Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2560521 C 2012/01/03

(11)(21) 2 560 521

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2005/02/18
(87) Date publication PCT/PCT Publication Date: 2006/07/06
(45) Date de délivrance/Issue Date: 2012/01/03
(85) Entrée phase nationale/National Entry: 2006/09/19
(86) N° demande PCT/PCT Application No.: US 2005/006133
(87) N° publication PCT/PCT Publication No.: 2006/071241
(30) Priorités/Priorities: 2004/02/18 (US60/545,425); 2004/04/05 (US60/559,754); 2004/12/03 (US60/632,862); 2004/12/22 (US60/639,068); 2005/01/28 (US60/648,188)

(51) Cl.Int./Int.Cl. *C07H 21/04* (2006.01), *C07H 21/00* (2006.01), *C12P 19/34* (2006.01), *C12Q 1/68* (2006.01), *C40B 40/06* (2006.01), *C12Q 1/06* (2006.01)

(72) Inventeurs/Inventors:
SAMPATH, RANGARAJAN, US;
HALL, THOMAS A., US;
ECKER, DAVID J., US;
ESHOO, MARK W., US;
MASSIRE, CHRISTIAN, US

(73) Propriétaire/Owner:
IBIS BIOSCIENCES, INC., US

(74) Agent: SMART & BIGGAR

(54) Titre : COMPOSITIONS DESTINEES A L'IDENTIFICATION DE BACTERIES
(54) Title: COMPOSITIONS FOR USE IN IDENTIFICATION OF BACTERIA

(57) Abrégé/Abstract:
The present invention provides oligonucleotide primers and compositions and kits containing the same for rapid identification of bacteria by amplification of a segment of bacterial nucleic acid followed by molecular mass analysis.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 July 2006 (06.07.2006)

PCT

(10) International Publication Number
WO 2006/071241 A2

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

(21) International Application Number:
PCT/US2005/006133

(22) International Filing Date:
18 February 2005 (18.02.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/545,425	18 February 2004 (18.02.2004)	US
60/559,754	5 April 2004 (05.04.2004)	US
60/632,862	3 December 2004 (03.12.2004)	US
60/639,068	22 December 2004 (22.12.2004)	US
60/648,188	28 January 2005 (28.01.2005)	US

(71) Applicant *(for all designated States except US)*: **ISIS PHARMACEUTICALS, INC.** [US/US]; 2292 Faraday Avenue, Carlsbad, CA 92008 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SAMPATH, Ramgarajan** [US/US]; 12223 Mannix Road, San Diego, CA 92129 (US). **HALL, Thomas, A.** [US/US]; 5239 Wohlford Street, Oceanside, CA 92056 (US). **ECKER, David, J.** [US/US]; 1041 Saxony Road, Encinitas, CA 92024 (US). **ESHOO, Mark, W.** [US/US]; 615 Glenmont Drive, Solana Beach, CA 92075 (US). **MASSIRE, Christian** [US/US]; 7498 Altiva Place, Carlsbad, CA 92009 (US).

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

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

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

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2006/071241 A2

(54) Title: COMPOSITIONS FOR USE IN IDENTIFICATION OF BACTERIA

(57) Abstract: The present invention provides oligonucleotide primers and compositions and kits containing the same for rapid identification of bacteria by amplification of a segment of bacterial nucleic acid followed by molecular mass analysis.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS COMPREND PLUS D'UN TOME.

CECI EST LE TOME __1___ DE __2 __

NOTE: Pour les tomes additionels, veillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE THAN ONE VOLUME.

THIS IS VOLUME __1___ OF __2___

NOTE: For additional volumes please contact the Canadian Patent Office.

COMPOSITIONS FOR USE IN IDENTIFICATION OF BACTERIA

[0001]

STATEMENT OF GOVERNMENT SUPPORT

[0002] This invention was made with United States Government support under DARPA/SPO contract BAA00-09. The United States Government may have certain rights in the invention.

FIELD OF THE INVENTION

[0003] The present invention relates generally to the field of genetic identification of bacteria and provides nucleic acid compositions and kits useful for this purpose when combined with molecular mass analysis.

BACKGROUND OF THE INVENTION

[0004] A problem in determining the cause of a natural infectious outbreak or a bioterrorist attack is the sheer variety of organisms that can cause human disease. There are over 1400 organisms infectious to humans; many of these have the potential to emerge suddenly in a natural epidemic or to be used in a malicious attack by bioterrorists (Taylor et al. Philos. Trans. R. Soc. London B. Biol. Sci., 2001, 356, 983-989). This number does not include numerous strain variants, bioengineered versions, or pathogens that infect plants or animals.

[0005] Much of the new technology being developed for detection of biological weapons incorporates a polymerase chain reaction (PCR) step based upon the use of highly specific primers and probes designed to selectively detect certain pathogenic organisms. Although this approach is appropriate for the most obvious bioterrorist organisms, like smallpox and anthrax, experience has shown that it is very difficult to predict which of hundreds of possible pathogenic organisms might be employed in a terrorist attack. Likewise, naturally emerging human disease that has caused devastating consequence in public health has come from unexpected families of

bacteria, viruses, fungi, or protozoa. Plants and animals also have their natural burden of infectious disease agents and there are equally important biosafety and security concerns for agriculture.

[0006] A major conundrum in public health protection, biodefense, and agricultural safety and security is that these disciplines need to be able to rapidly identify and characterize infectious agents, while there is no existing technology with the breadth of function to meet this need. Currently used methods for identification of bacteria rely upon culturing the bacterium to effect isolation from other organisms and to obtain sufficient quantities of nucleic acid followed by sequencing of the nucleic acid, both processes which are time and labor intensive.

[0007] Mass spectrometry provides detailed information about the molecules being analyzed, including high mass accuracy. It is also a process that can be easily automated. DNA chips with specific probes can only determine the presence or absence of specifically anticipated organisms. Because there are hundreds of thousands of species of benign bacteria, some very similar in sequence to threat organisms, even arrays with 10,000 probes lack the breadth needed to identify a particular organism.

[0008] There is a need for a method for identification of bioagents which is both specific and rapid, and in which no culture or nucleic acid sequencing is required. Disclosed in U.S. published Patents: 2003-0027135; 2003-0228571; 2004-0209260; 2004-0219517; 2009-0280471; 2005-0266397 and in U.S. issued Patent Nos.: 7,217,510; 7,226,739; 7,255,992, each of which is commonly owned, are methods for identification of bioagents (any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus) in an unbiased manner by molecular mass and base composition analysis of "bioagent identifying amplicons" which are obtained by amplification of segments of essential and conserved genes which are involved in, for example, translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like. Examples of these proteins include, but are not limited to, ribosomal RNAs, ribosomal proteins, DNA and RNA polymerases, elongation factors, tRNA synthetases, protein chain initiation factors, heat shock protein groEL, phosphoglycerate kinase, NADH dehydrogenase, DNA ligases, DNA gyrases and DNA topoisomerases, metabolic enzymes, and the like.

[0009] To obtain bioagent identifying amplicons, primers are selected to hybridize to conserved sequence regions which bracket variable sequence regions to yield a segment of nucleic acid which can be amplified and which is amenable to methods of molecular mass analysis. The variable sequence regions provide the variability of molecular mass which is used for bioagent identification. Upon amplification by PCR or other amplification methods with the specifically chosen primers, an amplification product that represents a bioagent identifying amplicon is obtained. The molecular mass of the amplification product, obtained by mass spectrometry for example, provides the means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass of the amplification product or the corresponding base composition (which can be calculated from the molecular mass of the amplification product) is compared with a database of molecular masses or base compositions and a match indicates the identity of the bioagent. Furthermore, the method can be applied to rapid parallel analyses (for example, in a multi-well plate format) the results of which can be employed in a triangulation identification strategy which is amenable to rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent identification.

[0010] The result of determination of a previously unknown base composition of a previously unknown bioagent (for example, a newly evolved and heretofore unobserved bacterium or virus) has downstream utility by providing new bioagent indexing information with which to populate base composition databases. The process of subsequent bioagent identification analyses is thus greatly improved as more base composition data for bioagent identifying amplicons becomes available.

[0011] The present invention provides oligonucleotide primers and compositions and kits containing the oligonucleotide primers, which define bacterial bioagent identifying amplicons and, upon amplification, produce corresponding amplification products whose molecular masses provide the means to identify bacteria, for example, at and below the species taxonomic level.

SUMMARY OF THE INVENTION

[0012] The present invention provides primers and compositions comprising pairs of primers, and kits containing the same for use in identification of bacteria. The primers are designed to produce bacterial bioagent identifying amplicons of DNA encoding genes essential to life such as, for example, 16S and 23S rRNA, DNA-directed RNA polymerase subunits (rpoB and rpoC),

valyl-tRNA synthetase (valS), elongation factor EF-Tu (TufB), ribosomal protein L2 (rplB), protein chain initiation factor (infB), and spore protein (sspE). The invention further provides drill-down primers, compositions comprising pairs of primers and kits containing the same, which are designed to provide sub-species characterization of bacteria.

[0013] The present invention also provides an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 97, or a composition comprising the same; an oligonucleotide primer 20 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 451, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 97, and a second oligonucleotide primer 20 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 451.

[0014] The present invention also provides an oligonucleotide primer 19 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 127, or a composition comprising the same; an oligonucleotide primer 14 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 482, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 19 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 127, and a second oligonucleotide primer 14 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 482.

[0015] The present invention also provides an oligonucleotide primer 19 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 174, or a composition comprising the same; an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 530, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 19 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 174, and a second oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 530.

[0016] The present invention also provides an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 310, or a composition

comprising the same; an oligonucleotide primer 19 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 668, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 310, and a second oligonucleotide primer 19 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 668.

[0017] The present invention also provides an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 313, or a composition comprising the same; an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 670, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 313, and a second oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 670.

[0018] The present invention also provides an oligonucleotide primer 17 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 277, or a composition comprising the same; an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 632, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 17 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 277, and a second oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 632.

[0019] The present invention also provides an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 285, or a composition comprising the same; an oligonucleotide primer 19 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 640, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 285, and a second oligonucleotide primer 19 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 640.

[0020] The present invention also provides an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 301, or a composition comprising the same; an oligonucleotide primer 21 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 656, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 301, and a second oligonucleotide primer 21 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 656.

[0021] The present invention also provides an oligonucleotide primer 18 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 308, or a composition comprising the same; an oligonucleotide primer 18 to 35 nucleobases in length comprising 70% to 100% sequence identity with SEQ ID NO: 663, or a composition comprising the same; a composition comprising both primers; and a composition comprising a first oligonucleotide primer 18 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 308, and a second oligonucleotide primer 18 to 35 nucleobases in length comprising between 70% to 100% sequence identity of SEQ ID NO: 663.

[0022] The present invention also provides compositions, such as those described herein, wherein either or both of the first and second oligonucleotide primers comprise at least one modified nucleobase, a non-templated T residue on the 5'-end, at least one non-template tag, or at least one molecular mass modifying tag, or any combination thereof.

[0023] The present invention also provides kits comprising any of the compositions described herein. The kits can comprise at least one calibration polynucleotide, or at least one ion exchange resin linked to magnetic beads, or both.

[0024] The present invention also provides methods for identification of an unknown bacterium. Nucleic acid from the bacterium is amplified using any of the compositions described herein to obtain an amplification product. The molecular mass of the amplification product is determined. Optionally, the base composition of the amplification product is determined from the molecular mass. The base composition or molecular mass is compared with a plurality of base compositions or molecular masses of known bacterial bioagent identifying amplicons, wherein a match between the base composition or molecular mass and a member of the plurality of base

compositions or molecular masses identifies the unknown bacterium. The molecular mass can be measured by mass spectrometry. In addition, the prrsence or absence of a particular clade, genus, species, or sub-species of a bioagent can be determined by the methods described herein.

[0025] The present invention also provides methods for determination of the quantity of an unknown bacterium in a sample. The sample is contacted with any of the compositions described herein and a known quantity of a calibration polynucleotide comprising a calibration sequence. Concurrently, nucleic acid from the bacterium in the sample is amplified with any of the compositions described herein and nucleic acid from the calibration polynucleotide in the sample is amplified with any of the compositions described herein to obtain a first amplification product comprising a bacterial bioagent identifying amplicon and a second amplification product comprising a calibration amplicon. The molecular mass and abundance for the bacterial bioagent identifying amplicon and the calibration amplicon is determined. The bacterial bioagent identifying amplicon is distinguished from the calibration amplicon based on molecular mass, wherein comparison of bacterial bioagent identifying amplicon abundance and calibration amplicon abundance indicates the quantity of bacterium in the sample. The method can also comprise determining the base composition of the bacterial bioagent identifying amplicon.

BRIEF DESCRIPTION OF THE DRAWINGS

[0026] Figure 1 is a represenataive pseudo-four dimensional plot of base compositions of bioagent identifying amplicons of enterobacteria obtained with a primer pair targeting the rpoB gene (primer pair no 14 (SEQ ID NOs: 37:362). The quantity each of the nucleobases A, G and C are represented on the three axes of the plot while the quantity of nucleobase T is represented by the diameter of the spheres. Base composition probability clouds surrounding the spheres are also shown.

[0027] Figure 2 is a represenataive diagram illustrating the primer selection process.

[0028] Figure 3 lists common pathogenic bacteria and primer pair coverage. The primer pair number in the upper right hand corner of each polygon indicates that the primer pair can produce a bioagent identifying amplicon for all species within that polygon.

[0029] Figure 4 is a represenataive 3D diagram of base composition (axes A, G and C) of bioagent identifying amplicons obtained with primer pair number 14 (a precursor of primer pair

number 348 which targets 16S rRNA). The diagram indicates that the experimentally determined base compositions of the clinical samples (labeled NHRC samples) closely match the base compositions expected for *Streptococcus pyogenes* and are distinct from the expected base compositions of other organisms.

[0030] Figure 5 is a represenataive mass spectrum of amplification products representing bioagent identifying amplicons of *Streptococcus pyogenes*, *Neisseria* meningitidis, and *Haemophilus influenzae* obtained from amplification of nucleic acid from a clinical sample with primer pair number 349 which targets 23S rRNA. Experimentally determined molecular masses and base compositions for the sense strand of each amplification product are shown.

[0031] Figure 6 is a represenataive mass spectrum of amplification products representing a bioagent identifying amplicon of *Streptococcus pyogenes*, and a calibration amplicon obtained from amplification of nucleic acid from a clinical sample with primer pair number 356 which targets rplB. The experimentally determined molecular mass and base composition for the sense strand of the *Streptococcus pyogenes* amplification product is shown.

[0032] Figure 7 is a represenataive process diagram for identification and determination of the quantity of a bioagent in a sample.

[0033] Figure 8 is a represenataive mass spectrum of an amplified nucleic acid mixture which contained the Ames strain of *Bacillus anthracis*, a known quantity of combination calibration polynucleotide (SEQ ID NO: 741), and primer pair number 350 which targets the capC gene on the virulence plasmid pX02 of *Bacillus anthracis*. Calibration amplicons produced in the amplification reaction are visible in the mass spectrum as indicated and abundance data (peak height) are used to calculate the quantity of the Ames strain of *Bacillus anthracis*.

DESCRIPTION OF EMBODIMENTS

[0034] The present invention provides oligonucleotide primers which hybridize to conserved regions of nucleic acid of genes encoding, for example, proteins or RNAs necessary for life which include, but are not limited to: 16S and 23S rRNAs, RNA polymerase subunits, t-RNA synthetases, elongation factors, ribosomal proteins, protein chain initiation factors, cell division proteins, chaperonin groEL, chaperonin dnaK, phosphoglycerate kinase, NADH dehydrogenase, DNA ligases, metabolic enzymes and DNA topoisomerases. These primers provide the

functionality of producing, for example, bacterial bioagent identifying amplicons for general identification of bacteria at the species level, for example, when contacted with bacterial nucleic acid under amplification conditions.

[0035] Referring to Figure 2, primers are designed as follows: for each group of organisms, candidate target sequences are identified (**200**) from which nucleotide alignments are created (**210**) and analyzed (**220**). Primers are designed by selecting appropriate priming regions (**230**) which allows the selection of candidate primer pairs (**240**). The primer pairs are subjected to *in silico* analysis by electronic PCR (ePCR) (**300**) wherein bioagent identifying amplicons are obtained from sequence databases such as, for example, GenBank or other sequence collections (**310**), and checked for specificity *in silico* (**320**). Bioagent identifying amplicons obtained from GenBank sequences (**310**) can also be analyzed by a probability model which predicts the capability of a particular amplicon to identify unknown bioagents such that the base compositions of amplicons with favorable probability scores are stored in a base composition database (**325**). Alternatively, base compositions of the bioagent identifying amplicons obtained from the primers and GenBank sequences can be directly entered into the base composition database (**330**). Candidate primer pairs (**340**) are validated by *in* vitro amplification by a method such as, for example, PCR analysis (**400**) of nucleic acid from a collection of organisms (**410**). Amplification products that are obtained are optionally analyzed to confirm the sensitivity, specificity and reproducibility of the primers used to obtain the amplification products (**420**).

[0036] Synthesis of primers is well known and routine in the art. The primers may be conveniently and routinely made through the well-known technique of solid phase synthesis. Equipment for such synthesis is sold by several vendors including, for example, Applied Biosystems (Foster City, CA). Any other means for such synthesis known in the art may additionally or alternatively be employed.

[0037] The primers can be employed as compositions for use in, for example, methods for identification of bacterial bioagents as follows. In some embodiments, a primer pair composition is contacted with nucleic acid of an unknown bacterial bioagent. The nucleic acid is amplified by a nucleic acid amplification technique, such as PCR for example, to obtain an amplification product that represents a bioagent identifying amplicon. The molecular mass of one strand or each strand of the double-stranded amplification product is determined by a molecular mass measurement technique such as, for example, mass spectrometry wherein the two strands of the

double-stranded amplification product are separated during the ionization process. In some embodiments, the mass spectrometry is electrospray Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS) or electrospray time of flight mass spectrometry (ESI-TOF-MS). A list of possible base compositions can be generated for the molecular mass value obtained for each strand and the choice of the correct base composition from the list is facilitated by matching the base composition of one strand with a complementary base composition of the other strand. The molecular mass or base composition thus determined is compared with a database of molecular masses or base compositions of analogous bioagent identifying amplicons for known bacterial bioagents. A match between the molecular mass or base composition of the amplification product from the unknown bacterial bioagent and the molecular mass or base composition of an analogous bioagent identifying amplicon for a known bacterial bioagent indicates the identity of the unknown bioagent.

[0038] In some embodiments, the primer pair used is one of the primer pairs of Table 1. In some embodiments, the method is repeated using a different primer pair to resolve possible ambiguities in the identification process or to improve the confidence level for the identification assignment.

[0039] In some embodiments, a bioagent identifying amplicon may be produced using only a single primer (either the forward or reverse primer of any given primer pair), provided an appropriate amplification method is chosen, such as, for example, low stringency single primer PCR (LSSP-PCR). Adaptation of this amplification method in order to produce bioagent identifying amplicons can be accomplished by one with ordinary skill in the art without undue experimentation.

[0040] In some embodiments, the oligonucleotide primers are "broad range survey primers" which hybridize to conserved regions of nucleic acid encoding RNA, such as ribosomal RNA (rRNA), of all, or at least 70%, at least 80%, at least 85%, at least 90%, or at least 95% of known bacteria and produce bacterial bioagent identifying amplicons. As used herein, the term "broad range survey primers" refers to primers that bind to nucleic acid encoding rRNAs of all, or at least 70%, at least 80%, at least 85%, at least 90%, or at least 95% known species of bacteria. In some embodiments, the rRNAs to which the primers hybridize are 16S and 23S rRNAs. In some embodiments, the broad range survey primer pairs comprise oligonucleotides ranging in length from 13 to 35 nucleobases, each of which have from 70% to 100% sequence identity with primer

pair numbers 3, 10, 11, 14, 16, and 17 which consecutively correspond to SEQ ID NOs: 6:369, 26:388, 29:391, 37:362, 48:404, and 58:414.

[0041] In some cases, the molecular mass or base composition of a bacterial bioagent identifying amplicon defined by a broad range survey primer pair does not provide enough resolution to unambiguously identify a bacterial bioagent at the species level. These cases benefit from further analysis of one or more bacterial bioagent identifying amplicons generated from at least one additional broad range survey primer pair or from at least one additional "division-wide" primer pair (*vide infra*). The employment of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as "triangulation identification" (*vide infra*).

[0042] In other embodiments, the oligonucleotide primers are "division-wide" primers which hybridize to nucleic acid encoding genes of broad divisions of bacteria such as, for example, members of the *Bacillus/Clostridia* group or members of the α-, β-, γ-, and ε-proteobacteria. In some embodiments, a division of bacteria comprises any grouping of bacterial genera with more than one genus represented. For example, the β-proteobacteria group comprises members of the following genera: *Eikenella, Neisseria, Achromobacter, Bordetella, Burkholderia,* and *Raltsonia.* Species members of these genera can be identified using bacterial bioagent identifying amplicons generated with primer pair 293 (SEQ ID NOs: 344:700) which produces a bacterial bioagent identifying amplicon from the tufB gene of β-proteobacteria. Examples of genes to which division-wide primers may hybridize to include, but are not limited to: RNA polymerase subunits such as rpoB and rpoC, tRNA synthetases such as valyl-tRNA synthetase (valS) and aspartyl-tRNA synthetase (aspS), elongation factors such as elongation factor EF-Tu (tufB), ribosomal proteins such as ribosomal protein L2 (rplB), protein chain initiation factors such as protein chain initiation factor infB, chaperonins such as groL and dnaK, and cell division proteins such as peptidase ftsH (hflB). In some embodiments, the division-wide primer pairs comprise oligonucleotides ranging in length from 13 to 35 nucleobases, each of which have from 70% to 100% sequence identity with primer pair numbers 34, 52, 66, 67, 71, 72, 289, 290 and 293 which consecutively correspond to SEQ ID NOs: 160:515, 261:624, 231:591, 235:587, 349:711, 240:596, 246:602, 256:620, 344:700.

[0043] In other embodiments, the oligonucleotide primers are designed to enable the identification of bacteria at the clade group level, which is a monophyletic taxon referring to a group of organisms which includes the most recent common ancestor of all of its members and

all of the descendants of that most recent common ancestor. The *Bacillus cereus* clade is an example of a bacterial clade group. In some embodiments, the clade group primer pairs comprise oligonucleotides ranging in length from 13 to 35 nucleobases, each of which have from 70% to 100% sequence identity with primer pair number 58 which corresponds to SEQ ID NOs: 322:686.

[0044] In other embodiments, the oligonucleotide primers are "drill-down" primers which enable the identification of species or "sub-species characteristics." Sub-species characteristics are herein defined as genetic characteristics that provide the means to distinguish two members of the same bacterial species. For example, *Escherichia coli* O157:H7 and *Escherichia coli* K12 are two well known members of the species *Escherichia coli*. *Escherichia coli* O157:H7, however, is highly toxic due to the its Shiga toxin gene which is an example of a sub-species characteristic. Examples of sub-species characteristics may also include, but are not limited to: variations in genes such as single nucleotide polymorphisms (SNPs), variable number tandem repeats (VNTRs). Examples of genes indicating sub-species characteristics include, but are not limited to, housekeeping genes, toxin genes, pathogenicity markers, antibiotic resistance genes and virulence factors. Drill-down primers provide the functionality of producing bacterial bioagent identifying amplicons for drill-down analyses such as strain typing when contacted with bacterial nucleic acid under amplification conditions. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of bacterial infections. Examples of pairs of drill-down primers include, but are not limited to, a trio of primer pairs for identification of strains of *Bacillus anthracis*. Primer pair 24 (SEQ ID NOs: 97:451) targets the capC gene of virulence plasmid pX02, primer pair 30 (SEQ ID NOs: 127:482) targets the cyA gene of virulence plasmid pX02, and primer pair 37 (SEQ ID NOs: 174:530) targets the lef gene of virulence plasmid pX02. Additional examples of drill-down primers include, but are not limited to, six primer pairs that are used for determining the strain type of group A *Streptococcus*. Primer pair 80 (SEQ ID NOs: 310:668) targets the gki gene, primer pair 81 (SEQ ID NOs: 313:670) targets the gtr gene, primer pair 86 (SEQ ID NOs: 227:632) targets the murI gene, primer pair 90 (SEQ ID NOs: 285:640) targets the mutS gene, primer pair 96 (SEQ ID NOs: 301:656) targets the xpt gene, and primer pair 98 (SEQ ID NOs: 308:663) targets the yqiL gene.

[0045] In some embodiments, the primers used for amplification hybridize to and amplify genomic DNA, DNA of bacterial plasmids, or DNA of DNA viruses.

[0046] In some embodiments, the primers used for amplification hybridize directly to ribosomal RNA or messenger RNA (mRNA) and act as reverse transcription primers for obtaining DNA from direct amplification of bacterial RNA or rRNA. Methods of amplifying RNA using reverse transcriptase are well known to those with ordinary skill in the art and can be routinely established without undue experimentation.

[0047] One with ordinary skill in the art of design of amplification primers will recognize that a given primer need not hybridize with 100% complementarity in order to effectively prime the synthesis of a complementary nucleic acid strand in an amplification reaction. Moreover, a primer may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (e.g., a loop structure or a hairpin structure). The primers of the present invention may comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% sequence identity with any of the primers listed in Table 1. Thus, in some embodiments of the present invention, an extent of variation of 70% to 100%, or any range therewithin, of the sequence identity is possible relative to the specific primer sequences disclosed herein. Determination of sequence identity is described in the following example: a primer 20 nucleobases in length which is otherwise identical to another 20 nucleobase primer but having two non-identical residues has 18 of 20 identical residues (18/20 = 0.9 or 90% sequence identity). In another example, a primer 15 nucleobases in length having all residues identical to a 15 nucleobase segment of primer 20 nucleobases in length would have 15/20 = 0.75 or 75% sequence identity with the 20 nucleobase primer.

[0048] Percent homology, sequence identity or complementarity, can be determined by, for example, the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison WI), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489). In some embodiments, homology, sequence identity, or complementarity of primers with respect to the conserved priming regions of bacterial nucleic acid, is at least 70%, at least 80%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or is 100%.

[0049] In some embodiments, the primers described herein comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or at least 99%, or 100% (or any range therewithin) sequence identity with the primer sequences specifically disclosed herein. Thus, for example, a primer may have between 70% and

100%, between 75% and 100%, between 80% and 100%, and between 95% and 100% sequence identity with SEQ ID NO: 26. Likewise, a primer may have similar sequence identity with any other primer whose nucleotide sequence is disclosed herein.

[0050] One with ordinary skill is able to calculate percent sequence identity or percent sequence homology and able to determine, without undue experimentation, the effects of variation of primer sequence identity on the function of the primer in its role in priming synthesis of a complementary strand of nucleic acid for production of an amplification product of a corresponding bioagent identifying amplicon.

[0051] In some embodiments of the present invention, the oligonucleotide primers are between 13 and 35 nucleobases in length (13 to 35 linked nucleotide residues). These embodiments comprise oligonucleotide primers 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 or 35 nucleobases in length, or any range therewithin.

[0052] In some embodiments, any given primer comprises a modification comprising the addition of a non-templated T residue to the 5' end of the primer (i.e., the added T residue does not necessarily hybridize to the nucleic acid being amplified). The addition of a non-templated T residue has an effect of minimizing the addition of non-templated A residues as a result of the non-specific enzyme activity of *Taq* polymerase (Magnuson et al. Biotechniques, 1996, 21, 700-709), an occurrence which may lead to ambiguous results arising from molecular mass analysis.

[0053] In some embodiments of the present invention, primers may contain one or more universal bases. Because any variation (due to codon wobble in the 3rd position) in the conserved regions among species is likely to occur in the third position of a DNA triplet, oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind to more than one nucleotide, referred to herein as a "universal nucleobase." For example, under this "wobble" pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal nucleobases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes et al., Nucleosides and Nucleotides, 1995, 14, 1001-1003), the degenerate nucleotides d**P** or d**K** (Hill et al.), an acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot et al., Nucleosides and Nucleotides, 1995, 14, 1053-1056) or the purine analog 1-(2-deoxy-β-D-ribofuranosyl)-imidazole-4-carboxamide (Sala *et* al., Nucl. Acids Res., 1996, 24, 3302-3306).

[0054] In some embodiments, to compensate for the somewhat weaker binding by the "wobble" base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, 5-propynyluracil which binds to adenine and 5-propynylcytosine and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is commonly owned. Propynylated primers are described in U.S. issued Patent 6,875,593 which is also commonly owned. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183.

[0055] In some embodiments, non-template primer tags are used to increase the melting temperature (T_m) of a primer-template duplex in order to improve amplification efficiency. A non-template tag is at least three consecutive A or T nucleotide residues on a primer which are not complementary to the template. In any given non-template tag, A can be replaced by C or G and T can also be replaced by C or G. Although Watson-Crick hybridization is not expected to occur for a non-template tag relative to the template, the extra hydrogen bond in a G-C pair relative to a A-T pair confers increased stability of the primer-template duplex and improves amplification efficiency for subsequent cycles of amplification when the primers hybridize to strands synthesized in previous cycles.

[0056] In other embodiments, propynylated tags may be used in a manner similar to that of the non-template tag, wherein two or more 5-propynylcytidine or 5-propynyluridine residues replace template matching residues on a primer. In other embodiments, a primer contains a modified internucleoside linkage such as a phosphorothioate linkage, for example.

[0057] In some embodiments, the primers contain mass-modifying tags. Reducing the total number of possible base compositions of a nucleic acid of specific molecular weight provides a means of avoiding a persistent source of ambiguity in determination of base composition of amplification products. Addition of mass-modifying tags to certain nucleobases of a given primer will result in simplification of *de novo* determination of base composition of a given bioagent identifying amplicon (*vide infra*) from its molecular mass.

[0058] In some embodiments of the present invention, the mass modified nucleobase comprises one or more of the following: for example, 7-deaza-2'-deoxyadenosine-5-triphosphate, 5-iodo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxycytidine-5'-triphosphate, 5-iodo-2'-deoxycytidine-5'-triphosphate, 5-hydroxy-2'-deoxyuridine-5'-triphosphate, 4-thiothymidine-5'-triphosphate, 5-aza-2'-deoxyuridine-5'-triphosphate, 5-fluoro-2'-deoxyuridine-5'-triphosphate, O6-methyl-2'-deoxyguanosine-5'-triphosphate, N2-methyl-2'-deoxyguanosine-5'-triphosphate, 8-oxo-2'-deoxyguanosine-5'-triphosphate or thiothymidine-5'-triphosphate. In some embodiments, the mass-modified nucleobase comprises ^{15}N or ^{13}C or both ^{15}N and ^{13}C.

[0059] In some embodiments of the present invention, at least one bacterial nucleic acid segment is amplified in the process of identifying the bioagent. Thus, the nucleic acid segments that can be amplified by the primers disclosed herein and that provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as "bioagent identifying amplicons." The term "amplicon" as used herein, refers to a segment of a polynucleotide which is amplified in an amplification reaction. In some embodiments of the present invention, bioagent identifying amplicons comprise from about 45 to about 200 nucleobases (i.e. from about 45 to about 200 linked nucleosides), from about 60 to about 150 nucleobases, from about 75 to about 125 nucleobases. One of ordinary skill in the art will appreciate that the invention embodies compounds of 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, and 200 nucleobases in length, or any range therewithin. It is the combination of the portions of the bioagent nucleic acid segment to which the primers hybridize (hybridization sites) and the variable region between the primer hybridization sites that comprises the bioagent identifying amplicon. Since genetic data provide the underlying basis for identification of bioagents by the methods of the present invention, it is prudent to select segments of nucleic acids which ideally provide enough variability to distinguish each individual bioagent and whose molecular mass is amenable to molecular mass determination.

[0060] In some embodiments, bioagent identifying amplicons amenable to molecular mass determination which are produced by the primers described herein are either of a length, size or mass compatible with the particular mode of molecular mass determination or compatible with a means of providing a predictable fragmentation pattern in order to obtain predictable fragments of a length compatible with the particular mode of molecular mass determination. Such means of providing a predictable fragmentation pattern of an amplification product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, for example. Methods of using restriction enzymes and cleavage primers are well known to those with ordinary skill in the art.

[0061] In some embodiments, amplification products corresponding to bacterial bioagent identifying amplicons are obtained using the polymerase chain reaction (PCR) which is a routine method to those with ordinary skill in the molecular biology arts. Other amplification methods may be used such as ligase chain reaction (LCR), low-stringency single primer PCR, and multiple strand displacement amplification (MDA) which are also well known to those with ordinary skill.

[0062] In the context of this invention, a "bioagent" is any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus. Examples of bioagents include, but are not limited, to cells, (including but not limited to human clinical samples, bacterial cells and other pathogens), viruses, fungi, protists, parasites, and pathogenicity markers (including but not limited to: pathogenicity islands, antibiotic resistance genes, virulence factors, toxin genes and other bioregulating compounds). Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores) and may be encapsulated or bioengineered. In the context of this invention, a "pathogen" is a bioagent which causes a disease or disorder.

[0063] In the context of this invention, the term "unknown bioagent" may mean either: (i) a bioagent whose existence is known (such as the well known bacterial species *Staphylococcus aureus* for example) but which is not known to be in a sample to be analyzed, or (ii) a bioagent whose existence is not known (for example, the SARS coronavirus was unknown prior to April 2003). For example, if the method for identification of coronaviruses disclosed in commonly owned U.S. published Patent 2005-0266397 was to be employed prior to April 2003 to identify the SARS coronavirus in a clinical sample, both meanings of "unknown" bioagent are applicable since the SARS coronavirus was unknown to

science prior to April, 2003 and since it was not known what bioagent (in this case a coronavirus) was present in the sample. On the other hand, if the method of U.S. published Patent 2005-0266397 was to be employed subsequent to April 2003 to identify the SARS coronavirus in a clinical sample, only the first meaning (i) of "unknown" bioagent would apply since the SARS coronavirus became known to science subsequent to April 2003 and since it was not known what bioagent was present in the sample.

