



- (51) International Patent Classification:
A61B 3/10 (2006.01)
- (21) International Application Number:
PCT/US2014/064890
- (22) International Filing Date:
10 November 2014 (10.11.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/902,432 11 November 2013 (11.11.2013) US
- (71) Applicant: THE TRANSLATIONAL GENOMICS RESEARCH INSTITUTE [US/US]; 445 North Fifth Street, Suite 600, Phoenix, Arizona 85004 (US).
- (72) Inventors: SCHUPP, James M.; 445 North Fifth Street, Suite 600, Phoenix, Arizona 85004 (US). COLMAN, Rebecca; 445 North Fifth Street, Suite 600, Phoenix, Arizona 85004 (US). ENGELTHALER, David; 445 North Fifth Street, Suite 600, Phoenix, Arizona 85004 (US). GILLECE, John; 445 North Fifth Street, Suite 600, Phoenix, Arizona 85004 (US). HICKS, Nathan; 445 North Fifth Street, Suite 600, Phoenix, Arizona 85004 (US). KEIM, Paul S.; 445 North Fifth Street, Suite 600, Phoenix, Arizona 85004 (US).
- (74) Agent: FULLER, Rodney J; Booth Udall Fuller, PLC, 1255 West Rio Salado Parkway, Suite 215, Tempe, Arizona 85281 (US).

- (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, 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).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: SYSTEMS AND METHODS FOR UNIVERSAL TAIL-BASED INDEXING STRATEGIES FOR AMPLICON SEQUENCING

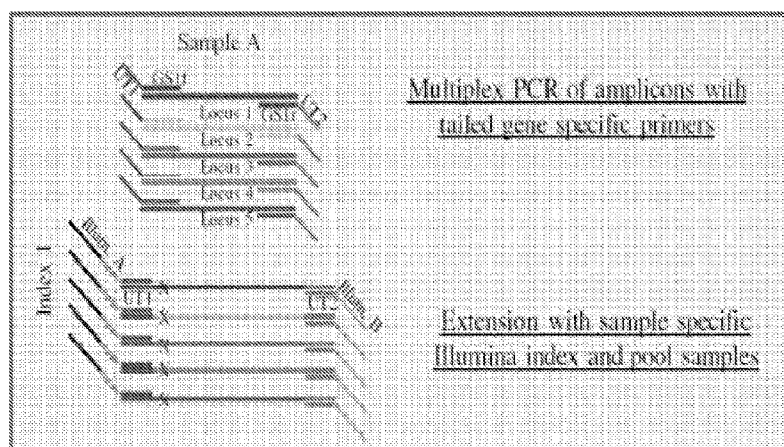


FIGURE 1A

(57) Abstract: Some embodiments of the invention include a method of preparing a sample for sequencing that includes receiving a sample and amplifying at least one marker within the sample. In some embodiments, amplification of the first marker may include mixing the sample with a first oligonucleotide that comprises a first universal tail sequence and a second oligonucleotide that comprises a second universal tail sequence. In some aspects of the invention, the first universal tail sequence and the second universal tail sequence are different sequences.

WO 2015/070187 A2

SYSTEMS AND METHODS FOR UNIVERSAL TAIL-BASED INDEXING STRATEGIES FOR AMPLICON SEQUENCING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Patent Application No. 61/902,432 filed November 11, 2013, which is hereby incorporated by reference in its entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0002] The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII-formatted sequence listing with a file named "91482_154_Sequence_Listing.txt" created on November 5, 2014, and having a size of 16 kilobytes, and is filed concurrently with the specification. The sequence listing contained in this ASCII-formatted document is part of the specification and is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0003] This application relates to systems and methods that can be used in conjunction with sequencing processes, and more particularly, this application relates to systems and methods of preparing amplicon products for sequencing through the use of a universal tail-based indexing strategy.

BACKGROUND OF THE INVENTION

[0004] High throughput amplicon sequencing of genomic regions of interest can be very useful for a variety of molecular genetic genotyping applications, including forensic and clinical sample analysis. Some conventional sequencing platforms provide for rapid, high yield sequence data, which can enable the sequencing of multiple amplicons from many samples in a short period of time. However, efficient use of at least some of these sequencing platforms requires the use of sample barcoding, which can be cumbersome and expensive when dealing with tens to thousands of samples. This complication is further compounded if multiple amplicons are to be sequenced from each sample. As such, there is a demonstrated need to develop systems and methods that enable the production of sequencing platform-ready amplicons in a multiplex fashion with a common set of

indexing oligonucleotide sequences that can be used in preparing any amplicon for sequencing.

[0005] For example, PCR reactions can be used to generate amplicons that are sequencing-platform ready for some systems (e.g., bacterial ribosomal gene sequencing) (JG Caporaso et al. (2012) ISME J. 6(8) 1621-1624). Moreover, some conventional methodologies include indexing, which comprises the use of a plasmid for production of the indexing oligonucleotide (Chubiz et al. (2012) PLoS One 7(10) e47959). Other conventional sequencing platforms may be compatible with custom amplicon sequencing products designed for different eukaryotic organisms, but these conventional systems utilize an initial hybridization of probes with an extension-ligation reaction followed by a PCR reaction requiring at least 50 nanograms (ng) of pure, high quality DNA template as starting material. As such, there is a demonstrated need for systems and methods that provide for substantially or completely simultaneous amplification of multiple amplicon targets from any organism in a PCR reaction with subsequent universal indexing addition that requires only a single set of common indexing oligonucleotides that can be used with any set of amplicons from any type of nucleic acid-based sample, regardless of the quality of the sample.

