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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 November 2008 (27.11.2008)

PCT

(10) International Publication Number
WO 2008/143627 A2

(51) International Patent Classification:
A61K 31/4741 (2006.01)

(21) International Application Number:
PCT/US2007/020045

(22) International Filing Date:
14 September 2007 (14.09.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/825,703 14 September 2006 (14.09.2006) US
60/946,367 26 June 2007 (26.06.2007) US

(71) Applicant *(for all designated States except US)*: **IBIS BIOSCIENCES, INC.** [US/US]; 1896 Rutherford Road, Carlsbad, CA 92008 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **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).

(74) Agents: **CASIMIR, David, A.** et al.; Casimir Jones S.C., 440 Science Drive, Suite 203, Madison, WI 53711 (US).

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

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

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

WO 2008/143627 A2

(54) **Title:** TARGETED WHOLE GENOME AMPLIFICATION METHOD FOR IDENTIFICATION OF PATHOGENS

(57) **Abstract:** The methods disclosed herein relate to methods and compositions for amplifying nucleic acid sequences, more specifically, from nucleic acid sequences of pathogens by targeted whole genome amplification.

TARGETED WHOLE GENOME AMPLIFICATION METHOD FOR IDENTIFICATION OF PATHOGENS

The present application claims priorty to United State Provisional Application Serial Numbers 60/825,703, filed September 14, 2006 and 60/946,367, filed June 26, 2007, the disclosures of each of which are herein incorporated by reference in their entiries.

FIELD OF THE INVENTION

[1] The methods disclosed herein relate to methods and compositions for amplifying nucleic acid sequences, more specifically, from specific nucleic acid sequences of pathogens.

GOVERNMENT SUPPORT STATEMENT

[2] This invention was made with United States Government support under HSARPA W81XWH-05-C-0116. The United States Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[3] In many fields of research such as genetic diagnosis, cancer research or forensic medicine, the scarcity of genomic DNA can be a severely limiting factor on the type and quantity of genetic tests that can be performed on a sample. One approach designed to overcome this problem is whole genome amplification. The objective is to amplify a limited DNA sample in a non-specific manner in order to generate a new sample that is indistinguishable from the original but with a higher DNA concentration. The aim of a typical whole genome amplification technique would be to amplify a sample up to a microgram level while respecting the original sequence representation.

[4] The first whole genome amplification methods were described in 1992, and were based on the principles of the polymerase chain reaction. Zhang and coworkers (Zhang, L., et al. Proc. Natl. Acad. Sci. USA, 1992, 89: 5847-5851) developed the primer extension PCR technique (PEP) and Telenius and collaborators (Telenius et al., Genomics. 1992, 13(3):718-25) designed the degenerate oligonucleotide-primed PCR method (DOP-PCR) Zhang et al., 1992).

[5] PEP involves a high number of PCR cycles; using *Taq* polymerase and 15 base random primers that anneal at a low stringency temperature. Although the PEP protocol has been improved in different ways, it still results in incomplete genome coverage, failing to amplify certain sequences such as repeats. Failure to prime and amplify regions containing repeats may lead to incomplete representation of a whole genome because consistent primer coverage across the length of the genome provides for optimal representation of the genome. This method also has limited efficiency on very small samples (such as single cells). Moreover, the use of *Taq* polymerase implies that the maximal product length is about 3 kb.

[6] DOP-PCR is a method which uses *Taq* polymerase and semi-degenerate oligonucleotides (such as CGACTCGAGNNNNNNATGTGG (SEQ ID NO: 1), for example, where N = A, T, C or G) that bind at a low annealing temperature at approximately one million sites within the human genome. The first cycles are followed by a large number of cycles with a higher annealing temperature, allowing only for the amplification of the fragments that were tagged in the first step. This leads to incomplete representation of a whole genome. DOP-PCR generates, like PEP, fragments that are in average 400-500 bp, with a maximum size of 3 kb, although fragments up to 10 kb have been reported. On the other hand, as noted for PEP, a low input of genomic DNA (less than 1 ng) decreases the fidelity and the genome coverage (Kittler et al., Anal. Biochem. 2002, 300(2), 237-44).

[7] Multiple displacement amplification (MDA, also known as strand displacement amplification; SDA) is a non-PCR-based isothermal method based on the annealing of random hexamers to denatured DNA, followed by strand-displacement synthesis at constant temperature (Blanco et al., 1989, J. Biol. Chem. 264:8935-40). It has been applied to small genomic DNA samples, leading to the synthesis of high molecular weight DNA with limited sequence representation bias (Lizardi et al., Nature Genetics 1998, 19, 225-232; Dean et al., Proc. Natl. Acad. Sci. U.S.A. 2002, 99, 5261-5266). As DNA is synthesized by strand displacement, a gradually increasing number of priming events occur, forming a network of hyper-branched DNA structures. The reaction can be catalyzed by the Phi29 DNA polymerase or by the large fragment of the *Bst* DNA polymerase. The Phi29 DNA polymerase possesses a proofreading activity resulting in error rates 100 times lower than the *Taq* polymerase.

[8] The methods described above generally produce amplification of whole genomes wherein all of the nucleic acid in a given sample is indiscriminately amplified. These methods cannot selectively amplify target genomes in the presence of background or contaminating genomes. Therefore, the results obtained from these methods have a problematically high amount of contaminating background nucleic acid. Purifying collected samples to isolate target genome(s) and remove background genome(s) will result in a further reduction in the amount of already scarce target genome.

[9] There is a long felt need for a method of targeted amplification of a whole genome relative to background or contaminating genomes. In certain cases where only small quantities of a nucleic acid sample to be tested for the presence of a given target nucleic acid sequence, it would be advantageous to introduce specificity into amplification of whole genomes so that a particular target genome is selectively amplified relative to other genomes present within a given sample. For example, in cases of microbial forensics or clinical diagnostics, it would be useful to selectively amplify a genome of a pathogen, or a class of pathogens relative to the genomes of organisms which are also present in the sample which contains a

small quantity of total nucleic acid. This would provide the quantities of nucleic acid of the pathogen that are necessary to identify the pathogen. The methods disclosed herein satisfy this long felt need.

SUMMARY OF THE INVENTION

[10] The methods disclosed herein include methods of designing targeted whole genome amplification primers and using the targeted whole genome amplification primers in selective whole genome amplification reactions of a sample to elevate the quantity of nucleic acid representing a pathogen genome in a given sample which may be a common diagnostic sample such as blood and fractions or components thereof, sputum, urine, cerebrals spinal fluid, hepatic cells, and tissue biopsies.

[11] Design of targeted whole genome amplification primers is accomplished by identifying at least one pathogen genome of interest and identifying at least one background genome of a bioagent suspected of being present in a sample that would contain the pathogen genome of interest. The next step is to identify all unique genome sequence segments of specified lengths within the pathogen genome sequence and to determine the frequency of occurrence of these genome sequence segments in the pathogen genome(s) and in the background genome(s). The next step is to calculate a selectivity ratio for the genome sequence segments by dividing the frequency of occurrence within the pathogen genome sequence by the frequency of occurrence of the plurality of genome sequence segments within the background genome sequences. A selectivity ratio threshold is chosen to a first subset of genome sequence segments that have selectivity ratios equal to or above the selectivity ratio threshold. This first sub-set of genome sequence segments is analyzed with respect to the pathogen genome(s) to determine the lengths of separation of the genome sequence segments along the pathogen genome. A second sub-set of genome sequence segments is chosen from the first sub-set such that the genome sequence segments of the second sub-set have a mean separation distance of less than a selected length of nucleobases. Next, targeted whole genome amplification primers are selected to hybridize to the genome sequence segments of the second sub-set such that the pathogen genome will be amplified selectively over the background genomes when subjected to whole genome amplification conditions.

[12] The elevated quantity of nucleic acid representing a pathogen genome obtained with the targeted whole genome amplification primers may then be used as template DNA for subsequent detailed analyses to identify the pathogen by producing amplification products corresponding to bioagent identifying amplicons. The molecular masses of the bioagent identifying amplicons are measured by mass spectrometry methods such as electrospray time-of-flight mass spectrometry for example. Base compositions of the bioagent identifying amplicons are calculated from the molecular masses. The molecular masses and/or base compositions are then compared with a database of molecular masses and/or base compositions of bioagent identifying amplicons of known bioagents which are defined by specifically designed primer pair, in order to identify the pathogen in the sample. In certain embodiments, the amplification products

corresponding to bioagent identifying amplicons are carried out in multiplexing reactions where more than one primer pair is included in a single reaction mixture.

[13] Also disclosed are diagnostic kits that include any or all of the following components: targeted whole genome amplification primers, a highly processive polymerase suitable for catalyzing a whole genome amplification reaction, deoxynucleotide triphosphates and primer pairs for producing amplification products corresponding to bioagent identifying amplicons. The kits may also include buffer components or additives and instructions for carrying out the amplification reactions such as for example, indications of specific combinations of primer pairs for multiplexing reactions.

[14] Disclosed herein are methods and related kits used for identification of pathogens implicated in septicemia and sepsis. Such methods and kits may include any of primer pairs of primer pair numbers 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 354 (SEQ ID NOs: 597:605), 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 2249 (SEQ ID NOs: 601:609), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627). These primer pairs are useful for obtaining amplification products corresponding to bioagent identifying amplicons which are used to identify pathogens causing septicemia or sepsis. These pathogens are bacteria that include, but are not limited to the following: *Escherichia coli, Klebsiella pneumoniae, Klebsiella oxytoca, Serratia marcescens, Enterobacter cloacae, Enterobacter aerogenes, Proteus mirabilis, Pseudomonas aeruginosa, Acinetobacter baumannii, Stenotrophomonas maltophilia, Staphylococcus aureus, Staphylococcus epidermidis, Staphylococcus haemolyticus, Streptococcus pneumoniae, Streptococcus pyogenes, Streptococcus agalactiae, Streptococcus mitis, Enterococcus faecium, Enterococcus faecalis, Candida albicans, Candida tropicalis, Candida parapsilosis, Candida krusei, Candida glabrata, Mycobacterium tuberculosis,* and *Aspergillus fumigatus.* After selection of appropriate targeted whole genome amplification primers to a reference sequence of any of the genomes of the bacteria including, but not limited to those listed above, which are implicated in sepsis and septicemia, targeted whole genome amplification reactions can be performed to obtain sufficient quantities of nucleic acid such that identification of a bacterium implicated in sepsis or septicemia at the genus, species or sub-species level can be rapidly confirmed using an appropriate combination of the primer pairs listed above, which are appropriate for identification of bacteria implicated in sepsis or septicemia. In some cases, a single primer pair selected from those listed above may be sufficient for identification of a bacterium implicated in sepsis or septicemia at the genus, species or sub-species level.

[15] Also disclosed herein are methods and kits for identification of *Mycobacterium tuberculosis* and drug-resistant strains thereof. Such methods and kits may include any of primer pair numbers 3600 (SEQ ID NOs: 692:715), 3546 (SEQ ID NOs: 670:694), 3547 (SEQ ID NOs: 671:695), 3548 (SEQ ID

NOs: 672:696), 3550 (SEQ ID NOs: 673:697), 3551 (SEQ ID NOs: 674:698), 3552 (SEQ ID NOs: 675:699), 3553 (SEQ ID NOs: 676:700), 3554 (SEQ ID NOs: 677:701), 3555 (SEQ ID NOs: 678:702), 3556 (SEQ ID NOs: 679:702), 3557 (SEQ ID NOs: 680:703), 3558 (SEQ ID NOs: 681:704), 3559 (SEQ ID NOs: 682:705), 3560 (SEQ ID NOs: 683:706), 3561 (SEQ ID NOs: 684:707), 3581 (SEQ ID NOs: 685:708), 3582 (SEQ ID NOs: 686:709), 3583 (SEQ ID NOs: 687:710), 3584 (SEQ ID NOs: 688:711), 3586 (SEQ ID NOs: 689:712), 3587 (SEQ ID NOs: 690:713), 3599 (SEQ ID NOs: 691:714), and 3601 (SEQ ID NOs: 692:715). After selection of appropriate targeted whole genome amplification primers to a reference sequence of *Mycobacterium tuberculosis,* targeted whole genome amplification reactions can be performed to obtain sufficient quantities of nucleic acid such that identification of individual strains or sub-species of *Mycobacterium tuberculosis*, such as drug-resistant strains, for example, can be rapidly confirmed using an appropriate combination of the primer pairs listed above. In some cases, a single primer pair selected from those listed above may be appropriate for identification of individual strains or sub-species of *Mycobacterium tuberculosis.*

[16] Also disclosed herein are methods and kits for identification of *Staphylococcus aureus*, and drug-resistant strains thereof. Such methods and kits may include any of primer pair numbers 879 (SEQ ID NOs: 717:727), 2056 (SEQ ID NOs: 718:728), 2081 (SEQ ID NOs: 719:729), 2086 (SEQ ID NOs: 720:730), 2095 (SEQ ID NOs: 721:731), 2256 (SEQ ID NOs: 722:732), 2313 (SEQ ID NOs: 723:733), 3005 (SEQ ID NOs: 724:734), 3016 (SEQ ID NOs: 725:735), 3106 (SEQ ID NOs: 726:736), 2738 (SEQ ID NOs: 737:740), 2739 (SEQ ID NOs: 738:741), 2740 (SEQ ID NOs: 738:742) and 2741 (SEQ ID NOs: 739:740). After selection of appropriate targeted whole genome amplification primers to a reference sequence of *Mycobacterium tuberculosis,* targeted whole genome amplification reactions can be performed to obtain sufficient quantities of nucleic acid such that identification of individual strains or sub-species of *Staphylococcus aureus*, such as drug-resistant strains, for example, can be rapidly confirmed using an appropriate combination of the primer pairs listed above. In some cases, a single primer pair selected from those listed above may be appropriate for identification of individual strains or sub-species of *Staphylococcus aureus.*

[17] Also disclosed herein are methods and kits for identification of influenza viruses, and drug-resistant strains thereof. Such methods and kits may include any of primer pair numbers 1261 (SEQ ID NOs: 639:647), 1266 (SEQ ID NOs: 640:648), 1275 (SEQ ID NOs: 641:649), 1279 (SEQ ID NOs: 642:650), 1287 (SEQ ID NOs: 643:651), 2775 (SEQ ID NOs: 644:652), 2777 (SEQ ID NOs: 645:653), and 2798 (SEQ ID NOs: 646:654). After selection of appropriate targeted whole genome amplification primers to a reference sequence for an influenza virus, targeted whole genome amplification reactions can be performed to obtain sufficient quantities of nucleic acid such that identification of individual strains or sub-species of influenza viruses, such as drug-resistant strains, for example, can be rapidly confirmed using an

appropriate combination of the primer pairs listed above. In some cases, a single primer pair selected from those listed above may be appropriate for identification of individual strains or sub-species of influenza viruses.

[18] Also disclosed herein are methods and kits for identification of hepatitis C viruses, and drug-resistant strains thereof. Such methods and kits may include any of primer pair numbers 3682 (SEQ ID NOs: 655:662), 3683 (SEQ ID NOs: 656:663), 3684 (SEQ ID NOs: 657:664), 3685 (SEQ ID NOs: 658:665), 3686 (SEQ ID NOs: 658:666), 3687 (SEQ ID NOs: 659:667), 3688 (SEQ ID NOs: 660:667), 3689 (SEQ ID NOs: 660:668) and 3691 (SEQ ID NOs: 661:669). After selection of appropriate targeted whole genome amplification primers to a reference sequence for a hepatitis C virus, targeted whole genome amplification reactions can be performed to obtain sufficient quantities of nucleic acid such that identification of individual strains or sub-species of hepatitis C viruses, such as drug-resistant strains, for example, can be rapidly confirmed using an appropriate combination of the primer pairs listed above. In some cases, a single primer pair selected from those listed above may be appropriate for identification of individual strains or sub-species of hepatitis C viruses.

[19] For example, in some embodiments, the present invention provides a method comprising: amplifying at least one pathogen genome from a sample suspected of comprising at least one pathogen genome and at least one background genome using a plurality of targeted whole genome amplification primers, thereby elevating the quantity of nucleic acid representing said at least one pathogen genome relative to the quantity of nucleic acid representing said at least one background genome, wherein said plurality of targeted whole genome amplification primers is selected by one or more or each of the steps of:

i. identifying at least one pathogen genome;

ii. identifying at least one background genome;

iii. identifying a plurality of genome sequence segments having unique sequences within said pathogen genome sequence;

iv. determining frequency of occurrence of members of said plurality of genome sequence segments within said pathogen genome sequence and determining frequency of occurrence of said plurality of genome sequence segments within said background genome sequences;

v. calculating a selectivity ratio for said members by dividing said frequency of occurrence within said pathogen genome sequence by said frequency of occurrence of said plurality of genome sequence segments within said background genome sequences;

vi. selecting a selectivity ratio threshold value, thereby defining a first sub-set of said plurality of genome sequence segments having selectivity ratios equal to or greater than said selectivity ratio threshold value;

vii. determining the lengths of pathogen genome sequence occurring between genome sequence segments of said first sub-set;

viii. selecting a second sub-set of genome sequence segments from said first sub-set wherein members of said second sub-set have a mean separation distance of less than a selected length of nucleobases; and

ix. selecting targeted whole genome amplification primers that hybridize to members of said second sub-set of genome sequence segments such that, under whole genome amplification conditions, said at least one pathogen genome is amplified selectively over said at least one background genomes.

[20] In some embodiments, the method further comprises the step of producing one or more amplification products representing bioagent identifying amplicons from said amplified pathogen genome using one or more primer pairs. In some embodiments, the method further comprises the step of measuring molecular masses of said amplification products by mass spectrometry. In some embodiments, the mass spectrometry is electrospray time-of-flight mass spectrometry. In some embodiments, the method further comprises the step of comparing said molecular masses with a database comprising molecular masses of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample. In some embodiments, the method further comprising the step of calculating base compositions of said amplification products from said molecular masses. In some embodiments, the method further comprises the step of comparing said base compositions with a database comprising base compositions of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample.

[21] In some embodiments, the amplification products are generated using a plurality of primer pairs that define bioagent identifying amplicons. In some embodiments, the plurality of primer pairs are used in a multiplex reaction to generate a plurality of amplification products. In some embodiments, the plurality of primer pairs comprises at least two primer pairs from the group consisting of primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 354 (SEQ ID NOs: 597:605), 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 2249 (SEQ ID NOs: 601:609), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627). In some embodiments, the plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 3346 (SEQ ID NOs: 616:631). In some embodiments, the plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), and 3361 (SEQ ID NOs: 620:635). In some embodiments, the plurality of primer pairs comprises primer pair numbers 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604) and at least one of the primer pairs selected from the group consisting of 354 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

[22] In some embodiments, a high processivity polymerase enzyme is used at said amplification step. In some embodiments, the high processivity polymerase enzyme is a recombinant polymerase

enzyme. In some embodiments, the high processivity polymerase enzyme is a genetically engineered polymerase enzyme. In some embodiments, the high processivity polymerase enzyme is phi29.

[23] In some embodiments, the sample comprises human whole blood. In some embodiments, the method further comprises the step of extracting total nucleic acid from said sample before carrying out said amplifying step. In some embodiments, the sample comprises human buffy coat. In some embodiments, the method comprises the step of extracting total nucleic acid from said sample before carrying out said amplifying step. In some embodiments, the sample comprises human serum. In some embodiments, the method further comprises the step of extracting total nucleic acid from said sample before carrying out said amplifying step. In some embodiments, the sample comprises human hepatic cells. In some embodiments, the method further comprises the step of extracting total nucleic acid from sample before carrying out said amplifying step. In some embodiments, the sample comprises sputum. In some embodiments, the method further comprises the step of extracting total nucleic acid from sample before carrying out said amplifying step. In some embodiments, the sample comprises urine. In some embodiments, the method further comprises the step of extracting total nucleic acid from sample before carrying out said amplifying step. In some embodiments, the sample comprises biopsy tissue. In some embodiments, the method further comprises the step of extracting total nucleic acid from sample before carrying out said amplifying step.

[24] In some embodiments, the at least one pathogen is a bacterium. In some embodiments, the bacterium is one or more of (e.g., is selected from the group consisting of): *Escherichia coli, Klebsiella pneumoniae, Klebsiella oxytoca, Serratia marcescens, Enterobacter cloacae, Enterobacter aerogenes, Proteus mirabilis, Pseudomonas aeruginosa, Acinetobacter baumannii, Stenotrophomonas maltophilia, Staphylococcus aureus, Staphylococcus epidermidis, Staphylococcus haemolyticus, Streptococcus pneumoniae, Streptococcus pyogenes, Streptococcus agalactiae, Streptococcus mitis, Enterococcus faecium, Enterococcus faecalis, Candida albicans, Candida tropicalis, Candida parapsilosis, Candida krusei, Candida glabrata, Mycobacterium tuberculosis,* and *Aspergillus fumigatus.*

[25] In some embodiments, the at least one background genome comprises a human nucleic acid. In some embodiments, the said identifying step indicates the presence of bacterial sepsis in a human patient. In some embodiments, the said identifying step indicates the presence of bacteremia in a human patient.

[26] In some embodiments, the at least one pathogen is a virus. In some embodiments, the virus is HIV. In some embodiments, the virus is HCV. In some embodiments, the virus is influenza virus.

[27] The present invention also provides a method comprising one or more of, or each of, the steps of:

a. extracting nucleic acids from a sample; and

b. mixing said nucleic acids with a plurality of targeted whole genome amplification primers, a high processivity polymerase enzyme to produce an amplification mixture, wherein said plurality of targeted whole genome amplification primers is selected by:

i. identifying at least one target genome suspected of being present in said sample;

ii. identifying at least one background genome suspected of being present in said sample;

iii. identifying a plurality of genome sequence segments having unique sequences within said target genome sequence;

iv. determining frequency of occurrence of members of said plurality of genome sequence segments within said target genome sequence and within said background genome sequences;

v. calculating a selectivity ratio for said members by dividing said frequency of occurrence within said target genome by said frequency of occurrence of said plurality of genome sequence segments within said background genome sequences;

vi. selecting a selectivity ratio threshold value, thereby defining a first sub-set of said plurality of genome sequence segments having selectivity ratios equal to or greater than said selectivity ratio threshold value;

vii. determining the lengths of target genome sequence occurring between genome sequence segments of said first sub-set;

viii. selecting a second sub-set of genome sequence segments from said first sub-set wherein members of said second sub-set have a mean separation of less than a selected length of nucleobases; and

ix. selecting targeted whole genome amplification primers that hybridize to members of said second sub-set of genome sequence segments such that said at least one target genome is amplified selectively over said at least one background genome.

[28] In some embodiments, the method further comprises the step of amplifying one or more of said extracted nucleic acids in said mixture of step b. In some embodiments, the amplifying step is a targeted whole genome amplification reaction. In some embodiments, the method further comprises the step of performing a second amplification step using at least one primer pair that defines a bioagent identifying amplicon to obtain at least a second amplification product. In some embodiments, the method further comprises the step of measuring the molecular mass of said second amplification product by mass spectrometry. In some embodiments, the mass spectrometry is electrospray time-of-flight mass spectrometry.

[29] In some embodiments, the method further comprises the step of comparing said molecular mass with a database comprising molecular masses of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample. In some embodiments, the method further comprises the step of calculating a base composition of said amplification products from said molecular mass. In some embodiments, the method further comprises the step of comparing said base compositions with a database comprising base compositions of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample.

[30] In some embodiments, the second amplification step comprises obtaining a plurality of amplification products generated using a plurality of primer pairs that define bioagent identifying

amplicons. In some embodiments, the plurality of primer pairs is used in one or more multiplex reactions to generate a plurality of amplification products. In some embodiments, the plurality of primer pairs comprises at least two primer pairs from the group consisting of primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 354 (SEQ ID NOs: 597:605), 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 2249 (SEQ ID NOs: 601:609), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627). In some embodiments, the plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 3346 (SEQ ID NOs: 616:631). In some embodiments, the plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), and 3361 (SEQ ID NOs: 620:635). In some embodiments, the plurality of primer pairs comprises primer pair numbers 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604) and at least one of the primer pairs selected from the group consisting of 354 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

[31] In some embodiments, a high processivity polymerase enzyme is used at said amplification step. In some embodiments, the high processivity polymerase enzyme is a recombinant polymerase enzyme. In some embodiments, the high processivity polymerase enzyme is a genetically engineered polymerase enzyme. In some embodiments, the high processivity polymerase enzyme is phi29.

[32] In some embodiments, the sample comprises human whole blood. In some embodiments, the method further comprises the step of extracting total nucleic acid from said sample before carrying out said amplifying step. In some embodiments, the sample comprises human buffy coat. In some embodiments, the method comprises the step of extracting total nucleic acid from said sample before carrying out said amplifying step. In some embodiments, the sample comprises human serum. In some embodiments, the method further comprises the step of extracting total nucleic acid from said sample before carrying out said amplifying step. In some embodiments, the sample comprises human hepatic cells. In some embodiments, the method further comprises the step of extracting total nucleic acid from sample before carrying out said amplifying step. In some embodiments, the sample comprises sputum. In some embodiments, the method further comprises the step of extracting total nucleic acid from sample before carrying out said amplifying step. In some embodiments, the sample comprises urine. In some embodiments, the method further comprises the step of extracting total nucleic acid from sample before carrying out said amplifying step. In some embodiments, the sample comprises biopsy tissue. In some embodiments, the method further comprisies the step of extracting total nucleic acid from sample before carrying out said amplifying step.

[33] In some embodiments, the at least one pathogen is a bacterium. In some embodiments, the bacterium is one or more of (e.g., is selected from the group consisting of): *Escherichia coli, Klebsiella*

pneumoniae, Klebsiella oxytoca, Serratia marcescens, Enterobacter cloacae, Enterobacter aerogenes, Proteus mirabilis, Pseudomonas aeruginosa, Acinetobacter baumannii, Stenotrophomonas maltophilia, Staphylococcus aureus, Staphylococcus epidermidis, Staphylococcus haemolyticus, Streptococcus pneumoniae, Streptococcus pyogenes, Streptococcus agalactiae, Streptococcus mitis, Enterococcus faecium, Enterococcus faecalis, Candida albicans, Candida tropicalis, Candida parapsilosis, Candida krusei, Candida glabrata, Mycobacterium tuberculosis, and *Aspergillus fumigatus.*

[34] In some embodiments, the at least one background genome comprises a human nucleic acid. In some embodiments, the said identifying step indicates the presence of bacterial sepsis in a human patient. In some embodiments, the said identifying step indicates the presence of bacteremia in a human patient.

[35] In some embodiments, the at least one pathogen is a virus. In some embodiments, the virus is HIV. In some embodiments, the virus is HCV. In some embodiments, the virus is influenza virus.

[36] The present invention also provides kits containing one or more components necessary for, useful for, or sufficient for performing any of the methods described above or elsewhere herein. In some embodiments, the kit comprises a high processivity polymerase enzyme and a plurality of purified targeted whole genome amplification primers. In some embodiments, the kit further comprises at least one primer pair that defines a bioagent identifying amplicon. In some embodiments, the plurality of primer pairs comprises at least two primer pairs from the group consisting of primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 354 (SEQ ID NOs: 597:605), 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 2249 (SEQ ID NOs: 601:609), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627). In some embodiments, the plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 3346 (SEQ ID NOs: 616:631). In some embodiments, the plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), and 3361 (SEQ ID NOs: 620:635). In some embodiments, the plurality of primer pairs comprises primer pair numbers 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604) and at least one of the primer pairs selected from the group consisting of 354 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627). In some embodiments, the high processivity enzyme is phi29.

BRIEF DESCRIPTION OF THE DRAWINGS

[37] Figure 1 is a plot indicating the relationships between sensitivity, selectivity and length of the genome sequence segments and primers hybridizing thereto.

[38] Figure 2 is a process diagram indicating the process steps for selection of genome sequence segments and primers hybridizing thereto.

[39] Figure 3A is a plot indicating the quantities of human DNA obtained from whole genome amplification (WGA) reactions performed with random hexamer primers (solid diamond) and the targeted whole genome amplification (TWGA) method using the primers of Table 3 (clear circle).

[40] Figure 3B is a plot indicating the quantity of *Bacillus anthracis* DNA obtained from whole genome amplification (WGA) reactions performed with random hexamer primers (solid diamond) and targeted whole genome amplification (TWGA) method using the primers of Table 3 (clear circle).

[41] Figure 4A is a plot indicating the quantities of human DNA obtained from whole genome amplification (WGA) reactions performed with random hexamer primers (solid diamond) and the targeted whole genome amplification (TWGA) method using the first generation primers of Table 3 (clear circle) and the second generation primers of Table 4 (clear square).

[42] Figure 4B is a plot indicating the quantity of *Bacillus anthracis* DNA obtained from whole genome amplification (WGA) reactions performed with random hexamer primers (solid diamond) and targeted whole genome amplification (TWGA) method using the primers of Table 3 (clear circle) and the second generation primers of Table 4 (clear square).

[43] Figures 5A and 5B are plots indicating the quantities of *Bacillus anthracis* DNA (target genome) and *Homo sapiens* DNA (background genome) obtained in targeted whole genome amplification reactions with the indicated quantity of background DNA and 200 femtograms (fg) of *Bacillus anthracis* DNA.

[44] Figures 6A and 6B are plots comparing the quantities of *Bacillus anthracis* DNA (target genome) and *Homo sapiens* DNA (background genome) obtained in a targeted whole genome amplification reaction (Figure 6A) vs. a conventional whole genome amplification reaction (Figure 6B).

[45] Figures 7A and 7B are plots of quantity of amplified DNA obtained in a range of concentrations of *Bacillus anthracis* DNA (target genome) with a constant concentration of *Homo sapiens* DNA (background genome). Figure 7A indicates the quantities of *Bacillus anthracis* DNA obtained in two different targeted whole genome amplification reactions and in a conventional whole genome amplification reaction. Figure 7B indicates the quantities of *Homo sapiens* DNA in the same three reactions.

[46] Figure 8 is a process diagram illustrating a representative primer pair selection process.

[47] Figure 9 is a process diagram illustrating an embodiment of the calibration method.

DEFINITIONS

[48] To facilitate an understanding of the methods disclosed herein, a number of terms and phrases are defined below:

[49] As used herein, the term "abundance" refers to an amount. The amount may be described in terms of concentration which are common in molecular biology such as "copy number," "pfu or plate-forming unit" which are well known to those with ordinary skill. Concentration may be relative to a known standard or may be absolute.

[50] The term "amplification," as used herein, refers to a process of multiplying an original quantity of a nucleic acid template in order to obtain greater quantities of the original nucleic acid.

[51] As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" also applies to the term "sample template."

[52] As used herein, the term "amplification reagents" refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification, excluding primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, micro-well, or other vessel).

[53] As used herein, the term "analogous" when used in context of comparison of bioagent identifying amplicons indicates that the bioagent identifying amplicons being compared are produced with the same pair of primers. For example, bioagent identifying amplicon "A" and bioagent identifying amplicon "B", produced with the same pair of primers are analogous with respect to each other. Bioagent identifying amplicon "C", produced with a different pair of primers is not analogous to either bioagent identifying amplicon "A" or bioagent identifying amplicon "B".

[54] As used herein, the term "anion exchange functional group" refers to a positively charged functional group capable of binding an anion through an electrostatic interaction. The most well known anion exchange functional groups are the amines, including primary, secondary, tertiary and quaternary amines.

[55] The term "background organisms," as used herein, refers to organisms typically present in a given sample which are not of interest and are thus considered to be contaminants.

[56] The term "background genome," as used herein refers to the genome of a background organism. Background organisms will vary according to the sample source. In a non-limiting example, for targeted whole genome amplification of a soil bioremediation bacterium in a soil sample, it would be advantageous to define the genomes of organisms native to soil such as *C. elegans*, as background genomes. In another non-limiting example, for whole genome amplification of a genome belonging to a target pathogen in a human tissue sample, it would be advantageous to define human DNA as a background genome.

[57] The term "bacteria" or "bacterium" refers to any member of the groups of eubacteria and archaebacteria.

[58] The term "bacteremia" refers to the presence of bacteria in the bloodstream. It is also known by the related terms "blood poisoning" or "toxemia." In the hospital, indwelling catheters are a frequent cause of bacteremia and subsequent nosocomial infections, because they provide a means by which bacteria normally found on the skin can enter the bloodstream. Other causes of bacteremia include dental procedures (occasionally including simple tooth brushing), herpes (including herpetic whitlow), urinary tract infections, intravenous drug use, and colorectal cancer. Bacteremia may also be seen in oropharyngeal, gastrointestinal or genitourinary surgery or exploration.

