embedded with identical digital holographic elements are used to create unique subsets of probe-labeled microbeads. After hybridization, the microbeads are excited by laser light and the microbead code and probe label are read in real time multiplex assay.

[0080] In some embodiments, a solution based assay system is employed as exemplified by the NanoString nCounter Analysis System (Geiss G et al. Direct multiplexed measure-

ment of gene expression with color-coded probe pairs. Nature Biotech. 26:317-325, 2008). With this methodology, a sample is mixed with a solution of reporter probes that recognize unique sequences and capture probes that allow the complexes formed between the nucleic acids in the sample and the reporter probes to be immobilized on a solid surface for data collection. Each reporter probe is color-coded and is detected through fluorescence.

[0081] In a further embodiment, branched DNA technology, as exemplified by Panomics QuantiGene Plex 2.0 assay system, is used. Branched DNA technology comprises a sandwich nucleic acid hybridization assay for RNA detection and quantification that amplifies the reporter signal rather than the sequence. By measuring the RNA at the sample source, the assay avoids variations or errors inherent to extraction and amplification of target polynucleotides. The QuantiGene Plex technology can be combined with multiplex bead based assay system such as the Luminex system described above to enable simultaneous quantification of multiple RNA targets directly from whole cells or purified RNA preparations.

Sample Sources and Nucleic Acid Preparation

[0082] In some embodiments, the sample used can be an ecosystems sample. Ecosystems include microbiomes associated with plants, animals, and humans. Animal and human associated microbiomes include those found in the gastrointestinal tract, respiratory system, nares, urogenital tract, mammary glands, oral cavity, auditory canal, feces, urine, and skin. In some embodiments, the sample can be any kind of clinical or medical sample. For example, samples from blood, urine, feces, nares, the lungs, the gut, other bodily fluids or excretions, materials derived therefrom, or combinations thereof of mammals can be assayed using the array system. Also, the probes selected by the methods disclosed herein and the array system of the present embodiments can be used to identify an infection in the blood of an animal. The probes selected by the methods disclosed herein and the array system of the present embodiments can also be used to assay medical samples that are directly or indirectly exposed to the outside of the body, such as the lungs, ear, nose, throat, the entirety of the digestive system or the skin of an animal. In some embodiments, a sample includes cell culture samples and/or bacterial culture samples. In some embodiments, a sample comprises a pulmonary sample from a subject, including but not limited to sputum, endotracheal aspirate, bronchoalveolar lavage sample, a swab of the endotrachea, materials derived therefrom, or combinations thereof.

[0083] Techniques and systems to obtain nucleic acids from multiple organisms in a sample, such as an ecosystem, medical, or clinical sample, are well known by persons skilled in the art. In some embodiments, a sample is treated with an agent that selectively modifies a nucleic acid of dead cells. In some embodiments, samples treated with an agent that selectively modifies a nucleic acid of dead cells are treated with a second agent, such as light (e.g. visible light), and optionally washed to remove excess selective agent before extracting nucleic acid. Many commercially available DNA extraction and purification kits can be used. Samples with lower than 2 pg purified DNA can require amplification, which can be performed using conventional techniques known in the art, such as a whole community genome amplification (WCGA) method (Wu et al., Appl. Environ. Microbiol. (2006) 72, 4931-4941). In some embodiments, highly conserved sequences such as those found in the 16S RNA gene, 23S RNA gene, 5S RNA gene, 5.8S rRNA gene, 12S rRNA gene, 18S rRNA gene, 28S rRNA gene, gyrB gene, rpoB gene, fusA gene, recA gene, coxl gene and nifD gene are amplified. Usually, amplification is performed using PCR, but other types of nucleic acid amplification can be employed. Generally, amplification is performed using a single pair of universal primers specific to a highly conserved sequence. For redundancy or for increased amount of total amplicon concentration, two or more universal probe pairs each specific to a different highly conserved sequence can be used. Representative PCR primers include: bacterial primers 27F and 1492R. In some embodiments, a nucleic acid sample is amplified using a collection of primers each comprising one or more nucleotide positions selected at random from two or more different nucleotides. In some embodiments, primers, nucleotides, or other reagents used in an amplification reaction are labeled to produced labeled amplification products.

[0084] A gel electrophoresis method can also be used to isolate community RNA (McGrath et al., J. Microbiol. Methods (2008) 75:172-176). Samples with lower than 5 pg purified RNA may require amplification, which can be performed using conventional techniques known in the art, such as a whole community RNA amplification approach (WCRA) (Gao et al., Appl. Environ. Microbiol. (2007) 73:563-571) to obtain cDNA. In some embodiments, sampling and DNA extraction are conducted as previously described (DeSantis et al., Microbial Ecology, 53(3):371-383, 2007).

[0085] In some embodiments, DNA; total RNA, or a fraction thereof, including rRNA, 16S rRNA, and 23S rRNA; or combinations thereof are directly labeled and used without any amplification.

Probe Preparations

[0086] Techniques and means for generating oligonucleotide probes to be used on analysis systems, beads or in other systems are well-known by persons skilled in the art. For example, the oligonucleotide probes can be generated by synthesis of synthetic polynucleotides or oligonucleotides, e.g., using N-phosphonate or phosphoramidite chemistries (Froehler et al., Nucleic Acid Res. 14:5399-5407 (1986); McBride et al., Tetrahedron Lett. 24:246-248 (1983)). Synthetic sequences are typically between about 10 and about 500 bases in length, more typically between about 15 and about 100 bases, and most preferably between about 20 and about 40 bases in length. In some embodiments, synthetic nucleic acids include non-natural bases, such as, but by no means limited to, inosine. An example of a suitable nucleic acid analogue is peptide nucleic acid (see, e.g., Egholm et al., Nature 363:566-568 (1993); U.S. Pat. No. 5,539,083). In some embodiments, at least 10, 25, 50, 100, 500, 1,000, 5,000, 10,000, 20,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000 100,000, 200,000, 500,000, 1,000,000 or 2,000,000 probes are included on the array. In further embodiments, each PM probe has one or more corresponding MM probe present on the array. Typically, each PM-MM probe pair is associated with an OTU. In some embodiments, at least 10, 25, 50, 100, 500, 1,000, 5,000, 10,000, 20,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000 100,000, 200,000 or 500,000 probe pairs are placed on the array. Generally, sets of probe pairs have at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 or 35 probe pairs present.

[0087] In some embodiments, positive control probes that are complementary to particular sequences in the target sequences (e.g., 16S rRNA gene) are used as internal quantification standards (QS) and included in the system. In other embodiments, positive control probes, also known as internal DNA quantification standards (QS) probes are probes that hybridize to spiked-in nucleic acid sequence targets. Usually, the sequences are from metabolic genes. In some embodiments, negative control (NC) probes, e.g., probes that are not complementary or do not appreciably hybridize to sequences in the target sequences (e.g., 16S rRNA gene) are included on the array. Unlike the QS probes, no target material is spiked into the sample mix for the NC probes, prior to sample processing.

[0088] In some embodiments, the probes are synthesized separately and then attached to a solid support or surface, which can be made, e.g., from glass, latex, plastic (e.g., polypropylene, nylon, polystyrene), polyacrylamide, nitrocellulose, gel, silicon, or other porous or nonporous material. In some embodiments, the surface is spherical or cylindrical as in the case of microbeads or rods. In other embodiments, the surface is planar, as in an array or microarray.

[0089] Oligonucleotides produced using techniques known in the art can be built on and/or coupled to microspheres, beads, microbeads, rods, or other microscopic particles for use in arrays, flow cytometry, and other multiplex assay systems. Numerous microparticles are commercially available from about 0.01 to 100 micrometers in diameter. Generally, microparticles from about 0.1-50 μm, about 1-20 μm, or about 3-10 μm are preferred. The size and shapes of microparticles can be uniform or they can vary. In some embodiments, sublots of different sizes, shapes or both are conjugated to probes before combining the sublots to make a final mixed lot of labeled microparticles. The individual sublots can therefore be distinguished and classified based on their size and shape. The size of the microparticles can be measured in practically any flow cytometry apparatus by so-called forward or small-angle scatter light. The shape of the particle can be also discriminated by flow cytometry, e.g., by high-resolution slit-scanning method.

[0090] Microparticles can be made out of any solid or semi-solid material including glass, glass composites, metals, ceramics, or polymers. Frequently, the microparticles are polystyrene or latex material, but any type of polymeric material is acceptable including but not limited to brominated polystyrene, polyacrylic acid, polyacrylonitrile, polyacrylamide, polyacrolein, polybutadiene, polydimethylsiloxane, polyisoprene, polyurethane, polyvinylacetate, polyvinylchloride, polyvinylpyridine, polyvinylbenzylchloride, polyvinyltoluene, polyvinylidene chloride, polydivinylbenzene, polymethylmethacrylate, or combinations thereof. Microparticles can be magnetic or non-magnetic and can also have a fluorescent dye, quantum dot, or other indicator material incorporated into the microparticle structure or attached to the surface of the microparticles. Frequently, microparticles can also contain 1 to 30% of a cross-linking agent, such as divinyl benzene, ethylene glycol dimethacrylate, trimethylol propane trimethacrylate, or N,N'methylene-bis-acrylamide or other functionally equivalent agents known in the art.

[0091] Embodiments disclosed herein are applicable for use in any analysis system, including but not limited to bead or solution multiplex reaction platforms, or across multiple platforms, for example, Affymetrix GeneChip® Arrays, Illumina BeadChip® Arrays, Luminex xMAP® Technology, Agilent Two-Channel Arrays, MAGIChips (Analysis systems of Gel-immobilized Compounds) or the NanoString nCounter Analysis System. The Affymetrix (Santa Clara, Calif., USA) platform DNA arrays can have the oligonucleotide probes (approximately 25mer) synthesized directly on the glass surface by a photolithography method at an approximate density of 10,000 molecules per μm^2 (Chee et al., Science (1996) 274:610-614). Spotted DNA arrays use oligonucleotides that are synthesized individually at a predefined concentration and are applied to a chemically activated glass surface. In general, oligonucleotide lengths can range from a few nucleotides to hundreds of bases in length, but are typically from about 10mer to 50mer, about 15mer to 40mer, or about 20mer to about 30mer in length.

Target Labeling

[0092] In some embodiments, the nucleic acid targets are labeled so that a laser scanner tuned to a specific wavelength of light can measure the number of fluorescent molecules that hybridized to a specific DNA probe. For arrays, the nucleic acid targets are typically fragmented to between 15 and 100 nucleotides in length and a biotinylated nucleotide is added to the end of the fragment by terminal DNA transferase. At a later stage, the biotinylated fragments that hybridize to the oligonucleotide probes are used as a substrate for the addition of multiple phycoerythrin fluorophores by a sandwich (Streptavidin) method. For some arrays, such as those made by AGILENT or NIMBLEGEN, the purified community DNA can be fluorescently labeled by random priming using the Klenow fragment of DNA polymerase and more than one fluorescent moiety can be used (e.g. controls could be labeled with Cy3, and experimental samples labeled with Cy5 for direct comparison by hybridization to a single analysis system). Some labeling methods incorporate the molecular label into the target during an amplification or enzymatic step to produce multiple labeled copies of the target.

[0093] In some embodiments, the detection system is able to measure the microbial diversity of complex communities without PCR amplification, and consequently, without the inherent biases associated with PCR amplification. In some embodiments, nucleic acid from dead cells is selectively removed before detection, such that detection is directed to determining the presence, absence, relative abundance, and/or quantity of live organisms in a sample, such as live members of one or more OTUs. Actively metabolizing cells typically contain about 20,000 or more ribosomes for protein assembly compared to quiescent or dead cells that have few. In some embodiments, rRNA can be purified directly from a sample and processed with no amplification step, thereby reducing or avoiding bias caused by preferential amplification of some sequences over others. Thus, in some embodiments, the signal from the analysis system can reflect the true number of rRNA molecules that are present in the samples. This can be expressed as the number of cells multiplied by the number of rRNA copies within each cell. The number of cells in a sample can then be inferred by several different methods, such as, for example, quantitative real-time PCR, or FISH (fluorescence in situ hybridization.). Then the average number of ribosomes within each cell can be calculated.

Hybridization

[0094] Hybridizations can be carried out under conditions well-known by persons skilled in the art. See Rhee et al.

(Appl. Environ. Microbiol. (2004) 70:4303-4317) and Wu et al. (Appl. Environ. Microbiol. (2006) 72:4931-4941). The temperature can be varied to reduce or increase stringency and allow the detection of more or less divergent sequences. Robotic hybridization and stringency wash stations can be used to give more consistent results and reduce processing time. In some embodiments, the hybridization and washing process can be accomplished in less than about half an hour, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 14 hours, 16 hours, 18 hours, 20 hours or 24 hours. Generally, hybridization and washing times are reduced for microparticle based detection systems owing to the greater accessibility of the probes to the target molecules. Generally, hybridization times can be reduced for low complexity assays and/or assays for which there is an excess of target analytes.

Signal Quantification

[0095] After hybridization, arrays can be scanned using any suitable scanning device. Non-limiting examples of conventional microarray scanners include GeneChip Scanner 3000 or GeneArray Scanner, (Affymetrix, Santa Clara, Calif.); and ProScan Array (Perkin Elmer, Boston, Mass.); and can be equipped with lasers having resolutions of 10 pm or finer. The scanned image displays can be captured as a pixel image, saved, and analyzed by quantifying the pixel density (intensity) of each spot on the array using image quantification software (e.g., GeneChip Analysis system Analysis Suite, version 5.1 Affymetrix, Santa Clara, Calif.; and ImaGene 6.0, Biodiscovery Inc. Los Angeles, Calif., USA). For each probe, an individual signal value can be obtained through imaging parsing and conversion to xy-coordinates. Intensity summaries for each feature can be created and variance estimations among the pixels comprising a feature can be calculated.