SUMMARY

[0006] Some embodiments of the invention include a method of preparing a sample for sequencing that includes receiving a sample and amplifying at least one marker within the sample. In some embodiments, amplification of the first marker may include mixing the sample with a first oligonucleotide that comprises a first universal tail sequence and a second oligonucleotide that comprises a second universal tail sequence. In some aspects of the invention, the first universal tail sequence and the second universal tail sequence are different sequences. In some embodiments of the invention, the sequences of the first and second universal tail sequences may be at least partially correlated with the guanine-cytosine content of the first marker, and the organism from which the first marker is to be amplified.

[0007] In some aspects, the method may further include amplifying a second marker by mixing the sample with a third oligonucleotide that includes the first universal tail sequence and a fourth oligonucleotide that includes the second universal tail sequence. For example, the first marker and the second marker can be amplified in the same reaction vessel such that the amplification of these two

markers occurs during a multiplex PCR assay. In some embodiments, more than two markers can be amplified during the multiplex PCR assay.

[0008] In addition, the amplification of the first marker and the second marker can result in the production of a first amplicon and the second amplicon that comprise the first and second universal tail sequences integrated therein. Some embodiments of the invention further include adding an index to the first amplicon and the second amplicon using at least one indexing oligonucleotide. For example, the indexing oligonucleotide comprises a complementary sequence that recognizes at least one of the first universal tail sequence and the second universal tail sequence that has been previously integrated within the first and second amplicons.

[0009] In some aspects of the invention, the sample can be from any source desirable for sequencing analysis. For example, the sample can be a portion of a pathogenic organism and the first and/or second markers may be antibiotic resistance genes. In other embodiments, the sample may be from a castor bean or a castor bean product. Moreover, the sample may also be from an animal, such as a human. In some aspects, the sample may be from a human and the marker may be affiliated with a known human condition, such as cystic fibrosis.

[0010] In certain aspects, the first universal tail sequence and the second universal tail sequence independently comprise a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, and SEQ ID NO: 4.

[0011] In other aspects, the at least one indexing oligonucleotide comprises a sequence selected from the group consisting of SEQ ID NOs: 9-44.

[0012] Some embodiments of the invention provide a method of assessing the presence of at least one antibiotic resistance gene within a population of organisms. The method may include initially receiving a sample of the population of organisms and then extracting a template from the sample. Moreover, the method may also include performing a multiplex polymerase chain reaction assay that includes amplifying a first marker from the template and a second marker from the template within the same reaction vessel.

[0013] In yet other embodiments, the sequencing comprises amplifying the first marker and/or second marker with a sequencing primer. The sequencing primer may comprise a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

[0014] In some aspects, amplifying the first marker can include mixing the template with a first oligonucleotide that includes a first universal tail sequence and a second oligonucleotide that includes a second universal tail sequence. Similarly, amplifying the second marker includes mixing the template with a third oligonucleotide comprising the first universal tail sequence and a fourth oligonucleotide comprising the second universal tail sequence. As a result of this amplification step, a first amplicon and a second amplicon are produced. In other embodiments, more than a first amplicon and a second amplicon can be produced.

[0015] In some aspects, the template can be any variety of molecules. For example, the template can be a crude nucleic acid extract from the population of organisms that may include DNA and/or RNA. In some aspects, the template can comprise double-stranded or single-stranded DNA. Moreover, in some embodiments, the template can comprise RNA.

[0016] In other embodiments, the present invention is directed to a method of determining gene copy number and/or quantifying gene expression in a sample, the method comprising the steps of: a) receiving a sample; b) amplifying at least a first marker within the sample to produce a first amplicon, wherein amplifying at least the first marker comprises mixing the sample with a first oligonucleotide comprising a first universal tail sequence and a second oligonucleotide comprising a second universal tail sequence, wherein the first universal tail sequence and the second oligonucleotide sequence are different; c) adding an index to the first amplicon using at least one indexing oligonucleotide, wherein the at least one indexing oligonucleotide comprises a complementary sequence that recognizes at least one of the first universal tail sequence and the second universal tail sequence; d) sequencing the first amplicon to produce a number of sequencing reads; and e) determining gene copy number and/or quantifying gene expression from differential target sequencing read counts wherein the sequencing read counts of the sample are compared with sequencing read counts of a reference sample.

[0017] In certain aspects, the reference sample is a sample from a wild-type organism. In other aspects, the reference sample is an internal standard comprising one or more markers of interest. In yet other aspects, the reference sample is derived from a healthy subject or from a healthy tissue in which disease has not been detected.

[0018] Additional objectives, advantages and novel features will be set forth in the description which follows or will become apparent to those skilled in the art upon examination of the drawings and detailed description which follows.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] **FIG. 1** is a simplified illustration detailing steps that can be performed in accord with some embodiments of the invention. **FIG. 1A** depicts multiplex PCR of amplicons with tailed gene specific primers. **FIG. 1B** shows pooled multi-sample/amplicon sequencing, and **FIG. 1C** shows demultiplexing of sample reads, alignment to reference sequences, and subsequent analysis.

[0020] **FIG. 2** is a listing of universal tail sequences and oligonucleotide sequences that can be used in conjunction with the universal tail sequences.

[0021] **FIG. 3** is a series of output graphs from an AGILENT BIOANALYZER[®] that detail the resulting concentrations of amplicons produced using 5 ng and 500 fg of input template using some embodiments of the invention.

[0022] **FIG. 4** presents an AGILENT BIOANALYZER[®] analysis with a six-plex *Mycobacterium tuberculosis* (Mtb) antibiotic resistance gene amplicon sequencing assay. The chromatogram depicts analysis of a multiplex amplification reaction after amplicon yield optimization.

[0023] **FIG. 5** depicts AGILENT BIOANALYZER[®] chromatograms from a universal tail (UT) amplicon six-plex assay for 11 diverse *Burkholderia pseudomallei* strains and 1 *Burkholderia mallei* strain. Arrows in the Bp_406e panel indicate peaks corresponding to each amplicon.