[0064] The employment of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as "triangulation identification." Triangulation identification is pursued by analyzing a plurality of bioagent identifying amplicons selected within multiple core genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen et al., J. Appl. Microbiol., 1999, 87, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

[0065] In some embodiments, the triangulation identification process can be pursued by characterization of bioagent identifying amplicons in a massively parallel fashion using the polymerase chain reaction (PCR), such as multiplex PCR where multiple primers are employed in the same amplification reaction mixture, or PCR in multi-well plate format wherein a different and unique pair of primers is used in multiple wells containing otherwise identical reaction mixtures. Such multiplex and multi-well PCR methods are well known to those with ordinary skill in the arts of rapid throughput amplification of nucleic acids.

[0066] In some embodiments, the molecular mass of a particular bioagent identifying amplicon is determined by mass spectrometry. Mass spectrometry has several advantages, not the least of which is high bandwidth characterized by the ability to separate (and isolate) many molecular peaks across a broad range of mass to charge ratio (m/z). Thus, mass spectrometry is intrinsically a parallel detection scheme without the need for radioactive or fluorescent labels, since every amplification product is identified by its molecular mass. The current state of the art in mass spectrometry is such that less than femtomole quantities of material can be readily analyzed to afford information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular

weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons.

[0067] In some embodiments, intact molecular ions are generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but are not limited to, electrospray ionization (ES), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a single mass spectrum affords an estimate of molecular mass of the bioagent identifying amplicon. Electrospray ionization mass spectrometry (ESI-MS) is particularly useful for very high molecular weight polymers such as proteins and nucleic acids having molecular weights greater than 10 kDa, since it yields a distribution of multiply-charged molecules of the sample without causing a significant amount of fragmentation.

[0068] The mass detectors used in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), time of flight (TOF), ion trap, quadrupole, magnetic sector, Q-TOF, and triple quadrupole.

[0069] In some embodiments, conversion of molecular mass data to a base composition is useful for certain analyses. As used herein, a "base composition" is the exact number of each nucleobase (A, T, C and G). For example, amplification of nucleic acid of *Neisseria meningitidis* with a primer pair that produces an amplification product from nucleic acid of 23S rRNA that has a molecular mass (sense strand) of 28480.75124, from which a base composition of A25 G27 C22 T18 is assigned from a list of possible base compositions calculated from the molecular mass using standard known molecular masses of each of the four nucleobases.

[0070] In some embodiments, assignment of base compositions to experimentally determined molecular masses is accomplished using "base composition probability clouds." Base compositions, like sequences, vary slightly from isolate to isolate within species. It is possible to manage this diversity by building "base composition probability clouds" around the composition constraints for each species. This permits identification of organisms in a fashion similar to sequence analysis. A "pseudo four-dimensional plot" (Figure 1) can be used to visualize the concept of base composition probability clouds. Optimal primer design requires optimal choice

of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual bioagents. Areas where clouds overlap indicate regions that may result in a misclassification, a problem which is overcome by a triangulation identification process using bioagent identifying amplicons not affected by overlap of base composition probability clouds.

[0071] In some embodiments, base composition probability clouds provide the means for screening potential primer pairs in order to avoid potential misclassifications of base compositions. In other embodiments, base composition probability clouds provide the means for predicting the identity of a bioagent whose assigned base composition was not previously observed and/or indexed in a bioagent identifying amplicon base composition database due to evolutionary transitions in its nucleic acid sequence. Thus, in contrast to probe-based techniques, mass spectrometry determination of base composition does not require prior knowledge of the composition or sequence in order to make the measurement.

[0072] The present invention provides bioagent classifying information similar to DNA sequencing and phylogenetic analysis at a level sufficient to identify a given bioagent. Furthermore, the process of determination of a previously unknown base composition for a given bioagent (for example, in a case where sequence information is unavailable) has downstream utility by providing additional bioagent indexing information with which to populate base composition databases. The process of future bioagent identification is thus greatly improved as more BCS indexes become available in base composition databases.

[0073] In one embodiment, a sample comprising an unknown bioagent is contacted with a pair of primers which provide the means for amplification of nucleic acid from the bioagent, and a known quantity of a polynucleotide that comprises a calibration sequence. The nucleic acids of the bioagent and of the calibration sequence are amplified and the rate of amplification is reasonably assumed to be similar for the nucleic acid of the bioagent and of the calibration sequence. The amplification reaction then produces two amplification products: a bioagent identifying amplicon and a calibration amplicon. The bioagent identifying amplicon and the calibration amplicon should be distinguishable by molecular mass while being amplified at essentially the same rate. Effecting differential molecular masses can be accomplished by choosing as a calibration sequence, a representative bioagent identifying amplicon (from a specific species of bioagent) and performing, for example, a 2 to 8 nucleobase deletion or

insertion within the variable region between the two priming sites. The amplified sample containing the bioagent identifying amplicon and the calibration amplicon is then subjected to molecular mass analysis by mass spectrometry, for example. The resulting molecular mass analysis of the nucleic acid of the bioagent and of the calibration sequence provides molecular mass data and abundance data for the nucleic acid of the bioagent and of the calibration sequence. The molecular mass data obtained for the nucleic acid of the bioagent enables identification of the unknown bioagent and the abundance data enables calculation of the quantity of the bioagent, based on the knowledge of the quantity of calibration polynucleotide contacted with the sample.

[0074] In some embodiments, the identity and quantity of a particular bioagent is determined using the process illustrated in Figure 7. For instance, to a sample containing nucleic acid of an unknown bioagent are added primers (**500**) and a known quantity of a calibration polynucleotide (**505**). The total nucleic acid in the sample is subjected to an amplification reaction (**510**) to obtain amplification products. The molecular masses of amplification products are determined (**515**) from which are obtained molecular mass and abundance data. The molecular mass of the bioagent identifying amplicon (**520**) provides the means for its identification (**525**) and the molecular mass of the calibration amplicon obtained from the calibration polynucleotide (**530**) provides the means for its identification (**535**). The abundance data of the bioagent identifying amplicon is recorded (**540**) and the abundance data for the calibration data is recorded (**545**), both of which are used in a calculation (**550**) which determines the quantity of unknown bioagent in the sample.

[0075] In some embodiments, construction of a standard curve where the amount of calibration polynucleotide spiked into the sample is varied, provides additional resolution and improved confidence for the determination of the quantity of bioagent in the sample. The use of standard curves for analytical determination of molecular quantities is well known to one with ordinary skill and can be performed without undue experimentation.

[0076] In some embodiments, multiplex amplification is performed where multiple bioagent identifying amplicons are amplified with multiple primer pairs which also amplify the corresponding standard calibration sequences. In this or other embodiments, the standard calibration sequences are optionally included within a single vector which functions as the

calibration polynucleotide. Multiplex amplification methods are well known to those with ordinary skill and can be performed without undue experimentation.

[0077] In some embodiments, the calibrant polynucleotide is used as an internal positive control to confirm that amplification conditions and subsequent analysis steps are successful in producing a measurable amplicon. Even in the absence of copies of the genome of a bioagent, the calibration polynucleotide should give rise to a calibration amplicon. Failure to produce a measurable calibration amplicon indicates a failure of amplification or subsequent analysis step such as amplicon purification or molecular mass determination. Reaching a conclusion that such failures have occurred is in itself, a useful event.

[0078] In some embodiments, the calibration sequence is inserted into a vector which then itself functions as the calibration polynucleotide. In some embodiments, more than one calibration sequence is inserted into the vector that functions as the calibration polynucleotide. Such a calibration polynucleotide is herein termed a "combination calibration polynucleotide." The process of inserting polynucleotides into vectors is routine to those skilled in the art and can be accomplished without undue experimentation. Thus, it should be recognized that the calibration method should not be limited to the embodiments described herein. The calibration method can be applied for determination of the quantity of any bioagent identifying amplicon when an appropriate standard calibrant polynucleotide sequence is designed and used. The process of choosing an appropriate vector for insertion of a calibrant is also a routine operation that can be accomplished by one with ordinary skill without undue experimentation.

[0079] The present invention also provides kits for carrying out, for example, the methods described herein. In some embodiments, the kit may comprise a sufficient quantity of one or more primer pairs to perform an amplification reaction on a target polynucleotide from a bioagent to form a bioagent identifying amplicon. In some embodiments, the kit may comprise from one to fifty primer pairs, from one to twenty primer pairs, from one to ten primer pairs, or from two to five primer pairs. In some embodiments, the kit may comprise one or more primer pairs recited in Table 1.

[0080] In some embodiments, the kit may comprise one or more broad range survey primer(s), division wide primer(s), clade group primer(s) or drill-down primer(s), or any combination thereof. A kit may be designed so as to comprise particular primer pairs for identification of a

particular bioagent. For example, a broad range survey primer kit may be used initially to identify an unknown bioagent as a member of the *Bacillus/Clostridia* group. Another example of a division-wide kit may be used to distinguish *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis* from each other. A clade group primer kit may be used, for example, to identify an unknown bacterium as a member of the *Bacillus cereus* clade group. A drill-down kit may be used, for example, to identify genetically engineered *Bacillus anthracis*. In some embodiments, any of these kits may be combined to comprise a combination of broad range survey primers and division-wide primers, clade group primers or drill-down primers, or any combination thereof, for identification of an unknown bacterial bioagent.

[0081] In some embodiments, the kit may contain standardized calibration polynucleotides for use as internal amplification calibrants.

[0082] In some embodiments, the kit may also comprise a sufficient quantity of reverse transcriptase (if an RNA virus is to be identified for example), a DNA polymerase, suitable nucleoside triphosphates (including any of those described above), a DNA ligase, and/or reaction buffer, or any combination thereof, for the amplification processes described above. A kit may further include instructions pertinent for the particular embodiment of the kit, such instructions describing the primer pairs and amplification conditions for operation of the method. A kit may also comprise amplification reaction containers such as microcentrifuge tubes and the like. A kit may also comprise reagents or other materials for isolating bioagent nucleic acid or bioagent identifying amplicons from amplification, including, for example, detergents, solvents, or ion exchange resins which may be linked to magnetic beads. A kit may also comprise a table of measured or calculated molecular masses and/or base compositions of bioagents using the primer pairs of the kit.

[0083] In order that the invention disclosed herein may be more efficiently understood, examples are provided below. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting the invention in any manner. Throughout these examples, molecular cloning reactions, and other standard recombinant DNA techniques, were carried out according to methods described in Maniatis et al., Molecular Cloning - A Laboratory Manual,

2nd ed., Cold Spring Harbor Press (1989), using commercially available reagents, except where otherwise noted.

EXAMPLES

[0084] Example 1: Selection of Primers That Define Bioagent Identifying Amplicons

[0085] For design of primers that define bacterial bioagent identifying amplicons, relevant sequences from, for example, GenBank are obtained, aligned and scanned for regions where pairs of PCR primers would amplify products of about 45 to about 200 nucleotides in length and distinguish species from each other by their molecular masses or base compositions. A typical process shown in Figure 2 is employed.

[0086] A database of expected base compositions for each primer region is generated using an *in silico* PCR search algorithm, such as (ePCR). An existing RNA structure search algorithm (Macke et al., Nuc. Acids Res., 2001, 29, 4724-4735)

has been modified to include PCR parameters such as hybridization conditions, mismatches, and thermodynamic calculations (SantaLucia, Proc. Natl. Acad. Sci. U.S.A., 1998, 95, 1460-1465). This also provides information on primer specificity of the selected primer pairs.

[0087] Table 1 represents a collection of primers (sorted by forward primer name) designed to identify bacteria using the methods herein described. The forward or reverse primer name indicates the gene region of bacterial genome to which the primer hybridizes relative to a reference sequence eg: the forward primer name 16S_EC_1077_1106 indicates that the primer hybridizes to residues 1077-1106 of the gene encoding 16S ribosomal RNA in an *E. coli* reference sequence represented by a sequence extraction of coordinates 4033120..4034661 from GenBank gi number 16127994 (as indicated in Table 2). As an additional example: the forward primer name BONTA_X52066_450_473 indicates that the primer hybridizes to residues 450-437 of the gene encoding *Clostridium botulinum* neurotoxin type A (BoNT/A) represented by GenBank Accession No. X52066 (primer pair name codes appearing in Table 1 are defined in Table 2). In Table 1, U^a = 5-propynyluracil; C^a = 5-propynylcytosine; * = phosphorothioate linkage. The primer pair number is an in-house database index number.

Table 1: Primer Pairs for Identification of Bacterial Bioagents

Primer pair number	For. primer name	Forward sequence	For. SEQ ID NO:	Rev. primer name	Reverse sequence	Rev. SEQ ID NO:
1	16S_EC_107	GTGAGATGTTGGGTTAA	1	16S_EC_1175	GACGTCATCCCCACCTTCC	368

	7_1106_F	GTCCCGTAACGAG		_1195_R	TC	
266	16S_EC_1082_1100_F	ATGTTGGGTTAAGTCCCGC	2	16S_EC_1177_1196_10G_11G_R	TGACGTCATGGCCACCTTCC	372
265	16S_EC_1082_1100_F	ATGTTGGGTTAAGTCCCGC	2	16S_EC_1177_1196_10G_R	TGACGTCATGCCCACCTTCC	373
230	16S_EC_1082_1100_F	ATGTTGGGTTAAGTCCCGC	2	16S_EC_1177_1196_R	TGACGTCATCCCCACCTTCC	374
263	16S_EC_1082_1100_F	ATGTTGGGTTAAGTCCCGC	2	16S_EC_1525_1541_R	AAGGAGGTGATCCAGCC	382
2	16S_EC_1082_1106_F	ATGTTGGGTTAAGTCCCGCAACGAG	3	16S_EC_1175_1197_R	TTGACGTCATCCCCACCTTCCTC	371
278	16S_EC_1090_1111_2_F	TTAAGTCCCGCAACGAGCGCAA	4	16S_EC_1175_1196_R	TGACGTCATCCCCACCTTCCTC	369
361	16S_EC_1090_1111_2_TMOD_F	TTTAAGTCCCGCAACGAGCGCAA	5	16S_EC_1175_1196_TMOD_R	TTGACGTCATCCCCACCTTCCTC	370
3	16S_EC_1090_1111_F	TTAAGTCCCGCAACGATCGCAA	6	16S_EC_1175_1196_R	TGACGTCATCCCCACCTTCCTC	369
256	16S_EC_1092_1109_F	TAGTCCCGCAACGAGCGC	7	16S_EC_1174_1195_R	GACGTCATCCCCACCTTCCTCC	367
159	16S_EC_1100_1116_F	CAACGAGCGCAACCCTT	8	16S_EC_1174_1188_R	TCCCCACCTTCCTCC	366
247	16S_EC_1195_1213_F	CAAGTCATCATGGCCCTTA	9	16S_EC_1525_1541_R	AAGGAGGTGATCCAGCC	382
4	16S_EC_1222_1241_F	GCTACACACGTGCTACAATG	10	16S_EC_1303_1323_R	CGAGTTGCAGACTGCGATCCG	376
232	16S_EC_1303_1323_F	CGGATTGGAGTCTGCAACTCG	11	16S_EC_1389_1407_R	GACGGGCGGTGTGTACAAG	378
5	16S_EC_1332_1353_F	AAGTCGGAATCGCTAGTAATCG	12	16S_EC_1389_1407_R	GACGGGCGGTGTGTACAAG	378
252	16S_EC_1367_1387_F	TACGGTGAATACGTTCCCGGG	13	16S_EC_1485_1506_R	ACCTTGTTACGACTTCACCCCA	379
250	16S_EC_1387_1407_F	GCCTTGTACACACCTCCCGTC	14	16S_EC_1494_1513_R	CACGGCTACCTTGTTACGAC	381
231	16S_EC_1389_1407_F	CTTGTACACACCGCCCGTC	15	16S_EC_1525_1541_R	AAGGAGGTGATCCAGCC	382
251	16S_EC_1390_1411_F	TTGTACACACCGCCCGTCATAC	16	16S_EC_1486_1505_R	CCTTGTTACGACTTCACCCC	380
6	16S_EC_30_54_F	TGAACGCTGGTGGCATGCTTAACAC	17	16S_EC_105_126_R	TACGCATTACTCACCCGTCCGC	361
243	16S_EC_314_332_F	CACTGGAACTGAGACACGG	18	16S_EC_556_575_R	CTTTACGCCCAGTAATTCCG	385
7	16S_EC_38_64_F	GTGGCATGCCTAATACATGCAAGTCG	19	16S_EC_101_120_R	TTACTCACCCGTCCGCCGCT	357
279	16S_EC_405_432_F	TGAGTGATGAAGGCCTTAGGGTTGTAAA	20	16S_EC_507_527_R	CGGCTGCTGGCACGAAGTTAG	384
8	16S_EC_49_68_F	TAACACATGCAAGTCGAACG	21	16S_EC_104_120_R	TTACTCACCCGTCCGCC	359
275	16S_EC_49_68_F	TAACACATGCAAGTCGAACG	21	16S_EC_1061_1078_R	ACGACACGAGCTGACGAC	364
274	16S_EC_49_68_F	TAACACATGCAAGTCGAACG	21	16S_EC_880_894_R	CGTACTCCCCAGGCG	390
244	16S_EC_518_536_F	CCAGCAGCCGCGGTAATAC	22	16S_EC_774_795_R	GTATCTAATCCTGTTTGCTCCC	387
226	16S_EC_556_575_F	CGGAATTACTGGGCGTAAAG	23	16S_EC_683_700_R	CGCATTTCACCGCTACAC	386
264	16S_EC_556_575_F	CGGAATTACTGGGCGTAAAG	23	16S_EC_774_795_R	GTATCTAATCCTGTTTGCTCCC	387
273	16S_EC_683_700_F	GTGTAGCGGTGAAATGCG	24	16S_EC_1303_1323_R	CGAGTTGCAGACTGCGATCCG	377
9	16S_EC_683_700_F	GTGTAGCGGTGAAATGCG	24	16S_EC_774_795_R	GTATCTAATCCTGTTTGCTCCC	387
158	16S_EC_683_700_F	GTGTAGCGGTGAAATGCG	24	16S_EC_880_894_R	CGTACTCCCCAGGCG	390
245	16S_EC_683_700_F	GTGTAGCGGTGAAATGCG	24	16S_EC_967_985_R	GGTAAGGTTCTTCGCGTTG	396
294	16S_EC_7_33_F	GAGAGTTTGATCCTGGCTCAGAACGAA	25	16S_EC_101_122_R	TGTTACTCACCCGTCTGCCACT	358
10	16S_EC_713_732_F	AGAACACCGATGGCGAAGGC	26	16S_EC_789_809_R	CGTGGACTACCAGGGTATCTA	388
346	16S_EC_713_732_TMOD_F	TAGAACACCGATGGCGAAGGC	27	16S_EC_789_809_TMOD_R	TCGTGGACTACCAGGGTATCTA	389
228	16S_EC_774	GGGAGCAAACAGGATTA	28	16S_EC_880_	CGTACTCCCCAGGCG	390

	_795_F	GATAC		894_R		
11	16S_EC_785_806_F	GGATTAGAGACCCTGGTAGTCC	29	16S_EC_880_897_R	GGCCGTACTCCCCAGGCG	391
347	16S_EC_785_806_TMOD_F	TGGATTAGAGACCCTGGTAGTCC	30	16S_EC_880_897_TMOD_R	TGGCCGTACTCCCCAGGCG	392
12	16S_EC_785_810_F	GGATTAGATACCCTGGTAGTCCACGC	31	16S_EC_880_897_2_R	GGCCGTACTCCCCAGGCG	391
13	16S_EC_789_810_F	TAGATACCCTGGTAGTCCACGC	32	16S_EC_880_894_R	CGTACTCCCCAGGCG	390
255	16S_EC_789_810_F	TAGATACCCTGGTAGTCCACGC	32	16S_EC_882_899_R	GCGACCGTACTCCCCAGG	393
254	16S_EC_791_812_F	GATACCCTGGTAGTCCACACCG	33	16S_EC_886_904_R	GCCTTGCGACCGTACTCCC	394
248	16S_EC_8_27_F	AGAGTTTGATCATGGCTCAG	34	16S_EC_1525_1541_R	AAGGAGGTGATCCAGCC	382
242	16S_EC_8_27_F	AGAGTTTGATCATGGCTCAG	34	16S_EC_342_358_R	ACTGCTGCCTCCCGTAG	383
253	16S_EC_804_822_F	ACCACGCCGTAAACGATGA	35	16S_EC_909_929_R	CCCCCGTCAATTCCTTTGAGT	395
246	16S_EC_937_954_F	AAGCGGTGGAGCATGTGG	36	16S_EC_1220_1240_R	ATTGTAGCACGTGTGTAGCCC	375
14	16S_EC_960_981_F	TTCGATGCAACGCGAAGAACCT	37	16S_EC_1054_1073_R	ACGAGCTGACGACAGCCATG	362
348	16S_EC_960_981_TMOD_F	TTTCGATGCAACGCGAAGAACCT	38	16S_EC_1054_1073_TMOD_R	TACGAGCTGACGACAGCCATG	363
119	16S_EC_969_985_1P_F	ACGCGAAGAACCTTAU[a]C	39	16S_EC_1061_1078_2P_R	ACGACACGAGU[a]C[a]GACGAC	364
15	16S_EC_969_985_F	ACGCGAAGAACCTTACC	39	16S_EC_1061_1078_R	ACGACACGAGCTGACGAC	364
272	16S_EC_969_985_F	ACGCGAAGAACCTTACC	40	16S_EC_1389_1407_R	GACGGGCGGTGTGTACAAG	378
344	16S_EC_971_990_F	GCGAAGAACCTTACCAGGTC	41	16S_EC_1043_1062_R	ACAACCATGCACCACCTGTC	360
120	16S_EC_972_985_2P_F	CGAAGAAU[a]U[a]TTACC	42	16S_EC_1064_1075_2P_R	ACACGAGU[a]C[a]GAC	365
121	16S_EC_972_985_F	CGAAGAACCTTACC	42	16S_EC_1064_1075_R	ACACGAGCTGAC	365
1073	23S_BRM_1110_1129_F	TGCGCGGAAGATGTAACGGG	43	23S_BRM_1176_1201_R	TCGCAGGCTTACAGAACGCTCTCCTA	397
1074	23S_BRM_515_536_F	TGCATACAAACAGTCGGAGCCT	44	23S_BRM_616_635_R	TCGGACTCGCTTTCGCTACG	398
241	23S_BS_-68_-44_F	AAACTAGATAACAGTAGACATCAC	45	23S_BS_5_21_R	GTGCGCCCTTTCTAACTT	399
235	23S_EC_1602_1620_F	TACCCCAAACCGACACAGG	46	23S_EC_1686_1703_R	CCTTCTCCCGAAGTTACG	402
236	23S_EC_1685_1703_F	CCGTAACTTCGGGAGAAGG	47	23S_EC_1828_1842_R	CACCGGGCAGGCGTC	403
16	23S_EC_1826_1843_F	CTGACACCTGCCCGGTGC	48	23S_EC_1906_1924_R	GACCGTTATAGTTACGGCC	404
349	23S_EC_1826_1843_TMOD_F	TCTGACACCTGCCCGGTGC	49	23S_EC_1906_1924_TMOD_R	TGACCGTTATAGTTACGGCC	405
237	23S_EC_1827_1843_F	GACGCCTGCCCGGTGC	50	23S_EC_1929_1949_R	CCGACAAGGAATTTCGCTACC	407
249	23S_EC_1831_1849_F	ACCTGCCCAGTGCTGGAAG	51	23S_EC_1919_1936_R	TCGCTACCTTAGGACCGT	406
234	23S_EC_187_207_F	GGGAACTGAAACATCTAAGTA	52	23S_EC_242_256_R	TTCGCTCGCCGCTAC	408
233	23S_EC_23_37_F	GGTGGATGCCTTGGC	53	23S_EC_115_130_R	GGGTTTCCCCATTCGG	401
238	23S_EC_2434_2456_F	AAGGTACTCCGGGGATAACAGGC	54	23S_EC_2490_2511_R	AGCCGACATCGAGGTGCCAAAC	409
257	23S_EC_2586_2607_F	TAGAACGTCGCGAGACAGTTCG	55	23S_EC_2658_2677_R	AGTCCATCCCGGTCCTCTCG	411
239	23S_EC_2599_2616_F	GACAGTTCGGTCCCTATC	56	23S_EC_2653_2669_R	CCGGTCCTCTCGTACTA	410
18	23S_EC_2645_2669_2_F	CTGTCCCTAGTACGAGAGGACCGG	57	23S_EC_2751_2767_R	GTTTCATGCTTAGATGCTTTCAGC	417
17	23S_EC_2645_2669_F	TCTGTCCCTAGTACGAGAGGACCGG	58	23S_EC_2744_2761_R	TGCTTAGATGCTTTCAGC	414
118	23S_EC_2646_2667_F	CTGTTCTTAGTACGAGAGGACC	59	23S_EC_2745_2765_R	TTCGTGCTTAGATGCTTTCAG	415
360	23S_EC_264	TCTGTTCTTAGTACGAG	60	23S_EC_2745	TTTCGTGCTTAGATGCTTT	416

	6_2667_TMOD_F	AGGACC		_2765_TMOD_R	CAG	
147	23S_EC_2652_2669_F	CTAGTACGAGAGGACCGG	61	23S_EC_2741_2760_R	ACTTAGATGCTTTCAGCGGT	413
240	23S_EC_2653_2669_F	TAGTACGAGAGGACCGG	62	23S_EC_2737_2758_R	TTAGATGCTTTCAGCACTTATC	412
20	23S_EC_493_518_2_F	GGGGAGTGAAAGAGATCCTGAAACCG	63	23S_EC_551_571_2_R	ACAAAAGGCACGCCATCACCC	418
19	23S_EC_493_518_F	GGGGAGTGAAAGAGATCCTGAAACCG	63	23S_EC_551_571_R	ACAAAAGGTACGCCGTCACCC	419
21	23S_EC_971_992_F	CGAGAGGGAAACAACCCAGACC	64	23S_EC_1059_1077_R	TGGCTGCTTCTAAGCCAAC	400
1158	AB_MLST-11-OIF007_1202_1225_F	TCGTGCCCGCAATTTGCATAAAGC	65	AB_MLST-11-OIF007_1266_1296_R	TAATGCCGGGTAGTGCAATCCATTCTTCTAG	420
1159	AB_MLST-11-OIF007_1202_1225_F	TCGTGCCCGCAATTTGCATAAAGC	65	AB_MLST-11-OIF007_1299_1316_R	TGCACCTGCGGTCGAGCG	421
1160	AB_MLST-11-OIF007_1234_1264_F	TTGTAGCACAGCAAGGCAAATTTCCTGAAAC	66	AB_MLST-11-OIF007_1335_1362_R	TGCCATCCATAATCACGCCATACTGACG	422
1161	AB_MLST-11-OIF007_1327_1356_F	TAGGTTTACGTCAGTATGGCGTGATTATGG	67	AB_MLST-11-OIF007_1422_1448_R	TGCCAGTTTCCACATTTCACGTTCGTG	423
1162	AB_MLST-11-OIF007_1345_1369_F	TCGTGATTATGGATGGCAACGTGAA	68	AB_MLST-11-OIF007_1470_1494_R	TCGCTTGAGTGTAGTCATGATTGCG	424
1163	AB_MLST-11-OIF007_1351_1375_F	TTATGGATGGCAACGTGAAACGCGT	69	AB_MLST-11-OIF007_1470_1494_R	TCGCTTGAGTGTAGTCATGATTGCG	424
1164	AB_MLST-11-OIF007_1387_1412_F	TCTTTGCCATTGAAGATGACTTAAGC	70	AB_MLST-11-OIF007_1470_1494_R	TCGCTTGAGTGTAGTCATGATTGCG	424
1165	AB_MLST-11-OIF007_1542_1569_F	TACTAGCGGTAAGCTTAAACAAGATTGC	71	AB_MLST-11-OIF007_1656_1680_R	TGAGTCGGGTTCACTTTACCTGGCA	425
1166	AB_MLST-11-OIF007_1566_1593_F	TTGCCAATGATATTCGTTGGTTAGCAAG	72	AB_MLST-11-OIF007_1656_1680_R	TGAGTCGGGTTCACTTTACCTGGCA	425
1167	AB_MLST-11-OIF007_1611_1638_F	TCGGCGAAATCCGTATTCCTGAAAATGA	73	AB_MLST-11-OIF007_1731_1757_R	TACCGGAAGCACCAGCGACATTAATAG	427
1168	AB_MLST-11-OIF007_1726_1752_F	TACCACTATTAATGTCGCTGGTGCTTC	74	AB_MLST-11-OIF007_1790_1821_R	TGCAACTGAATAGATTGCAGTAAGTTATAAGC	428
1169	AB_MLST-11-OIF007_1792_1826_F	TTATAACTTACTGCAATCTATTCAGTTGCTTGGTG	75	AB_MLST-11-OIF007_1876_1909_R	TGAATTATGCAAGAAGTGATCAATTTTCTCACGA	429
1170	AB_MLST-11-OIF007_1792_1826_F	TTATAACTTACTGCAATCTATTCAGTTGCTTGGTG	75	AB_MLST-11-OIF007_1895_1927_R	TGCCGTAACTAACATAAGAGAATTATGCAAGAA	430
1152	AB_MLST-11-OIF007_185_214_F	TATTGTTTCAAATGTACAAGGTGAAGTGCG	76	AB_MLST-11-OIF007_291_324_R	TCACAGGTTCTACTTCATCAATAATTTCCATTGC	432
1171	AB_MLST-11-OIF007_1970_2002_F	TGGTTATGTACCAAATACTTTGTCTGAAGATGG	77	AB_MLST-11-OIF007_2097_2118_R	TGACGGCATCGATACCACCGTC	431
1154	AB_MLST-11-OIF007_206_239_F	TGAAGTGCGTGATGATATCGATGCACTTGATGTA	78	AB_MLST-11-OIF007_318_344_R	TCCGCCAAAAACTCCCCTTTTCACAGG	433

1153	AB_MLST-11-OIF007_260_289_F	TGGAACGTTATCAGGTGCCCCAAAAATTCG	79	AB_MLST-11-OIF007_364_393_R	TTGCAATCGACATATCCATTTCACCATGCC	434
1155	AB_MLST-11-OIF007_522_552_F	TCGGTTTAGTAAAAGAACGTATTGCTCAACC	80	AB_MLST-11-OIF007_587_610_R	TTCTGCTTGAGGAATAGTGCGTGG	435
1156	AB_MLST-11-OIF007_547_571_F	TCAACCTGACTGCGTGAATGGTTGT	81	AB_MLST-11-OIF007_656_686_R	TACGTTCTACGATTTCTTCATCAGGTACATC	436
1157	AB_MLST-11-OIF007_601_627_F	TCAAGCAGAAGCTTTGGAAGAAGAAGG	82	AB_MLST-11-OIF007_710_736_R	TACAACGTGATAAACACGACCAGAAGC	437
1151	AB_MLST-11-OIF007_62_91_F	TGAGATTGCTGAACATTTAATGCTGATTGA	83	AB_MLST-11-OIF007_169_203_R	TTGTACATTTGAAACAATATGCATGACATGTGAAT	426
1100	ASD_FRT_1_29_F	TTGCTTAAAGTTGGTTTTATTGGTTGGCG	84	ASD_FRT_86_116_R	TGAGATGTCGAAAAAAACGTTGGCAAAATAC	439
1101	ASD_FRT_43_76_F	TCAGTTTTAATGTCTCGTATGATCGAATCAAAAG	85	ASD_FRT_129_156_R	TCCATATTGTTGCATAAAACCTGTTGGC	438
291	ASPS_EC_405_422_F	GCACAACCTGCGGCTGCG	86	ASPS_EC_521_538_R	ACGGCACGAGGTAGTCGC	440
485	BONTA_X52066_450_473_F	TCTAGTAATAATAGGACCCTCAGC	87	BONTA_X52066_517_539_R	TAACCATTTCGCGTAAGATTCAA	441
486	BONTA_X52066_450_473P_F	T*U^a*C^{a}AGTAATAATAGGA*U^a*U^a*U^a*C^a*U^{a}AGC	87	BONTA_X52066_517_539P_R	TAACCA*C^a*C^a*C^a*U^{a}GCGTAAGA*C^a*C^a*U^{a}AA	441
481	BONTA_X52066_538_552_F	TATGGCTCTACTCAA	88	BONTA_X52066_647_660_R	TGTTACTGCTGGAT	443
482	BONTA_X52066_538_552P_F	TA*C^{a}GGC*C^a*U^a*C^{a}A*U^a*C^a*U^{a}AA	88	BONTA_X52066_647_660P_R	TG*C^a*C^{a}A*U^a*C^{a}G*U^a*C^{a}GGAT	443
487	BONTA_X52066_591_620_F	TGAGTCACTTGAAGTTGATACAAATCCTCT	89	BONTA_X52066_644_671_R	TCATGTGCTAATGTTACTGCTGGATCTG	442
483	BONTA_X52066_701_720_F	GAATAGCAATTAATCCAAAT	90	BONTA_X52066_759_775_R	TTACTTCTAACCCACTC	444
484	BONTA_X52066_701_720P_F	GAA*C^{a}AG*U^{a}AA*C^a*C^{a}AA*C^a*U^a*U^{a}AAAT	90	BONTA_X52066_759_775P_R	TTA*U^a*C^a*C^a*U^a*C^{a}AA*U^a*U^a*U^{a}A*U^a*C^{a}C	444
774	CAF1_AF053947_33407_33430_F	TCAGTTCCGTTATCGCCATTGCAT	91	CAF1_AF053947_33494_33514_R	TGCGGGCTGGTTCAACAAGAG	445
776	CAF1_AF053947_33435_33457_F	TGGAACTATTGCAACTGCTAATG	92	CAF1_AF053947_33499_33517_R	TGATGCGGGCTGGTTCAAC	446
775	CAF1_AF053947_33515_33541_F	TCACTCTTACATATAAGGAAGGCGCTC	93	CAF1_AF053947_33595_33621_R	TCCTGTTTTATAGCCGCCAAGAGTAAG	447
777	CAF1_AF053947_33687_33716_F	TCAGGATGGAAATAACCACCAATTCACTAC	94	CAF1_AF053947_33755_33782_R	TCAAGGTTCTCACCGTTTACCTTAGGAG	448
22	CAPC_BA_104_131_F	GTTATTTAGCACTCGTTTTTAATCAGCC	95	CAPC_BA_180_205_R	TGAATCTTGAAACACCATACGTAACG	449
23	CAPC_BA_114_133_F	ACTCGTTTTTAATCAGCCCG	96	CAPC_BA_185_205_R	TGAATCTTGAAACACCATACG	450
24	CAPC_BA_274_303_F	GATTATTGTTATCCTGTTATGCCATTTGAG	97	CAPC_BA_349_376_R	GTAACCCTTGTCTTTGAATTGTATTTGC	451
350	CAPC_BA_274_303_TMOD_F	TGATTATTGTTATCCTGTTATGCCATTTGAG	98	CAPC_BA_349_376_TMOD_R	TGTAACCCTTGTCTTTGAATTGTATTTGC	452
25	CAPC_BA_276_296_F	TTATTGTTATCCTGTTATGCC	99	CAPC_BA_358_377_R	GGTAACCCTTGTCTTTGAAT	453
26	CAPC_BA_281_301_F	GTTATCCTGTTATGCCATTTG	100	CAPC_BA_361_378_R	TGGTAACCCTTGTCTTTG	454
27	CAPC_BA_315_334_F	CCGTGGTATTGGAGTTATTG	101	CAPC_BA_361_378_R	TGGTAACCCTTGTCTTTG	454
1053	CJST_CJ_10	TTGAGGGTATGCACCGT	102	CJST_CJ_116	TCCCCTCATGTTTAAATGA	456