[59] As used herein, a "base composition" is the exact number of each nucleobase (for example, A, T, C and G) in a segment of nucleic acid. For example, amplification of nucleic acid of strain 5170 of *Mycobacterium tuberculosis* using primer pair number 3550 (SEQ ID NOs: 673:697) produces an amplification product 129 nucleobases in length from nucleic acid of the embB gene that has a base composition of A21 G37 C44 T27 (by convention - with reference to the sense strand of the amplification product). Because the molecular masses of each of the four natural nucleotides and chemical modifications thereof are known (if applicable), a measured molecular mass can be deconvoluted to a list of possible base compositions. Identification of a base composition of a sense strand which is complementary to the corresponding antisense strand in terms of base composition provides a confirmation of the true base composition of an unknown amplification product. For example, the base composition of the antisense strand of the 129 nucleobase amplification product described above is A27 G44 C37 T21.

[60] As used herein, a "base composition probability cloud" is a representation of the diversity in base composition resulting from a variation in sequence that occurs among different isolates of a given

species. The "base composition probability cloud" represents the base composition constraints for each species and is typically visualized using a pseudo four-dimensional plot.

[61] As used herein, 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. As used herein, a "pathogen" is a bioagent which causes a disease or disorder. A pathogen that infects a human is known as a "human pathogen." Non-human pathogens may infect specific animals but not humans. Human pathogens are of interest for clinical reasons and non-human pathogen identification is of interest in veterinary applications of the methods disclosed herein.

[62] As used herein, a "bioagent division" is defined as group of bioagents above the species level and includes but is not limited to, orders, families, classes, clades, genera or other such groupings of bioagents above the species level.

[63] As used herein, the term "bioagent identifying amplicon" refers to a polynucleotide that is amplified from nucleic acid of a bioagent in an amplification reaction and which 1) provides sufficient variability to distinguish among bioagents from whose nucleic acid the bioagent identifying amplicon is produced and 2) whose molecular mass is amenable to a rapid and convenient molecular mass determination modality such as mass spectrometry, for example. *In silico* representations of bioagent identifying amplicons are particularly useful for inclusion in databases used for identification of bioagents. Bioagent identifying amplicons are defined by a pair of primers that hybridize to regions of nucleic acid of a given bioagent.

[64] As used herein, the term "biological product" refers to any product originating from an organism. Biological products are often products of processes of biotechnology. Examples of biological products include, but are not limited to: cultured cell lines, cellular components, antibodies, proteins and other cell-derived biomolecules, growth media, growth harvest fluids, natural products and bio-pharmaceutical products.

[65] The terms "biowarfare agent" and "bioweapon" are synonymous and refer to a bacterium, virus, fungus or protozoan that could be deployed as a weapon to cause bodily harm to individuals. Military or terrorist groups may be implicated in deployment of biowarfare agents.

[66] As used herein, the term "broad range survey primer pair" refers to a primer pair designed to produce bioagent identifying amplicons across different broad groupings of bioagents. For example, the ribosomal RNA-targeted primer pairs are broad range survey primer pairs which have the capability of producing bacterial bioagent identifying amplicons for essentially all known bacteria. With respect to broad range primer pairs employed for identification of bacteria, a broad range survey primer pair for bacteria such as 16S rRNA primer pair number 346 (SEQ ID NOs: 594:602) for example, will produce an bacterial bioagent identifying amplicon for essentially all known bacteria.

[67] The term "calibration amplicon" refers to a nucleic acid segment representing an amplification product obtained by amplification of a calibration sequence with a pair of primers designed to produce a bioagent identifying amplicon.

[68] The term "calibration sequence" refers to a polynucleotide sequence to which a given pair of primers hybridizes for the purpose of producing an internal (i.e.: included in the reaction) calibration standard amplification product for use in determining the quantity of a bioagent in a sample. The calibration sequence may be expressly added to an amplification reaction, or may already be present in the sample prior to analysis.

[69] The term "clade primer pair" refers to a primer pair designed to produce bioagent identifying amplicons for species belonging to a clade group. A clade primer pair may also be considered as a "speciating" primer pair which is useful for distinguishing among closely related species.

[70] The term "codon" refers to a set of three adjoined nucleotides (triplet) that codes for an amino acid or a termination signal.

[71] As used herein, the term "codon base composition analysis," refers to determination of the base composition of an individual codon by obtaining a bioagent identifying amplicon that includes the codon. The bioagent identifying amplicon will at least include regions of the target nucleic acid sequence to which the primers hybridize for generation of the bioagent identifying amplicon as well as the codon being analyzed, located between the two primer hybridization regions.

[72] As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (i.e., a sequence of nucleotides such as an oligonucleotide or a target nucleic acid) related by the base-pairing rules. For example, the sequence 5'-A-G-T-3', is complementary to the sequence 3'-T-C-A-5'. Complementarity may be "partial," in which only some of the nucleic acids' bases are matched

according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids. Either term may also be used in reference to individual nucleotides, especially within the context of polynucleotides. For example, a particular nucleotide within an oligonucleotide may be noted for its complementarity, or lack thereof, to a nucleotide within another nucleic acid strand, in contrast or comparison to the complementarity between the rest of the oligonucleotide and the nucleic acid strand. But in this sense, complementarity either exists or does not exist i.e.: there is no partial complementarity.

[73] The term "complement of a nucleic acid sequence" as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in "antiparallel association." Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids disclosed herein and include, for example, inosine and 7-deazaguanine. Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs. Where a first oligonucleotide is complementary to a region of a target nucleic acid and a second oligonucleotide has complementary to the same region (or a portion of this region) a "region of overlap" exists along the target nucleic acid. The degree of overlap will vary depending upon the extent of the complementarity.

[74] The term "degenerate primers," as used herein refers to a mixture of similar, but not identical, primers having one or more residues substituted relative to the other primer(s) in the mixture. Degenerate nucleotide codes include R, K, S, Y, M, W, B, H, N, D, V and I. The corresponding combinations are listed in 37 CFR §1.821. For example, the sequence AAATTT**R**CCCGGG (SEQ ID NO: 2) actually refers to a combination of primers having the following sequences: AAATTT**A**CCCGGG (SEQ ID NO: 3), and AAATTT**G**CCCGGG (SEQ ID NO: 4) because R = A or G.

[75] As used herein, the term "division-wide primer pair" refers to a primer pair designed to produce bioagent identifying amplicons within sections of a broader spectrum of bioagents For example, primer pair number 354 (SEQ ID NOs: 597:605), a division-wide primer pair, is designed to produce bacterial bioagent identifying amplicons for members of the *Bacillus* group of bacteria which comprises, for example, members of the genera *Streptococcus*, *Enterococcus*, and *Staphylococcus*. Other division-wide

primer pairs may be used to produce bacterial bioagent identifying amplicons for other groups of bacterial bioagents.

[76] As used herein, the term "concurrently amplifying" used with respect to more than one amplification reaction refers to the act of simultaneously amplifying more than one nucleic acid in a single reaction mixture.

[77] As used herein, the term "drill-down primer pair" refers to a primer pair designed to produce bioagent identifying amplicons for identification of sub-species characteristics or confirmation of a species assignment. For example, primer pair number 897 (SEQ ID NOs: 717:727), a drill-down *Staphylococcus aureus* genotyping primer pair, is designed to produce *Staphylococcus aureus* genotyping amplicons. Other drill-down primer pairs may be used to produce bioagent identifying amplicons for *Staphylococcus aureus* and other bacterial species.

[78] The term "duplex" refers to the state of nucleic acids in which the base portions of the nucleotides on one strand are bound through hydrogen bonding the their complementary bases arrayed on a second strand. The condition of being in a duplex form reflects on the state of the bases of a nucleic acid. By virtue of base pairing, the strands of nucleic acid also generally assume the tertiary structure of a double helix, having a major and a minor groove. The assumption of the helical form is implicit in the act of becoming duplexed.

[79] As used herein, the term "etiology" refers to the causes or origins, of diseases or abnormal physiological conditions.

[80] The term "frequency of occurrence" as used herein, refers to the number of different coordinates where a given genome sequence segment occurs within a given genome. The frequency of occurrence of a given genome sequence segment provides a means of defining the sensitivity of a primer designed to hybridize to the genome sequence segment. The frequency of occurrence of a given genome sequence segment is also used in the calculation of selectivity ratios.

[81] The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for the production of an RNA having a non-coding function (e.g., a ribosomal or transfer RNA), a polypeptide or a precursor. The RNA or polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or function is retained.

[82] The term "genome," as used herein, generally refers to the complete set of genetic information in the form of one or more nucleic acid sequences, including text or *in silico* representations thereof. A genome may include either DNA or RNA, depending upon its organism of origin. Most organisms have DNA genomes while some viruses have RNA genomes. As used herein, the term "genome" need not comprise the complete set of genetic information. The term may also refer to at least a majority portion of a genome such as at least 50% to 100% of an entire genome or any whole or fractional percentage therebetween.

[83] The term "genome sequence segment," as used herein, refers to a portion of a genome sequence which is initially defined as a primer hybridization candidate for the purpose of the targeted whole genome amplification methods disclosed herein. The related term "unique genome sequence segment" refers to a genome sequence segment that occurs at least once in a given genome. For example, a simplified hypothetical 8 nucleobase genome consisting of the following sequence: aattccgg (SEQ ID NO: 5) has four unique genome sequence segments of five nucleobase lengths (aattc (SEQ ID NO: 6); attcc (SEQ ID NO: 7); ttccg (SEQ ID NO: 8); and tccgg (SEQ ID NO: 9)). This same simplified hypothetical 8 nucleobase genome also has three unique genome sequence segments of six nucleobase lengths: (aattcc (SEQ ID NO: 10); attccg (SEQ ID NO: 11); and ttccgg (SEQ ID NO: 12)). This same simplified hypothetical 8 nucleobase genome also has two unique genome sequence segments of seven nucleobase lengths: (aattccg (SEQ ID NO: 13); and attccgg (SEQ ID NO: 14)). This same simplified hypothetical 8 nucleobase genome also has one unique genome sequence segment which is 8 nucleobases in length : (aattccgg (SEQ ID NO: 5). In another example, a simplified hypothetical 8 nucleobase genome consisting of the following sequence: aaaaaaaa (SEQ ID NO: 15) obviously only has a single unique genome sequence segment which is five nucleobases in length (occurring 4 times), as well as a single unique genome sequence segment which is six nucleobases in length (occurring 3 times), a single unique genome sequence segment which is seven nucleobases in length (occurring twice) and a single unique genome sequence segment which is eight nucleobases in length (occurring once).

[84] The term "genotype," as used herein, refers to the genetic makeup of an organism. Members of the same species of organism having genetic differences are said to have different genotypes.

[85] The terms "homology," "homologous" and "sequence identity" refer to a degree of identity. There may be partial homology or complete homology. A partially homologous sequence is one that is less than 100% identical to another sequence. 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 a primer 20 nucleobases in length would have 15/20 = 0.75 or 75% sequence identity with the 20 nucleobase primer. As used herein, sequence identity is meant to be properly determined when the query sequence and the subject sequence are both described and aligned in the 5' to 3' direction. Sequence alignment algorithms such as BLAST, will return results in two different alignment orientations. In the Plus/Plus orientation, both the query sequence and the subject sequence are aligned in the 5' to 3' direction. On the other hand, in the Plus/Minus orientation, the query sequence is in the 5' to 3' direction while the subject sequence is in the 3' to 5' direction. It should be understood that with respect to the primers disclosed herein, sequence identity is properly determined when the alignment is designated as Plus/Plus. Sequence identity may also encompass alternate or modified nucleobases that perform in a functionally similar manner to the regular nucleobases adenine, thymine, guanine and cytosine with respect to hybridization and primer extension in amplification reactions. In a non-limiting example, if the 5-propynyl pyrimidines propyne C and/or propyne T replace one or more C or T residues in one primer which is otherwise identical to another primer in sequence and length, the two primers will have 100% sequence identity with each other. In another non-limiting example, Inosine (I) may be used as a replacement for G or T and effectively hybridize to C, A or U (uracil). Thus, if inosine replaces one or more C, A or U residues in one primer which is otherwise identical to another primer in sequence and length, the two primers will have 100% sequence identity with each other. Other such modified or universal bases may exist which would perform in a functionally similar manner for hybridization and amplification reactions and will be understood to fall within this definition of sequence identity.

[86] As used herein, "housekeeping gene" refers to a gene encoding a protein or RNA involved in basic functions required for survival and reproduction of a bioagent. Housekeeping genes include, but are not limited to genes encoding RNA or proteins involved in translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like.

[87] The term "hybridization," as used herein refers to the process of joining two complementary strands of DNA or one each of DNA and RNA to form a double-stranded molecule.

[88] The term "*in silico*" refers to processes taking place via computer calculations. For example, electronic PCR (ePCR) is a process analogous to ordinary PCR except that it is carried out using nucleic acid sequences and primer pair sequences stored on a computer formatted medium.

[89] The term "*in vitro* method," as used herein, describes a biochemical process performed in a test-tube or other laboratory apparatus. An amplification reaction performed on a nucleic acid sample in a microtube or a well of a multi-well plate is an example of an *in vitro* method.

[90] The "ligase chain reaction" (LCR; sometimes referred to as "Ligase Amplification Reaction" (LAR) described by Barany, Proc. Natl. Acad. Sci., 88:189 (1991); Barany, PCR Methods and Applic., 1:5 (1991); and Wu and Wallace, Genomics 4:560 (1989) has developed into a well-recognized alternative method for amplifying nucleic acids. In LCR, four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, that hybridize to the opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation, hybridization and ligation amplify a short segment of DNA. LCR has also been used in combination with PCR to achieve enhanced detection of single-base changes. However, because the four oligonucleotides used in this assay can pair to form two short ligatable fragments, there is the potential for the generation of target-independent background signal. The use of LCR for mutant screening is limited to the examination of specific nucleic acid positions.

[91] The term "locked nucleic acid" or "LNA" refers to a nucleic acid analogue containing one or more 2'-O, 4'-C-methylene-β-D-ribofuranosyl nucleotide monomers in an RNA mimicking sugar conformation. LNA oligonucleotides display unprecedented hybridization affinity toward complementary single-stranded RNA and complementary single- or double-stranded DNA. LNA oligonucleotides induce A-type (RNA-like) duplex conformations. The primers disclosed herein may contain LNA modifications.

[92] As used herein, the term "mass-modifying tag" refers to any modification to a given nucleotide which results in an increase in mass relative to the analogous non-mass modified nucleotide. Mass-modifying tags can include heavy isotopes of one or more elements included in the nucleotide such as carbon-13 for example. Other possible modifications include addition of substituents such as iodine or bromine at the 5 position of the nucleobase for example.

[93] The term "mass spectrometry" refers to measurement of the mass of atoms or molecules. The molecules are first converted to ions, which are separated using electric or magnetic fields according to the ratio of their mass to electric charge. The measured masses are used to identity the molecules.

[94] The term "mean" as used herein refers to the arithmetic average; the sum of the data divided by the sample size.

[95] The term "microorganism" as used herein means an organism too small to be observed with the unaided eye and includes, but is not limited to bacteria, virus, protozoans, fungi; and ciliates.

[96] The term "multi-drug resistant" or multiple-drug resistant" refers to a microorganism which is resistant to more than one of the antibiotics or antimicrobial agents used in the treatment of said microorganism.

[97] The term "multiplex PCR" refers to a PCR reaction where more than one primer set is included in the reaction pool allowing 2 or more different DNA targets to be amplified by PCR in a single reaction tube.

[98] The term "non-template tag" refers to a stretch of at least three guanine or cytosine nucleobases of a primer used to produce a bioagent identifying amplicon which are not complementary to the template. A non-template tag is incorporated into a primer for the purpose of increasing the primer-duplex stability of later cycles of amplification by incorporation of extra G-C pairs which each have one additional hydrogen bond relative to an A-T pair.

[99] The term "nucleic acid sequence" as used herein refers to the linear composition of the nucleic acid residues A, T, C or G or any modifications thereof, within an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single or double stranded, and represent the sense or antisense strand

[100] As used herein, the term "nucleobase" is synonymous with other terms in use in the art including "nucleotide," "deoxynucleotide," "nucleotide residue," "deoxynucleotide residue," "nucleotide triphosphate (NTP)," or deoxynucleotide triphosphate (dNTP).

[101] The term "nucleotide analog" as used herein refers to modified or non-naturally occurring nucleotides such as 5-propynyl pyrimidines (i.e., 5-propynyl-dTTP and 5-propynyl-dTCP), 7-deaza purines (i.e., 7-deaza-dATP and 7-deaza-dGTP). Nucleotide analogs include base analogs and comprise modified forms of deoxyribonucleotides as well as ribonucleotides.

[102] The term "oligonucleotide" as used herein is defined as a molecule comprising two or more deoxyribonucleotides or ribonucleotides, preferably at least 5 nucleotides, more preferably at least about 13 to 35 nucleotides. The exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, PCR, or a combination thereof. Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a

phosphodiester linkage, an end of an oligonucleotide is referred to as the "5'-end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3'-end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. A first region along a nucleic acid strand is said to be upstream of another region if the 3' end of the first region is before the 5' end of the second region when moving along a strand of nucleic acid in a 5' to 3' direction. All oligonucleotide primers disclosed herein are understood to be presented in the 5' to 3' direction when reading left to right. When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the "upstream" oligonucleotide and the latter the "downstream" oligonucleotide. Similarly, when two overlapping oligonucleotides are hybridized to the same linear complementary nucleic acid sequence, with the first oligonucleotide positioned such that its 5' end is upstream of the 5' end of the second oligonucleotide, and the 3' end of the first oligonucleotide is upstream of the 3' end of the second oligonucleotide, the first oligonucleotide may be called the "upstream" oligonucleotide and the second oligonucleotide may be called the "downstream" oligonucleotide.

[103] The term "organism," as used herein, refers to humans, animals, plants, protozoa, bacteria, fungi and viruses.

[104] As used herein, a "pathogen" is a bioagent which causes a disease or disorder.

[105] As used herein, the terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

[106] The term "peptide nucleic acid" ("PNA") as used herein refers to a molecule comprising bases or base analogs such as would be found in natural nucleic acid, but attached to a peptide backbone rather than the sugar-phosphate backbone typical of nucleic acids. The attachment of the bases to the peptide is such as to allow the bases to base pair with complementary bases of nucleic acid in a manner similar to that of an oligonucleotide. These small molecules, also designated anti gene agents, stop transcript elongation by binding to their complementary strand of nucleic acid (Nielsen, et al. Anticancer Drug Des. 1993, 8, 53-63). The primers disclosed herein may comprise PNAs.

[107] The term "polymerase" refers to an enzyme having the ability to synthesize a complementary strand of nucleic acid from a starting template nucleic acid strand and free dNTPs.

[108] As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, hereby incorporated by reference, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified." With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of 32P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

[109] The term "polymerization means" or "polymerization agent" refers to any agent capable of facilitating the addition of nucleoside triphosphates to an oligonucleotide. Preferred polymerization means comprise DNA and RNA polymerases.

[110] The term "primer," as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double

stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer, use of the method, and the parameters used for primer design, as disclosed herein. Primers disclosed herein fall into two general categories; (i) primer pairs, generally ranging in length from about 12 to about 35 nucleobases in length, that define bioagent identifying amplicons which are useful for preparing amplification products corresponding to bioagent identifying amplicons; and (ii) targeted whole genome amplification primers which are designed to hybridize at positions across essentially the entire genome of a bioagent of interest. Targeted whole genome amplification primers are not matched up in pairs and are typically of lengths ranging from about 5 to about 13 nucleobases in length.

[111] As used herein, the terms "pair of primers," or "primer pair" are synonymous. A primer pair is used for amplification of a nucleic acid sequence. A pair of primers comprises a forward primer and a reverse primer. The forward primer hybridizes to a sense strand of a target gene sequence to be amplified and primes synthesis of an antisense strand (complementary to the sense strand) using the target sequence as a template. A reverse primer hybridizes to the antisense strand of a target gene sequence to be amplified and primes synthesis of a sense strand (complementary to the antisense strand) using the target sequence as a template.

[112] The primer pairs are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. In some embodiments, the highly conserved sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity, or between about 99-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus design of the primers requires selection of a variable region with appropriate variability to resolve the identity of a given bioagent. Bioagent identifying amplicons are ideally specific to the identity of the bioagent.

[113] Properties of the primers may include any number of properties related to structure including, but not limited to: nucleobase length which may be contiguous (linked together) or non-contiguous (for example, two or more contiguous segments which are joined by a linker or loop moiety), modified or universal nucleobases (used for specific purposes such as for example, increasing hybridization affinity, preventing non-templated adenylation and modifying molecular mass) percent complementarity to a given target sequences.

[114] Properties of the primers also include functional features including, but not limited to, orientation of hybridization (forward or reverse) relative to a nucleic acid template. The coding or sense strand is the strand to which the forward priming primer hybridizes (forward priming orientation) while the reverse priming primer hybridizes to the non-coding or antisense strand (reverse priming orientation). The functional properties of a given primer pair also include the generic template nucleic acid to which the primer pair hybridizes. For example, in the case of primer pairs, identification of bioagents can be accomplished at different levels using primers suited to resolution of each individual level of identification. Broad range survey primers are designed with the objective of identifying a bioagent as a member of a particular division (e.g., an order, family, genus or other such grouping of bioagents above the species level of bioagents). In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species or sub-species level. Other primers may have the functionality of producing bioagent identifying amplicons for members of a given taxonomic genus, clade, species, sub-species or genotype (including genetic variants which may include presence of virulence genes or antibiotic resistance genes or mutations). Additional functional properties of primer pairs include the functionality of performing amplification either singly (single primer pair per amplification reaction vessel) or in a multiplex fashion (multiple primer pairs and multiple amplification reactions within a single reaction vessel).

[115] The term “processivity,” as used herein, refers to the ability of an enzyme to repetitively continue its catalytic function without dissociating from its substrate. For example, Phi29 polymerase is a highly processive polymerase due to its tight binding of the template DNA substrate.

[116] As used herein, the terms "purified" or "substantially purified" refer to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated. An "isolated polynucleotide" or "isolated oligonucleotide" is therefore a substantially purified polynucleotide.

[117] The term "reverse transcriptase" refers to an enzyme having the ability to transcribe DNA from an RNA template. This enzymatic activity is known as reverse transcriptase activity. Reverse transcriptase activity is desirable in order to obtain DNA from RNA viruses which can then be amplified and analyzed by the methods disclosed herein.

[118] The term "ribosomal RNA” or "rRNA" refers to the primary ribonucleic acid constituent of ribosomes. Ribosomes are the protein-manufacturing organelles of cells and exist in the cytoplasm. Ribosomal RNAs are transcribed from the DNA genes encoding them.

[119] The term "sample" in the present specification and claims is used in its broadest sense. On the one hand it is meant to include a specimen or culture (e.g., microbiological cultures). On the other hand, it is meant to include both biological and environmental samples. A sample may include a specimen of synthetic origin. Biological samples may be animal, including human, fluid, solid (e.g., stool) or tissue, as well as liquid and solid food and feed products and ingredients such as dairy items, vegetables, meat and meat by-products, and waste. Biological samples may be obtained from all of the various families of domestic animals, as well as feral or wild animals, including, but not limited to, such animals as ungulates, bear, fish, lagamorphs, rodents, etc. Environmental samples include environmental material such as surface matter, soil, water, air and industrial samples, as well as samples obtained from food and dairy processing instruments, apparatus, equipment, utensils, disposable and non-disposable items. These examples are not to be construed as limiting the sample types applicable to the methods disclosed herein. The term "source of target nucleic acid" refers to any sample that contains nucleic acids (RNA or DNA). Particularly preferred sources of nucleic acids are biological samples including, but not limited to blood, saliva, urine, cerebral spinal fluid, pleural fluid, milk, lymph, sputum and semen. In particular, different fractions of blood samples exist such as serum or plasma (the liquid component of blood which contains various vital proteins), and buffy coat (a centrifuged fraction of blood that contains white blood cells and platelets). Other preferred sources of nucleic acids are specific cell types such as, hepatic cells for example. Other preferred sources of nucleic acids are tissue biopsies. Methods of handing such samples are well within the technical skill of an ordinary practitioner in the art.

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

[121] A "segment" is defined herein as a region of nucleic acid within a nucleic acid sequence.

[122] The term "selectivity," as used herein, is a measure which indicates the frequency of occurrence of a given genome sequence segment in a target relative to the frequency of occurrence of the same genome sequence segment in background genomes. The related term "selectivity ratio," as used herein, is a number calculated by dividing the frequency of occurrence of a given genome sequence segment in a target genome by its frequency of occurrence in background genomes.

[123] The "self-sustained sequence replication reaction" (3SR) (Guatelli et al., Proc. Natl. Acad. Sci. 1990, 87:1874-1878, with an erratum at Proc. Natl. Acad. Sci. 1990, 87:7797) is a transcription-based in vitro amplification system (Kwok et al., Proc. Natl. Acad. Sci. 1989, 86:1173-1177) that can exponentially amplify RNA sequences at a uniform temperature. The amplified RNA can then be utilized for mutation detection (Fahy et al., 1991, PCR Meth. Appl., 1:25-33). In this method, an oligonucleotide primer is used to add a phage RNA polymerase promoter to the 5′ end of the sequence of interest. In a cocktail of enzymes and substrates that includes a second primer, reverse transcriptase, RNase H, RNA polymerase and ribo- and deoxyribonucleoside triphosphates, the target sequence undergoes repeated rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the area of interest. The use of 3SR to detect mutations is kinetically limited to screening small segments of DNA (e.g., 200-300 base pairs).

[124] As used herein, the term "sequence alignment" refers to a listing of multiple DNA or amino acid sequences and aligns them to highlight their similarities. The listings can be made using bioinformatics computer programs.

[125] The term "sensitivity," as used herein, is a measure which indicates the frequency of occurrence of a given genome sequence segment within a target genome.

[126] The term "separation distance," as used herein, refers to the intervening distance along a given genome sequence between two genome sequence segments chosen as primer hybridization sites. For example, a first genome sequence segment having genome coordinates 100-107 and a second genome sequence segment having genome coordinates of 200-207 have a separation distance of 92 nucleobases (genome coordinates 108 to 199).

[127] The term "sepsis," as used herein, refers to a serious medical condition resulting from the immune response to a severe infection. The related term "septicemia" is a sepsis of the bloodstream caused by bacteremia (the presence of bacteria in the bloodstream). The associated term "sepsis-causing organisms" refers to organisms that are frequently found in the blood when in the state of sepsis. Although the majority of sepsis-causing organisms are bacteria, fungi have also been identified in the blood of individuals with sepsis.

[128] As used herein, the term "speciating primer pair" refers to a primer pair designed to produce a bioagent identifying amplicon with the diagnostic capability of identifying species members of a group of genera or a particular genus of bioagents. Primer pair number 2249 (SEQ ID NOs: 601:609), for example, is

a speciating primer pair used to distinguish *Staphylococcus aureus* from other species of the genus *Staphylococcus*.

[129] The terms "stopping criterion" and "stopping criteria" refer to a chosen minimal acceptable criterion or criteria of collections of genome sequence segments for inclusion in the set of selected genome sequence segments to which primers will be designed. Examples of stopping criteria include, but are not limited to values reflecting mean separation distance or maximum separation distance. These stopping criteria can be chosen to act as the final step in a method for primer design of primers useful with targeted whole genome amplification.

[130] As used herein, a "sub-species characteristic" is a genetic characteristic that provides the means to distinguish two members of the same bioagent species. For example, one viral strain could be distinguished from another viral strain of the same species by possessing a genetic change (e.g., for example, a nucleotide deletion, addition or substitution) in one of the viral genes, such as the RNA-dependent RNA polymerase. Sub-species characteristics such as virulence genes and drug-are responsible for the phenotypic differences among the different strains of bacteria.

[131] The term "target genome," as used herein, refers to a genome of interest acting as the subject of analysis of the methods disclosed herein. For example, it is desirable to produce large quantities of a "target genome" while minimizing production of "background genomes."

[132] The terms "threshold criterion" and "threshold criteria," as used herein refer to values reflecting characteristics of genome sequence segments at which selections of sub-sets of genome sequence segments are made. For example, sub-sets of genome sequence segments can be chosen using a threshold criterion of a selectivity ratio at or above the mean selectivity ratio.

[133] As used herein, the term "targeted whole genome amplification primers" refers to primers collected in a set which are useful for selectively amplifying one or more target genome relative to one or more background genomes. Targeted whole genome amplification primers are designed according methods disclosed herein.

[134] As used herein, the term "target genome sequence segment" refers to a portion of specified length (typically about six to about twelve nucleobases in length) of a genome which is desired to be selectively amplified relative to one or more background genomes. Primers are selected to hybridize as selectively as possible to target genome sequence segments while minimizing hybridization to one or more background genomes.

[135] The term "template" refers to a strand of nucleic acid on which a complementary copy is built from nucleoside triphosphates through the activity of a template-dependent nucleic acid polymerase. Within a duplex the template strand is, by convention, depicted and described as the "bottom" strand. Similarly, the non-template strand is often depicted and described as the "top" strand.

[136] The term "triangulation genotyping analysis" refers to a method of genotyping a bioagent by measurement of molecular masses or base compositions of amplification products, corresponding to bioagent identifying amplicons, obtained by amplification of regions of more than one gene. In this sense, the term "triangulation" refers to a method of establishing the accuracy of information by comparing three or more types of independent points of view bearing on the same findings. Triangulation genotyping analysis carried out with a plurality of triangulation genotyping analysis primers yields a plurality of base compositions that then provide a pattern or "barcode" from which a species type can be assigned. The species type may represent a previously known sub-species or strain, or may be a previously unknown strain having a specific and previously unobserved base composition barcode indicating the existence of a previously unknown genotype.

[137] As used herein, the term "triangulation genotyping analysis primer pair" is a primer pair designed to produce bioagent identifying amplicons for determining species types in a triangulation genotyping analysis.

[138] 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 produced with different primer pairs. 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.

[139] As used herein, 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. Patent Serial No. 10/829,826 (incorporated herein by reference in its entirety) 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. Patent Serial No. 10/829,826 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.

[140] The term "variable sequence" as used herein refers to differences in nucleic acid sequence between two nucleic acids. For example, the genes of two different bacterial species may vary in sequence by the presence of single base substitutions and/or deletions or insertions of one or more nucleotides. These two forms of the structural gene are said to vary in sequence from one another. As used herein, the term "viral nucleic acid" includes, but is not limited to, DNA, RNA, or DNA that has been obtained from viral RNA, such as, for example, by performing a reverse transcription reaction. Viral RNA can either be single-stranded (of positive or negative polarity) or double-stranded.

[141] The term "virus" refers to obligate, ultramicroscopic, parasites that are incapable of autonomous replication (i.e., replication requires the use of the host cell's machinery). Viruses can survive outside of a host cell but cannot replicate.

[142] The term "viremia" refers to a condition where viruses enter the bloodstream. It is similar to bacteremia, a condition where bacteria enter the bloodstream, and septicemia. Active viremia refers to the capability of the virus to replicate in blood. There are two types of viremia: primary viremia, which is the initial spread of virus in the blood; and secondary viremia, where the primary viremia has resulted in infection of additional tissues, in which the virus has replicated and once more entered the circulation.

[143] The term "wild-type" refers to a gene or a gene product that has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified", "mutant" or "polymorphic" refers to a gene or gene product that displays modifications in sequence and or functional properties (i.e., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

[144] As used herein, a "wobble base" is a variation in a codon found at the third nucleotide position of a DNA triplet. Variations in conserved regions of sequence are often found at the third nucleotide position due to redundancy in the amino acid code.

DESCRIPTION OF EMBODIMENTS

Overview

[145] Disclosed herein are methods and compositions for amplifying a target genome of interest in the presence of background genomes. In the sense that one or more target genomes is selected to be amplified from a sample containing background genomes, the method may be considered as a method for "targeted whole genome amplification." The problem being solved using the disclosed compositions and methods is the production of larger quantities of genomic nucleic acid of an organism of interest than of the genomic or other nucleic acid originating from the background organisms.