[0024] **FIG. 6** shows the potential of the UT amplicon multiplex assays for quantitative analysis. **FIG. 6A** depicts a quantitative DNA Sequencing (qDNA-Seq) analysis on a mock variable copy number template. **FIG. 6B** depicts a comparison of transcriptome RNAseq analysis vs. targeted UT amplicon sequencing RNAseq analysis in a *Burkholderia pseudomallei* six-plex amplicon multiplex assay.

[0025] **FIG. 7** depicts an internal positive PCR and sequencing control for universal tail amplicon sequencing.

[0026] The headings used in the figures should not be interpreted to limit the scope of the claims.

DETAILED DESCRIPTION

[0027] Some embodiments of the invention include systems and methods of preparing samples for one or more downstream processes that can be used for assessing one or more markers. Some embodiments of the invention may comprise a universal indexing sequencing strategy for use in downstream sequencing platform processes. By way of example only, some embodiments of the invention comprise a universal indexing sequencing strategy that can be used to amplify multiple genomic regions (e.g., markers, as described below) from a DNA sample simultaneously in a single reaction for the sequencing of one or more amplicons. One or more embodiments of the invention can be used with any desired sequencing platform, such as the ILLUMINA® Next Generation Sequencing (e.g., MiSEQ) platform, Life Technologies' Ion Torrent System, or any other sequencing system now known or developed in the future.

[0028] Moreover, some embodiments of the invention can be configured to be used with non-DNA samples (e.g., RNA samples) for applications that may comprise functionalities such as quantification of expression of one or more markers. For example, RNA can be extracted from a sample, converted to cDNA using techniques known in the art, and the cDNA can function as the template for additional processes (e.g., multiplex polymerase chain reaction assays and subsequent sequencing).

[0029] Some embodiments may be configured to enable relatively simple, inexpensive, and efficient preparation of samples for use on, in, and/or with downstream sequencing platforms. For example, this improvement over some conventional systems can originate from the use of a sequence coupled to one or more oligonucleotides / primers (as used herein, oligonucleotides and primers are used interchangeably). More specifically, one or more amplicons per sample can be generated using a hybrid oligonucleotide that is designed for amplification of a marker and incorporation of at least one universal tail sequence into the resulting amplicon. As a result, additional steps that may be conventionally required to prepare samples for sequencing can be limited or removed entirely.

[0030] In addition, at least some embodiments of the invention comprise the use of PCR before sequencing such that only limited amounts of starting material are necessary and the starting material need not be of high quality (e.g., genomic DNA, crude DNA extracts, single stranded DNA, RNA, cDNA, etc.).

In contrast, many conventional sample preparation systems may require relatively large amounts of starting material of relatively high quality, which can limit the use of these systems.

[0031] Some embodiments of the invention can be used for and/or in complement with high-throughput amplicon sequencing of markers, which can be very useful for a variety of molecular genetic genotyping applications, including forensic and clinical sample analysis. For example, use of the systems and methods of the invention can be employed with sequencing platforms to provide rapid, high-yield sequence data, which can enable the sequencing of multiple markers/amplicons from many samples in a relatively short period of time. In particular, many sequencing platforms may require the use of sample barcoding, which can be cumbersome and expensive when dealing with tens to hundreds or thousands of samples, which is a difficulty that can be further compounded if multiple amplicons are to be sequenced from each sample. As such, some embodiments of the invention can include systems and methods that comprise the use of universal tail sequences that can enable the production of sequencing platform-ready amplicons in a multiplex fashion. As described in greater detail below, one or more sets of indexing oligonucleotide sequences can in connection with amplicons that have one or more universal tail sequences incorporated therein.

[0032] Generally, some embodiments of the present invention can be used to detect, identify, assess, sequence, or otherwise evaluate a marker. A marker may be any molecular structure produced by a cell, expressed inside the cell, accessible on the cell surface, or secreted by the cell. A marker may be any protein, carbohydrate, fatty acid, nucleic acid, catalytic site, or any combination of these such as an enzyme, glycoprotein, cell membrane, virus, a particular cell, or other uni- or multimolecular structure. A marker may be represented by a sequence of a nucleic acid or any other molecules derived from the nucleic acid. Examples of such nucleic acids include miRNA, tRNA, siRNA, mRNA, cDNA, genomic DNA sequences, single-stranded DNA, or complementary sequences thereof. Alternatively, a marker may be represented by a protein sequence. The concept of a marker is not limited to the exact nucleic acid sequence or protein sequence or products thereof, rather it encompasses all molecules that may be detected by a method of assessing the marker. Without being limited by the theory, the detection, identification, assessment, sequencing, or any other evaluation of the marker may encompass an

assessment of a change in copy number (e.g., copy number of a gene or other forms of nucleic acid) or in the detection of one or more translocations. Moreover, in some embodiments, the marker may be relevant to a particular phenotype or genotype. By way of example only, in some embodiments, the marker may be related to phenotypes including antibiotic resistance, virulence, or any other phenotype.

[0033] Therefore, examples of molecules encompassed by a marker represented by a particular sequence further include alleles of the gene used as a marker. An allele includes any form of a particular nucleic acid that may be recognized as a form of the particular nucleic acid on account of its location, sequence, or any other characteristic that may identify it as being a form of the particular gene. Alleles include but need not be limited to forms of a gene that include point mutations, silent mutations, deletions, frameshift mutations, single nucleotide polymorphisms (SNPs), inversions, translocations, heterochromatic insertions, and differentially methylated sequences relative to a reference gene, whether alone or in combination. An allele of a gene may or may not produce a functional protein; may produce a protein with altered function, localization, stability, dimerization, or protein-protein interaction; may have overexpression, underexpression or no expression; may have altered temporal or spatial expression specificity; or may have altered copy number (e.g., greater or less numbers of copies of the allele). An allele may also be called a mutation or a mutant. An allele may be compared to another allele that may be termed a wild type form of an allele. In some cases, the wild type allele is more common than the mutant.