	80_1110_F	CTTTTTGATTCTTT		6_1198_R	TCAGGATAAAAAGC	
1063	CJST_CJ_1268_1299_F	AGTTATAAACACGGCTTTCCTATGGCTTATCC	103	CJST_CJ_1349_1379_R	TCGGTTTAAGCTCTACATGATCGTAAGGATA	457
1050	CJST_CJ_1290_1320_F	TGGCTTATCCAAATTTAGATCGTGGTTTTAC	104	CJST_CJ_1406_1433_R	TTTGCTCATGATCTGCATGAAGCATAAA	458
1058	CJST_CJ_1643_1670_F	TTATCGTTTGTGGAGCTAGTGCTTATGC	105	CJST_CJ_1724_1752_R	TGCAATGTGTGCTATGTCAGCAAAAAGAT	459
1045	CJST_CJ_1668_1700_F	TGCTCGAGTGATTGACTTTGCTAAATTTAGAGA	106	CJST_CJ_1774_1799_R	TGAGCGTGTGGAAAAGGACTTGGATG	460
1064	CJST_CJ_1680_1713_F	TGATTTTGCTAAATTTAGAGAAATTGCGGATGAA	107	CJST_CJ_1795_1822_R	TATGTGTAGTTGAGCTTACTACATGAGC	461
1056	CJST_CJ_1880_1910_F	TCCCAATTAATTCTGCCATTTTTCCAGGTAT	108	CJST_CJ_1981_2011_R	TGGTTCTTACTTGCTTTGCATAAACTTTCCA	462
1054	CJST_CJ_2060_2090_F	TCCCGGACTTAATATCAATGAAAATTGTGGA	109	CJST_CJ_2148_2174_R	TCGATCCGCATCACCATCAAAAGCAAA	463
1059	CJST_CJ_2165_2194_F	TGCGGATCGTTTGGTGGTTGTAGATGAAAA	110	CJST_CJ_2247_2278_R	TCCACACTGGATTGTAATTTACCTTGTTCTTT	464
1046	CJST_CJ_2171_2197_F	TCGTTTGGTGGTGGTAGATGAAAAAGG	111	CJST_CJ_2283_2313_R	TCTCTTTCAAAGCACCATTGCTCATTATAGT	465
1057	CJST_CJ_2185_2212_F	TAGATGAAAAGGGCGAAGTGGCTAATGG	112	CJST_CJ_2283_2316_R	TGAATTCTTTCAAAGCACCATTGCTCATTATAGT	466
1049	CJST_CJ_2636_2668_F	TGCCTAGAAGATCTTAAAAATTTCCGCCAACTT	113	CJST_CJ_2753_2777_R	TTGCTGCCATAGCAAAGCCTACAGC	467
1062	CJST_CJ_2678_2703_F	TCCCCAGGACACCCTGAAATTTCAAC	114	CJST_CJ_2760_2787_R	TGTGCTTTTTTTGCTGCCATAGCAAAGC	468
1065	CJST_CJ_2857_2887_F	TGGCATTTCTTATGAAGCTTGTTCTTTAGCA	115	CJST_CJ_2965_2998_R	TGCTTCAAAACGCATTTTTACATTTTCGTTAAAG	469
1055	CJST_CJ_2869_2895_F	TGAAGCTTGTTCTTTAGCAGGACTTCA	116	CJST_CJ_2979_3007_R	TCCTCCTTGTGCCTCAAAACGCATTTTTA	470
1051	CJST_CJ_3267_3293_F	TTTGATTTTACGCCGTCCTCCAGGTCG	117	CJST_CJ_3356_3385_R	TCAAAGAACCCGCACCTAATTCATCATTTA	471
1061	CJST_CJ_360_393_F	TCCTGTTATCCCTGAAGTAGTTAATCAAGTTTGT	118	CJST_CJ_443_477_R	TACAACTGGTTCAAAAACATTAAGCTGTAATTGTC	473
1048	CJST_CJ_360_394_F	TCCTGTTATCCCTGAAGTAGTTAATCAAGTTTGTT	119	CJST_CJ_442_476_R	TCAACTGGTTCAAAAACATTAAGTTGTAATTGTCC	472
1052	CJST_CJ_5_39_F	TAGGCGAAGATATACAAAGAGTATTAGAAGCTAGA	120	CJST_CJ_104_137_R	TCCCTTATTTTTCTTTCTACTACCTTCGGATAAT	455
1047	CJST_CJ_584_616_F	TCCAGGACAAATGTATGAAAAATGTCCAAGAAG	121	CJST_CJ_663_692_R	TTCATTTTCTGGTCCAAAGTAAGCAGTATC	474
1060	CJST_CJ_599_632_F	TGAAAAATGTCCAAGAAGCATAGCAAAAAAAGCA	122	CJST_CJ_711_743_R	TCCCGAACAATGAGTTGTATCAACTATTTTTAC	475
1096	CTXA_VBC_117_142_F	TCTTATGCCAAGAGGACAGAGTGAGT	123	CTXA_VBC_194_218_R	TGCCTAACAAATCCCGTCTGAGTTC	476
1097	CTXA_VBC_351_377_F	TGTATTAGGGGCATACAGTCCTCATCC	124	CTXA_VBC_441_466_R	TGTCATCAAGCACCCCAAAATGAACT	477
28	CYA_BA_1055_1072_F	GAAAGAGTTCGGATTGGG	125	CYA_BA_1112_1130_R	TGTTGACCATGCTTCTTAG	479
277	CYA_BA_1349_1370_F	ACAACGAAGTACAATACAAGAC	126	CYA_BA_1426_1447_R	CTTCTACATTTTTAGCCATCAC	480
30	CYA_BA_1353_1379_F	CGAAGTACAATACAAGACAAAAGAAGG	127	CYA_BA_1448_1467_R	TGTTAACGGCTTCAAGACCC	482
351	CYA_BA_1353_1379_TMOD_F	TCGAAGTACAATACAAGACAAAAGAAGG	128	CYA_BA_1448_1467_TMOD_R	TTGTTAACGGCTTCAAGACCC	483
31	CYA_BA_1359_1379_F	ACAATACAAGACAAAAGAAGG	129	CYA_BA_1447_1461_R	CGGCTTCAAGACCCC	481
32	CYA_BA_914_937_F	CAGGTTTAGTACCAGAACATGCAG	130	CYA_BA_999_1026_R	ACCACTTTTAATAAGGTTTGTAGCTAAC	484
33	CYA_BA_916_935_F	GGTTTAGTACCAGAACATGC	131	CYA_BA_1003_1025_R	CCACTTTTAATAAGGTTTGTAGC	478
115	DNAK_EC_428_449_F	CGGCGTACTTCAACGACAGCCA	132	DNAK_EC_503_522_R	CGCGGTCGGCTCGTTGATGA	485
1102	GALE_FRT_168_199_F	TTATCAGCTAGACCTTTTAGGTAAAGCTAAGC	133	GALE_FRT_241_269_R	TCACCTACAGCTTTAAAGCCAGCAAAATG	486
1104	GALE_FRT_308_339_F	TCCAAGGTACACTAAACTTACTTGAGCTAATG	134	GALE_FRT_390_422_R	TCTTCTGTAAAGGGTGGTTTATTATTCATCCCA	487
1103	GALE_FRT_834_865_F	TCAAAAAGCCCTAGGTAAAGAGATTCCATATC	135	GALE_FRT_901_925_R	TAGCCTTGGCAACATCAGCAAAACT	488
1092	GLTA_RKP_1023_1055_F	TCCGTTCTTACAAATAGCAATAGAACTTGAAGC	136	GLTA_RKP_1129_1156_R	TTGGCGACGGTATACCCATAGCTTTATA	489
1093	GLTA_RKP_1043_1072_2_F	TGGAGCTTGAAGCTATCGCTCTTAAAGATG	137	GLTA_RKP_1138_1162_R	TGAACATTTGCGACGGTATACCCAT	490

1094	GLTA_RKP_1043_1072_3_F	TGGAACTTGAAGCTCTCGCTCTTAAAGATG	138	GLTA_RKP_1138_1164_R	TGTGAACATTTGCGACGGTATACCCAT	492
1090	GLTA_RKP_1043_1072_F	TGGGACTTGAAGCTATCGCTCTTAAAGATG	139	GLTA_RKP_1138_1162_R	TGAACATTTGCGACGGTATACCCAT	491
1091	GLTA_RKP_400_428_F	TCTTCTCATCCTATGGCTATTATGCTTGC	140	GLTA_RKP_499_529_R	TGGTGGGTATCTTAGCAATCATTCTAATAGC	493
1095	GLTA_RKP_400_428_F	TCTTCTCATCCTATGGCTATTATGCTTGC	140	GLTA_RKP_505_534_R	TGCGATGGTAGGTATCTTAGCAATCATTCT	494
224	GROL_EC_219_242_F	GGTGAAAGAAGTTGCCTCTAAAGC	141	GROL_EC_328_350_R	TTCAGGTCCATCGGGTTCATGCC	496
280	GROL_EC_496_518_F	ATGGACAAGGTTGGCAAGGAAGG	142	GROL_EC_577_596_R	TAGCCGCGGTCGAATTGCAT	498
281	GROL_EC_511_536_F	AAGGAAGGCGTGATCACCGTTGAAGA	143	GROL_EC_571_593_R	CCGCGGTCGAATTGCATGCCTTC	497
220	GROL_EC_941_959_F	TGGAAGATCTGGGTCAGGC	144	GROL_EC_1039_1060_R	CAATCTGCTGACGGATCTGAGC	495
924	GYRA_AF100557_4_23_F	TCTGCCCGTGTCGTTGGTGA	145	GYRA_AF100557_119_142_R	TCGAACCGAAGTTACCCTGACCAT	499
925	GYRA_AF100557_70_94_F	TCCATTGTTCGTATGGCTCAAGACT	146	GYRA_AF100557_178_201_R	TGCCAGCTTAGTCATACGGACTTC	500
926	GYRB_AB008700_19_40_F	TCAGGTGGCTTACACGGCGTAG	147	GYRB_AB008700_111_140_R	TATTGCGGATCACCATGATGATATTCTTGC	501
927	GYRB_AB008700_265_292_F	TCTTTCTTGAATGCTGGTGTACGTATCG	148	GYRB_AB008700_369_395_R	TCGTTGAGATGGTTTTTACCTTCGTTG	502
928	GYRB_AB008700_368_394_F	TCAACGAAGGTAAAAACCATCTCAACG	149	GYRB_AB008700_466_494_R	TTTGTGAAACAGCGAACATTTTCTTGGTA	503
929	GYRB_AB008700_477_504_F	TGTTCGCTGTTTCACAAACAACATTCCA	150	GYRB_AB008700_611_632_R	TCACGCGCATCATCACCAGTCA	504
949	GYRB_AB008700_760_787_F	TACTTACTTGAGAATCCACAAGCTGCAA	151	GYRB_AB008700_862_888_2_R	TCCTGCAATATCTAATGCACTCTTACG	505
930	GYRB_AB008700_760_787_F	TACTTACTTGAGAATCCACAAGCTGCAA	151	GYRB_AB008700_862_888_R	ACCTGCAATATCTAATGCACTCTTACG	506
222	HFLB_EC_1082_1102_F	TGGCGAACCTGGTGAACGAAGC	152	HFLB_EC_1144_1168_R	CTTTCGCTTTCTCGAACTCAACCAT	507
1128	HUPB_CJ_113_134_F	TAGTTGCTCAAACAGCTGGGCT	153	HUPB_CJ_157_188_R	TCCCTAATAGTAGAAATAACTGCATCAGTAGC	509
1130	HUPB_CJ_76_102_F	TCCCGGAGCTTTTATGACTAAAGCAGAT	154	HUPB_CJ_114_135_R	TAGCCCAGCTGTTTGAGCAACT	508
1129	HUPB_CJ_76_102_F	TCCCGGAGCTTTTATGACTAAAGCAGAT	154	HUPB_CJ_157_188_R	TCCCTAATAGTAGAAATAACTGCATCAGTAGC	510
1079	ICD_CXB_176_198_F	TCGCCGTGGAAAAATCCTACGCT	155	ICD_CXB_224_247_R	TAGCCTTTTCTCCGGCGTAGATCT	512
1078	ICD_CXB_92_120_F	TTCCTGACCGACCCATTATTCCCTTTATC	156	ICD_CXB_172_194_R	TAGGATTTTTCCACGGCGGCATC	510
1077	ICD_CXB_93_120_F	TCCTGACCGACCCATTATTCCCTTTATC	157	ICD_CXB_172_194_R	TAGGATTTTTCCACGGCGGCATC	511
221	INFB_EC_1103_1124_F	GTCGTGAAAACGAGCTGGAAGA	158	INFB_EC_1174_1191_R	CATGATGGTCACAACCGG	513
964	INFB_EC_1347_1367_F	TGCGTTTACCGCAATGCGTGC	159	INFB_EC_1414_1432_R	TCGGCATCACGCCGTCGTC	514
34	INFB_EC_1365_1393_F	TGCTCGTGGTGCACAAGTAACGGATATTA	160	INFB_EC_1439_1467_R	TGCTGCTTTCGCATGGTTAATTGCTTCAA	515
352	INFB_EC_1365_1393_TMOD_F	TTGCTCGTGGTGCACAAGTAACGGATATTA	161	INFB_EC_1439_1467_TMOD_R	TTGCTGCTTTCGCATGGTTAATTGCTTCAA	516
223	INFB_EC_1969_1994_F	CGTCAGGGTAAATTCCGTGAAGTTAA	162	INFB_EC_2038_2058_R	AACTTCGCCTTCGGTCATGTT	517
781	INV_U22457_1558_1581_F	TGGTAACAGAGCCTTATAGGCGCA	163	INV_U22457_1619_1643_R	TTGCGTTGCAGATTATCTTTACCAA	518
778	INV_U22457_515_539_F	TGGCTCCTTGGTATGACTCTGCTTC	164	INV_U22457_571_598_R	TGTTAAGTGTGTTGCGGCTGTCTTTATT	519
779	INV_U22457_699_724_F	TGCTGAGGCCTGGACCGATTATTTAC	165	INV_U22457_753_776_R	TCACGCGACGAGTGCCATCCATTG	520
780	INV_U22457_834_858_F	TTATTTACCTGCACTCCCACAACTG	166	INV_U22457_942_966_R	TGACCCAAAGCTGAAAGCTTTACTG	521

1106	IPAH_SGF_113_134_F	TCCTTGACCGCCTTTCCGATAC	167	IPAH_SGF_172_191_R	TTTTCCAGCCATGCAGCGAC	522
1105	IPAH_SGF_258_277_F	TGAGGACCGTGTCGCGCTCA	168	IPAH_SGF_301_327_R	TCCTTCTGATGCCTGATGGACCAGGAG	523
1107	IPAH_SGF_462_486_F	TCAGACCATGCTCGCAGAGAAACTT	169	IPAH_SGF_522_540_R	TGTCACTCCCGACACGCCA	524
1080	IS1111A_NC002971_6866_6891_F	TCAGTATGTATCCACCGTAGCCAGTC	170	IS1111A_NC002971_6928_6954_R	TAAACGTCCGATACCAATGGTTCGCTC	525
1081	IS1111A_NC002971_7456_7483_F	TGGGTGACATTCATCAATTTCATCGTTC	171	IS1111A_NC002971_7529_7554_R	TCAACAACACCTCCTTATTCCCACTC	526
35	LEF_BA_1033_1052_F	TCAAGAAGAAAAAGAGC	172	LEF_BA_1119_1135_R	GAATATCAATTTGTAGC	527
36	LEF_BA_1036_1066_F	CAAGAAGAAAAAGAGCTTCTAAAAAGAATAC	173	LEF_BA_1119_1149_R	AGATAAAGAATCACGAATATCAATTTGTAGC	528
37	LEF_BA_756_781_F	AGCTTTTGCATATTATATCGAGCCAC	174	LEF_BA_843_872_R	TCTTCCAAGGATAGATTTATTTCTTGTTCG	530
353	LEF_BA_756_781_TMOD_F	TAGCTTTTGCATATTATATCGAGCCAC	175	LEF_BA_843_872_TMOD_R	TTCTTCCAAGGATAGATTTATTTCTTGTTCG	531
38	LEF_BA_758_778_F	CTTTTGCATATTATATCGAGC	176	LEF_BA_843_865_R	AGGATAGATTTATTTCTTGTTCG	529
39	LEF_BA_795_813_F	TTTACAGCTTTATGCACCG	177	LEF_BA_883_900_R	TCTTGACAGCATCCGTTG	532
40	LEF_BA_883_899_F	CAACGGATGCTGGCAAG	178	LEF_BA_939_958_R	CAGATAAAGAATCGCTCCAG	533
782	LL_NC003143_2366996_2367019_F	TGTAGCCGCTAAGCACTACCATCC	179	LL_NC003143_2367073_2367097_R	TCTCATCCCGATATTACCGCCATGA	534
783	LL_NC003143_2367172_2367194_F	TGGACGGCATCACGATTCTCTAC	180	LL_NC003143_2367249_2367271_R	TGGCAACAGCTCAACACCTTTGG	535
878	MECA_Y14051_3645_3670_F	TGAAGTAGAAATGACTGAACGTCCGA	181	MECA_Y14051_3690_3719_R	TGATCCTGAATGTTTATATCTTTAACGCCT	536
877	MECA_Y14051_3774_3802_F	TAAAACAAACTACGGTAACATTGATCGCA	182	MECA_Y14051_3828_3854_R	TCCCAATCTAACTTCCACATACCATCT	537
879	MECA_Y14051_4507_4530_F	TCAGGTACTGCTATCCACCCTCAA	183	MECA_Y14051_4555_4581_R	TGGATAGACGTCATATGAAGGTGTGCT	538
880	MECA_Y14051_4510_4530_F	TGTACTGCTATCCACCCTCAA	184	MECA_Y14051_4586_4610_R	TATTCTTCGTTACTCATGCCATACA	539
882	MECA_Y14051_4520_4530P_F	TU^{a}U^{a}AU^{a}U^{a}U^{a}C^{a}U^{a}AA	185	MECA_Y14051_4590_4600P_R	C^{a}AU^{a}C^{a}U^{a}AC^{a}GU^{a}U^{a}A	540
883	MECA_Y14051_4520_4530P_F	TU^{a}U^{a}AU^{a}U^{a}U^{a}C^{a}U^{a}AA	185	MECA_Y14051_4600_4610P_R	C^{a}AC^{a}C^{a}U^{a}C^{a}C^{a}U^{a}GC^{a}T	541
881	MECA_Y14051_4669_4698_F	TCACCAGGTTCAACTCAAAAAATATTAACA	186	MECA_Y14051_4765_4793_R	TAACCACCCCAAGATTTATCTTTTTGCCA	542
876	MECIA_Y14051_3315_3341_F	TTACACATATCGTGAGCAATGAACTGA	187	MECIA_Y14051_3367_3393_R	TGTGATATGGAGGTGTAGAAGGTGTTA	543
914	OMPA_AY485227_272_301_F	TTACTCCATTATTGCTTGGTTACACTTTCC	188	OMPA_AY485227_364_388_R	GAGCTGCGCCAACGAATAAATCGTC	544
916	OMPA_AY485227_311_335_F	TACACAACAATGGCGGTAAAGATGG	189	OMPA_AY485227_424_453_R	TACGTCGCCTTTAACTTGGTTATATTCAGC	545
915	OMPA_AY485227_379_401_F	TGCGCAGCTCTTGGTATCGAGTT	190	OMPA_AY485227_492_519_R	TGCCGTAACATAGAAGTTACCGTTGATT	546
917	OMPA_AY485227_415_441_F	TGCCTCGAAGCTGAATATAACCAAGTT	191	OMPA_AY485227_514_546_R	TCGGGCGTAGTTTTTAGTAATTAAATCAGAAGT	547
918	OMPA_AY485227_494_520_F	TCAACGGTAACTTCTATGTTACTTCTG	192	OMPA_AY485227_569_596_R	TCGTCGTATTTATAGTGACCAGCACCTA	548
919	OMPA_AY485227_551_577_F	TCAAGCCGTACGTATTATTAGGTGCTG	193	OMPA_AY485227_658_680_R	TTTAAGCGCCAGAAAGCACCAAC	550

920	OMPA_AY485227_555_581_F	TCCGTACGTATTATTAGGTGCTGGTCA	194	OMPA_AY485227_635_662_R	TCAACACCAGCGTTACCTAAAGTACCTT	549
921	OMPA_AY485227_556_583_F	TCGTACGTATTATTAGGTGCTGGTCACT	195	OMPA_AY485227_659_683_R	TCGTTTAAGCGCCAGAAAGCACCAA	551
922	OMPA_AY485227_657_679_F	TGTTGGTGCTTTCTGGCGCTTAA	196	OMPA_AY485227_739_765_R	TAAGCCAGCAAGAGCTGTATAGTTCCA	552
923	OMPA_AY485227_660_683_F	TGGTGCTTTCTGGCGCTTAAACGA	197	OMPA_AY485227_786_807_R	TACAGGAGCAGCAGGCTTCAAG	553
1088	OMPB_RKP_1192_1221_F	TCTACTGATTTTGGTAATCTTGCAGCACAG	198	OMPB_RKP_1288_1315_R	TAGCAGCAAAAGTTATCACACCTGCAGT	554
1089	OMPB_RKP_3417_3440_F	TGCAAGTGGTACTTCAACATGGGG	199	OMPB_RKP_3520_3550_R	TGGTTGTAGTTCCTGTAGTTGTTGCATTAAC	555
1087	OMPB_RKP_860_890_F	TTACAGGAAGTTTAGGTGGTAATCTAAAAGG	200	OMPB_RKP_972_996_R	TCCTGCAGCTCTACCTGCTCCATTA	556
41	PAG_BA_122_142_F	CAGAATCAAGTTCCCAGGGG	201	PAG_BA_190_209_R	CCTGTAGTAGAAGAGGTAAC	558
42	PAG_BA_123_145_F	AGAATCAAGTTCCCAGGGGTTAC	202	PAG_BA_187_210_R	CCCTGTAGTAGAAGAGGTAACCAC	557
43	PAG_BA_269_287_F	AATCTGCTATTTGGTCAGG	203	PAG_BA_326_344_R	TGATTATCAGCGGAAGTAG	559
44	PAG_BA_655_675_F	GAAGGATATACGGTTGATGTC	204	PAG_BA_755_772_R	CCGTGCTCCATTTTTCAG	560
45	PAG_BA_753_772_F	TCCTGAAAAATGGAGCACGG	205	PAG_BA_849_868_R	TCGGATAAGCTGCCACAAGG	561
46	PAG_BA_763_781_F	TGGAGCACGGCTTCTGATC	206	PAG_BA_849_868_R	TCGGATAAGCTGCCACAAGG	562
912	PARC_X95819_123_147_F	GGCTCAGCCATTTAGTTACCGCTAT	207	PARC_X95819_232_260_R	TCGCTCAGCAATAATTCACTATAAGCCGA	566
913	PARC_X95819_43_63_F	TCAGCGCGTACAGTGGGTGAT	208	PARC_X95819_143_170_R	TTCCCCTGACCTTCGATTAAAGGATAGC	563
911	PARC_X95819_87_110_F	TGGTGACTCGGCATGTTATGAAGC	209	PARC_X95819_192_219_R	GGTATAACGCATCGCAGCAAAAGATTTA	564
910	PARC_X95819_87_110_F	TGGTGACTCGGCATGTTATGAAGC	209	PARC_X95819_201_222_R	TTCGGTATAACGCATCGCAGCA	565
773	PLA_AF053945_7186_7211_F	TTATACCGGAAACTTCCCGAAAGGAG	210	PLA_AF053945_7257_7280_R	TAATGCGATACTGGCCTGCAAGTC	567
770	PLA_AF053945_7377_7402_F	TGACATCCGGCTCACGTTATTATGGT	211	PLA_AF053945_7434_7462_R	TGTAAATTCCGCAAAGACTTTGGCATTAG	568
771	PLA_AF053945_7382_7404_F	TCCGGCTCACGTTATTATGGTAC	212	PLA_AF053945_7482_7502_R	TGGTCTGAGTACCTCCTTTGC	569
772	PLA_AF053945_7481_7503_F	TGCAAAGGAGGTACTCAGACCAT	213	PLA_AF053945_7539_7562_R	TATTGGAAATACCGGCAGCATCTC	570
909	RECA_AF251469_169_190_F	TGACATGCTTGTCCGTTCAGGC	214	RECA_AF251469_277_300_R	TGGCTCATAAGACGCGCTTGTAGA	572
908	RECA_AF251469_43_68_F	TGGTACATGTGCCTTCATTGATGCTG	215	RECA_AF251469_140_163_R	TTCAAGTGCTTGCTCACCATTGTC	571
1072	RNASEP_BDP_574_592_F	TGGCACGGCCATCTCCGTG	216	RNASEP_BDP_616_635_R	TCGTTTCACCCTGTCATGCCG	573
1070	RNASEP_BKM_580_599_F	TGCGGGTAGGGAGCTTGAGC	217	RNASEP_BKM_665_686_R	TCCGATAAGCCGGATTCTGTGC	574
1071	RNASEP_BKM_616_637_F	TCCTAGAGGAATGGCTGCCACG	218	RNASEP_BKM_665_687_R	TGCCGATAAGCCGGATTCTGTGC	575
1112	RNASEP_BRM_325_347_F	TACCCCAGGGAAAGTGCCACAGA	219	RNASEP_BRM_402_428_R	TCTCTTACCCCACCCTTTCACCCTTAC	576
1172	RNASEP_BRM_461_488_F	TAAACCCCATCGGGAGCAAGACCGAATA	220	RNASEP_BRM_542_561_2_R	TGCCTCGTGCAACCCACCCG	577
1111	RNASEP_BRM_461_488_F	TAAACCCCATCGGGAGCAAGACCGAATA	220	RNASEP_BRM_542_561_R	TGCCTCGCGCAACCTACCCG	578
258	RNASEP_BS_43_61_F	GAGGAAAGTCCATGCTCGC	221	RNASEP_BS_363_384_R	GTAAGCCATGTTTTGTTCCATC	579
259	RNASEP_BS_43_61_F	GAGGAAAGTCCATGCTCGC	221	RNASEP_BS_363_384_R	GTAAGCCATGTTTTGTTCCATC	578
258	RNASEP_BS_43_61_F	GAGGAAAGTCCATGCTCGC	221	RNASEP_EC_345_362_R	ATAAGCCGGGTTCTGTCG	581

258	RNASEP_BS_43_61_F	GAGGAAAGTCCATGCTCGC	221	RNASEP_SA_358_379_R	ATAAGCCATGTTCTGTTCCATC	584
1076	RNASEP_CLB_459_487_F	TAAGGATAGTGCAACAGAGATATACCGCC	222	RNASEP_CLB_498_522_R	TTTACCTCGCCTTTCCACCCTTACC	579
1075	RNASEP_CLB_459_487_F	TAAGGATAGTGCAACAGAGATATACCGCC	222	RNASEP_CLB_498_526_R	TGCTCTTACCTCACCGTTCCACCCTTACC	580
258	RNASEP_EC_61_77_F	GAGGAAAGTCCGGGCTC	223	RNASEP_BS_363_384_R	GTAAGCCATGTTTTGTTCCATC	578
258	RNASEP_EC_61_77_F	GAGGAAAGTCCGGGCTC	223	RNASEP_EC_345_362_R	ATAAGCCGGGTTCTGTCG	581
260	RNASEP_EC_61_77_F	GAGGAAAGTCCGGGCTC	223	RNASEP_EC_345_362_R	ATAAGCCGGGTTCTGTCG	581
258	RNASEP_EC_61_77_F	GAGGAAAGTCCGGGCTC	223	RNASEP_SA_358_379_R	ATAAGCCATGTTCTGTTCCATC	584
1085	RNASEP_RKP_264_287_F	TCTAAATGGTCGTGCAGTTGCGTG	224	RNASEP_RKP_295_321_R	TCTATAGAGTCCGGACTTTCCTCGTGA	582
1082	RNASEP_RKP_419_448_F	TGGTAAGAGCGCACCGGTAAGTTGGTAACA	225	RNASEP_RKP_542_565_R	TCAAGCGATCTACCCGCATTACAA	583
1083	RNASEP_RKP_422_443_F	TAAGAGCGCACCGGTAAGTTGG	226	RNASEP_RKP_542_565_R	TCAAGCGATCTACCCGCATTACAA	583
1086	RNASEP_RKP_426_448_F	TGCATACCGGTAAGTTGGCAACA	227	RNASEP_RKP_542_565_R	TCAAGCGATCTACCCGCATTACAA	583
1084	RNASEP_RKP_466_491_F	TCCACCAAGAGCAAGATCAAATAGGC	228	RNASEP_RKP_542_565_R	TCAAGCGATCTACCCGCATTACAA	583
258	RNASEP_SA_31_49_F	GAGGAAAGTCCATGCTCAC	229	RNASEP_BS_363_384_R	GTAAGCCATGTTTTGTTCCATC	578
258	RNASEP_SA_31_49_F	GAGGAAAGTCCATGCTCAC	229	RNASEP_EC_345_362_R	ATAAGCCGGGTTCTGTCG	581
258	RNASEP_SA_31_49_F	GAGGAAAGTCCATGCTCAC	229	RNASEP_SA_358_379_R	ATAAGCCATGTTCTGTTCCATC	584
262	RNASEP_SA_31_49_F	GAGGAAAGTCCATGCTCAC	229	RNASEP_SA_358_379_R	ATAAGCCATGTTCTGTTCCATC	584
1098	RNASEP_VBC_331_349_F	TCCGCGGAGTTGACTGGGT	230	RNASEP_VBC_388_414_R	TGACTTTCCTCCCCCTTATCAGTCTCC	585
66	RPLB_EC_650_679_F	GACCTACAGTAAGAGGTTCTGTAATGAACC	231	RPLB_EC_739_762_R	TCCAAGTGCTGGTTTACCCCATGG	591
356	RPLB_EC_650_679_TMOD_F	TGACCTACAGTAAGAGGTTCTGTAATGAACC	232	RPLB_EC_739_762_TMOD_R	TTCCAAGTGCTGGTTTACCCCATGG	592
73	RPLB_EC_669_698_F	TGTAATGAACCCTAATGACCATCCACACGG	233	RPLB_EC_735_761_R	CCAAGTGCTGGTTTACCCCATGGAGTA	586
74	RPLB_EC_671_700_F	TAATGAACCCTAATGACCATCCACACGGTG	234	RPLB_EC_737_762_R	TCCAAGTGCTGGTTTACCCCATGGAG	590
67	RPLB_EC_688_710_F	CATCCACACGGTGGTGGTGAAGG	235	RPLB_EC_736_757_R	GTGCTGGTTTACCCCATGGAGT	587
70	RPLB_EC_688_710_F	CATCCACACGGTGGTGGTGAAGG	235	RPLB_EC_743_771_R	TGTTTTGTATCCAAGTGCTGGTTTACCCC	593
357	RPLB_EC_688_710_TMOD_F	TCATCCACACGGTGGTGGTGAAGG	236	RPLB_EC_736_757_TMOD_R	TGTGCTGGTTTACCCCATGGAGT	588
449	RPLB_EC_690_710_F	TCCACACGGTGGTGGTGAAGG	237	RPLB_EC_737_758_R	TGTGCTGGTTTACCCCATGGAG	589
113	RPOB_EC_1336_1353_F	GACCACCTCGGCAACCGT	238	RPOB_EC_1438_1455_R	TTCGCTCTCGGCCTGGCC	594
963	RPOB_EC_1527_1549_F	TCAGCTGTCGCAGTTCATGGACC	239	RPOB_EC_1630_1649_R	TCGTCGCGGACTTCGAAGCC	595
72	RPOB_EC_1845_1866_F	TATCGCTCAGGCGAACTCCAAC	240	RPOB_EC_1909_1929_R	GCTGGATTCGCCTTTGCTACG	596
359	RPOB_EC_1845_1866_TMOD_F	TTATCGCTCAGGCGAACTCCAAC	241	RPOB_EC_1909_1929_TMOD_R	TGCTGGATTCGCCTTTGCTACG	597
962	RPOB_EC_2005_2027_F	TCGTTCCTGGAACACGATGACGC	242	RPOB_EC_2041_2064_R	TTGACGTTGCATGTTCGAGCCCAT	598
69	RPOB_EC_3762_3790_F	TCAACAACCTCTTGGAGGTAAAGCTCAGT	243	RPOB_EC_3836_3865_R	TTTCTTGAAGAGTATGAGCTGCTCCGTAAG	600
111	RPOB_EC_3775_3803_F	CTTGGAGGTAAGTCTCATTTTGGTGGGCA	244	RPOB_EC_3829_3858_R	CGTATAAGCTGCACCATAAGCTTGTAATGC	599
940	RPOB_EC_3798_3821_F	TGGGCAGCGTTTCGGCGAAATGGA	245	RPOB_EC_3862_3889_2_R	TGTCCGACTTGACGGTTAGCATTTCCTG	604
939	RPOB_EC_3798_3821_F	TGGGCAGCGTTTCGGCGAAATGGA	245	RPOB_EC_3862_3889_R	TGTCCGACTTGACGGTCAGCATTTCCTG	605
289	RPOB_EC_3799_3821_F	GGGCAGCGTTTCGGCGAAATGGA	246	RPOB_EC_3862_3888_R	GTCCGACTTGACGGTCAACATTTCCTG	602
362	RPOB_EC_3799_3821_TM	TGGGCAGCGTTTCGGCGAAATGGA	245	RPOB_EC_3862_3888_TMOD	TGTCCGACTTGACGGTCAACATTTCCTG	603