[146] The greater quantities of nucleic acid representing the organism of interest are then available for further analyses, such as analyses conducted toward determining the genotype of a given microorganism, for example. Such analyses may encompass any type of nucleic acid characterization such as probe detection analysis by real time PCR, microarray analysis, sequencing analysis or analysis by methods disclosed herein which include determination of molecular mass and/or base composition of amplification products corresponding to bioagent identifying amplicons. The methods are particularly useful for obtaining increased quantities of nucleic acid of pathogens existing in human samples such as blood and fractions thereof, including serum and buffy coat, hepatic cells, sputum, urine and tissue biopsies. Pathogens that may be identified in such samples are implicated in bacteremia, septicemia and sepsis as well as viremia.

Target Genomes for Design of Targeted Whole Genome Amplification Primers

[147] In some preferred embodiments, one or more target genomes are chosen. The choice of target genomes is dictated by the objective of the analysis. For example, if the desired outcome of the targeted whole genome amplification process is to obtain nucleic acid representing the genome of a biowarfare organism such as *Bacillus anthracis*, which is suspected of being present in a soil sample at the scene of a biowarfare attack, one may choose to select the genome of *Bacillus anthracis* as the one and only target genome. If, on the other hand, the desired outcome of the targeted whole genome amplification process is to obtain nucleic acid representing a group of bacteria, such as, a group of potential biowarfare agents, more than one target genome may be selected such as, a group comprising any or all of the following bacteria: *Bacillus anthracis, Francisella tularensis, Yersinia pestis, Brucella sp., Burkholderia mallei, Rickettsia prowazekii, and Escherichia coli* 0157. Likewise, a different genome or group of genomes could be selected as the target genome(s) for other purposes. For example, a human genome or mitochondrial DNA may be the target over common genomes found in a soil sample or other sample environments where a crime may have taken place. Thus, the current methods and compositions can be

applied and the human genome (target) selectively amplified over the background genomes. Other examples could include the genomes of group of viruses that cause respiratory illness, pathogens that cause sepsis, or a group of fungi known to contaminate households.

Background Genomes for Design of Targeted Whole Genome Amplification Primers

[148] Background genomes may be selected based on the likelihood of the nucleic acid of certain organisms being present. For example, a soil sample which was handled by a human would be expected to contain nucleic acid representing the genomes of organisms including, but not limited to: *Homo sapiens, Gallus gallus, Guillardia theta, Oryza sativa, Arabidopsis thaliana, Yarrowia lipolytica, Saccharomyces cerevisiae, Debaryomyces hansenii, Kluyveromyces lactis, Schizosaccharomyces pom, Aspergillus fumigatus, Cryptococcus neoformans, Encephalitozoon cuniculi, Eremothecium gossypii, Candida glabrata, Apis mellifera, Drosophila melanogaster, Tribolium castaneum, Anopheles gambiae,* and *Caenorhabditis elegans*. Any or all of these genomes are appropriate to estimate as background genomes in the sample. The organisms actually in any particular sample will vary for each sample based upon the source and/or environment. Therefore, background genomes may be selected based upon the identities of organisms actually present in the sample. The composition of a sample can be determined using any of a number of techniques known to those ordinarily skilled in the art. In a further embodiment, the primers can be designed based upon actual identification of one or more background organisms in the sample, and based upon likelihood of any further one or more background organisms being in the sample.

Identification of Unique Genome Sequence Segments as Primer Hybridization Sites

[149] Once the target and background genomes of a sample are determined, the next step is to identify genome sequence segments within the target genome which are useful as primer hybridization sites. The efficiency of a given targeted whole genome amplification is dependent on effective use of primers. To produce an amplification product representative of a whole genome, the primer hybridization sites should have appropriate separation across the length of the genome. Preferably the mean separation distance between the primer hybridization sites is about 1000 nucleobases or less. More preferably the mean separation is about 800 nucleobases in length or less. Even more preferably, the mean separation is about 600 nucleobases in length or less. Most preferably, the mean separation between primer hybridization sites is about 500 nucleobases in length or less.

[150] One with ordinary skill in the art will recognize that effective priming for whole genome amplification depends upon several factors such as the fidelity and processivity of the polymerase enzyme used for primer extension. A longer mean separation distance between primer hybridization sites becomes more acceptable if the polymerase enzyme has high processivity. This indicates that the polymerase binds tightly to the nucleic acid template. This is a desirable characteristic for targeted whole genome amplification because it enables the polymerase to remain bound to the template nucleic acid and continue to extend the complementary nucleic acid strand being synthesized. Examples of polymerase enzymes

having high processivity include, but are not limited to Phi29 polymerase and *Taq* polymerase. Protein engineering strategies have been used to produce high processivity polymerase enzymes, for example, by covalent linkage of a polymerase to a DNA-binding protein (Wang et al., Nucl. Acids Res., 2004, 32(3) 1197-1207). As polymerases with improved processivity become available, longer mean separation distances, even greatly exceeding 1000 nucleobases may be acceptable for targeted whole genome amplification.

Hybridization Sensitivity and Selectivity

[151] For the purpose of targeted whole genome amplification, the choice of length of the primer hybridization sites (genome sequence segments) and the lengths of the corresponding primers hybridizing thereto, preferably will balance two factors; (1) sensitivity, which indicates the frequency of binding of a given primer to the target genome, and (2) selectivity, which indicates the extent to which a given primer hybridizes to the target genome with greater frequency than it hybridizes to background genomes. Generally, longer primers tend toward greater selectivity and lesser sensitivity while the converse holds for shorter primers. The relationship between primer length, selectivity and sensitivity is graphically represented in Figure 1. Preferably primers of about 5 to about 13 nucleobases in length are useful for targeted whole genome amplification; however, primer lengths falling outside of this range can be used as well. One will recognize that this range comprises primers having lengths of 5, 6, 7, 8, 9, 10, 11, 12 and 13 nucleobases. Primer size affects the balance between selectivity of the primer and sensitivity of the primer. Optimal primer length is determined for each sample with this balance in mind. Primers with lengths less than 5 nucleobases or greater than 13 nucleobases are also useful if the selectivity and sensitivity can be optimally maintained for that sample. Choosing a plurality of primers having various lengths provide broad priming across the target genome sequence(s) while also providing preferential binding of the primers to the target genome sequence(s) relative to the background genome sequences.

Selection Threshold Criteria

[152] In some embodiments, it is preferable to determine a suitable sub-set of the total unique genome sequence segments in order to reduce the total number of primers in the targeted whole genome amplification set in order to reduce the costs and complexity of the primer set. In some embodiments, determination of the suitable sub-set of unique genome sequence segments entails choosing one or more threshold criteria which indicate a useful and practical cut-off point for sensitivity and/or selectivity of a given genome sequence segment. Examples of such criteria include, but are not limited to, a selected threshold frequency of occurrence (a frequency of occurrence threshold value), and a selected selectivity ratio (a selectivity ratio threshold value).

[153] In some embodiments, it is useful to rank the total unique genome sequence segments according to the criteria. For example, the total unique genome sequence segments are ranked according to frequency of occurrence with the #1 rank indicating the greatest frequency of occurrence and the lowest

rank indicating the lowest frequency of occurrence. A threshold frequency of occurrence can then be chosen from the ranks. The threshold frequency of occurrence serves as the dividing line between members of the sub-set chosen for further analysis and the members that will not be further analyzed.

[154] In a non-limiting example, the mean "frequency of occurrence" can be calculated from the frequency of occurrence of the total genome sequence segments and this mean frequency of occurrence can be selected as a threshold criterion. The "frequency of occurrence" is defined in the "Definitions" section and also described in detail in Example 1. In one embodiment, genome sequence segments having a frequency of occurrence equal to or greater than the mean frequency of occurrence for all genome sequences being analyzed are chosen as a sub-set for further analysis. In other examples, the frequency of occurrence threshold criterion can be chosen above the mean frequency of occurrence or below the mean frequency of occurrence. In other examples, the sub-set is chosen with a frequency of occurrence threshold criterion that defines the sub-set as consisting of 80%, 70%, 60% or 50% of the total unique genome sequence segments or any whole or fractional number therebetween.

[155] In another non-limiting example, a "selectivity ratio" is chosen as the threshold criterion. The selectivity ratio is defined in the "Definitions" section and also described in detail in Example 1. In one embodiment, all genome sequence segments having a selectivity ratio equal to or greater than the mean selectivity ratio are chosen as a sub-set for further analysis. In other examples, the selectivity ratio threshold criterion can be chosen above the mean selectivity ratio or below the mean selectivity ratio. In other examples, the sub-set is chosen with a selectivity ratio threshold criterion that defines the sub-set as consisting of 80%, 70%, 60% or 50% of the total unique genome sequence segments or any whole or fractional number therebetween.

[156] In some embodiments, choosing the target genome sequence segments that are useful as primer hybridization sites is facilitated by the identification of most, if not all, of the unique genome sequence segments with lengths of 5, 6, 7, 8, 9, 10, 11, 12 and 13 nucleobases from which the primer hybridization sites will be chosen. Identification of unique sequence segments within genome sequences itself is a procedure that is well known to those with ordinary skill in bioinformatics. Furthermore, determination of the frequency of occurrence of a given genome sequence segment can be determined routinely using BLAST programs (basic local alignment search tools) and PowerBLAST programs known in the art (Altschul et al., J. Mol. Biol., 1990, 215, 403-410; Zhang and Madden, Genome Res., 1997, 7, 649-656). One with ordinary skill will recognize that improvements in polymerase processivity through, for example, protein engineering, discovery of new polymerases or improvements in amplification reagents and methods will allow for a shift in the balance between selectivity and sensitivity toward selectivity because a polymerase with improved processivity can synthesize longer stretches of primer extension products

without the need for high frequency of occurrence of shorter genome sequence segments acting as hybridization sites for shorter primers. Thus, primer lengths above 13 nucleobases are also practical for use in targeted whole genome amplification.

[157] Example 1 provides a demonstration of identification of unique genome sequence segments within a target genome, determination of the frequencies of occurrence of the genome sequence segments within the target genome sequence and determination of the frequencies of occurrence of the genome sequence segments within the background genome sequences. The example further describes calculation and ranking of selectivity ratios using the frequencies of occurrence of genome sequence segments within the target genomes and within the background genomes. In brief, selectivity ratios provide a description of the selectivity of a given genome sequence segment towards the target genome(s) with respect to the background genomes. A selectivity ratio is calculated for a given genome sequence segment simply by dividing the frequency of occurrence of the genome sequence segment within the target genome(s) by the frequency of occurrence of the genome sequence segment in the background genomes. A high selectivity ratio for a given genome sequence segment is favorable because it indicates that a primer designed to hybridize to the genome sequence segment will hybridize to the target genome(s) more frequently than it will hybridize to the background genomes, thus, accomplishing one objective for selective priming of the target genome. Selectivity ratios can be calculated either for a single target genome or for a plurality of target genomes. It is advantageous to consider the frequency of occurrence of all genome sequence segments in all of the chosen background genome segments to obtain useful selectivity ratios but, depending on the objective of the targeted whole genome amplification, it is not typically necessary to consider all possible target genomes in calculation of selectivity ratios. For example, in a simplified system consisting of two target genomes (target genome A and target genome B) and three background genomes (background genomes C, D and E), the selectivity ratio for genome sequence segment X which occurs once (frequency of occurrence = 1) in A, B, C, D and E, the target genome A selectivity ratio would be calculated as follows:

$1(A) / (1(C) + 1(D) + 1(E)) = 0.333$

In contrast, the total target genome (A+B) selectivity ratio would be calculated as follows:

$1(A) + 1(B) / (1(C) + 1(D) + 1 (E) = 0.667$

Design of Primers

[158] The primers that are designed to hybridize to the selected genome sequence segments are preferably 100% complementary to the genome sequence segments. In other embodiments, the primers that are designed to hybridize to the selected genome sequence segments are at least about 70% to about 100% complementary to the genome sequence segments, or any whole or fractional number therebetween. In general terms, design of primers for hybridization to selected nucleic acid sequences is well known to those with skill in the art and can be aided by commercially available computer programs. It is generally

preferable to design a given primer such that it is the same length as the genome sequence segment which was analyzed and chosen as a primer hybridization site. However, in some cases it may be advantageous to alter the length of the primer relative to the primer hybridization site. For example, if the primer is analyzed and found to have an unfavorable melting temperature and would benefit from elongation at the 5' or 3' end to produce a primer having an improved affinity for the target genome sequence. The length of the primer can be either increased or decreased. One with ordinary skill will recognize that alteration of the primer length also alters the primer hybridization site so that it no longer identical to the originally selected genome sequence segment. In some cases, it may be beneficial to analyze the genome sequence segment which corresponds to the hybridization site of a given length-altered primer. This analysis may be done by examination of data including but not limited to: frequency of occurrence and selectivity ratio and may also be done by actual *in vitro* testing of the length-altered primer.

[159] In some embodiments, in cases where it may be advantageous to design a primer to be less than 100% complementary to its corresponding genome sequence segment, it is also advantageous to examine the complement of the re-calculate selection criteria (such as frequency of occurrence and selectivity ratio) for a hypothetical genome sequence segment that is 100% complementary to the primer which is less than 100% complementary to its corresponding original genome sequence segment. If the selection criteria are unfavorable, it would be advantageous to consider design of an alternate primer sequence having improved selection criteria.

[160] In some embodiments, degenerate primers are designed in cases where there is ambiguity in the genome sequence or there is the possibility of occurrence of a single nucleotide polymorphism.

[161] In some embodiments, one or more phosphorothioate linkages are incorporated into the primers at the 3' end for the purpose of making the primers more resistant to nuclease activity.

[162] In some embodiments, the primers comprise chemically modified nucleobases which enhance affinity of hybridization and promote amplification efficiency. Such chemical modifications include, but are not limited to: 5-propynyl pyrimidines, phenoxazines, G-clamps, 2,6-diaminopurines and the like. One with ordinary skill in the art of making nucleotide modifications is capable of producing appropriate modifications to enhance the affinity of primers designed by the methods disclosed herein.

[163] In some embodiments, the primers are designed based upon the methods disclosed herein, synthesized and tested in targeted whole genome amplification under *in vitro* conditions where the efficiency of the targeted whole genome amplification can be assessed with respect to efficiency and/or bias toward the target genome(s) with respect to the background genomes. If the efficiency and/or bias is found

to be sub-optimal, redesign of selected primers may then be made by modifying them to correct potential defects such as poor affinity for template nucleic acid, occurrence of secondary structure and formation of primer dimers. In some embodiments, the redesigned primers are subjected to one or more additional rounds of *in vitro* testing in targeted whole genome amplification reactions to confirm their collective efficiency and/or bias toward the target genome(s) with respect to the background genomes. In some embodiments, if the efficiency and/or bias is found to be sub-optimal after a round of *in vitro* testing, the process of selection of primers is repeated using altered selection criteria which may include a higher selectivity ratio threshold value or one or more altered stopping criteria values which may include altered values for mean separation distance or maximum separation distance. One with ordinary skill will recognize that alteration of the selectivity ratio threshold value and the stopping criteria will result in a different set of primers being selected. The different sets of primers selected as a result of alteration of the selectivity ratio threshold value and/or stopping criteria may then be subjected to *in vitro* testing and additional rounds of alterations of the selection criteria for selection of an improved set of primers as needed.

Targeted Whole Genome Amplification Primer Kits

[164] Some embodiments also comprise kits that include targeted whole genome amplification primers designed according to the methods disclosed herein. In some embodiments, the kits comprise primers designed for general targeted whole genome amplification of bacteria from one or more collections of background genomes. For example, a targeted whole genome amplification kit for identification of bacteria in soil will have primers selected based on the genomes of typical background organisms found in soil. In another example, a targeted whole genome amplification kit for genotyping of viruses causing respiratory illness might be assembled with primers selected based on the target genomes of the respiratory pathogens and background genomes including the human genome and the genomes of commensal organisms found in human mucus, or other fluids. In another example, a targeted whole genome amplification kit for genotyping of sepsis-causing bacteria might be assembled with primers selected based on the target genomes of the sepsis-causing bacteria and background genomes including the human genome. Since human blood generally does not contain significant quantities of bacteria under non-sepsis conditions, bacterial genomes generally not be included in the primer selection process for this kit.

[165] In some embodiments, the kits comprise a sufficient quantity of a polymerase enzyme having high processivity. In some embodiments, the high processivity polymerase is Phi29 polymerase or *Taq* polymerase. In other embodiments, the high processivity polymerase is a genetically engineered polymerase whose processivity is increased relative to the native polymerase from which it was constructed.

[166] In some embodiments, the kits further comprise deoxynucleotide triphosphates, buffers, buffer additives such as magnesium salts, trehalose and betaine at concentrations optimized for targeted whole genome amplification.

[167] In some embodiments, the kits further comprise instructions for carrying out targeted whole genome amplification reactions.

[168] In one embodiment, the kits comprise at least a majority of the primers of the group consisting of SEQ ID NOs: 203-402 (see Table 3) or preferably at least a majority of the primers of the group consisting of SEQ ID NOs: 204:593 (see Table 4).

Bioagent Identifying Amplicons

[169] Disclosed herein are methods for detection and identification of unknown bioagents using bioagent identifying amplicons. Primers are selected to hybridize to conserved sequence regions of nucleic acids derived from a bioagent, and which bracket variable sequence regions to yield a bioagent identifying amplicon, which can be amplified and which is amenable to molecular mass determination. The molecular mass then provides a means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass or corresponding base composition signature of the amplification product is then matched against a database of molecular masses or base composition signatures. A match is obtained when an experimentally-determined molecular mass or base composition of an analyzed amplification product is compared with known molecular masses or base compositions of known bioagent identifying amplicons and the experimentally determined molecular mass or base composition is the same as the molecular mass or base composition of one of the known bioagent identifying amplicons. Alternatively, the experimentally-determined molecular mass or base composition may be within experimental error of the molecular mass or base composition of a known bioagent identifying amplicon and still be classified as a match. In some cases, the match may also be classified using a probability of match model such as the models described in U.S. Serial No. 11/073,362, which is commonly owned and incorporated herein by reference in entirety. Furthermore, the method can be applied to rapid parallel multiplex analyses, the results of which can be employed in a triangulation identification strategy. The present method provides rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent detection and identification.

[170] Despite enormous biological diversity, all forms of life on earth share sets of essential, common features in their genomes. Since genetic data provide the underlying basis for identification of bioagents by the methods disclosed herein, it is necessary 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.

[171] Unlike bacterial genomes, which exhibit conservation of numerous genes (i.e. housekeeping genes) across all organisms, viruses do not share a gene that is essential and conserved among all virus families. Therefore, viral identification is achieved within smaller groups of related viruses, such as members of a particular virus family or genus. For example, RNA-dependent RNA polymerase is present in all single-stranded RNA viruses and can be used for broad priming as well as resolution within the virus family.

[172] In some embodiments, at least one bacterial nucleic acid segment is amplified in the process of identifying the bacterial 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.

[173] In some embodiments, bioagent identifying amplicons comprise from about 27 to about 200 nucleobases (i.e. from about 39 to about 200 linked nucleosides), although both longer and short regions may be used. One of ordinary skill in the art will appreciate that these embodiments include compounds of 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 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 or 200 nucleobases in length, or any range therewithin.

[174] 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. Thus, it can be said that a given bioagent identifying amplicon is "defined by" a given pair of primers.

[175] 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 chemical reagents, restriction enzymes or cleavage primers, for example. Thus, in some embodiments, bioagent identifying amplicons are larger than 200 nucleobases and are amenable to molecular mass determination following restriction digestion. Methods of using restriction enzymes and cleavage primers are well known to those with ordinary skill in the art.

[176] In some embodiments, amplification products corresponding to bioagent identifying amplicons are obtained using the polymerase chain reaction (PCR) that 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). These methods are also known to those with ordinary skill.

Primer Pairs that Define Bioagent Identifying Amplicons

[177] In some embodiments, the primers are designed to bind to conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which provide variability sufficient to distinguish each individual bioagent, and which are amenable to molecular mass analysis. In some embodiments, the highly conserved sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity, or between about 99-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus, design of the primers involves selection of a variable region with sufficient variability to resolve the identity of a given bioagent. In some embodiments, bioagent identifying amplicons are specific to the identity of the bioagent.

[178] In some embodiments, identification of bioagents is accomplished at different levels using primers suited to resolution of each individual level of identification. Broad range survey primers are designed with the objective of identifying a bioagent as a member of a particular division (e.g., an order, family, genus or other such grouping of bioagents above the species level of bioagents). In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species or sub-species level. Examples of broad range survey primers include, but are not limited to: primer pair numbers: 346 (SEQ ID NOs: 594:602), and 348 (SEQ ID NOs: 595:603) which target DNA encoding 16S rRNA, and primer pair number 349 (SEQ ID NOs: 596:604) which targets DNA encoding 23S rRNA. Additional broad range survey primer pairs are disclosed in U.S. Serial No. 11/409,535 which is incorporated herein by reference in entirety.

[179] In some embodiments, drill-down primers are designed with the objective of identifying a bioagent at the sub-species level (including strains, subtypes, variants and isolates) based on sub-species characteristics which may, for example, include single nucleotide polymorphisms (SNPs), variable number

tandem repeats (VNTRs), deletions, drug resistance mutations or any other modification of a nucleic acid sequence of a bioagent relative to other members of a species having different sub-species characteristics. Drill-down intelligent primers are not always required for identification at the sub-species level because broad range survey intelligent primers may, in some cases provide sufficient identification resolution to accomplishing this identification objective. Examples of drill-down primers are disclosed in U.S. Patent Application Serial No. 11/409,535 which is incorporated herein by reference in entirety.

[180] A representative process flow diagram used for primer selection and validation process is outlined in Figure 8. For each group of organisms, candidate target sequences are identified (**200**) from which nucleotide alignments are created (**210**) and analyzed (**220**). Primers are then designed by selecting appropriate priming regions (**230**) to facilitate the selection of candidate primer pairs (**240**). The primer pairs are then subjected to in silico analysis by electronic PCR (ePCR) (**300**) wherein bioagent identifying amplicons are obtained from sequence databases such as 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 given amplicon to identify unknown bioagents such that the base compositions of amplicons with favorable probability scores are then 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 (**240**) are validated by testing their ability to hybridize to target nucleic acid by an in vitro amplification by a method such as PCR analysis (**400**) of nucleic acid from a collection of organisms (**410**). Amplification products thus obtained are analyzed by gel electrophoresis or by mass spectrometry to confirm the sensitivity, specificity and reproducibility of the primers used to obtain the amplification products (**420**).

[181] Many important pathogens, including the organisms of greatest concern as biowarfare agents, have been completely sequenced. This effort has greatly facilitated the design of primers for the detection of unknown bioagents. The combination of broad-range priming with division-wide and drill-down priming has been used very successfully in several applications of the technology, including environmental surveillance for biowarfare threat agents and clinical sample analysis for medically important pathogens.

[182] 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. However, it

should be noted that "synthesis" of primers does not equate with "design" of primers. The primers disclosed herein have been designed by the methods disclosed herein and then synthesized by the known methods.

[183] In some embodiments, primers are employed as compositions for use in methods for identification of bacterial bioagents as follows: a primer pair composition is contacted with nucleic acid (such as, for example, bacterial DNA or DNA reverse transcribed from the rRNA) of an unknown bacterial bioagent. The nucleic acid is then 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 each strand of the double-stranded amplification product is determined by a molecular mass measurement technique such as mass spectrometry for example, 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 then 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 and the molecular mass or base composition of an analogous bioagent identifying amplicon for a known viral bioagent indicates the identity of the unknown bacterial bioagent. In some embodiments, the method is repeated using one or more different primer pairs to resolve possible ambiguities in the identification process or to improve the confidence level for the identification assignment.

[184] 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.

[185] 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 or below 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. The employment

of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as triangulation identification.

[186] In other embodiments, the oligonucleotide primers are division-wide primers which hybridize to nucleic acid encoding genes of species within a genus of bacteria. In other embodiments, the oligonucleotide primers are drill-down primers which enable the identification of sub-species characteristics. Drill down primers provide the functionality of producing bioagent identifying amplicons for drill-down analyses such as strain typing when contacted with nucleic acid under amplification conditions. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of viral infections. In some embodiments, sub-species characteristics are identified using only broad range survey primers and division-wide and drill-down primers are not used.

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

[188] In some embodiments, various computer software programs may be used to aid in design of primers for amplification reactions such as Primer Premier 5 (Premier Biosoft, Palo Alto, CA) or OLIGO Primer Analysis Software (Molecular Biology Insights, Cascade, CO). These programs allow the user to input desired hybridization conditions such as melting temperature of a primer-template duplex for example. In some embodiments, an *in silico* PCR search algorithm, such as (ePCR) is used to analyze primer specificity across a plurality of template sequences which can be readily obtained from public sequence databases such as GenBank for example. An existing RNA structure search algorithm (Macke et al., Nucl. Acids Res., 2001, 29, 4724-4735, which is incorporated herein by reference in its entirety) 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, which is incorporated herein by reference in its entirety). This also provides information on primer specificity of the selected primer pairs. In some embodiments, the hybridization conditions applied to the algorithm can limit the results of primer specificity obtained from the algorithm. In some embodiments, the melting temperature threshold for the primer template duplex is specified to be 35°C or a higher temperature. In some embodiments the number of acceptable mismatches is specified to be seven mismatches or less. In some embodiments, the buffer components and concentrations and primer concentrations may be specified and incorporated into the algorithm, for example, an appropriate primer concentration is about 250 nM and appropriate buffer components are 50 mM sodium or potassium and 1.5 mM Mg^{2+}.

[189] 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., for example, a loop structure or a hairpin structure). The primers 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 2 of U.S. Serial No. 11/409,535, which is incorporated herein by reference in entirety. Thus, in some embodiments, 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 identical to another 20 nucleobase primer 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.

[190] 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, complementarity of primers with respect to the conserved priming regions of viral nucleic acid is between about 70% and about 75% 80%. In other embodiments, homology, sequence identity or complementarity, is between about 75% and about 80%. In yet other embodiments, homology, sequence identity or complementarity, is at least 85%, 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%.

[191] 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.

[192] 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.

[193] In one embodiment, the primers are at least 13 nucleobases in length. In another embodiment, the primers are less than 36 nucleobases in length.

[194] In some embodiments, the oligonucleotide primers are 13 to 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. The methods disclosed herein contemplate use of both longer and shorter primers. Furthermore, the primers may also be linked to one or more other desired moieties, including, but not limited to, affinity groups, ligands, regions of nucleic acid that are not complementary to the nucleic acid to be amplified, labels, etc. Primers may also form hairpin structures. For example, hairpin primers may be used to amplify short target nucleic acid molecules. The presence of the hairpin may stabilize the amplification complex (see e.g., TAQMAN MicroRNA Assays, Applied Biosystems, Foster City, California).

[195] In some embodiments, any oligonucleotide primer pair may have one or both primers with less then 70% sequence homology with a corresponding member of any of the primer pairs of Table 2 of U.S. Serial No. 11/409,535, if the primer pair has the capability of producing an amplification product corresponding to a bioagent identifying amplicon. In other embodiments, any oligonucleotide primer pair may have one or both primers with a length greater than 35 nucleobases if the primer pair has the capability of producing an amplification product corresponding to a bioagent identifying amplicon.

[196] In some embodiments, the function of a given primer may be substituted by a combination of two or more primers segments that hybridize adjacent to each other or that are linked by a nucleic acid loop structure or linker which allows a polymerase to extend the two or more primers in an amplification reaction.

[197] In some embodiments, the primer pairs used for obtaining bioagent identifying amplicons are the primer pairs of Table 2 of U.S. Serial No. 11/409,535. In other embodiments, other combinations of primer pairs are possible by combining certain members of the forward primers with certain members of the reverse primers. An example can be seen in Table 2 of U.S. Serial No. 11/409,535, for two primer pair combinations of forward primer 16S_EC_789_810_F with the reverse primers 16S_EC_880_894_R or 16S_EC_882_899_R. Arriving at a favorable alternate combination of primers in a primer pair depends upon the properties of the primer pair, most notably the size of the bioagent identifying amplicon that is defined by the primer pair, which preferably is between about 39 to about 200 nucleobases in length. Alternatively, a bioagent identifying amplicon longer than 200 nucleobases in length could be cleaved into smaller segments by cleavage reagents such as chemical reagents, or restriction enzymes, for example.

[198] In some embodiments, the primers are configured to amplify nucleic acid of a bioagent to produce amplification products that can be measured by mass spectrometry and from whose molecular

masses candidate base compositions can be readily calculated.

[199] 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 adenosine 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.

[200] In some embodiments, 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 (or RNA) 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 dP or dK (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).

[201] 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 that 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 (also known as propynylated thymine) 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 and incorporated herein by reference in its entirety. Propynylated primers are described in U.S Pre-Grant Publication No. 2003-0170682, which is also commonly owned and incorporated herein by reference in its entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096, each of which is incorporated herein by reference in its entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

[202] In some embodiments, primer hybridization is enhanced using primers containing 5-propynyl deoxycytidine and deoxythymidine nucleotides. These modified primers offer increased affinity and base pairing selectivity.

[203] In some embodiments, non-template primer tags are used to increase the melting temperature (Tm) 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 an 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.

[204] 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.

[205] 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 from its molecular mass.

[206] In some embodiments, 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 15N or 13C or both 15N and 13C.

[207] In some embodiments, multiplex amplification is performed where multiple bioagent identifying amplicons are amplified with a plurality of primer pairs. The advantages of multiplexing are that fewer reaction containers (for example, wells of a 96- or 384-well plate) are needed for each molecular

mass measurement, providing time, resource and cost savings because additional bioagent identification data can be obtained within a single analysis. Multiplex amplification methods are well known to those with ordinary skill and can be developed without undue experimentation. However, in some embodiments, one useful and non-obvious step in selecting a plurality candidate bioagent identifying amplicons for multiplex amplification is to ensure that each strand of each amplification product will be sufficiently different in molecular mass that mass spectral signals will not overlap and lead to ambiguous analysis results. In some embodiments, a 10 Da difference in mass of two strands of one or more amplification products is sufficient to avoid overlap of mass spectral peaks.

[208] In some embodiments, as an alternative to multiplex amplification, single amplification reactions can be pooled before analysis by mass spectrometry. In these embodiments, as for multiplex amplification embodiments, it is useful to select a plurality of candidate bioagent identifying amplicons to ensure that each strand of each amplification product will be sufficiently different in molecular mass that mass spectral signals will not overlap and lead to ambiguous analysis results.

Determination of Molecular Mass of Bioagent Identifying Amplicons

[209] In some embodiments, the molecular mass of a given 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.

[210] 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.

[211] The mass detectors used in the methods described herein 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.

Base Compositions of Bioagent Identifying Amplicons

[212] Although the molecular mass of amplification products obtained using intelligent primers provides a means for identification of bioagents, conversion of molecular mass data to a base composition signature is useful for certain analyses. As used herein, "base composition" is the exact number of each nucleobase (A, T, C and G) determined from the molecular mass of a bioagent identifying amplicon. In some embodiments, a base composition provides an index of a specific organism. Base compositions can be calculated from known sequences of known bioagent identifying amplicons and can be experimentally determined by measuring the molecular mass of a given bioagent identifying amplicon, followed by determination of all possible base compositions which are consistent with the measured molecular mass within acceptable experimental error. The following example illustrates determination of base composition from an experimentally obtained molecular mass of a 46-mer amplification product originating at position 1337 of the 16S rRNA of *Bacillus anthracis*. The forward and reverse strands of the amplification product have measured molecular masses of 14208 and 14079 Da, respectively. The possible base compositions derived from the molecular masses of the forward and reverse strands for the *Bacillus anthracis* products are listed in Table 1.

