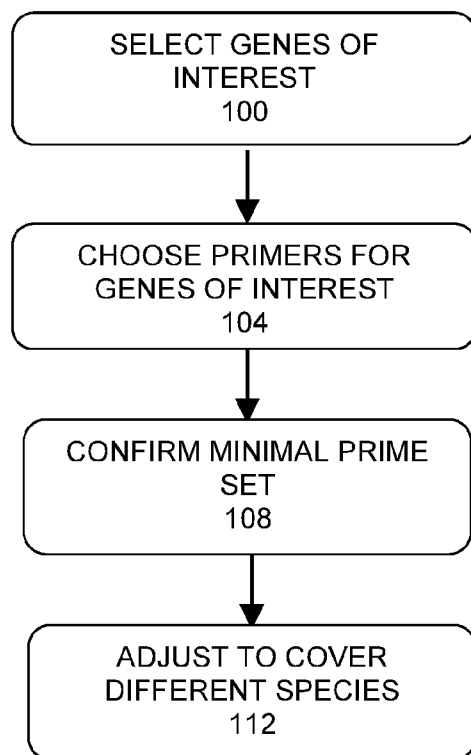




- (51) International Patent Classification:
C12Q 1/68 (2006.01) *G06F 19/22* (2011.01)
G06F 19/18 (2011.01)
- (21) International Application Number: PCT/IB2016/052852
- (22) International Filing Date: 17 May 2016 (17.05.2016)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 62/167,996 29 May 2015 (29.05.2015) US
- (71) Applicant: **KONINKLIJKE PHILIPS N.V.** [NL/NL];
High Tech Campus 5, 5656 AE Eindhoven (NL).
- (72) Inventors: **CHOTHANI, Sonia**; c/o High Tech Campus,
Building 5, 5656 AE Eindhoven (NL).
KAMALAKARAN, Sitharthan; c/o High Tech Campus,
Building 5, 5656 AE Eindhoven (NL).
- (74) Agents: **BELOBORODOV, Mark** et al.; High Tech Cam-
pus, Building 5, 5656 AE Eindhoven (NL).
- (81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, 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, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,

[Continued on next page]

(54) Title: DEGENERATE PRIMER SETS



(57) Abstract: A set of broad ranging primers to allow rapid amplification of any pathogenic bacteria as well as a methodology for the design of these primers. One approach to providing such a set in accord with the present invention is to design primers for regions/genes that are conserved enough for universal primer binding and at the same time have enough variability to give distinguishing capacity. The set may be selected based on optimizing for degeneracy and coverage.

FIG 1





SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

DEGENERATE PRIMER SETS

FIELD

[0001] The present invention generally relates to primer sets for amplifying pathogenic bacteria, and more specifically to a degenerate primer set and the design of the same.

BACKGROUND

[0002] There is an urgent clinical need to rapidly and accurately identify infection-causing pathogens to avoid unwanted antibiotic usage and hence the spread of antibiotic resistance. Currently, a symptoms based elimination strategy is used to identify the infection, which can take a considerable amount of time.

[0003] Next generation sequencing technologies permit infectious disease diagnostics to be performed in a matter of hours rather than days or weeks, provided that the proper primers are utilized in the sequencing process. There are, however, a large number of pathogenic species, and some primers are better suited than others for the amplification of particular species.

[0004] Accordingly, there is a need for an improved primer set that is able to amplify some part of any pathogenic bacterial genome and hence identify the bacteria causing infection in an accurate, timely and cost-effective manner.

SUMMARY

[0005] This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description section. This summary is not intended to identify or exclude key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0006] Embodiments of the present invention provide a set of broad ranging primers to allow rapid amplification of any pathogenic bacteria as well as a methodology for the design of these primers. One approach to find such a set in accord with the present invention is to design primers for regions/genes that are conserved enough for universal primer binding and at the same

time have enough variability to give distinguishing capacity. The set may be selected based on optimizing for degeneracy and coverage.