	OD_F			_R		
288	RPOB_EC_3802_3821_F	CAGCGTTTCGGCGAAATGGA	247	RPOB_EC_3862_3885_R	CGACTTGACGGTTAACATTTCCTG	601
48	RPOC_EC_1018_1045_2_F	CAAAACTTATTAGGTAAGCGTGTTGACT	248	RPOC_EC_1095_1124_2_R	TCAAGCGCCATCTCTTTCGGTAATCCACAT	610
47	RPOC_EC_1018_1045_F	CAAAACTTATTAGGTAAGCGTGTTGACT	248	RPOC_EC_1095_1124_R	TCAAGCGCCATTTCTTTTGGTAAACCACAT	611
68	RPOC_EC_1036_1060_F	CGTGTTGACTATTCGGGGCGTTCAG	249	RPOC_EC_1097_1126_R	ATTCAAGAGCCATTTCTTTTGGTAAACCAC	612
49	RPOC_EC_114_140_F	TAAGAAGCCGGAAACCATCAACTACCG	250	RPOC_EC_213_232_R	GGCGCTTGTACTTACCGCAC	617
227	RPOC_EC_1256_1277_F	ACCCAGTGCTGCTGAACCGTGC	251	RPOC_EC_1295_1315_R	GTTCAAATGCCTGGATACCCA	613
292	RPOC_EC_1374_1393_F	CGCCGACTTCGACGGTGACC	252	RPOC_EC_1437_1455_R	GAGCATCAGCGTGCGTGCT	614
364	RPOC_EC_1374_1393_TMOD_F	TCGCCGACTTCGACGGTGACC	253	RPOC_EC_1437_1455_TMOD_R	TGAGCATCAGCGTGCGTGCT	615
229	RPOC_EC_1584_1604_F	TGGCCCGAAAGAAGCTGAGCG	254	RPOC_EC_1623_1643_R	ACGCGGGCATGCAGAGATGCC	616
978	RPOC_EC_2145_2175_F	TCAGGAGTCGTTCAACTCGATCTACATGATG	255	RPOC_EC_2228_2247_R	TTACGCCATCAGGCCACGCA	622
290	RPOC_EC_2146_2174_F	CAGGAGTCGTTCAACTCGATCTACATGAT	256	RPOC_EC_2227_2245_R	ACGCCATCAGGCCACGCAT	620
363	RPOC_EC_2146_2174_TMOD_F	TCAGGAGTCGTTCAACTCGATCTACATGAT	257	RPOC_EC_2227_2245_TMOD_R	TACGCCATCAGGCCACGCAT	621
51	RPOC_EC_2178_2196_2_F	TGATTCCGGTGCCCGTGGT	258	RPOC_EC_2225_2246_2_R	TTGGCCATCAGACCACGCATAC	618
50	RPOC_EC_2178_2196_F	TGATTCTGGTGCCCGTGGT	259	RPOC_EC_2225_2246_R	TTGGCCATCAGGCCACGCATAC	619
53	RPOC_EC_2218_2241_2_F	CTTGCTGGTATGCGTGGTCTGATG	260	RPOC_EC_2313_2337_2_R	CGCACCATGCGTAGAGATGAAGTAC	623
52	RPOC_EC_2218_2241_F	CTGGCAGGTATGCGTGGTCTGATG	261	RPOC_EC_2313_2337_R	CGCACCGTGGGTTGAGATGAAGTAC	624
354	RPOC_EC_2218_2241_TMOD_F	TCTGGCAGGTATGCGTGGTCTGATG	262	RPOC_EC_2313_2337_TMOD_R	TCGCACCGTGGGTTGAGATGAAGTAC	625
958	RPOC_EC_2223_2243_F	TGGTATGCGTGGTCTGATGGC	263	RPOC_EC_2329_2352_R	TGCTAGACCTTTACGTGCACCGTG	626
960	RPOC_EC_2334_2357_F	TGCTCGTAAGGGTCTGGCGGATAC	264	RPOC_EC_2380_2403_R	TACTAGACGACGGGTCAGGTAACC	627
55	RPOC_EC_808_833_2_F	CGTCGTGTAATTAACCGTAACAACCG	265	RPOC_EC_865_891_R	ACGTTTTTCGTTTTGAACGATAATGCT	629
54	RPOC_EC_808_833_F	CGTCGGGTGATTAACCGTAACAACCG	266	RPOC_EC_865_889_R	GTTTTTCGTTGCGTACGATGATGTC	628
961	RPOC_EC_917_938_F	TATTGGACAACGGTCGTCGCGG	267	RPOC_EC_1009_1034_R	TTACCGAGCAGGTTCTGACGGAAACG	607
959	RPOC_EC_918_938_F	TCTGGATAACGGTCGTCGCGG	268	RPOC_EC_1009_1031_R	TCCAGCAGGTTCTGACGGAAACG	606
57	RPOC_EC_993_1019_2_F	CAAAGGTAAGCAAGGACGTTTCCGTCA	269	RPOC_EC_1036_1059_2_R	CGAACGGCCAGAGTAGTCAACACG	608
56	RPOC_EC_993_1019_F	CAAAGGTAAGCAAGGTCGTTTCCGTCA	270	RPOC_EC_1036_1059_R	CGAACGGCCTGAGTAGTCAACACG	609
75	SP101_SPET11_1_29_F	AACCTTAATTGGAAAGAAACCCAAGAAGT	271	SP101_SPET11_92_116_R	CCTACCCAACGTTCACCAAGGGCAG	676
446	SP101_SPET11_1_29_TMOD_F	TAACCTTAATTGGAAAGAAACCCAAGAAGT	272	SP101_SPET11_92_116_TMOD_R	TCCTACCCAACGTTCACCAAGGGCAG	677
85	SP101_SPET11_1154_1179_F	CAATACCGCAACAGCGGTGGCTTGGG	273	SP101_SPET11_1251_1277_R	GACCCCAACCTGGCCTTTTGTCGTTGA	630
424	SP101_SPET11_1154_1179_TMOD_F	TCAATACCGCAACAGCGGTGGCTTGGG	274	SP101_SPET11_1251_1277_TMOD_R	TGACCCCAACCTGGCCTTTTGTCGTTGA	631
76	SP101_SPET11_118_147_F	GCTGGTGAAAATAACCCAGATGTCGTCTTC	275	SP101_SPET11_213_238_R	TGTGGCCGATTTCACCACCTGCTCCT	644
425	SP101_SPET11_118_147_TMOD_F	TGCTGGTGAAAATAACCCAGATGTCGTCTTC	276	SP101_SPET11_213_238_TMOD_R	TTGTGGCCGATTTCACCACCTGCTCCT	645
86	SP101_SPET	CGCAAAAAAATCCAGCT	277	SP101_SPET1	AAACTATTTTTTTAGCTAT	632

	11_1314_1336_F	ATTAGC		1_1403_1431_R	ACTCGAACAC	
426	SP101_SPET11_1314_1336_TMOD_F	TCGCAAAAAAATCCAGCTATTAGC	278	SP101_SPET11_1403_1431_TMOD_R	TAAACTATTTTTTTAGCTATACTCGAACAC	633
87	SP101_SPET11_1408_1437_F	CGAGTATAGCTAAAAAAATAGTTTATGACA	279	SP101_SPET11_1486_1515_R	GGATAATTGGTCGTAACAAGGGATAGTGAG	634
427	SP101_SPET11_1408_1437_TMOD_F	TCGAGTATAGCTAAAAAAATAGTTTATGACA	280	SP101_SPET11_1486_1515_TMOD_R	TGGATAATTGGTCGTAACAAGGGATAGTGAG	635
88	SP101_SPET11_1688_1716_F	CCTATATTAATCGTTTACAGAAACTGGCT	281	SP101_SPET11_1783_1808_R	ATATGATTATCATTGAACTGCGGCCG	636
428	SP101_SPET11_1688_1716_TMOD_F	TCCTATATTAATCGTTTACAGAAACTGGCT	282	SP101_SPET11_1783_1808_TMOD_R	TATATGATTATCATTGAACTGCGGCCG	637
89	SP101_SPET11_1711_1733_F	CTGGCTAAAACTTTGGCAACGGT	283	SP101_SPET11_1808_1835_R	GCGTGACGACCTTCTTGAATTGTAATCA	638
429	SP101_SPET11_1711_1733_TMOD_F	TCTGGCTAAAACTTTGGCAACGGT	284	SP101_SPET11_1808_1835_TMOD_R	TGCGTGACGACCTTCTTGAATTGTAATCA	639
90	SP101_SPET11_1807_1835_F	ATGATTACAATTCAAGAAGGTCGTCACGC	285	SP101_SPET11_1901_1927_R	TTGGACCTGTAATCAGCTGAATACTGG	640
430	SP101_SPET11_1807_1835_TMOD_F	TATGATTACAATTCAAGAAGGTCGTCACGC	286	SP101_SPET11_1901_1927_TMOD_R	TTTGGACCTGTAATCAGCTGAATACTGG	641
91	SP101_SPET11_1967_1991_F	TAACGGTTATCATGGCCCAGATGGG	287	SP101_SPET11_2062_2083_R	ATTGCCCAGAAATCAAATCATC	642
431	SP101_SPET11_1967_1991_TMOD_F	TTAACGGTTATCATGGCCCAGATGGG	288	SP101_SPET11_2062_2083_TMOD_R	TATTGCCCAGAAATCAAATCATC	643
77	SP101_SPET11_216_243_F	AGCAGGTGGTGAAATCGGCCACATGATT	289	SP101_SPET11_308_333_R	TGCCACTTTGACAACTCCTGTTGCTG	654
432	SP101_SPET11_216_243_TMOD_F	TAGCAGGTGGTGAAATCGGCCACATGATT	290	SP101_SPET11_308_333_TMOD_R	TTGCCACTTTGACAACTCCTGTTGCTG	655
92	SP101_SPET11_2260_2283_F	CAGAGACCGTTTTATCCTATCAGC	291	SP101_SPET11_2375_2397_R	TCTGGGTGACCTGGTGTTTTAGA	646
433	SP101_SPET11_2260_2283_TMOD_F	TCAGAGACCGTTTTATCCTATCAGC	292	SP101_SPET11_2375_2397_TMOD_R	TTCTGGGTGACCTGGTGTTTTAGA	647
93	SP101_SPET11_2375_2399_F	TCTAAAACACCAGGTCACCCAGAAG	293	SP101_SPET11_2470_2497_R	AGCTGCTAGATGAGCTTCTGCCATGGCC	648
434	SP101_SPET11_2375_2399_TMOD_F	TTCTAAAACACCAGGTCACCCAGAAG	294	SP101_SPET11_2470_2497_TMOD_R	TAGCTGCTAGATGAGCTTCTGCCATGGCC	649
94	SP101_SPET11_2468_2487_F	ATGGCCATGGCAGAAGCTCA	295	SP101_SPET11_2543_2570_R	CCATAAGGTCACCGTCACCATTCAAAGC	650
435	SP101_SPET11_2468_2487_TMOD_F	TATGGCCATGGCAGAAGCTCA	296	SP101_SPET11_2543_2570_TMOD_R	TCCATAAGGTCACCGTCACCATTCAAAGC	651
78	SP101_SPET11_266_295_F	CTTGTACTTGTGGCTCACACGGCTGTTTGG	297	SP101_SPET11_355_380_R	GCTGCTTTGATGGCTGAATCCCCTTC	661
436	SP101_SPET11_266_295_TMOD_F	TCTTGTACTTGTGGCTCACACGGCTGTTTGG	298	SP101_SPET11_355_380_TMOD_R	TGCTGCTTTGATGGCTGAATCCCCTTC	662
95	SP101_SPET11_2961_2984_F	ACCATGACAGAAGGCATTTTGACA	299	SP101_SPET11_3023_3045_R	GGAATTTACCAGCGATAGACACC	652
437	SP101_SPET11_2961_2984_TMOD_F	TACCATGACAGAAGGCATTTTGACA	300	SP101_SPET11_3023_3045_TMOD_R	TGGAATTTACCAGCGATAGACACC	653
96	SP101_SPET11_3075_3103_F	GATGACTTTTTAGCTAATGGTCAGGCAGC	301	SP101_SPET11_3168_3196_R	AATCGACGACCATCTTGGAAAGATTTCTC	656
438	SP101_SPET	TGATGACTTTTTAGCTA	302	SP101_SPET1	TAATCGACGACCATCTTGG	657

	11_3075_3103_TMOD_F	ATGGTCAGGCAGC		1_3168_3196_TMOD_R	AAAGATTTCTC	
448	SP101_SPET11_3085_3104_F	TAGCTAATGGTCAGGCAGCC	303	SP101_SPET11_3170_3194_R	TCGACGACCATCTTGGAAAGATTTC	658
79	SP101_SPET11_322_344_F	GTCAAAGTGGCACGTTTACTGGC	304	SP101_SPET11_423_441_R	ATCCCCTGCTTCTGCTGCC	665
439	SP101_SPET11_322_344_TMOD_F	TGTCAAAGTGGCACGTTTACTGGC	305	SP101_SPET11_423_441_TMOD_R	TATCCCCTGCTTCTGCTGCC	666
97	SP101_SPET11_3386_3403_F	AGCGTAAAGGTGAACCTT	306	SP101_SPET11_3480_3506_R	CCAGCAGTTACTGTCCCCTCATCTTTG	659
440	SP101_SPET11_3386_3403_TMOD_F	TAGCGTAAAGGTGAACCTT	307	SP101_SPET11_3480_3506_TMOD_R	TCCAGCAGTTACTGTCCCCTCATCTTTG	660
98	SP101_SPET11_3511_3535_F	GCTTCAGGAATCAATGATGGAGCAG	308	SP101_SPET11_3605_3629_R	GGGTCTACACCTGCACTTGCATAAC	663
441	SP101_SPET11_3511_3535_TMOD_F	TGCTTCAGGAATCAATGATGGAGCAG	309	SP101_SPET11_3605_3629_TMOD_R	TGGGTCTACACCTGCACTTGCATAAC	664
80	SP101_SPET11_358_387_F	GGGGATTCAGCCATCAAAGCAGCTATTGAC	310	SP101_SPET11_448_473_R	CCAACCTTTTCCACAACAGAATCAGC	668
442	SP101_SPET11_358_387_TMOD_F	TGGGGATTCAGCCATCAAAGCAGCTATTGAC	311	SP101_SPET11_448_473_TMOD_R	TCCAACCTTTTCCACAACAGAATCAGC	669
447	SP101_SPET11_364_385_F	TCAGCCATCAAAGCAGCTATTG	312	SP101_SPET11_448_471_R	TACCTTTTCCACAACAGAATCAGC	667
81	SP101_SPET11_600_629_F	CCTTACTTCGAACTATGAATCTTTTGGAAG	313	SP101_SPET11_686_714_R	CCCATTTTTTCACGCATGCTGAAAATATC	670
443	SP101_SPET11_600_629_TMOD_F	TCCTTACTTCGAACTATGAATCTTTTGGAAG	314	SP101_SPET11_686_714_TMOD_R	TCCCATTTTTTCACGCATGCTGAAAATATC	671
82	SP101_SPET11_658_684_F	GGGGATTGATATCACCGATAAGAAGAA	315	SP101_SPET11_756_784_R	GATTGGCGATAAAGTGATATTTTCTAAAA	672
444	SP101_SPET11_658_684_TMOD_F	TGGGGATTGATATCACCGATAAGAAGAA	316	SP101_SPET11_756_784_TMOD_R	TGATTGGCGATAAAGTGATATTTTCTAAAA	673
83	SP101_SPET11_776_801_F	TCGCCAATCAAAACTAAGGGAATGGC	317	SP101_SPET11_871_896_R	GCCCACCAGAAAGACTAGCAGGATAA	674
445	SP101_SPET11_776_801_TMOD_F	TTCGCCAATCAAAACTAAGGGAATGGC	318	SP101_SPET11_871_896_TMOD_R	TGCCCACCAGAAAGACTAGCAGGATAA	675
84	SP101_SPET11_893_921_F	GGGCAACAGCAGCGGATTGCGATTGCGCG	319	SP101_SPET11_988_1012_R	CATGACAGCCAAGACCTCACCCACC	678
423	SP101_SPET11_893_921_TMOD_F	TGGGCAACAGCAGCGGATTGCGATTGCGCG	320	SP101_SPET11_988_1012_TMOD_R	TCATGACAGCCAAGACCTCACCCACC	679
706	SSPE_BA_114_137_F	TCAAGCAAACGCACAATCAGAAGC	321	SSPE_BA_196_222_R	TTGCACGTCTGTTTCAGTTGCAAATTC	683
612	SSPE_BA_114_137P_F	TCAAGCAAACGCACAAC^{a}U^{a}AGAAGC	321	SSPE_BA_196_222P_R	TTGCACGTU^{a}C^{a}GTTTCAGTTGCAAATTC	684
58	SSPE_BA_115_137_F	CAAGCAAACGCACAATCAGAAGC	322	SSPE_BA_197_222_R	TGCACGTCTGTTTCAGTTGCAAATTC	686
355	SSPE_BA_115_137_TMOD_F	TCAAGCAAACGCACAATCAGAAGC	321	SSPE_BA_197_222_TMOD_R	TTGCACGTCTGTTTCAGTTGCAAATTC	687
215	SSPE_BA_121_137_F	AACGCACAATCAGAAGC	323	SSPE_BA_197_216_R	TCTGTTTCAGTTGCAAATTC	685
699	SSPE_BA_123_153_F	TGCACAATCAGAAGCTAAGAAAGCGCAAGCT	324	SSPE_BA_202_231_R	TTTCACAGCATGCACGTCTGTTTCAGTTGC	688
704	SSPE_BA_146_168_F	TGCAAGCTTCTGGTGCTAGCATT	325	SSPE_BA_242_267_R	TTGTGATTGTTTTGCAGCTGATTGTG	689
702	SSPE_BA_150_168_F	TGCTTCTGGTGCTAGCATT	326	SSPE_BA_243_264_R	TGATTGTTTTGCAGCTGATTGT	691
610	SSPE_BA_150_168P_F	TGCTTCTGGC^{a}GU^{a}C^{a}AGU^{a}ATT	326	SSPE_BA_243_264P_R	TGATTGTTTTGU^{a}AGU^{a}TGAC^{a}C^{a}GT	691

700	SSPE_BA_156_168_F	TGGTGCTAGCATT	327	SSPE_BA_243255_R	TGCAGCTGATTGT	690
608	SSPE_BA_156_168P_F	TGGC^{a}GU^{a}C^{a}AGU^{a}ATT	327	SSPE_BA_243255P_R	TGU^{a}AGU^{a}TGAC^{a}C^{a}GT	690
705	SSPE_BA_6389_F	TGCTAGTTATGGTACAGAGTTTGCGAC	328	SSPE_BA_163191_R	TCATAACTAGCATTTGTGCTTTGAATGCT	682
703	SSPE_BA_7289_F	TGGTACAGAGTTTGCGAC	329	SSPE_BA_163182_R	TCATTTGTGCTTTGAATGCT	681
611	SSPE_BA_7289P_F	TGGTAU^{a}AGAGC^{a}C^{a}C^{a}GU^{a}GAC	329	SSPE_BA_163182P_R	TCATTTGTGCC^{a}C^{a}C^{a}GAAC^{a}GU^{a}T	681
701	SSPE_BA_7589_F	TACAGAGTTTGCGAC	330	SSPE_BA_163177_R	TGTGCTTTGAATGCT	680
609	SSPE_BA_7589P_F	TAU^{a}AGAGC^{a}C^{a}C^{a}CGU^{a}GAC	330	SSPE_BA_163177P_R	TGTGCC^{a}C^{a}C^{a}GAAC^{a}GU^{a}T	680
1099	TOXR_VBC_135_158_F	TCGATTAGGCAGCAACGAAAGCCG	331	TOXR_VBC_221_246_R	TTCAAAACCTTGCTCTCGCCAAACAA	692
905	TRPE_AY094355_1064_1086_F	TCGACCTTTGGCAGGAACTAGAC	332	TRPE_AY094355_1171_1196_R	TACATCGTTTCGCCCAAGATCAATCA	693
904	TRPE_AY094355_1278_1303_F	TCAAATGTACAAGGTGAAGTGCGTGA	333	TRPE_AY094355_1392_1418_R	TCCTCTTTTCACAGGCTCTACTTCATC	694
903	TRPE_AY094355_1445_1471_F	TGGATGGCATGGTGAAATGGATATGTC	334	TRPE_AY094355_1551_1580_R	TATTTGGGTTTCATTCCACTCAGATTCTGG	695
902	TRPE_AY094355_1467_1491_F	ATGTCGATTGCAATCCGTACTTGTG	335	TRPE_AY094355_1569_1592_R	TGCGCGAGCTTTTATTTGGGTTTC	696
906	TRPE_AY094355_666_688_F	GTGCATGCGGATACAGAGCAGAG	336	TRPE_AY094355_769_791_R	TTCAAAATGCGGAGGCGTATGTG	697
907	TRPE_AY094355_757_776_F	TGCAAGCGCGACCACATACG	337	TRPE_AY094355_864_883_R	TGCCCAGGTACAACCTGCAT	698
114	TUFB_EC_225_251_F	GCACTATGCACACGTAGATTGTCCTGG	338	TUFB_EC_284309_R	TATAGCACCATCCATCTGAGCGGCAC	706
60	TUFB_EC_239_259_2_F	TTGACTGCCCAGGTCACGCTG	339	TUFB_EC_283303_2_R	GCCGTCCATTTGAGCAGCACC	704
59	TUFB_EC_239_259_F	TAGACTGCCCAGGACACGCTG	340	TUFB_EC_283303_R	GCCGTCCATCTGAGCAGCACC	705
942	TUFB_EC_251_278_F	TGCACGCCGACTATGTTAAGAACATGAT	341	TUFB_EC_337360_R	TATGTGCTCACGAGTTTGCGGCAT	707
941	TUFB_EC_275_299_F	TGATCACTGGTGCTGCTCAGATGGA	342	TUFB_EC_337362_R	TGGATGTGCTCACGAGTCTGTGGCAT	708
117	TUFB_EC_757_774_F	AAGACGACCTGCACGGGC	343	TUFB_EC_849867_R	GCGCTCCACGTCTTCACGC	709
293	TUFB_EC_957_979_F	CCACACGCCGTTCTTCAACAACT	344	TUFB_EC_1034_1058_R	GGCATCACCATTTCCTTGTCCTTCG	700
367	TUFB_EC_957_979_TMOD_F	TCCACACGCCGTTCTTCAACAACT	345	TUFB_EC_1034_1058_TMOD_R	TGGCATCACCATTTCCTTGTCCTTCG	701
62	TUFB_EC_976_1000_2_F	AACTACCGTCCTCAGTTCTACTTCC	346	TUFB_EC_1045_1068_2_R	GTTGTCACCAGGCATTACCATTTC	702
61	TUFB_EC_976_1000_F	AACTACCGTCCGCAGTTCTACTTCC	347	TUFB_EC_1045_1068_R	GTTGTCGCCAGGCATAACCATTTC	703
63	TUFB_EC_985_1012_F	CCACAGTTCTACTTCCGTACTACTGACG	348	TUFB_EC_1033_1062_R	TCCAGGCATTACCATTTCTACTCCTTCTGG	699
225	VALS_EC_1105_1124_F	CGTGGCGGCGTGGTTATCGA	349	VALS_EC_1195_1214_R	ACGAACTGGATGTCGCCGTT	710
71	VALS_EC_1105_1124_F	CGTGGCGGCGTGGTTATCGA	349	VALS_EC_1195_1218_R	CGGTACGAACTGGATGTCGCCGTT	711
358	VALS_EC_1105_1124_TMOD_F	TCGTGGCGGCGTGGTTATCGA	350	VALS_EC_1195_1218_TMOD_R	TCGGTACGAACTGGATGTCGCCGTT	712
965	VALS_EC_1128_1151_F	TATGCTGACCGACCAGTGGTACGT	351	VALS_EC_1231_1257_R	TTCGCGCATCCAGGAGAAGTACATGTT	713
112	VALS_EC_1833_1850_F	CGACGCGCTGCGCTTCAC	352	VALS_EC_1920_1943_R	GCGTTCCACAGCTTGTTGCAGAAG	714
116	VALS_EC_1920_1943_F	CTTCTGCAACAAGCTGTGGAACGC	353	VALS_EC_1948_1970_R	TCGCAGTTCATCAGCACGAAGCG	715
295	VALS_EC_610_649_F	ACCGAGCAAGGAGACCAGC	354	VALS_EC_705_727_R	TATAACGCACATCGTCAGGGTGA	716
931	WAAA_Z96925_2_29_F	TCTTGCTCTTTCGTGAGTTCAGTAAATG	355	WAAA_Z96925_115_138_R	CAAGCGGTTTGCCTCAAATAGTCA	717
932	WAAA_Z9692	TCGATCTGGTTTCATGC	356	WAAA_Z96925	TGGCACGAGCCTGACCTGT	718

	5_286_311_F	TGTTTCAGT		_394_412_R		

[0088] Primer pair name codes and reference sequences are shown in Table 2. The primer name code typically represents the gene to which the given primer pair is targeted. The primer pair name includes coordinates with respect to a reference sequence defined by an extraction of a section of sequence or defined by a GenBank gi number, or the corresponding complementary sequence of the extraction, or the entire GenBank gi number as indicated by the label "no extraction." Where "no extraction" is indicated for a reference sequence, the coordinates of a primer pair named to the reference sequence are with respect to the GenBank gi listing. Gene abbreviations are shown in bold type in the "Gene Name" column.

Table 2: Primer Name Codes and Reference Sequences

Primer name code	Gene Name	Organism	Reference GenBank gi number	Extracted gene coordinates of gi number	Extraction or entire gene SEQ ID NO:
16S_EC	**16S** rRNA (16S ribosomal RNA gene)	*Escherichia coli*	16127994	4033120..4034661	719
23S_EC	**23S** rRNA (23S ribosomal RNA gene)	*Escherichia coli*	16127994	4166220..4169123	720
CAPC_BA	**capC** (capsule biosynthesis gene)	*Bacillus anthracis*	6470151	Complement (55628..56074)	721
CYA_BA	**cya** (cyclic AMP gene)	*Bacillus anthracis*	4894216	Complement (154288..156626)	722
DNAK_EC	**dnaK** (chaperone dnaK gene)	*Escherichia coli*	16127994	12163..14079	723
GROL_EC	**groL** (chaperonin groL)	*Escherichia coli*	16127994	4368603..4370249	724
HFLB_EC	**hflb** (cell division protein peptidase ftsH)	*Escherichia coli*	16127994	Complement (3322645..3324576)	725
INFB_EC	**infB** (protein chain initiation factor infB gene)	*Escherichia coli*	16127994	Complement (3310983..3313655)	726
LEF_BA	**lef** (lethal factor)	*Bacillus anthracis*	21392688	Complement (149357..151786)	727
PAG_BA	**pag** (protective antigen)	*Bacillus anthracis*	21392688	143779..146073	728
RPLB_EC	**rplB** (50S ribosomal protein L2)	*Escherichia coli*	16127994	3449001..3448180	729
RPOB_EC	**rpoB** (DNA-directed RNA polymerase beta chain)	*Escherichia coli*	6127994	Complement 4178823..4182851	730
RPOC_EC	**rpoC** (DNA-directed RNA polymerase beta' chain)	*Escherichia coli*	16127994	4182928..4187151	731
SP101ET_SPET_11	Concatenation comprising: **gki** (glucose kinase) **gtr** (glutamine transporter protein) **murI** (glutamate racemase) **mutS** (DNA mismatch	Artificial Sequence* - partial gene sequences of *Streptococcus pyogenes*	15674250	Complement (1258294..1258791) complement (1236751..1237200) 312732..313169 Complement	732

	repair protein)			(1787602..1788007)	
	xpt (xanthine phosphoribosyl transferase)			930977..931425	
	yqiL (acetyl-CoA-acetyl transferase)			129471..129903	
	tkt (transketolase)			1391844..1391386	
SSPE_BA	**sspE** (small acid-soluble spore protein)	*Bacillus anthracis*	30253828	226496..226783	733
TUFB_EC	**tufB** (Elongation factor Tu)	*Escherichia coli*	16127994	4173523..4174707	734
VALS_EC	**valS** (Valyl-tRNA synthetase)	*Escherichia coli*	16127994	Complement (4481405..4478550)	735
ASPS_EC	**aspS** (Aspartyl-tRNA synthetase)	*Escherichia coli*	16127994	complement(1946777..1948546)	736
CAF1_AF053947	**caf1** (capsular protein caf1)	*Yersinia pestis*	2996286	No extraction - GenBank coordinates used	-
INV_U22457	**inv** (invasin)	*Yersinia pestis*	1256565	74..3772	737
LL_NC003143	*Y. pestis* specific chromosomal genes - difference region	*Yersinia pestis*	16120353	No extraction - GenBank coordinates used	-
BONTA_X52066	**BoNT/A** (neurotoxin type A)	*Clostridium botulinum*	40381	77..3967	738
MECA_Y14051	**mecA** methicillin resistance gene	*Staphylococcus aureus*	2791983	No extraction - GenBank coordinates used	739
TRPE_AY094355	**trpE** (anthranilate synthase (large component))	*Acinetobacter baumanii*	20853695	No extraction - GenBank coordinates used	740
RECA_AF251469	**recA** (recombinase A)	*Acinetobacter baumanii*	9965210	No extraction - GenBank coordinates used	741
GYRA_AF100557	**gyrA** (DNA gyrase subunit A)	*Acinetobacter baumanii*	4240540	No extraction - GenBank coordinates used	742
GYRB_AB008700	**gyrB** (DNA gyrase subunit B)	*Acinetobacter baumanii*	4514436	No extraction - GenBank coordinates used	743
WAAA_Z96925	**waaA** (3-deoxy-D-manno-octulosonic-acid transferase)	*Acinetobacter baumanii*	2765828	No extraction - GenBank coordinates used	744
CJST_CJ	Concatenation comprising: **tkt** (transketolase) **glyA** (serine hydroxymethyltransferase) **gltA** (citrate synthase) **aspA** (aspartate ammonia lyase) **glnA** (glutamine synthase) **pgm** (phosphoglycerate mutase)	Artificial Sequence* - partial gene sequences of *Campylobacter jejuni*	15791399	1569415..1569873 367573..368079 complement (1604529..1604930) 96692..97168 complement (657609..658085) 327773..328270	745

	uncA (ATP synthetase alpha chain)			112163..112651	
RNASEP_BDP	**RNase P** (ribonuclease P)	*Bordetella pertussis*	33591275	Complement (3226720..3227933)	746
RNASEP_BKM	**RNase P** (ribonuclease P)	*Burkholderia mallei*	53723370	Complement (2527296..2528220)	747
RNASEP_BS	**RNase P** (ribonuclease P)	*Bacillus subtilis*	16077068	Complement (2330250..2330962)	748
RNASEP_CLB	**RNase P** (ribonuclease P)	*Clostridium perfringens*	18308982	Complement (2291757..2292584)	749
RNASEP_EC	**RNase P** (ribonuclease P)	*Escherichia coli*	16127994	Complement (3267457..3268233	750
RNASEP_RKP	**RNase P** (ribonuclease P)	*Rickettsia prowazekii*	15603881	complement(605276..606109)	751
RNASEP_SA	**RNase P** (ribonuclease P)	*Staphylococcus aureus*	15922990	complement(1559869..1560651)	752
RNASEP_VBC	**RNase P** (ribonuclease P)	*Vibrio cholerae*	15640032	complement(2580367..2581452)	753
ICD_CXB	**icd** (isocitrate dehydrogenase)	*Coxiella burnetii*	29732244	complement(1143867..1144235)	754
IS1111A	multi-locus IS1111A insertion element	*Acinetobacter baumannii*	29732244	No extraction	-
OMPA_AY485227	**ompA** (outer membrane protein A)	*Rickettsia prowazekii*	40287451	No extraction	755
OMPB_RKP	**ompB** (outer membrane protein B)	*Rickettsia prowazekii*	15603881	complement(881264..886195)	756
GLTA_RKP	**gltA** (citrate synthase)	*Vibrio cholerae*	15603881	complement(1062547..1063857)	757
TOXR_VBC	**toxR** (transcription regulator toxR)	*Francisella tularensis*	15640032	complement(1047143..1048024)	758
ASD_FRT	**asd** (Aspartate semialdehyde dehydrogenase)	*Francisella tularensis*	56707187	complement(438608..439702)	759
GALE_FRT	**galE** (UDP-glucose 4-epimerase)	*Shigella flexneri*	56707187	809039..810058	760
IPAH_SGF	**ipaH** (invasion plasmid antigen)	*Campylobacter jejuni*	30061571	2210775..2211614	761
HUPB_CJ	**hupB** (DNA-binding protein Hu-beta)	*Coxiella burnetii*	15791399	complement(849317..849819)	762
AB_MLST	Concatenation comprising: **trpE** (anthranilate synthase component I)) **adk** (adenylate kinase) **mutY** (adenine glycosylase) **fumC** (fumarate hydratase) **efp** (elongation factor p) **ppa** (pyrophosphate phospho-hydratase	Artificial Sequence* - partial gene sequences of *Acinetobacter baumannii*	-	Sequenced in-house	763

[0089] * Note: These artificial reference sequences represent concatenations of partial gene extractions from the indicated reference gi number. Partial sequences were used to create the concatenated sequence because complete gene sequences were not necessary for primer design. The stretches of arbitrary residues "N"s were added for the convenience of separation of the partial gene extractions (100N for SP101_SPET11 (SEQ ID NO: 732); 50N for CJST_CJ (SEQ ID NO: 745); and 40N for AB_MLST (SEQ ID NO: 763)).