Table 1

Possible Base Compositions for *B. anthracis* 46mer Amplification Product

Calc. Mass Forward Strand	Mass Error orward Stran	Base omposition of Forwar Strand	Calc. Mass Reverse Strand	Mass Error Reverse Stran	Base Composition o Reverse Strand
14208.2935	0.079520	A1 G17 C10 T18	14079.2624	0.080600	A0 G14 C13 T19
14208.3160	0.056980	A1 G20 C15 T10	14079.2849	0.058060	A0 G17 C18 T11
14208.3386	0.034440	A1 G23 C20 T2	14079.3075	0.035520	A0 G20 C23 T3
14208.3074	0.065560	A6 G11 C3 T26	14079.2538	0.089180	A5 G5 C1 T35
14208.3300	0.043020	A6 G14 C8 T18	14079.2764	0.066640	A5 G8 C6 T27
14208.3525	0.020480	A6 G17 C13 T10	14079.2989	0.044100	A5 G11 C11 T19
14208.3751	0.002060	A6 G20 C18 T2	14079.3214	0.021560	A5 G14 C16 T11
14208.3439	0.029060	A11 G8 C1 T26	14079.3440	0.000980	A5 G17 C21 T3
14208.3665	0.006520	A11 G11 C6 T18	14079.3129	0.030140	A10 G5 C4 T27
14208.3890	**0.016020**	**A11 G14 C11 T10**	14079.3354	0.007600	A10 G8 C9 T19
14208.4116	0.038560	A11 G17 C16 T2	**14079.3579**	**0.014940**	**A10 G11 C14 T11**
14208.4030	0.029980	A16 G8 C4 T18	14079.3805	0.037480	A10 G14 C19 T3

14208.4255	0.052520	A16 G11 C9 T10	14079.3494	0.006360	A15 G2 C2 T27
14208.4481	0.075060	A16 G14 C14 T2	14079.3719	0.028900	A15 G5 C7 T19
14208.4395	0.066480	A21 G5 C2 T18	14079.3944	0.051440	A15 G8 C12 T11
14208.4620	0.089020	A21 G8 C7 T10	14079.4170	0.073980	A15 G11 C17 T3
-	-	-	14079.4084	0.065400	A20 G2 C5 T19
-	-	-	14079.4309	0.087940	A20 G5 C10 T13

[213] Among the 16 possible base compositions for the forward strand and the 18 possible base compositions for the reverse strand that were calculated, only one pair (shown in bold) are complementary base compositions, which indicates the true base composition of the amplification product. It should be recognized that this logic is applicable for determination of base compositions of any bioagent identifying amplicon, regardless of the class of bioagent from which the corresponding amplification product was obtained.

[214] In some embodiments, assignment of previously unobserved base compositions (also known as "true unknown base compositions") to a given phylogeny can be accomplished via the use of pattern classifier model algorithms. Base compositions, like sequences, vary slightly from strain to strain within species, for example. In some embodiments, the pattern classifier model is the mutational probability model. On other embodiments, the pattern classifier is the polytope model. The mutational probability model and polytope model are both commonly owned and described in U.S. Patent application Serial No. 11/073,362 which is incorporated herein by reference in entirety.

[215] In one embodiment, 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" 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.

[216] 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.

[217] The methods disclosed herein provide 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 base composition indexes become available in base composition databases.

Triangulation Identification

[218] In some cases, a molecular mass of a single bioagent identifying amplicon alone does not provide enough resolution to unambiguously identify a given bioagent. 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 determining the molecular masses of a plurality of bioagent identifying amplicons selected within a plurality of housekeeping 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.

[219] 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. In other related embodiments, one PCR reaction per well or container may be carried out, followed by an amplicon pooling step wherein the amplification products of different wells are combined in a single well or container which is then subjected to molecular mass analysis. The combination of pooled amplicons can be chosen such that the expected ranges of molecular masses of individual amplicons are not overlapping and thus will not complicate identification of signals.

Codon Base Composition Analysis

[220] In some embodiments, one or more nucleotide substitutions within a codon of a gene of an infectious organism confer drug resistance upon an organism which can be determined by codon base composition analysis. The organism can be a bacterium, virus, fungus or protozoan.

[221] In some embodiments, the amplification product containing the codon being analyzed is of a length of about 39 to about 200 nucleobases. The primers employed in obtaining the amplification product can hybridize to upstream and downstream sequences directly adjacent to the codon, or can hybridize to upstream and downstream sequences one or more sequence positions away from the codon. The primers may have between about 70% to 100% sequence complementarity with the sequence of the gene containing the codon being analyzed.

[222] In some embodiments, the codon analysis is undertaken for the purpose of investigating genetic disease in an individual. In other embodiments, the codon analysis is undertaken for the purpose of investigating a drug resistance mutation or any other deleterious mutation in an infectious organism such as a bacterium, virus, fungus or protozoan. In some embodiments, the bioagent is a bacterium identified in a biological product.

[223] In some embodiments, the molecular mass of an amplification product containing the codon being analyzed is measured by mass spectrometry. The mass spectrometry can be either electrospray (ESI) mass spectrometry or matrix-assisted laser desorption ionization (MALDI) mass spectrometry. Time-of-flight (TOF) is an example of one mode of mass spectrometry compatible with the methods disclosed herein.

[224] The methods disclosed herein can also be employed to determine the relative abundance of drug resistant strains of the organism being analyzed. Relative abundances can be calculated from amplitudes of mass spectral signals with relation to internal calibrants. In some embodiments, known quantities of internal amplification calibrants can be included in the amplification reactions and abundances of analyte amplification product estimated in relation to the known quantities of the calibrants.

[225] In some embodiments, upon identification of one or more drug-resistant strains of an infectious organism infecting an individual, one or more alternative treatments can be devised to treat the individual.

Determination of the Quantity of a Bioagent Using a Calibration Amplicon

[226] In some embodiments, the identity and quantity of an unknown bioagent can be determined using the process illustrated in Figure 9. Primers (**500**) and a known quantity of a calibration polynucleotide (**505**) are added to a sample containing nucleic acid of an unknown bioagent. The total nucleic acid in the sample is then 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.

[227] A sample comprising an unknown bioagent is contacted with a pair of primers that 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 2-8 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.

[228] 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.

[229] 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.

[230] 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.

[231] In some embodiments, the calibration sequence is comprised of DNA. In some embodiments, the calibration sequence is comprised of RNA.

[232] In some embodiments, the calibration sequence is inserted into a vector that 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.

Identification of Bacteria Using Bioagent Identifying Amplicons

[233] In other embodiments, the primer pairs produce bioagent identifying amplicons defined by priming regions at stable and highly conserved regions of nucleic acid of bacteria. The advantage to characterization of an amplicon defined by priming regions that fall within a highly conserved region is that there is a low probability that the region will evolve past the point of primer recognition, in which case, the primer hybridization of the amplification step would fail. Such a primer pair is thus useful as a broad range survey-type primer pair. In another embodiment, the intelligent primers produce bioagent identifying amplicons including a region which evolves more quickly than the stable region described above. The advantage of characterization bioagent identifying amplicon corresponding to an evolving genomic region is that it is useful for distinguishing emerging strain variants or the presence of virulence genes, drug resistance genes, or codon mutations that induce drug resistance.

[234] The methods disclosed herein have significant advantages as a platform for identification of diseases caused by emerging bacterial strains such as, for example, drug-resistant strains of *Staphylococcus*

aureus. The methods disclosed herein eliminate the need for prior knowledge of bioagent sequence to generate hybridization probes. This is possible because the methods are not confounded by naturally occurring evolutionary variations occurring in the sequence acting as the template for production of the bioagent identifying amplicon. Measurement of molecular mass and determination of base composition is accomplished in an unbiased manner without sequence prejudice.

[235] Another embodiment also provides a means of tracking the spread of a bacterium, such as a particular drug-resistant strain when a plurality of samples obtained from different locations are analyzed by the methods described above in an epidemiological setting. In one embodiment, a plurality of samples from a plurality of different locations is analyzed with primer pairs which produce bioagent identifying amplicons, a subset of which contains a specific drug-resistant bacterial strain. The corresponding locations of the members of the drug-resistant strain subset indicate the spread of the specific drug-resistant strain to the corresponding locations.

[236] Another embodiment provides the means of identifying a sepsis-causing bacterium. The sepsis-causing bacterium is identified in samples including, but not limited to blood and fractions thereof (including but not limited to serum and buffy coat), sputum, urine, specific cell types including but not limited to hepatic cells, and various tissue biopsies.

[237] Sepsis-causing bacteria include, but are not limited to the following bacteria: *Prevotella denticola, Porphyromonas gingivalis, Borrelia burgdorferi, Mycobacterium tuberculosis, Mycobacterium fortuitum, Corynebacterium jeikeium, Propionibacterium acnes, Mycoplasma pneumoniae, Streptococcus agalactiae, Streptococcus pneumoniae, Streptococcus mitis, Streptococcus pyogenes, Listeria monocytogenes, Enterococcus faecalis, Enterococcus faecium, Staphylococcus aureus, Staphylococcus coagulase-negative, Staphylococcus epidermis, Staphylococcus hemolyticus, Campylobacter jejuni, Bordatella pertussis, Burkholderia cepacia, Legionella pneumophila, Acinetobacter baumannii, Acinetobacter calcoaceticus, Pseudomonas aeruginosa, Aeromonas hydrophila, Enterobacter aerogenes, Enterobacter cloacae, Klebsiella pneumoniae, Moxarella catarrhalis, Morganella morganii, Proteus mirabilis, Proteus vulgaris, Pantoea agglomerans, Bartonella henselae, Stenotrophomonas maltophila, Actinobacillus actinomycetemcomitans, Haemophilus influenzae, Escherichia coli, Klebsiella oxytoca, Serratia marcescens,* and *Yersinia enterocolitica*.

[238] In some embodiments, identification of a sepsis-causing bacterium provides the information required to choose an antibiotic with which to treat an individual infected with the sepsis-causing bacterium and treating the individual with the antibiotic. Treatment of humans with antibiotics is well known to medical practitioners with ordinary skill.

Kits for Producing Bioagent Identifying Amplicons

[239] Also provided are kits for carrying out 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 2 of U.S. Serial No. 11/409,535.

[240] In some embodiments, the kit comprises one or more broad range survey primer(s), division wide primer(s), or drill-down primer(s), or any combination thereof. If a given problem involves identification of a specific bioagent, the solution to the problem may require the selection of a particular combination of primers to provide the solution to the problem. A kit may be designed so as to comprise particular primer pairs for identification of a particular bioagent. A drill-down kit may be used, for example, to distinguish different genotypes or strains, drug-resistant, or otherwise. In some embodiments, the primer pair components of any of these kits may be additionally combined to comprise additional combinations of broad range survey primers and division-wide primers so as to be able to identify a bacterium.

[241] In some embodiments, the kit contains standardized calibration polynucleotides for use as internal amplification calibrants. Internal calibrants are described in commonly owned PCT Publication Number WO 2005/098047 which is incorporated herein by reference in its entirety.

[242] In some embodiments, the kit comprises a sufficient quantity of reverse transcriptase (if RNA is to be analyzed for example), a DNA polymerase, suitable nucleoside triphosphates (including alternative dNTPs such as inosine or modified dNTPs such as the 5-propynyl pyrimidines or any dNTP containing molecular mass-modifying tags such as 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.

[243] Some embodiments are kits that contain one or more survey bacterial primer pairs represented by primer pair compositions wherein each member of each pair of primers has 70% to 100%

sequence identity with the corresponding member from the group of primer pairs represented by any of the primer pairs of Table 2 of U.S. Serial No. 11/409,535. The survey primer pairs may include broad range primer pairs which hybridize to ribosomal RNA, and may also include division-wide primer pairs which hybridize to housekeeping genes such as rplB, tufB, rpoB, rpoC, valS, and infB, for example.

[244] In some embodiments, a kit may contain one or more survey bacterial primer pairs and one or more triangulation genotyping analysis primer pairs such as the primer pairs of Tables 8, 12, 14, 19, 21, 23, or 24 of U.S. Serial No. 11/409,535. In some embodiments, the kit may represent a less expansive genotyping analysis but include triangulation genotyping analysis primer pairs for more than one genus or species of bacteria. For example, a kit for surveying nosocomial infections at a health care facility may include, for example, one or more broad range survey primer pairs, one or more division wide primer pairs, one or more *Acinetobacter baumannii* triangulation genotyping analysis primer pairs and one or more *Staphylococcus aureus* triangulation genotyping analysis primer pairs. One with ordinary skill will be capable of analyzing *in silico* amplification data to determine which primer pairs will be able to provide optimal identification resolution for the bacterial bioagents of interest.

[245] In some embodiments, a kit may be assembled for identification of sepsis-causing bacteria. An example of such a kit embodiment is a kit comprising one or more of the primer pairs of Table 25 of U.S. Serial No. U.S. Serial No. 11/409,535, which provide for a broad survey of sepsis-causing bacteria.

[246] Some embodiments of the kits are 96-well or 384-well plates with a plurality of wells containing any or all of the following components: dNTPs, buffer salts, Mg^{2+}, betaine, and primer pairs. In some embodiments, a polymerase is also included in the plurality of wells of the 96-well or 384-well plates.

[247] Some embodiments of the kit contain instructions for PCR and mass spectrometry analysis of amplification products obtained using the primer pairs of the kits.

[248] Some embodiments of the kit include a barcode which uniquely identifies the kit and the components contained therein according to production lots and may also include any other information relative to the components such as concentrations, storage temperatures, etc. The barcode may also include analysis information to be read by optical barcode readers and sent to a computer controlling amplification, purification and mass spectrometric measurements. In some embodiments, the barcode provides access to a subset of base compositions in a base composition database which is in digital communication with base composition analysis software such that a base composition measured with primer pairs from a given kit can be compared with known base compositions of bioagent identifying amplicons defined by the primer pairs of that kit.

[249] In some embodiments, the kit contains a database of base compositions of bioagent identifying amplicons defined by the primer pairs of the kit. The database is stored on a convenient computer readable medium such as a compact disk or USB drive, for example.

[250] In some embodiments, the kit includes a computer program stored on a computer formatted medium (such as a compact disk or portable USB disk drive, for example) comprising instructions which direct a processor to analyze data obtained from the use of the primer pairs disclosed herein. The instructions of the software transform data related to amplification products into a molecular mass or base composition which is a useful concrete and tangible result used in identification and/or classification of bioagents. In some embodiments, the kits contain all of the reagents sufficient to carry out one or more of the methods described herein.

Combination Kits Including Targeted Whole Genome Amplification Primers and Primer Pairs for Obtaining Bioagent Identifying Amplicons

[251] In some embodiments, kits are provided that include targeted whole genome amplification primers and primer pairs for production of bioagent identifying amplicons. These kits are for use in applications where a bioagent such as a human pathogen for example, is present only in small quantities in a human clinical sample. An example of such a kit could include a set of targeted whole genome amplification primers for selective amplification of a bacterium implicated in septicemia. The targeted whole genome amplification primers are designed with human genomic DNA chosen as a background genome, for the purpose of detection of an infection of an individual with *Bacillus anthracis.* The kit would also include one or more broad range survey primer pairs and/or division-wide primer pairs for production of amplification products corresponding to bioagent identifying amplicons for identification of the bacterium. Optionally one or more drill-down primer pairs are included in the kit for determining sub-species characteristics of the septicemia by analysis of additional bioagent identifying amplicons.

[252] These combination kits may also include a plurality of polymerase enzymes whose members are specialized for a PCR type amplification reaction, such as *Taq* polymerase, for example, to obtain amplification products corresponding to bioagent identifying amplicons, and such as Phi29 polymerase which is a high processivity polymerase suitable for catalysis of multiple displacement amplification reactions for targeted whole genome amplification reactions carried out for elevating the quantity of a target genome of interest.

[253] The combination kits may also include amplification reagents including but not limited to: deoxynucleotide triphosphates, compatible solutes such as betaine and trehalose, buffer components, and salts such as magnesium chloride.

[254] While the present invention has been described with specificity in accordance with certain of its embodiments, the following examples serve only to illustrate the invention and are not intended to limit the same. 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.

Example 1: Identification and Ranking of Genome Sequence Segments

[255] This example illustrates the process of identification of unique genome sequence segments of 6 to 12 nucleobases in length, as well as determination of frequency of occurrence and selectivity ratio values for a simplified hypothetical genome model system consisting of a single target genome having the sequence: aaaaaaaaaattttttttttccccccccccgggggggggg ((SEQ ID NO: 16) base composition of A10 T10 C10 and G10) with two background genomes having the following sequences aaaaaaaattttttttccccccccgggggggg (SEQ ID NO: 17) Bkg 1: base composition of A8 T8 C8 G8) and aaaaaaaaaattttttttt (SEQ ID NO: 18) Bkg 2: base composition of A10 T10 C0 G0). Table 2 provides a list of all unique genome sequence segments for the target genome and indicates the frequency of occurrence of each genome sequence segment in the target genome and in the background genomes. For example, the genome sequence segment having the sequence of eight consecutive c residues **cccccccc** (SEQ ID NO:445) occurs 3 times (**bold**) within the 10 nucleobase stretch of c residues in the simplified hypothetical target genome:

aaaaaaaaaattttttttt**cccccccc**ccgggggggggg (SEQ ID NO: 16);

aaaaaaaaaattttttttt**cccccccc**cgggggggggg (SEQ ID NO: 16); aaaaaaaaaattttttttt**cccccccc**ccgggggggggg (SEQ ID NO: 16);

(c residue stretch underlined) but only once in the background genomes (the genome sequence segment appears once in Bkg 1 and does not appear in Bkg 2). The selectivity ratio for this genome sequence segment is 3.00 as determined by dividing the frequency of occurrence in the target genome by the frequency of occurrence in the background genomes. The data in Table 2 are sorted according to the selectivity ratio rank. A selectivity ratio of infinity (∞) indicates that the genome sequence segment does not occur in the background genomes (Bkg 1 and Bkg 2). The mean frequency of occurrence of the genome sequence segments in the target genome was calculated to be 1.22 and the mean selectivity ratio was calculated to be 0.76. If desired, these values could be used as threshold values for selection of one or more sub-sets of genome sequence segments for further characterization by processes such as the process shown in Figure 2 for example. Alternatively, threshold values greater than or less than the mean frequency of occurrence or the mean selectivity ratio could be chosen.

Table 2: Frequency of Occurrence of Genome Sequence Segments in a Hypothetical Target Genome and Two Hypothetical Background Genomes

Genome Sequence Segment	SEQ ID NO:	Frequency in Target	Frequency in Bkg 1	Frequency in Bkg 2	Total Background	Selectivity Ratio	Selectivity Ratio Rank

ccccccccc	19	2	0	0	0	Infinity	1
gggggggggg	20	2	0	0	0	Infinity	1
cccccccccc	21	1	0	0	0	Infinity	1
ccccccccccg	22	1	0	0	0	Infinity	1
cggggggggg	23	1	0	0	0	Infinity	1
ggggggggggg	24	1	0	0	0	Infinity	1
tccccccccc	25	1	0	0	0	Infinity	1
tttttttttc	26	1	0	0	0	Infinity	1
ccccccccccg	27	1	0	0	0	Infinity	1
cccccccccgg	28	1	0	0	0	Infinity	1
ccgggggggggg	29	1	0	0	0	Infinity	1
cgggggggggg	30	1	0	0	0	Infinity	1
tcccccccccc	31	1	0	0	0	Infinity	1
ttcccccccccc	32	1	0	0	0	Infinity	1
ttttttttcc	33	1	0	0	0	Infinity	1
tttttttttc	34	1	0	0	0	Infinity	1
atttttttttc	35	1	0	0	0	Infinity	1
ccccccccccgg	36	1	0	0	0	Infinity	1
cccccccccggg	37	1	0	0	0	Infinity	1
cccgggggggggg	38	1	0	0	0	Infinity	1
ccgggggggggg	39	1	0	0	0	Infinity	1
tcccccccccg	40	1	0	0	0	Infinity	1
ttcccccccccc	41	1	0	0	0	Infinity	1
tttcccccccccc	42	1	0	0	0	Infinity	1
ttttttttccc	43	1	0	0	0	Infinity	1
tttttttttcc	44	1	0	0	0	Infinity	1
cccccccc	45	3	1	0	1	3.00	2
gggggggg	46	3	1	0	1	3.00	2
ggggggg	47	4	2	0	2	2.00	3
cccccc	48	5	3	0	3	1.67	4
gggggg	49	5	3	0	3	1.67	4
ccccg	50	1	1	0	1	1.00	5
ccccgg	51	1	1	0	1	1.00	5
cccggg	52	1	1	0	1	1.00	5
ccgggg	53	1	1	0	1	1.00	5
cggggg	54	1	1	0	1	1.00	5
tcccc	55	1	1	0	1	1.00	5

ttcccc	56	1	1	0	1	1.00	5
tttccc	57	1	1	0	1	1.00	5
ttttcc	58	1	1	0	1	1.00	5
tttttc	59	1	1	0	1	1.00	5
ccccccg	60	1	1	0	1	1.00	5
cccccgg	61	1	1	0	1	1.00	5
ccccggg	62	1	1	0	1	1.00	5
cccgggg	63	1	1	0	1	1.00	5
ccggggg	64	1	1	0	1	1.00	5
cgggggg	65	1	1	0	1	1.00	5
tccccccc	66	1	1	0	1	1.00	5
ttccccc	67	1	1	0	1	1.00	5
tttcccc	68	1	1	0	1	1.00	5
ttttccc	69	1	1	0	1	1.00	5
tttttcc	70	1	1	0	1	1.00	5
ttttttc	71	1	1	0	1	1.00	5
cccccccg	72	1	1	0	1	1.00	5
ccccccgg	73	1	1	0	1	1.00	5
cccccggg	74	1	1	0	1	1.00	5
ccccgggg	75	1	1	0	1	1.00	5
cccggggg	76	1	1	0	1	1.00	5
ccgggggg	77	1	1	0	1	1.00	5
cgggggggg	78	1	1	0	1	1.00	5
tcccccccc	79	1	1	0	1	1.00	5
ttcccccc	80	1	1	0	1	1.00	5
tttccccc	81	1	1	0	1	1.00	5
ttttcccc	82	1	1	0	1	1.00	5
tttttccc	83	1	1	0	1	1.00	5
ttttttcc	84	1	1	0	1	1.00	5
tttttttc	85	1	1	0	1	1.00	5
aaaaaaaaa	86	2	0	2	2	1.00	5
ccccccccg	87	1	1	0	1	1.00	5
cccccccgg	88	1	1	0	1	1.00	5
ccccccggg	89	1	1	0	1	1.00	5
cccccgggg	90	1	1	0	1	1.00	5
ccccggggg	91	1	1	0	1	1.00	5
cccgggggg	92	1	1	0	1	1.00	5

ccgggggggg	93	1	1	0	1	1.00	5
cgggggggg	94	1	1	0	1	1.00	5
tcccccccc	95	1	1	0	1	1.00	5
ttcccccccc	96	1	1	0	1	1.00	5
tttcccccc	97	1	1	0	1	1.00	5
ttttccccc	98	1	1	0	1	1.00	5
tttttcccc	99	1	1	0	1	1.00	5
ttttttccc	100	1	1	0	1	1.00	5
tttttttcc	101	1	1	0	1	1.00	5
ttttttttc	102	1	1	0	1	1.00	5
ttttttttt	103	2	0	2	2	1.00	5
aaaaaaaaaa	104	1	0	1	1	1.00	5
aaaaaaaaat	105	1	0	1	1	1.00	5
attttttttt	106	1	0	1	1	1.00	5
ccccccccgg	107	1	1	0	1	1.00	5
cccccccggg	108	1	1	0	1	1.00	5
ccccccgggg	109	1	1	0	1	1.00	5
cccccggggg	110	1	1	0	1	1.00	5
ccccgggggg	111	1	1	0	1	1.00	5
cccgggggggg	112	1	1	0	1	1.00	5
ccgggggggg	113	1	1	0	1	1.00	5
ttcccccccc	114	1	1	0	1	1.00	5
tttcccccc	115	1	1	0	1	1.00	5
ttttccccc	116	1	1	0	1	1.00	5
tttttcccc	117	1	1	0	1	1.00	5
ttttttcccc	118	1	1	0	1	1.00	5
tttttttccc	119	1	1	0	1	1.00	5
ttttttttcc	120	1	1	0	1	1.00	5
tttttttttt	121	1	0	1	1	1.00	5
aaaaaaaaaat	122	1	0	1	1	1.00	5
aaaaaaaaatt	123	1	0	1	1	1.00	5
aatttttttt	124	1	0	1	1	1.00	5
attttttttt	125	1	0	1	1	1.00	5
ccccccccggg	126	1	1	0	1	1.00	5
cccccccgggg	127	1	1	0	1	1.00	5
ccccccgggggg	128	1	1	0	1	1.00	5
cccccgggggg	129	1	1	0	1	1.00	5

ccccgggggggg	130	1	1	0	1	1.00	5
cccgggggggg	131	1	1	0	1	1.00	5
tttccccccccc	132	1	1	0	1	1.00	5
ttttccccccc	133	1	1	0	1	1.00	5
tttttcccccc	134	1	1	0	1	1.00	5
ttttttccccc	135	1	1	0	1	1.00	5
tttttttcccc	136	1	1	0	1	1.00	5
ttttttttccc	137	1	1	0	1	1.00	5
aaaaaaaaaatt	138	1	0	1	1	1.00	5
aaaaaaaaattt	139	1	0	1	1	1.00	5
aaatttttttt	140	1	0	1	1	1.00	5
aatttttttttt	141	1	0	1	1	1.00	5
cccccccgggg	142	1	1	0	1	1.00	5
ccccccggggg	143	1	1	0	1	1.00	5
cccccgggggg	144	1	1	0	1	1.00	5
ccccgggggggg	145	1	1	0	1	1.00	5
ccccgggggggg	146	1	1	0	1	1.00	5
ttttccccccc	147	1	1	0	1	1.00	5
tttttcccccc	148	1	1	0	1	1.00	5
ttttttccccc	149	1	1	0	1	1.00	5
tttttttcccc	150	1	1	0	1	1.00	5
ttttttttccc	151	1	1	0	1	1.00	5
aaaaaaaa	15	3	1	3	4	0.75	6
ttttttt	153	3	1	3	4	0.75	6
aaaaaaa	154	4	2	4	6	0.67	7
ccccccc	155	4	2	4	6	0.67	7
tttttt	156	4	2	4	6	0.67	7
aaaaaa	157	5	3	5	8	0.63	8
ttttt	158	5	3	5	8	0.63	8
aaaaat	159	1	1	1	2	0.50	9
aaaatt	160	1	1	1	2	0.50	9
aaattt	161	1	1	1	2	0.50	9
aatttt	162	1	1	1	2	0.50	9
attttt	163	1	1	1	2	0.50	9
aaaaaat	164	1	1	1	2	0.50	9
aaaaatt	165	1	1	1	2	0.50	9
aaaattt	166	1	1	1	2	0.50	9

aaatttt	167	1	1	1	2	0.50	9
aattttt	168	1	1	1	2	0.50	9
atttttt	169	1	1	1	2	0.50	9
aaaaaaat	170	1	1	1	2	0.50	9
aaaaaatt	171	1	1	1	2	0.50	9
aaaaattt	172	1	1	1	2	0.50	9
aaaatttt	173	1	1	1	2	0.50	9
aaattttt	174	1	1	1	2	0.50	9
aatttttt	175	1	1	1	2	0.50	9
atttttt	176	1	1	1	2	0.50	9
aaaaaaaat	177	1	1	1	2	0.50	9
aaaaaaatt	178	1	1	1	2	0.50	9
aaaaaattt	179	1	1	1	2	0.50	9
aaaaatttt	180	1	1	1	2	0.50	9
aaaattttt	181	1	1	1	2	0.50	9
aaatttttt	182	1	1	1	2	0.50	9
aattttttt	183	1	1	1	2	0.50	9
atttttttt	184	1	1	1	2	0.50	9
aaaaaaaatt	185	1	1	1	2	0.50	9
aaaaaaattt	186	1	1	1	2	0.50	9
aaaaaatttt	187	1	1	1	2	0.50	9
aaaaattttt	188	1	1	1	2	0.50	9
aaaatttttt	189	1	1	1	2	0.50	9
aaatttttt	190	1	1	1	2	0.50	9
aattttttt	191	1	1	1	2	0.50	9
aaaaaaaattt	192	1	1	1	2	0.50	9
aaaaaaatttt	193	1	1	1	2	0.50	9
aaaaaattttt	194	1	1	1	2	0.50	9
aaaaatttttt	195	1	1	1	2	0.50	9
aaaattttttt	196	1	1	1	2	0.50	9
aaatttttttt	197	1	1	1	2	0.50	9
aaaaaaaatttt	198	1	1	1	2	0.50	9
aaaaaaattttt	199	1	1	1	2	0.50	9
aaaaaatttttt	200	1	1	1	2	0.50	9
aaaaatttttt	201	1	1	1	2	0.50	9
aaaatttttttt	202	1	1	1	2	0.50	9

Example 2: *In Silico* Method for Design of Primers for Targeted Whole Genome Amplification

[256] Some embodiments of the methods disclosed herein are *in silico* methods for selecting primers for targeted whole genome amplification. The primers are selected by first defining the target genome(s) and background genome(s). For the target genome(s), all unique genome sequence segments of lengths of about 5 to about 13 nucleobases in length are determined by a set of computer executable instructions stored on a computer-readable medium.

[257] In some embodiments, the target and background genome segments are obtained from public databases such as GenBank, for example. The frequency of occurrence values of members of the genome sequence segments in the target genome(s) and background genome(s) are determined by computer executable instructions such as a BLAST algorithm for example. The selectivity ratio values of members of the genome sequence segments are determined by computer executable mathematical instructions. In some embodiments, the *in silico* method ranks the genome sequence segments according to frequency of occurrence and/or selectivity ratio. In some embodiments, a frequency of occurrence threshold value is chosen to define a sub-set of genome sequence segments to carry forward.

[258] In some embodiments, a selectivity ratio threshold value is chosen to define a sub-set of genome sequence segments to carry forward. In some embodiments, the selectivity ratio threshold value is any whole or fractional percentage between about 25% above or about 25% below the mean selectivity ratio. For example, if the mean selectivity ratio is 55, the chosen selectivity ratio threshold value may be any whole or fractional number between about 41.25 and about 68.75. In other embodiments, both a frequency of occurrence threshold value and a selectivity ratio threshold value are chosen and both of these threshold values are used to define the sub-set of genome sequence segments to carry forward. The genome sequence segments are ranked according to the chosen threshold value.

[259] At this point, a process such as the process outlined in Figure 2 may be followed wherein the top ranked genome sequence segment is selected and added to the sub-set of genome sequence segments (**1000**). Then the next highest ranking genome sequence segment is selected (**2000**) and subjected to a first computer executable query (**3000**) which determines whether or not the next ranked genome sequence segment originates from within the largest remaining separation distance (remaining portion of the genome which has not had a genome sequence segment selected). If the next highest ranking genome sequence segment does not originate within the largest separation distance, it is skipped (but remains in with the same rank in the group of unselected genome sequence segments) and the process reverts to step **2000**. If the next highest ranking genome sequence segment does originate from within the largest separation distance it is selected and added to the set of genome sequence segments to which primers will be designed (**4000**). An example of operation of steps **1000** to **5000** (including cycling between steps **2000** and **5000**) of Figure 2 follows: the top ranked genome sequence segment (#1) is selected by default in step **1000**. As a result of selection of genome sequence segment #1, only two separation distances remain on the target genome. One

of the two separation distances stretches from the 5' end of the #1 genome sequence segment to the 5' end of the genome and the other of the two separation distances stretches from the 3' end of the #1 genome sequence segment to the 5' end of the genome. It is assumed in this example that the 5' end of the genome to the 5' end of the #1 genome sequence segment has the longest separation distance. In step **2000**, the next highest ranked genome sequence segment (#2 in this case) is selected. At step **3000** (query 1) it is determined whether or not the #2 ranked genome sequence segment is located within this longest separation distance between the 5' end of the genome and the 5' end of the #1 genome sequence segment. If the #2 ranked genome sequence segment is not located within this longest separation distance, it is not selected and remains in the unselected group while the process reverts to step **2000** where the next highest ranked genome sequence segment (#3) is selected from the list of ranked genome sequence segments. In performing step **3000** on genome sequence segment #3, it is determined that this genome sequence segment is located within the largest separation distance. Thus genome sequence segment #3 is added to the sub-set in step **4000**. At this point, only genome sequence segments #1 and #3 have been added to the sub-set. In step **5000**, it is confirmed that the predetermined quantity of genome sequence segments (for example 200 genome sequence segments) has not been obtained (because only 2 genome sequence segments have been selected thus far). The answer to query 2 (**5000**) is "no" and the process cycles back to step **2000** where the next ranked genome sequence segment is selected. In this example, the next ranked genome sequence segment is #2 because it was skipped in the previous cycle. In step **3000** query 1 determines that genome sequence segment now does fall within the largest separation distance (because the largest separation distance in the previous cycle is no longer the largest in the current cycle due to the appearance of genome sequence segment #3). Thus genome sequence segment #2 is added to the sub-set in step **4000**. Step **5000** is then performed and the answer to query 2 is "no" because only 3 genome sequence segments have been selected thus far. Again the process cycles back to step **2000** and continues cycling between steps **2000** and **5000**, selecting the next highest ranked genome sequence segments in each cycle and performing the queries of step **3000** and step **5000** until the predetermined quantity of genome sequence segments is obtained.