[0007] In one aspect, embodiments of the present invention relate to a method of amplifying nucleic acid in a sample. The method includes selecting at least one primer from Table 1 for the amplification of nucleic acid in a sample and producing an amplicon using the selected at least one primer.

[0008] In one embodiment, the method further includes sequencing the amplicon and obtaining an amplicon sequence; comparing the amplicon sequence to a list of nucleic acids; and identifying the presence of at least one pathogen based on the comparison, wherein the list of nucleic acids corresponds to nucleic acid sequences found in at least one pathogen.

[0009] In one embodiment, the identification of the at least one pathogen is performed in fewer than 10 hours. In one embodiment, the pathogen is bacterial. In one embodiment, the method further includes determining treatment based on the presence of the at least one identified pathogen; the sample may be obtained from a human and treatment may comprise administering an antibiotic to decrease the presence of the at least one identified pathogen.

[0010] In one embodiment, the at least one primer from Table 1 comprises a primer with the sequence ACTCCTACGGGAGGCAGC. In one embodiment, the at least one primer from Table 1 comprises a primer set comprising a primer with the sequence ACTCCTACGGGAGGCAGC; a primer with the sequence CCAGCAGYYGCGGTAATA; and a primer with the sequence TCCTAAGGTAGCGAAATTCCT.

[0011] In one embodiment, the at least one primer is optimized for the highest coverage of pathogens with the minimum number of primers. In one embodiment, the at least one primer is optimized for melting temperature and guanine-cytosine content. In one embodiment, the at least one primer is designed using phylogenetic conservation analysis. In one embodiment, the at least one primer is optimized based on at least one of degeneracy to cover a plurality of species of pathogens; and variability to distinguish among said plurality of species of pathogens.

[0012] In another aspect, embodiments of the present invention relate to a system for amplifying nucleic acid in a sample, the system comprising a primer set comprising at least one primer from Table 1.

[0013] In yet another aspect, embodiments of the present invention relate to a system for diagnosing the presence of at least one pathogen. The system includes a primer set and a database of nucleic acids. The primer set includes at least one primer from Table 1, the primer set optimized to produce an amplicon of nucleic acid in a sample. The nucleic acids in the database correspond to the at least one pathogen, wherein the presence of the at least one pathogen in a sample is identifiable based on a comparison of the sequence of the amplicon to the database of nucleic acids corresponding to the at least one pathogen.

[0014] In one embodiment, the primer set comprises a primer with the sequence ACTCCTACGGGAGGCAGC, a primer with the sequence CCAGCAGYYGCGGTAATA, and a primer with the sequence TCCTAAGGTAGCGAAATTCCT. In one embodiment, the at least one primer is further optimized based on at least one of: degeneracy to cover a plurality of species of pathogens; and variability to distinguish among said plurality of species of pathogens. In one embodiment, the system further includes a bioinformatics engine configured to sequence and compare the amplicon with the database of nucleic acids corresponding to the at least one pathogen; the bioinformatics engine may be further configured to produce a report including at least one of species ID, substrain ID, antibiotic for treatment, resistance value, and actionable treatment.

[0015] In still yet another aspect, embodiments of the present invention relate to a method of obtaining a primer set. The method includes identifying a plurality of pathogens of interest; identifying genes of said plurality of pathogens of interest; generating at least one primer for at least one of said identified genes; and repeating the foregoing steps to generate at least one additional primer for at least one other of said identified genes of said plurality of pathogens of interest until a primer is generated that corresponds to at least a majority of said plurality of pathogens of interest. The at least one primer is generated based on an increased level of degeneracy among the identified genes of said plurality of pathogens of interest and variability among said identified genes of said plurality of pathogens of interest wherein said variability is sufficient to enable differentiation among said plurality of pathogens of interest.

[0016] These and other features and advantages, which characterize the present non-limiting embodiments, will be apparent from a reading of the following detailed description and a review of the associated drawings. It is to be understood that both the foregoing general description and

the following detailed description are explanatory only and are not restrictive of the non-limiting embodiments as claimed.