[0096] With flow cytometry based detection systems, a representative fraction of microparticles in each sublot of microparticles can be examined. The individual sublots, also known as subsets, can be prepared so that microparticles within a sublot are relatively homogeneous, but differ in at least one distinguishing characteristic from microparticles in any other sublot. Therefore, the sublot to which a microparticle belongs can readily be determined from different sublots using conventional flow cytometry techniques as described in U.S. Pat. No. 6,449,562. Typically, a laser is shined on individual microparticles and at least three known classification parameter values measured: forward light scatter (C_1) which generally correlates with size and refractive index; side light scatter (C_2) which generally correlates with size; and fluorescent emission in at least one wavelength (C_3) which generally results from the presence of fluorochrome incorporated into the labeled target sequence. Because microparticles from different subsets differ in at least one of the above listed classification parameters, and the classification parameters for each subset are known, a microparticle's sublot identity can be verified during flow cytometric analysis of the pool of microparticles in a single assay step and in real-time. For each sublot of microparticles representing a particular probe, the intensity of the hybridization signal can be calculated along with signal variance estimations after performing background subtraction.

Data Processing and Statistical Analysis

[0097] Simultaneous detection of at least 500, 1000, 5000, 10000, 20000, 30000, 40000, 50000, 60000, or more taxa with a high level of confidence can incorporate techniques to de-convolute the signal intensity of numerous probe sets into probability estimates. In some embodiments, the methods, compositions and systems enable detection in one assay the presence or absence of a microorganism in a community of microorganisms, such as an environmental or clinical sample when the microorganism comprises less than 0.05% of the total population of microorganisms. In some embodiments, detection includes determining the quantity of the microorganism, e.g., the percentage of the microorganism in the total microorganism population. De-convolution techniques can include the incorporation of NC probe pairs into the analysis system and the use of the data to fit the hybridization signals from the QS probe pairs to the hybridization distribution of the NC probe pairs.

[0098] De-convolution techniques can allow the detection and quantification of nucleic acids in a sample and by inference, the detection and quantification of microorganisms in a sample. In one aspect of the present application, a system is provided for determining the presence or quantity of a microorganism in a sample comprising contacting a sample with a plurality of probes, detecting the hybridization signals of the sample nucleic acids with the probes and de-convoluting the signals to determine the presence, absence and/or quantity of a particular nucleic acid present in a population of nucleic acids where the particular nucleic acid is present at less than 0.01% of the total nucleic acid population. In some embodiments, the particular nucleic acid is at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96% or 97% homologous to other nucleic acids in the population.

[0099] In some embodiments, the data output from an imaged or scanned sample is de-convoluted and analyzed using the following methods. Using an array as an illustrative example, the hybridization signals are converted to xy-coordinates with intensity summaries and variance estimates generated for the pixels using commercial software. The data is outputted using a standard data format like a CEL file (Affymetrix), or a Feature Report file (NimbleGen).

[0100] The hybridization signals undergo background subtraction. Typically, the background intensity is computed independently for each quadrant as the average signal intensity of the least intense 2% of the probes in the quadrant. Other threshold values can also be used, e.g., 0.5%, 1%, 3%, 4%, 5% or 10%. Background intensity is then subtracted from all probes in a quadrant before further computation is performed. This noise removal procedure can be done on a quadrant-by-quadrant basis or across a whole array.

[0101] In some embodiment, array signals are normalized to allow for the comparison of results achieved in different experiments or for the comparison of replicate experiments. Normalization can be achieved by a number of methods. In some embodiments, reproducibility between different probes for the same target are evaluated using a Position Dependent Nearest Neighbor (PDNN) model as described in Zhang L. et al., A model of molecular interactions on short oligonucleotide analysis systems, Nat. Biotechnol. 2003, 21(7):818-821. The PDNN model allows estimation of the sequence specific noise signal and a non-specific background signal, and thus enables estimation of the true intensity for the probes.

[0102] In other embodiments, per-array models of signal and background distributions using responses observed from comparison of the PM and MM probe pairs and the internal DNA quantification standards (QS) probe pairs are created. In

some embodiments, the probability that each probe pair is "positive" is determined by calculating a difference score, d, for each probe pair. d can be defined as:

$$d = 1 - \left(\frac{PM - MM}{PM + MM}\right) \quad \text{Eqn. 1}$$

[0103] wherein:

[0104] PM=scaled intensity of the perfect match probe;

[0105] MM=scaled intensity of the mismatch probe; and,

[0106] d=pair difference score.

The value of d can range from 0 to 2. When PM>>MM, the value of d approaches 0; when PM=MM, d=1; and when PM<<MM, the value of d approaches 2.

[0107] In some embodiments, the internal DNA quantification standards (QS) and negative control (NC) probe pairs are binned and sorted by attributes of the probes. Examples of the attributes of the probes that can be used in the embodiments disclosed herein include, but are not limited to binding energy; base composition, including A+T count, G+C count, and T count; sequence complexity; cross-hybridization binding energy; secondary structure; hair-pin forming potential; melting temperature; and length of the probe. These attributes of the probes can affect hybridization properties of the probes, for example, A+T count may affect hydrogen bonding of the probe, and T count may affect the length and base composition of the fragments produced by the use of DNase. Fragmentation with other enzyme systems may be influenced by the composition of other bases.

[0108] In some embodiments, QS and NC probe pairs are binned and sorted based on the individual probe's A+T count and T count. For each bin (A+T count by T count), the d values from the negative control probes are fit to a normal distribution to derive the scale (mean) and shape (standard deviation). Then, the d values from QS are fit to a gamma distribution to derive scale and shape. For each array, multiple density plots are produced by this process. In general, even one extra T can result in appreciable difference in the probe gamma scale parameter.

[0109] The parameters derived from gamma and normal distributions are used to derive a pair response score, r, for each probe pair. r is an indicator of the probability that a probe pair is positive, i.e., the probability for a probe pair to be responsive to the target sequence. r may be defined as:

$$r = \left(\frac{pdf_{\gamma}(X = d)}{pdf_{\gamma}(X = d) + pdf_{norm}(X = d)}\right) \quad \text{Eqn. 2}$$

[0110] where:

r=response score to measure the potential that a specific probe pair is binding a target sequence and not a background signal, i.e. the probability of the probe pair being positive for the specific target sequence;

pdf_{γ} (X=d)=probability that d could be drawn from the gamma distribution estimated for the target class ATx Ty;

pdf_{norm} (X=d)=probability that d could be drawn from the normal distribution estimated for the target class ATx Ty.

r can range from 0 to 1. r approaches 1 when PM>>MM, and r approaches 0 when PM<<MM.

[0111] Each set of interrogation probe pairs, e.g., an OTU, can be scored based on pair response scores, cross-hybridization relationships or both. In some embodiments, the system removes data from at least a subset of probe pair sets before making a final call on the presence or quantity of said microorganisms. In some embodiments, the data is removed based on interrogation probe cross hybridization potential. In some embodiments, the scoring of probe pairs is performed by a two-stage process as discussed below.

[0112] For example, a two stage analysis can be performed wherein only probe pairs that pass a first stage are analyzed in the next stage. In the first stage, the distribution of r across each set of probe pairs, R, is determined. For each set of probe pairs that is associated with an OTU, the r values of all probe pairs are ranked within the set, and percentage of probe pairs that meet one or more threshold r values are determined. Frequently, three threshold determinations are made at 25% increments across the total range of ranked probe pairs (interquartile Q1, Q2, and Q3); however, any number of threshold determinations or percentage increments can be used. For example, a determination can use one increment at 70% in which probe pairs must pass a threshold value of 80%.

[0113] Typically, to differentiate signal from noise, an OTU is considered to pass Stage 1 if Q1, Q2, and Q3 of the set of probe pairs that is associated with this OTU surpass the threshold of $Q1_{min}$, $Q2_{min}$, and $Q3_{min}$, respectively. That is, for an OTU to pass Stage 1, the r value of 75% of the probe pairs in the set of probe pairs that is associated with that OTU has to be at least $Q1_{min}$, the r value of 50% of the probe pairs in that set of probe pairs have to be at least $Q2_{min}$, and the r value of 25% of the probe pairs in that set of probe pairs have to be at least $Q3_{min}$. $Q1_{min}$ is at least about 0.5, about 0.55, about 0.6, about 0.65, about 0.7, about 0.75, about 0.8, about 0.82, about 0.84, about 0.86, about 0.88, about 0.90, about 0.91, about 0.92, about 0.93, about 0.94, about 0.95, about 0.96, about 0.97, about 0.98, or about 0.99. $Q2_{min}$ is at least about 0.5, about 0.55, about 0.6, about 0.65, about 0.7, about 0.75, about 0.8, about 0.82, about 0.84, about 0.86, about 0.88, about 0.90, about 0.91, about 0.92, about 0.93, about 0.94, about 0.95, about 0.96, about 0.97, about 0.98, or about 0.99. $Q3_{min}$ is at least about 0.5, about 0.55, about 0.6, about 0.65, about 0.7, about 0.75, about 0.8, about 0.82, about 0.84, about 0.86, about 0.88, about 0.90, about 0.91, about 0.92, about 0.93, about 0.94, about 0.95, about 0.96, about 0.97, about 0.98, about 0.99, about 0.992, about 0.994, about 0.996, about 0.998, or about 0.999. In some embodiments, $Q1_{min}$, $Q2_{min}$, and $Q3_{min}$ are determined empirically from spike-in experiments. For example, $Q1_{min}$, $Q2_{min}$, and $Q3_{min}$ are chosen to allow 2 pM amplicon concentration to pass. In some embodiments, $Q1_{min}$, $Q2_{min}$, and $Q3_{min}$ are 0.98, 0.97, and 0.82, respectively. These threshold numbers were empirically derived using DNase to fragment the sample sequences. Since DNase has a T-bias, the use of other enzymes can require a shift in the threshold numbers and can be empirically derived.

[0114] In the second stage only the OTUs passing the first are considered as potential sources of cross-hybridization. In some embodiments, for each OTU, only probe-pairs with r>0.5 (these are the probe pairs considered as to be likely responsive to the target sequence) are further analyzed. In other instances, only probe pairs with r>0.6, 0.7, 0.8, or 0.9 are considered responsive and are further analyzed. Probe pairs that are unlikely to be responsive (i.e., r<0.5) are not analyzed further even if their set R, was responsive overall. $R_{0.5}$ represents the subset of probe pairs in which all probe pairs have r>0.5. Typically, based on the interquartile Q1, Q2

and Q3 values chosen at Stage 1, most of the probe pairs in the OTUs passing Stage 1 are analyzed. In other embodiments, only the probe-pairs with r>0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, or 0.90 are further analyzed.

[0115] For each probe pair in the $R_{0.5}$ subset, the count of putatively cross-hybridizing OTUs (i.e., the number of OTUs with which the probe pair can cross-hybridize) is determined. In this process, only the OTUs that have passed Stage 1 are considered as potential sources of cross-hybridization. Each probe pair in the $R_{0.5}$ subset is penalized by dividing its r value by the count of putatively cross-hybridizing OTUs to determine its modified possibility of being positive. The modified possibility of being positive for a probe pair may be represented by a r_x value. r_x may be defined as:

$$r_x = \frac{r}{scalarS_{1x}} \quad \text{Eqn. 3}$$

[0116] where

[0117] S_1=Set of OTUs passing Stage1; and,

[0118] S_{1x}=Set of OTUs passing Stage 1 with cross hybridization potential to the given probe pair.

[0119] r_x is proportional to the response of the probe pair and the specificity of the probe pair given the community observed during the first stage. r_x value can range from 0 to 1. For each set of probe pairs associated with an OTU, r_x are calculated for each probe pair and ranked within the set. Interquartile Q1, Q2, Q3 values for the distribution of r_x value in each set of probe pairs are determined. The taxon represented by the OTU is considered to be present if Q1 is greater than Q_{x1}, Q2 is greater than Q_{x2}, or Q3 is greater than Q_{x3}. Q_{x1} is at least about 0.5, at least about 0.55, at least about 0.6, at least about 0.65, at least about 0.7 at least about 0.75, at least at least about 0.8, at least about 0.85, at least about 0.90, at least about 0.95, or at least about 0.97. Q_{x2} is at least about 0.5, at least about 0.55, at least about 0.6, at least about 0.65, at least about 0.7 at least about 0.75, at least at least about 0.8, at least about 0.85, at least about 0.90, at least about 0.95, or at least about 0.97. Q_{x3} is at least about 0.5, at least about 0.55, at least about 0.6, at least about 0.65, at least about 0.7 at least about 0.75, at least at least about 0.8, at least about 0.85, at least about 0.90, at least about 0.95, or at least about 0.97. In some embodiments, Q_{x1} is at least 0.66, that is, 75% of the probe pairs in the set of the probe pairs have a r_x value that is at least 0.66.

[0120] A two stage hybridization signal analysis procedure can be performed on hybridization signals from any array or microparticle generated data set, including data generated from the use of any combination of probes selected using the disclosed methodologies. In some embodiments, the second stage of the procedure penalizes probes based on the number of cross-hybridizations, the intensity of the cross-hybridization signals or a combination of the two.

[0121] The method disclosed herein is useful for hierarchical probe set scoring. An OTU may be present at a node at any hierarchical level on a clustering tree. As used herein, an OTU is a group of one or more organisms, such as a domain, a sub-domain, a kingdom, a sub¬ kingdom, a phylum, a subphylum, a class, a sub-class, an order, a sub-order, a family, a subfamily, a genus, a subgenus, a species, or any cluster. In some embodiments, a $R_{0.5}$ set is collected for each node on the phylogenetic tree and consists of all unique probes from subordinate $R_{0.5}$ sets. For example, for calculating r_x values for probe pairs in a $R_{0.5}$ set for an OTU representing an "order," the count of putatively cross-hybridizing equally-ranked taxa (i.e., "order" node) containing at least one sequence with cross-hybridization potential is used as the denominator in Eqn. 3.