[0034] Some embodiments of the invention may comprise the use of one or more methods of amplifying a nucleic acid-based starting material (i.e., a template, including genomic DNA, crude DNA extract, single-stranded DNA, double-stranded DNA, cDNA, RNA, or any other single-stranded or double-stranded nucleic acids). Nucleic acids may be selectively and specifically amplified from a template nucleic acid contained in a sample. In some nucleic acid amplification methods, the copies are generated exponentially. Examples of nucleic acid amplification methods known in the art include: polymerase chain reaction (PCR), ligase chain reaction (LCR), self-sustained sequence replication (3SR), nucleic acid sequence based amplification (NASBA), strand displacement amplification (SDA), amplification with Q β replicase, whole genome amplification with enzymes such as ϕ 29, whole genome PCR, in vitro transcription with T7 RNA polymerase or any other RNA

polymerase, or any other method by which copies of a desired sequence are generated.

[0035] In addition to genomic DNA, any polynucleotide sequence can be amplified with an appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

[0036] PCR generally involves the mixing of a nucleic acid sample, two or more primers or oligonucleotides (primers and oligonucleotides are used interchangeably herein) that are designed to recognize the template DNA, a DNA polymerase, which may be a thermostable DNA polymerase such as Taq or Pfu, and deoxyribose nucleoside triphosphates (dNTP's). In some embodiments, the DNA polymerase used can comprise a high fidelity Taq polymerase such that the error rate of incorrect incorporation of dNTPs is less than one per 1,000 base pairs. Reverse transcription PCR, quantitative reverse transcription PCR, and quantitative real time reverse transcription PCR are other specific examples of PCR. In general, the reaction mixture is subjected to temperature cycles comprising a denaturation stage (typically 80-100° C), an annealing stage with a temperature that is selected based on the melting temperature (T_m) of the primers and the degeneracy of the primers, and an extension stage (for example 40-75° C). In real-time PCR analysis, additional reagents, methods, optical detection systems, and devices known in the art are used that allow a measurement of the magnitude of fluorescence in proportion to concentration of amplified template. In such analyses, incorporation of fluorescent dye into the amplified strands may be detected or measured.

[0037] Either primers or primers along with probes allow a quantification of the amount of specific template DNA present in the initial sample. In addition, RNA may be detected by PCR analysis by first creating a DNA template from RNA through a reverse transcriptase enzyme (i.e., the creation of cDNA). The marker expression may be detected by quantitative PCR analysis facilitating genotyping analysis of the samples.

[0038] "Amplification" is a special case of nucleic acid replication involving template specificity. Amplification may be a template-specific replication or a non-template-specific replication (i.e., replication may be specific template-dependent or not). Template specificity is here distinguished from fidelity of replication (synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or

deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out. The amplification process may result in the production of one or more amplicons.

[0039] The term "template" refers to nucleic acid originating from a sample that is analyzed for the presence of one or more markers. In contrast, "background template" or "control" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified out of the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

[0040] In addition to primers and probes, template specificity is also achieved in some amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under the conditions in which they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. Other nucleic acid sequences will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin et al. (1970) *Nature* (228):227). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace (1989) *Genomics* (4):560). Finally, Taq and Pfu polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H. A. Erlich (ed.) (1989) *PCR Technology*, Stockton Press).

[0041] The term "amplifiable nucleic acid" refers to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template." The terms "PCR product," "PCR fragment," "amplification product," and "amplicon" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing

and extension. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

[0042] In some forms of PCR assays, quantification of a target in an unknown sample is often required. Such quantification may be determined in reference to the quantity of a control sample. The control sample starting material / template may be co-amplified in the same tube in a multiplex assay or may be amplified in a separate tube. Generally, the control sample contains template at a known concentration. The control sample template may be a plasmid construct comprising only one copy of the amplification region to be used as quantification reference. To calculate the quantity of a target in an unknown sample, various mathematical models are established. Calculations are based on the comparison of the distinct cycle determined by various methods, e.g., crossing points (CP) and cycle threshold values (Ct) at a constant level of fluorescence; or CP acquisition according to established mathematic algorithm.

[0043] Some embodiments of the invention may comprise a multiplex assay. As used herein, the term "multiplex" refers to the production of more than one amplicon, PCR product, PCR fragment, amplification product, etc. in a single reaction vessel. In other words, multiplex is to be construed as the amplification of more than one target-specific sequences within a PCR reaction or assay within the same PCR assay mixture (e.g., more than one amplicon is produced within a single vessel that contains all of the reagents necessary to perform a PCR reaction). In some embodiments, a step prior to performing the PCR (or RT-PCR, quantitative RT-PCR, etc.) reaction can occur such that sets of primers and/or primers and probes are designed, produced, and optimized within a given set of reaction conditions to ensure proper amplicon production during the performance of the PCR.

[0044] The algorithm for Ct values in real time-PCR calculates the cycle at which each PCR amplification reaches a significant threshold. The calculated Ct value is proportional to the number of marker copies present in the sample, and the Ct value is a precise quantitative measurement of the copies of the marker found in any sample. In other words, Ct values represent the presence of respective marker that the primer sets are designed to recognize. If the marker is missing in a sample, there should be no amplification in the Real Time-PCR reaction.