BRIEF DESCRIPTION OF DRAWINGS

[0017] Non-limiting and non-exhaustive embodiments are described with reference to the following figures in which:

[0018] FIG. 1 depicts an example of one embodiment of a method for primer design in accord with the present invention; and

[0019] FIG. 2 illustrates a block diagram of an exemplary system for primer design according to the present invention.

[0020] In the drawings, like reference characters generally refer to corresponding parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed on the principles and concepts of operation.

DETAILED DESCRIPTION

[0021] Various embodiments are described more fully below with reference to the accompanying drawings, which form a part hereof, and which show specific exemplary embodiments. However, embodiments may be implemented in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the embodiments to those skilled in the art. Embodiments may be practiced as methods, systems or devices. Accordingly, embodiments may take the form of a hardware implementation, an entirely software implementation or an implementation combining software and hardware aspects. The following detailed description is, therefore, not to be taken in a limiting sense.

[0022] Reference in the specification to “one embodiment” or to “an embodiment” means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least one embodiment of the invention. The appearances of the

phrase “in one embodiment” in various places in the specification are not necessarily all referring to the same embodiment.

[0023] Some portions of the description that follow are presented in terms of symbolic representations of operations on non-transient signals stored within a computer memory. These descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. Such operations typically require physical manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical, magnetic or optical signals capable of being stored, transferred, combined, compared and otherwise manipulated. It is convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers, or the like. Furthermore, it is also convenient at times, to refer to certain arrangements of steps requiring physical manipulations of physical quantities as modules or code devices, without loss of generality.

[0024] However, all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated otherwise as apparent from the following discussion, it is appreciated that throughout the description, discussions utilizing terms such as “processing” or “computing” or “calculating” or “determining” or “displaying” or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0025] Certain aspects of the present invention include process steps and instructions that could be embodied in software, firmware or hardware, and when embodied in software, could be downloaded to reside on and be operated from different platforms used by a variety of operating systems.

[0026] The present invention also relates to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage

medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, solid state memory, application specific integrated circuits (ASICs), or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus or enterprise service bus. Furthermore, the computers referred to in the specification may include a single processor or may be architectures employing multiple processor designs for increased computing capability in a distributed manner.

[0027] The processes and displays presented herein are not inherently related to any particular computer or other apparatus. Various general purpose systems may also be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these systems will appear from the description below. In addition, the present invention is not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the present invention as described herein, and any references below to specific languages are provided for disclosure of enablement and best mode of the present invention.

[0028] In addition, the language used in the specification has been principally selected for readability and instructional purposes, and may not have been selected to delineate or circumscribe the inventive subject matter. Accordingly, the disclosure of the present invention is intended to be illustrative, but not limiting, of the scope of the invention, which is set forth in the claims.

Overview

[0029] Embodiments of the present invention concern a “universal” primer set to further reduce the turn-around time required for infectious disease diagnostics and thereby provide timely and accurate treatment to an infected patient. The primer set encompasses most (i.e., > 90%) of these pathogens. Once the pathogen of interest is amplified using the primer set, it can be sequenced and mapped to a reference sequence to quickly identify the species in the infection.

[0030] In ordinary application these primers are appended with adaptors to comply with Illumina index for sequencing. Using multiplex PCR and NextGen sequencing the amplicon

sequence is obtained. A local BLAST may then be run on the amplicon sequence to identify the pathogen associated with the infection.

[0031] A primer is a short nucleic acid sequence used as a starting point for DNA synthesis. In PCR, primers are used to determine the DNA fragment to be amplified by the PCR process. A primer may be designed to be specific a particular organism's gene sequence or it may be degenerate, i.e., designed for a gene sequence that is independent of a particular organism.

[0032] With reference to Figure 1, embodiments of the present invention focus on regions of genes (Step 100) that are conserved enough across pathogens to permit the selection of a common set of primers to enable the sequencing of the maximum number of pathogens, while having enough variability to differentiate species based on amplicons.