[0122] In some embodiments, the OTUs at the leaf level (e.g., species, subgenus or genus) are first analyzed. Then each successive level of nodes in the clustering tree is analyzed. In some embodiments, the analysis is performed up to the domain level. In some embodiments, the analysis is performed up to the phylum level. In some other embodiments, the analysis is performed up to the kingdom level. Penalization for cross-hybridization in Eqn. 3 is only performed for probes on the same taxonomy level. All present taxa are quantified using the mean scaled PM probe intensity after discarding the highest and lowest value of the set R (HybScore). In some embodiments, only taxa present at a first level are analyzed further.

[0123] In some embodiments, a summary abundance score is determined. Corrected abundance scores are created based on G+C content and uracil incorporation. Generally, probes with higher G+C content produce a higher hybridization signal that is typically compensated for correcting the abundance scores.

[0124] The probability of detection for each taxonomic node is determined by summarizing terminal node detection and the breadth of cross-hybridization relationships. Hierarchical probes are scored for evidence of novel organisms based on cluster analysis.

[0125] In some embodiments, the system is capable of analyzing other data in conjunction with that obtained from the analysis of probe hybridization signal strength. In some embodiments, the system can analyze sequencing reaction data including that obtained with high-through put sequencing techniques. In some embodiments, the sequencing data is from same regions of the same highly conserved sequence analyzed by the method disclosed herein using probes.

High Capacity Analysis System Applications

[0126] Numerous subject-derived samples can be assayed to determine the sample's microbiome composition. By having an assay system capable of detecting in a single assay the presence and optionally quantity of at least 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 200,000, 500,000 or 1,000,000 bacterial or archeal taxa, a complete picture of a microbiotic ecosystem, such as the living microorganisms in the ecosystem, can be achieved quickly and at relatively low cost providing the ability to examine numerous subjects.

[0127] The elucidation of a specific microbiome associated with an ecosystem, animal, human, organ system, condition, and the like allows for the generation of a "signature," "biosignature," or "fingerprint" of the particular environment sampled, terms used interchangeably herein. If the biosignature is from a normal or healthy system or subject, or is from a subject free from a condition under examination, then the associated biosignature can be used as a reference for the comparison of later samples from the same or other subjects to monitor for changes that are associated with an abnormal or unhealthy state or condition. For example, if a later biosignature of a subject shows that the microbiome has shifted away from that associated with a healthy pulmonary status, then preemptive measures could be taken to prevent a continued shift.

[0128] Similarly, a biosignature of an environment can be compared to a biosignature generated from a pool of samples that represent an average or normal biosignature for a population or collection of environments. For example, a sample from an unhealthy individual or environment could be assayed and the microbial biosignature compared to the biosignature seen in a healthy population at large or unaffected environment. If one or more microorganisms are detected in the unhealthy individual that are either not seen in the general population or not seen at the same prevalence then therapeutic measures can be taken to selectively eliminate or reduce in number the microorganisms associated with the unhealthy state. Once a relationship is known between the prevalence of a particular microorganism or group of microorganisms (e.g. one or more OTUs that consist essentially of viable organisms from a sample) and a disease state, then disease progression or treatment response can also be monitored, diagnosed, and/or predicted using the present systems and methods.

[0129] Numerous microbiomes of animals or humans can be analyzed with the present systems and methods including the gut, respiratory system, urogenital tract, mammary glands, skin, oral cavity, auditory canal, and skin. Clinical samples such as blood, sputum, nares, feces, and urine can be used with the method. From the analysis of normal individuals and those suffering from a disease or condition, a large database of fingerprints or biosignatures can be assembled. By comparing the biosignatures between healthy and disease related states, associations can be made as to the influence and importance of individual components of the microbiome.

[0130] Once these associations are made, treatments can be designed and tested to alter the composition of the microbiota seen in the disease state. Additionally, by regularly monitoring the microbial composition of an affected organ system in a diseased individual, disease progress or response to therapy can be observed and if need, additional therapeutic measures taken to alter the microbiome composition to one that is more representative of that seen in a healthy population.

[0131] An interesting property of bacteria that has great importance in healthcare, water quality and food safety is quorum sensing. Many bacteria are able to sense the presence of other members of their species or related species and upon reaching a specific density the bacteria start producing various virulence or pathogenicity factors. In other words, the bacteria's gene expression is coordinated as a group. For example, some bacteria produce exopolysaccharides that are known as "slime layers." The secretion of exopolysaccharidse can decrease the ability of white blood cells to phagocytize the microorganisms and make the microorganisms more resistant to therapeutics or cleaning agents. Traditional methodologies require the detection of specific gene expression in order to detect or study quorum sensing and other population induced effects. The present systems and methods can be used to understand the changes that occur in a microbiome that are associated with a given effect such as biofilm formation or toxicity production. One can develop protocols with the present systems and methods to look for and determine conditions that lead to quorum sensing. For example, testing samples at various timepoints and under varying conditions can lead to determining how and when to intervene or reverse population induced expression of virulence or pathogenicity factors.

[0132] In some embodiments, a method is provided to identify a new indicator species for an environmental or health condition with the present systems and methods. The condition can be that of a normal or healthy state. Alternatively, the indicator species can be for an unhealthy or abnormal condition. To identify a new indicator species, a normal sample is simultaneously assayed to determine the presence or quantity of each OTU associated with all known bacteria, archae, or fungi; this test result is compared to the results achieved in the simultaneous assay of sample from the environment of the condition where the presence or quantity of each OTU associated with all known bacteria, archae, or fungi was determined. Microorganisms that change in abundance at least 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold or 100-fold, either increasing in abundance or decreasing in abundance represent putative indicator species for a condition.

[0133] In other embodiments, methods are provided for identifying indicators species associated with a disease state, disease progression, treatment regimen, probiotic administration, including progression of disease. In some embodiments, methods are provided for monitoring a change in the environment or health status associated with introducing one or more new microorganisms into a community. For example, measures to increase a particular microorganism's percentage of the gut microbiome in an individual, such as feeding a person yogurt or a food supplement containing *L. casei*, can be monitored using the present methods and systems. In some embodiments, the presence, absence, relative abundance, and/or quantity of desired and/or undesired live microorganisms in food and food supplements can be evaluated by the methods disclosed herein. For example, the viability over time of microorganisms in a probiotic supplement can be determined and used as a measure of shelf-life and/or potency of the probiotic supplement.

Combined Analysis

[0134] The ability to identify and quantitate the microorganisms in a sample (e.g. the live microorganisms in a sample comprising live and dead microorganisms) can be combined with a gene expression technology such as a functional gene array to correlate populations with observed gene expression. Similarly, microbiome composition analysis can be correlated with the presence of chemicals, proteins including enzymes, toxins, drugs, antibiotics or other sample constituents. For instance, nucleic acids isolated from a soil sample can be analyzed to elucidate the microbiome composition (e.g. biosignature) and also to identify expressed genes. In the bare, nutrient-poor soils on the Antarctic, this analysis associated chitinase and mannanase expression with Bacteroidetes and CH_4-related genes with Alphaproteobacteria. (Yergeau et al., Environmental microarray analyses of Antarctic soil microbial communities. ISME J. 3:340-351, 2009). Significant correlations were also found between taxon abundances and C- and N-cycle gene abundance. From this data, one can predict that certain organisms or groups of organisms are required or account for the majority of an expected or observed enzymatic or degradative process. For example, members of the Bacteroidetes phylum probably degrade the majority of environmental chitin, a major constituent of exoskeletons of insect and arthropods and also of fungi cell walls, at the sample locale.

[0135] This methodology can be used to identify new antibiotic producing organisms, even ones that are unculturable. For instance, soil extracts can be tested for antibiotic activity. If a positive extract is found, a sample of the soil from which a portion was extracted for antibiotic can be analyzed for

microbial composition and perhaps gene expression. Major constituents of the microbiome could be correlated with antibiotic activity with the correlation strengthened through gene expression data allowing one to predict that a particular organism or group of organisms is responsible for the observed antibiotic activity.

[0136] In some embodiments, a method is provided for making a prediction about a sample comprising a) determining microorganism population data as the probability of the presence or absence of at least 100 OTUs of microorganisms in said sample; b) determining gene expression data of one or more genes by said microorganisms in said sample and c) using said expression data and population data to make a prediction about said sample. In some embodiments, the prediction entails the identity of a microorganism responsible for a characteristic or condition observed in an environment.

[0137] Other combined analysis methods include the use of a diffusion chamber to retain microorganisms in a sample while one or more constituents or parameters of the sample are changed. For instance, the salinity or pH of the sample can be changed abruptly or gradually over time. Following specific time intervals, the microbiome of the sample in the diffusion chamber can be determined. Microorganisms that cannot tolerate the new environment conditions will die, become reduced in number due to unfavorable conditions or predation, or remain static in their numbers. In contrast, microorganisms that can tolerate the new conditions will at least maintain their number or thrive, perhaps becoming a dominant population. Use of a diffusion chamber coupled with a system capable of detecting the presence or quantity of at least 10,000 OTUs can allow the identification of microorganisms that perish or fail to thrive when placed in a new environment. Such microorganisms are termed "transient", meaning that their percent composition of the microbiome changes quickly. The identification of transient microorganisms can be used to ascertain the time and/or place they were introduced into an environment. Different transient microorganisms can have different half-lives for a particular condition.

[0138] Diffusion chambers can also take the form of a semi-permeable capsule, tube, rod, or sphere or other solid or semi-solid object. A microbiome or a select group of bacteria can be placed inside the capsule, that is then sealed and introduced into an environment for a specified period of time. Upon removal, the capsule is opened and the microbiome or select group of bacteria sampled to ascertain changes in the presence or quantity of the individual constituents. The capsule can be removed once or periodically to sample the microbiome. Alternatively, multiple single use capsules with identical quantities of the microbiome can be used, each one removed and sampled at a different time point. Microbiomes placed in capsules or other semi-permeable containers can be introduced into a living organism, usually through an orifice, to measure changes to the microbiome composition associated with a particular organ or system environment. For example, a semi-permeable capsule or tube containing a microbiome can be introduced into the gastrointestinal system through the mouth or anus. A microbiome from a healthy individual can be introduced in this manner into an unhealthy individual, such as a patient suffering from Crohn's disease or irritable bowel syndrome to ascertain the effect of the unhealthy condition on the normal, healthy individual associated microbiome. In this manner, the efficacy of drug effectiveness and treatment protocols could also be evaluated based on the effects of the gut ecology on a known microbiome.

Low Density-Special Purpose Detection Systems

[0139] In some embodiments, probes are selected for constructing special purpose systems including those with arrays or microparticles. Typically, special purpose "low density" systems, are designed for use in a specific environment or for a particular application and usually feature a reduced number of probes, "down-selected" probes, that are specific to organisms that are known or expected to be present in the particular environment, such as associated with a particular biosignature. In some cases the biosignature is fecal contamination. Typically, a low density system comprises no more than 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000 or 10,000 down selected probes or 5, 10, 25, 50, 100, 250, 500, 1,000, 2,500 or 5,000 down selected probes probe pairs (PM and MM probes). In some embodiments, only 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 probes are used per OTU. In further embodiments, only PM probes are used. Generally, these down-selected probes have robust hybridization signals and few or no cross hybridizations. In some embodiments, the collection of down selected probes have a median cross hybridization potential number of less than 20, 15, 10, 8, 7, 6, 5, 4, 3, 2, or 1 per probe. Frequently the down selected probes belong to OTUs that have reduced numbers of probes. In some embodiments, the OTUs of a down select probe collection have a median number of less than 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3 or 2 probes per OTU. Generally, low density systems feature probes that recognize no more than 10, 25, 50, 100, 250, 500, 1,000, 2,000, or 5,000 taxa. For a set number of probes, a number of design strategies can be employed for low density systems. One approach is to maximize the number of OTUs identified, e.g., use one probe per OTU with no mismatch probes. Another approach is to select probes based on the desired confidence level. Here, multiple probes for each OTU along with corresponding mismatch probes can be required to achieve at least 95% confidence level for the presence and quantity of each OTU. The probes for a particular low density application can be selected by applying a sample from an appropriate environment to a high density analysis system, e.g., a detection system that can in a single assay determine the probability of the presence or quantity of at least 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 250,000, 500,00 or 1,000,000 OTUs of a single domain, such as bacteria, archea, or fungi, or alternatively, for each known OTU of a single domain. Probes associated with prevalent OTUs can be selected for a low density system. Alternately, the OTUs seen in a sample of interest can be compared with a control sample and shared OTUs subtracted out with the probes associated with the remaining OTUs selected for the low density system. Additionally, probes can be selected based on a change in prevalence of OTUs between the environment of interest and a control environment. For example, OTUs that are at least 2-fold 5-fold, 10-fold, 100-fold or 1,000-fold more abundant in the sample of interest compared to the control sample are included in the down selected probe set. Using this information, a down selected array, bead multiplex system or other low density assay system is designed.

[0140] "Low density" assays systems can be used to identify select microorganisms and determine the percentage composition of various select microorganisms in relation to

each other. Low density assay systems can be constructed using probes selected through the disclosed methodologies. These low density systems can identify at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500, 1000 or more microorganisms. Representative microorganisms to be identified and optionally quantified are listed in Table 2.