[0045] Alternatively, the Cp value may be utilized. A Cp value represents the cycle at which the increase of fluorescence is highest and where the

logarithmic phase of a PCR begins. The LightCycler® 480 Software calculates the second derivatives of entire amplification curves and determines where this value is at its maximum. By using the second-derivative algorithm, data obtained are more reliable and reproducible, even if fluorescence is relatively low.

[0046] The various and non-limiting embodiments of the PCR-based method detecting marker expression level as described herein may comprise one or more probes and/or primers. Generally, the probe or primer contains a sequence complementary to a sequence specific to a region of the nucleic acid of the marker gene. A sequence having less than 60% 70%, 80%, 90%, 95%, 99% or 100% identity to the identified gene sequence may also be used for probe or primer design if it is capable of binding to its complementary sequence of the desired target sequence in marker nucleic acid.

[0047] Some embodiments of the invention may include a method of comparing a marker in a sample relative to one or more control samples. A control may be any sample with a previously determined level of expression. A control may comprise material within the sample or material from sources other than the sample. Alternatively, the expression of a marker in a sample may be compared to a control that has a level of expression predetermined to signal or not signal a cellular or physiological characteristic. This level of expression may be derived from a single source of material including the sample itself or from a set of sources.

[0048] The sample in this method is preferably a biological sample from a subject. The term "sample" or "biological sample" is used in its broadest sense. Depending upon the embodiment of the invention, for example, a sample may comprise a bodily fluid including whole blood, serum, plasma, urine, saliva, cerebral spinal fluid, semen, vaginal fluid, pulmonary fluid, tears, perspiration, mucus and the like; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print, or any other material isolated in whole or in part from a living subject or organism. Biological samples may also include sections of tissues such as biopsy and autopsy samples, and frozen sections taken for histologic purposes such as blood, plasma, serum, sputum, stool, tears, mucus, hair, skin, and the like. Biological samples also include explants and primary and/or transformed cell cultures derived from patient tissues. In some embodiments, sample may comprise

a portion of a non-animal organism, such as a plant (e.g., castor beans or derivatives thereof).

[0049] In some embodiments, sample or biological sample may include a bodily tissue, fluid, or any other specimen that may be obtained from a living organism that may comprise additional living organisms. By way of example only, in some embodiments, sample or biological sample may include a specimen from a first organism (e.g., a human) that may further comprise an additional organism (e.g., bacteria, including pathogenic or non-pathogenic / commensal bacteria, viruses, parasites, fungi, including pathogenic or non-pathogenic fungi, etc.). In some embodiments of the invention, the additional organism may be separately cultured after isolation of the sample to provide additional starting materials for downstream analyses. In some embodiments, the sample or biological sample may comprise a direct portion of the additional, non-human organism and the host organism (e.g., a biopsy or sputum sample that contains human cells and bacteria).

[0050] The invention may further comprise the step of sequencing the amplicon. Methods of sequencing include but need not be limited to any form of DNA sequencing including Sanger, next-generation sequencing, pyrosequencing, SOLiD sequencing, massively parallel sequencing, pooled, and barcoded DNA sequencing or any other sequencing method now known or yet to be disclosed.

[0051] In Sanger Sequencing, a single-stranded DNA template, a primer, a DNA polymerase, nucleotides and a label such as a radioactive label conjugated with the nucleotide base or a fluorescent label conjugated to the primer, and one chain terminator base comprising a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP, are added to each of four reaction (one reaction for each of the chain terminator bases). The sequence may be determined by electrophoresis of the resulting strands. In dye terminator sequencing, each of the chain termination bases is labeled with a fluorescent label of a different wavelength that allows the sequencing to be performed in a single reaction.

[0052] In pyrosequencing, the addition of a base to a single-stranded template to be sequenced by a polymerase results in the release of a pyrophosphate upon nucleotide incorporation. An ATP sulfurylase enzyme converts pyrophosphate into ATP that in turn catalyzes the conversion of luciferin to oxyluciferin which results

in the generation of visible light that is then detected by a camera or other sensor capable of capturing visible light.

[0053] In SOLiD sequencing, the molecule to be sequenced is fragmented and used to prepare a population of clonal magnetic beads (in which each bead is conjugated to a plurality of copies of a single fragment) with an adaptor sequence and alternatively a barcode sequence. The beads are bound to a glass surface. Sequencing is then performed through 2-base encoding.

[0054] In massively parallel sequencing, randomly fragmented targeted nucleic acids and/or amplicons are attached to a surface. The fragments / amplicons are extended and bridge amplified to create a flow cell with clusters, each with a plurality of copies of a single fragment sequence. The templates are sequenced by synthesizing the fragments in parallel. Bases are indicated by the release of a fluorescent dye correlating to the addition of the particular base to the fragment. Nucleic acid sequences may be identified by the IUAPC letter code which is as follows: A - Adenine base; C- Cytosine base; G - guanine base; T or U - thymine or uracil base. M - A or C; R - A or G; W - A or T; S - C or G; Y - C or T; K - G or T; V - A or C or G; H - A or C or T; D - A or G or T; B - C or G or T; N or X - A or C or G or T. Note that T or U may be used interchangeably depending on whether the nucleic acid is DNA or RNA. A sequence having less than 60%, 70%, 80%, 90%, 95%, 99% or 100% identity to the identifying sequence may still be encompassed by the invention if it is able of binding to its complimentary sequence and/or facilitating nucleic acid amplification of a desired target sequence. In some embodiments, as previously mentioned, the method may include the use of massively parallel sequencing, as detailed in U.S. Patent Nos. 8,431,348 and 7,754,429, which are hereby incorporated by reference in their entirety.