[0033] In some embodiments, ribosomal genes are the target gene regions for the primer design process. Particular ribosomal genes of interest are the 16S and 23S genes, but there are also some other genes like rpsB, rpsM, rpIP, etc., that have such properties. In other embodiments, some protein coding marker genes are considered conserved enough to determine phylogeny and are used as the target gene regions; such coding marker genes may be identified using a tool for phylogenetic analysis of genomes and metagenomes such as PHYLOSIFT, available from <https://phylosift.wordpress.com/>.

[0034] While conserved behavior is important when selecting a target gene region, it is important to balance the variability in the gene region against the conserved behavior in choosing the final primer set. For example, it may be theoretically possible to have one primer that covers all pathogenic species, but the variability among amplicons may not be enough to differentiate among species.

[0035] Once a particular target gene region has been selected, commercially available software tools such as HYDEN (available from <http://acgt.cs.tau.ac.il/hyden/>) and PRIMUX (available from <http://sourceforge.net/projects/primux/>) can be applied to the bacterial genome multiple sequence alignment data to output a set of primers (Step 104) that cover the maximum number of bacteria while satisfying specified criteria such as degeneracy, expected amplicon size, melting temperature, etc.

[0036] The suggested primer set (Step 104) can then be evaluated to ensure that it is the minimum primer set (Step 108). In one embodiment, a table consisting of the primers and their percentage degeneracy with respect to each gene is generated and sorted by degeneracy percentage.

5 [0037] The minimum primer set (Step 108) may be adjusted to cover additional species (Step 112). In one embodiment, this may consist of taking the additional species and repeating Steps 100–108 for the additional species to obtain one or more additional primers to add to the primer set.

10 [0038] The resulting primer set may be reviewed to identify the minimum number of primers that cover the maximum number of species, and additional primers may be added to the minimal set until the point of diminishing returns is reached. The result is a primer set combination that amplifies at least some region of all known pathogenic species, allowing amplification of a region given any infectious sample which can be further sequenced to identify the species in the infection. The maximum number of primers allowed in the PCR cycle may serve as a natural
15 check on the number of primers that can be added to the set. Table 1 presents a list of several primers that have been identified for use in various embodiments of the Universal Primer Set (UPS).

[0039] One example of a UPS in accord with the present invention is developed by selecting 16S, 23S, rpIB, and rpIP as the genes of interest (Step 100). The primers for the genes of interest
20 are, with reference to Table 1, Primer 1, Primer 4, Primer 6 and Primer 7 (Step 104). This combination covers 172 out of 181 pathogenic species, leaving only nine species uncovered.

[0040] These four primers can be used as a UPS, amplifying 95% of pathogenic species. Overhang adapter sequences may be added to the primers (e.g., CAGTA on the 5' end and making sure there is no primer dimer formation) to promote compatibility with commercially
25 available sequencing technologies (e.g., Illumina sequencing).

[0041] In another embodiment, the genes of interest may be selected using a clustering method such as multiple sequence alignment (MSA) to identify regions of conservation in one or more subsets of the species of interest (Step 100'). In some embodiments the genes of interest have regions that have a short conserved section, followed by a high information content section.

[0042] The least degenerate sequence may be located with a sliding window approach. Primers may then be chosen for each subset separately (Step 104). The minimal primer set may be confirmed by, e.g., generating a table listing the primers and their percentage degeneracy with respect to each gene sorting by degeneracy percentage.

- 5 [0043] One the universal primer set (UPS) is designed and a multiplex PCR experiment is carried out, the result is an amplicon to be sequenced. Once the amplicon is sequenced, it is mapped back to its source genome using, e.g., a local BLAST search on the pathogen genome database.