TABLE 2

Representative Microorganisms Recognized by Low Density Assay Systems

Species	Application
Listeria monocytogenes	Food safety, environmental surveillance of food processing plants
Salmonella enterica subsp. *enterica* serovar Enteritidis	Food safety, environmental surveillance of food processing plants
Pseudomonas aeruginosa	Pulmonary health

[0141] Low density assays systems are useful for numerous environmental and clinical applications. Exemplary applications are listed in Table 2. Medical conditions that can be identified, diagnosed, prognosed, tracked, or treated based on data obtained with a low density system include but are not limited to, cystic fibrosis, chronic obstructive pulmonary disease, Crohn's Disease, irritable bowel syndrome, cancer, rhinitis, stomach ulcers, colitis, atopy, asthma, neonatal necrotizing enterocolitis, obesity, periodontal disease and any disease or disorder caused by, aggravated by or related to the presence, absence or population change of a microorganism. Through the judicious selection of OTUs to be included in a system, the system becomes a diagnostic device capable of diagnosing one or more conditions or diseases with a high level of confidence producing very low rates of false positive or false negative readings.

[0142] In some embodiments, the low density systems also feature confirmatory probes that are specific (complimentary) for genes or sequences expressed in specific organisms. For example, the caf1 virulence gene of *Yersinia pestis* and the zonula occludens toxin (zot) gene of *Vibrio cholerae* and also confirmatory probes to *Y. pestis* or *V. cholerae*.

Kits

[0143] As used herein a "kit" refers to any delivery system for delivering materials or reagents for carrying out a method disclosed herein. In the context of assays, such delivery systems include systems that allow for the storage, transport, or delivery of arrays or beads with probes, reaction reagents (e.g., probes, enzymes, etc. in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions for performing the assay etc.) from one location to another. For example, kits include one or more enclosures (e.g., boxes) containing the relevant reaction reagents and/or supporting materials for assays disclosed herein.

[0144] In one aspect of the present application, kits for analysis of nucleic acid targets are provided. According to some embodiments, a kit includes a plurality of probes capable of determining the presence or quantity over 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 30000, 40000 50000 or 60000 different OTUs in a single assay. Such probes can be coupled to, for example, an array or plurality of microbeads. In some aspects a kit comprises at least 5, 10, 15, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000, 500000, 1000000 or 2000000 interrogation probes selected using the disclosed methodologies and/or for use in the identification and/or comparison of a biosignature of one or more samples.

[0145] The kit can also include reagents for sample processing, including one or more agents described herein, alone or in combination. In some embodiments, the reagents comprise an agent that selectively modifies nucleic acid of dead cells. In some embodiments, the reagents comprise reagents for the PCR amplification of sample nucleic acids including primers to amplify regions of a highly conserved sequence, such as regions of the 16S rRNA gene. In some embodiments, the reagents comprise reagents for the direct labeling of RNA, such as rRNA. In further embodiments, the kit includes instructions for using the kit. In other embodiments, the kit includes a password or other permission for the electronic access to a remote data analysis and manipulation software program. Such kits will have a variety of uses, including environmental monitoring, diagnosing disease, monitoring disease progress or response to treatment, and identifying a contamination source and/or the presence, absence, or amount of one or more contaminants.

Computer Implemented Methods

[0146] FIG. **7** illustrates an example of a suitable computing system environment or architecture in which computing subsystems may provide processing functionality to execute software embodiments of the present application, including probe selection, analysis of samples, and remote networking. The method or system disclosed herein may also operational with numerous other general purpose or special purpose computing system including personal computers, server computers, hand-held or laptop devices, multiprocessor systems, and the like.

[0147] The method or system may be described in the general context of computer-executable instructions, such as program modules, being executed by a computer. The method or system may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network.

[0148] With reference to FIG. **7**, an exemplary system for implementing the method or system includes a general purpose computing device in the form of a computer **102**.

[0149] Components of computer **102** may include, but are not limited to, a processing unit **104**, a system memory **106**, and a system bus **108** that couples various system components including the system memory to the processing unit **104**.

[0150] Computer **102** typically includes a variety of computer readable media. Computer readable media includes both volatile and nonvolatile media, removable and non-removable media and a may comprise computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices.

[0151] The system memory **106** includes computer storage media in the form of volatile and/or nonvolatile memory such as read only memory (ROM) **110** and random access memory (RAM) **112**. A basic input/output system **114** (BIOS), containing the basic routines that help to transfer information between elements within computer **102**, such as during start-up, is typically stored in ROM **110**. RAM **112** typically contains data and/or program modules that are immediately

accessible to and/or presently being operated on by processing unit **104**. FIG. **7** illustrates operating system **132**, application programs **134** such as sequence analysis, probe selection, signal analysis and cross-hybridization analysis programs, other program modules **136**, and program data **138**.

[0152] The computer **102** can also include other removable/non-removable, volatile/nonvolatile computer storage media. By way of example only, FIG. **7** illustrates a hard disk drive **116** that reads from or writes to non-removable, nonvolatile magnetic media, a magnetic disk drive **118** that reads from or writes to a removable, nonvolatile magnetic disk **120**, and an optical disk drive **122** that reads from or writes to a removable, nonvolatile optical disk **124** such as a CD ROM or other optical media. Other removable/non-removable, volatile/nonvolatile computer storage media that can be used in the exemplary operating environment include magnetic tape cassettes, flash memory cards, digital versatile disks, digital video tape, solid state RAM, solid state ROM, and the like. The hard disk drive **116** is typically connected to the system bus **108** through a non-removable memory interface such as interface **126**, and magnetic disk drive **118** and optical disk drive **122** are typically connected to the system bus **108** by a removable memory interface, such as interface **128** or **130**.

[0153] The drives and their associated computer storage media discussed above and illustrated in FIG. **7**, provide storage of computer readable instructions, data structures, program modules and other data for the computer **102**. In FIG. **7**, for example, hard disk drive **116** is illustrated as storing operating system **132**, application programs **134**, other program modules **136**, and program data **138**. A user may enter commands and information into the computer **102** through input devices such as a keyboard **140** and a mouse, trackball or touch pad **142**. These and other input devices are often connected to the processing unit **104** through a user input interface **144** that is coupled to the system bus, but may be connected by other interface and bus structures, such as a parallel port or a universal serial bus (USB). A monitor **158** or other type of display device is also connected to the system bus **108** via an interface, such as a video interface or graphics display interface **156**. In addition to the monitor **158**, computers can also include other peripheral output devices such as speakers (not shown) and printer (not shown), which can be connected through an output peripheral interface (not shown).

[0154] The computer **102** can be integrated into an analysis system, such as a microarray or other probe system described herein. Alternatively, the data generated by an analysis system can be imported into the computer system using various means known in the art.

[0155] The computer **102** can operate in a networked environment using logical connections to one or more remote computers or analysis systems. The remote computer can be a personal computer, a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements described above relative to the computer **102**. The logical connections depicted in FIG. **7** include a local area network (LAN) **148** and a wide area network (WAN) **150**, but can also include other networks. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet. When used in a LAN networking environment, the computer **102** is connected to the LAN **148** through a network interface or adapter **152**. When used in a WAN networking environment, the computer **102** typically includes a modem **154** or other means for establishing communications over the WAN **150**, such as the Internet. The modem **154**, which can be internal or external, can be connected to the system bus **108** via the user input interface **144**, or other appropriate mechanism. In a networked environment, program modules depicted relative to the computer **102**, or portions thereof, can be stored in the remote memory storage device.

[0156] In further aspects of the present application, computer-implemented methods are provided for analyzing the presence or quantity of over 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 30,000, 40,000 50,000 or 60,000 different OTUs in a single assay. In some embodiments, computer executable logic is provided for determining the presence or quantity of one or more microorganisms in a sample comprising: logic for analyzing intensities from a set of probes that selectively binds each of at least 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 30,000, 40,000 50,000 or 60,000 unique and highly conserved polynucleotides and determining the presence of at least 97% of all species present in said sample with at least 90%, 95%, 96%, 97%, 98%, 99% or 99.5% confidence level.

[0157] In some embodiments, computer executable logic is provided for determining probability that one or more organisms, from a set of different organisms, are present in a sample. The computer logic comprises processes or instructions for determining the likelihood that individual interrogation probe intensities are accurate based on comparison with intensities of negative control probes and positive control probes; a process or instructions for determining likelihood that an individual OTU is present based on intensities of interrogation probes from OTUs that pass a first quantile threshold; and a process or instructions for penalizing one or more OTUs that have passed the first quantile threshold based on their potential for cross-hybridizing with other probes that have also passed the first quantile threshold.

[0158] In some embodiments, computer executable logic is provided for determining the presence of one or more microorganisms in a sample. The logic allows for the analysis of a set of at least 1000 different interrogation perfect probes. The logic further provides for the discarding of information from at least 10% of the interrogation perfect match probes in the process of making the determination. In some embodiments, the computer executable logic is stored on computer readable media and represents a computer software product.

[0159] In some embodiments, computer software products are provided wherein computer executable logic embodying aspects of the invention is stored on computer media like hard drives or optical drives. In some embodiments, the computer software products comprise instructions that when executed perform the methods described herein for determining candidate probes.

[0160] In some further embodiments, computer systems are provided that can perform the methods of the inventions. In some embodiments, the computer system is integrated into and is part of an analysis system, like a flow cytometer or a microarray imaging device. In other embodiments, the computer system is connected to or ported to an analysis system. In some embodiments, the computer system is connected to an analysis system by a network connection. In one example of a system that employs a network connection, a sample is imaged using a commercially available imaging system and software. The data is outputted using a standard data format like a CEL file (AFFYMETRIX®), or a Feature Report file

(NIMBLEGEN®). Then the data is sent to a remote or central location for analysis using a method disclosed herein. In some embodiments, a standardized analysis is performed providing signal normalization, OTU quantification, and visual analytics. In other embodiments, a customized analysis is performed using a fixed protocol designed for the user's particular needs. In still other embodiments, a user configurable analysis is used, include a protocol that allows for the user to adjust at least one variable before each analysis run.

[0161] After processing, the results are stored in an exchangeable binary format for later use or sharing. Additionally, hybridization scores and OTU probability values can be exported to a tab delimited file or in a format compatible with UniFrac (Lozupone, et al., UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context, BMC Bioinformatics, 7, 371; 2006) for further statistical analysis of the detected sample communities.

[0162] In some embodiments, multiple, interactive views of the data are available, including taxonomic trees, heatmaps, hierarchical clustering, parallel coordinates (time series), bar plots, and multidimensional scaling scatterplots. In some embodiments, the taxonomy tree displays the mean intensities for each detected OTU and displays the leaves of the tree as a heatmap of samples. The tree can be dynamically pruned by filtering OTUs below a certain intensity or probability threshold. Additionally, the tree can be summarized at any level from phylum to subfamily. In other embodiments, the user can hierarchically cluster both OTUs and samples using any of the standard distance and linkage methods from the integrated C Clustering Library (de Hoon, et al., Open source clustering software, Bioinformatics, 20, 1453-1454; 2004), and the resulting dendrograms displayed in a secondary heatmap window. In some embodiments, a third window is provided that displays interactive bar plots of differential OTU intensities to facilitate pairwise comparison of samples. For any two samples, the height of the difference bars displays either the absolute or relative difference in mean intensity between OTUs. The bars can be grouped and sorted along the horizontal axis by any taxonomic rank for easy identification and comparison. Synchronized selection and filtering affords users the unique ability to seamlessly navigate between multiple views of the data. For example, users can select a cluster in the hierarchical clustering window and simultaneously view the selected organisms in the taxonomy tree, immediately revealing both their phylogenetic and environmental relationship. In further embodiments, the data from the analysis system, i.e., analysis system or flow cytometer, can be co-analyzed and displayed with high-throughput sequencing data. In some embodiments, for each organism identified as present in the sample, the user is able to view a list of other environments where the particular organism is found.

[0163] In some embodiments, the screen displays are dynamic and synchronized to allow the selection or filtration of OTUs with changes to any view simultaneously reflected in all other views. Additionally, OTUs confirmed by 16S rRNA gene, 18S rRNA gene, or 23S rRNA gene sequencing can be co-displayed in all views.

Business Methods

[0164] In some aspects of the present application, a business method is provided wherein a client images an array or scans a lot of microparticles and sends a file containing the data to a service provider for analysis. The service provider analyzes the data and provides a report to the user in return for financial compensation. In some embodiments, the user has access to the service provider's analysis system and can manipulate and adjust the analysis parameters or the display of the results.

[0165] In another aspect of the present application, a business method is provided wherein a client sends a sample to be processed, imaged or scanned and the data analyzed for the presence or quantity of organisms. The service provider sends a report to the client in return for financial compensation. In some embodiments, the client has access to a suite of data analysis and display programs for the further analysis and viewing of the data. In further embodiments, the service provider first provides a system or kit to the client. The kit can include a system to assay a majority, or the entirety of the microbiome present or the system can contain "down-selected" probes designed for particular applications. After sample processing and imaging, the client sends the data for analysis by the service provider. In some embodiments, the client report is electronic. In other embodiments, the client is provided access to a suite of data analysis and display programs for the further viewing, manipulation, comparison and analysis of the data. In some embodiments, the client is provided access to a proprietary database in which to compare results. In other embodiments, the client is provided access to one or more public databases, or a combination of private and public database for the comparison of results. In some embodiments, the proprietary database includes the pooled results (fingerprints, biosignatures) for normal samples or the pooled results from particular abnormal situations such as a disease state. In some embodiments, the biosignatures are continuously and automatically updated upon receipt of a new sample analysis.