[0090] Example 2: DNA isolation and Amplification

[0091] Genomic materials from culture samples or swabs were prepared using the DNeasy® 96 Tissue Kit (Qiagen, Valencia, CA). All PCR reactions are assembled in 50 µl reactions in the 96 well microtiter plate format using a Packard MPII liquid handling robotic platform and MJ Dyad® thermocyclers (MJ research, Waltham, MA). The PCR reaction consisted of 4 units of Amplitaq Gold®, 1x buffer II (Applied Biosystems, Foster City, CA), 1.5 mM $MgCl_2$, 0.4 M betaine, 800 µM dNTP mix, and 250 nM of each primer.

[0092] The following PCR conditions were used to amplify the sequences used for mass spectrometry analysis: 95C for 10 minutes followed by 8 cycles of 95C for 30 seconds, 48C for 30 seconds, and 72C for 30 seconds, with the 48C annealing temperature increased 0.9C after each cycle. The PCR was then continued for 37 additional cycles of 95C for 15 seconds, 56C for 20 seconds, and 72C for 20 seconds.

[0093] Example 3: Solution Capture Purification of PCR Products for Mass Spectrometry with Ion Exchange Resin-Magnetic Beads

[0094] For solution capture of nucleic acids with ion exchange resin linked to magnetic beads, 25 µl of a 2.5 mg/mL suspension of BioClon amine terminated supraparamagnetic beads were added to 25 to 50 µl of a PCR reaction containing approximately 10 pM of a typical PCR amplification product. The above suspension was mixed for approximately 5 minutes by vortexing or pipetting, after which the liquid was removed after using a magnetic separator. The beads containing bound PCR amplification product were then washed 3x with 50mM ammonium bicarbonate/50% MeOH or 100mM ammonium bicarbonate/50% MeOH, followed by three more washes with 50% MeOH. The bound PCR amplicon was eluted with 25mM piperidine, 25mM imidazole, 35% MeOH, plus peptide calibration standards.

[0095] Example 4: Mass Spectrometry and Base Composition Analysis

[0096] The ESI-FTICR mass spectrometer is based on a Bruker Daltonics (Billerica, MA) Apex II 70e electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer that employs an actively shielded 7 Tesla superconducting magnet. The active shielding constrains the majority of the fringing magnetic field from the superconducting magnet to a relatively small volume. Thus, components that might be adversely affected by stray magnetic fields, such as CRT monitors, robotic components, and other electronics, can operate in close proximity to the FTICR spectrometer. All aspects of pulse sequence control and data acquisition were performed on a 600 MHz Pentium II data station running Bruker's Xmass software under Windows NT 4.0 operating system. Sample aliquots, typically 15 μl, were extracted directly from 96-well microtiter plates using a CTC HTS PAL autosampler (LEAP Technologies, Carrboro, NC) triggered by the FTICR data station. Samples were injected directly into a 10 μl sample loop integrated with a fluidics handling system that supplies the 100 μl /hr flow rate to the ESI source. Ions were formed via electrospray ionization in a modified Analytica (Branford, CT) source employing an off axis, grounded electrospray probe positioned approximately 1.5 cm from the metalized terminus of a glass desolvation capillary. The atmospheric pressure end of the glass capillary was biased at 6000 V relative to the ESI needle during data acquisition. A counter-current flow of dry N_2 was employed to assist in the desolvation process. Ions were accumulated in an external ion reservoir comprised of an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode, prior to injection into the trapped ion cell where they were mass analyzed. Ionization duty cycles > 99% were achieved by simultaneously accumulating ions in the external ion reservoir during ion detection. Each detection event consisted of 1M data points digitized over 2.3 s. To improve the signal-to-noise ratio (S/N), 32 scans were co-added for a total data acquisition time of 74 s.

[0097] The ESI-TOF mass spectrometer is based on a Bruker Daltonics MicroTOF™. Ions from the ESI source undergo orthogonal ion extraction and are focused in a reflectron prior to detection. The TOF and FTICR are equipped with the same automated sample handling and fluidics described above. Ions are formed in the standard MicroTOF™ ESI source that is equipped with the same off-axis sprayer and glass capillary as the FTICR ESI source. Consequently, source conditions were the same as those described above. External ion accumulation was also employed to improve ionization duty cycle during data acquisition. Each detection event on the TOF was comprised of 75,000 data points digitized over 75 μs.

[0098] The sample delivery scheme allows sample aliquots to be rapidly injected into the electrospray source at high flow rate and subsequently be electrosprayed at a much lower flow rate for improved ESI sensitivity. Prior to injecting a sample, a bolus of buffer was injected at a high flow rate to rinse the transfer line and spray needle to avoid sample contamination/carryover. Following the rinse step, the autosampler injected the next sample and the flow rate was switched to low flow. Following a brief equilibration delay, data acquisition commenced. As spectra were co-added, the autosampler continued rinsing the syringe and picking up buffer to rinse the injector and sample transfer line. In general, two syringe rinses and one injector rinse were required to minimize sample carryover. During a routine screening protocol a new sample mixture was injected every 106 seconds. More recently a fast wash station for the syringe needle has been implemented which, when combined with shorter acquisition times, facilitates the acquisition of mass spectra at a rate of just under one spectrum/minute.

[0099] Raw mass spectra were post-calibrated with an internal mass standard and deconvoluted to monoisotopic molecular masses. Unambiguous base compositions were derived from the exact mass measurements of the complementary single-stranded oligonucleotides. Quantitative results are obtained by comparing the peak heights with an internal PCR calibration standard present in every PCR well at 500 molecules per well for the ribosomal DNA-targeted primers and 100 molecules per well for the protein-encoding gene targets. Calibration methods are commonly owned and disclosed in U.S. Provisional Patent Application Serial No. 60/545,425.

[0100] Example 5: *De Novo* Determination of Base Composition of Amplification Products using Molecular Mass Modified Deoxynucleotide Triphosphates

[0101] Because the molecular masses of the four natural nucleobases have a relatively narrow molecular mass range (A = 313.058, G = 329.052, C = 289.046, T = 304.046 – See Table 3), a persistent source of ambiguity in assignment of base composition can occur as follows: two nucleic acid strands having different base composition may have a difference of about 1 Da when the base composition difference between the two strands is G ↔ A (-15.994) combined with C ↔ T (+15.000). For example, one 99-mer nucleic acid strand having a base composition of $A_{27}G_{30}C_{21}T_{21}$ has a theoretical molecular mass of 30779.058 while another 99-mer nucleic acid strand having a base composition of $A_{26}G_{31}C_{22}T_{20}$ has a theoretical molecular mass of 30780.052. A 1 Da difference in molecular mass may be within the experimental error of a

molecular mass measurement and thus, the relatively narrow molecular mass range of the four natural nucleobases imposes an uncertainty factor.

[0102] The present invention provides for a means for removing this theoretical 1 Da uncertainty factor through amplification of a nucleic acid with one mass-tagged nucleobase and three natural nucleobases. The term "nucleobase" as used herein is synonymous with other terms in use in the art including "nucleotide," "deoxynucleotide," "nucleotide residue," "deoxynucleotide residue," "nucleotide triphosphate (NTP)," or deoxynucleotide triphosphate (dNTP).

[0103] Addition of significant mass to one of the 4 nucleobases (dNTPs) in an amplification reaction, or in the primers themselves, will result in a significant difference in mass of the resulting amplification product (significantly greater than 1 Da) arising from ambiguities arising from the G ↔ A combined with C ↔ T event (Table 3). Thus, the same the G ↔ A (-15.994) event combined with 5-Iodo-C ↔ T (-110.900) event would result in a molecular mass difference of 126.894. If the molecular mass of the base composition $A_{27}G_{30}$ **5-Iodo-**$C_{21}T_{21}$ (33422.958) is compared with $A_{26}G_{31}$**5-Iodo-**$C_{22}T_{20}$, (33549.852) the theoretical molecular mass difference is +126.894. The experimental error of a molecular mass measurement is not significant with regard to this molecular mass difference. Furthermore, the only base composition consistent with a measured molecular mass of the 99-mer nucleic acid is $A_{27}G_{30}$**5-Iodo-**$C_{21}T_{21}$. In contrast, the analogous amplification without the mass tag has 18 possible base compositions.

Table 3: Molecular Masses of Natural Nucleobases and the Mass-Modified Nucleobase 5-Iodo-C and Molecular Mass Differences Resulting from Transitions

Nucleobase	Molecular Mass	Transition	Δ Molecular Mass
A	313.058	A-->T	-9.012
A	313.058	A-->C	-24.012
A	313.058	A-->5-Iodo-C	101.888
A	313.058	A-->G	15.994
T	304.046	T-->A	9.012
T	304.046	T-->C	-15.000
T	304.046	T-->5-Iodo-C	110.900
T	304.046	T-->G	25.006
C	289.046	C-->A	24.012
C	289.046	C-->T	15.000

C	289.046	C-->G	40.006
5-Iodo-C	414.946	5-Iodo-C-->A	-101.888
5-Iodo-C	414.946	5-Iodo-C-->T	-110.900
5-Iodo-C	414.946	5-Iodo-C-->G	-85.894
G	329.052	G-->A	-15.994
G	329.052	G-->T	-25.006
G	329.052	G-->C	-40.006
G	329.052	G-->5-Iodo-C	85.894

[0104] Example 6: Data Processing

[0105] Mass spectra of bioagent identifying amplicons are analyzed independently using a maximum-likelihood processor, such as is widely used in radar signal processing. This processor, referred to as GenX, first makes maximum likelihood estimates of the input to the mass spectrometer for each primer by running matched filters for each base composition aggregate on the input data. This includes the GenX response to a calibrant for each primer.

[0106] The algorithm emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of *a priori* expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database is used to define the mass base count matched filters. The database contains the sequences of known bacterial bioagents and includes threat organisms as well as benign background organisms. The latter is used to estimate and subtract the spectral signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a running-sum estimate of the noise covariance. Background signal strengths are estimated and used along with the matched filters to form signatures which are then subtracted. the maximum likelihood process is applied to this "cleaned up" data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

[0107] The amplitudes of all base compositions of bioagent identifying amplicons for each primer are calibrated and a final maximum likelihood amplitude estimate per organism is made based upon the multiple single primer estimates. Models of all system noise are factored into this two-stage maximum likelihood calculation. The processor reports the number of molecules of each base composition contained in the spectra. The quantity of amplification product

corresponding to the appropriate primer set is reported as well as the quantities of primers remaining upon completion of the amplification reaction.

[0108] Example 7: Use of Broad Range Survey and Division Wide Primer Pairs for Identification of Bacteria in an Epidemic Surveillance Investigation

[0109] This investigation employed a set of 16 primer pairs which is herein designated the "surveillance primer set" and comprises broad range survey primer pairs, division wide primer pairs and a single *Bacillus* clade primer pair. The surveillance primer set is shown in Table 4 and consists of primer pairs originally listed in Table 1. This surveillance set comprises primers with T modifications (note TMOD designation in primer names) which constitutes a functional improvement with regard to prevention of non-templated adenylation (*vide supra*) relative to originally selected primers which are displayed below in the same row. Primer pair 449 (non-T modified) has been modified twice. Its predecessors are primer pairs 70 and 357, displayed below in the same row. Primer pair 360 has also been modified twice and its predecessors are primer pairs 17 and 118.

Table 4: Bacterial Primer Pairs of the Surveillance Primer Set

Primer Pair No.	Forward Primer Name	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer (SEQ ID NO:)	Target Gene
346	16S_EC_713_732_TMOD_F	27	16S_EC_789_809_TMOD_R	389	16S rRNA
10	16S_EC_713_732_F	26	16S_EC_789_809	388	16S rRNA
347	16S_EC_785_806_TMOD_F	30	16S_EC_880_897_TMOD_R	392	16S rRNA
11	16S_EC_785_806_F	29	16S_EC_880_897_R	391	16S rRNA
348	16S_EC_960_981_TMOD_F	38	16S_EC_1054_1073_TMOD_R	363	16S rRNA
14	16S_EC_960_981_F	37	16S_EC_1054_1073_R	362	16S rRNA
349	23S_EC_1826_1843_TMOD_F	49	23S_EC_1906_1924_TMOD_R	405	23S rRNA
16	23S_EC_1826_1843_F	48	23S_EC_1906_1924_R	404	23S rRNA
352	INFB_EC_1365_1393_TMOD_F	161	INFB_EC_1439_1467_TMOD_R	516	infB
34	INFB_EC_1365_1393_F	160	INFB_EC_1439_1467_R	515	infB
354	RPOC_EC_2218_2241_TMOD_F	262	RPOC_EC_2313_2337_TMOD_R	625	rpoC
52	RPOC_EC_2218_2241_F	261	RPOC_EC_2313_2337_R	624	rpoC
355	SSPE_BA_115_137_TMOD_F	321	SSPE_BA_197_222_TMOD_R	687	sspE
58	SSPE_BA_115_137_F	322	SSPE_BA_197_222_R	686	sspE
356	RPLB_EC_650_679_TMOD_F	232	RPLB_EC_739_762_TMOD_R	592	rplB
66	RPLB_EC_650_679_F	231	RPLB_EC_739_762_R	591	rplB
358	VALS_EC_1105_1124_TMOD_F	350	VALS_EC_1195_1218_TMOD_R	712	valS
71	VALS_EC_1105_1124_F	349	VALS_EC_1195_1218_R	711	valS
359	RPOB_EC_1845_1866_TMOD_F	241	RPOB_EC_1909_1929_TMOD_R	597	rpoB
72	RPOB_EC_1845_1866_F	240	RPOB_EC_1909_1929_R	596	rpoB
360	23S_EC_2646_2667_TMOD_F	60	23S_EC_2745_2765_TMOD_R	416	23S rRNA
118	23S_EC_2646_2667_F	59	23S_EC_2745_2765_R	415	23S rRNA
17	23S_EC_2645_2669_F	58	23S_EC_2744_2761_R	414	23S rRNA

361	16S_EC_1090_1111_2_TMOD_F	5	16S_EC_1175_1196_TMOD_R	370	16S rRNA
3	16S_EC_1090_1111_2_F	6	16S_EC_1175_1196_R	369	16S rRNA
362	RPOB_EC_3799_3821_TMOD_F	245	RPOB_EC_3862_3888_TMOD_R	603	rpoB
289	RPOB_EC_3799_3821_F	246	RPOB_EC_3862_3888_R	602	rpoB
363	RPOC_EC_2146_2174_TMOD_F	257	RPOC_EC_2227_2245_TMOD_R	621	rpoC
290	RPOC_EC_2146_2174_F	256	RPOC_EC_2227_2245_R	620	rpoC
367	TUFB_EC_957_979_TMOD_F	345	TUFB_EC_1034_1058_TMOD_R	701	tufB
293	TUFB_EC_957_979_F	344	TUFB_EC_1034_1058_R	700	tufB
449	RPLB_EC_690_710_F	237	RPLB_EC_737_758_R	589	rplB
357	RPLB_EC_688_710_TMOD_F	236	RPLB_EC_736_757_TMOD_R	588	rplB
67	RPLB_EC_688_710_F	235	RPLB_EC_736_757_R	587	rplB

[0110] The 16 primer pairs of the surveillance set are used to produce bioagent identifying amplicons whose base compositions are sufficiently different amongst all known bacteria at the species level to identify, at a reasonable confidence level, any given bacterium at the species level. As shown in Tables 6A-E, common respiratory bacterial pathogens can be distinguished by the base compositions of bioagent identifying amplicons obtained using the 16 primer pairs of the surveillance set. In some cases, triangulation identification improves the confidence level for species assignment. For example, nucleic acid from *Streptococcus pyogenes* can be amplified by nine of the sixteen surveillance primer pairs and *Streptococcus pneumoniae* can be amplified by ten of the sixteen surveillance primer pairs. The base compositions of the bioagent identifying amplicons are identical for only one of the analogous bioagent identifying amplicons and differ in all of the remaining analogous bioagent identifying amplicons by up to four bases per bioagent identifying amplicon. The resolving power of the surveillance set was confirmed by determination of base compositions for 120 isolates of respiratory pathogens representing 70 different bacterial species and the results indicated that natural variations (usually only one or two base substitutions per bioagent identifying amplicon) amongst multiple isolates of the same species did not prevent correct identification of major pathogenic organisms at the species level.

[0111] *Bacillus anthracis* is a well known biological warfare agent which has emerged in domestic terrorism in recent years. Since it was envisioned to produce bioagent identifying amplicons for identification of *Bacillus anthracis*, additional drill-down analysis primers were designed to target genes present on virulence plasmids of *Bacillus anthracis* so that additional confidence could be reached in positive identification of this pathogenic organism. Three drill-down analysis primers were designed and are listed in Tables 1 and 5. In Table 5 the drill-down set comprises primers with T modifications (note TMOD designation in primer names) which

constitutes a functional improvement with regard to prevention of non-templated adenylation (*vide supra*) relative to originally selected primers which are displayed below in the same row.

Table 5: Drill-Down Primer Pairs for Confirmation of Identification of *Bacillus anthracis*

Primer Pair No.	Forward Primer Name	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer (SEQ ID NO:)	Target Gene
350	CAPC_BA_274_303_TMOD_F	98	CAPC_BA_349_376_TMOD_R	452	capC
24	CAPC_BA_274_303_F	97	CAPC_BA_349_376_R	451	capC
351	CYA_BA_1353_1379_TMOD_F	128	CYA_BA_1448_1467_TMOD_R	483	cyA
30	CYA_BA_1353_1379_F	127	CYA_BA_1448_1467_R	482	cyA
353	LEF_BA_756_781_TMOD_F	175	LEF_BA_843_872_TMOD_R	531	lef
37	LEF_BA_756_781_F	174	LEF_BA_843_872_R	530	lef

[0112] Phylogenetic coverage of bacterial space of the sixteen surveillance primers of Table 4 and the three *Bacillus* anthracis drill-down primers of Table 5 is shown in Figure 3 which lists common pathogenic bacteria. Figure 3 is not meant to be comprehensive in illustrating all species identified by the primers. Only pathogenic bacteria are listed as representative examples of the bacterial species that can be identified by the primers and methods of the present invention. Nucleic acid of groups of bacteria enclosed within the polygons of Figure 3 can be amplified to obtain bioagent identifying amplicons using the primer pair numbers listed in the upper right hand corner of each polygon. Primer coverage for polygons within polygons is additive. As an illustrative example, bioagent identifying amplicons can be obtained for *Chlamydia trachomatis* by amplification with, for example, primer pairs 346-349, 360 and 361, but not with any of the remaining primers of the surveillance primer set. On the other hand, bioagent identifying amplicons can be obtained from nucleic acid originating from *Bacillus anthracis* (located within 5 successive polygons) using, for example, any of the following primer pairs: 346-349, 360, 361 (base polygon), 356, 449 (second polygon), 352 (third polygon), 355 (fourth polygon), 350, 351 and 353 (fifth polygon). Multiple coverage of a given organism with multiple primers provides for increased confidence level in identification of the organism as a result of enabling broad triangulation identification.

[0113] In Tables 6A-E, base compositions of respiratory pathogens for primer target regions are shown. Two entries in a cell, represent variation in ribosomal DNA operons. The most predominant base composition is shown first and the minor (frequently a single operon) is indicated by an asterisk (*). Entries with NO DATA mean that the primer would not be expected to prime this species due to mismatches between the primer and target region, as determined by theoretical PCR.

Table 6A – Base Compositions of Common Respiratory Pathogens for Bioagent Identifying Amplicons Corresponding to Primer Pair Nos: 346, 347 and 348

Organism	Strain	Primer 346 [A G C T]	Primer 347 [A G C T]	Primer 348 [A G C T]
Klebsiella pneumoniae	MGH78578	[29 32 25 13] [29 31 25 13]*	[23 38 28 26] [23 37 28 26]*	[26 32 28 30] [26 31 28 30]*
Yersinia pestis	CO-92 Biovar Orientalis	[29 32 25 13]	[22 39 28 26]	[29 30 28 29] [30 30 27 29]*
Yersinia pestis	KIM5 P12 (Biovar Mediaevalis)	[29 32 25 13]	[22 39 28 26]	[29 30 28 29]
Yersinia pestis	91001	[29 32 25 13]	[22 39 28 26]	[29 30 28 29] [30 30 27 29]*
Haemophilus influenzae	KW20	[28 31 23 17]	[24 37 25 27]	[29 30 28 29]
Pseudomonas aeruginosa	PAO1	[30 31 23 15]	[26 36 29 24] [27 36 29 23]*	[26 32 29 29]
Pseudomonas fluorescens	Pf0-1	[30 31 23 15]	[26 35 29 25]	[28 31 28 29]
Pseudomonas putida	KT2440	[30 31 23 15]	[28 33 27 27]	[27 32 29 28]
Legionella pneumophila	Philadelphia-1	[30 30 24 15]	[33 33 23 27]	[29 28 28 31]
Francisella tularensis	schu 4	[32 29 22 16]	[28 38 26 26]	[25 32 28 31]
Bordetella pertussis	Tohama I	[30 29 24 16]	[23 37 30 24]	[30 32 30 26]
Burkholderia cepacia	J2315	[29 29 27 14]	[27 32 26 29]	[27 36 31 24] [20 42 35 19]*
Burkholderia pseudomallei	K96243	[29 29 27 14]	[27 32 26 29]	[27 36 31 24]
Neisseria gonorrhoeae	FA 1090, ATCC 700825	[29 28 24 18]	[27 34 26 28]	[24 36 29 27]
Neisseria meningitides	MC58 (serogroup B)	[29 28 26 16]	[27 34 27 27]	[25 35 30 26]
Neisseria meningitides	serogroup C, FAM18	[29 28 26 16]	[27 34 27 27]	[25 35 30 26]
Neisseria meningitides	Z2491 (serogroup A)	[29 28 26 16]	[27 34 27 27]	[25 35 30 26]
Chlamydophila pneumoniae	TW-183	[31 27 22 19]	NO DATA	[32 27 27 29]
Chlamydophila pneumoniae	AR39	[31 27 22 19]	NO DATA	[32 27 27 29]
Chlamydophila pneumoniae	CWL029	[31 27 22 19]	NO DATA	[32 27 27 29]
Chlamydophila pneumoniae	J138	[31 27 22 19]	NO DATA	[32 27 27 29]
Corynebacterium diphtheriae	NCTC13129	[29 34 21 15]	[22 38 31 25]	[22 33 25 34]
Mycobacterium avium	k10	[27 36 21 15]	[22 37 30 28]	[21 36 27 30]
Mycobacterium avium	104	[27 36 21 15]	[22 37 30 28]	[21 36 27 30]
Mycobacterium tuberculosis	CSU#93	[27 36 21 15]	[22 37 30 28]	[21 36 27 30]
Mycobacterium tuberculosis	CDC 1551	[27 36 21 15]	[22 37 30 28]	[21 36 27 30]
Mycobacterium tuberculosis	H37Rv (lab strain)	[27 36 21 15]	[22 37 30 28]	[21 36 27 30]
Mycoplasma pneumoniae	M129	[31 29 19 20]	NO DATA	NO DATA
Staphylococcus aureus	MRSA252	[27 30 21 21]	[25 35 30 26]	[30 29 30 29] [29 31 30 29]*
Staphylococcus aureus	MSSA476	[27 30 21 21]	[25 35 30 26]	[30 29 30 29] [30 29 29 30]*
Staphylococcus aureus	COL	[27 30 21 21]	[25 35 30 26]	[30 29 30 29] [30 29 29 30]*
Staphylococcus aureus	Mu50	[27 30 21 21]	[25 35 30 26]	[30 29 30 29] [30 29 29 30]*
Staphylococcus aureus	MW2	[27 30 21 21]	[25 35 30 26]	[30 29 30 29] [30 29 29 30]*

Staphylococcus aureus	N315	[27 30 21 21]	[25 35 30 26]	[30 29 30 29] [30 29 29 30]*
Staphylococcus aureus	NCTC 8325	[27 30 21 21]	[25 35 30 26] [25 35 31 26]*	[30 29 30 29] [30 29 29 30]
Streptococcus agalactiae	NEM316	[26 32 23 18]	[24 36 31 25] [24 36 30 26]*	[25 32 29 30]
Streptococcus equi	NC_002955	[26 32 23 18]	[23 37 31 25]	[29 30 25 32]
Streptococcus pyogenes	MGAS8232	[26 32 23 18]	[24 37 30 25]	[25 31 29 31]
Streptococcus pyogenes	MGAS315	[26 32 23 18]	[24 37 30 25]	[25 31 29 31]
Streptococcus pyogenes	SSI-1	[26 32 23 18]	[24 37 30 25]	[25 31 29 31]
Streptococcus pyogenes	MGAS10394	[26 32 23 18]	[24 37 30 25]	[25 31 29 31]
Streptococcus pyogenes	Manfredo (M5)	[26 32 23 18]	[24 37 30 25]	[25 31 29 31]
Streptococcus pyogenes	SF370 (M1)	[26 32 23 18]	[24 37 30 25]	[25 31 29 31]
Streptococcus pneumoniae	670	[26 32 23 18]	[25 35 28 28]	[25 32 29 30]
Streptococcus pneumoniae	R6	[26 32 23 18]	[25 35 28 28]	[25 32 29 30]
Streptococcus pneumoniae	TIGR4	[26 32 23 18]	[25 35 28 28]	[25 32 30 29]
Streptococcus gordonii	NCTC7868	[25 33 23 18]	[24 36 31 25]	[25 31 29 31]
Streptococcus mitis	NCTC 12261	[26 32 23 18]	[25 35 30 26]	[25 32 29 30] [24 31 35 29]*
Streptococcus mutans	UA159	[24 32 24 19]	[25 37 30 24]	[28 31 26 31]

Table 6B – Base Compositions of Common Respiratory Pathogens for Bioagent Identifying Amplicons Corresponding to Primer Pair Nos: 349, 360, and 356

Organism	Strain	Primer 349 [A G C T]	Primer 360 [A G C T]	Primer 356 [A G C T]
Klebsiella pneumoniae	MGH78578	[25 31 25 22]	[33 37 25 27]	NO DATA
Yersinia pestis	CO-92 Biovar Orientalis	[25 31 27 20] [25 32 26 20]*	[34 35 25 28]	NO DATA
Yersinia pestis	KIM5 P12 (Biovar Mediaevalis)	[25 31 27 20] [25 32 26 20]*	[34 35 25 28]	NO DATA
Yersinia pestis	91001	[25 31 27 20]	[34 35 25 28]	NO DATA
Haemophilus influenzae	KW20	[28 28 25 20]	[32 38 25 27]	NO DATA
Pseudomonas aeruginosa	PAO1	[24 31 26 20]	[31 36 27 27] [31 36 27 28]*	NO DATA
Pseudomonas fluorescens	Pf0-1	NO DATA	[30 37 27 28] [30 37 27 28]	NO DATA
Pseudomonas putida	KT2440	[24 31 26 20]	[30 37 27 28]	NO DATA
Legionella pneumophila	Philadelphia-1	[23 30 25 23]	[30 39 29 24]	NO DATA
Francisella tularensis	schu 4	[26 31 25 19]	[32 36 27 27]	NO DATA
Bordetella pertussis	Tohama I	[21 29 24 18]	[33 36 26 27]	NO DATA
Burkholderia cepacia	J2315	[23 27 22 20]	[31 37 28 26]	NO DATA
Burkholderia pseudomallei	K96243	[23 27 22 20]	[31 37 28 26]	NO DATA
Neisseria gonorrhoeae	FA 1090, ATCC 700825	[24 27 24 17]	[34 37 25 26]	NO DATA
Neisseria meningitidis	MC58 (serogroup B)	[25 27 22 18]	[34 37 25 26]	NO DATA
Neisseria meningitidis	serogroup C, FAM18	[25 26 23 18]	[34 37 25 26]	NO DATA
Neisseria	Z2491 (serogroup A)	[25 26 23 18]	[34 37 25 26]	NO DATA

meningitidis				
Chlamydophila pneumoniae	TW-183	[30 28 27 18]	NO DATA	NO DATA
Chlamydophila pneumoniae	AR39	[30 28 27 18]	NO DATA	NO DATA
Chlamydophila pneumoniae	CWL029	[30 28 27 18]	NO DATA	NO DATA
Chlamydophila pneumoniae	J138	[30 28 27 18]	NO DATA	NO DATA
Corynebacterium diphtheriae	NCTC13129	NO DATA	[29 40 28 25]	NO DATA
Mycobacterium avium	k10	NO DATA	[33 35 32 22]	NO DATA
Mycobacterium avium	104	NO DATA	[33 35 32 22]	NO DATA
Mycobacterium tuberculosis	CSU#93	NO DATA	[30 36 34 22]	NO DATA
Mycobacterium tuberculosis	CDC 1551	NO DATA	[30 36 34 22]	NO DATA
Mycobacterium tuberculosis	H37Rv (lab strain)	NO DATA	[30 36 34 22]	NO DATA
Mycoplasma pneumoniae	M129	[28 30 24 19]	[34 31 29 28]	NO DATA
Staphylococcus aureus	MRSA252	[26 30 25 20]	[31 38 24 29]	[33 30 31 27]
Staphylococcus aureus	MSSA476	[26 30 25 20]	[31 38 24 29]	[33 30 31 27]
Staphylococcus aureus	COL	[26 30 25 20]	[31 38 24 29]	[33 30 31 27]
Staphylococcus aureus	Mu50	[26 30 25 20]	[31 38 24 29]	[33 30 31 27]
Staphylococcus aureus	MW2	[26 30 25 20]	[31 38 24 29]	[33 30 31 27]
Staphylococcus aureus	N315	[26 30 25 20]	[31 38 24 29]	[33 30 31 27]
Staphylococcus aureus	NCTC 8325	[26 30 25 20]	[31 38 24 29]	[33 30 31 27]
Streptococcus agalactiae	NEM316	[28 31 22 20]	[33 37 24 28]	[37 30 28 26]
Streptococcus equi	NC_002955	[28 31 23 19]	[33 38 24 27]	[37 31 28 25]
Streptococcus pyogenes	MGAS8232	[28 31 23 19]	[33 37 24 28]	[38 31 29 23]
Streptococcus pyogenes	MGAS315	[28 31 23 19]	[33 37 24 28]	[38 31 29 23]
Streptococcus pyogenes	SSI-1	[28 31 23 19]	[33 37 24 28]	[38 31 29 23]
Streptococcus pyogenes	MGAS10394	[28 31 23 19]	[33 37 24 28]	[38 31 29 23]
Streptococcus pyogenes	Manfredo (M5)	[28 31 23 19]	[33 37 24 28]	[38 31 29 23]
Streptococcus pyogenes	SF370 (M1)	[28 31 23 19] [28 31 22 20]*	[33 37 24 28]	[38 31 29 23]
Streptococcus pneumoniae	670	[28 31 22 20]	[34 36 24 28]	[37 30 29 25]
Streptococcus pneumoniae	R6	[28 31 22 20]	[34 36 24 28]	[37 30 29 25]
Streptococcus pneumoniae	TIGR4	[28 31 22 20]	[34 36 24 28]	[37 30 29 25]
Streptococcus gordonii	NCTC7868	[28 32 23 20]	[34 36 24 28]	[36 31 29 25]
Streptococcus mitis	NCTC 12261	[28 31 22 20] [29 30 22 20]*	[34 36 24 28]	[37 30 29 25]
Streptococcus mutans	UA159	[26 32 23 22]	[34 37 24 27]	NO DATA

Table 6C – Base Compositions of Common Respiratory Pathogens for Bioagent Identifying Amplicons Corresponding to Primer Pair Nos: 449, 354, and 352

Organism	Strain	Primer 449 [A G C T]	Primer 354 [A G C T]	Primer 352 [A G C T]
Klebsiella pneumoniae	MGH78578	NO DATA	[27 33 36 26]	NO DATA
Yersinia pestis	CO-92 Biovar Orientalis	NO DATA	[29 31 33 29]	[32 28 20 25]
Yersinia pestis	KIM5 P12 (Biovar Mediaevalis)	NO DATA	[29 31 33 29]	[32 28 20 25]
Yersinia pestis	91001	NO DATA	[29 31 33 29]	NO DATA
Haemophilus influenzae	KW20	NO DATA	[30 29 31 32]	NO DATA
Pseudomonas aeruginosa	PAO1	NO DATA	[26 33 39 24]	NO DATA
Pseudomonas fluorescens	Pf0-1	NO DATA	[26 33 34 29]	NO DATA
Pseudomonas putida	KT2440	NO DATA	[25 34 36 27]	NO DATA
Legionella pneumophila	Philadelphia-1	NO DATA	NO DATA	NO DATA
Francisella tularensis	schu 4	NO DATA	[33 32 25 32]	NO DATA
Bordetella pertussis	Tohama I	NO DATA	[26 33 39 24]	NO DATA
Burkholderia cepacia	J2315	NO DATA	[25 37 33 27]	NO DATA
Burkholderia pseudomallei	K96243	NO DATA	[25 37 34 26]	NO DATA
Neisseria gonorrhoeae	FA 1090, ATCC 700825	[17 23 22 10]	[29 31 32 30]	NO DATA
Neisseria meningitidis	MC58 (serogroup B)	NO DATA	[29 30 32 31]	NO DATA
Neisseria meningitidis	serogroup C, FAM18	NO DATA	[29 30 32 31]	NO DATA
Neisseria meningitidis	Z2491 (serogroup A)	NO DATA	[29 30 32 31]	NO DATA
Chlamydophila pneumoniae	TW-183	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	AR39	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	CWL029	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	J138	NO DATA	NO DATA	NO DATA
Corynebacterium diphtheriae	NCTC13129	NO DATA	NO DATA	NO DATA
Mycobacterium avium	k10	NO DATA	NO DATA	NO DATA
Mycobacterium avium	104	NO DATA	NO DATA	NO DATA
Mycobacterium tuberculosis	CSU#93	NO DATA	NO DATA	NO DATA
Mycobacterium tuberculosis	CDC 1551	NO DATA	NO DATA	NO DATA
Mycobacterium tuberculosis	H37Rv (lab strain)	NO DATA	NO DATA	NO DATA
Mycoplasma pneumoniae	M129	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	MRSA252	[17 20 21 17]	[30 27 30 35]	[36 24 19 26]
Staphylococcus aureus	MSSA476	[17 20 21 17]	[30 27 30 35]	[36 24 19 26]
Staphylococcus aureus	COL	[17 20 21 17]	[30 27 30 35]	[35 24 19 27]
Staphylococcus aureus	Mu50	[17 20 21 17]	[30 27 30 35]	[36 24 19 26]
Staphylococcus aureus	MW2	[17 20 21 17]	[30 27 30 35]	[36 24 19 26]