Sr.No	Gene	Forward Primer	Reverse Primer	Degeneracy	No. of genomes	Avg amplicon length
1	16S	ACTCCTACGGGAGGCAGC	TTACCGCGGCTGCTGGCA	0	1299	179
2	16S	CCAGCAGYYGCGGTAATA	GACTACCAGGGTATCTAATCCTG	4/0	1266	287.99
3	23S	TCCTAAGGTAGCGAAATTCCT	DACCGCCCCAGTHAAACT	0/4	1308	335
4	23S	SCKCCBAAAADGTAACGG	CGACATCGAGGTGCCAAA	16/0	P1+113	249
5	23S	AAAGACCCYRTGVASCTT	CTGTTATCCCYRGVGTAVCTTT	16/32	P2+204	398
6	rplB	GGYGGYGGHCAYAARCRN	RCSRCCRCCVTGNNGRTG	384/384	804	584
7	rplP	CARATHGARKCNGCNCGT	CYTTRCCNKHACCCATNC	384/384	683	127
8	rpsC	TGGGNCARAARRTHMAYC	RAYCCANACYTTNACDCC	384/384	636	610
9	rpsM	GGCNCGTATHGCNCGNRT	YTTDCGNGTRCGNGCRTT	384/384	629	329
10	rpsB	CAYTTYGGNCAYMARACH	GMRCGVATNGCRTCTCR	384/384	604	584
11	rplF	ATGTCNCGTRTHGSWAAN	CCYTDDCCYTTRTANGGN	384/384	581	482
12	rpsI	GGHACHGGHCGHCGYAAR	TTARCGYTDSWRWAYTG	324/384	541	372
13	rpsL	ATGSCWACNATYMABCAR	RTCYTTNACNCGDCCRCC	384/384	512	300

Table 1 – Primers for Universal Primer Set (UPS)

- 10 [0044] Figure 2 is a block diagram of an exemplary system for universal primer design in accord with the present invention. In this embodiment, a computing unit 200 is in communication with a source of pathogen genomic data 208 and a source of sequencing data 204.

- [0045] The computing unit 200 may take a variety of forms in various embodiments.
- 15 Exemplary computing units suitable for use with the present invention include desktop computers, laptop computers, virtual computers, server computers, smartphones, tablets, phablets, etc. Data sources 204, 208 may also take a variety of forms, including but not limited to structured databases (e.g., SQL databases), unstructured databases (e.g., Hadoop clusters,

NoSQL databases), or other data sources running on a variety of computing units (e.g., desktop computers, laptop computers, virtual computers, server computers, smartphones, tablets, phablets, etc.). The computing units may be heterogeneous or homogeneous in various embodiments of the present invention. In some embodiments, the data source 204 may be a
5 piece of sequencing equipment that sequences the genome of at least one microorganism in a sample. In some embodiments, the data source 208 may be a publicly or privately accessible database of pathogenic genomic data.

[0046] The components of the systems may be interconnected using a variety of network technologies being heterogeneous or homogenous in various embodiments. Suitable network
10 technologies include but are not limited to wired network connections (e.g., Ethernet, gigabit Ethernet, token ring, etc.) and wireless network connections (e.g., Bluetooth, 802.11x, 3G/4G wireless technologies, etc.).

[0047] In operation, the computing unit 200 queries the sequencing data source 204 for sequencing data for one or more microorganisms from a pathogen sample. The sequencing data
15 source 204 may have such information because it has performed such a test on the sample, or it may have received such information directly or indirectly (i.e., through data entry or transmission) from a piece of equipment that performed such testing.

[0048] In operation, the computing unit 200 queries the pathogen genomic data source 208 for information concerning the genomes for one or more pathogens identified by the sequencing
20 data source 204. The pathogen genomic data source 208 may have such information stored locally, or it may contact other computing units to obtain the relevant genomic information as necessary.

[0049] As discussed above, the computing unit 200 may receive sequencing data from the sequencing data source 204 for one or more amplicons generated using the UPS and query the
25 pathogen genomic data source 208 to identify the pathogen associated with the sequenced amplicon.