[0055] Some embodiments of the invention comprise multiple steps and/or processes that are carried out to execute the universal tail indexing strategy to prepare amplicons for sequencing. In some embodiments, one or more markers for a given sample or template can be selected. Some embodiments of the invention can be used in conjunction with an analysis of one or more markers (e.g., genes / alleles) associated with a particular phenotype (e.g., resistance to one or more pharmaceuticals, such as antibiotics). By way of example only, some embodiments of the invention can be used to detect and/or quantify the development of antibiotic resistance in populations of patients infected with an organism (e.g., *Mycobacterium*

tuberculosis). As such, in some aspects, prior to performing additional steps, an investigator can assess the markers present within the genome of the organism to determine which markers are implicated in the development of antibiotic resistance. For example, markers can be selected that may contain a SNP or other change or alteration that can confer at least partial antibiotic resistance. In other aspects of the invention, markers can be selected that are not implicated in antibiotic resistance, but are associated with other phenotypes / genotypes that are desirable for further analysis. Regardless of the markers selected (e.g., markers related to antibiotic resistance or markers that are not related to antibiotic resistance), the markers may comprise sequence variations that comprise a degree of differentiability relative to a wild-type or control version of the marker such that a comparison of the marker and the wild-type/control sequence of the marker will enable an investigator to assess the genotype of the organism being tested.

[0056] After selection of the markers, marker-specific primers / oligonucleotides can be designed for the amplification of the markers to produce the desired amplicons. As is known in the art, a forward and a reverse marker-specific primer can be designed to amplify the marker from a nucleic acid sample. In some embodiments, the forward and reverse primers can be designed to produce an amplicon (e.g., some or all of the sequence of the marker) of a desired length. For example, the length of the amplicon may comprise approximately 50 base pairs (bp), 100 bp, 150 bp, 200 bp, 250 bp, 300 bp, 350 bp, 400 bp, 450 bp, 500 bp, 1,000 bp, or any size amplicon therebetween.

[0057] As previously mentioned, some embodiments of the invention may include a multiplex PCR reaction. For example, marker-specific primers can be designed for multiple markers or multiple regions of the same marker such that multiple amplicons of between about 50 bp and 1,000 bp are being produced within a single PCR reaction vessel. In other words, the forward and reverse primers can be designed to function within a given set of temperature parameters such that more than one amplicon can be successfully amplified from a given template within a single PCR reaction mixture. As such, multiple amplicons can be prepared using the universal tail indexing strategy for sequencing preparation. In some embodiments, the multiplex PCR reaction can result in the production of between two and 10 amplicons. In other embodiments, the multiplex PCR reaction can result in the

production of more than 10 amplicons. In some embodiments, the multiplex PCR reaction can be optimized to produce about five amplicons.

[0058] In some embodiments, the forward and reverse primers that have been designed for each of the markers can be modified to include a universal tail. For example, the universal tail sequences can be relatively or completely unique sequences of nucleotides that are coupled to the 5' ends of some or all of the forward and reverse marker-specific primers. In some aspects, the universal tail sequences can be selected such that there is little to no overlap in sequence between portions of the markers that are being amplified and the universal tail sequences. Moreover, the universal tail sequences can comprise a length between ten and twenty nucleotides in length. In some embodiments, the universal tail sequences can be any other length, as desired by the user to meet the needs and requirements of the reaction. As such, the universal tail sequences can exhibit a relatively negligible impact on binding of the forward and reverse marker-specific primers to the template sequence to enable amplification. Moreover, as a result of being included on the 5' end of the forward and reverse marker-specific primers, the universal tail sequences will form a portion of the resulting amplicons. In addition, in some aspects of the invention, the sequences selected for the universal tail sequences can be at least partially correlated with the chemical composition of the template nucleic acids. For example, in some aspects, the sequences selected for the universal tail sequences can be at least partially correlated with the G-C content of the organism from which the template is isolated.

[0059] In some aspects, some or all of the universal tail sequences can be at least partially unique. In some embodiments, each of the 5' ends of all of the forward marker-specific primers within a given PCR assay mixture can comprise the same or a similar universal tail sequence (e.g., a first universal tail sequence or UT1). Similarly, each of the 5' ends of all of the reverse marker-specific primers within the same PCR assay mixture can comprise a second universal tail sequence (UT2) that differs from the first universal tail sequence. As such, each respective sample from which a template sequence is used in the multiplex PCR assay will have two unique universal tail sequences. Accordingly, each forward and reverse marker-specific primer within a multiplex PCR mixture will include a unique universal tail sequence. For example, if the PCR includes 35 different samples, 35 universal tail sequences can be employed for the forward primers in each of the 35 unique

reactions (i.e., not including technical replicates) and 35 universal tail sequences can be employed for the reverse primers in each of the 35 unique reactions (i.e., not including technical replicates). Overall, the forward and reverse marker-specific primers that each comprise the universal tail sequences can comprise a generally short length (e.g., 25-50 bp), which can facilitate simultaneous amplification of multiple targets in a single reaction.

[0060] In addition, some embodiments of the invention may comprise performing quantitative PCR to optimize the multiplex PCR assay. For example, after design of the forward and reverse marker-specific primers that each include a universal tail sequence, the contemplated multiplex PCR assays can be performed using quantitative PCR (e.g., using DNA as a template) to assess relative quantities of the amplicons produced. Accordingly, the sequence coverage of each amplicon is considered to be equal if the quantities of the amplicons produced by the multiplex quantitative PCR appear to be equal. If the quantities of the amplicons produced by the multiplex quantitative PCR do not appear to be equal, the forward and/or reverse marker-specific primers can be altered and re-optimized until adequate quantities of amplicons are produced.