Staphylococcus aureus	N315	[17 20 21 17]	[30 27 30 35]	[36 24 19 26]
Staphylococcus aureus	NCTC 8325	[17 20 21 17]	[30 27 30 35]	[35 24 19 27]
Streptococcus agalactiae	NEM316	[22 20 19 14]	[26 31 27 38]	[29 26 22 28]
Streptococcus equi	NC_002955	[22 21 19 13]	NO DATA	NO DATA
Streptococcus pyogenes	MGAS8232	[23 21 19 12]	[24 32 30 36]	NO DATA
Streptococcus pyogenes	MGAS315	[23 21 19 12]	[24 32 30 36]	NO DATA
Streptococcus pyogenes	SSI-1	[23 21 19 12]	[24 32 30 36]	NO DATA
Streptococcus pyogenes	MGAS10394	[23 21 19 12]	[24 32 30 36]	NO DATA
Streptococcus pyogenes	Manfredo (M5)	[23 21 19 12]	[24 32 30 36]	NO DATA
Streptococcus pyogenes	SF370 (M1)	[23 21 19 12]	[24 32 30 36]	NO DATA
Streptococcus pneumoniae	670	[22 20 19 14]	[25 33 29 35]	[30 29 21 25]
Streptococcus pneumoniae	R6	[22 20 19 14]	[25 33 29 35]	[30 29 21 25]
Streptococcus pneumoniae	TIGR4	[22 20 19 14]	[25 33 29 35]	[30 29 21 25]
Streptococcus gordonii	NCTC7868	[21 21 19 14]	NO DATA	[29 26 22 28]
Streptococcus mitis	NCTC 12261	[22 20 19 14]	[26 30 32 34]	NO DATA
Streptococcus mutans	UA159	NO DATA	NO DATA	NO DATA

Table 6D – Base Compositions of Common Respiratory Pathogens for Bioagent Identifying Amplicons Corresponding to Primer Pair Nos: 355, 358, and 359

Organism	Strain	Primer 355 [A G C T]	Primer 358 [A G C T]	Primer 359 [A G C T]
Klebsiella pneumoniae	MGH78578	NO DATA	[24 39 33 20]	[25 21 24 17]
Yersinia pestis	CO-92 Biovar Orientalis	NO DATA	[26 34 35 21]	[23 23 19 22]
Yersinia pestis	KIM5 P12 (Biovar Mediaevalis)	NO DATA	[26 34 35 21]	[23 23 19 22]
Yersinia pestis	91001	NO DATA	[26 34 35 21]	[23 23 19 22]
Haemophilus influenzae	KW20	NO DATA	NO DATA	NO DATA
Pseudomonas aeruginosa	PAO1	NO DATA	NO DATA	NO DATA
Pseudomonas fluorescens	Pf0-1	NO DATA	NO DATA	NO DATA
Pseudomonas putida	KT2440	NO DATA	[21 37 37 21]	NO DATA
Legionella pneumophila	Philadelphia-1	NO DATA	NO DATA	NO DATA
Francisella tularensis	schu 4	NO DATA	NO DATA	NO DATA
Bordetella pertussis	Tohama I	NO DATA	NO DATA	NO DATA
Burkholderia cepacia	J2315	NO DATA	NO DATA	NO DATA
Burkholderia pseudomallei	K96243	NO DATA	NO DATA	NO DATA
Neisseria gonorrhoeae	FA 1090, ATCC 700825	NO DATA	NO DATA	NO DATA
Neisseria meningitidis	MC58 (serogroup B)	NO DATA	NO DATA	NO DATA
Neisseria meningitidis	serogroup C, FAM18	NO DATA	NO DATA	NO DATA

Neisseria meningitidis	Z2491 (serogroup A)	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	TW-183	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	AR39	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	CWL029	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	J138	NO DATA	NO DATA	NO DATA
Corynebacterium diphtheriae	NCTC13129	NO DATA	NO DATA	NO DATA
Mycobacterium avium	k10	NO DATA	NO DATA	NO DATA
Mycobacterium avium	104	NO DATA	NO DATA	NO DATA
Mycobacterium tuberculosis	CSU#93	NO DATA	NO DATA	NO DATA
Mycobacterium tuberculosis	CDC 1551	NO DATA	NO DATA	NO DATA
Mycobacterium tuberculosis	H37Rv (lab strain)	NO DATA	NO DATA	NO DATA
Mycoplasma pneumoniae	M129	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	MRSA252	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	MSSA476	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	COL	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	Mu50	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	MW2	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	N315	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	NCTC 8325	NO DATA	NO DATA	NO DATA
Streptococcus agalactiae	NEM316	NO DATA	NO DATA	NO DATA
Streptococcus equi	NC 002955	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	MGAS8232	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	MGAS315	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	SSI-1	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	MGAS10394	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	Manfredo (M5)	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	SF370 (M1)	NO DATA	NO DATA	NO DATA
Streptococcus pneumoniae	670	NO DATA	NO DATA	NO DATA
Streptococcus pneumoniae	R6	NO DATA	NO DATA	NO DATA
Streptococcus pneumoniae	TIGR4	NO DATA	NO DATA	NO DATA
Streptococcus gordonii	NCTC7868	NO DATA	NO DATA	NO DATA
Streptococcus mitis	NCTC 12261	NO DATA	NO DATA	NO DATA
Streptococcus mutans	UA159	NO DATA	NO DATA	NO DATA

Table 6E – Base Compositions of Common Respiratory Pathogens for Bioagent Identifying Amplicons Corresponding to Primer Pair Nos: 362, 363, and 367

Organism	Strain	Primer 362 [A G C T]	Primer 363 [A G C T]	Primer 367 [A G C T]
Klebsiella pneumoniae	MGH78578	[21 33 22 16]	[16 34 26 26]	NO DATA
Yersinia pestis	CO-92 Biovar Orientalis	[20 34 18 20]	NO DATA	NO DATA
Yersinia pestis	KIM5 P12 (Biovar Mediaevalis)	[20 34 18 20]	NO DATA	NO DATA
Yersinia pestis	91001	[20 34 18 20]	NO DATA	NO DATA
Haemophilus influenzae	KW20	NO DATA	NO DATA	NO DATA
Pseudomonas aeruginosa	PAO1	[19 35 21 17]	[16 36 28 22]	NO DATA
Pseudomonas fluorescens	Pf0-1	NO DATA	[18 35 26 23]	NO DATA
Pseudomonas putida	KT2440	NO DATA	[16 35 28 23]	NO DATA
Legionella pneumophila	Philadelphia-1	NO DATA	NO DATA	NO DATA
Francisella tularensis	schu 4	NO DATA	NO DATA	NO DATA
Bordetella pertussis	Tohama I	[20 31 24 17]	[15 34 32 21]	[26 25 34 19]
Burkholderia cepacia	J2315	[20 33 21 18]	[15 36 26 25]	[25 27 32 20]
Burkholderia pseudomallei	K96243	[19 34 19 20]	[15 37 28 22]	[25 27 32 20]
Neisseria gonorrhoeae	FA 1090, ATCC 700825	NO DATA	NO DATA	NO DATA
Neisseria meningitidis	MC58 (serogroup B)	NO DATA	NO DATA	NO DATA
Neisseria meningitidis	serogroup C, FAM18	NO DATA	NO DATA	NO DATA
Neisseria meningitidis	Z2491 (serogroup A)	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	TW-183	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	AR39	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	CWL029	NO DATA	NO DATA	NO DATA
Chlamydophila pneumoniae	J138	NO DATA	NO DATA	NO DATA
Corynebacterium diphtheriae	NCTC13129	NO DATA	NO DATA	NO DATA
Mycobacterium avium	k10	[19 34 23 16]	NO DATA	[24 26 35 19]
Mycobacterium avium	104	[19 34 23 16]	NO DATA	[24 26 35 19]
Mycobacterium tuberculosis	CSU#93	[19 31 25 17]	NO DATA	[25 25 34 20]
Mycobacterium tuberculosis	CDC 1551	[19 31 24 18]	NO DATA	[25 25 34 20]
Mycobacterium tuberculosis	H37Rv (lab strain)	[19 31 24 18]	NO DATA	[25 25 34 20]
Mycoplasma pneumoniae	M129	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	MRSA252	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	MSSA476	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	COL	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	Mu50	NO DATA	NO DATA	NO DATA
Staphylococcus aureus	MW2	NO DATA	NO DATA	NO DATA
Staphylococcus	N315	NO DATA	NO DATA	NO DATA

aureus				
Staphylococcus aureus	NCTC 8325	NO DATA	NO DATA	NO DATA
Streptococcus agalactiae	NEM316	NO DATA	NO DATA	NO DATA
Streptococcus equi	NC 002955	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	MGAS8232	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	MGAS315	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	SSI-1	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	MGAS10394	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	Manfredo (M5)	NO DATA	NO DATA	NO DATA
Streptococcus pyogenes	SF370 (M1)	NO DATA	NO DATA	NO DATA
Streptococcus pneumoniae	670	NO DATA	NO DATA	NO DATA
Streptococcus pneumoniae	R6	[20 30 19 23]	NO DATA	NO DATA
Streptococcus pneumoniae	TIGR4	[20 30 19 23]	NO DATA	NO DATA
Streptococcus gordonii	NCTC7868	NO DATA	NO DATA	NO DATA
Streptococcus mitis	NCTC 12261	NO DATA	NO DATA	NO DATA
Streptococcus mutans	UA159	NO DATA	NO DATA	NO DATA

[0114] Four sets of throat samples from military recruits at different military facilities taken at different time points were analyzed using the primers of the present invention. The first set was collected at a military training center from November 1 to December 20, 2002 during one of the most severe outbreaks of pneumonia associated with group A *Streptococcus* in the United States since 1968. During this outbreak, fifty-one throat swabs were taken from both healthy and hospitalized recruits and plated on blood agar for selection of putative group A *Streptococcus* colonies. A second set of 15 original patient specimens was taken during the height of this group A *Streptococcus* -associated respiratory disease outbreak. The third set were historical samples, including twenty-seven isolates of group A *Streptococcus*, from disease outbreaks at this and other military training facilities during previous years. The fourth set of samples was collected from five geographically separated military facilities in the continental U.S. in the winter immediately following the severe November/December 2002 outbreak.

[0115] Pure colonies isolated from group A *Streptococcus*-selective media from all four collection periods were analyzed with the surveillance primer set. All samples showed base compositions that precisely matched the four completely sequenced strains of *Streptococcus pyogenes*. Shown in Figure 4 is a 3D diagram of base composition (axes A, G and C) of bioagent identifying amplicons obtained with primer pair number 14 (a precursor of primer pair

number 348 which targets 16S rRNA). The diagram indicates that the experimentally determined base compositions of the clinical samples closely match the base compositions expected for *Streptococcus pyogenes* and are distinct from the expected base compositions of other organisms.

[0116] In addition to the identification of *Streptococcus pyogenes*, other potentially pathogenic organisms were identified concurrently. Mass spectral analysis of a sample whose nucleic acid was amplified by primer pair number 349 (SEQ ID NOs: 49 and 405) exhibited signals of bioagent identifying amplicons with molecular masses that were found to correspond to analogous base compositions of bioagent identifying amplicons of *Streptococcus pyogenes* (A27 G32 C24 T18), *Neisseria meningitidis* (A25 G27 C22 T18), and *Haemophilus influenzae* (A28 G28 C25 T20) (see Figure 5 and Table 6B). These organisms were present in a ratio of 4:5:20 as determined by comparison of peak heights with peak height of an internal PCR calibration standard as described in commonly owned U.S. Patent Application Serial No: 60/545,425.

[0117] Since certain division-wide primers that target housekeeping genes are designed to provide coverage of specific divisions of bacteria to increase the confidence level for identification of bacterial species, they are not expected to yield bioagent identifying amplicons for organisms outside of the specific divisions. For example, primer pair number 356 (SEQ ID NOs: 232:592) primarily amplifies the nucleic acid of members of the classes *Bacilli* and *Clostridia* and is not expected to amplify proteobacteria such as *Neisseria meningitidis* and *Haemophilus influenzae*. As expected, analysis of the mass spectrum of amplification products obtained with primer pair number 356 does not indicate the presence of *Neisseria meningitidis* and *Haemophilus influenzae* but does indicate the presence of *Streptococcus pyogenes* (Figures 3 and 6, Table 6B). Thus, these primers or types of primers can confirm the absence of particular bioagents from a sample.

[0118] The 15 throat swabs from military recruits were found to contain a relatively small set of microbes in high abundance. The most common were *Haemophilus influenza*, *Neisseria meningitides*, and *Streptococcus pyogenes*. *Staphylococcus epidermidis*, *Moraxella cattarhalis*, *Corynebacterium pseudodiphtheriticum*, and *Staphylococcus aureus* were present in fewer samples. An equal number of samples from healthy volunteers from three different geographic locations, were identically analyzed. Results indicated that the healthy volunteers have bacterial

flora dominated by multiple, commensal non-beta-hemolytic *Streptococcal* species, including the viridans group *streptococci* (*S. parasangunis*, *S. vestibularis*, *S. mitis, S. oralis* and *S. pneumoniae*; data not shown), and none of the organisms found in the military recruits were found in the healthy controls at concentrations detectable by mass spectrometry. Thus, the military recruits in the midst of a respiratory disease outbreak had a dramatically different microbial population than that experienced by the general population in the absence of epidemic disease.

[0119] Example 8: Drill-down Analysis for Determination of emm-Type of *Streptococcus pyogenes* in Epidemic Surveillance

[0120] As a continuation of the epidemic surveillance investigation of Example 7, determination of sub-species characteristics (genotyping) of *Streptococcus pyogenes*, was carried out based on a strategy that generates strain-specific signatures according to the rationale of Multi-Locus Sequence Typing (MLST). In classic MLST analysis, internal fragments of several housekeeping genes are amplified and sequenced (Enright et al. Infection and Immunity, 2001, 69, 2416-2427). In classic MLST analysis, internal fragments of several housekeeping genes are amplified and sequenced. In the present investigation, bioagent identifying amplicons from housekeeping genes were produced using drill-down primers and analyzed by mass spectrometry. Since mass spectral analysis results in molecular mass, from which base composition can be determined, the challenge was to determine whether resolution of *emm* classification of strains of *Streptococcus pyogenes* could be determined.

[0121] An alignment was constructed of concatenated alleles of seven MLST housekeeping genes (glucose kinase (gki), glutamine transporter protein (gtr), glutamate racemase (murI), DNA mismatch repair protein (mutS), xanthine phosphoribosyl transferase (xpt), and acetyl-CoA acetyl transferase (yqiL)) from each of the 212 previously *emm*-typed strains of *Streptococcus pyogenes*. From this alignment, the number and location of primer pairs that would maximize strain identification via base composition was determined. As a result, 6 primer pairs were chosen as standard drill-down primers for determination of *emm*-type of *Streptococcus pyogenes*. These six primer pairs are displayed in Table 7. This drill-down set comprises primers with T modifications (note TMOD designation in primer names) which constitutes a functional improvement with regard to prevention of non-templated adenylation (*vide supra*) relative to originally selected primers which are displayed below in the same row.

Table 7: Group A *Streptococcus* Drill-Down Primer Pairs

Primer Pair No.	Forward Primer Name	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer (SEQ ID NO:)	Target Gene
442	SP101_SPET11_358_387_TMOD_F	311	SP101_SPET11_448_473_TMOD_R	669	gki
80	SP101_SPET11_358_387_F	310	SP101_SPET11_448_473_TMOD_R	668	gki
443	SP101_SPET11_600_629_TMOD_F	314	SP101_SPET11_686_714_TMOD_R	671	gtr
81	SP101_SPET11_600_629_F	313	SP101_SPET11_686_714_R	670	gtr
426	SP101_SPET11_1314_1336_TMOD_F	278	SP101_SPET11_1403_1431_TMOD_R	633	murI
86	SP101_SPET11_1314_1336_F	277	SP101_SPET11_1403_1431_R	632	murI
430	SP101_SPET11_1807_1835_TMOD_F	286	SP101_SPET11_1901_1927_TMOD_R	641	mutS
90	SP101_SPET11_1807_1835_F	285	SP101_SPET11_1901_1927_R	640	mutS
438	SP101_SPET11_3075_3103_TMOD_F	302	SP101_SPET11_3168_3196_TMOD_R	657	xpt
96	SP101_SPET11_3075_3103_F	301	SP101_SPET11_3168_3196_R	656	xpt
441	SP101_SPET11_3511_3535_TMOD_F	309	SP101_SPET11_3605_3629_TMOD_R	664	yqiL
98	SP101_SPET11_3511_3535_F	308	SP101_SPET11_3605_3629_R	663	yqiL

[0122] The primers of Table 7 were used to produce bioagent identifying amplicons from nucleic acid present in the clinical samples. The bioagent identifying amplicons which were subsequently analyzed by mass spectrometry and base compositions corresponding to the molecular masses were calculated.

[0123] Of the 51 samples taken during the peak of the November/December 2002 epidemic (Table 8A-C rows 1-3), all except three samples were found to represent *emm3*, a Group A *Streptococcus* genotype previously associated with high respiratory virulence. The three outliers were from samples obtained from healthy individuals and probably represent non-epidemic strains. Archived samples (Tables 8A-C rows 5-13) from historical collections showed a greater heterogeneity of base compositions and *emm* types as would be expected from different epidemics occurring at different places and dates. The results of the mass spectrometry analysis and *emm* gene sequencing were found to be concordant for the epidemic and historical samples.

Table 8A: Base Composition Analysis of Bioagent Identifying Amplicons of Group A *Streptococcus* samples from Six Military Installations Obtained with Primer Pair Nos. 426 and 430

# of Instances	emm-type by Mass Spectrometry	emm-Gene Sequencing	Location (sample)	Year	murI (Primer Pair No. 426)	mutS (Primer Pair No. 430)
48	3	3	MCRD San Diego (Cultured)	2002	A39 G25 C20 T34	A38 G27 C23 T33
2	6	6			A40 G24 C20 T34	A38 G27 C23 T33
1	28	28			A39 G25 C20 T34	A38 G27 C23 T33
15	3	ND			A39 G25 C20 T34	A38 G27 C23 T33
6	3	3	NHRC San Diego-Archive (Cultured)	2003	A39 G25 C20 T34	A38 G27 C23 T33
3	5,58	5			A40 G24 C20 T34	A38 G27 C23 T33
6	6	6			A40 G24 C20 T34	A38 G27 C23 T33
1	11	11			A39 G25 C20 T34	A38 G27 C23 T33
3	12	12			A40 G24 C20 T34	A38 G26 C24 T33
1	22	22			A39 G25 C20 T34	A38 G27 C23 T33
3	25,75	75			A39 G25 C20 T34	A38 G27 C23 T33
4	44/61,82,9	44/61			A40 G24 C20 T34	A38 G26 C24 T33
2	53,91	91			A39 G25 C20 T34	A38 G27 C23 T33
1	2	2	Ft. Leonard Wood (Cultured)	2003	A39 G25 C20 T34	A38 G27 C24 T32
2	3	3			A39 G25 C20 T34	A38 G27 C23 T33
1	4	4			A39 G25 C20 T34	A38 G27 C23 T33
1	6	6			A40 G24 C20 T34	A38 G27 C23 T33
11	25 or 75	75			A39 G25 C20 T34	A38 G27 C23 T33
1	25,75, 33, 34,4,52,84	75			A39 G25 C20 T34	A38 G27 C23 T33
1	44/61 or 82 or 9	44/61			A40 G24 C20 T34	A38 G26 C24 T33
2	5 or 58	5			A40 G24 C20 T34	A38 G27 C23 T33
3	1	1	Ft. Sill (Cultured)	2003	A40 G24 C20 T34	A38 G27 C23 T33
2	3	3			A39 G25 C20 T34	A38 G27 C23 T33
1	4	4			A39 G25 C20 T34	A38 G27 C23 T33
1	28	28			A39 G25 C20 T34	A38 G27 C23 T33
1	3	3	Ft. Benning (Cultured)	2003	A39 G25 C20 T34	A38 G27 C23 T33
1	4	4			A39 G25 C20 T34	A38 G27 C23 T33
3	6	6			A40 G24 C20 T34	A38 G27 C23 T33
1	11	11			A39 G25 C20 T34	A38 G27 C23 T33
1	13	94**			A40 G24 C20 T34	A38 G27 C23 T33
1	44/61 or 82 or 9	82			A40 G24 C20 T34	A38 G26 C24 T33
1	5 or 58	58			A40 G24 C20 T34	A38 G27 C23 T33
1	78 or 89	89			A39 G25 C20 T34	A38 G27 C23 T33
2	5 or 58	ND	Lackland AFB (Throat Swabs)	2003	A40 G24 C20 T34	A38 G27 C23 T33
1	2				A39 G25 C20 T34	A38 G27 C24 T32
1	81 or 90				A40 G24 C20 T34	A38 G27 C23 T33
1	78				A38 G26 C20 T34	A38 G27 C23 T33
3***	No detection				No detection	No detection
7	3	ND	MCRD San Diego (Throat Swabs)	2002	A39 G25 C20 T34	A38 G27 C23 T33
1	3	ND			No detection	A38 G27 C23 T33
1	3	ND			No detection	No detection
1	3	ND			No detection	No detection
2	3	ND			No detection	A38 G27 C23 T33
3	No detection	ND			No detection	No detection

Table 8B: Base Composition Analysis of Bioagent Identifying Amplicons of Group A *Streptococcus* samples from Six Military Installations Obtained with Primer Pair Nos. 438 and 441

# of Instances	emm-type by Mass Spectrometry	emm-Gene Sequencing	Location (sample)	Year	xpt (Primer Pair No. 438)	yqiL (Primer Pair No. 441)
48	3	3	MCRD San Diego (Cultured)	2002	A30 G36 C20 T36	A40 G29 C19 T31
2	6	6			A30 G36 C20 T36	A40 G29 C19 T31
1	28	28			A30 G36 C20 T36	A41 G28 C18 T32
15	3	ND			A30 G36 C20 T36	A40 G29 C19 T31
6	3	3	NHRC San Diego-Archive (Cultured)	2003	A30 G36 C20 T36	A40 G29 C19 T31
3	5,58	5			A30 G36 C20 T36	A40 G29 C19 T31
6	6	6			A30 G36 C20 T36	A40 G29 C19 T31
1	11	11			A30 G36 C20 T36	A40 G29 C19 T31
3	12	12			A30 G36 C19 T37	A40 G29 C19 T31
1	22	22			A30 G36 C20 T36	A40 G29 C19 T31
3	25,75	75			A30 G36 C20 T36	A40 G29 C19 T31
4	44/61,82,9	44/61			A30 G36 C20 T36	A41 G28 C19 T31
2	53,91	91			A30 G36 C19 T37	A40 G29 C19 T31
1	2	2	Ft. Leonard Wood (Cultured)	2003	A30 G36 C20 T36	A40 G29 C19 T31
2	3	3			A30 G36 C20 T36	A40 G29 C19 T31
1	4	4			A30 G36 C19 T37	A41 G28 C19 T31
1	6	6			A30 G36 C20 T36	A40 G29 C19 T31
11	25 or 75	75			A30 G36 C20 T36	A40 G29 C19 T31
1	25,75, 33, 34,4,52,84	75			A30 G36 C19 T37	A40 G29 C19 T31
1	44/61 or 82 or 9	44/61			A30 G36 C20 T36	A41 G28 C19 T31
2	5 or 58	5			A30 G36 C20 T36	A40 G29 C19 T31
3	1	1	Ft. Sill (Cultured)	2003	A30 G36 C19 T37	A40 G29 C19 T31
2	3	3			A30 G36 C20 T36	A40 G29 C19 T31
1	4	4			A30 G36 C19 T37	A41 G28 C19 T31
1	28	28			A30 G36 C20 T36	A41 G28 C18 T32
1	3	3	Ft. Benning (Cultured)	2003	A30 G36 C20 T36	A40 G29 C19 T31
1	4	4			A30 G36 C19 T37	A41 G28 C19 T31
3	6	6			A30 G36 C20 T36	A40 G29 C19 T31
1	11	11			A30 G36 C20 T36	A40 G29 C19 T31
1	13	94**			A30 G36 C20 T36	A41 G28 C19 T31
1	44/61 or 82 or 9	82			A30 G36 C20 T36	A41 G28 C19 T31
1	5 or 58	58			A30 G36 C20 T36	A40 G29 C19 T31
1	78 or 89	89			A30 G36 C20 T36	A41 G28 C19 T31
2	5 or 58	ND	Lackland AFB (Throat Swabs)	2003	A30 G36 C20 T36	A40 G29 C19 T31
1	2				A30 G36 C20 T36	A40 G29 C19 T31
1	81 or 90				A30 G36 C20 T36	A40 G29 C19 T31
1	78				A30 G36 C20 T36	A41 G28 C19 T31
3***	No detection				No detection	No detection
7	3	ND	MCRD San Diego (Throat Swabs)	2002	A30 G36 C20 T36	A40 G29 C19 T31
1	3	ND			A30 G36 C20 T36	A40 G29 C19 T31
1	3	ND			A30 G36 C20 T36	No detection
1	3	ND			No detection	A40 G29 C19 T31
2	3	ND			A30 G36 C20 T36	A40 G29 C19 T31
3	No detection	ND			No detection	No detection

Table 8C: Base Composition Analysis of Bioagent Identifying Amplicons of Group A *Streptococcus* samples from Six Military Installations Obtained with Primer Pair Nos. 438 and 441

# of Instances	emm-type by Mass Spectrometry	emm-Gene Sequencing	Location (sample)	Year	gki (Primer Pair No. 442)	gtr ((Primer Pair No. 443)
48	3	3	MCRD San Diego (Cultured)	2002	A32 G35 C17 T32	A39 G28 C16 T32
2	6	6			A31 G35 C17 T33	A39 G28 C15 T33
1	28	28			A30 G36 C17 T33	A39 G28 C16 T32
15	3	ND			A32 G35 C17 T32	A39 G28 C16 T32
6	3	3	NHRC San Diego-Archive (Cultured)	2003	A32 G35 C17 T32	A39 G28 C16 T32
3	5,58	5			A30 G36 C20 T30	A39 G28 C15 T33
6	6	6			A31 G35 C17 T33	A39 G28 C15 T33
1	11	11			A30 G36 C20 T30	A39 G28 C16 T32
3	12	12			A31 G35 C17 T33	A39 G28 C15 T33
1	22	22			A31 G35 C17 T33	A38 G29 C15 T33
3	25,75	75			A30 G36 C17 T33	A39 G28 C15 T33
4	44/61,82,9	44/61			A30 G36 C18 T32	A39 G28 C15 T33
2	53,91	91			A32 G35 C17 T32	A39 G28 C16 T32
1	2	2	Ft. Leonard Wood (Cultured)	2003	A30 G36 C17 T33	A39 G28 C15 T33
2	3	3			A32 G35 C17 T32	A39 G28 C16 T32
1	4	4			A31 G35 C17 T33	A39 G28 C15 T33
1	6	6			A31 G35 C17 T33	A39 G28 C15 T33
11	25 or 75	75			A30 G36 C17 T33	A39 G28 C15 T33
1	25,75, 33, 34,4,52,84	75			A30 G36 C17 T33	A39 G28 C15 T33
1	44/61 or 82 or 9	44/61			A30 G36 C18 T32	A39 G28 C15 T33
2	5 or 58	5			A30 G36 C20 T30	A39 G28 C15 T33
3	1	1	Ft. Sill (Cultured)	2003	A30 G36 C18 T32	A39 G28 C15 T33
2	3	3			A32 G35 C17 T32	A39 G28 C16 T32
1	4	4			A31 G35 C17 T33	A39 G28 C15 T33
1	28	28			A30 G36 C17 T33	A39 G28 C16 T32
1	3	3	Ft. Benning (Cultured)	2003	A32 G35 C17 T32	A39 G28 C16 T32
1	4	4			A31 G35 C17 T33	A39 G28 C15 T33
3	6	6			A31 G35 C17 T33	A39 G28 C15 T33
1	11	11			A30 G36 C20 T30	A39 G28 C16 T32
1	13	94**			A30 G36 C19 T31	A39 G28 C15 T33
1	44/61 or 82 or 9	82			A30 G36 C18 T32	A39 G28 C15 T33
1	5 or 58	58			A30 G36 C20 T30	A39 G28 C15 T33
1	78 or 89	89			A30 G36 C18 T32	A39 G28 C15 T33
2	5 or 58	ND	Lackland AFB (Throat Swabs)	2003	A30 G36 C20 T30	A39 G28 C15 T33
1	2				A30 G36 C17 T33	A39 G28 C15 T33
1	81 or 90				A30 G36 C17 T33	A39 G28 C15 T33
1	78				A30 G36 C18 T32	A39 G28 C15 T33
3***	No detection				No detection	No detection
7	3	ND	MCRD San Diego (Throat Swabs)	2002	A32 G35 C17 T32	A39 G28 C16 T32
1	3	ND			No detection	No detection
1	3	ND			A32 G35 C17 T32	A39 G28 C16 T32
1	3	ND			A32 G35 C17 T32	No detection
2	3	ND			A32 G35 C17 T32	No detection
3	No detection	ND			No detection	No detection

[0124] Example 9: Design of Calibrant Polynucleotides based on Bioagent Identifying Amplicons for Identification of Species of Bacteria (Bacterial Bioagent Identifying Amplicons)

[0125] This example describes the design of 19 calibrant polynucleotides based on bacterial bioagent identifying amplicons corresponding to the primers of the broad surveillance set (Table 4) and the *Bacillus anthracis* drill-down set (Table 5).

[0126] Calibration sequences were designed to simulate bacterial bioagent identifying amplicons produced by the T modified primer pairs shown in Table 4 (primer names have the designation "TMOD"). The calibration sequences were chosen as a representative member of the section of bacterial genome from specific bacterial species which would be amplified by a given primer pair. The model bacterial species upon which the calibration sequences are based are also shown in Table 9. For example, the calibration sequence chosen to correspond to an amplicon produced by primer pair no. 361 is SEQ ID NO: 722. In Table 9, the forward (_F) or reverse (_R) primer name indicates the coordinates of an extraction representing a gene of a standard reference bacterial genome to which the primer hybridizes e.g.: the forward primer name 16S_EC_713_732_TMOD_F indicates that the forward primer hybridizes to residues 713-732 of the gene encoding 16S ribosomal RNA in an *E. coli* reference sequence (in this case, the reference sequence is an extraction consisting of residues 4033120-4034661 of the genomic sequence of *E. coli* K12 (GenBank gi number 16127994). Additional gene coordinate reference information is shown in Table 10. The designation "TMOD" in the primer names indicates that the 5' end of the primer has been modified with a non-matched template T residue which prevents the PCR polymerase from adding non-templated adenosine residues to the 5' end of the amplification product, an occurrence which may result in miscalculation of base composition from molecular mass data (*vide supra*).

[0127] The 19 calibration sequences described in Tables 9 and 10 were combined into a single calibration polynucleotide sequence (SEQ ID NO: 741 - which is herein designated a "combination calibration polynucleotide") which was then cloned into a pCR®-Blunt vector (Invitrogen, Carlsbad, CA). This combination calibration polynucleotide can be used in conjunction with the primers of Table 9 as an internal standard to produce calibration amplicons for use in determination of the quantity of any bacterial bioagent. Thus, for example, when the combination calibration polynucleotide vector is present in an amplification reaction mixture, a calibration amplicon based on primer pair 346 (16S rRNA) will be produced in an amplification

reaction with primer pair 346 and a calibration amplicon based on primer pair 363 (rpoC) will be produced with primer pair 363. Coordinates of each of the 19 calibration sequences within the calibration polynucleotide (SEQ ID NO: 783) are indicated in Table 10.