[0166] In some embodiments, the database further comprises highly conserved sequence listings. In some embodiments, the database is updated automatically as new sequence information becomes available, for instance, from the National Institutes of Health's Human Microbiome Project. In further embodiments, probe sets are automatically updated based on the new sequence information. Continuous upgrading of the sequence information and refinement of the probe sets allow for increasing accuracy and resolution in determining the composition of microbiomes and the quantity of their individual constituents. In some embodiments, the system compares earlier microbiome biosignatures with later microbiome biosignatures from the same or substantially similar environments and analyzes the changes in probe set composition and hybridization signal analysis parameters for information that is useful in improving or refining the discrimination between related OTUs, identification and quantification of microbiome constituents, or increasing accuracy of the determinations.

[0167] In some embodiments, the database compiles information about specific microbiomes, for example, the microbiota associated with healthy and unhealthy human intestinal microflora including, age, gender and general health status of host, geographical location of host, host's diet (i.e., Western, Asian or vegetarian), water source, host's occupation or social status, host's housing status.

[0168] In some embodiments, the reference healthy/normal signatures for adults, male and female, and children can be used as benchmarks to identify presymptomatic and symptomatic disease states, response to treatments/therapies, infection, and/or secondary infection associated with disease.

In some embodiments, the client is provided with a diagnosis or treatment recommendation based on the comparison between the client's sample microbiome and one or more reference microbiome.

[0169] In some embodiments, the present application contemplates using 454 pyrosequencing for detection of OTUs in a sample. The OTUs are detected using barcoded amplicons. The amplification can occur on beads. 454 sequencing protocols are described in, for example, (Poinar H N, et. al Science. 2006 Jan. 20; 311(5759) pages 392-4) which is incorporated herein by reference.

EXAMPLES

[0170] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example 1

Exclusion of Dead Cell DNA from Detection

[0171] FIG. **1** illustrates one non-limiting embodiment of the present application. To briefly summarize, a test sample suspected of containing a mixture of vegetative cells, live, and dead bacterial endospores is first heat shocked at 80° C. for 15 minutes (Step 1). This step will eliminate any vegetative cells present in the sample, whereas spores will remain unaffected. The sample is then treated with PMA followed by incubation in the dark. Thereafter, the sample is exposed to visible light for 5 minutes, so that the DNA from dead cells only will be cross-linked. In general, excess PMA is removed from a sample after the treatment step and prior to or as part of subsequent nucleic acid extractions steps, such as by centrifugation and washing. Following the initial PMA treatment, the sample is concentrated by centrifugation and the fluorescence of the resulting reaction mixture, which contains dead cells and spores, is measured. The sample is then divided into two. The first aliquot is heated to 100° C. for 18 minutes (Step 2a) to kill all microbes, including endospores, and the fluorescence of the reaction mixture is measured as mentioned before. The Step 2a fluorescence measured is subtracted from Step 1 fluorescence to determine total viable spores in the test sample. In Step 2b, qPCR is used to provide confirmatory quantitation of the total viable spores by detecting DNA concentration, since the PMA intercalated DNA from dead cells/spores would be unavailable for PCR amplification. In some embodiments, the total detection time, including PCR reaction for low-biomass samples (1 to 10^3 spores per reaction), will be within 60 minutes.

[0172] A simplified work flow of the PMA-Microarray approach is shown in FIG. **2**. To summarize, appropriate samples are taken according to established sampling protocols. Each sample is then divided into three portions and processed separately. The first sample is treated with PMA, light inactivated, and DNA is extracted using the Maxwell 16 automated system (Promega Corp.). In the second sample, DNA is extracted using an automated DNA extracting system.

[0173] In some embodiments, DNA from both the first and second samples is then used for 16S rRNA gradient PCR, the product of which is further used for PhyloChip analysis; the 16S rRNA qPCR is used to further measure the total microbial concentration with and without PMA treatment; and the third and last sample is used for microscopy analysis. The BacLight stain (Invitrogen Crop.) is used to distinguish between viable and nonviable cells present in a sample; samples that are stained green are viable and samples that are stained red are nonviable. In some embodiments, these results will then be used to correlate with the PhyloChip, and qPCR results.

[0174] In some embodiments, the methods are used for the detection and identification of only viable microbes in clean rooms or satellites, such as for the purposes of planetary protection.

Example 2

PMA Method

[0175] An objective was to evaluate a rapid and sensitive spore detection concept that will estimate viable microbial spores.

[0176] Real-time quantitative-PCR methods can be used to estimate microbial populations. This unique method is based on the use of a fluorescent DNA intercalating agent propidium monoazide (PMA), which can only penetrate the membranes of dead cells. The combination approach described in this example is referred to as the "PMA-method." The present example describes the development of a reliable and efficient method for extracting DNA from spores. The spore structure is complex and made up of several layers that impart resistance to extreme environmental conditions. These several layers which include the cortex and coat structures together are probably responsible for difficulties in releasing DNA from spores. Using a urea-based extraction buffer, it was possible to degrade the outer spore coats. With the spore coats removed the spores were more susceptible to lytic degradation. After lytic degradation, DNA extraction of both viable and dead spores was consistent in yield and amplification efficiency. Several alternative DNA extraction methods were tried including the method by Rawsthorne et al (Rawsthorne and Phister 2009, Letters in Applied Microbiology, 49: 652-654), which utilized DTT, and were found to be less efficient.

[0177] DNA intercalating agents have been used to selectively distinguish between viable and dead bacterial cells, as they can only penetrate the membrane of dead cells (Rudi et al. 2005, Applied and Environmental Microbiology, 71(2): 1018-1024; Nogva et al. 2003, BioTechniques, 34: 804-813). Once the DNA of dead cells is crosslinked, it is unavailable for PCR amplification (Nocker et al. 2007, Applied and Environmental Microbiology, 73(16): 5111-5117). Recently, Rawsthorne et al (2009) demonstrated that PMA can be used to selectively distinguish between viable and non-viable *B. subtilis* spores.

[0178] DNA isolated from *E. coli* was treated with PMA to show that PMA can be used to intercalate DNA and make it unavailable for PCR amplification (FIG. **3**).

[0179] Using the method of Rawsthorne et al (2009), *B. pumilus* SAFR-032 spores were heat inactivated at 121° C. for 15 min in two consecutive cycles. DNA was extracted using both the urea and DTT based methods. When compared to viable spores, DNA after one and two consecutive auto-

clave cycles was significantly degraded, and remaining DNA was barely detectable by qPCR. This means that PMA pre-treatment of thermally inactivated *B. pumilus* spores may not be necessary when using rigorous time/temperature combinations.

[0180] Alternative killing treatments including a lower heating temperature (90° C.) and the use of UV were utilized to see how the DNA from *B. pumilus* is affected and if PMA treatment is necessary. FIGS. **5** and **6** show plate count compared to qPCR results. Based on the qPCR results, if the alternative treatments are used, spores will have to be treated with PMA as the DNA is not degraded.

[0181] A comprehensive census of the viable microbes on the surfaces of spacecraft can be used to maximize detection of resident microbiota. For example, PMA-based quantitation can be used to elucidate the viable microbial community that could survive to reach other planets.

Example 3

Evaluation of PMA-Treated Sample by Phylogenetic Array

[0182] A phylogenetic array is fabricated with some of the organism-specific and OTU-specific 16s rRNA probes selected by the methods described herein and in co-pending international application PCT/US2010/040106. The phylogenetic array in this example consists of 1,016,064 probe features, arranged as a grid of 1,008 rows and columns. Of these features, ~90% are oligonucleotide perfect match (PM) or mis-match (MM) probes with exact or inexact complementarity, respectively, to 16s rRNA genes. In general, MM probes have one or more sequence differences with respect to PM probes, resulting in one or more mismatches with a target sequence. Each PM probe is paired with a mismatch control probe to distinguish target-specific hybridization from background and non-target cross-hybridization. The remaining probes are used for image orientation, normalization controls, or for pathogen-specific signature amplicon detection using additional targeted regions of a chromosome. Each high-density 16s rRNA gene microarray is designed with additional probes that (1) target amplicons of prokaryotic metabolic genes spiked into the 16s rRNA gene amplicon mix in defined quantities just prior to fragmentation and (2) are complementary to pre-labelled oligonucleotides added into the hybridization mix. The first control collectively tests the target fragmentation, labeling by biotinylation, array hybridization, and staining/scanning efficiency. It also allows the overall fluorescent intensity to be normalized across all the arrays in an experiment. The second control directly assays the hybridization, staining and scanning.

[0183] A microbial sample is treated with PMA followed by incubation in the dark. Thereafter, the sample is exposed to visible light for 5 minutes, so that the DNA from dead cells only will be cross-linked. Following the initial PMA treatment, the sample is concentrated by centrifugation, and DNA is extracted using the Maxwell 16 automated system (Promega Corp.). Extracted DNA is PCR amplified, and DNA is hybridized to the phylogenetic array to detect and optionally quantify OTUs represented by viable cells and/or spores from the sample.

[0184] Complementary targets to the probe sequences hybridize to the array and fluorescent signals are captured as pixel images using standard AFFYMETRIX® software (GeneChip Microarray Analysis Suite, version 5.1) that reduce the data to an individual signal value for each probe and is typically exported as a human readable CEL' file. Background probes are identified from the CEL file as those producing intensities in the lowest 2% of all intensities. The average intensity of the background probes is subtracted from the fluorescence intensity of all probes. The noise value (N) is the variation in pixel intensity signals observed by the scanner as it reads the array surface. The standard deviation of the pixel intensities within each of the identified background probe intensities is divided by the square root of the number of pixels comprising that feature. The average of the resulting quotients was used for N in the calculations described below.

[0185] Using previous methods, probe pairs scored as positive are those that meet two criteria: (i) the fluorescence intensity from the perfectly matched probe is at least 1.3 times greater than the intensity from the mismatched control probe, and (ii) the difference in intensity, PM minus MM, is at least 130 times greater than the squared noise value (>130 N2). The positive fraction (PosFrac) is calculated for each probe set as the number of positive probe pairs divided by the total number of probe pairs in a probe set. An OTU is considered 'present' when its PosFrac for the corresponding probe set was >0.92 (based on empirical data from clone library analyses). Replicate arrays can be used collectively in determining the presence of each OTU by requiring each to exceed a PosFrac threshold. Present calls are propagated upwards through the taxonomic hierarchy by considering any node (subfamily, family, order, etc.) as 'present' if at least one of its subordinate OTUs is present.

[0186] Hybridization intensity us the measure of OTU abundance and is calculated in arbitrary units for each probe set as the trimmed average (maximum and minimum values removed before averaging) of the PM minus MM intensity differences across the probe pairs in a given probe set. All intensities<1 are shifted to 1 to avoid errors in subsequent logarithmic transformations.

Example 4

Phylogenetic Array Analysis

[0187] Following sample preparation, application, incubation and washing, using standard techniques, phylogenetic arrays were scanned using a GeneArray Scanner from Affymetrix. The scan was captured as a pixel image using standard AFFYMETRIX® software (GCOS v1.6 using parameter: Percentile v6.0) that reduces the data to an individual row in a text-encoded table for each probe. See Table 3.

TABLE 3

Exemplary Display of Array Data
[INTENSITY]
NumberCells = 506944
CellHeader = XY

NPIXELS	MEAN	STDV
047.9	025	167.0
11060.2	025	4293.0
243.7	036	179.3
3681.5	025	4437.0

[0188] Each analysis system had approximately 1,016,000 cells, with 1 probe sequence per cell. The analysis system scanner recorded the signal intensity across the array, which ranges from 0 to 65,000 arbitrary units (a.u) in a regular grid

with ~30-45 pixels per cell. A 2 pixel margin was used between adjacent cells, leaving approximately 25-40 pixels per probe of usable signal. From these pixels, the AFFYMETRIX® software computed the 75th percentile average pixel intensity (denoted as the "MEAN"), the standard deviation of signal intensity among the about 25-40 pixels (denoted as the "STDV"), and the number of pixels used per cell (denoted as "NPIXELS"). Any cells that had pixels that were three standard deviations apart in signal intensity were classified as outliers.

[0189] The analysis systems were divided into a user-defined number of horizontal and vertical divisions. By default, four horizontal and four vertical divisions were created resulting in 16 regularly spaced sectors for independent background subtraction. The background intensity was computed independently for each quadrant, as the average signal intensity of the least intense 2% (by default) of probes in that quadrant. The background intensity was then subtracted from all probes before further computation.

[0190] The noise value was estimated according to recommendations in the AFFYMETRIX® GeneChip User Guide v3.3. Noise (N) was due to variations in pixel intensity signals observed by the scanner as it read the array surface and was calculated as the standard deviation of the pixel intensities within each of the identified background cells divided by the square root of the number of pixels comprising that cell. The average of the resulting quotients was used for N in the calculations described below:

$$N = \frac{\sum_{i \in B} \frac{Si}{\sqrt{pix_i}}}{scalarB}$$

[0191] where

[0192] B is a background cell

[0193] S_i is the standard deviation among the pixels in B

[0194] pix_i is the count of pixels in B

[0195] scalarB is the count of all background cells, cumulative

[0196] The intensities of all probes were then scaled so that the average observed signal intensity of the spiked in probes had a pre-determined signal strength. This was accomplished by finding a scaling factor (Sf) in order to force the mean response of the corresponding PM probes to a target mean using the equation below:

$$Sf = \bar{e}_t \Big/ \frac{\sum_{i \in Kpm} e_i}{scalarK_{pm}}$$

[0197] where

[0198] $\bar{e}_t$=targeted mean intensity (default: 2500)

[0199] $scalarK_{pm}$=count of probes complementing any spike-in

[0200] S_f=scaling factor

[0201] Typically, the pre-determined signal strengths ranged from about 0 to about 65,000. Once the scaling factor was derived, all cell intensities were multiplied by the scaling factor.

[0202] The noise (N) was scaled by the same factor: $N_s = N \times S_f$, where N_s=scaled noise, N=unscaled noise, and S_f=scaling factor.