[260] In some embodiments, the predetermined number of genome sequence segments is sufficient to provide consistently dispersed coverage of the genome by primers hybridizing to the selected genome sequence segments. In some embodiments, this predetermined number of genome sequence segments is between about 100 to about 300 genome sequence segments, including any number therebetween.

[261] The predetermined number will depend upon the length of the target genome(s). For example, longer genomes may require additional primer coverage and thus selecting a larger predetermined number of genome sequence segments to serve as primer hybridization sites may be advantageous. In some embodiments, after a group of genome sequence segments have been selected, statistical measures such as

those presented in Table 5 may be used to evaluate the likelihood that a group of primers designed to hybridize to the genome sequence segments will produce efficient and biased amplification of the target genome(s) of interest. If the statistics are deemed inefficient, it may be advantageous to consider revising the predetermined number of genome sequence segments to a larger number to provide greater coverage of the target genome(s). This statistical evaluation process is useful because it avoids the unnecessary expense of *in vitro* testing of entire groups of primers.

[262] Continuing now in the process of Figure 2, when the answer to the second query (**5000**) is "yes," the predetermined quantity of genome sequence segments has been obtained. At that point, a third computer executable query (**6000**) is performed to determine whether or not the "stopping criterion/criteria" has or have been met. The "stopping criterion/criteria" represent the final threshold value(s) relating to genome sequence segment coverage over which the *in silico* method must pass before the method instructions and queries of the *in silico* end (**7000**). If the stopping criteria have not been met, the process cycles back to step **2000** with an adjustment of the selectivity threshold value if necessary (**6500**).

[263] In some embodiments, a single stopping criterion used. In other embodiments, more than one stopping criteria are used. In one embodiment one stopping criterion is a value reflecting the mean separation distance between genome sequence segments within the target genome sequence(s). For example, a mean distance between genome sequence segments is a whole or fractional number less or equal to about 500, 600, 700, 900, or 1000 nucleobases or any whole or fractional number therebetween. In other embodiments, the stopping criterion is the mean distance between genome sequence segments within the target genome sequence(s) or a value above or below the mean distance between genome sequence segments within the target genome sequence(s).

[264] In other embodiments, a stopping criterion is the maximum distance between any two of the selected genome sequence segments within the target genome sequence(s). For example, an appropriate maximum distance between any two genome sequence segments might be less than or equal to about 5,000, 6,000, 7,000, 8,000, 9,000 or 10,000 nucleobases or any number therebetween.

[265] In some embodiments, after the stopping criterion or criteria have been met and the computer executable instructions are complete, the *in silico* method produces an output report comprising a list of genome sequence segments. The report may be a print-out or a display on a graphical interface or any other means for displaying the results of the selection process. The *in silico* method may also provide a means for designing primers that hybridize to the genome sequence segments.

Example 3: Selection of Primer Sets for Targeted Whole Genome Amplification

[266] In a first example for targeted whole genome amplification, *Bacillus anthracis* Ames was chosen as a single target genome. The set of background genomes included the genomes of: *Homo sapiens, Gallus gallus, Guillardia theta, Oryza sativa, Arabidopsis thaliana, Yarrowia lipolytica, Saccharomyces cerevisiae, Debaryomyces hansenii, Kluyveromyces lactis, Schizosaccharomyces pom, Aspergillus fumigatus, Cryptococcus neoformans, Encephalitozoon cuniculi, Eremothecium gossypii, Candida glabrata, Apis mellifera, Drosophila melanogaster, Tribolium castaneum, Anopheles gambiae,* and *Caenorhabditis elegans.* These background genomes were chosen because they would be expected to be present in a typical soil sample handled by a human.

[267] Unique genome sequence segments 7 to 12 nucleobases in length were identified. Frequency of occurrence and selectivity ratio values were determined. As a result, 200 genome sequence segments were identified. In most cases, the primers designed to hybridize with 100% complementarity to its corresponding genome sequence segment. In a few other cases, degenerate primers were prepared. The degenerate bases of the primers occur at positions complementary to positions having ambiguity within the target *Bacillus anthracis* genome or complementary to positions known or thought to be susceptible to single nucleotide polymorphisms. The 200 primers (Table 3) designed to hybridize to the genome sequence segments were found to have a combined total of 12822 hybridization sites. The mean separation distance of the genome sequence segments and the primers hybridizing thereto was found to be 815 nucleobases in length. The maximum distance between the genome sequence segments and the primers hybridizing thereto was found to be 5420 nucleobases in length. The mean "frequency bias" of hybridization of a primer to the target genome relative to the background genomes was calculated to be 3.31, indicating that the average primer hybridizes at 3.31 different positions on the target genome sequence for each single position it hybridizes to a background genome sequence.

[268] In an experiment designed to test the efficiency of the targeted whole genome amplification reaction vs. traditional whole genome amplification, reactions were carried out using 50, 100, 200, and 400 femtograms of *Bacillus anthracis* Sterne genomic DNA in the presence of 100 nanograms of human genomic DNA. Amplified quantities of DNA were determined and it was found that the targeted whole genome amplification reactions resulted in much greater specificity toward amplification of *Bacillus anthracis* Sterne genomic DNA than human genomic DNA. Figure 3A indicates that ordinary whole genome amplification using random primers 6 nucleobases in length under the conditions listed above results in production of larger quantities of human genomic DNA, as would be expected. Figure 3B, on the other hand indicates that the 200 primers described above selectively amplify the *Bacillus anthracis* Sterne genomic DNA relative to the human DNA, even though the quantity of *Bacillus anthracis* Sterne genomic DNA was much lower than the human genomic DNA.

[269] A second experiment was conducted where additional target genomes were selected for the primer design process. The group of total target genomes included the genomes of the following potential biowarfare agents: *Bacillus anthracis*, *Francisella tularensis, Yersinia pestis, Brucella sp., Burkholderia mallei, Rickettsia prowazekii, and Escherichia coli* 0157. The group of background genomes was expanded. An exact match BLAST was used to determine the frequency of occurrence of genome sequence segments in the background genomes. A larger number of genome sequence segments was analyzed and query 3 (Figure 2 - **6000**) was automated.

[270] The 200 primers designed in the first experiment are shown in Table 3 and the 191 primers designed in the second experiment are shown in Table 4. In Tables 3 and 4, an asterisk (*) indicates a phosphorothioate linkage and degenerate nucleobases codes are as follows: r = a or g; k = g or t; s = g or c; y = c or t; m = a or c, and w = a or t.

Table 3: First Generation Targeted Whole Genome Amplification Primer Set

Sequence	SEQ ID NO:
aaaaaagc*g*g	203
aaaacg*c*t	204
aaaagaagtt*a*t	205
aaaaggc*g*g	206
aaaccgc*c*a	207
aaaccgt*a*t	208
aaaccgt*t*a	209
aaagaagaag*t*t	210
aaagaagctt*t*a	211
aaagaagtat*t*a	212
aaagccg*a*t	213
aaagcgtggg*g*a	214
aaagtagaag*a*a	215
aaataacg*a*t	216
aaatacg*c*t	217
aaatcattaa*a*g	218
aaattag*c*g	219
aaccgcc*t*t	220
aacgat*t*g	221
aacgata*t*t	222
aacgctt*c*w	223
aacgtga*a*c	224

aacttctttt*t*c	225
aagaaac*g*c	226
aagarttaaa*a*g	227
aagataaaga*t*g	228
aagatgtaaa*a*g	229
aagcatctaa*g*c	230
aagcgat*c*a	231
aagcggt*t*c	232
aagtaac*g*a	233
aataacg*c*a	234
aatattggac*a*a	235
aatcattaat*a*t	236
aatccag*c*g	237
aatcgcc*c*a	238
aatcgta*t*c	239
aatcgtt*a*a	240
aatcgtt*g*c	241
aatctggtgg*t*a	242
aatgcg*g*t	243
aattaa*c*g	244
aatttcatct*a*a	245
accgata*a*t	246
accgcat*c*a	247
acgaatg*a*t	248
acgatgt*t*g	249
acggtta*t*c	250
acggttt*t*a	251
acgrtaa*a*a	252
acgttt*a*t	253
actttttat*c*t	254
agaattatta*a*a	255
agataaa*c*g	256
agatgaaaat*g*g	257
agcaatc*g*c	258
agcagttgca*g*c	259
agcgcaa*t*c	260
agcttgt*t*g	261

agttgat*c*g	262
ataaaaaaag*c*g	263
ataaaaaagg*t*a	264
ataaagaaga*t*g	265
ataaagatat*t*a	266
ataacga*a*g	267
ataactaata*a*a	268
ataatagaag*a*a	269
ataccatttt*t*a	270
atacgat*a*a	271
atagatgaaa*a*t	272
atagcga*t*a	273
atatcgt*a*a	274
atatcttttt*c*a	275
atattaaa*g*c	276
atattgaaga*a*g	277
atattgat*a*c	278
atcagct*a*c	279
atcatgc*c*g	280
atcgcac*c*g	281
atcgcctt*c*a	282
atcgtaa*t*a	283
atcgtga*a*g	284
atcgtta*a*a	285
atcttca*c*g	286
atcttcttta*a*t	287
attaata*c*c	288
attacaa*c*g	289
attacaac*a*a	290
attacc*g*c	291
attagaagaa*a*t	292
attatc*g*g	293
attatcg*t*a	294
attcatc*g*g	295
attgatat*t*a	296
attgatataa*a*t	297
attgatgaa*g*c	298

attgatgatt*t*a	299
attgcagc*a*a	300
atttagataa*a*t	301
atttagatga*a*g	302
atttatca*g*c	303
atttattatt*a*g	304
atttctttat*c*a	305
caatcgg*t*g	306
caatcgy*t*a	307
cacctttttt*a*a	308
cagcgat*t*a	309
cagcttttt*t*a	310
catcgct*t*c	311
catctaaaat*a*a	312
catcttc*c*g	313
ccaatcg*g*c	314
cccgctt*c*a	315
ccggtaa*t*a	316
cgataat*g*a	317
cgattaa*a*g	318
cgattg*c*g	319
cgcctct*t*c	320
cgctaaa*t*a	321
cgcttta*t*a	322
cggcgcgctg*a*a	323
cggtatt*g*a	324
cgtaaag*a*a	325
cgtaaat*a*c	326
cgtgatc*a*a	327
cgtttat*t*a	328
cgwtaat*a*a	329
ctaattcttc*t*a	330
ctacttttttc*c*a	331
ctgtagaaga*a*g	332
ctgttttaga*a*g	333
cttcacg*a*a	334
cttcatca*a*c	335

cttcatctaa*t*a	336
cttcttctaa*a*a	337
cttcttcttt*a*a	338
cttctttc*g*c	339
ctttagaaaa*t*a	340
ctttatataa*a*r	341
ctttatcaat*a*a	342
ctttcgct*t*c	343
cttttatata*a*a	344
ctttttcwtc*t*a	345
gaaaaaggat*t*a	346
gaaacga*t*c	347
gaaacgt*t*a	348
gaaattgctg*a*c	349
gaagaagyga*a*a	350
gaagatgaaa*a*a	351
gaagatttat*t*a	352
gaagtattaa*a*a	353
gaatatgaag*a*a	354
gatattgata*a*a	355
gatgaagata*a*a	356
gatttattat*t*a	357
gatttcacga*a*a	358
gcaata*a*c	359
gccttt*a*c	360
gcgaaag*a*a	361
gcgattt*t*a	362
gcggtat*t*a	363
gcgttaa*t*a	364
gcgttta*a*a	365
gcgtttt*g*a	366
gckgatt*t*a	367
gctaaaaaag*a*a	368
gctattttat*t*a	369
gctcgcgcga*c*a	370
gcttctttta*t*a	371
gctttttcat*c*a	372

ggcatt*a*c	373
ggcggta*a*a	374
ggttgaa*a*c	375
ggttta*a*c	376
gtaaaac*g*a	377
gtaaagcttt*c*a	378
gtgacga*a*a	379
gttatcg*c*a	380
gttgttttac*c*a	381
sttccgc*a*a	382
taaaatgggt*g*a	383
taaagcaatt*a*a	384
taaatcatct*a*a	385
taacgaa*g*a	386
taactcttct*a*a	387
taatgctt*c*a	388
tacatcat*c*a	389
tatcatc*g*a	390
tatcattaat*a*a	391
tatcctcttc*c*a	392
tcttctaata*a*a	393
tcttctaatt*c*a	394
tcttcttcta*a*a	395
tctttttta*c*a	396
tgacgat*a*a	397
tgatgcg*a*a	398
tgcttctttt*a*a	399
ttagatgaag*a*a	400
ttagctaaag*a*a	401
ttattagaag*a*a	402

Table 4: Second Generation Targeted Whole Genome Amplification Primer Set

Sequence	SEQ ID NO:
aaaacaat*t*g	403
aaaacgtt*t*a	404
aaaagaat*t*a	405
aaaaggta*t*t	406
aaaaggtg*a*a	407

aaataacg*a*t	216
aaatcgttga*t*a	409
aaatggtga*a*g	410
aacaccaa*t*t	411
aacgaaag*a*t	412
aacgaaagaa*g*a	413
aacgaat*a*a	414
aagaagcga*a*g	415
aagaagtaaa*a*g	416
aagcg*g*a	417
aatcgc*t*a	418
aatcgcaa*t*t	419
aatcgcygat*a*t	420
aatcgttt*c*a	421
acaacga*t*t	422
accgataa*t*a	423
acgaagc*a*a	424
agaagcgat*g*a	425
agcgaaaga*a*g	426
atacga*t*g	427
atacgg*a*a	428
atataaaa*g*a	429
atatg*c*g	430
atattatc*g*t	431
atcarcgatt*t*t	432
atcata*c*g	433
atccgt*t*a	434
atgaag*c*g	435
atgtaac*g*a	436
attaaagat*g*g	437
attaac*g*c	438
attacaaa*a*g	439
attacgat*a*a	440
attacgt*t*a	441
attacttg*t*a	442
attatatg*a*a	443
attattat*c*g	444

attgaaaaag*c*a	445
attgaaac*g*a	446
attgcttc*t*t	447
attgtcg*t*t	448
atttatcg*t*a	449
caacttct*t*t	450
caatcgt*a*t	451
caattaat*a*c	452
caattgga*a*t	453
caccaatt*a*c	454
caccaatt*g*t	455
cacctttta*c*a	456
catacg*a*a	457
catataa*c*g	458
catcaattg*t*t	459
ccgct*t*t	460
cgacttaccg*a*c	461
cgata*a*c	462
cgataaag*a*a	463
cgatataat*t*t	464
cgatg*t*a	465
cgattga*a*g	466
cgatttttc*a*a	467
cgcaa*t*a	468
cgcttttta*t*t	469
cggat*a*t	470
cggtaa*a*t	471
cggttta*a*t	472
cgtaat*a*t	473
cgtata*a*c	474
cgttaat*t*g	475
cgttatg*a*a	476
ctatcg*t*a	477
ctgattaaag*t*t	478
cttccata*a*t	479
cttcgt*a*a	480
cttctata*t*a	481

cttctgca*a*t	482
cttcttca*c*g	483
cttcttcttt*c*g	484
cttcttta*a*t	485
cttctttc*g*c	339
cttctttcg*g*a	487
ctttcgct*t*t	488
ctttcgcttc*t*t	489
cttttaattc*t*t	490
cttttgtaa*t*a	491
ctttttcg*t*a	492
cttttttc*a*t	493
ctttttya*t*c	494
gaaacgat*t*g	495
gaagaagcga*a*a	496
gaagaagt*a*a	497
gaagaagta*g*c	498
gatacgaa*a*g	499
gatgaatt*a*g	500
gatta*c*g	501
gattaaagtt*t*c	502
gcaattgaaa*a*a	503
gcaattgt*a*t	504
gcaattgt*t*g	505
gcgaaagaa*g*c	506
gcgtaa*t*a	507
gctacttt*a*t	508
gcttcttt*c*g	509
gctttttta*t*t	510
gtattaaaa*g*a	511
gttaattg*a*a	512
gttcg*t*a	513
gttgc*g*a	514
taaagataa*t*g	515
taaagcg*t*t	516
taaagtgaaa*c*t	517
taaatcttc*t*a	518

taacagaa*g*a	519
taacgaaaga*a*g	520
taacgga*a*a	521
taactcttc*t*t	522
taatam*c*g	523
taatcg*y*a	524
taatgaag*a*a	525
taattgct*t*c	526
tacaattt*c*a	527
taccgt*t*a	528
tacgaaaga*a*g	529
tacgaatg*a*t	530
tactcg*t*t	531
tagaagaa*g*t	532
tagaagaag*c*g	533
tagaagc*g*a	534
tatatcgact*t*a	535
tatatcrgcg*a*t	536
tatcggcgat*t*t	537
tatgtaa*c*g	538
tattag*c*g	539
tattcg*c*t	540
tattgatg*a*a	541
tawtacga*a*a	542
tcaattgc*a*a	543
tcaattgct*t*c	544
tcattac*g*a	545
tccaattg*a*a	546
tccgaaag*a*a	547
tccgct*a*a	548
tccgt*a*t	549
tcctgtta*c*a	550
tcgca*t*a	551
tcgcttta*t*t	552
tcgtat*t*g	553
tcgttaca*a*t	554
tctacaat*t*a	555

tctactaa*t*t	556
tcttcaat*a*t	557
tcttctaa*c*g	558
tctttata*t*g	559
tctttatat*t*c	560
tctttcgc*t*a	561
tctttttc*g*c	562
tgaaaaag*c*g	563
tgaaacaat*t*g	564
tgaaacga*a*t	565
tgaagcga*t*t	566
tgcaa*c*g	567
tgcgaaaga*a*a	568
tgcttcttc*t*a	569
tgtaaaag*g*t	570
tgtcggtaag*t*c	571
tgttctttc*g*t	572
ttaacgaaa*g*a	573
ttaacgg*a*a	574
ttacgaaa*g*a	575
ttagaaga*t*g	576
ttattatc*g*g	577
ttcaata*c*g	578
ttcacgaa*t*a	579
ttccgt*a*a	580
ttcgtaaa*t*t	581
ttcttta*c*g	582
ttctttcg*c*a	583
ttctttcgtt*a*a	584
ttctttta*t*a	585
ttgcaatt*g*c	586
ttgtaatt*g*g	587
ttgtcggta*a*g	588
tttattaga*t*g	589
tttcgtat*a*t	590
tttcgtta*t*a	591
tttwtcgt*a*a	592

twacgat*t*g	593

[271] Table 5 shows a comparison of statistics obtained from the first and second experiments. The statistics indicate the likelihood that more selective and efficient priming of the target *Bacillus anthracis* genome would be expected under the conditions of the second generation proof-of-concept experiment.

Table 5: Statistical Comparison of First and Second Experiments

Statistic	First Generation Experiment	Second Generation Experiment
Total Frequency of Occurrence of all Selected Genome Sequence Segments	12822	25822
Mean Separation Distance Between Selected Genome Sequence Segments	815	404
Maximum Separation Distance Between Selected Genome Sequence Segments	5420	3477
Average Frequency Bias to Target Genome Over Background Genomes	3.31	4.67

[272] The results of the second generation experiment are shown in Figures 4A and 4B. It is readily apparent that the modifications to the selection process added in the second experiment result in a more efficient targeted whole genome amplification reaction which is biased toward amplification of the *Bacillus anthracis* target genome. The primers of Table 4 produce less human DNA and more *Bacillus anthracis* DNA than the traditional whole genome amplification (WGA) and the first generation primer set (Table 3). Furthermore, the frequency bias was found to be even higher for the remaining target genomes as shown in Table 6.

Table 6: Statistical Comparison of Genome Sequence Segments for the Target Genomes of the Second Experiment

Target Genome	Total Frequency of Occurrence of Segments	Mean Separation Distance	Maximum Distance Between Segments	Mean Frequency Bias
Bacillus anthracis	25822	404.84	3477	4.67
Rickettsia prowazekii	5606	396.41	2265	5.44
Escherichia coli	23501	467.89	4822	22.70
Yersinia pestis	18597	500.43	4616	35.69
Brucella sp.	13442	490.10	3527	41.96
Francisella tularensis	7925	477.56	3179	50.08
Burkholderia mallei	25218	462.73	4062	291.13

Example 4: Targeted Whole Genome Amplification Protocol

[273] The targeted whole genome amplification reaction mixture consisted of: 5 microliters of template DNA, and 0.04025 M TRIS HCl, 0.00975 M TRIS base, 0.012 M $MgCl_2$, 0.01 M $(NH_4)_2SO_4$, 0.8 M betaine, 0.8 M trehalose, 25 mM of each deoxynucleotide triphosphate (Bioline, Randolph, MA, U.S.A),

0.004 M dithiothreitol, 0.05 mM of primers of the selected primer set, and 0.5 units of Phi29 polymerase enzyme per microliter of reaction mixture.

[274] The thermal cycling conditions for the amplification reaction were as follows:

1. 30°C for 4 minutes
2. 15 °C for 15 seconds
3. repeat steps 1 and 2 x 150
4. hold at 95 °C for 10 minutes
5. hold at 4 °C until ready for analysis

Example 5: Targeted Whole Genome Amplification of Sepsis-Causing Microorganisms

[275] This example is directed toward design of a kit for targeted whole genome amplification of organisms which are known to cause sepsis. A collection of target genomes is assembled, comprising the genomes of the following microorganisms known to cause bloodstream infections: *Escherichia coli, Klebsiella pneumoniae, Klebsiella oxytoca, Serratia marcescens, Enterobacter cloacae, Enterobacter aerogenes, Proteus mirabilis, Pseudomonas aeruginosa, Acinetobacter baumannii, Stenotrophomonas maltophilia, Staphylococcus aureus, Staphylococcus epidermidis, Staphylococcus haemolyticus, Streptococcus pneumoniae, Streptococcus pyogenes, Streptococcus agalactiae, Streptococcus mitis, Enterococcus faecium, Enterococcus faecalis, Candida albicans, Candida tropicalis, Candida parapsilosis, Candida krusei, Candida glabrata* and *Aspergillus fumigatus*. Because the healthy human bloodstream generally does not contain microorganisms or parasites, only the human genome is chosen as a single background genome. Alternatively, if a human was known to be infected with a virus such as HIV or HCV for example, the genomes of HIV or HCV could be included as background genomes during the primer design process. Genomes commonly found in the human bloodstream are considered background genomes.

[276] The target and background genomes are obtained from a genomics database such as GenBank. The target genomes are scanned by a computer program to identify all unique genome sequence segments between 5 and 13 nucleobases in length. The computer program further determines and records the frequency of occurrence of each of the unique genome sequence segments within each of the target genomes.

[277] The human genome is then scanned to determine the frequency of occurrence of the genome sequence segments. Optionally, the entire list of genome sequence segments is reduced by removing genome sequence segments that have low frequencies of occurrence by choosing an arbitrary frequency of occurrence threshold criterion such as, for example, the mean frequency of occurrence or any frequency of occurrence 25% above or below the mean frequency of occurrence or any whole or fractional percentage therebetween. For example, if the mean frequency of occurrence is 100, 25% above 100 equals 125 and

25% below 100 equals 75 and the frequency of occurrence threshold criterion may be any whole or fractional number between about 75 and about 125. When this step is complete, a subset of the original list of unique genome sequence segments remains. At this point, the subset of genome sequence subsets is analyzed by the computer program to determine the frequency of occurrence of each of the genome sequence segments within the human genome. Upon completion of this step, the genome sequence segments of the subset are associated with the following data; the frequency of occurrence within each of the target genomes and the frequency of occurrence within the human genome. A value indicating the total target frequency of occurrence is calculated by adding the frequency of occurrence of the genome sequence segments in each of the target genomes.

[278] The selectivity ratio is calculated by the computer program for the genome sequence segments of the subset by dividing the total target frequency of occurrence by the background frequency of occurrence. When the series of selectivity ratio calculations are complete, the genome sequence segments are ranked by their selectivity ratio values such that the highest selectivity ratio receives the highest rank. The ranked genome sequence segments are then subjected to the process described Example 2 and illustrated in Figure 2.

[279] The process of Example 2 and Figure 2 ends when the pre-determined quantity of 200 genome sequence segments is reached and when the stopping criteria are met. The stopping criteria are the following: the mean distance between the selected genome sequence segments on the target genomes is less than 500 nucleobases and the maximum distance between the selected genome sequence segments on the target genomes is less than 5000 nucleobases. These values are calculated by the computer program from the known coordinates of the target genomes and the selected genome sequence segments.

[280] The primer design step begins after completion of the selection process of the genome sequence segments. The genome sequence segments represent primer hybridization sites and a primer is designed to bind to each of the selected genome sequence segments. For an initial round of primer design and testing, primers are designed to be 100% complementary to each of the selected genome sequence segments. Optionally, the primers can be subjected to an *in silico* analysis to determine if they unfavorable characteristics. Unfavorable characteristics may include poor affinity (as measured by melting temperature) for their corresponding target genome sequence segment, primer dimer formation, or presence of secondary structure. Upon identification of unfavorable characteristics in a given primer, the primer is redesigned by alteration of length or by incorporation of modified nucleobases.

[281] Once primer design (and redesign if necessary) is complete, the primers are synthesized and subjected to *in vitro* testing by amplification of the target genomes in the presence of human DNA

(representing the background human genome) to determine the amplification efficiency and bias toward the target genomes. Analyses such as those shown in Figures 3 and 4 are useful for determining these measures. In addition, analyses of statistics such as those shown in Table 6 are useful for obtaining an estimation of bias toward the target genomes relative to the background human genome.

[282] When the primer design and testing is complete, kits are assembled. The kits contain the primers, deoxynucleotide triphosphates, a processive polymerase, buffers and additives useful for improving the yield of amplified genomes. These kits are used to amplify genomic DNA of sepsis-causing organisms from blood samples of individuals exhibiting symptoms of sepsis. The amplified DNA is then available for further testing for the purpose of genotyping. Such tests include real-time PCR, microarray analysis and triangulation genotyping analysis by mass spectrometry of bioagent identifying amplicons as described herein (Examples 6-12). Additionally, genotyping of sepsis-causing organisms is useful in determining an appropriate course of treatment with antibiotics and alerting authorities of the presence of potentially drug-resistant strains of sepsis-causing organisms. Such genotyping analyses can be developed using methods described herein as well as those disclosed in commonly owned U.S. Application Serial No 11/409,535 which is incorporated herein by reference in entirety.

Example 6: Design and Validation of Primer Pairs that Define Bioagent Identifying Amplicons for Identification of Bacteria

[283] For design of primers that define bacterial bioagent identifying amplicons, a series of bacterial genome segment sequences are obtained, aligned and scanned for regions where pairs of PCR primers would amplify products of about 39 to about 200 nucleotides in length and distinguish subgroups and/or individual strains from each other by their molecular masses or base compositions. A typical process shown in Figure 8 is employed for this type of analysis.

[284] 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., Nucl. Acids Res., 2001, 29, 4724-4735, which is incorporated herein by reference in its entirety) has been modified to include PCR parameters such as hybridization conditions, mismatches, and thermodynamic calculations (Santa Lucia, Proc. Natl. Acad. Sci. U.S.A., 1998, 95, 1460-1465, which is incorporated herein by reference in its entirety). This also provides information on primer specificity of the selected primer pairs. An example of a collection of such primer pairs is disclosed in U.S. Application Serial No. 11/409,535 which is incorporated herein by reference in entirety.

Example 7: Sample Preparation and PCR

[285] Genomic DNA id prepared from samples using the DNeasy Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer's protocols.

[286] PCR reactions are assembled in 50 μL reaction volumes in a 96-well microtiter plate format using a Packard MPII liquid handling robotic platform and M.J. Dyad thermocyclers (MJ research, Waltham, MA) or Eppendorf Mastercycler thermocyclers (Eppendorf, Westbury, NY). The PCR reaction mixture includes 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 mixture and 250 nM of each primer. The following typical PCR conditions are used: 95°C for 10 min followed by 8 cycles of 95°C for 30 seconds, 48°C for 30 seconds, and 72°C 30 seconds with the 48°C annealing temperature increasing 0.9°C with each of the eight cycles. The PCR reaction is then continued for 37 additional cycles of 95°C for 15 seconds, 56°C for 20 seconds, and 72°C 20 seconds.

Example 8: Purification of PCR Products for Mass Spectrometry with Ion Exchange Resin-Magnetic Beads

[287] For solution capture of nucleic acids with ion exchange resin linked to magnetic beads, 25 μl of a 2.5 mg/mL suspension of BioClone amine-terminated superparamagnetic beads is added to 25 to 50 μl of a PCR (or RT-PCR) reaction containing approximately 10 pM of a typical PCR amplification product. The above suspension is mixed for approximately 5 minutes by vortexing or pipetting, after which the liquid is removed after using a magnetic separator. The beads containing bound PCR amplification product are then washed three times with 50 mM ammonium bicarbonate/50% MeOH or 100mM ammonium bicarbonate/50% MeOH, followed by three more washes with 50% MeOH. The bound PCR amplification product is eluted with a solution of 25 mM piperidine, 25 mM imidazole, 35% MeOH which includes peptide calibration standards.

Example 9: Mass Spectrometry and Base Composition Analysis

[288] 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, are extracted directly from 96-well microtiter plates using a CTC HTS PAL autosampler (LEAP Technologies, Carrboro, NC) triggered by the FTICR data station. Samples are 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 are 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 metallized terminus of a glass desolvation capillary. The atmospheric pressure end of the glass capillary is biased at 6000 V relative to the ESI needle during data acquisition. A counter-current flow of dry N_2 is employed to assist in the desolvation process.

Ions are 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 are mass analyzed. Ionization duty cycles greater than 99% are achieved by simultaneously accumulating ions in the external ion reservoir during ion detection. Each detection event includes 1M data points digitized over 2.3 s. To improve the signal-to-noise ratio (S/N), 32 scans are co-added for a total data acquisition time of 74 s.

[289] 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 is also employed to improve ionization duty cycle during data acquisition. Each detection event on the TOF includes 75,000 data points digitized over 75 μs.

[290] 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 is 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 injects the next sample and the flow rate is switched to low flow. Following a brief equilibration delay, data acquisition commenced. As spectra are 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 are required to minimize sample carryover. During a routine screening protocol a new sample mixture is 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.

[291] Raw mass spectra are post-calibrated with an internal mass standard and deconvoluted to monoisotopic molecular masses. Unambiguous base compositions are 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. Calibration methods are commonly owned and disclosed in PCT Publication Number WO 2005/098047 which is incorporated herein by reference in entirety.

Example 10: *De Novo* Determination of Base Composition of Amplification Products using Molecular Mass Modified Deoxynucleotide Triphosphates

[292] 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 7), 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.

[293] The methods provide 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).

[294] 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 7). 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 7: 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

[295] 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.

[296] 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.

[297] 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.

[298] Base count blurring can be carried out as follows. "Electronic PCR" can be conducted on nucleotide sequences of the desired bioagents to obtain the different expected base counts that could be obtained for each primer pair. See for example, ncbi.nlm.nih.gov/sutils/e-pcr/; Schuler, *Genome Res.* 7:541-50, 1997. In one illustrative embodiment, one or more spreadsheets, such as Microsoft Excel workbooks contain a plurality of worksheets. First in this example, there is a worksheet with a name similar to the workbook name; this worksheet contains the raw electronic PCR data. Second, there is a worksheet named "filtered bioagents base count" that contains bioagent name and base count; there is a separate record for each strain after removing sequences that are not identified with a genus and species and removing all sequences for bioagents with less than 10 strains. Third, there is a worksheet that contains the frequency of substitutions, insertions, or deletions for this primer pair. This data is generated by first creating a pivot table from the data in the "filtered bioagents base count" worksheet and then executing an Excel VBA macro. The macro creates a table of differences in base counts for bioagents of the same species, but different strains. One of ordinary skill in the art may understand additional pathways for obtaining similar table differences without undo experimentation.

[299] Application of an exemplary script, involves the user defining a threshold that specifies the fraction of the strains that are represented by the reference set of base counts for each bioagent. The reference set of base counts for each bioagent may contain as many different base counts as are needed to meet or exceed the threshold. The set of reference base counts is defined by taking the most abundant strain's base type composition and adding it to the reference set and then the next most abundant strain's base type composition is added until the threshold is met or exceeded. The current set of data was obtained using a threshold of 55%, which was obtained empirically.