[0050] The computing unit 200 may access either data source 204, 208 first or access both data sources contemporaneously. In some embodiments, computing unit 200 is local to an operator, i.e., being located on a local area network accessed by the operator. In other

embodiments, computing unit 200 is accessed by an operator over yet another network connection (not shown), such as a wide area network or the Internet, and the graphical presentation is delivered to the operator over such network connection. In these embodiments, the computing unit 200 includes security and web server functionality customary to such remotely-accessed devices.

[0051] Embodiments of the present disclosure, for example, are described above with reference to block diagrams and/or operational illustrations of methods, systems, and computer program products according to embodiments of the present disclosure. The functions/acts noted in the blocks may occur out of the order as shown in any flowchart. For example, two blocks shown in succession may in fact be executed substantially concurrent or the blocks may sometimes be executed in the reverse order, depending upon the functionality/acts involved. Additionally, not all of the blocks shown in any flowchart need to be performed and/or executed. For example, if a given flowchart has five blocks containing functions/acts, it may be the case that only three of the five blocks are performed and/or executed. In this example, any of the three of the five blocks may be performed and/or executed.

[0052] The description and illustration of one or more embodiments provided in this application are not intended to limit or restrict the scope of the present disclosure as claimed in any way. The embodiments, examples, and details provided in this application are considered sufficient to convey possession and enable others to make and use the best mode of the claimed embodiments. The claimed embodiments should not be construed as being limited to any embodiment, example, or detail provided in this application. Regardless of whether shown and described in combination or separately, the various features (both structural and methodological) are intended to be selectively included or omitted to produce an embodiment with a particular set of features. Having been provided with the description and illustration of the present application, one skilled in the art may envision variations, modifications, and alternate embodiments falling within the spirit of the broader aspects of the general inventive concept embodied in this application that do not depart from the broader scope of the claimed embodiments.

CLAIMS

What is claimed is:

1. A method of amplifying nucleic acid in a sample, the method comprising:
selecting at least one primer from Table 1 for the amplification of nucleic acid in a
5 sample; and
producing an amplicon using the selected at least one primer.
2. The method of claim 1 further comprising:
sequencing said amplicon and obtaining an amplicon sequence;
comparing said amplicon sequence to a list of nucleic acids, wherein said list of nucleic
10 acids corresponds to nucleic acid sequences found in at least one pathogen; and
identifying the presence of at least one pathogen based on said comparison.
3. The method of claim 2 wherein identifying the at least one pathogen is performed in
fewer than 10 hours.
4. The method of claim 2 wherein said pathogen is bacterial.
- 15 5. The method of claim 2 further comprising determining treatment based on said presence
of said at least one identified pathogen.
6. The method of claim 5 wherein the sample is obtained from a human and treatment
comprises administration of an antibiotic to decrease the presence of the at least one identified
pathogen.
- 20 7. The method of claim 1 wherein said at least one primer from Table 1 comprises a primer
with the sequence ACTCCTACGGGAGGCAGC.
8. The method of claim 1 wherein said at least one primer from Table 1 comprises a primer
set comprising:
a primer with the sequence ACTCCTACGGGAGGCAGC;
25 a primer with the sequence CCAGCAGYYGCGGTAATA; and
a primer with the sequence TCCTAAGGTAGCGAAATTCCT.

9. The method of claim 1 wherein said at least one primer is optimized for the highest coverage of pathogens with the minimum number of primers.
10. The method of claim 1 where said at least one primer is optimized for melting temperature and guanine-cytosine content.
- 5 11. The method of claim 1 wherein the at least one primer is designed using phylogenetic conservation analysis.
12. The method of claim 1 wherein the at least one primer is optimized based on at least one of:
- 10 degeneracy to cover a plurality of species of pathogens; and
- variability to distinguish among said plurality of species of pathogens.
13. A system for amplifying nucleic acid in a sample, the system comprising:
- a primer set comprising at least one primer from Table 1.
14. A system for diagnosing the presence of at least one pathogen, the system comprising:
- 15 a primer set comprising at least one primer from Table 1, said primer set optimized to produce an amplicon of nucleic acid in a sample; and
- a database of nucleic acids corresponding to the at least one pathogen, wherein the presence of said at least one pathogen in a sample is identifiable based on a comparison of the sequence of the amplicon to the database of nucleic acids corresponding to the at least one pathogen.
- 20 15. The system of claim 14 wherein said primer set comprises:
- a primer with the sequence ACTCCTACGGGAGGCAGC;
- a primer with the sequence CCAGCAGYYGCGGTAATA; and
- a primer with the sequence TCCTAAGGTAGCGAAATTCCT...
- 25 16. The system of claim 14 wherein the at least one primer is further optimized based on at least one of:
- degeneracy to cover a plurality of species of pathogens; and
- variability to distinguish among said plurality of species of pathogens.