[0203] As an alternative or optional step, MM probes with high hybridization signal responses were identified and the probe pairs were eliminated where:

$$\left[\left(\frac{MM}{PM} > srt_\gamma\right) \wedge (MM - PM > N_s \times sdtm_\gamma)\right] \vee [PM \in O] \vee [MM \in O]$$

[0204] where:

[0205] N_s=scaled noise

[0206] O=outlier set

The remaining probe pairs were scored by:

$$\left(\frac{PM}{MM} > srt\right) \wedge (PM - MM > N_s^2 \times sdtm)$$

[0207] where

[0208] PM=sealed intensity of the perfect match probe

[0209] MM=scaled intensity of the perfect match probe

[0210] srt=standard ratio threshold (default: 1.3)

[0211] sdtm=standard difference threshold multiplier (default: 130)

[0212] N_s=scaled noise

[0213] After classifying an OTU as "present", the present call was propagated upwards through the taxonomic hierarchy by considering any node (subfamily, family, order, etc.) as 'present' if at least one of its subordinate OTUs was present.

[0214] Hybridization intensity was the measure of OTU abundance and was calculated in arbitrary units for each probe set as the trimmed average (maximum and minimum values removed before averaging) of the PM minus MM intensity differences across the probe pairs in a given probe set.

[0215] While preferred embodiments of the present application have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

Example 5

Clean Room Sampling, PMA Processing and Phylogenetic Array Analysis

[0216] The two Class 100K certified as per Fed-Std-209E (equivalent ISO 14644-1 Class 8) cleanroom facilities were examined. Mission critical hardware and components were assembled in the first cleanroom site, Jet Propulsion Laboratory's Spacecraft Assembly Facility [JPL-SAF], whereas the second cleanroom site, JPL cleanroom 144, did not support mission critical activities but maintained its Class 100K certification. Both of these cleanroom facilities were operated at

a positive pressure, with temperatures in the range of 20+4° C., and relative humidity ranging from 30 to 50%. In both cases, the total hydrocarbon content of the facility air (gases and vapors) was below 15 ppm (calculated as methane). The Class 100K environment typically did not exceed 10K particulate obstruction levels during 30 days of assembly/integration activities. Measured non-volatile residue levels did not exceed 0.3 μg/cm^2.

[0217] For convenience and differentiation, 454 pyrosequencing-based sequence discrimination results in the clustering in what are referred to as "molecular operational taxonomic unit (MOTU)." Variation in DNA sequence among microbes can arise via naturally occurring evolutionary events and/or methodological errors (e.g., homopolymer repetition in pyrosequencing). It is the goal of the MOTU-based classification and clustering system presented here, and in detail elsewhere (Blaxter & Floyd, TRENDS in Ecology and Evolution 18: 268-269, 2003), to separate these two sources of sequence variation based on known error rates in sequencing and measured levels of difference across various taxonomical schemes. The accuracy and specificity of a MOTU-based system can be derived from measured levels of between-taxa and within-group variation from well-defined populations, and of observational error obtained by re-sequencing (Blaxter & Floyd, 2003). In a similar vein, PhyloChip DNA microarrays employ multiple ~25 nt probes which collectively represent the full length ~1.5-kb 16S rRNA gene of each taxon. In this study, PhyloChip-derived taxonomic units (OTU) are delineated in accordance with the hybridization scores of a given set of 37 or more specific 25-mer probes, which have been previously designed based on the prevalence of members of a given OTU, and dissimilarity in DNA sequences outside of the given OTU. Ultimately, a microorganism can be assigned to only one given MOTU via either similarity within a homologous sequenced DNA fragment or hybridization-based OTU, but neither MOTU nor OTU need be congruent with other taxonomic schemes.

[0218] PhyloChip G3 (DNA Microarray, Generation 3) Analysis:

[0219] Bacterial 16S rRNA genes were amplified from DNA preparations from each sample using the primers 27F (5'-AGA GTT TGA TCC TGG CTC AG) and 1492R (GGT TAC CTT GTT ACG ACT T). PCR conditions were as follows: 1 cycle of initial melting for 3 min at 95° C., followed by 35 cycles of 30 sec melting at 95° C., 30 sec annealing over a 48-57.5° C. gradient, and 2 min extension at 72° C., with a final 10 min incubation at 72° C. To maximize observed diversity, four separate PCR reactions were performed for each sample using a gradient of annealing temperatures (48° C., 50.1° C., 54.4° C., and 57.5° C.). Archaeal amplification of 16S rRNA genes was performed with primers 4aF and 1492uR, and resulting amplicons were purified via gel-excision if/when quantifiable amounts of product had been generated. A maximum of 450 ng of PCR product from each sample was used for phylogenetic microarray analysis. A detailed explanation of the processing of the PhyloChip assay has been described elsewhere (Hazen, et al., Science 330: 204-208, 2010). Briefly, pooled PCR products from each sampling event were spiked with known amounts of synthetic control 16S rRNA gene fragments and non-16S rRNA gene fragments. Fluorescence intensities arising from these controls were used as standards for normalization among samples. Target fragmentation, biotin labeling, PhyloChip hybridization, scanning and staining, as well as background subtraction, noise calculation and detection, and quantification criteria, were performed as previously reported (DeSantis, et al., Microb. Ecol. 53: 371-383, 2007; Hazen, et al., 2010). OTUs were deemed present if quartiles of the ranked r scores (response score to determine the potential of a probe pair responding to a target and not to the background) met the following criteria: rQ1≥0.70, rQ2≥0.95, rQ3≥0.98 (stage 1 analysis). Subfamilies that had an rxQ3 value (cross-hybridization adjusted response score) of ≥0.6 were deemed present following an additional stage 2 analysis requirement for OTU calling within this subfamily.

[0220] Phylochip G3 Data Processing:

[0221] Following stage 2 analysis, hybridization intensities were log 2*1000 transformed for linearization. Subfamily filtering was performed manually by choosing representative OTUs of each subfamily depending on how often a OTU was called present throughout the samples. OTUs and subfamilies were subsequently clustered and classified in accordance with the recently released Greengenes taxonomy (McDonald, et al., ISME J, 2011). In order to better understand the phylogenetic relationships of detected OTUs/subfamilies, multiple sequence alignments of representative sequences were retrieved from the SILVA database (Pruesse, et al., Nucleic Acids Res. 35: 7188-7196, 2007). Neighbor joining phylogenetic trees were compiled based on the 70,000 character alignments of each OTU via MEGA4 software (Tamura, et al., Mol. Biol. Evol. 24: 1596-1599, 2007). Subsequently, trees were mated with corresponding heatmaps (either presence/absence or relative abundance) and rendered in iTOL (Letunic & Bork, Bioinformatics 23: 127-128, 2007). Environmental clustering (Non-metric Multi-Dimensional Scaling, N-MDS) based on abundance scores and Bray Curtis distance was performed using the R programming environment (www.r-project.org, Vegan and MASS package). Weighted principal component analysis was performed using the FAST Unifrac interface, where OTUs were grouped into subfamilies and the number of distinct OTUs per subfamily served as weighting (Hamady, et al., ISME J 4: 17-27, 2010). Pearson correlation of abundance values between observed OTUs and various environmental factors (pH, temperature) were generated with Microsoft Excel (2007). Correlations with r>|.877| for comparison among three samples and r>|.811| for comparison among four samples were considered statistically significant (null hypothesis: OTUs do not correlate with the environmental factor tested). Graphs were constructed with the SigmaPlot 10.0 software suite.

[0222] PhyloChip G3 DNA microarray analyses and associated biostatistical processing were carried out as described recently (Hazen et al., "Deep-sea oil plume enriches indigenous oil-degrading bacteria," Science 330(6001):204-208 (2010); Cooper et al., "Comparison of innovative molecular approaches and standard spore assays for assessment of surface cleanliness." Appl. Environ. Microbiol., 77(15):5438-5444 (2011), the contents of which are hereby expressly incorporated by reference). OTUs that were called present in negative control samples (e.g., sampling blanks, handling controls) were removed from the entire data set. Non-metric multi dimensional scaling was performed on abundance values of resulting OTUs using the R programming environment (Vegan package, http://www.r-project.org/). Phylogenetic classification of OTUs was performed using GreenGenes (DeSantis, et al., Appl Environ Microbiol 72: 5069-5072, 2006) in combination with SILVA (Pruesse, et al., 2007) and RDP II (Cole, et al., 2005). OTUs that represents majority of

the living bacterial population were identified based on two criteria: A) the OTU should be present based on presence/absence call by the phyloChip analysis B) OUT should have a higher abundance value in PMA treated sample as compared to non-PMA treated sample. The total OTUs (1277) detected were grouped into 126 OTUs at genus level by selecting only those which exhibit higher abundance in PMA treated samples. Phylogenetic trees were constructed in MEGA 4 (Tamura, et al., 2007) based on an 70,000 character alignment (SILVA, (Pruesse, et al., 2007) and the neighbour joining method. Heat maps of abundance ratios and trees were rendered in iTOL (Letunic & Bork, 2007).

[0223] 454 Pyrosequencing and Data Analysis:

[0224] The 16S universal bacterial primers 28F-5' GAG TTT GAT CNT GGC TCA G 3' (SEQ ID NO: 101) and 519R 5'-GTN TTA CNG CGG CKG CTG-3' (SEQ ID NO:102) were used to PCR amplify the ~500 bp V1 to V3 hypervariable region of the 16S rRNA gene. Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed as described previously (Dowd, et al., 2008). In preparation for FLX-Titanium sequencing (Roche, Nutley, N.J.), template DNA fragment size and concentration were accurately measured with a TBS-380 Fluorometer using picogreen fluorescent dyes (Turner Biosystems, CA, USA). Total volume of PCR amplicon used for emulsion PCR was 2 uL for strong positive (>15 pg/uL); 5 uL for weak positive (5 and 15 pg/uL), and 20 uL for negative (<5 pg/uL) PCR products from individual samples. This normalization step helped to minimize bias in resulting sequence and/or tag frequency in favor of strong PCR products. A sample of 9.6×106 double-stranded DNA molecules/ml (mean=625 bp) were combined with 9.6× 106 DNA capture beads, and then amplified by emulsion PCR. Following bead recovery and enrichment, the bead attached DNAs were denatured with NaOH, and sequencing primers were annealed. A four-region 454 sequencing run was performed on a GS PicoTiterPlate (PTP) using the Genome Sequencer FLX System (Roche, Nutley, N.J.). A maximum of 40 distinctly barcoded samples were analyzed on each quarter region of the PTP. All FLX procedures were performed according to the Genome Sequencer FLX System's manufacturer's instructions (Roche, Nutley, N.J.). All bTEFAP procedures were performed at the Research and Testing Laboratory (RTL; Lubbock, Tex.) in accordance with well established protocols (www.researchandtesting.com).

[0225] bTEFAP-Derived Bacterial Diversity and Data Analysis:

[0226] All resulting 454 pyrosequences were processed and analyzed using the MOTHUR software package (Schloss, et al., Appl Environ Microbiol 75: 7537-7541, 2009). The AmpliconNoise algorithm was implemented to reduce sequencing error. Sequences were removed from the analysis if they i) did not contain the primer and/or barcode sequence, ii) had a total sequence length<200 bp, or iii) had a quality score<25. The filtered sequences were assigned to individual samples according to their previously engineered 12-nt barcodes. Unique sequences were aligned using the SILVA reference alignment (Schloss, 2009), and only the overlapping region was considered for analyses. After subtracting all chimeras from consideration, remaining high-quality sequences were clustered and classified according to the recently released Greengenes training set and taxonomy (McDonald, et al., 2011, Werner, et al., 2012). Sequences were clustered into molecular operational taxonomic units (MOTU) at the 0.03 level of resolution.

[0227] Controls:

[0228] Very low amounts of PCR products were obtained for BisKit blank, negative control (<2 pg/uL). This indicates that DNA associated with sampling devices and reagents used in this study is in extremely low concentration and extraneous nature rather than from intact bacterial cells. This extraneous DNA is not available for PCR reaction after PMA treatment hence yields further reduced PCR products.

[0229] Lab controls and field blanks were also treated with PMA to assess relative percent viability of contaminant microbiota. For all of these cleanroom control and blank samples rrn copy numbers remained at 10^2 irrespective of whether they had been subjected to PMA treatment. The signals detected in these control samples, implying a presence of <100 viable contaminant organisms, were present at insignificant levels in relation to those arising from the actual samples of interest, and were thus ignored

[0230] Results:

[0231] Total bacterial burden as assessed by universal qPCR, and the various characteristics of the samples collected during this study were measured. When PMA was omitted from the molecular processing, the qPCR-derived total bacterial burden (viable+non-viable) of the floors and GSE samples collected in the mission-critical JPL-SAF was ~10^6 rrn copies/m^2. After treatment with PMA, the viable bacterial population of JPL-SAF samples as measured via PMA-qPCR was ~7% and 41% for floors and GSE, respectively. Samples collected from the floor and GSE of the non-mission critical JPL-144 facility yielded much higher rrn copy numbers (more than one-log) both with and without PMA-treatment. However, the percent viable portion of the total bacterial population present in the JPL-144 samples was ~21% and 7% for floors and GSE samples, respectively. Upon comparative analysis, JPL-144 floors housed a ~66.7-fold greater total bacterial burden than their mission-critical JPL-SAF counterparts. Similarly, JPL-144 GSE samples were ~15% more laden with total bacteria (viable+non-viable) than the JPL-SAF GSE materials.