[0061] After design and adequate optimization of the multiplex PCR assay comprising multiple forward and reverse marker-specific primers that each include universal tail sequences, the multiplex PCR can be performed to obtain the amplicons. In some embodiments, template that has been previously isolated from a sample can be used for the amplification of the amplicons. In some aspects, multiple PCR reaction replicates can be performed for each sample template and one or more control templates. For example, in embodiments configured to assess / quantify the development of antibiotic resistance in a population of organisms (e.g., bacteria), DNA can be isolated from one or more representative members of the population (e.g., colonies) to serve as the template. Moreover, DNA may also be isolated from a population of similar organisms that are known to not have a genotype that confers some level of antibiotic resistance to function as a control (e.g., a negative control). In addition, DNA may also be isolated from a population of similar organisms that are known to have a genotype that confers some level of antibiotic resistance to function as a control (e.g., a positive control). In other embodiments, the markers associated with a particular phenotype (e.g., antibiotic

resistance) may already be known such that one or both of the control samples may not be necessary for comparison.

[0062] In certain aspects, the presence of antibiotic resistance in a population of organisms is indicated by detection of a particular gene or allele in the population. Non-limiting examples of genes that confer antibiotic resistance include *bla_{tem}*, *bla_{shv}*, *bla_{rob}*, *bla_{oxa}*, *bla_Z*, *aadB*, *aacC1*, *aacC2*, *aacC3*, *aac6'-IIa*, *aacA4*, *aad(6')*, *vanA*, *vanB*, *vanC*, *msrA*, *sarA*, *aac(6') aph(2'')*, *vat*, *vga*, *ermA*, *ermB*, *ermC*, *mecA*, *int*, *sul*, *mecA*, *aac2ia*, *aac2ib*, *aac2ic*, *aac2id*, *aac2i*, *aac3ia*, *aac3iia*, *aac3iib*, *aac3iii*, *aac3iv*, *aac3ix*, *aac3vi*, *aac3viii*, *aac3vii*, *aac3x*, *aac6i*, *aac6ia*, *aac6ib*, *aac6ic*, *aac6ie*, *aac6if*, *aac6ig*, *aac6iia*, *aac6iib*, *aad9*, *aad9ib*, *aadd*, *acra*, *acrb*, *adea*, *adeb*, *adec*, *amra*, *amrb*, *ant2ia*, *ant2ib*, *ant3ia*, *ant4iia*, *ant6ia*, *aph33ia*, *aph33ib*, *aph3ia*, *aph3ib*, *aph3ic*, *aph3iia*, *aph3iva*, *aph3va*, *aph3vb*, *aph3via*, *aph3vii*, *aph4ib*, *aph6ia*, *aph6ib*, *aph6ic*, *aph6id*, *arna*, *backa*, *bcra*, *bcrc*, *bl1_acc*, *bl1_ampc*, *bl1_asba*, *bl1_ceps*, *bl1_cmy2*, *bl1_ec*, *bl1_fox*, *bl1_mox*, *bl1_och*, *bl1_pao*, *bl1_pse*, *bl1_sm*, *bl2a_1*, *bl2a_exo*, *bl2a_iii2*, *bl2a_iii*, *bl2a_kcc*, *bl2a_nps*, *bl2a_okp*, *bl2a_pc*, *bl2be_ctxm*, *bl2be_oxy1*, *bl2be_per*, *bl2be_shv2*, *bl2b_rob*, *bl2b_tem1*, *bl2b_tem2*, *bl2b_tem*, *bl2b_tle*, *bl2b_ula*, *bl2c_bro*, *bl2c_pse1*, *bl2c_pse3*, *bl2d_lcr1*, *bl2d_moxa*, *bl2d_oxa10*, *bl2d_oxa1*, *bl2d_oxa2*, *bl2d_oxa5*, *bl2d_oxa9*, *bl2d_r39*, *bl2e_cbla*, *bl2e_cepa*, *bl2e_cfxa*, *bl2e_fpm*, *bl2e_y56*, *bl2f_nmca*, *bl2f_sme1*, *bl2_ges*, *bl2_kpc*, *bl2_len*, *bl2_veb*, *bl3_ccra*, *bl3_cit*, *bl3_cpha*, *bl3_gim*, *bl3_imp*, *bl3_l*, *bl3_shw*, *bl3_sim*, *bl3_vim*, *ble*, *blt*, *bmr*, *cara*, *cata10*, *cata11*, *cata12*, *cata13*, *cata14*, *cata15*, *cata16*, *cata1*, *cata2*, *cata3*, *cata4*, *cata5*, *cata6*, *cata7*, *cata8*, *cata9*, *catb1*, *catb2*, *catb3*, *catb4*, *catb5*, *ceoa*, *ceob*, *cml_e1*, *cml_e2*, *cml_e3*, *cml_e4*, *cml_e5*, *cml_e6*, *cml_e7*, *cml_e8*, *dfra10*, *dfra12*, *dfra13*, *dfra14*, *dfra15*, *dfra16*, *dfra17*, *dfra19*, *dfra1*, *dfra20*, *dfra21*, *dfra22*, *dfra23*, *dfra24*, *dfra25*, *dfra25*, *dfra25*, *dfra26*, *dfra5*, *dfra7*, *dfrb1*, *dfrb2*, *dfrb3*, *dfrb6*, *emea*, *emrd*, *emre*, *erea*, *ereb*, *erma*, *ermb*, *ermc*, *emrd*, *erme*, *ermf*, *ermg*, *ermh*, *ermn*, *ermo*, *ermq*, *ermr*, *erms*, *ermt*, *ermu*, *ermv*, *ermw*, *ermx*, *ermy*, *fosa*, *fosb*, *fosc*, *fosx*, *fusb*, *fush*, *ksga*, *lmra*, *lmrb*, *lnua*, *lnub*, *lsa*, *maca*, *macb*, *mdte*, *mdtf*, *mdtg*, *mdth*, *mdtk*, *mdtl*, *mdtm*, *mdtn*, *mdto*, *mdtp*, *meca*, *mecr1*, *mefa*, *mepa*, *mexa*, *mexb*, *mexc*, *mexd*, *mexe*, *mexf*, *mexh*, *mexi*, *mexw*, *mexx*, *mexy*, *mfpa*, *mpha*, *mphb*, *mphc*, *msra*, *norm*, *oleb*, *opcm*, *opra*, *oprd*, *oprj*, *oprm*, *oprn*, *otra*, *otrb*, *pbp1a*, *pbp1b*, *pbp2b*, *pbp2*, *pbp2x*, *pmra*, *qac*, *qaca*, *qacb*, *qnra*, *qnrb*, *qnrs*, *rosa*, *rosb*, *smea*, *smeb*, *smec*, *smed*, *smee*, *smef*, *smb*, *sta*, *str*, *sul1*, *sul2*, *sul3*, *tcma*, *tc3*, *tet30*,