17. The system of claim 14 further comprising:

a bioinformatics engine configured to sequence and compare said amplicon with said database of nucleic acids corresponding to the at least one pathogen.

5 18. The system of claim 17 wherein the bioinformatics engine is further configured to produce a report comprising at least one of species ID, substrain ID, antibiotic for treatment; resistance value; and actionable treatment.

19. A method of obtaining a primer set comprising:

(a) identifying a plurality of pathogens of interest;

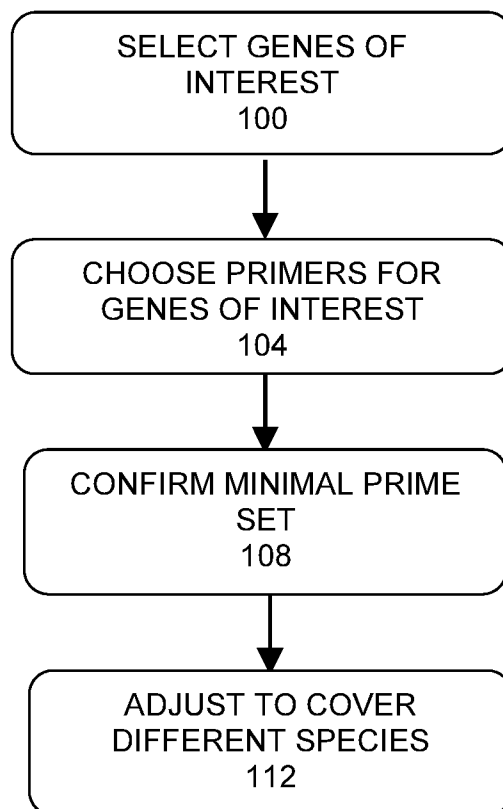
(b) identifying genes of said plurality of pathogens of interest;

10 (c) generating at least one primer for at least one of said identified genes wherein said at least one primer is generated based on:

an increased level of degeneracy among said identified genes of said plurality of pathogens of interest; and

15 variability among said identified genes of said plurality of pathogens of interest wherein said variability is sufficient to enable differentiation among said plurality of pathogens of interest; and

(d) repeating steps (a)-(c) to generate at least one additional primer for at least one other of said identified genes of said plurality of pathogens of interest until a primer is generated that corresponds to at least a majority of said plurality of pathogens of interest.

**FIG. 1**

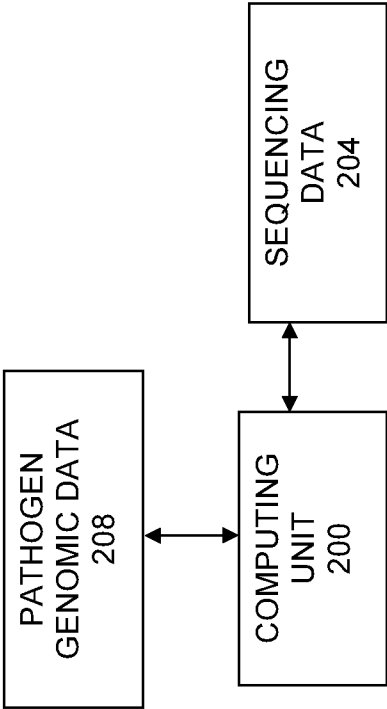


FIG. 2