[0232] 454 Pyrosequencing:

[0233] The concentration of 500-bp amplicons resulting from pre-454 pyrosequencing PCR was <3.75 pg/uL for all samples treated with PMA, and >5 pg/uL in all samples when PMA treatment was omitted. A very strong 17 pg/uL PCR product resulted from one untreated JPL-144 floor sample (GI-42-1), which was 2.3-fold higher than that arising from its corresponding JPL-SAF floor sample (GI-36-4). The untreated JPL-144 GSE material (GI-42-2) exhibited 1.6-fold greater PCR amplicon concentrations than their untreated mission-critical JPL-SAF GSE counterparts (GI-36-3). After having been pretreated with PMA, even the mission critical SAC floor sample yielded fewer PCR products (5.4 pg/uL) than the GSE material housed in the JPL-144 cleanroom (8.9 pg/uL). In general, irrespective of the type of sample analyzed, a considerable decrease in PCR yield was observed for all samples pretreated with PMA.

[0234] A breakdown of the number of MOTUs observed in the various samples was examined over the course of this study. Overall, the JPL-144 GSE samples housed 4.7-fold more bacterial MOTUs than the mission-critical GSE, whereas the cleanroom floor surfaces from each of these cleanrooms yielded a roughly equivalent number of MOTUs. The mission critical JPL-SAF floor was rich in MOTUs affiliated with physiologically recalcitrant bacteria (Actinobacteria, Deinococci, Acidobacteria, and Firmicutes), whereas the

JPL-144 cleanroom floor harbored predominantly Proteobacterial MOTUs. It was particularly apparent that the few acidobacterial MOTUs (4 MOTUs) in JPL-SAF floor samples were present in great abundance (112 sequences). However, the GSE samples exhibited no such corollaries between MOTU numbers and sequence occurrence.

[0235] Anywhere from one to 528 high-quality pyrosequences (>250-bp) were obtained from samples that had been pretreated with PMA, whereas 1,783 to 15,914 high-quality pyrosequences were recovered from untreated samples. Regardless of sample type, PMA treated samples consistently yielded significantly fewer pyrosequences that their untreated counterparts. Even when PMA treatment was omitted, the mission critical JPL-SAF cleanroom floor sample (GI-36-4) yielded far fewer pyrosequences (4,318 pyrosequences) than the non-mission JPL-144 cleanroom floor sample (GI-42-1; 15,914 pyrosequences). Approximately 59% of all of the pyrosequences retrieved from these two distinct facility floor samples were present in both, though this shared fraction represented only 7.9% of the total observed MOTUs (69 out of 872). The relative abundance of pyrosequences retrieved from the untreated JPL-SAF samples was plotted as a Venn diagram. Even though equivalent surface areas were sampled from the floor and GSE of the JPL-SAF (18 m2 each), only 6,101 pyrosequences were retrieved from the GSE samples whereas the floor gave rise to 20,232 sequences. Between the JPL-SAF floor and GSE samples, ~38% of the total pyrosequences observed were shared, but constituted a mere ~7.5% of the total MOTUs (43 out of 574).

[0236] The MOTUs shared between two distinct sample sets, for example floors of JPL-SAF vs. JPL-144 or JPL-SAF floors vs. JPL-SAF GSE, were comparatively analyzed. Whenever the difference in the number of representative sequences per MOTU was less than 10, the comparison between MOTU sets was deemed moot and thus omitted from consideration. Following this subtraction, the cleanroom floor comparative sample set retained ~92% of the total pyrosequences observed, however, 87.5% of the total MOTUs (711 out of 872) were eliminated from this calculation. Similarly, ~40% of the total pyrosequences compared in the JPL-SAF floor vs. GSE sample set were eliminated, corresponding to a 77.5% loss of total MOTUs (445 out of 574).

[0237] Closer observation of sequences from JPL-SAF cleanroom floor and GSE samples indicates that they predominantly belong to hardier organisms such as *Acidobacteria* sp., *Actinobacteria* sp., *Arsenicicoccus* sp., *Arthrobacter* sp., *Corynebacterium* sp., *Kineococcus* sp., *Propionibacterium* sp., *Nocardioides* sp., *Streptomyces* sp., *Bacillus* sp., *Clostridium* sp., *Lactobacillus* sp., *Deinococcus* sp., *Staphylococcus* sp.

[0238] When treated with PMA, marked reduction in total number of sequences (89 and 1) and OTUs (17 and 1) was observed for JPL-SAF and 144 cleanroom floor samples respectively. The PMA effect was significantly higher in non-mission critical floor samples (447 to 1 MOTU) indicating large number of dead cells or extraneous DNA present on these floors. This viable population belongs to taxa Firmicutes, Actinobacteria and Proteobacteria. Similar trend was observed in PMA treated GSE samples (4 and 42 MOTUs) and (14 and 528 sequences) from JPL-SAF and 144 respectively. The hardy organisms observed in cleanroom floor samples without PMA treatment were absent or in very low number in PMA treated samples.

[0239] As described above, the system (e.g., PMA-microarrays) and the methods disclosed herein can be used to accurately and sensitively detect and quantify viable organisms in a sample.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 102

<210> SEQ ID NO 1
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 265003; MM probe is SEQ ID NO:2

<400> SEQUENCE: 1

aatggaattt gatagaacct ctgga 25

<210> SEQ ID NO 2
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 265003; PM probe is SEQ ID NO:1

<400> SEQUENCE: 2

aatggaattt gaaagaacct ctgga 25

<210> SEQ ID NO 3

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 231625; MM probe is SEQ ID NO:4

<400> SEQUENCE: 3

cagtaaggca tgagtaactt accgt 25

<210> SEQ ID NO 4
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 231625; PM probe is SEQ ID NO:3

<400> SEQUENCE: 4

cagtaaggca tgggtaactt accgt 25

<210> SEQ ID NO 5
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 186992; MM probe is SEQ ID NO:6

<400> SEQUENCE: 5

aaacgatgac cactagacgt tgggg 25

<210> SEQ ID NO 6
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 186992; PM probe is SEQ ID NO:5

<400> SEQUENCE: 6

aaacgatgac cattagacgt tgggg 25

<210> SEQ ID NO 7
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 327194; MM probe is SEQ ID NO:8

<400> SEQUENCE: 7

ataaagttac ttgcctgcat ttggt 25

<210> SEQ ID NO 8
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 327194; PM probe is SEQ ID NO:7

-continued

<400> SEQUENCE: 8

ataaagttac ttccctgcat ttggt 25

<210> SEQ ID NO 9
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 151981; MM probe is SEQ ID NO:10

<400> SEQUENCE: 9

ccgtagcgta cgtgatctac gatta 25

<210> SEQ ID NO 10
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 151981; PM probe is SEQ ID NO:9

<400> SEQUENCE: 10

ccgtagcgta cgggatctac gatta 25

<210> SEQ ID NO 11
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 191212; MM probe is SEQ ID NO:12

<400> SEQUENCE: 11

taaaatttcg atggatacgg attgg 25

<210> SEQ ID NO 12
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 191212; PM probe is SEQ ID NO:11

<400> SEQUENCE: 12

taaaatttcg atcgatacgg attgg 25

<210> SEQ ID NO 13
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 276480; MM probe is SEQ ID NO:14

<400> SEQUENCE: 13

caggcacgaa aacgtgggga tccaa 25

<210> SEQ ID NO 14

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 276480; PM probe is SEQ ID NO:13

<400> SEQUENCE: 14

caggcacgaa aaggtgggga tccaa 25

<210> SEQ ID NO 15
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 300668; MM probe is SEQ ID NO:16

<400> SEQUENCE: 15

cggttctttt tgatggcgac cggcg 25

<210> SEQ ID NO 16
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 300668; PM probe is SEQ ID NO:15

<400> SEQUENCE: 16

cggttctttt tgttggcgac cggcg 25

<210> SEQ ID NO 17
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 25579; MM probe is SEQ ID NO:18

<400> SEQUENCE: 17

ggccttagcc cgtgctgccg tacga 25

<210> SEQ ID NO 18
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 25579; PM probe is SEQ ID NO:17

<400> SEQUENCE: 18

ggccttagcc cgagctgccg tacga 25

<210> SEQ ID NO 19
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 176744; MM probe is SEQ ID NO:20

-continued

<400> SEQUENCE: 19

cgcatgtcga gtatgttaaa tatcc 25

<210> SEQ ID NO 20
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 176744; PM probe is SEQ ID NO:19

<400> SEQUENCE: 20

cgcatgtcga gtttgttaaa tatcc 25

<210> SEQ ID NO 21
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 263709; MM probe is SEQ ID NO:22

<400> SEQUENCE: 21

cccaggttcg atagtaccaa ttagg 25

<210> SEQ ID NO 22
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 263709; PM probe is SEQ ID NO:21

<400> SEQUENCE: 22

cccaggttcg atggtaccaa ttagg 25

<210> SEQ ID NO 23
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 289612; MM probe is SEQ ID NO:24

<400> SEQUENCE: 23

ccaaggacgt tcgtgagccg ttcga 25

<210> SEQ ID NO 24
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 289612; PM probe is SEQ ID NO:23

<400> SEQUENCE: 24

ccaaggacgt tcttgagccg ttcga 25

<210> SEQ ID NO 25

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 4284; MM probe is SEQ ID NO:26

<400> SEQUENCE: 25

aaacgaccca ctgggactgt cctga 25

<210> SEQ ID NO 26
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 4284; PM probe is SEQ ID NO:25

<400> SEQUENCE: 26

aaacgaccca ctaggactgt cctga 25

<210> SEQ ID NO 27
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 111799; MM probe is SEQ ID NO:28

<400> SEQUENCE: 27

aaaattgccc gaaggaggat ctatt 25

<210> SEQ ID NO 28
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 111799; PM probe is SEQ ID NO:27

<400> SEQUENCE: 28

aaaattgccc gatggaggat ctatt 25

<210> SEQ ID NO 29
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 57769; MM probe is SEQ ID NO:30

<400> SEQUENCE: 29

atattactcc ccataacatt cgtcc 25

<210> SEQ ID NO 30
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 57769; PM probe is SEQ ID NO:29

-continued

<400> SEQUENCE: 30

atattactcc ccctaacatt cgtcc 25

<210> SEQ ID NO 31
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 277222; MM probe is SEQ ID NO:32

<400> SEQUENCE: 31

aacttcctta gctccaaata cattc 25

<210> SEQ ID NO 32
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 277222; PM probe is SEQ ID NO:31

<400> SEQUENCE: 32

aacttcctta gcaccaaata cattc 25

<210> SEQ ID NO 33
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 39458; MM probe is SEQ ID NO:34

<400> SEQUENCE: 33

aagtcccgta cgagcgcaac ccttg 25

<210> SEQ ID NO 34
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 39458; PM probe is SEQ ID NO:33

<400> SEQUENCE: 34

aagtcccgta cgcgcgcaac ccttg 25

<210> SEQ ID NO 35
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 298577; MM probe is SEQ ID NO:36

<400> SEQUENCE: 35

aaggcccggc actaactatg accgg 25

<210> SEQ ID NO 36

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 298577; PM probe is SEQ ID NO:35

<400> SEQUENCE: 36

aaggcccggc acgaactatg accgg 25

<210> SEQ ID NO 37
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 126538; MM probe is SEQ ID NO:38

<400> SEQUENCE: 37

acagtactta aagaataagc accgg 25

<210> SEQ ID NO 38
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 126538; PM probe is SEQ ID NO:37

<400> SEQUENCE: 38

acagtactta aaaaataagc accgg 25

<210> SEQ ID NO 39
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 252982; MM probe is SEQ ID NO:40

<400> SEQUENCE: 39

aaatattggg gctcaacctt aagca 25

<210> SEQ ID NO 40
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 252982; PM probe is SEQ ID NO:39

<400> SEQUENCE: 40

aaatattggg gcgcaacctt aagca 25

<210> SEQ ID NO 41
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 197225; MM probe is SEQ ID NO:42

-continued

<400> SEQUENCE: 41

cgtatagtac ggtcagtttt ctaca 25

<210> SEQ ID NO 42
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 197225; PM probe is SEQ ID NO:41

<400> SEQUENCE: 42

cgtatagtac gggcagtttt ctaca 25

<210> SEQ ID NO 43
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 162084; MM probe is SEQ ID NO:44

<400> SEQUENCE: 43

ccgaaggacc ggttctcgct cgact 25

<210> SEQ ID NO 44
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 162084; PM probe is SEQ ID NO:43

<400> SEQUENCE: 44

ccgaaggacc ggctctcgct cgact 25

<210> SEQ ID NO 45
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 228980; MM probe is SEQ ID NO:46

<400> SEQUENCE: 45

cctagttacc gatcttgccc taggc 25

<210> SEQ ID NO 46
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 228980; PM probe is SEQ ID NO:45

<400> SEQUENCE: 46

cctagttacc gagcttgccc taggc 25

<210> SEQ ID NO 47

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 123763; MM probe is SEQ ID NO:48

<400> SEQUENCE: 47

cacagatttc cctgtgttat tcccc 25

<210> SEQ ID NO 48
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 123763; PM probe is SEQ ID NO:47

<400> SEQUENCE: 48

cacagatttc ccggtgttat tcccc 25

<210> SEQ ID NO 49
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 240310; MM probe is SEQ ID NO:50

<400> SEQUENCE: 49

tagcaataca tcaggtagtt agtgg 25

<210> SEQ ID NO 50
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 240310; PM probe is SEQ ID NO:49

<400> SEQUENCE: 50

tagcaataca tccggtagtt agtgg 25

<210> SEQ ID NO 51
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 16702; MM probe is SEQ ID NO:52

<400> SEQUENCE: 51

gactcaacag gcatcacatc agata 25

<210> SEQ ID NO 52
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 16702; PM probe is SEQ ID NO:51

-continued

<400> SEQUENCE: 52

gactcaacag gcttcacatc agata 25

<210> SEQ ID NO 53
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 16703; MM probe is SEQ ID NO:54