[300] For each base count not included in the reference base count set for that bioagent, the script then proceeds to determine the manner in which the current base count differs from each of the base counts in the reference set. This difference may be represented as a combination of substitutions, Si=Xi, and insertions, Ii=Yi, or deletions, Di=Zi. If there is more than one reference base count, then the reported difference is chosen using rules that aim to minimize the number of changes and, in instances with the same number of changes, minimize the number of insertions or deletions. Therefore, the primary rule is to identify the difference with the minimum sum (Xi+Yi) or (Xi+Zi), *e.g.*, one insertion rather than two substitutions. If there are two or more differences with the minimum sum, then the one that will be reported is the one that contains the most substitutions.

[301] Differences between a base count and a reference composition are categorized as one, two, or more substitutions, one, two, or more insertions, one, two, or more deletions, and combinations of

substitutions and insertions or deletions. The different classes of nucleobase changes and their probabilities of occurrence have been delineated in U.S. Patent Application Publication No. 2004209260 which is incorporated herein by reference in entirety.

Example 11: Selection and Use of Primer Pairs for Identification of Species of Bacteria involved in Sepsis

[302] In this example, identification of bacteria known to cause sepsis was accomplished using a panel of primer pairs chosen specifically with the aim of identifying these bacteria (Table 8). In this current example, the more specific group of bacteria known to be involved in causing sepsis is to be surveyed. Therefore, in development of this current panel of primer pairs, certain established surveillance primer pairs of U.S. Application Serial No. 11/409,535 have been combined with an additional primer pair, primer pair number 2249. The primer members of primer pair 2249 hybridize to the tufB gene and produce a bioagent identifying amplicon for members of the family *Staphylococcaceae* which includes the genus *Staphylococcus*.

Table 8: Names of Primer Pairs in Panel for Characterization of Septicemia Pathogens

Primer Pair No.	Forward Primer Name	Forward Primer Sequence	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer Sequence	Reverse Primer (SEQ ID NO:)
346	16S_EC_713_732_TMOD_F	TAGAACACCG ATGGCGAAGGC	594	16S_EC_789_809_TMOD_R	TCGTGGACT ACCAGGGT ATCTA	602
348	16S_EC_785_806_TMOD_F	TTTCGATGCA ACGCGAAGA ACCT	595	16S_EC_880_897_TMOD_R	TACGAGCTG ACGACAGC CATG	603
349	23S_EC_1826_1843_TMOD_F	TCTGACACCT GCCCGGTGC	596	23S_EC_1906_1924_TMOD_R	TGACCGTT ATAGTTAC GGCC	604
354	RPOC_EC_2218_2241_TMOD_F	TCTGGCAGGT ATGCGTGGTC TGATG	597	RPOC_EC_2313_2337_TMOD_R	TCGCACCG TGGGTTGAG ATGAAGTAC	605
358	VALS_EC_1105_1124_TMOD_F	TCGTGGCGGCG TGGTTATCGA	598	VALS_EC_1195_1218_TMOD_R	TCGGTACGA ACTGGATGT CGCCGTT	606
359	RPOB_EC_1845_1866_TMOD_F	TTATCGCTCAGG CGAACTCCAAC	599	RPOB_EC_1909_1929_TMOD_R	TGCTGGATT CGCCTTTG CTACG	607
449	RPLB_EC_690_710_F	TCCACACGGTG GTGGTGAAGG	600	RPLB_EC_737_758_R	TGTGCTGGT TTACCCCA TGGAG	608
2249	TUFB_NC002758-615038-616222_696_725_F	TGAACGTGGTC AAATCAAAGTT GGTGAAGA	601	TUFB_NC002758-615038-616222_793_820_R	TGTCACCAG CTTCAGCGTA GTCTAATAA	609

[303] To test for potential interference of human DNA with the present assay, varying amounts of bacterial DNA from *E. coli* 0157 and *E. coli* K-12 were spiked into samples of human DNA at various concentration levels. Amplification was carried out using primer pairs 346, 348, 349, 354, 358 and 359 and the amplified samples were subjected to gel electrophoresis. Smearing was absent on the gel, indicating that

the primer pairs are specific for amplification of the bacterial DNA and that performance of the primer pairs is not appreciably affected in the presence of high levels of human DNA such as would be expected in blood samples. Measurement of the amplification products indicated that *E. coli* 0157 could be distinguished from *E. coli* K-12 by the base compositions of amplification products of primer pairs 358 and 359. This is a useful result because *E. coli* 0157 is a sepsis pathogen and because *E. coli* K-12 is a low-level contaminant of the commercially obtained *Taq* polymerase used for the amplification reactions.

[304] A test of 9 blinded mixture samples was conducted as an experiment designed to simulate a potential clinical situation where bacteria introduced via skin or oral flora contamination could confound the detection of sepsis pathogens. The samples contained mixtures of sepsis-relevant bacteria at different concentrations, whose identities were not known prior to measurements. Tables 9A and 9B show the results of the observed base compositions of the amplification products produced by the primer pairs of Table 8 which were used to identify the bacteria in each sample. Without prior knowledge of the bacteria included in the 9 samples provided, it was found that samples 1-5 contained *Proteus mirabilis, Staphylococcus aureus,* and *Streptococcus pneumoniae* at variable concentration levels as indicated in Tables 9A and 9B. Sample 6 contained only *Staphylococcus aureus.* Sample 7 contained only *Streptococcus pneumoniae.* Sample 8 contained only *Proteus mirabilis.* Sample 9 was blank.

[305] Quantitation of the three species of bacteria was carried out using calibration polynucleotides as described herein. The levels of each bacterium quantitated for each sample was found to be consistent with the levels expected.

[306] This example indicates that the panel of primer pairs indicated in Table 8 is useful for identification of bacteria that cause sepsis.

[307] In another experiment, two blinded samples were provided. The first sample, labeled "Germ A" contained *Enterococcus faecalis* and the second sample, labeled "Germ B" contained other *Klebsiella pneumoniae.* For "Germ A" the panel of primer pairs of Table 8 produced four bioagent identifying amplicons from bacterial DNA and primer pair numbers 347, 348, 349 and 449 whose base compositions indicated the identity of "Germ A" as *Enterococcus faecalis.* For "Germ B" the panel of primer pairs of Table 8 produced six bioagent identifying amplicons from bacterial DNA and primer pair numbers 347, 348, 349, 358, 359 and 354 whose base compositions indicated the identity of "Germ B" as *Klebsiella pneumoniae.*

[308] One with ordinary skill in the art will recognize that one or more of the primer pairs of Table 8 could be replaced with one or more different primer pairs should the analysis require modification such

that it would benefit from additional bioagent identifying amplicons that provide bacterial identification resolution for different species of bacteria and strains thereof.

Table 9A: Observed Base Compositions of Blinded Samples of Amplification Products Produced with Primer Pair Nos. 346, 348, 349 and 449

Sample	Organism Component	Organism Concentration (genome copies)	Primer Pair Number 346	Primer Pair Number 348	Primer Pair Number 349	Primer Pair Number 449
1	*Proteus mirabilis*	470	A29G32C25T13	-	-	-
1	*Staphylococcus aureus*	>1000	-	A30G29C30T29	A26G3C25T20	-
1	*Streptococcus pneumoniae*	>1000	-	A26G32C28T30	A28G31C22T20	A22G20C19T14
2	*Staphylococcus aureus*	>1000	A27G30C21T21	A30G29C30T29	A26G30C25T20	-
2	*Streptococcus pneumoniae*	>1000	-	-	-	A22G20C19T14
2	*Proteus mirabilis*	390	-	-	-	-
3	*Proteus mirabilis*	>10000	A29G32C25T13	A29G30C28T29	A25G31C27T20	-
3	*Streptococcus pneumoniae*	675	-	-	-	A22G20C19T14
3	*Staphylococcus aureus*	110	-	-	-	-
4	*Proteus mirabilis*	2130	A29G32C25T13	A29G30C28T29	A25G31C27T20	-
4	*Streptococcus pneumoniae*	>3000	-	A26G32C28T30	A28G31C22T20	A22G20C19T14
4	*Staphylococcus aureus*	335	-	-	-	-
5	*Proteus mirabilis*	>10000	A29G32C25T13	A29G30C28T29	A25G31C27T20	-
5	*Streptococcus pneumoniae*	77	-	-	-	A22G20C19T14
5	*Staphylococcus aureus*	>1000				
6	*Staphylococcus aureus*	266	A27G30C21T21	A30G29C30T29	A26G30C25T20	-
6	*Streptococcus pneumoniae*	0		-	-	-
6	*Proteus mirabilis*	0	-	-	-	-
7	*Streptococcus pneumoniae*	125	-	A26G32C28T30	A28G31C22T20	A22G20C19T14
7	*Staphylococcus aureus*	0	-	-	-	-
7	*Proteus mirabilis*	0	-	-	-	-
8	*Proteus mirabilis*	240	A29G32C25T13	A29G30C28T29	A25G31C27T20	-
8	*Streptococcus pneumoniae*	0	-	-	-	-
8	*Staphylococcus aureus*	0	-	-	-	-
9	*Proteus mirabilis*	0	-	-	-	-
9	*Streptococcus pneumoniae*	0	-	-	-	-
9	*Staphylococcus aureus*	0	-	-	-	-

WO 2008/143627 PCT/US2007/020045

Table 9B: Observed Base Compositions of Blinded Samples of Amplification Products Produced with Primer Pair Nos. 358, 359, 354 and 2249

Sample	Organism Component	Organism Concentration (genome copies)	Primer Pair Number 358	Primer Pair Number 359	Primer Pair Number 354	Primer Pair Number 2249
1	*Proteus mirabilis*	470	-	-	A29G29C35T29	-
1	*Staphylococcus aureus*	>1000	-	-	A30G27C30T35	A43G28C19T35
1	*Streptococcus pneumoniae*	>1000	-	-	-	-
2	*Staphylococcus aureus*	>1000	-	-	A30G27C30T35	A43G28C19T35
2	*Streptococcus pneumoniae*	>1000	-	-	-	-
2	*Proteus mirabilis*	390	-	-	A29G29C35T29	-
3	*Proteus mirabilis*	>10000	-	-	A29G29C35T29	-
3	*Streptococcus pneumoniae*	675	-	-	-	-
3	*Staphylococcus aureus*	110	-	-	-	A43G28C19T35
4	*Proteus mirabilis*	2130	-	-	A29G29C35T29	-
4	*Streptococcus pneumoniae*	>3000	-	-	-	-
4	*Staphylococcus aureus*	335	-	-	-	A43G28C19T35
5	*Proteus mirabilis*	>10000	-	-	A29G29C35T29	-
5	*Streptococcus pneumoniae*	77	-	-	-	-
5	*Staphylococcus aureus*	>1000	-	-	-	A43G28C19T35
6	*Staphylococcus aureus*	266	-	-	-	A43G28C19T35
6	*Streptococcus pneumoniae*	0	-	-	-	-
6	*Proteus mirabilis*	0	-	-	-	-
7	*Streptococcus pneumoniae*	125	-	-	-	-
7	*Staphylococcus aureus*	0	-	-	-	-
7	*Proteus mirabilis*	0	-	-	-	-
8	*Proteus mirabilis*	240	-	-	A29G29C35T29	-
8	*Streptococcus pneumoniae*	0	-	-	-	-
8	*Staphylococcus aureus*	0	-	-	-	-
9	*Proteus mirabilis*	0	-	-	-	-
9	*Streptococcus pneumoniae*	0	-	-	-	-
9	*Staphylococcus aureus*	0	-	-	-	-

Example 12: Design and Validation of Primer Pairs Designed for Production of Amplification Products from DNA of Sepsis-Causing Bacteria

[309] The following primer pairs of Table 10 were designed to provide an improved collection of bioagent identifying amplicons for the purpose of identifying sepsis-causing bacteria.

Table 10: Primer Pairs for Producing Bioagent Identifying Amplicons of Sepsis-Causing Bacteria

Primer Pair Number	Forward Primer Name	Forward Sequence	Forward SEQ ID NO:	Reverse Primer Name	Reverse Sequence	Reverse SEQ ID NO:
3346	RPOB_ NC000913_ 3704_3731_F	TGAACCACT TGGTTGACGA CAAGATGCA	616	RPOB_ NC000913_ 3793_3815_R	TCACCGAAACGC TGACCACCGAA	627
3347	RPOB_ NC000913_ 3704_3731_F	TGAACCACTT GGTTGACGA CAAGATGCA	616	RPOB_ NC000913_ 3796_3821_R	TCCATCTCACCG AAACGCTGA CCACC	632
3348	RPOB_ NC000913_ 3714_3740_F	TGTTGATGA CAAGATGCA CGCGCGTTC	623	RPOB_ NC000913_ 3796_3821_R	TCCATCTCACC GAAACGCTGA CCACC	632
3349	RPOB_ NC000913_ 3720_3740_F	TGACAAGA TGCACGCG CGTTC	619	RPOB_ NC000913_ 3796_3817_R	CTCACCGAAACGCT ACCACC	636
3350	RPLB_EC_ 690_710_F	TCCACACGG TGGTGGT GAAGG	614	RPLB_ NC000913_ 739_762_R	TCCAAGCGCAG GTTTACCCC ATGG	630
3351	RPLB_EC_ 690_710_F	TCCACACGG TGGTGGT GAAGG	614	RPLB_ NC000913_ 742_762_R	TCCAAGCGCAG GTTTACCCCA	628
3352	RPLB_ NC000913_ 674_698_F	TGAACCCTA ATGATCAC CCACACGG	618	RPLB_ NC000913_ 739_762_R	TCCAAGCGCAGG TTTACCCCATGG	630
3353	RPLB_ NC000913_ 674_698_2_F	TGAACCCTAA CGATCACC CACACGG	617	RPLB_ NC000913_ 742_762_R	TCCAAGCGCA GGTTTACCCCA	629
3354	RPLB_EC_ 690_710_F	TCCACACGG TGGTGGTG AAGG	614	RPLB_ NC000913_ 742_762_2_R	TCCAAGCGCT GGTTTACCCCA	631
3355	PLB_NC000913_6 _680_F	TCCAACTGTTC GTGGTTCTGT AATGAACCC	613	RPLB_ NC000913_ 739_762_R	TCCAAGCGCAG GTTTACCCC ATGG	630
3356	RPOB_ NC000913_ 3789_3812_F	TCAGTTCGGT GGCCAGCGC TTCGG	610	RPOB_ NC000913_ 3868_3894_R	TACGTCGTCCG ACTTGACCG TCAGCAT	625
3357	RPOB_ NC000913_ 3789_3812_F	TCAGTTCGG TGGCCAGC GCTTCGG	610	RPOB_ NC000913_ 3862_3887_R	TCCGACTTGAC CGTCAGCAT CTCCTG	633
3358	RPOB_ NC000913_ 3789_3812_2_F	TCAGTTCGG TGGTCAGCG CTTCGG	611	RPOB_ NC000913_ 3862_3890_R	TCGTCGGACTT GATGGTCAGC AGCTCCTG	635
3359	RPOB_ NC000913_ 3739_3761_F	TCCACCGGTC CGTACTCC ATGAT	615	RPOB_ NC000913_ 3794_3812_R	CCGAAGCGCTG GCCACCGA	624
3360	GYRB_ NC002737_ 852_879_F	TCATACTCA TGAAGGTGG AACGCATGAA	612	GYRB_ NC002737_ 973_996_R	TGCAGTCAAGC CTTCACGAA CATC	637
3361	TUFB_ NC002758_ 275_298_F	TGATCACTG GTGCTGCTC AAATGG	620	TUFB_ NC002758_ 337_362_R	TGGATGTGTTC ACGAGTTTGA GGCAT	638

3362	VALS_ NC000913_ 1098_1115_F	TGGCGACCG TGGCGGCGT	621	VALS_ NC000913_ 1198_1226_R	TACTGCTTCGG GACGAACTG GATGTCGCC	626
3363	VALS_ NC000913_ 1105_1127_F	TGTGGCGGCG TGGTTATCG AACC	622	VALS_ NC000913_ 1207_1229_R	TCGTACTGCTT CGGGACGA ACTG	634

[310] Primer pair numbers 3346-3349, and 3356-3359 have forward and reverse primers that hybridize to the rpoB gene of sepsis-causing bacteria. The reference gene sequence used in design of these primer pairs is an extraction of nucleotide residues 4179268 to 4183296 from the genomic sequence of *E. coli* K12 (GenBank Accession No. NC_000913.2, gi number 49175990). All coordinates indicated in the primer names are with respect to this sequence extraction. For example, the forward primer of primer pair number 3346 is named RPOB_NC000913_3704_3731_F (SEQ ID NO: 616). This primer hybridizes to positions 3704 to 3731 of the extraction or positions 4182972 to 4182999 of the genomic sequence. Of this group of primer pairs, primer pair numbers 3346-3349 were designed to preferably hybridize to the rpoB gene of sepsis-causing gamma proteobacteria. Primer pairs 3356 and 3357 were designed to preferably hybridize to the rpoB gene of sepsis-causing beta proteobacteria, including members of the genus *Neisseria*. Primer pairs 3358 and 3359 were designed to preferably hybridize to the rpoB gene of members of the genera *Corynebacterium* and *Mycobacterium*.

[311] Primer pair numbers 3350-3355 have forward and reverse primers that hybridize to the rplB gene of gram positive sepsis-causing bacteria. The forward primer of primer pair numbers 3350, 3351 and 3354 is RPLB_EC_690_710_F (SEQ ID NO: 614). This forward primer had been previously designed to hybridize to GenBank Accession No. NC_000913.1, gi number 16127994. The reference gene sequence used in design of the remaining primers of primer pair numbers 3350-3355 is the reverse complement of an extraction of nucleotide residues 3448565 to 3449386 from the genomic sequence of *E. coli* K12 (GenBank Accession No. NC_000913.2, gi number 49175990). All coordinates indicated in the primer names are with respect to the reverse complement of this sequence extraction. For example, the forward primer of primer pair number 3352 is named RPLB_NC000913_674_698_F (SEQ ID NO: 634). This primer hybridizes to positions 674-698 of the reverse complement of the extraction or positions 3449239 to 3449263 of the reverse complement of the genomic sequence. This primer pair design example demonstrates that it may be useful to prepare new combinations of primer pairs using previously existing forward or reverse primers.

[312] Primer pair number 3360 has a forward primer and a reverse primer that both hybridize to the gyrB gene of sepsis-causing bacteria, preferably members of the genus *Streptococcus*. The reference gene sequence used in design of these primer pairs is an extraction of nucleotide residues 581680 to 583632 from the genomic sequence of *Streptococcus pyogenes* M1 GAS (GenBank Accession No. NC_002737.1, gi number 15674250). All coordinates indicated in the primer names are with respect to this sequence extraction. For example, the forward primer of primer pair number 3360 is named GYRB_NC002737_852_879_F (SEQ ID NO: 612). This primer hybridizes to positions 852 to 879 of the extraction.

[313] Primer pair number 3361 has a forward primer and a reverse primer that both hybridize to the tufB gene of sepsis-causing bacteria, preferably gram positive bacteria. The reference gene sequence used in design of these primer pairs is an extraction of nucleotide residues 615036 ... 616220 from the genomic sequence of *Staphylococcus aureus subsp. aureus* Mu50 (GenBank Accession No. NC_002758.2, gi number 57634611). All coordinates indicated in the primer names are with respect to this sequence extraction. For example, the forward primer of primer pair number 3361 is named TUFB_NC002758_275_298_F (SEQ ID NO: 612). This primer hybridizes to positions 275 to 298 of the extraction.

[314] Primer pair numbers 3362 and 3363 have forward and reverse primers that hybridize to the valS gene of sepsis-causing bacteria, preferably including *Klebsiella pneumoniae* and strains thereof. The reference gene sequence used in design of these primer pairs is the reverse complement of an extraction of nucleotide residues 4479005 to 4481860 from the genomic sequence of *E. coli* K12 (GenBank Accession No. NC_000913.2, gi number 49175990). All coordinates indicated in the primer names are with respect to the reverse complement of this sequence extraction. For example, the forward primer of primer pair number 3362 is named VALS_NC000913_1098_1115_F (SEQ ID NO: 621). This primer hybridizes to positions 1098 to 1115 of the reverse complement of the extraction.

[315] In a validation experiment, samples containing known quantities of known sepsis-causing bacteria were prepared. Total DNA was extracted and purified in the samples and subjected to amplification by PCR according to Example 2 and using the primer pairs described in this example. The three sepsis-causing bacteria chosen for this experiment were *Enterococcus faecalis, Klebsiella pneumoniae*, and *Staphylococcus aureus*. Following amplification, samples of the amplified mixture were purified by the method described in Example 3 subjected to molecular mass and base composition analysis as described in Example 4.

[316] Amplification products corresponding to bioagent identifying amplicons for *Enterococcus faecalis* were expected for primer pair numbers 3346-3355, 3360 and 3361. Amplification products were obtained and detected for all of these primer pairs.

[317] Amplification products corresponding to bioagent identifying amplicons for *Klebsiella pneumoniae* were expected and detected for primer pair numbers 3346-3349, 3356, 3358, 3359, 3362 and 3363. Amplification products corresponding to bioagent identifying amplicons for *Klebsiella pneumoniae* were detected for primer pair numbers 3346-3349 and 3358.

[318] Amplification products corresponding to bioagent identifying amplicons for *Staphylococcus aureus* were expected and detected for primer pair numbers 3348, 3350-3355, 3360, and 3361. Amplification products corresponding to bioagent identifying amplicons for *Klebsiella pneumoniae* were detected for primer pair numbers 3350-3355 and 3361.

Example 13: Selection of Primer Pairs for Genotyping of Members of the Bacterial Genus *Mycobacterium* and for Identification of Drug-Resistant Strains of *Mycobacterium tuberculosis*

[319] To combine the power of high-throughput mass spectrometric analysis of bioagent identifying amplicons with the sub-species characteristic resolving power provided by genotyping analysis and codon base composition analysis, a panel of twenty-four genotyping analysis primer pairs was selected. The primer pairs are designed to produce bioagent identifying amplicons within sixteen different housekeeping genes indicated by primer name codes in Table 11; rpoB, embB, fabG-inhA, katG, gyrA, rpsL, pncA, rv2109c, rv2348c, rv3815c, rv0041, rv00147, rv1814, rv0005gyrB, and rv0260c. The primer sequences are listed in Table 11.

[320] In *Mycobacterium tuberculosis*, the acquisition of drug resistance is mostly associated with the emergence of discrete key mutations that can be unambiguously determined using the methods disclosed herein.

[321] The evolution of the *Mycobacterium tuberculosis* genome is essentially clonal, thus allowing strain typing through the query of distinct genomic markers that are lineage-specific and only vertically inherited. Co-infections of mixed populations of genotypes of *Mycobacterium tuberculosis* can be revealed simultaneously in the mass spectra of amplification products produced using the primers of Table 11. The high G+C content and of the *Mycobacterium tuberculosis* genome itself greatly facilitates the development of short, efficient primers which are appropriate for multiplexing (inclusion of a plurality of primers in each amplification reaction mixture).

Table 11: Primer Pairs for Genotyping and Determination of Drug Resistance of Strains of *Mycobacterium tuberculosis*

Primer Pair No.	Forward Primer Name	Forward Primer Sequence	Forward Primer (SEQ ID NO:)	Reverse Primer Name	Reverse Primer Sequence	Reverse Primer (SEQ ID NO:)
3546	RPOB_ L27989-1-5084_ 2333_2351_F	TGTGGCCGCG ATCAAGGAG	670	RPOB_L27989-1-5084_2458_2474_R	TAGCCCGGC ACGCTCAC	694
3547	RPOB_ L27989-1-5084_ 2362_2384_F	TCAGCCAGC TGAGCCAATT CATG	671	RPOB_L27989-1-5084_2388_2407_R	TCCGACAG CGGGTTGTTCTG	695
3548	RPOB_ L27989-1-5084_ 2397_2414_F	TCGCTGTCGGG GTTGACC	672	RPOB_L27989-1-5084_2418_2434_R	TCCGACAGT CGGCGCTT	696
3550	EMBB_ AY727532-1-	TGCTCTGGCAT GTCATCGGC	673	EMBB_AY727532-1-344_209_228_R	TGAAGGGAT CCTCCGGGCTG	697

	344_100_119_F					
3551	EMBB_ AY727532-1-344_134_152_F	TGACGGCTACA TCCTGGGC	674	EMBB_AY727532-1-344_160_176_R	TGCGTGGTC GGCGACTC	698
3552	FABG-INHA-PROMOTER_ U66801-1-993_169_191_F	TGCTCGTGGAC ATACCGA TTTCG	675	FABG-INHA-PROMOTER_ U66801-1-993_224_243_R	TCAGTGGCTGT GGCAGTCAC	699
3553	KATG_U06268-1-2324_991_1010_F	TCGGTAAGGAC GCGATCACC	676	KATG_U06268-1-2324_1014_1034_R	TGTCCATACG ACCTCGATGCC	700
3554	KATG_U06268-1-2324_1433_1454_F	TGCCAGCCTTA AGAGCCAGATC	677	KATG_U06268-1-2324_1458_1480_R	TGTGAGACAGTC AATCCCGATGC	701
3555	GYRA_AF400983-1-385_69_84_F	TCACCCGCAC GGCGAC	678	GYRA_AF400983-1-385_103_119_R	TGGGCCA TGCGCACCAG	702
3556	GYRA_AF400983-1-385_80_99_F	TCGACGCGTCG ATCTACGAC	679	GYRA_AF400983-1-385_103_119_R	TGGGCCATG CGCACCAG	702
3557	RPSL_AY156733-1-375_65_82_F	TGGCTCTGAAG GGCAGCC	680	RPSL_AY156733-1-375_177_195_R	TGCCGTGACCT CGACCTGA	703
3558	PNCA_AL123456.2 _gi41353971-1-4411532_2289165_ 2289181_F (RC)	TCTGTGGCTGC CGCGTC	681	PNCA_AL123456. 2_gi41353971-1-4411532_2289303 2289287_R (RC)	TCGGCGCCA CCGGTTAC	704
3559	PNCA_AL123456.2 _gi41353971-1-4411532_2288970_ 2288989_F (RC)	TCATCACGTCG TGGCAACCA	682	PNCA_AL123456.2 _gi41353971-1-4411532_2289119 2289098_R (RC)	TACGTGTCCAG ACTGGGATGGA	705
3560	PNCA_AL123456.2 _gi41353971-1-4411532_2288815_ 2288832_F (RC)	TGTGCCTACAC CGGAGCG	683	PNCA_AL123456.2 _gi41353971-1-4411532_2288953 2288933_R (RC)	TCGTCTGGCGC ACACAATGAT	706
3561	PNCA_AL123456.2 _gi41353971-1-4411532_2288710_ 2288729_F (RC)	TCCGATCATTG TGTGCGCCA	684	PNCA_AL123456.2 _gi41353971-1-4411532_2288839 2288821_R (RC)	TGGTGCGCATC TCCTCCAG	707
3581	RV2109C_ AL123456.2 _gi41353971-1-4411532_2369291 2369316_F	TCGACCCGTC GTAGGTAATA CGATAC	685	RV2109C_AL123456.2 _gi41353971-1-4411532_2369342 _2369358_R	TGCCGAGGT GGCGCATT	708
3582	RV2348C_ AL123456.2 _gi41353971-1-4411532_2627916_ 2627940_F	TGCCTGTTTGA AACTGCCCA CATAC	686	RV2348C_AL123456.2 _gi41353971-1-4411532 _2627954_2627974_R	TCGGGCTCAACG ACACTTCCT	709
3583	RV3815C_ NC000962-1-4411532_4280680_ 4280699_F	TGCCTTGGTCG GGCACATTC	687	RV3815C_AL123456.2 _gi41353971-1-4411532 _4280716_4280734_R	TCCACCGGAA CCCGGATCA	710
3584	RV0041_AL123456.2 _gi41353971-1-4411532_43921_ 43939_F	TCTGCCCGCCG AGCAATAC	688	RV0041_AL123456.2 _gi41353971-1-4411532 _43960_43976_R	TGGTCCGGGT ACGCGGA	711
3586	RV0147_AL123456.2 _gi41353971-1-4411532_174655_ 174678_F	TCCGTAAGTC GGTGTTGA CCAAAC	689	RV0147_AL123456.2 _gi41353971 -1-411532_174694 174716_R	TGGCGGGTAGA TAAAGCTGGACA	712
3587	RV1814_AL123456.2 _gi41353971-1-4411532_2057117_	TCGGGTCCACC ACGGAATG	690	RV1814_AL123456.2 _gi41353971-1-4411532_2057151	TGGATGCCGCC ATAGTTCTTGTC	713

	2057135_F			_2057173_R		
3599	RV0083_AL123456.2 _gi41353971-1-4411532_92169_ 92187_F	TGCCGACGCGA TCGAACAG	691	RV0083_AL123456. 2_gi41353971-1-4411532_ 92220_92238_R	TAACAGCTCGG CCATGGCG	714
3600	RV0005GYRB_ AL123456.2 _gi41353971-1-4411532_6348_ 6368_F	TGACCAA GACC AAGTTGGGCA	692	RV0005GYRB _AL123456.2 _gi41353971-1-4411532 6457_6478_R	TGAGGACACAG CC TTGTTCACA	715
3601	RV0260C_AL123456. 2_gi41353971-1-4411532_311588_ 311604_F	TGCCCAGAGC CGTTCGT	693	RV0260C_AL123456.2 _gi41353971-1-4411532_311623_ 311639_2_R	TACACCCACGCC GTGGA	716

[322] The panel of 24 primer pairs is designed to be multiplexed into 8 amplification reactions. Thirteen primer pairs were designed with the objective of identifying mutations associated with resistance to drugs including rifampin (primer pair numbers 3546, 3547 and 3548), ethambutol (primer pair numbers 3550 and 3551), isoniazid (primer pair numbers 3353 and 3354), fluoroquinolone (primer pair number 3556), streptomycin (primer pair number 3557) and pyrazinamide (primer pair numbers 3558, 3558, 3560 and 3561). Four of these thirteen primer pairs were specifically designed to provide bioagent identifying amplicons for base composition analysis of single codons (primer pair numbers 3547 (rpoB codon D526), 3548 (rpoB codon H516), 3551 (embB codon M306), and 3553 (katG codon S315)). In any of these bioagent identifying amplicons used for base composition analysis, detection of a mutation identifies a drug-resistant strain of *Mycobacterium tuberculosis*. The remaining nine primer pairs define larger bioagent identifying amplicons that contain secondary drug resistance-conferring sites which are more rare than the four codons discussed above, but certain of these nine primer pairs define bioagent identifying amplicons that also contain some of these four codons (for example, primer pair 3546 contains two rpoB codons; D526 and H516).

[323] Shown in Table 12 are classifications of members of the bacterial genus *Mycobacterium* according to principal genetic group (PGG, determined using primer pair numbers X and X), genotype of *Mycobacterium tuberculosis*, or species of selected other members of the genus *Mycobacterium* (determined using primer pair numbers X, Y, Z), and drug resistance to rifampin, ethambutol, isoniazid, fluoroquinolone, streptomycin, and pyrazinamide. The primer pairs used to define the bioagent identifying amplicons for each PPG group, genotype or drug resistant strain are shown in the column headings. In the drug resistance columns, codon mutations are indicated by the amino acid single letter code and codon position convention which is well known to those with ordinary skill in the art. For example, when nucleic acid of *Mycobacterium tuberculosis* strain 13599 is amplified using primer pair number 3555, and the molecular mass or base composition is determined, mutation of codon 90 from alanine (A) to valine (V) is indicated and the conclusion is drawn that strain 13599 is resistant to the drug fluoroquinolone.

[324] Primer pair number 3600 is a speciation primer pair which is useful for distinguishing members of *Mycobacterium tuberculosis* PPG1 (including genotypes I, II and IIA) from other species of the

genus *Mycobacterium* (such as for example, *Mycobacterium africanum, Mycobacterium bovis, Mycobacterium microti,* and *Mycobacterium canettii*).