Table 9: Bacterial Primer Pairs for Production of Bacterial Bioagent Identifying Amplicons and Corresponding Representative Calibration Sequences

Primer Pair No.	Forward Primer Name	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer (SEQ ID NO:)	Calibration Sequence Model Species	Calibration Sequence (SEQ ID NO:)
361	16S_EC_1090_1111_2_TMOD_F	5	16S_EC_1175_1196_TMOD_R	370	*Bacillus anthracis*	764
346	16S_EC_713_732_TMOD_F	27	16S_EC_789_809_TMOD_R	389	*Bacillus anthracis*	765
347	16S_EC_785_806_TMOD_F	30	16S_EC_880_897_TMOD_R	392	*Bacillus anthracis*	766
348	16S_EC_960_981_TMOD_F	38	16S_EC_1054_1073_TMOD_R	363	*Bacillus anthracis*	767
349	23S_EC_1826_1843_TMOD_F	49	23S_EC_1906_1924_TMOD_R	405	*Bacillus anthracis*	768
360	23S_EC_2646_2667_TMOD_F	60	23S_EC_2745_2765_TMOD_R	416	*Bacillus anthracis*	769
350	CAPC_BA_274_303_TMOD_F	98	CAPC_BA_349_376_TMOD_R	452	*Bacillus anthracis*	770
351	CYA_BA_1353_1379_TMOD_F	128	CYA_BA_1448_1467_TMOD_R	483	*Bacillus anthracis*	771
352	INFB_EC_1365_1393_TMOD_F	161	INFB_EC_1439_1467_TMOD_R	516	*Bacillus anthracis*	772
353	LEF_BA_756_781_TMOD_F	175	LEF_BA_843_872_TMOD_R	531	*Bacillus anthracis*	773
356	RPLB_EC_650_679_TMOD_F	232	RPLB_EC_739_762_TMOD_R	592	*Clostridium botulinum*	774
449	RPLB_EC_690_710_F	237	RPLB_EC_737_758_R	589	*Clostridium botulinum*	775
359	RPOB_EC_1845_1866_TMOD_F	241	RPOB_EC_1909_1929_TMOD_R	597	*Yersinia Pestis*	776
362	RPOB_EC_3799_3821_TMOD_F	245	RPOB_EC_3862_3888_TMOD_R	603	*Burkholderia mallei*	777
363	RPOC_EC_2146_2174_TMOD_F	257	RPOC_EC_2227_2245_TMOD_R	621	*Burkholderia mallei*	778
354	RPOC_EC_2218_2241_TMOD_F	262	RPOC_EC_2313_2337_TMOD_R	625	*Bacillus anthracis*	779
355	SSPE_BA_115_137_TMOD_F	321	SSPE_BA_197_222_TMOD_R	687	*Bacillus anthracis*	780
367	TUFB_EC_957_979_TMOD_F	345	TUFB_EC_1034_1058_TMOD_R	701	*Burkholderia mallei*	781
358	VALS_EC_1105_1124_TMOD_F	350	VALS_EC_1195_1218_TMOD_R	712	*Yersinia Pestis*	782

Table 10: Primer Pair Gene Coordinate References and Calibration Polynucleotide Sequence Coordinates within the Combination Calibration Polynucleotide

Bacterial Gene and Species	Gene Extraction Coordinates of Genomic or Plasmid Sequence	Reference GenBank GI No. of Genomic (G) or Plasmid (P) Sequence	Primer Pair No.	Coordinates of Calibration Sequence in Combination Calibration Polynucleotide (SEQ ID NO: 783)
16S *E. coli*	4033120..4034661	16127994 (G)	346	16..109
16S *E. coli*	4033120..4034661	16127994 (G)	347	83..190
16S *E. coli*	4033120..4034661	16127994 (G)	348	246..353
16S *E. coli*	4033120..4034661	16127994 (G)	361	368..469
23S *E. coli*	4166220..4169123	16127994 (G)	349	743..837
23S *E. coli*	4166220..4169123	16127994 (G)	360	865..981
rpoB *E. coli*.	4178823..4182851 (complement strand)	16127994 (G)	359	1591..1672
rpoB *E. coli*	4178823..4182851 (complement strand)	16127994 (G)	362	2081..2167
rpoC *E. coli*	4182928..4187151	16127994 (G)	354	1810..1926
rpoC *E. coli*	4182928..4187151	16127994 (G)	363	2183..2279
infB *E. coli*	3313655..3310983 (complement strand)	16127994 (G)	352	1692..1791
tufB *E. coli*	4173523..4174707	16127994 (G)	367	2400..2498
rplB *E. coli*	3449001..3448180	16127994 (G)	356	1945..2060
rplB *E. coli*	3449001..3448180	16127994 (G)	449	1986..2055
valS *E. coli*	4481405..4478550 (complement strand)	16127994 (G)	358	1462..1572

capC *B. anthracis*	56074..55628 (complement strand)	6470151 (P)	350	2517..2616
cya *B. anthracis*	156626..154288 (complement strand)	4894216 (P)	351	1338..1449
lef *B. anthracis*	127442..129921	4894216 (P)	353	1121..1234
sspE *B. anthracis*	226496..226783	30253828 (G)	355	1007-1104

[0128] Example 10: Use of a Calibration Polynucleotide for Determining the Quantity of *Bacillus Anthracis* in a Sample Containing a Mixture of Microbes

[0129] The process described in this example is shown in Figure 7. The capC gene is a gene involved in capsule synthesis which resides on the pX02 plasmid of *Bacillus anthracis*. Primer pair number 350 (see Tables 9 and 10) was designed to identify *Bacillus anthracis* via production of a bacterial bioagent identifying amplicon. Known quantities of the combination calibration polynucleotide vector described in Example 3 were added to amplification mixtures containing bacterial bioagent nucleic acid from a mixture of microbes which included the Ames strain of *Bacillus anthracis*. Upon amplification of the bacterial bioagent nucleic acid and the combination calibration polynucleotide vector with primer pair no. 350, bacterial bioagent identifying amplicons and calibration amplicons were obtained and characterized by mass spectrometry. A mass spectrum measured for the amplification reaction is shown in Figure 8). The molecular masses of the bioagent identifying amplicons provided the means for identification of the bioagent from which they were obtained (Ames strain of *Bacillus anthracis*) and the molecular masses of the calibration amplicons provided the means for their identification as well. The relationship between the abundance (peak height) of the calibration amplicon signals and the bacterial bioagent identifying amplicon signals provides the means of calculation of the copies of the pX02 plasmid of the Ames strain of *Bacillus anthracis*. Methods of calculating quantities of molecules based on internal calibration procedures are well known to those of ordinary skill in the art.

[0130] Averaging the results of 10 repetitions of the experiment described above, enabled a calculation that indicated that the quantity of Ames strain of *Bacillus anthracis* present in the sample corresponds to approximately 10 copies of pX02 plasmid.

[0131] Example 11: Drill-down Genotyping of *Campylobacter* Species

[0132] A series of drill-down primers were designed as described in Example 1 with the objective of identification of different strains of *Campylobacter jejuni*. The primers are listed in Table 11 with the designation "CJST_CJ." Housekeeping genes to which the primers hybridize and produce bioagent identifying amplicons include: tkt (transketolase), glyA (serine

hydroxymethyltransferase), gltA (citrate synthase), aspA (aspartate ammonia lyase), glnA (glutamine synthase), pgm (phosphoglycerate mutase), and uncA (ATP synthetase alpha chain).

Table 11: *Campylobacter* Drill-down Primer Pairs

Primer Pair No.	Forward Primer Name	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer (SEQ ID NO:)	Target Gene
1053	CJST_CJ_1080_1110_F	102	CJST_CJ_1166_1198_R	456	gltA
1064	CJST_CJ_1680_1713_F	107	CJST_CJ_1795_1822_R	461	glyA
1054	CJST_CJ_2060_2090_F	109	CJST_CJ_2148_2174_R	463	pgm
1049	CJST_CJ_2636_2668_F	113	CJST_CJ_2753_2777_R	467	tkt
1048	CJST_CJ_360_394_F	119	CJST_CJ_442_476_R	472	aspA
1047	CJST_CJ_584_616_F	121	CJST_CJ_663_692_R	474	glnA

[0133] The primers were used to amplify nucleic acid from 50 food product samples provided by the USDA, 25 of which contained *Campylobacter jejuni* and 25 of which contained *Campylobacter coli*. Primers used in this study were developed primarily for the discrimination of *Campylobacter jejuni* clonal complexes and for distinguishing *Campylobacter jejuni* from *Campylobacter coli*. Finer discrimination between *Campylobacter coli* types is also possible by using specific primers targeted to loci where closely-related *Campylobacter coli* isolates demonstrate polymorphisms between strains. The conclusions of the comparison of base composition analysis with sequence analysis are shown in Tables 12A-C.

Table 12A – Results of Base Composition Analysis of 50 *Campylobacter* Samples with Drill-down MLST Primer Pair Nos: 1048 and 1047

Group	Species	Isolate origin	MLST type or Clonal Complex by Base Composition analysis	MLST Type or Clonal Complex by Sequence analysis	Strain	Base Composition of Bioagent Identifying Amplicon Obtained with Primer Pair No: 1048 (aspA)	Base Composition of Bioagent Identifying Amplicon Obtained with Primer Pair No: 1047 (glnA)
J-1	*C. jejuni*	Goose	ST 690 /692/707/991	ST 991	RM3673	A30 G25 C16 T46	A47 G21 C16 T25
J-2	*C. jejuni*	Human	Complex 206/48/353	ST 356, complex 353	RM4192	A30 G25 C16 T46	A48 G21 C17 T23
J-3	*C. jejuni*	Human	Complex 354/179	ST 436	RM4194	A30 G25 C15 T47	A48 G21 C18 T22
J-4	*C. jejuni*	Human	Complex 257	ST 257, complex 257	RM4197	A30 G25 C16 T46	A48 G21 C18 T22
J-5	*C. jejuni*	Human	Complex 52	ST 52, complex 52	RM4277	A30 G25 C16 T46	A48 G21 C17 T23
J-6	*C. jejuni*	Human	Complex 443	ST 51, complex 443	RM4275	A30 G25 C15 T47	A48 G21 C17 T23
					RM4279	A30 G25 C15 T47	A48 G21 C17 T23
J-7	*C. jejuni*	Human	Complex 42	ST 604, complex 42	RM1864	A30 G25 C15 T47	A48 G21 C18 T22
J-8	*C. jejuni*	Human	Complex 42/49/362	ST 362, complex 362	RM3193	A30 G25 C15 T47	A48 G21 C18 T22
J-9	*C. jejuni*	Human	Complex 45/283	ST 147, Complex 45	RM3203	A30 G25 C15 T47	A47 G21 C18 T23
	C. jejuni	Human	Consistent	ST 828	RM4183	A31 G27 C20 T39	A48 G21 C16 T24

			with 74 closely related sequence types (none belong to a clonal complex)	ST 832	RM1169	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1056	RM1857	A31 G27 C20 T39	A48 G21 C16 T24
				ST 889	RM1166	A31 G27 C20 T39	A48 G21 C16 T24
				ST 829	RM1182	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1050	RM1518	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1051	RM1521	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1053	RM1523	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1055	RM1527	A31 G27 C20 T39	A48 G21 C16 T24
		Poultry		ST 1017	RM1529	A31 G27 C20 T39	A48 G21 C16 T24
C-1	*C. coli*			ST 860	RM1840	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1063	RM2219	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1066	RM2241	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1067	RM2243	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1068	RM2439	A31 G27 C20 T39	A48 G21 C16 T24
		Swine		ST 1016	RM3230	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1069	RM3231	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1061	RM1904	A31 G27 C20 T39	A48 G21 C16 T24
		Unknown		ST 825	RM1534	A31 G27 C20 T39	A48 G21 C16 T24
				ST 901	RM1505	A31 G27 C20 T39	A48 G21 C16 T24
C-2	*C. coli*	Human	ST 895	ST 895	RM1532	A31 G27 C19 T40	A48 G21 C16 T24
C-3	*C. coli*	Poultry	Consistent with 63 closely related sequence types (none belong to a clonal complex)	ST 1064	RM2223	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1082	RM1178	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1054	RM1525	A31 G27 C20 T39	A48 G21 C16 T24
				ST 1049	RM1517	A31 G27 C20 T39	A48 G21 C16 T24
		Marmoset		ST 891	RM1531	A31 G27 C20 T39	A48 G21 C16 T24

Table 12B – Results of Base Composition Analysis of 50 *Campylobacter* Samples with Drill-down MLST Primer Pair Nos: 1053 and 1064

Group	Species	Isolate origin	MLST type or Clonal Complex by Base Composition analysis	MLST Type or Clonal Complex by Sequence analysis	Strain	Base Composition of Bioagent Identifying Amplicon Obtained with Primer Pair No: 1053 (gltA)	Base Composition of Bioagent Identifying Amplicon Obtained with Primer Pair No: 1064 (glyA)
J-1	*C. jejuni*	Goose	ST 690 /692/707/991	ST 991	RM3673	A24 G25 C23 T47	A40 G29 C29 T45
J-2	*C. jejuni*	Human	Complex 206/48/353	ST 356, complex 353	RM4192	A24 G25 C23 T47	A40 G29 C29 T45
J-3	*C. jejuni*	Human	Complex 354/179	ST 436	RM4194	A24 G25 C23 T47	A40 G29 C29 T45
J-4	*C. jejuni*	Human	Complex 257	ST 257, complex 257	RM4197	A24 G25 C23 T47	A40 G29 C29 T45
J-5	*C. jejuni*	Human	Complex 52	ST 52, complex 52	RM4277	A24 G25 C23 T47	A39 G30 C26 T48
J-6	*C. jejuni*	Human	Complex 443	ST 51, complex 443	RM4275	A24 G25 C23 T47	A39 G30 C28 T46
					RM4279	A24 G25 C23 T47	A39 G30 C28 T46
J-7	*C. jejuni*	Human	Complex 42	ST 604, complex 42	RM1864	A24 G25 C23 T47	A39 G30 C26 T48

J-8	*C. jejuni*	Human	Complex 42/49/362	ST 362, complex 362	RM3193	A24 G25 C23 T47	A38 G31 C28 T46
J-9	*C. jejuni*	Human	Complex 45/283	ST 147, Complex 45	RM3203	A24 G25 C23 T47	A38 G31 C28 T46
C-1	*C. coli*	Human	Consistent with 74 closely related sequence types (none belong to a clonal complex)	ST 828	RM4183	A23 G24 C26 T46	A39 G30 C27 T47
				ST 832	RM1169	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1056	RM1857	A23 G24 C26 T46	A39 G30 C27 T47
		Poultry		ST 889	RM1166	A23 G24 C26 T46	A39 G30 C27 T47
				ST 829	RM1182	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1050	RM1518	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1051	RM1521	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1053	RM1523	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1055	RM1527	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1017	RM1529	A23 G24 C26 T46	A39 G30 C27 T47
				ST 860	RM1840	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1063	RM2219	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1066	RM2241	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1067	RM2243	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1068	RM2439	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1016	RM3230	A23 G24 C26 T46	A39 G30 C27 T47
		Swine		ST 1069	RM3231	A23 G24 C26 T46	NO DATA
				ST 1061	RM1904	A23 G24 C26 T46	A39 G30 C27 T47
		Unknown		ST 825	RM1534	A23 G24 C26 T46	A39 G30 C27 T47
				ST 901	RM1505	A23 G24 C26 T46	A39 G30 C27 T47
C-2	*C. coli*	Human	ST 895	ST 895	RM1532	A23 G24 C26 T46	A39 G30 C27 T47
C-3	*C. coli*	Poultry	Consistent with 63 closely related sequence types (none belong to a clonal complex)	ST 1064	RM2223	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1082	RM1178	A23 G24 C26 T46	A39 G30 C27 T47
				ST 1054	RM1525	A23 G24 C25 T47	A39 G30 C27 T47
				ST 1049	RM1517	A23 G24 C26 T46	A39 G30 C27 T47
		Marmoset		ST 891	RM1531	A23 G24 C26 T46	A39 G30 C27 T47

Table 12C – Results of Base Composition Analysis of 50 *Campylobacter* Samples with Drill-down MLST Primer Pair Nos: 1054 and 1049

Group	Species	Isolate origin	MLST type or Clonal Complex by Base Composition analysis	MLST Type or Clonal Complex by Sequence analysis	Strain	Base Composition of Bioagent Identifying Amplicon Obtained with Primer Pair No: 1054 (pgm)	Base Composition of Bioagent Identifying Amplicon Obtained with Primer Pair No: 1049 (tkt)
J-1	*C. jejuni*	Goose	ST 690 /692/707/991	ST 991	RM3673	A26 G33 C18 T38	A41 G28 C35 T38
J-2	*C. jejuni*	Human	Complex 206/48/353	ST 356, complex 353	RM4192	A26 G33 C19 T37	A41 G28 C36 T37
J-3	*C. jejuni*	Human	Complex 354/179	ST 436	RM4194	A27 G32 C19 T37	A42 G28 C36 T36

J-4	*C. jejuni*	Human	Complex 257	ST 257, complex 257	RM4197	A27 G32 C19 T37	A41 G29 C35 T37
J-5	*C. jejuni*	Human	Complex 52	ST 52, complex 52	RM4277	A26 G33 C18 T38	A41 G28 C36 T37
J-6	*C. jejuni*	Human	Complex 443	ST 51, complex 443	RM4275	A27 G31 C19 T38	A41 G28 C36 T37
					RM4279	A27 G31 C19 T38	A41 G28 C36 T37
J-7	*C. jejuni*	Human	Complex 42	ST 604, complex 42	RM1864	A27 G32 C19 T37	A42 G28 C35 T37
J-8	*C. jejuni*	Human	Complex 42/49/362	ST 362, complex 362	RM3193	A26 G33 C19 T37	A42 G28 C35 T37
J-9	*C. jejuni*	Human	Complex 45/283	ST 147, Complex 45	RM3203	A28 G31 C19 T37	A43 G28 C36 T35
	C. jejuni	Human	Consistent with 74 closely related sequence types (none belong to a clonal complex)	ST 828	RM4183	A27 G30 C19 T39	A46 G28 C32 T36
C-1	*C. coli*			ST 832	RM1169	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1056	RM1857	A27 G30 C19 T39	A46 G28 C32 T36
		Poultry		ST 889	RM1166	A27 G30 C19 T39	A46 G28 C32 T36
				ST 829	RM1182	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1050	RM1518	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1051	RM1521	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1053	RM1523	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1055	RM1527	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1017	RM1529	A27 G30 C19 T39	A46 G28 C32 T36
				ST 860	RM1840	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1063	RM2219	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1066	RM2241	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1067	RM2243	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1068	RM2439	A27 G30 C19 T39	A46 G28 C32 T36
		Swine		ST 1016	RM3230	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1069	RM3231	A27 G30 C19 T39	A46 G28 C32 T36
				ST 1061	RM1904	A27 G30 C19 T39	A46 G28 C32 T36
		Unknown		ST 825	RM1534	A27 G30 C19 T39	A46 G28 C32 T36
				ST 901	RM1505	A27 G30 C19 T39	A46 G28 C32 T36
C-2	*C. coli*	Human	ST 895	ST 895	RM1532	A27 G30 C19 T39	A45 G29 C32 T36
C-3	*C. coli*	Poultry	Consistent with 63 closely related sequence types (none belong to a clonal complex)	ST 1064	RM2223	A27 G30 C19 T39	A45 G29 C32 T36
				ST 1082	RM1178	A27 G30 C19 T39	A45 G29 C32 T36
				ST 1054	RM1525	A27 G30 C19 T39	A45 G29 C32 T36
				ST 1049	RM1517	A27 G30 C19 T39	A45 G29 C32 T36
		Marmoset		ST 891	RM1531	A27 G30 C19 T39	A45 G29 C32 T36

[0134] The base composition analysis method was successful in identification of 12 different strain groups. *Campylobacter jejuni* and *Campylobacter coli* are generally differentiated by all

loci. Ten clearly differentiated *Campylobacter jejuni* isolates and 2 major *Campylobacter coli* groups were identified even though the primers were designed for strain typing of *Campylobacter jejuni*. One isolate (RM4183) which was designated as *Campylobacter jejuni* was found to group with *Campylobacter coli* and also appears to actually be *Campylobacter coli* by full MLST sequencing.

[0135] Example 12: Identification of *Acinetobacter baumannii* Using Broad Range Survey and Division-Wide Primers in Epidemiological Surveillance

[0136] To test the capability of the broad range survey and division-wide primer sets of Table 4 in identification of *Acinetobacter* species, 183 clinical samples were obtained from individuals participating in, or in contact with individuals participating in Operation Iraqi Freedom (including US service personnel, US civilian patients at the Walter Reed Army Institute of Research (WRAIR), medical staff, Iraqi civilians and enemy prisoners). In addition, 34 environmental samples were obtained from hospitals in Iraq, Kuwait, Germany, the United States and the USNS Comfort, a hospital ship.

[0137] Upon amplification of nucleic acid obtained from the clinical samples, primer pairs 346-349, 360, 361, 354, 362 and 363 (Table 4) all produced bacterial bioagent amplicons which identified *Acinetobacter baumannii* in 215 of 217 samples. The organism *Klebsiella pneumoniae* was identified in the remaining two samples. In addition, 14 different strain types (containing single nucleotide polymorphisms relative to a reference strain of *Acinetobacter baumannii*) were identified and assigned arbitrary numbers from 1 to 14. Strain type 1 was found in 134 of the sample isolates and strains 3 and 7 were found in 46 and 9 of the isolates respectively.

[0138] The epidemiology of strain type 7 of *Acinetobacter baumannii* was investigated. Strain 7 was found in 4 patients and 5 environmental samples (from field hospitals in Iraq and Kuwait). The index patient infected with strain 7 was a pre-war patient who had a traumatic amputation in March of 2003 and was treated at a Kuwaiti hospital. The patient was subsequently transferred to a hospital in Germany and then to WRAIR. Two other patients from Kuwait infected with strain 7 were found to be non-infectious and were not further monitored. The fourth patient was diagnosed with a strain 7 infection in September of 2003 at WRAIR. Since the fourth patient was not related involved in Operation Iraqi Freedom, it was inferred that the fourth patient was the subject of a nosocomial infection acquired at WRAIR as a result of the spread of strain 7 from the index patient.

[0139] The epidemiology of strain type 3 of *Acinetobacter baumannii* was also investigated. Strain type 3 was found in 46 samples, all of which were from patients (US service members, Iraqi civilians and enemy prisoners) who were treated on the USNS Comfort hospital ship and subsequently returned to Iraq or Kuwait. The occurrence of strain type 3 in a single locale may provide evidence that at least some of the infections at that locale were a result of a nosocomial infections.

[0140] This example thus illustrates an embodiment of the present invention wherein the methods of analysis of bacterial bioagent identifying amplicons provide the means for epidemiological surveillance.

[0141] Example 13: Selection and Use of MLST *Acinetobacter baumanii* Drill-down Primers

[0142] To combine the power of high-throughput mass spectrometric analysis of bioagent identifying amplicons with the sub-species characteristic resolving power provided by multi-locus sequence typing (MLST) such as the MLST methods of the MLST Databases at the Max-Planck Institute for Infectious Biology (web.mpiib-berlin.mpg.de/mlst/dbs/Mcatarrhalis/documents/primersCatarrhalis_html), an additional 21 primer pairs were selected based on analysis of housekeeping genes of the genus *Acinetobacter*. Genes to which the drill-down MLST analogue primers hybridize for production of bacterial bioagent identifying amplicons include anthranilate synthase component I (trpE), adenylate kinase (adk), adenine glycosylase (mutY), fumarate hydratase (fumC), and pyrophosphate phospho-hydratase (ppa). These 21 primer pairs are indicated with reference to sequence listings in Table 13. Primer pair numbers 1151-1154 hybridize to and amplify segments of trpE. Primer pair numbers 1155-1157 hybridize to and amplify segments of adk. Primer pair numbers 1158-1164 hybridize to and amplify segments of mutY. Primer pair numbers 1165-1170 hybridize to and amplify segments of fumC. Primer pair number 1171 hybridizes to and amplifies a segment of ppa. The primer names given in Table 13 indicates the coordinates to which the primers hybridize to a reference sequence which comprises a concatenation of the genes TrpE, efp (elongation factor p), adk, mutT, fumC, and ppa. For example, the forward primer of primer pair 1151 is named AB_MLST-11-OIF007_62_91_F because it hybridizes to the *Acinetobacter* MLST primer reference sequence of strain type 11 in sample 007 of Operation Iraqi Freedom (OIF) at positions 62 to 91.

Table 13: MLST Drill-Down Primers for Identification of Sub-species characteristics (Strain Type) of Members of the Bacterial Genus *Acinetobacter*

Primer Pair No.	Forward Primer Name	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer (SEQ ID NO:)
1151	AB_MLST-11-OIF007_62_91_F	83	AB_MLST-11-OIF007_169_203_R	426
1152	AB_MLST-11-OIF007_185_214_F	76	AB_MLST-11-OIF007_291_324_R	432
1153	AB_MLST-11-OIF007_260_289_F	79	AB_MLST-11-OIF007_364_393_R	434
1154	AB_MLST-11-OIF007_206_239_F	78	AB_MLST-11-OIF007_318_344_R	433
1155	AB_MLST-11-OIF007_522_552_F	80	AB_MLST-11-OIF007_587_610_R	435
1156	AB_MLST-11-OIF007_547_571_F	81	AB_MLST-11-OIF007_656_686_R	436
1157	AB_MLST-11-OIF007_601_627_F	82	AB_MLST-11-OIF007_710_736_R	437
1158	AB_MLST-11-OIF007_1202_1225_F	65	AB_MLST-11-OIF007_1266_1296_R	420
1159	AB_MLST-11-OIF007_1202_1225_F	65	AB_MLST-11-OIF007_1299_1316_R	421
1160	AB_MLST-11-OIF007_1234_1264_F	66	AB_MLST-11-OIF007_1335_1362_R	422
1161	AB_MLST-11-OIF007_1327_1356_F	67	AB_MLST-11-OIF007_1422_1448_R	423
1162	AB_MLST-11-OIF007_1345_1369_F	68	AB_MLST-11-OIF007_1470_1494_R	424
1163	AB_MLST-11-OIF007_1351_1375_F	69	AB_MLST-11-OIF007_1470_1494_R	424
1164	AB_MLST-11-OIF007_1387_1412_F	70	AB_MLST-11-OIF007_1470_1494_R	424
1165	AB_MLST-11-OIF007_1542_1569_F	71	AB_MLST-11-OIF007_1656_1680_R	425
1166	AB_MLST-11-OIF007_1566_1593_F	72	AB_MLST-11-OIF007_1656_1680_R	425
1167	AB_MLST-11-OIF007_1611_1638_F	73	AB_MLST-11-OIF007_1731_1757_R	427
1168	AB_MLST-11-OIF007_1726_1752_F	74	AB_MLST-11-OIF007_1790_1821_R	428
1169	AB_MLST-11-OIF007_1792_1826_F	75	AB_MLST-11-OIF007_1876_1909_R	429
1170	AB_MLST-11-OIF007_1792_1826_F	75	AB_MLST-11-OIF007_1895_1927_R	430
1171	AB_MLST-11-	77	AB_MLST-11-OIF007_2097_2118_R	431

	OIF007_1970_2002_F			

[0143] Analysis of bioagent identifying amplicons obtained using the primers of Table 13 for over 200 samples from Operation Iraqi Freedom resulted in the identification of 50 distinct strain type clusters. The largest cluster, designated strain type 11 (ST11) includes 42 sample isolates, all of which were obtained from US service personnel and Iraqi civilians treated at the 28th Combat Support Hospital in Baghdad. Several of these individuals were also treated on the hospital ship USNS Comfort. These observations are indicative of significant epidemiological correlation/linkage.

[0144] All of the sample isolates were tested against a broad panel of antibiotics to characterize their antibiotic resistance profiles. As an example of a representative result from antibiotic susceptibility testing, ST11 was found to consist of four different clusters of isolates, each with a varying degree of sensitivity/resistance to the various antibiotics tested which included penicillins, extended spectrum penicillins, cephalosporins, carbipenem, protein synthesis inhibitors, nucleic acid synthesis inhibitors, anti-metabolites, and anti-cell membrane antibiotics. Thus, the genotyping power of bacterial bioagent identifying amplicons, particularly drill-down bacterial bioagent identifying amplicons, has the potential to increase the understanding of the transmission of infections in combat casualties, to identify the source of infection in the environment, to track hospital transmission of nosocomial infections, and to rapidly characterize drug-resistance profiles which enable development of effective infection control measures on a time-scale previously not achievable.

[0145] Various modifications of the invention, in addition to those described herein, will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS COMPREND PLUS D'UN TOME.

CECI EST LE TOME __1___ DE __2___

NOTE: Pour les tomes additionels, veillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE THAN ONE VOLUME.

THIS IS VOLUME __1___ OF __2___

NOTE: For additional volumes please contact the Canadian Patent Office.

Claims:

1. An oligonucleotide primer pair composition selected from the group consisting of: a forward primer 26 to 36 nucleobases in length comprising 85% to 100% sequence identity with the forward primer of Primer Pair No. 350 SEQ ID NO: 98, and a reverse primer 24 to 34 nucleobases in length comprising 85% to 100% sequence identity with the reverse primer of Primer Pair No. 350 SEQ ID NO: 452; a forward primer 23 to 33 nucleobases in length comprising 85% to 100% sequence identity with the forward primer of Primer Pair No. 351 SEQ ID NO: 128, and a reverse primer 17 to 25 nucleobases in length comprising 85% to 100% sequence identity with the reverse primer of Primer Pair 351 SEQ ID NO: 483; a forward primer 25 to 35 nucleobases in length comprising 85% to 100% sequence identity with the forward primer of Primer Pair No. 352 SEQ ID NO: 161, and a reverse primer 21 to 31 nucleobases in length comprising 85% to 100% sequence identity with the reverse primer of Primer Pair No. 352 SEQ ID NO: 516; a forward primer 22 to 32 nucleobases in length comprising 85% to 100% sequence identity with the forward primer of Primer Pair No. 353 SEQ ID NO: 175, and a reverse primer 18 to 26 nucleobases in length comprising 85% to 100% sequence identity with the reverse primer of Primer Pair No. 353 SEQ ID NO: 531; and a forward primer 20 to 28 nucleobases in length comprising 85% to 100% sequence identity with the forward primer of Primer Pair No. 355 SEQ ID NO: 321, and a reverse primer 22 to 32 nucleobases in length comprising 85% to 100% sequence identity with the reverse primer of Primer Pair No. 355 SEQ ID NO: 687.

2. A composition comprising two or more of the oligonucleotide primer pair compositions of claim 1.

3. The composition of claim 1 or 2, wherein one of said primers comprises at least one modified nucleobase comprising: 7-deaza-2'-deoxyadenosine-5-triphosphate, 5-iodo-2'- deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'- deoxycytidine-5 '-triphosphate, 5-iodo-2'-deoxycytidine-5'-triphosphate, 5-hydroxy-2'- deoxyuridine-5'-triphosphate, 4-thiothymidine-5'-triphosphate, 5-aza-2'-deoxyuridine-5'- triphosphate, 5-fluoro-2'-deoxyuridine-5'-triphosphate, O6-methyl-2'-deoxyguanosine-5'- triphosphate, N2-methyl-2'-

deoxyguanosine-5'-triphosphate, 8-oxo-2'-deoxyguanosine-5'- triphosphate, thiothymidine-5'-triphosphate, or a nucleobase comprising ^{15}N or ^{13}C.

4. The composition of claim 1 or 2, wherein both of said forward primer and said reverse primer comprise at least one modified nucleobase comprising: 7-deaza-2'-deoxyadenosine-5-triphosphate, 5-iodo-2'- deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'- deoxycytidine-5 '-triphosphate, 5-iodo-2'-deoxycytidine-5'-triphosphate, 5-hydroxy-2'- deoxyuridine-5'-triphosphate, 4-thiothymidine-5'-triphosphate, 5-aza-2'-deoxyuridine-5'-triphosphate, 5-fluoro-2'-deoxyuridine-5'-triphosphate, O6-methyl-2'-deoxyguanosine-5'-triphosphate, N2-methyl-2'-deoxyguanosine-5'-triphosphate, 8-oxo-2'-deoxyguanosine-5'-triphosphate, thiothymidine-5'-triphosphate, or a nucleobase comprising ^{15}N or ^{13}C.

5. The composition of claim 1 or 2, wherein one of said primers comprises a non-templated T residue on the 5 '-end.

6. The composition of claim 1 or 2, wherein both of said forward primer and said reverse primer comprise a non-templated T residue on the 5 '-end.

7. The composition of claim 1 or 2, wherein one of said primers comprises at least one non-template tag.

8. The composition of claim 1 or 2, wherein both of said forward primer and said reverse primer comprise at least one non-template tag.

9. The composition of any one of claims 1 to 8, wherein one of said primers comprises at least one molecular mass modifying tag.

10. The composition of any one of claims 1 to 8, wherein both of said forward primer and said reverse primer comprise at least one molecular mass modifying tag.