<400> SEQUENCE: 53

tatctgatgt gatgcctgtt gagtc 25

<210> SEQ ID NO 54
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 16703; PM probe is SEQ ID NO:53

<400> SEQUENCE: 54

tatctgatgt gaagcctgtt gagtc 25

<210> SEQ ID NO 55
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 24979; MM probe is SEQ ID NO:56

<400> SEQUENCE: 55

tcgggcgttg ttaaatcaat caaga 25

<210> SEQ ID NO 56
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 24979; PM probe is SEQ ID NO:55

<400> SEQUENCE: 56

tcgggcgttg tttaatcaat caaga 25

<210> SEQ ID NO 57
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 24980; MM probe is SEQ ID NO:58

<400> SEQUENCE: 57

tcttgattga tttaacaacg cccga 25

<210> SEQ ID NO 58

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 24980; PM probe is SEQ ID NO:57

<400> SEQUENCE: 58

tcttgattga ttaaacaacg cccga 25

<210> SEQ ID NO 59
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 41927; MM probe is SEQ ID NO:60

<400> SEQUENCE: 59

aagcatctga ctcaacaggc atcac 25

<210> SEQ ID NO 60
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 41927; PM probe is SEQ ID NO:59

<400> SEQUENCE: 60

aagcatctga ctgaacaggc atcac 25

<210> SEQ ID NO 61
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 41928; MM probe is SEQ ID NO:62

<400> SEQUENCE: 61

gtgatgcctg ttgagtcaga tgctt 25

<210> SEQ ID NO 62
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 41928; PM probe is SEQ ID NO:61

<400> SEQUENCE: 62

gtgatgcctg ttcagtcaga tgctt 25

<210> SEQ ID NO 63
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 44977; MM probe is SEQ ID NO:64

-continued

<400> SEQUENCE: 63

ctgatcgtat gtatggtcta tggcc 25

<210> SEQ ID NO 64
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 44977; PM probe is SEQ ID NO:63

<400> SEQUENCE: 64

ctgatcgtat gtttggtcta tggcc 25

<210> SEQ ID NO 65
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 44978; MM probe is SEQ ID NO:66

<400> SEQUENCE: 65

ggccatagac catacatacg atcag 25

<210> SEQ ID NO 66
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 44978; PM probe is SEQ ID NO:65

<400> SEQUENCE: 66

ggccatagac caaacatacg atcag 25

<210> SEQ ID NO 67
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 46837; MM probe is SEQ ID NO:68

<400> SEQUENCE: 67

taaatcaatc aagacatccg gttgt 25

<210> SEQ ID NO 68
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 46837; PM probe is SEQ ID NO:67

<400> SEQUENCE: 68

taaatcaatc aacacatccg gttgt 25

<210> SEQ ID NO 69

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 46838; MM probe is SEQ ID NO:70

<400> SEQUENCE: 69

acaaccggat gtcttgattg attta 25

<210> SEQ ID NO 70
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 46838; PM probe is SEQ ID NO:69

<400> SEQUENCE: 70

acaaccggat gtgttgattg attta 25

<210> SEQ ID NO 71
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 48919; MM probe is SEQ ID NO:72

<400> SEQUENCE: 71

cgatgatggc tccgatccct ttttc 25

<210> SEQ ID NO 72
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 48919; PM probe is SEQ ID NO:71

<400> SEQUENCE: 72

cgatgatggc tcggatccct ttttc 25

<210> SEQ ID NO 73
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 48920; MM probe is SEQ ID NO:74

<400> SEQUENCE: 73

gaaaaaggga tcggagccat catcg 25

<210> SEQ ID NO 74
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 48920; PM probe is SEQ ID NO:73

-continued

<400> SEQUENCE: 74

gaaaaaggga tccgagccat catcg 25

<210> SEQ ID NO 75
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 74275; MM probe is SEQ ID NO:76

<400> SEQUENCE: 75

cagtatcgca ccgagcgcaa aattt 25

<210> SEQ ID NO 76
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 74275; PM probe is SEQ ID NO:75

<400> SEQUENCE: 76

cagtatcgca cccagcgcaa aattt 25

<210> SEQ ID NO 77
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 74276; MM probe is SEQ ID NO:78

<400> SEQUENCE: 77

aaattttgcg ctcggtgcga tactg 25

<210> SEQ ID NO 78
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 74276; PM probe is SEQ ID NO:77

<400> SEQUENCE: 78

aaattttgcg ctgggtgcga tactg 25

<210> SEQ ID NO 79
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 80839; MM probe is SEQ ID NO:80

<400> SEQUENCE: 79

aatcagcttc tgagaaaccg gttgt 25

<210> SEQ ID NO 80

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 80839; PM probe is SEQ ID NO:79

<400> SEQUENCE: 80

aatcagcttc tgtgaaaccg gttgt 25

<210> SEQ ID NO 81
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 80840; MM probe is SEQ ID NO:82

<400> SEQUENCE: 81

acaaccggtt tctcagaagc tgatt 25

<210> SEQ ID NO 82
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 80840; PM probe is SEQ ID NO:81

<400> SEQUENCE: 82

acaaccggtt tcacagaagc tgatt 25

<210> SEQ ID NO 83
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 83957; MM probe is SEQ ID NO:84

<400> SEQUENCE: 83

ctcaacaggc atcacatcag ataat 25

<210> SEQ ID NO 84
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 83957; PM probe is SEQ ID NO:83

<400> SEQUENCE: 84

ctcaacaggc atgacatcag ataat 25

<210> SEQ ID NO 85
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 83958; MM probe is SEQ ID NO:86

-continued

<400> SEQUENCE: 85

attatctgat gtgatgcctg ttgag 25

<210> SEQ ID NO 86
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 83958; PM probe is SEQ ID NO:85

<400> SEQUENCE: 86

attatctgat gtcatgcctg ttgag 25

<210> SEQ ID NO 87
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 93455; MM probe is SEQ ID NO:88

<400> SEQUENCE: 87

tttctgtaaa acaggcatgg atatc 25

<210> SEQ ID NO 88
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 93455; PM probe is SEQ ID NO:87

<400> SEQUENCE: 88

tttctgtaaa actggcatgg atatc 25

<210> SEQ ID NO 89
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 93456; MM probe is SEQ ID NO:90

<400> SEQUENCE: 89

gatatccatg cctgttttac agaaa 25

<210> SEQ ID NO 90
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 93456; PM probe is SEQ ID NO:89

<400> SEQUENCE: 90

gatatccatg ccagttttac agaaa 25

<210> SEQ ID NO 91

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 140612; MM probe is SEQ ID NO:92

<400> SEQUENCE: 91

ctcaacatcc tcaaaatagt tggca 25

<210> SEQ ID NO 92
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 140612; PM probe is SEQ ID NO:91

<400> SEQUENCE: 92

ctcaacatcc tctaaatagt tggca 25

<210> SEQ ID NO 93
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 140613; MM probe is SEQ ID NO:94

<400> SEQUENCE: 93

tgccaactat tttgaggatg ttgag 25

<210> SEQ ID NO 94
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 140613; PM probe is SEQ ID NO:93

<400> SEQUENCE: 94

tgccaactat ttagaggatg ttgag 25

<210> SEQ ID NO 95
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 193149; MM probe is SEQ ID NO:96

<400> SEQUENCE: 95

atttcatcat cagtatcgca ccgag 25

<210> SEQ ID NO 96
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 193149; PM probe is SEQ ID NO:95

-continued

<400> SEQUENCE: 96

atttcatcat cactatcgca ccgag 25

<210> SEQ ID NO 97
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 193150; MM probe is SEQ ID NO:98

<400> SEQUENCE: 97

ctcggtgcga tactgatgat gaaat 25

<210> SEQ ID NO 98
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 193150; PM probe is SEQ ID NO:97

<400> SEQUENCE: 98

ctcggtgcga tagtgatgat gaaat 25

<210> SEQ ID NO 99
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 195043; MM probe is SEQ ID NO:100

<400> SEQUENCE: 99

taaaacaggc atggatatct gtgta 25

<210> SEQ ID NO 100
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
Pair ID: 195043; PM probe is SEQ ID NO:99

<400> SEQUENCE: 100

taaaacaggc atcgatatct gtgta 25

<210> SEQ ID NO 101
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 16S universal bacterial primer 28F
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 11
<223> OTHER INFORMATION: N = A, C, T, or G

<400> SEQUENCE: 101

gagtttgatc ntggctcag 19

-continued

<210> SEQ ID NO 102
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 16S universal bacterial primer 519R
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 3
<223> OTHER INFORMATION: N = A, C, T, or G
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 8
<223> OTHER INFORMATION: N = A, C, T, or G

<400> SEQUENCE: 102

gtnttacngc ggckgctg 18

What is claimed is:

1. A method for detecting live cells in a sample comprising:

selectively amplifying a nucleic acid from live cells from a sample comprising live and dead cells; and

detecting the presence, absence, relative abundance, and/or quantity of one or more operational taxon units (OTUs) in the sample based on hybridization of amplified nucleic acid to a plurality of probes complementary to 16s rRNA sequences, wherein said one or more OTUs consist essentially of live cells from said sample.

2. The method of claim **1**, wherein said nucleic acid is DNA.

3. The method of claim **1**, wherein said selectively amplifying comprises pre-treating the sample with an agent that selectively modifies a nucleic acid of dead cells.

4. The method of claim **3**, wherein the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent.

5. The method of claim **4**, wherein said DNA intercalating agent is propidium monoazide.

6. The method of claim **1**, wherein said probes are used to detect the presence, absence, relative abundance, and/or quantity of at least 10,000 different OTUs in a single assay.

7. The method of claim **1**, wherein said presence, absence, relative abundance, and/or quantity is detected with a confidence level greater than 95%.

8. The method of claim **1**, further comprising quantifying the number of live cells in the sample.

9. A method for detecting live cells in a sample comprising:

selectively amplifying a nucleic acid from live cells from a sample comprising live and dead cells; and

determining the presence, absence, relative abundance, and/or quantity of at least 1,000 different OTUs in a single assay, wherein said OTUs consist essentially of live cells from said sample.

10. A method for detecting live cells in a sample comprising:

(a) selectively amplifying a nucleic acid from live cells from a sample comprising live and dead cells;

(b) hybridizing amplified nucleic acid to a plurality of probes;

(c) determining hybridization signal strength distributions for a plurality of different interrogation probes, each of which is complementary to a section within one or more highly conserved polynucleotides in one or more target OTUs;

(d) determining hybridization signal strength distributions for a plurality of mismatch probes, wherein for each interrogation probe, one or more different corresponding mismatch probes comprising one or more nucleotide mismatches with said section within said one or more highly conserved polynucleotides are included in the plurality of mismatch probes; and

(e) using the hybridization signal strengths of the interrogation probes and mismatch probes to determine the probability that the hybridization signal for the different interrogation probes represents the presence, absence, relative abundance, and/or quantity of said one or more OTUs, wherein said one or more OTUs consist essentially of live cells from said sample.

11. The method of claim **9** or **10**, wherein said selectively amplifying comprises pre-treating the sample with an agent that selectively modifies a nucleic acid of dead cells.

12. The method of claim **11**, wherein the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent.

13. The method of claim **12**, wherein said DNA intercalating agent is propidium monoazide.

14. The method of claim **9** or **10**, wherein contacting said sample with an agent that selectively modifies a nucleic acid of dead cells in said sample is followed by exposure to visible light.

15. The method of claim **10**, wherein said highly conserved polynucleotides are selected from the group consisting of 16S rRNA gene, 23S rRNA gene, 5S rRNA gene, 5.8S rRNA gene, 12S rRNA gene, 18S rRNA gene, 28S rRNA gene, gyrB gene, rpoB gene, fusA gene, recA gene, coxl gene, nif13 gene, RNA molecules derived therefrom, and a combination thereof.

16. The method of claim **10**, wherein each interrogation probe has 4 or more corresponding mismatch probes in the plurality of mismatch probes.

17. The method of claim **10**, wherein said probes are used to detect the presence, absence, relative abundance, and/or quantity of at least 10,000 different OTUs in a single assay.

18. The method of claim **10**, wherein said probes are attached to a substrate.

19. The method of claim **18**, wherein said substrate comprises a bead or a microsphere.

20. The method of claim **18**, wherein said substrate comprises glass, plastic, or silicon

21. The method of claim **9** or **10**, wherein said presence, absence, relative abundance, and/or quantity is detected with a confidence level greater than 95%.

22. The method of claim **9**, further comprising quantifying the number of live cells in the sample.

23. A kit for detecting live cells in a sample comprising:

(a) an agent that selectively modifies a nucleic acid of dead cells;

(b) a plurality of different interrogation probes, each of which is complementary to a section within one or more highly conserved polynucleotides in one or more target operational taxon units (OTUs); and,

(c) a plurality of mismatch probes, wherein for each interrogation probe, one or more different corresponding mismatch probes comprising one or more nucleotide mismatches with said section within said one or more highly conserved polynucleotides are included in the plurality of mismatch probes.

24. The kit of claim **23**, wherein the agent that selectively modifies a nucleic acid of dead cells is a DNA intercalating agent.

25. The kit of claim **23**, wherein said DNA intercalating agent is propidium monoazide.

26. The kit of claim **23**, wherein each interrogation probe has 4 or more corresponding mismatch probes in the plurality of mismatch probes.

27. The kit of claim **23**, wherein said highly conserved polynucleotides are selected from the group consisting of 16S rRNA gene, 23S rRNA gene, 5S rRNA gene, 5.8S rRNA gene, 12S rRNA gene, 18S rRNA gene, 28S rRNA gene, gyrB gene, rpoB gene, fusA gene, recA gene, coxl gene, nif13 gene, RNA molecules derived therefrom, and a combination thereof.

* * * * *