Table 12: Classification and Drug Resistance Profiles of Strains of Members of the Genus *Mycobacterium* and Genotypes of *Mycobacterium tuberculosis*

Strain	Principal Genetic Group (PGG) --- Primer Pair Numbers: 3554,3556	Genotype --- imer Pair Numbe 3581,3582, 3583, 3584, 3586, 3587, 3599,3600,3601	Drug Resistance to Rifampin --- Primer Pair Numbers: 3546, 3547,3548	Drug Resistance to Ethambutol --- Primer Pair Numbers: 3550,3551	Drug Resistance to Isoniazid --- Primer Pair Numbers: 3553 3552	Drug Resistance to Fluoroquinolone --- Primer Pair Number: 3555	Drug Resistance to Streptomycin --- Primer Pair Number: 3557	Drug Resistance to Pyrazinamide --- Primer Pair Numbers: 3558,3559, 3560, 3561
19422	PGG-1	*M africanum* or *M. microti*	wild type	wt	wt wt	wt	wt	wt
10130	PGG-1	*M. bovis*	wt	wt	wt wt	wt	wt	[part2] C>G
35737 (BCG)	PGG-1	*M. bovis*	wt	wt	wt wt	wt	wt	wt
M. Canettii	PGG-1	*M. canettii*	wt	wt	wt wt	wt	wt	[part2] C>G
14157, 15042	PGG-1	**I**	wt	wt	wt wt	wt	wt	wt
16116	PGG-1	**IIA**	wt	wt	wt wt	wt	wt	wt
15021	PGG-1	**IIA**	wt	wt	wt wt	wt	wt	[part2] C>T
5116	PGG-1	**IIA**	wt	wt	S315T wt	wt	wt	wt
12360, 13876, 14149	PGG-1	**II**	wt	wt	wt wt	wt	wt	wt
13599	PGG-1	**II**	wt	wt	wt C-15T	A90V	wt	[part2] A>G
13598	PGG-1	**II**	H528Y	M306V	S315(N/T) wt	wt	K43R	wt
10545	PGG-1	**II**	wt	M306I	S315T wt	wt	wt	wt
13632	PGG-1	**II**	transition	M306I	S315T wt	wt	wt	[part2] C>T, [part3] G>C
14207	PGG-1	**III**	wt	wt	wt wt	wt	wt	wt
13866, 13874, 14038	PGG-2	**III or IV**	wt	wt	wt wt	wt	wt	wt
12578, 12590	PGG-2	**III or IV**	wt	wt	S315T wt	wt	wt	[part3] G>C
14404	PGG-2	**IV**	wt	wt	wt wt	wt	wt	wt

14831	PGG-2	**IV**	wt	wt	S315T T-8C	wt	wt	wt
5170, 13672, 13699, 14424	PGG-2	**V**	wt	wt	wt wt	wt	wt	wt
13679, 14399	PGG-2	**VI**	wt	wt	wt wt	wt	wt	wt
13592	PGG-2	**VI**	wt	wt	S315T wt	wt	wt	wt
13594, 13658, 13869	PGG-3	**VII**	wt	wt	wt wt	T95S	wt	wt
13821	PGG-3	**VIII**	wt	wt	wt wt	T95S	wt	wt
35837 (H37Rv7)	PGG-3	**VIII**	wt	M306V	wt wt	T95S	wt	wt

Example 14: Validation of the Panel of 24 Primer Pairs

[309] Each primer pair was individually validated using the reference *Mycobacterium tuberculosis* strain H37Rv. Dilution To Extinction (DTE) experiments yielded the expected base composition down to 16 genomic copies per well. A multiplexing scheme was then determined in order to spread into different wells the primer pairs targeting the same gene, to spread within a single well the expected amplicon masses, and to avoid cross-formation of primer duplexes. The multiplexing scheme is shown in Table 13 where multiplexed amplification reactions are indicated in headings numbered A through H and the primer pairs utilized for each reaction are shown below.

Table 13: Multiplexing Scheme for Panel of 24 Primer Pairs

Reaction A	Reaction B	Reaction C	Reaction D	Reaction E	Reaction F	Reaction G	Reaction H
3547	3548	3601	3551	3553	3554	3555	3556
3581	3584	3599	3582	3583	3587	3552	3586
3550	3600	3559	3560	3546	3558	3561	3557

[310] An example of an experimentally determined table of base compositions is shown in Table 14. Base compositions of amplification products obtained from nucleic acid isolated from *Mycobacterium tuberculosis* strain 5170 using the primer pair multiplex reactions indicated in Table 13 are shown. Molecular masses of the amplification products were measured by electrospray time of flight mass spectrometry in order to calculate the base compositions. It should be noted that the lengths of the amplification products within each reaction mixture vary greatly in length in order to avoid overlap of molecular masses during the measurements. For example, reaction A has three amplification products which have lengths of 46 (A13 T11 C15 G07), 68 (A14 T18 C21 G15) and 129 (A21 T37 C44 G27).

Table 14: Base Compositions Obtained in the Multiplex Amplification Reactions of Nucleic Acid of *Mycobacterium tuberculosis* Strain 5170

Reaction	Primer Pair No.	Base Composition (A G C T)
A	3547	13 11 15 07
A	3581	14 18 21 15
A	3550	21 37 44 27
B	3548	06 13 12 07
B	3584	13 13 24 06
B	3600	37 34 35 25
C	3601	07 20 15 10
C	3599	10 26 22 12
C	3559	26 34 53 28
D	3551	08 13 16 06
D	3582	13 15 17 14
D	3560	28 48 37 26
E	3553	11 15 11 07
E	3583	06 19 16 14
E	3546	-

F	3554	11 13 14 10
F	3587	15 16 16 10
F	3558	-
G	3555	09 14 21 07
G	3552	13 26 22 14
G	3561	22 48 39 21
H	3556	07 11 15 07
H	3586	15 11 23 13
H	3557	26 44 39 22

[311] Dilution to extinction experiments were then carried out with the chosen triplets of primer pairs in multiplex conditions. Base compositions expected on the basis of the known sequence of the reference strain were observed down to 32 genomic copies per well on average. The assay was finally tested using a collection of 36 diverse strains from the Public Health Research Institute. As expected, the base compositions results were in accordance with the genotyping and drug-resistance profiles already determined for these reference strains.

Example 15: Primer Pairs that Define Bioagent Identifying Amplicons for Hepatitis C Viruses

[312] For design of primers that define hepatitis c virus strain identifying amplicons, a series of hepatitis C virus genome sequences were obtained, aligned and scanned for regions where pairs of PCR primers would amplify products of about 27 to about 200 nucleotides in length and distinguish strains and quasispecies from each other by their molecular masses or base compositions.

[313] Table 15 represents a collection of primers (sorted by primer pair number) designed to identify hepatitis C viruses using the methods described herein. The primer pair number is an in-house database index number. The forward or reverse primer name shown in Table 15 indicates the gene region of the viral genome to which the primer hybridizes relative to a reference sequence. In Table 15, for example, the forward primer name HCVUTR5_NC001433-1-9616_9250_9273_F indicates that the forward primer (_F) hybridizes to residues 9250-9275 of the UTR (untranslated region) of a hepatitis C virus reference sequence represented by an extraction of nucleotides 1 to 9616 of GenBank Accession No. NC_001433.1. One with ordinary skill will know how to obtain individual gene sequences or portions thereof from genomic sequences present in GenBank.

Table 15: Primer Pairs for Identification of Strains of Hepatitis C Viruses

Primer Pair No.	Forward Primer Name	Forward Sequence	Forward SEQ ID NO:	Reverse Primer Name	Reverse Sequence	Reverse SEQ ID NO:
3682	HCVUTR5_ NC001433-1-9616_ 9250_9273_F	TCAGCGGA GGTGACAT GTATCACA	655	HCVUTR5_ NC001433-1-9616_ 9313_9337_R	TACTCCTCC TTTCGGTA GCGGTAGA	662
3683	HCVUTR5_ NC001433-1-9616_ 9177_9200_F	TCGACCAAC CTTAAACG CACTCCA	656	HCVUTR5_ NC001433-1-9616_ 9261_9285_R	GACATGTAT CACAACCT GTCGCACA	663
3684	HCVUTR5_ NC001433-1-9616_ 3644_3662_F	TTAGCACC TCGACGG CTGG	657	HCVUTR5_ NC001433-1-9616_ 3735_3756_R	CATGCTAAT GTCGTTCC GGCGA	664
3685	HCVUTR5_ NC001433-1-9616_ 3708_3731_F	TGCTCGGA CCTTTACT TGGTCACG	658	HCVUTR5_ NC001433-1-9616_ 3735_3757_R	CATGCTGAT GTCATTCCG GTGCA	665
3686	HCVUTR5_ NC001433-1-9616_ 3708_3731_F	TGCTCGGA CCTTTAC TTGGTCACG	658	HCVUTR5_ NC001433-1-9616_ 3822_3840_R	TCGGGTGGTC CACTGCTCA	666
3687	HCVUTR5_ NC001433-1-9616_ 3796_3817_F	TGCCCGT CTCCTAC TTGAAGGG	659	HCVUTR5_ NC001433-1-9616_ 3876_3893_R	GCTGTGTACAC CCGGCGA	667
3688	HCVUTR5_ NC001433-1-9616_ 3855_3872_F	TTTGCGG GCACCTT CCGG	660	HCVUTR5_ NC001433-1-9616_ 3876_3893_R	GCTGTGTACAC CCGGCGA	667
3689	HCVUTR5_ NC001433-1-9616_ 3855_3872_F	TTTGCGGG CACCTT CCGG	660	HCVUTR5_ NC001433-1-9616_ 3942_3962_2_R	ATGCGGTATCC GGTCCTCACA	668
3691	HCVUTR5_ NC001433-1-9616_1974_1996_2_F	TGGCTCGG TTGTACAG GGATGAA	661	HCVUTR5_ NC001433-1-9616_2070_2091	TGCCCAACGGA CTACTTCCTGA	669

Example 16: Primer Pairs that Define Bioagent Identifying Amplicons for Identification of Strains of Influenza Viruses

[314] For design of primers that define bioagent identifying amplicons for identification of strains of influenza viruses, a series of influenza virus genome sequences were obtained, aligned and scanned for regions where pairs of PCR primers would amplify products of about 27 to about 200 nucleotides in length and distinguish influenza virus strains of from each other by their molecular masses or base compositions.

[315] Table 16 represents a collection of primers (sorted by primer pair number) designed to identify hepatitis C viruses using the methods described herein. The primer pair number is an in-house database index number. The forward or reverse primer name shown in Table 16 indicates the gene region of the influenza virus genome to which the primer hybridizes relative to a reference sequence. In Table 16, for example, the forward primer name FLUBPB2_NC002205_603_629_F indicates that the forward primer (_F) hybridizes to residues 603-629 of an influenza reference sequence represented by an extraction of nucleotides from GenBank Accession No. NC_002205. One with ordinary skill will know how to obtain individual gene sequences or portions thereof from genomic sequences present in GenBank.

Table 16: Primer Pairs for Identification of Strains of Influenza Viruses

Primer Pair Number	Forward Primer Name	Forward Sequence	Forward SEQ ID NO:	Reverse Primer Name	Reverse Sequence	Reverse SEQ ID NO:

1261	FLUBPB2_NC002205_603_629_F	TCCCATTGTACTGGCATACATGCTTGA	639	FLUBPB2_NC002205_667_693_R	TATGAACTCAGCTGATGTTGCTCCTGC	647
1266	FLUANUC_J02147_118_148_F	TACATCCAGATGTGCACTGAACTCAAACTCA	640	FLUANUC_J02147_188_218_R	TCGTCAAATGCAGAGAGCACCATTCTCTCTA	648
1275	FLUBNUC_NC002208_90_116_F	TCCAATCATCAGACCAGCAACCCTTGC	641	FLUBNUC_NC002208_164_189_R	TCCGATATCAGCTTCACTGCTTGTGG	649
1279	FLUAM1_NC004524_369_396_F	TCTTGCCAGTTGTATGGGCCTCATATAC	642	FLUAM1_NC004524_451_473_R	TGGGAGTCAGCAATCTGCTCACA	650
1287	FLUAPA_NC004520_562_584_F	TGGGATTCCTTTCGTCAGTCCGA	643	FLUAPA_NC004520_647_673_R	TGGAGAAGTTCGGTGGGAGACTTTGGT	651
2775	FLUANS1_NC004525_1_19_F	TCCAGGACATACTGATGAGGATGTCAAAAATGCA	644	FLUANS1_NC004525_29_52_R	TGCTTCCCCAAGCGAATCTCTGTA	652
2777	FLUANS2_NC004525_47_74_F	TGTCAAAAATGCAATTGGGGTCCTCATC	645	FLUANS2_NC004525_121_151_R	TCATTACTGCTTCTCCAAGCGAATCTCTGTA	653
2798	FLUPB1_J02151_1210_1235_F	TGTCCTGGAATGATGATGGGCATGTT	646	FLU_ALL_PB1_J02151_1313_1337_R	TCATCAGAGGATTGGAGTCCATCCC	654
1261	FLUBPB2_NC002205_603_629_F	TCCCATTGTACTGGCATACATGCTTGA	639	FLUBPB2_NC002205_667_693_R	TATGAACTCAGCTGATGTTGCTCCTGC	647

Example 17: Primer Pairs that Define Bioagent Identifying Amplicons for Identification of Strains of *Staphylococcus aureus*

[316] For design of primers that define bioagent identifying amplicons for identification of strains of *Staphylococcus aureus*, a series of *Staphylococcus aureus* virus genome sequences were obtained, aligned and scanned for regions where pairs of PCR primers would amplify products of about 27 to about 200 nucleotides in length and distinguish *Staphylococcus aureus* strains of from each other by their molecular masses or base compositions.

[317] Table 17 represents a collection of primers (sorted by primer pair number) designed to identify *Staphylococcus aureus* strains using the methods described herein. The primer pair number is an in-house database index number. The forward or reverse primer name shown in Table 17 indicates the gene region of the influenza virus genome to which the primer hybridizes relative to a reference sequence. In Table 17, for example, the forward primer name MECA_Y14051_4507_4530_F indicates that the forward primer (_F) hybridizes to residues 4507-4530 of the mecA gene of *Staphylococcus aureus* sequence represented by GenBank Accession No. Y14051. One with ordinary skill will know how to obtain individual gene sequences or portions thereof from genomic sequences present in GenBank.

Table 17: Primer Pairs for Identification of Strains of *Staphylococcus aureus*

Primer Pair Number	Forward Primer Name	Forward Sequence	Forward SEQ ID NO:	Reverse Primer Name	Reverse Sequence	Reverse SEQ ID NO:
879	MECA_Y14051_	TCAGGTACTG	717	MECA_Y14051	TGGATAGACGT	727

	4507_4530_F	CTATCCACCC TCAA		_4555_4581_R	CATATGAAG GTGTGCT	
2056	MECI-R_ NC003923-41798-41609_33_60_F	TTTACACATAT CGTGAGCAAT GAACTGA	718	MECI-R_ NC003923-41798-41609_86_113_R	TGTGATATGGAGGT TAGAAGGTGTTA	728
2081	ERMA_ NC002952-55890-56621_366_395_F	AGCTATCTTATCGT AGAAGGGATTTG	719	ERMA_ NC002952-55890-56621_438_465_R	TGAGCATTTTTA TATCCATCT CCACCAT	729
2086	ERMC_ NC005908-2004-2738_85_116_F	TCTGAACATGA TAATATCTTTGA AATCGGCTC	720	ERMC_ NC005908-2004-2738_173_206_R	TCCGTAGTTTTG CATAATTTATG GTCTATTTCAA	730
2095	PVLUK_ NC003923-1529595-1531285_688_713_F	TGAGCTGCATC AACTGTATT GGATAG	721	PVLUK_ NC003923-1529595-1531285_ 775_804_R	TGGAAAACTCA TGAAATTAAA GTGAAAGGA	731
2256	NUC_NC002758 -894288-894974_316_345_F	TACAAAGGTC AACCAATGAC ATTCAGACTA	722	NUC_NC002758-894288-894974_ 396_421_R	TAAATGCACTT GCTTCAGGG CCATAT	732
2313	MUPR_X75439_ 2486_2516_F	TAATTGGGCTC TTTCTCGCTTA AACACCTTA	723	MUPR_X75439 _2548_2574_R	TAATCTGGCTGCGC AGTGAAATCGT	733
3005	TUFB_NC002758-615038-616222_ 688_710_F	TGCCGTGTTG AACGTGGTC AAAT	724	TUFB_NC002758-615038-616222_ 783_813_R	TGCTTCAGCGT AGTCTAATAAT TTACGGAAC	734
3016	MUPR_X75439_ 2482_2510_F	TAGATAATTG GGCTCTTTCTC GCTTAAAC	725	MUPR_X75439 _2551_2573_R	AATCTGGCTGCGGA GTGAAAT	735
3106	TSST1_NC002758.2 _519_546_F	TCGTCATCAG CTAACTCAAA TACATGGA	726	TSST1_ NC002758.2 593_620_R	TCACTTTGATAT GTGGATCCGT CATTCA	736
2738	GYRA_NC002953 -7005-9668_ 166_195_F	TAAGGTATGAC ACCGGATAAA TCATATAAA	737	GYRA_ NC002953-7005-9668_265_287_R	TCTTGAGCCATA CGTACCATTGC	740
2739	GYRA_NC002953 -7005-9668_221 _249_F	TAATGGGTAAA TATCACCCTC ATGGTGAC	738	GYRA_ NC002953-7005-9668_316_343_R	TATCCATTGAAC CAAAGTTACCT TGGCC	741
2740	GYRA_NC002953 -7005-9668_ 221_249_F	TAATGGGTAAA TATCACCCTC ATGGTGAC	738	GYRA_ NC002953-7005-9668_253_283_R	TAGCCATACGTA CCATTGCTTCA TAAATAGA	742
2741	GYRA_NC002953 -7005-9668_ 234_261_F	TCACCCTCATG GTGACTCATC TATTTAT	739	GYRA_ NC002953-7005-9668_265_287_R	TCTTGAGCCATA CGTACCATTGC	740

Example 18: Comparison of Targeted Whole Genome Amplification Method with an Unbiased Whole Genome Amplification Method

[318] A set of algorithms was developed for the design of TWGA primer sets favoring amplification of target DNA from a DNA mixture as described in Example 2. As a test case, a TWGA primer set consisting of approximately 200 primers was designed for the preferential amplification of *Bacillus anthracis* genomic DNA from a mixture of background genomes. The primer set showed high representation of the *Bacillus anthracis* genome and under-representation in a panel of eukaryotic genomes selected from mammals, insects, plants, birds, and nematodes. The primer set was designed with consistent binding of the primers along the *Bacillus anthracis* genome, maintaining representation across the entire genome during amplification. To demonstrate the preferential amplification of target DNA from a DNA

mixture, mixtures of *Bacillus anthracis* Sterne DNA and human DNA were amplified using targeted whole genome amplification, and the resulting products were quantified by Quantitative Real-Time PCR-based detection of distinctive genomic sequences. As shown in Figure 5A, 175-fold amplification of B. anthracis DNA was observed in the presence of a ten million-fold excess of human background DNA, with minimal amplification of the background DNA itself. A 3000-fold amplification of target DNA was observed when background was reduced slightly, to a million-fold excess relative to the target DNA levels, again with minimal amplification of background DNA (Figure 5B).

[319] Results obtained from the targeted whole genome amplification reaction are contrasted with results of an unbiased whole genome amplification reaction in Figure 6. Target genome was prepared in a million-fold excess of background DNA and amplified by targeted whole genome amplification or by unbiased whole genome amplification. In contrast to targeted whole genome amplification, unbiased whole genome amplification uses random priming which should result in similar amplification of both target DNA and background DNA. In Figure 6A it can be seen that targeted whole genome amplification favored amplification of the target DNA. In contrast, whole genome amplification produced similar levels of amplification of both components of the DNA mixture (Figure 6B).

[320] In Figure 7, it is evident that targeted whole genome amplification increases the sensitivity of detection of target DNA from a mixture, in comparison to unbiased whole genome amplification. Reactions were prepared with human DNA present at 0.1 micrograms per reaction and with *Bacillus anthracis* genomic DNA incremented from 50 to 400 femtograms. Preferential amplification with targeted whole genome amplification primers was compared to unbiased amplification using random unbiased whole genome amplification primers. As shown above, targeted whole genome amplification gave higher yields of *Bacillus anthracis* DNA and lower yields of human DNA than unbiased whole genome amplification (Figure 7A and 7B). Significantly, targeted whole genome amplification gave detectable *Bacillus anthracis* product with 50 femtograms of starting material, whereas unbiased whole genome amplification did not.

[321] Targeted whole genome amplification primer sets were developed for six additional target organisms and a cocktail of the primer sets were run in the targeted whole genome amplification reactions. Similar results were obtained when targeted whole genome amplification was formulated with this pool of primer sets or with the *Bacillus anthracis*-specific targeted whole genome amplification primer set, indicating that targeted whole genome amplification can be multiplexed (targeted whole genome amplification seven-set primers vs. TWGA single-set primers, Figure 7).

CONCLUDING STATEMENTS

[322] The present invention includes any combination of the various species and subgeneric groupings falling within the generic disclosure. This invention therefore includes the generic description of

the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[323] While in accordance with the patent statutes, description of the various embodiments and examples have been provided, the scope of the invention is not to be limited thereto or thereby. Modifications and alterations of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the present invention.

[324] Therefore, it will be appreciated that the scope of this invention is to be defined by the appended claims, rather than by the specific examples which have been presented by way of example.

[325] Each reference (including, but not limited to, journal articles, U.S. and non-U.S. patents, patent application publications, international patent application publications, gene bank gi or accession numbers, internet web sites, and the like) cited in the present application is incorporated herein by reference in its entirety.

CLAIMS

What is claimed is:

1. A method comprising:

amplifying at least one pathogen genome from a sample suspected of comprising at least one pathogen genome and at least one background genome using a plurality of targeted whole genome amplification primers, thereby elevating the quantity of nucleic acid representing said at least one pathogen genome relative to the quantity of nucleic acid representing said at least one background genome, wherein said plurality of targeted whole genome amplification primers is selected by:

i. identifying at least one pathogen genome;

ii. identifying at least one background genome;

iii. identifying a plurality of genome sequence segments having unique sequences within said pathogen genome sequence;

iv. determining frequency of occurrence of members of said plurality of genome sequence segments within said pathogen genome sequence and determining frequency of occurrence of said plurality of genome sequence segments within said background genome sequences;

v. calculating a selectivity ratio for said members by dividing said frequency of occurrence within said pathogen genome sequence by said frequency of occurrence of said plurality of genome sequence segments within said background genome sequences;

vi. selecting a selectivity ratio threshold value, thereby defining a first sub-set of said plurality of genome sequence segments having selectivity ratios equal to or greater than said selectivity ratio threshold value;

vii. determining the lengths of pathogen genome sequence occurring between genome sequence segments of said first sub-set;

viii. selecting a second sub-set of genome sequence segments from said first sub-set wherein members of said second sub-set have a mean separation distance of less than a selected length of nucleobases; and

ix. selecting targeted whole genome amplification primers that hybridize to members of said second sub-set of genome sequence segments such that, under whole genome amplification conditions, said at least one pathogen genome is amplified selectively over said at least one background genomes.

2. The method of claim 1 further comprising the step of producing one or more amplification products representing bioagent identifying amplicons from said amplified pathogen genome using one or more primer pairs.

3. The method of claim 2 further comprising the step of measuring molecular masses of said amplification products by mass spectrometry.

4. The method of claim 3 wherein said mass spectrometry is electrospray time-of-flight mass spectrometry.

5. The method of claim 3 further comprising the step of comparing said molecular masses with a database comprising molecular masses of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample.

6. The method of claim 3 further comprising the step of calculating base compositions of said amplification products from said molecular masses.

7. The method of claim 6 further comprising the step of comparing said base compositions with a database comprising base compositions of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample.

8. The method of claim 2 wherein said amplification products are generated using a plurality of primer pairs that define bioagent identifying amplicons.

9. The method of claim 8 wherein said plurality of primer pairs are used in a multiplex reaction to generate a plurality of amplification products.

10. The method of claim 8 wherein said plurality of primer pairs comprises at least two primer pairs from the group consisting of primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 354 (SEQ ID NOs: 597:605), 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 2249 (SEQ ID NOs: 601:609), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

11. The method of claim 8 wherein said plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 3346 (SEQ ID NOs: 616:631).

12. The method of claim 8 wherein said plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), and 3361 (SEQ ID NOs: 620:635).

13. The method of claim 8 wherein said plurality of primer pairs comprises primer pair numbers 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604) and at least one of the primer pairs selected from the group consisting of 354 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

14. The method of claim 1 wherein a high processivity polymerase enzyme is used at said amplification step.

15. The method of claim 14 wherein said high processivity polymerase enzyme is a recombinant polymerase enzyme.

16. The method of claim 14 wherein said high processivity polymerase enzyme is a genetically engineered polymerase enzyme.

17. The method of claim 14 wherein said high processivity polymerase enzyme is phi29.

18. The method of claim 1, wherein said sample comprises human whole blood.

19. The method of claim 18 further comprising the step of extracting total nucleic acid from said sample before carrying out said amplifying step.

20. The method of claim 1 wherein said sample comprises human buffy coat.

21. The method of claim 20 further comprising the step of extracting total nucleic acid from said sample before carrying out said amplifying step.

22. The method of claim 1 wherein said sample comprises human serum.

23. The method of claim 22 further comprising the step of extracting total nucleic acid from said sample before carrying out said amplifying step.

24. The method of claim 1 wherein said sample comprises human hepatic cells.

25. The method of claim 24 further comprising the step of extracting total nucleic acid from sample before carrying out said amplifying step.

26. The method of claim 1 wherein said sample comprises sputum.

27. The method of claim 26 further comprising the step of extracting total nucleic acid from sample before carrying out said amplifying step.

28. The method of claim 1 wherein said sample comprises urine.

29. The method of claim 28 further comprising the step of extracting total nucleic acid from sample before carrying out said amplifying step.

30. The method of claim 1 wherein said sample comprises biopsy tissue.

31. The method of claim 30 further comprising the step of extracting total nucleic acid from sample before carrying out said amplifying step

32. The method of claim 1 wherein said at least one pathogen is a bacterium.

33. The method of claim 32 wherein said bacterium is selected from the group consisting of: *Escherichia coli, Klebsiella pneumoniae, Klebsiella oxytoca, Serratia marcescens, Enterobacter cloacae, Enterobacter aerogenes, Proteus mirabilis, Pseudomonas aeruginosa, Acinetobacter baumannii, Stenotrophomonas maltophilia, Staphylococcus aureus, Staphylococcus epidermidis, Staphylococcus haemolyticus, Streptococcus pneumoniae, Streptococcus pyogenes, Streptococcus agalactiae, Streptococcus mitis, Enterococcus faecium, Enterococcus faecalis, Candida albicans, Candida tropicalis, Candida parapsilosis, Candida krusei, Candida glabrata, Mycobacterium tuberculosis,* and *Aspergillus fumigatus.*

34. The method of claim 1 wherein said at least one background genome comprises a human nucleic acid.

35. The method of claims 5 or 7 wherein said identifying step indicates the presence of bacterial sepsis in a human patient.

36. The method of claims 5 or 7 wherein said identifying step indicates the presence of bacteremia in a human patient.

37. The method of claim 1 wherein said pathogen is a virus.

38. The method of claim 37 wherein said virus is HIV.

39. The method of claim 37 wherein said virus is HCV.

40. The method of claim 37 wherein said virus is influenza virus.

41. A diagnostic kit for performing any of the methods of claims 1, 5, 7, 32, 33 and 37-41.

42. A diagnostic kit comprising a high processivity polymerase enzyme and a plurality of purified targeted whole genome amplification primers.

43. The kit of claim 42 further comprising at least one primer pair that defines a bioagent identifying amplicon.

44. The kit of claim 43 wherein said plurality of primer pairs comprises at least two primer pairs from the group consisting of primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 354 (SEQ ID NOs: 597:605), 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 2249 (SEQ ID NOs: 601:609), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

45. The kit of claim 43 wherein said plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 3346 (SEQ ID NOs: 616:631).

46. The kit of claim 43 wherein said plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), and 3361 (SEQ ID NOs: 620:635).

47. The kit of claim 43 wherein said plurality of primer pairs comprises primer pair numbers 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604) and at least one of the primer pairs selected from the group consisting of 354 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

48. The kit of claim 43 wherein said high processivity enzyme is phi29.

49. A method comprising the steps of:

a. extracting nucleic acids from a sample; and

b. mixing said nucleic acids with a plurality of targeted whole genome amplification primers, a high processivity polymerase enzyme to produce an amplification mixture, wherein said plurality of targeted whole genome amplification primers is selected by:

i. identifying at least one target genome suspected of being present in said sample;

ii. identifying at least one background genome suspected of being present in said sample;

iii. identifying a plurality of genome sequence segments having unique sequences within said target genome sequence;

iv. determining frequency of occurrence of members of said plurality of genome sequence segments within said target genome sequence and within said background genome sequences;

v. calculating a selectivity ratio for said members by dividing said frequency of occurrence within said target genome by said frequency of occurrence of said plurality of genome sequence segments within said background genome sequences;

vi. selecting a selectivity ratio threshold value, thereby defining a first sub-set of said plurality of genome sequence segments having selectivity ratios equal to or greater than said selectivity ratio threshold value;

vii. determining the lengths of target genome sequence occurring between genome sequence segments of said first sub-set;

viii. selecting a second sub-set of genome sequence segments from said first sub-set wherein members of said second sub-set have a mean separation of less than a selected length of nucleobases; and

ix. selecting targeted whole genome amplification primers that hybridize to members of said second sub-set of genome sequence segments such that said at least one target genome is amplified selectively over said at least one background genome.

50. The method of claim 49 further comprising the step of amplifying one or more of said extracted nucleic acids in said mixture of step b.

51. The method of claim 49 wherein said amplifying step is a targeted whole genome amplification reaction.

52. The method of claim 51 further comprising the step of performing a second amplification step using at least one primer pair that defines a bioagent identifying amplicon to obtain at least a second amplification product.

53. The method of claim 52 further comprising the step of measuring the molecular mass of said second amplification product by mass spectrometry.

54. The method of claim 53 wherein said mass spectrometry is electrospray time-of-flight mass spectrometry.

55. The method of claim 52 further comprising the step of comparing said molecular mass with a database comprising molecular masses of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample.

56. The method of claim 53 further comprising the step of calculating a base composition of said amplification products from said molecular mass.

57. The method of claim 56 further comprising the step of comparing said base compositions with a database comprising base compositions of bioagent identifying amplicons of pathogens produced with said primer pairs, thereby identifying said pathogen in said sample.

58. The method of claim 52 wherein said second amplification step comprises obtaining a plurality of amplification products generated using a plurality of primer pairs that define bioagent identifying amplicons.

59. The method of claim 58 wherein said plurality of primer pairs is used in one or more multiplex reactions to generate a plurality of amplification products.

60. The method of claim 58 or 59 wherein said plurality of primer pairs comprises at least two primer pairs from the group consisting of primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 354 (SEQ ID NOs: 597:605), 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 2249 (SEQ ID NOs: 601:609), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

61. The method of claim 58 or 59 wherein said plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), 3346 (SEQ ID NOs: 616:631).

62. The method of claim 58 or 59 wherein said plurality of primer pairs comprises primer pair numbers: 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604), and 3361 (SEQ ID NOs: 620:635).

63. The method of claim 58 or 59 wherein said plurality of primer pairs comprises primer pair numbers 346 (SEQ ID NOs: 594:602), 348 (SEQ ID NOs: 595:603), 349 (SEQ ID NOs: 596:604) and at least one of the primer pairs selected from the group consisting of 354 358 (SEQ ID NOs: 598:606), 359 (SEQ ID NOs: 599:607), 3346 (SEQ ID NOs: 616:631), 449 (SEQ ID NOs: 600:608), 3350 (SEQ ID NOs: 614:629), 3361 (SEQ ID NOs: 620:635), and 3360 (SEQ ID NOs: 612:627).

64. The method of claim 49 wherein said high processivity polymerase enzyme is a recombinant polymerase enzyme.

65. The method of claim 49 wherein said high processivity polymerase enzyme is a genetically engineered polymerase enzyme.

66. The method of claim 49 wherein said high processivity polymerase enzyme is phi29.

67. The method of claim 49, wherein said sample comprises human whole blood.

68. The method of claim 67 further comprising the step of extracting total nucleic acid from said sample before carrying out said amplifying step.

69. The method of claim 49 wherein said sample comprises human buffy coat.

70. The method of claim 69 further comprising the step of extracting total nucleic acid from said sample before carrying out said amplifying step.

71. The method of claim 49 wherein said sample comprises human serum.

72. The method of claim 71 further comprising the step of extracting total nucleic acid from said sample before carrying out said amplifying step.

73. The method of claim 49 wherein said sample comprises human hepatic cells.

74. The method of claim 73 further comprising the step of total nucleic acid from sample before carrying out said amplifying step.