Figure 1

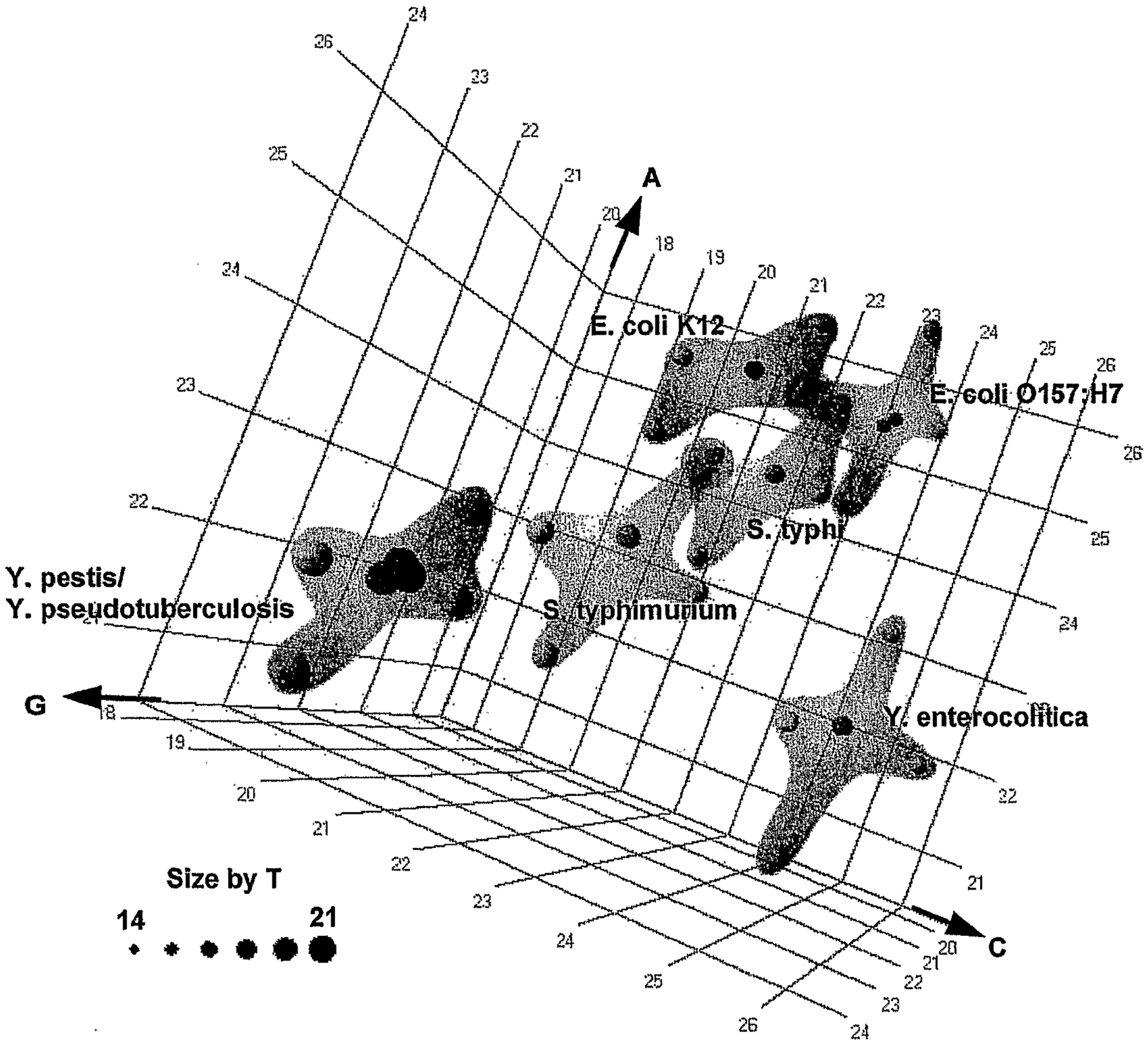

A
E. coli K12
E. coli O157:H7
S. typhi
S. typhimurium
Y. pestis/
Y. pseudotuberculosis
Y. enterocolitica
G
C
Size by T
14
21

Figure 2

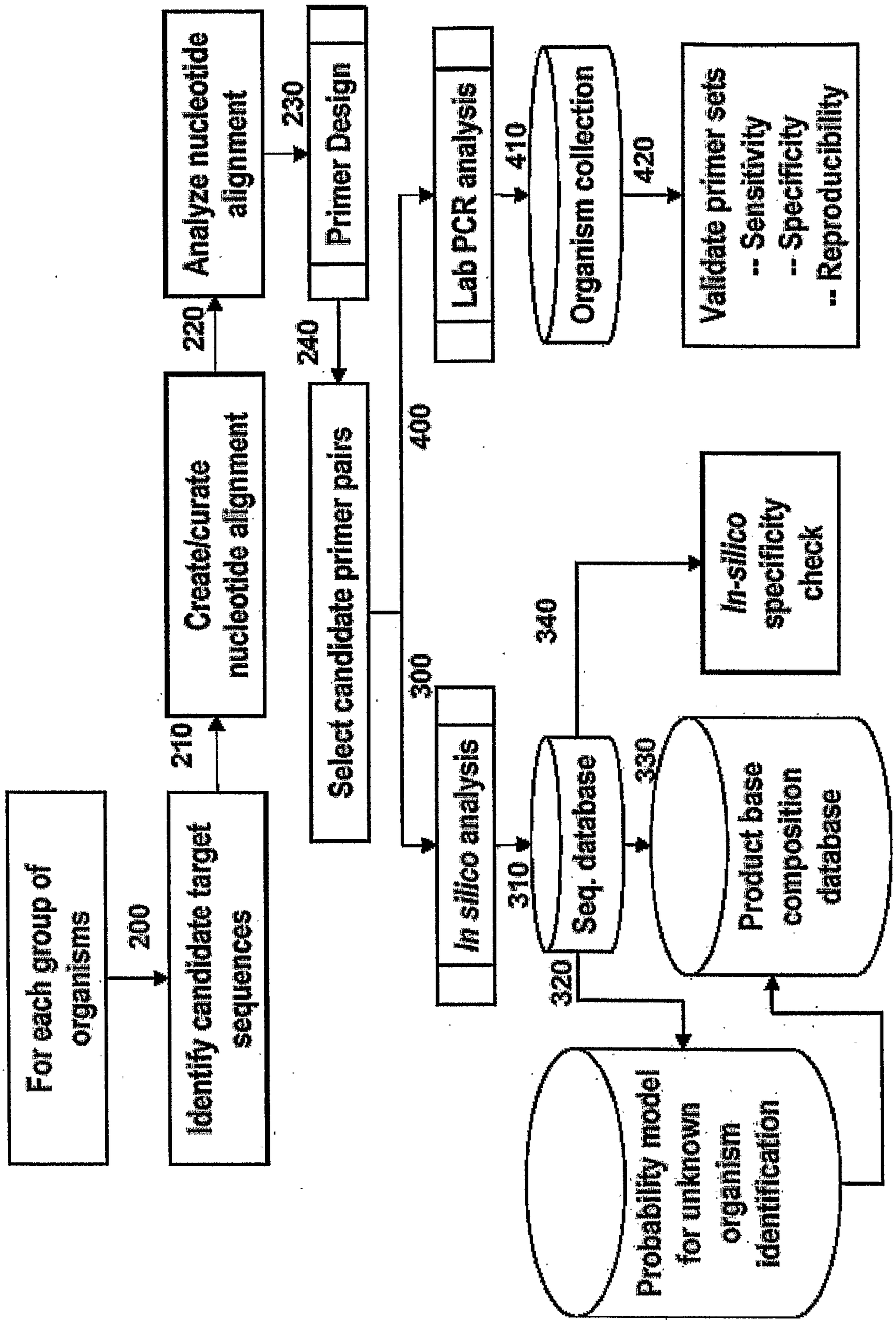

For each group of organisms
200
Identify candidate target sequences
210
Create/curate nucleotide alignment
220
Analyze nucleotide alignment
230
Primer Design
240
Select candidate primer pairs
300
In silico analysis
310
Seq. database
320
Probability model for unknown organism identification
325
Product base composition database
330
In-silico specificity check
340
Lab PCR analysis
400
410
Organism collection
420
Validate primer sets
-- Sensitivity
-- Specificity
-- Reproducibility

Figure 3

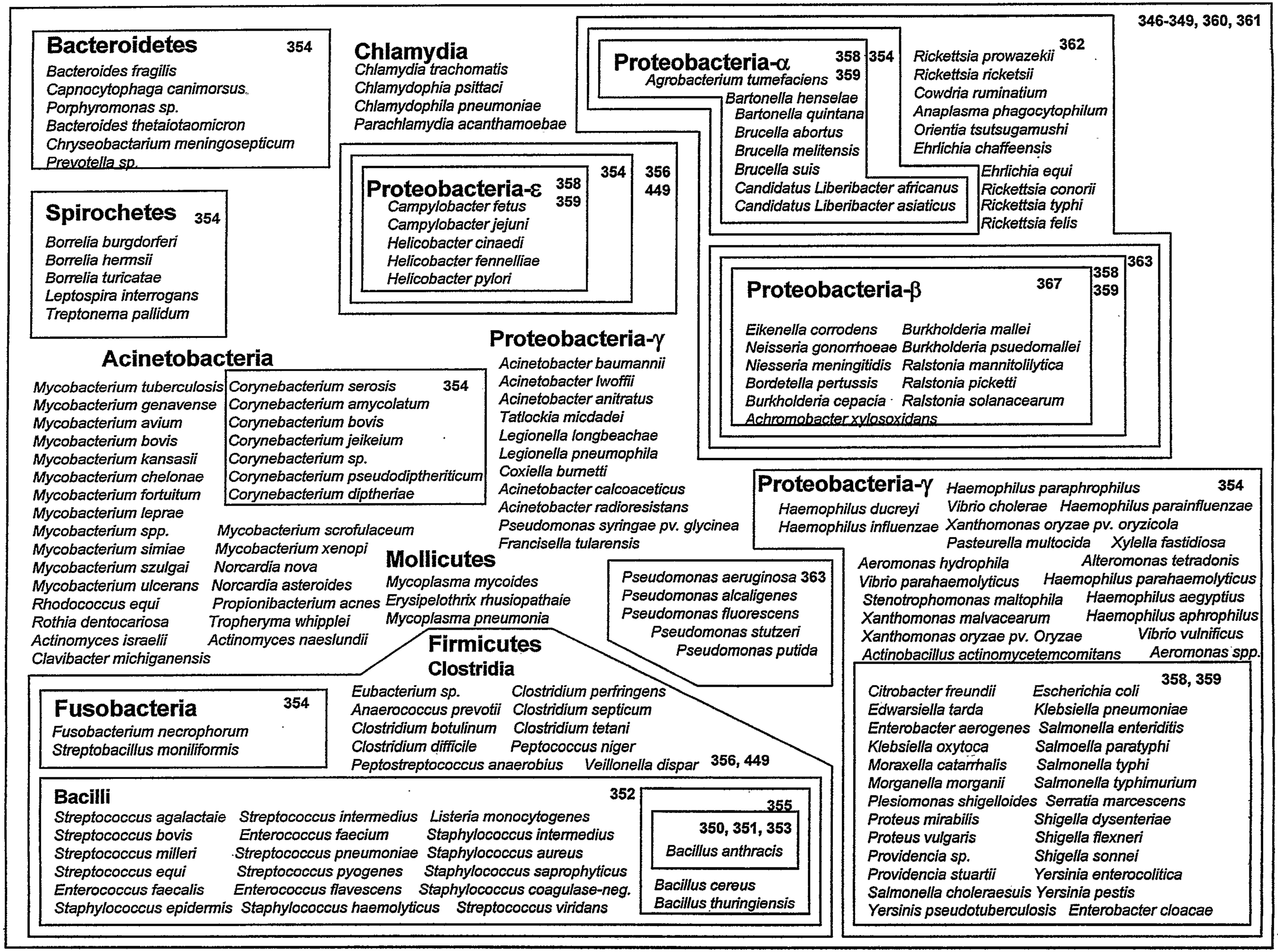

Bacteroidetes 354
Bacteroides fragilis
Capnocytophaga canimorsus
Porphyromonas sp.
Bacteroides thetaiotaomicron
Chryseobactarium meningosepticum
Prevotella sp.
Spirochetes 354
Borrelia burgdorferi
Borrelia hermsii
Borrelia turicatae
Leptospira interrogans
Treptonema pallidum
Chlamydia
Chlamydia trachomatis
Chlamydophia psittaci
Chlamydophila pneumoniae
Parachlamydia acanthamoebae
Proteobacteria-ε 358 359 354 356 449
Campylobacter fetus
Campylobacter jejuni
Helicobacter cinaedi
Helicobacter fennelliae
Helicobacter pylori
346-349, 360, 361
Proteobacteria-α 358 359 354
Agrobacterium tumefaciens
Bartonella henselae
Bartonella quintana
Brucella abortus
Brucella melitensis
Brucella suis
Candidatus Liberibacter africanus
Candidatus Liberibacter asiaticus
362
Rickettsia prowazekii
Rickettsia ricketsii
Cowdria ruminatium
Anaplasma phagocytophilum
Orientia tsutsugamushi
Ehrlichia chaffeensis
Ehrlichia equi
Rickettsia conorii
Rickettsia typhi
Rickettsia felis
363
Proteobacteria-β 367 358 359
Eikenella corrodens
Neisseria gonorrhoeae
Niesseria meningitidis
Bordetella pertussis
Burkholderia cepacia
Achromobacter xylosoxidans
Burkholderia mallei
Burkholderia psuedomallei
Ralstonia mannitolilytica
Ralstonia picketti
Ralstonia solanacearum
Acinetobacteria
Mycobacterium tuberculosis
Mycobacterium genavense
Mycobacterium avium
Mycobacterium bovis
Mycobacterium kansasii
Mycobacterium chelonae
Mycobacterium fortuitum
Mycobacterium leprae
Mycobacterium spp.
Mycobacterium simiae
Mycobacterium szulgai
Mycobacterium ulcerans
Rhodococcus equi
Rothia dentocariosa
Actinomyces israelii
Clavibacter michiganensis
Corynebacterium serosis 354
Corynebacterium amycolatum
Corynebacterium bovis
Corynebacterium jeikeium
Corynebacterium sp.
Corynebacterium pseudodiptheriticum
Corynebacterium diptheriae
Mycobacterium scrofulaceum
Mycobacterium xenopi
Norcardia nova
Norcardia asteroides
Propionibacterium acnes
Tropheryma whipplei
Actinomyces naeslundii
Proteobacteria-γ
Acinetobacter baumannii
Acinetobacter lwoffii
Acinetobacter anitratus
Tatlockia micdadei
Legionella longbeachae
Legionella pneumophila
Coxiella burnetti
Acinetobacter calcoaceticus
Acinetobacter radioresistans
Pseudomonas syringae pv. glycinea
Francisella tularensis
Mollicutes
Mycoplasma mycoides
Erysipelothrix rhusiopathaie
Mycoplasma pneumonia
Pseudomonas aeruginosa 363
Pseudomonas alcaligenes
Pseudomonas fluorescens
Pseudomonas stutzeri
Pseudomonas putida
Proteobacteria-γ
Haemophilus paraphrophilus 354
Haemophilus ducreyi
Haemophilus influenzae
Vibrio cholerae
Haemophilus parainfluenzae
Xanthomonas oryzae pv. oryzicola
Pasteurella multocida
Xylella fastidiosa
Aeromonas hydrophila
Alteromonas tetradonis
Vibrio parahaemolyticus
Haemophilus parahaemolyticus
Stenotrophomonas maltophila
Haemophilus aegyptius
Xanthomonas malvacearum
Haemophilus aphrophilus
Xanthomonas oryzae pv. Oryzae
Vibrio vulnificus
Actinobacillus actinomycetemcomitans
Aeromonas spp.
358, 359
Citrobacter freundii
Edwarsiella tarda
Enterobacter aerogenes
Klebsiella oxytoca
Moraxella catarrhalis
Morganella morganii
Plesiomonas shigelloides
Proteus mirabilis
Proteus vulgaris
Providencia sp.
Providencia stuartii
Salmonella choleraesuis
Yersinis pseudotuberculosis
Escherichia coli
Klebsiella pneumoniae
Salmonella enteriditis
Salmoella paratyphi
Salmonella typhi
Salmonella typhimurium
Serratia marcescens
Shigella dysenteriae
Shigella flexneri
Shigella sonnei
Yersinia enterocolitica
Yersinia pestis
Enterobacter cloacae
Firmicutes
Clostridia
Eubacterium sp.
Anaerococcus prevotii
Clostridium botulinum
Clostridium difficile
Peptostreptococcus anaerobius
Clostridium perfringens
Clostridium septicum
Clostridium tetani
Peptococcus niger
Veillonella dispar 356, 449
Fusobacteria 354
Fusobacterium necrophorum
Streptobacillus moniliformis
Bacilli 352
Streptococcus agalactaie
Streptococcus bovis
Streptococcus milleri
Streptococcus equi
Enterococcus faecalis
Staphylococcus epidermis
Streptococcus intermedius
Enterococcus faecium
Streptococcus pneumoniae
Streptococcus pyogenes
Enterococcus flavescens
Staphylococcus haemolyticus
Listeria monocytogenes
Staphylococcus intermedius
Staphylococcus aureus
Staphylococcus saprophyticus
Staphylococcus coagulase-neg.
Streptococcus viridans
355
350, 351, 353
Bacillus anthracis
Bacillus cereus
Bacillus thuringiensis

Figure 4

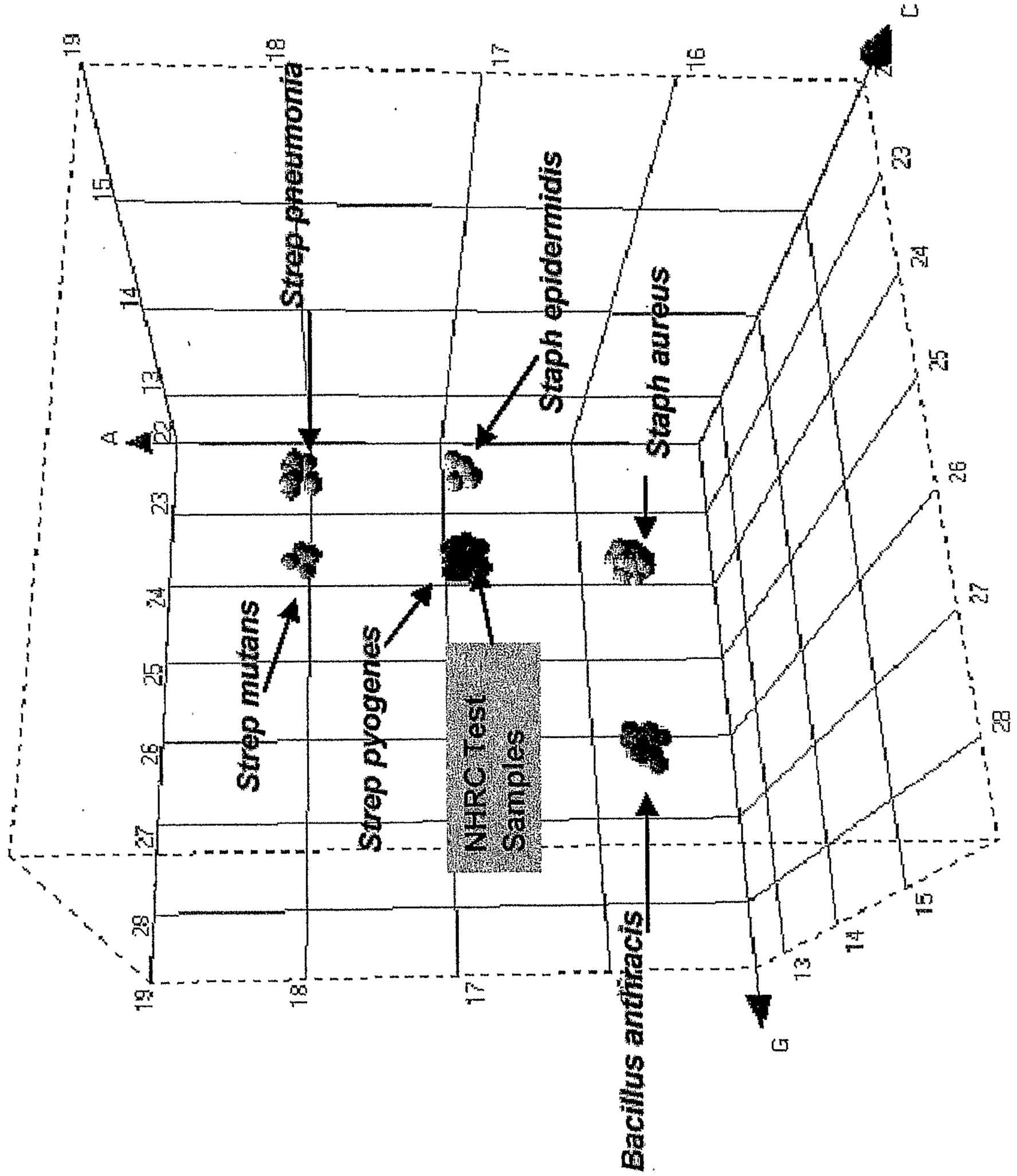

Base Composition Signatures from primer pair 14 (16S rRNA)
Strep pneumonia
Staph epidermidis
Staph aureus
Strep mutans
Strep pyogenes
NHRC Test Samples
cillus anthracis
A
G

Figure 5

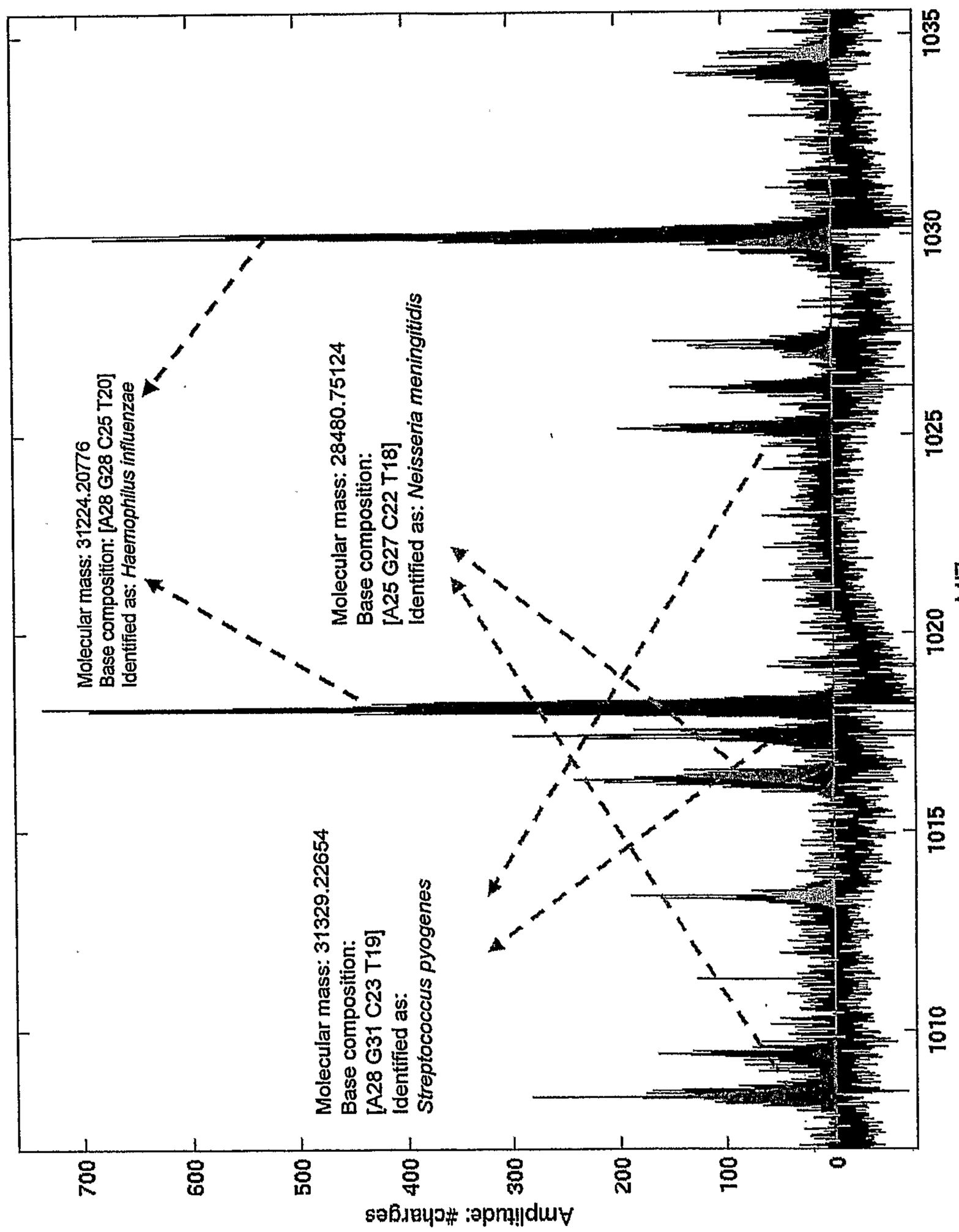
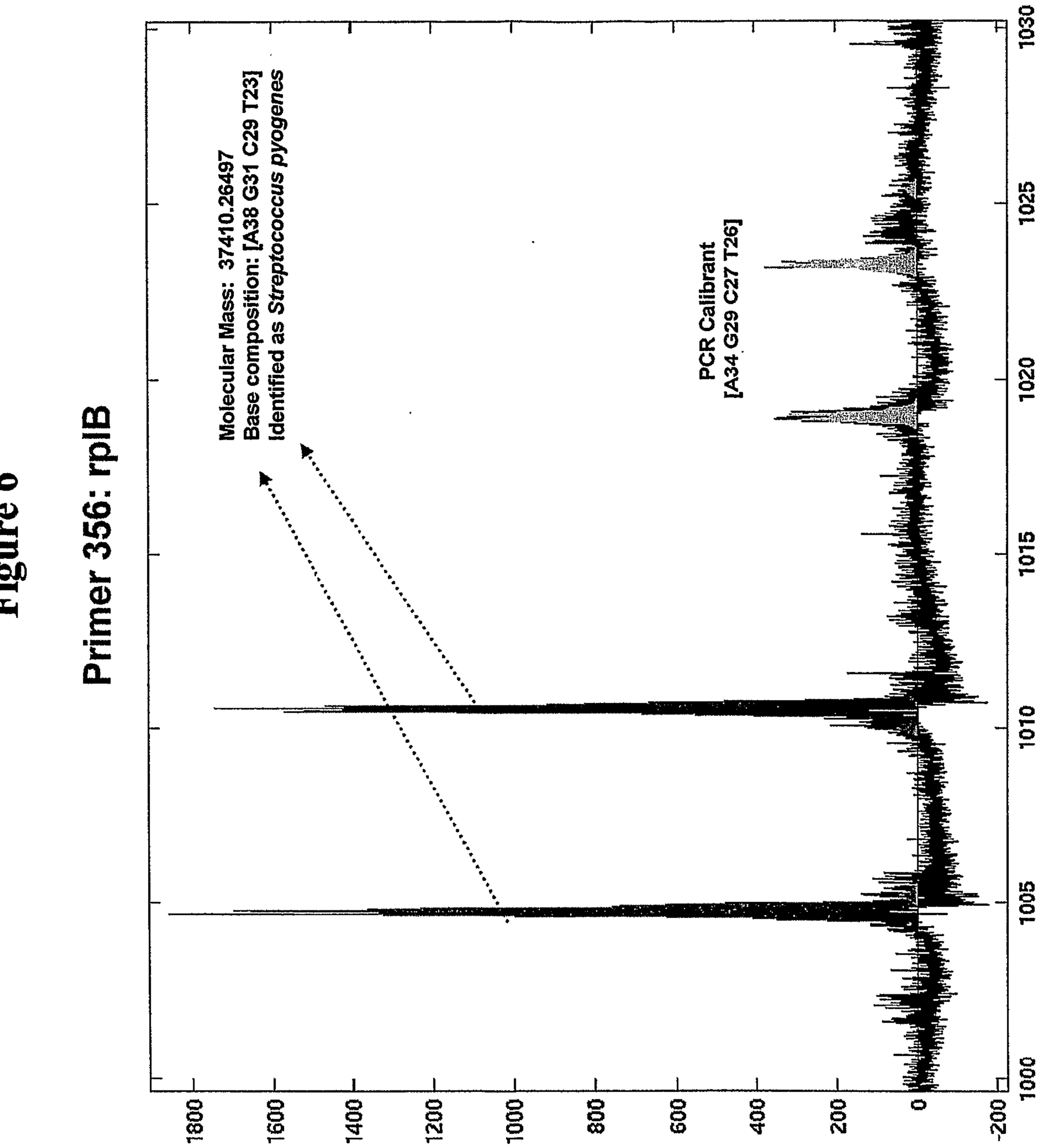

Primer Pair 349: 23S rDNA
Molecular mass: 31224.20776
Base composition: [A28 G28 C25 T20]
Identified as: Haemophilus influenzae
Molecular mass: 28480.75124
Base composition:
[A25 G27 C22 T18]
Identified as: Neisseria meningitidis
Molecular mass: 31329.22654
Base composition:
[A28 G31 C23 T19]
Identified as:
Streptococcus pyogenes
Amplitude: #charges
700
600
500
400
300
200
100
0
1010
1015
1020
1025
1030
1035
M/Z

Figure 6


Primer 356: rplB
Molecular Mass: 37410.26497
Base composition: [A38 G31 C29 T23]
Identified as Streptococcus pyogenes
PCR Calibrant
[A34 G29 C27 T26]
Amplitude: #charges
MZ
1800
1600
1400
1200
1000
800
600
400
200
0
-200
1000
1005
1010
1015
1020
1025
1030

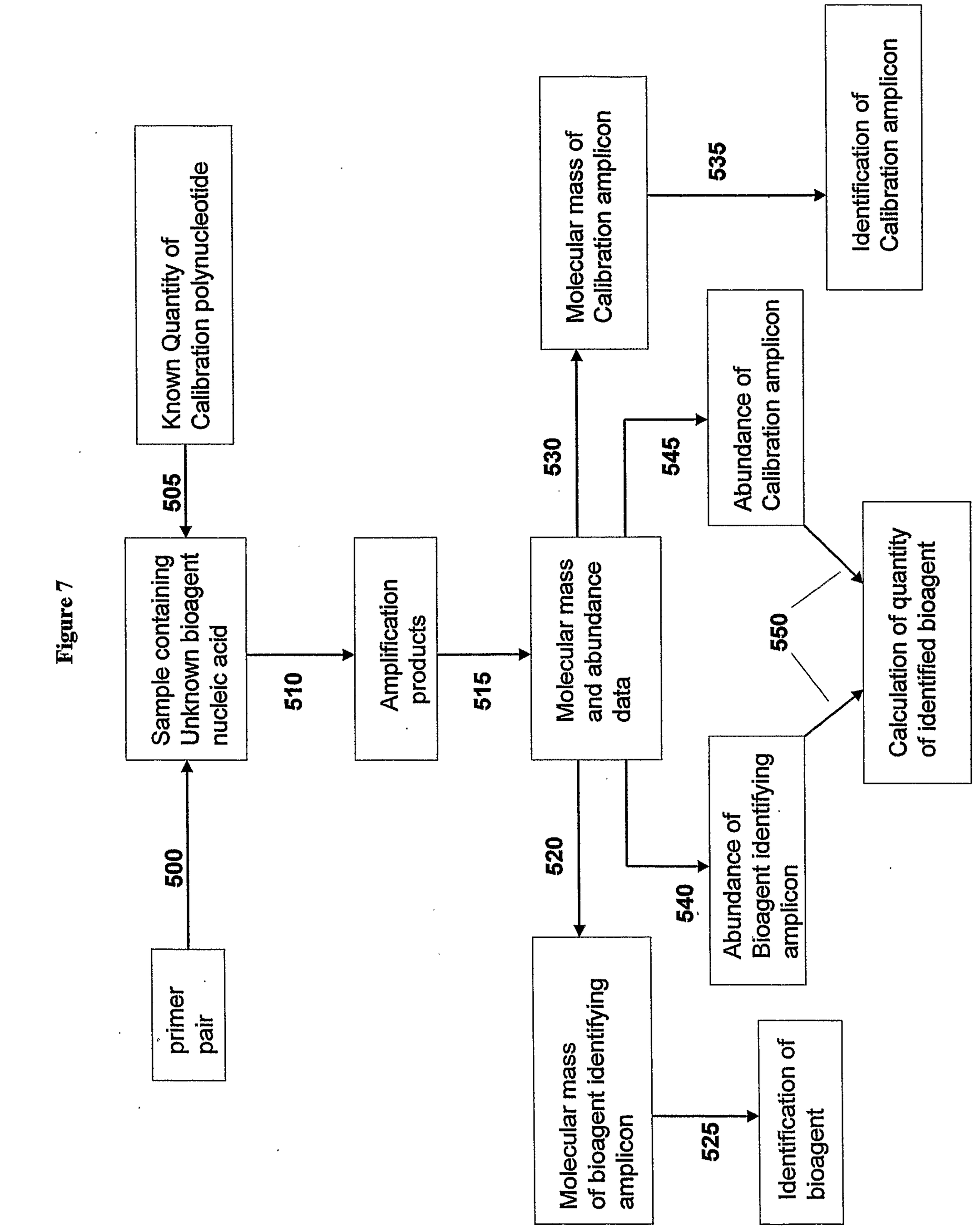
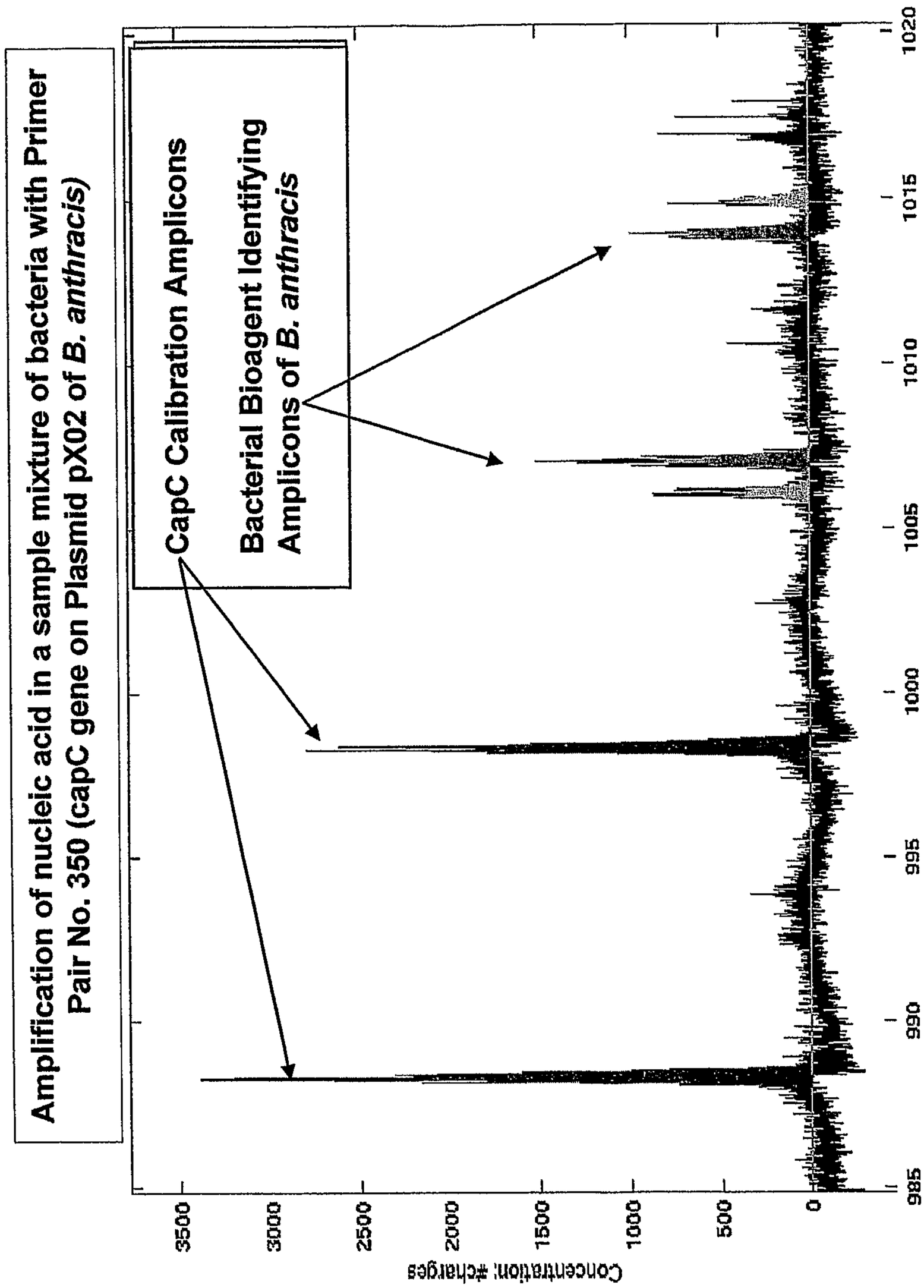

primer pair
500
Sample containing Unknown bioagent nucleic acid
505
Known Quantity of Calibration polynucleotide
510
Amplification products
515
Molecular mass and abundance data
520
Molecular mass of bioagent identifying amplicon
525
Identification of bioagent
530
Molecular mass of Calibration amplicon
535
Identification of Calibration amplicon
540
Abundance of Bioagent identifying amplicon
545
Abundance of Calibration amplicon
550
Calculation of quantity of identified bioagent

Figure 7

Figure 8


Amplification of nucleic acid in a sample mixture of bacteria with Primer Pair No. 350 (capC gene on Plasmid pX02 of B. anthracis)
CapC Calibration Amplicons
Bacterial Bioagent Identifying Amplicons of B. anthracis
Concentration: #charges
MZ