75. The method of claim 49 wherein said sample comprises sputum.

76. The method of claim 75 further comprising the step of extracting total nucleic acid from sample before carrying out said amplifying step.

77. The method of claim 49 wherein said sample comprises urine.

78. The method of claim 77 further comprising the step of extracting total nucleic acid from sample before carrying out said amplifying step.

79. The method of claim 49 wherein said sample comprises biopsy tissue.

80. The method of claim 79 further comprising the step of extracting total nucleic acid from sample before carrying out said amplifying step

81. The method of claim 49 wherein said sample comprises a bacterium.

82. The method of claim 81 wherein said bacterium is selected from the group consisting of: *Escherichia coli, Klebsiella pneumoniae, Klebsiella oxytoca, Serratia marcescens, Enterobacter cloacae, Enterobacter aerogenes, Proteus mirabilis, Pseudomonas aeruginosa, Acinetobacter baumannii, Stenotrophomonas maltophilia, Staphylococcus aureus, Staphylococcus epidermidis, Staphylococcus haemolyticus, Streptococcus pneumoniae, Streptococcus pyogenes, Streptococcus agalactiae, Streptococcus mitis, Enterococcus faecium, Enterococcus faecalis, Candida albicans, Candida tropicalis, Candida parapsilosis, Candida krusei, Candida glabrata* and *Aspergillus fumigatus.*

83. The method of claim 49 wherein said at least one background genome comprises a human nucleic acid.

84. The method of claim 56 or 57 wherein said identifying step indicates the presence of bacterial sepsis in a human.

85. The method of claim 56 or 57 wherein said pathogen is a virus.

86. The method of claim 85 wherein said virus is HIV.

87. The method of claim 85 wherein said virus is HCV.

88. The method of claim 85 wherein said virus is influenza virus.

89. A diagnostic kit for performing any of the methods of claims 49, 81, 82, and 85-88, said kit comprising a plurality of targeted whole genome amplification primers.

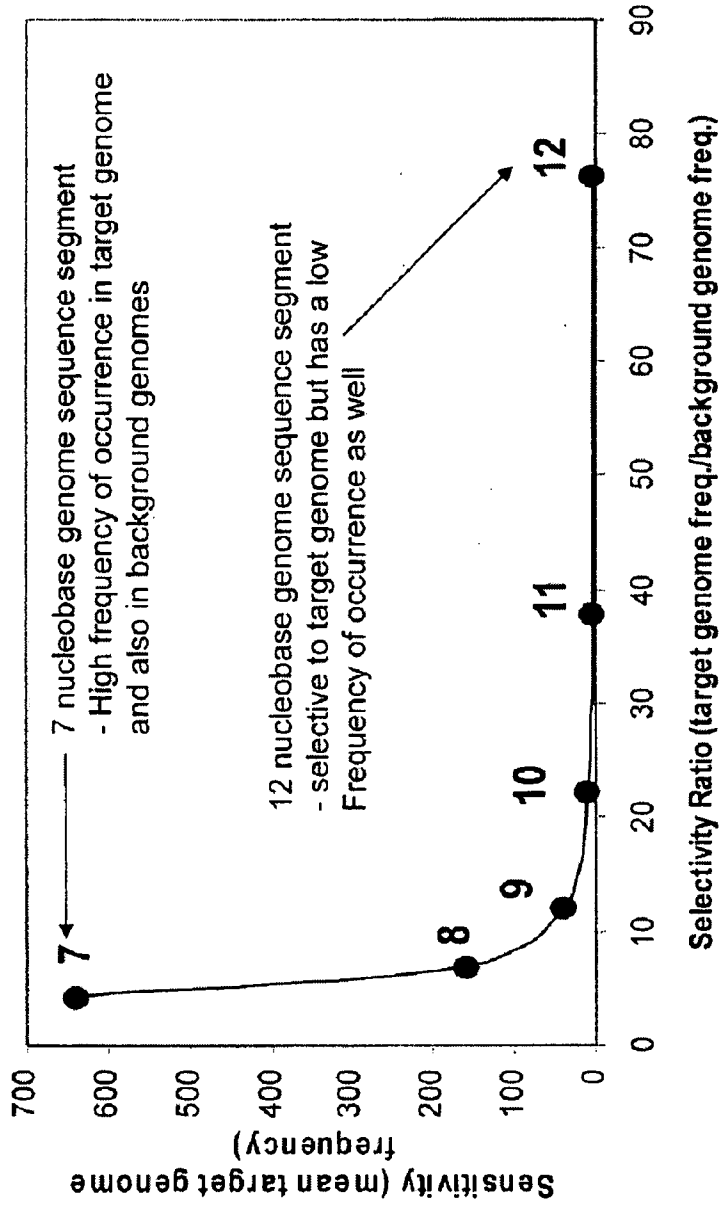
7 nucleobase genome sequence segment
- High frequency of occurrence in target genome
and also in background genomes
12 nucleobase genome sequence segment
- selective to target genome but has a low
Frequency of occurrence as well
7
8
9
10
11
12
Sensitivity (mean target genome frequency)
Selectivity Ratio (target genome freq./background genome freq.)
0
100
200
300
400
500
600
700
0
10
20
30
40
50
60
70
80
90

Figure 1

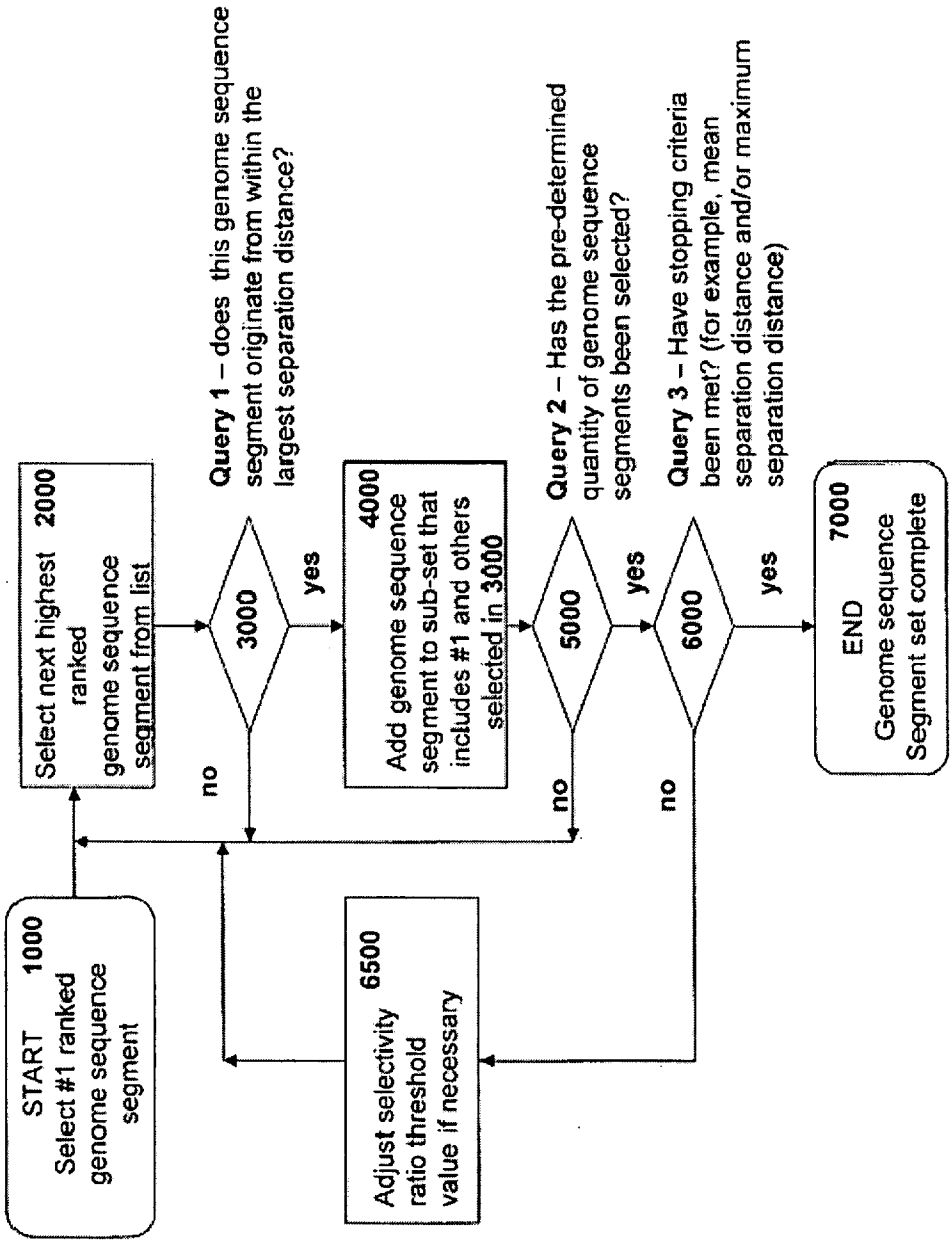
START 1000
Select #1 ranked genome sequence segment
Select next highest ranked genome sequence segment from list 2000
Query 1 – does this genome sequence segment originate from within the largest separation distance?
3000
yes
no
4000
Add genome sequence segment to sub-set that includes #1 and others selected in 3000
Query 2 – Has the pre-determined quantity of genome sequence segments been selected?
5000
yes
no
Query 3 – Have stopping criteria been met? (for example, mean separation distance and/or maximum separation distance)
6000
yes
no
END 7000
Genome sequence Segment set complete
Adjust selectivity ratio threshold value if necessary 6500

Figure 2

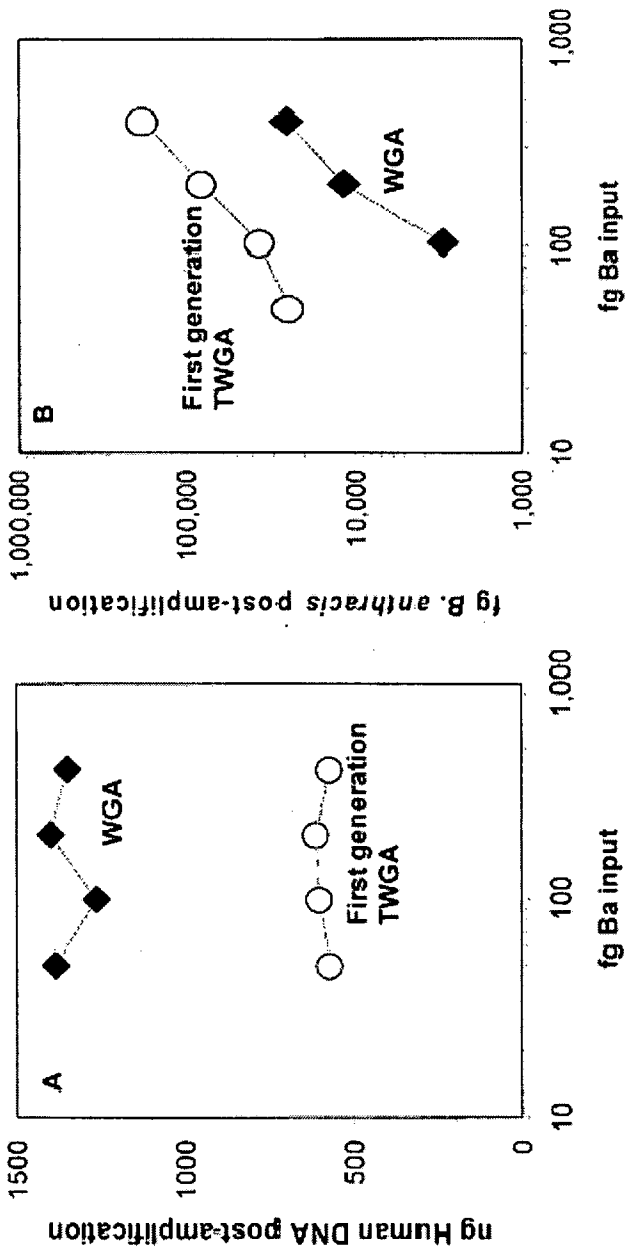
B
First generation TWGA
WGA
1,000,000
100,000
10,000
1,000
10
100
1,000
fg Ba input
fg B. anthracis post-amplification
A
WGA
First generation TWGA
1500
1000
500
0
10
100
1,000
fg Ba input
ng Human DNA post-amplification

Figure 3

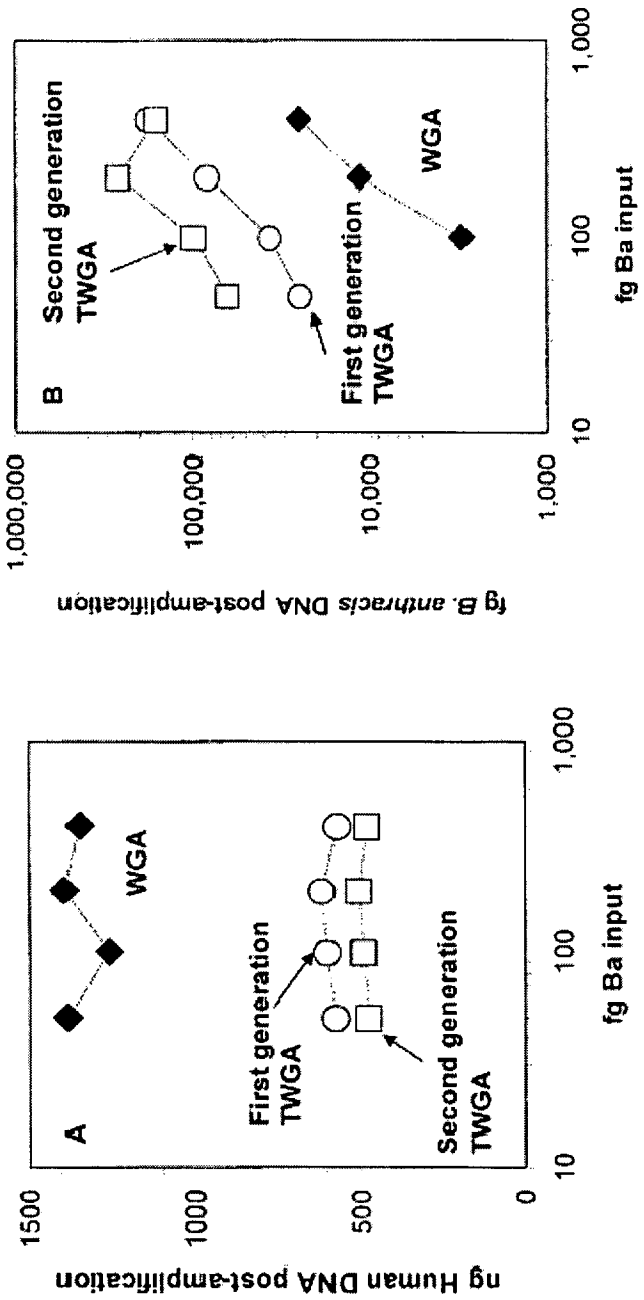
B
Second generation TWGA
First generation TWGA
WGA
fg B. anthracis DNA post-amplification
1,000,000
100,000
10,000
1,000
10
100
1,000
fg Ba input
A
WGA
First generation TWGA
Second generation TWGA
ng Human DNA post-amplification
1500
1000
500
0
10
100
1,000
fg Ba input

Figure 4

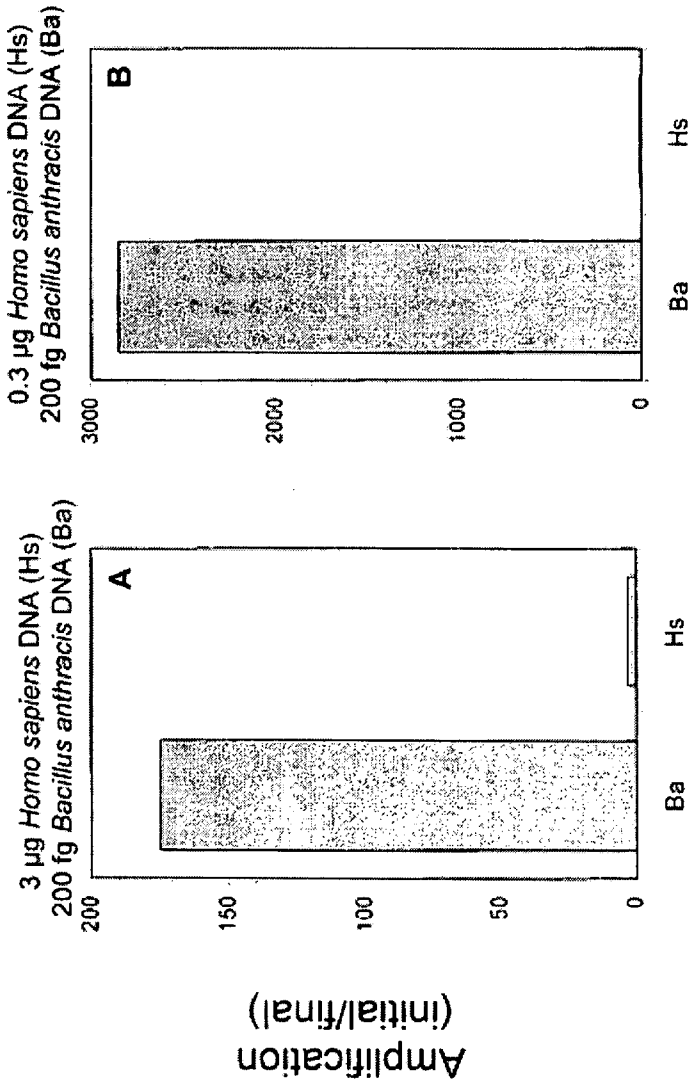
0.3 µg Homo sapiens DNA (Hs)
200 fg Bacillus anthracis DNA (Ba)
B
3000
2000
1000
0
Ba
Hs
3 µg Homo sapiens DNA (Hs)
200 fg Bacillus anthracis DNA (Ba)
A
200
150
100
50
0
Ba
Hs
Amplification
(initial/final)

Figure 5

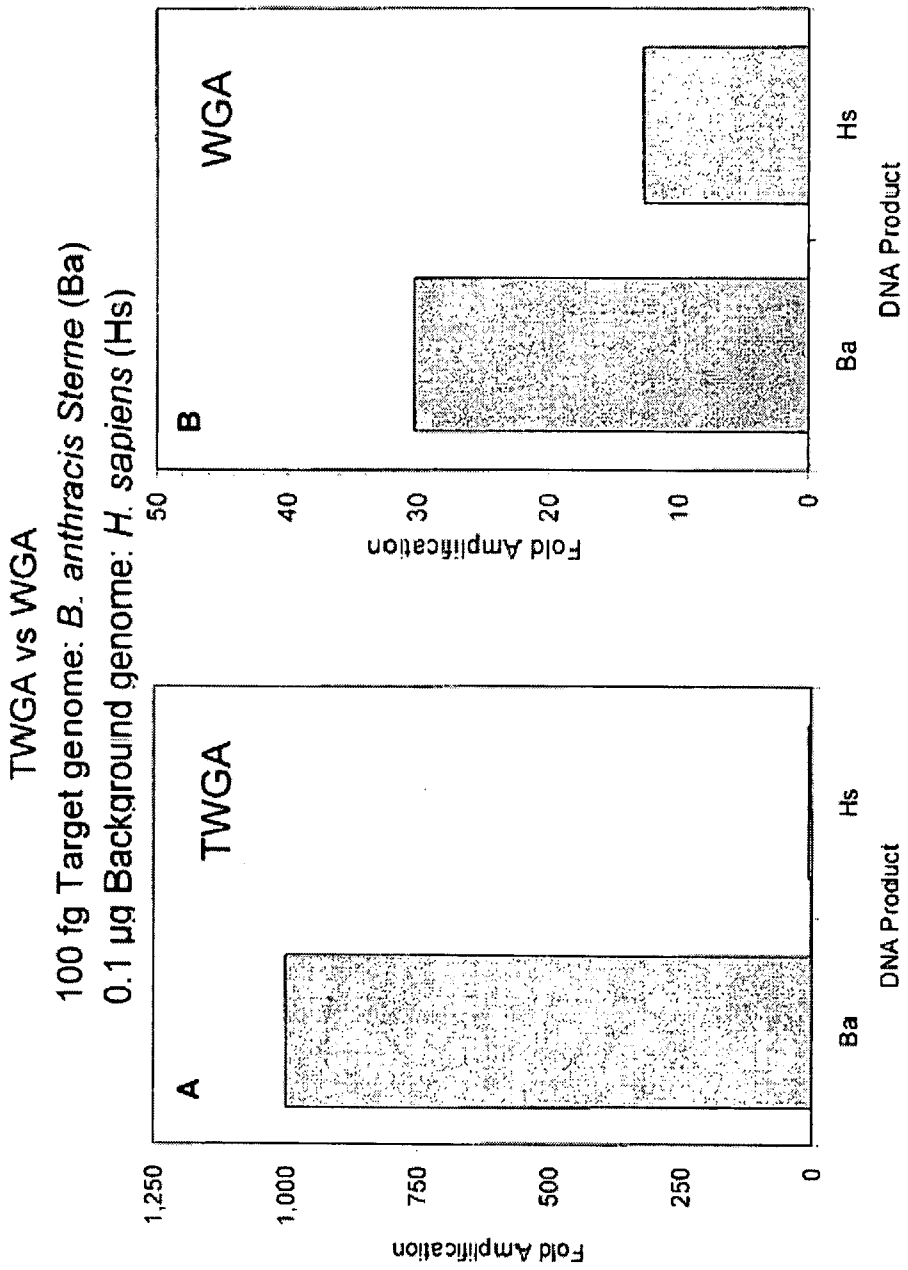
TWGA vs WGA
100 fg Target genome: B. anthracis Sterne (Ba)
0.1 µg Background genome: H. sapiens (Hs)
A
TWGA
Fold Amplification
1,250
1,000
750
500
250
0
Ba
Hs
DNA Product
B
WGA
Fold Amplification
50
40
30
20
10
0
Ba
Hs
DNA Product

Figure 6

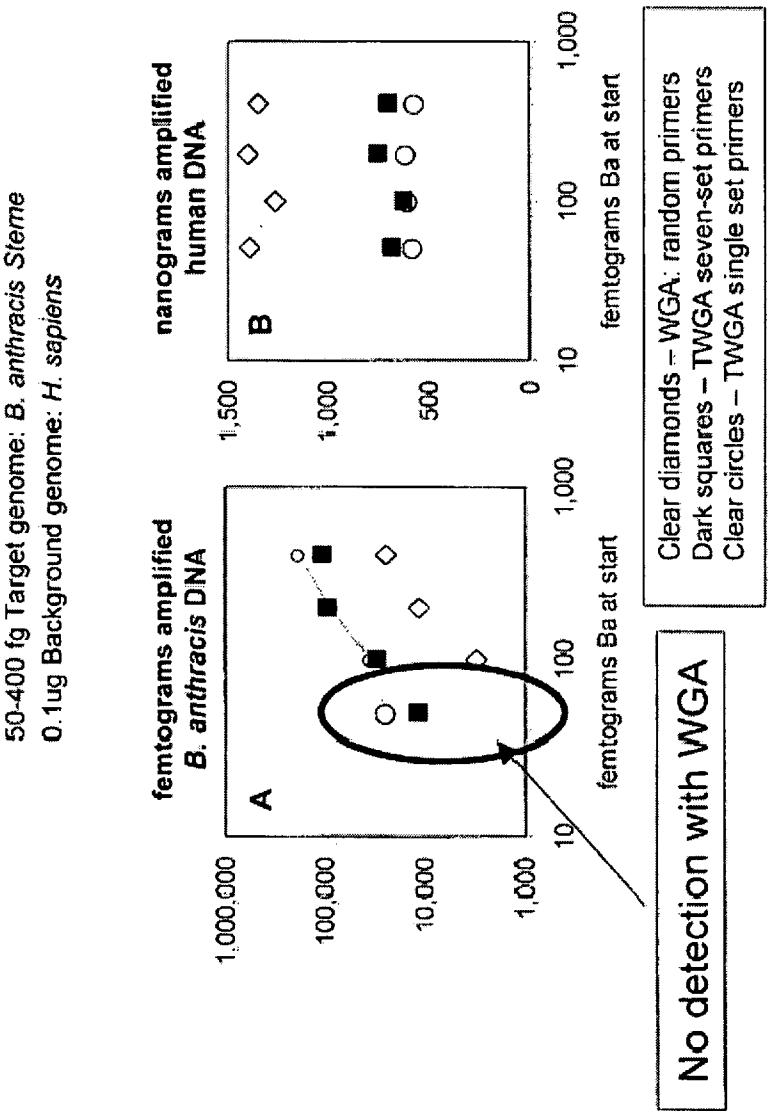
50-400 fg Target genome: B. anthracis Sterne
0.1ug Background genome: H. sapiens
femtograms amplified B. anthracis DNA
A
1,000,000
100,000
10,000
1,000
10
100
1,000
femtograms Ba at start
nanograms amplified human DNA
B
1,500
1,000
500
0
10
100
1,000
femtograms Ba at start
Clear diamonds – WGA: random primers
Dark squares – TWGA seven-set primers
Clear circles – TWGA single set primers
No detection with WGA

Figure 7

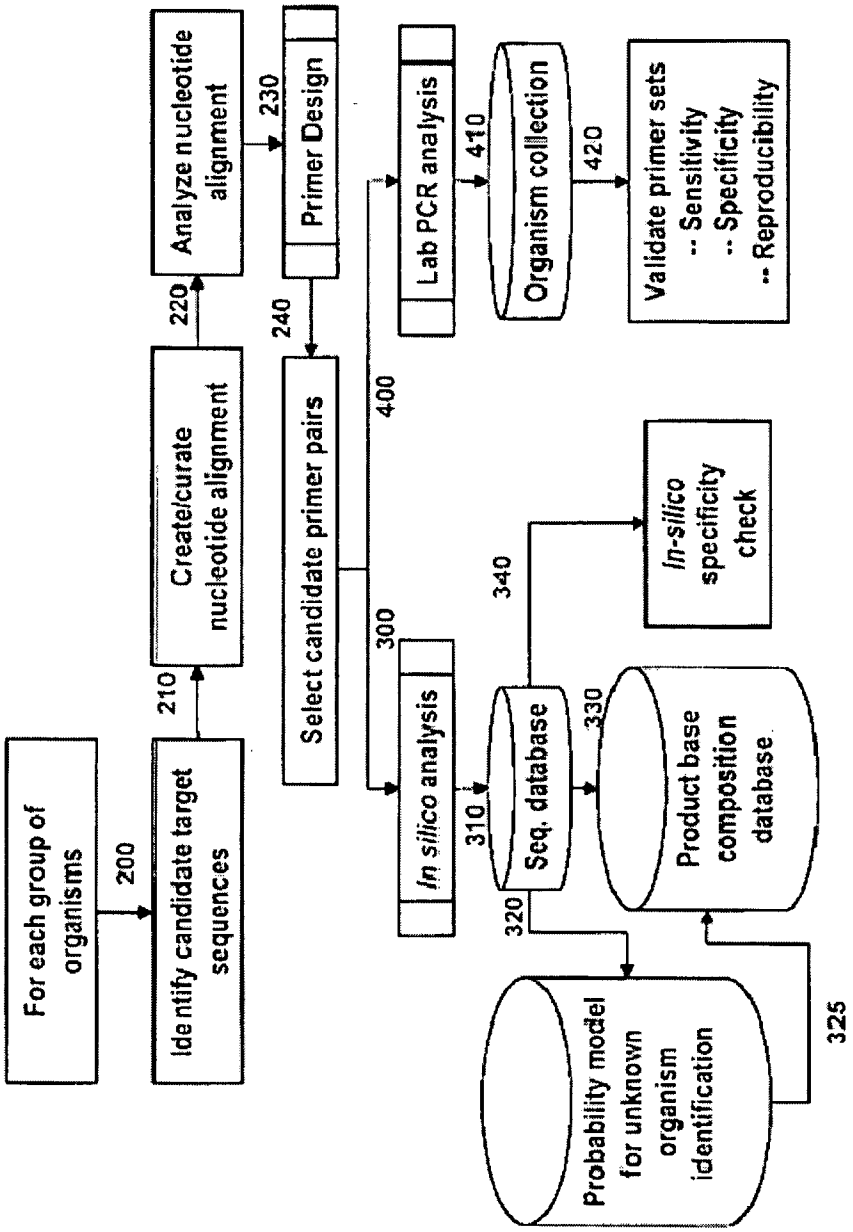
For each group of organisms
200
Identify candidate target sequences
210
Create/curate nucleotide alignment
220
Analyze nucleotide alignment
230
Primer Design
240
Select candidate primer pairs
300
In silico analysis
310
Seq. database
320
Probability model for unknown organism identification
325
Product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

Figure 8

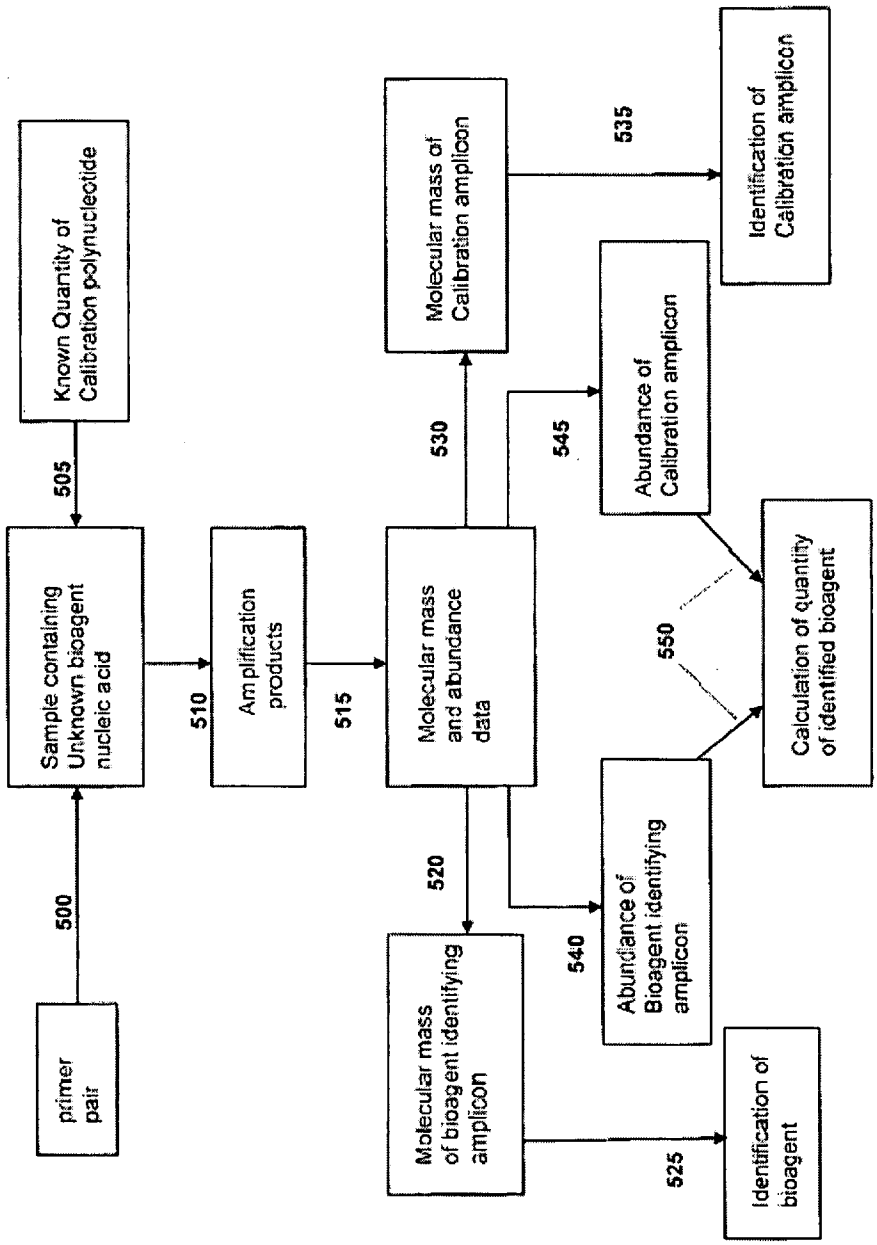
primer pair
500
Sample containing Unknown bioagent nucleic acid
505
Known Quantity of Calibration polynucleotide
510
Amplification products
515
Molecular mass and abundance data
520
Molecular mass of bioagent identifying amplicon
525
Identification of bioagent
530
Molecular mass of Calibration amplicon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

Figure 9