tet31, tet32, tet33, tet34, tet36, tet37, tet38, tet39, tet40, teta, tetb, tetc, tetd, tete, tetg, teth, tetj, tetk, tell, tetm, teto, tetpa, tetpb, tet, tetq, tets, tett, tetu, tetv, tetw, text, tety, tetz, tlrc, lmrb, lolc, tsnr, vana, vanb, vanc, vand, vane, vang, vanha, vanhb, vanhd, vanra, vanrb, vanrc, vanrd, vanre, vanrg, vansa, vansb, vansc, vansd, vane, vansg, vant, vante, vantg, vanug, vanwb, vanwg, vanxa, vanxb, vanxd, vanxyc, vanxye, vanxyg, vanya, vanyb, vanyd, vanyg, vanz, vaia, vatb, vatc, vald, vate, vgaa, vgab, vgba, vgbb, vph, ykkc, and ykkd (see the Antibiotic Resistance Genes Database (ARDB) available online).

[0063] In some embodiments, after successful production of the amplicons during the multiplex PCR assay, the resulting amplicons can be further processed to provide sequencing-ready amplicons. For example, some embodiments of the invention may comprise an indexing extension step. In some aspects, the indexing extension step may comprise extending the optimized multiplex amplicons using a set of indexing and common primers that recognize the respective universal tail sequences used for the particular group of amplicons in a minimal cycle PCR assay (e.g., 5-10 total cycles). In particular, each multiplex set of amplicons to be sequenced can be extended with a different set of index oligonucleotides and common oligonucleotides that recognize UT1 and UT2, respectively. In some aspects, the index sequence of the index oligonucleotides can be custom designed to allow for the selection of an index sequence from potentially thousands of different index sequences.

[0064] After this step, the resulting products include a set of amplicons for each sample / template that comprise the same index and any necessary sequences that may be required for a particular sequencing platform (e.g., platform sequences associated with the ILLUMINA® Next Generation sequencing platform). Thereafter, the resulting extension-reaction products can be quantified, pooled, and sequenced using a desired platform. In some aspects, the inclusion of the universal tail sequences on the index and common primers can coincide with the use of genomic and index read primers in the mixture of sequencing primer reagents. For example, some embodiments of the invention are capable of pooling multiple amplicons with multiple indices in a single sequencing run to provide 40,000x-95,000x coverage across the amplicons. In other embodiments, the systems and methods associated with the invention can be configured to provide any level of sequencing coverage that is desirable to the user (e.g., higher or lower than the

coverage levels discussed above). In some embodiments, after sequencing and generation of the sequence data, the resulting data can be demultiplexed and the sequence files can be aligned to the appropriate reference sequences for subsequent sequence analyses.

[0065] Some embodiments of the invention may comprise other applications. For example, some embodiments comprise an application of the universal tail sequences that could be used to detect clinically relevant RNA transcripts in a multiplex fashion. For example, RNA can be extracted from a sample, converted to cDNA using techniques known in the art, and the cDNA can function as the template for additional processes (e.g., multiplex PCR assays and subsequent sequencing). In some aspects, the amplicons resulting from the multiplex PCR reaction can be sequenced, in a manner as previously mentioned, and the resulting sequences can be aligned. As a result, differential numbers of sequence reads generated by the sequencing process (i.e., when aligned to the amplicon reference sequences), can provide data regarding the different copy numbers in the original RNA sample. As a result of this process, clinicians can gain an insight into the actual expression of a gene that is present. Specifically, virulence or resistance markers may display differential expression in a clinical sample. As such, understanding differential expression of these markers could be clinically relevant. Moreover, some embodiments could also be used to characterize population diversity using a relatively small set of markers from a very large number of samples, even within a single sample, which can enhance forensic applications.

[0066] In view of some conventional systems that require the addition of the index via ligation or an analogous method, some embodiments of the invention offer advantages. For example, by decoupling the marker-specific amplification from the addition of the indexes or indices, the marker-specific primers can be shorter and less prone to interactions with other primers, which can facilitate a true multiplex PCR reaction. As such, efficient amplification of multiple targets from very low quality and quantity DNA samples is enabled by this feature, which can be important for clinical and forensic samples. As a result of the use of some embodiments of the invention, clinical laboratories, using accessible sequencing technologies, would be able to rapidly detect the presence/absence of gene sequences within a patient sample that may impact treatment choice and/or clinical outcome, such as bacterial virulence genes or antibiotic resistance genes.

[0067] Embodiments of the invention offer additional advantages relative to conventional systems. For example, some embodiments of the invention comprise the use of PCR before sequencing such that only limited amounts of starting material are necessary and the starting material need not be of high quality (e.g., genomic DNA, crude DNA extracts, single stranded DNA, RNA, cDNA, etc.). In contrast, many conventional sample preparation systems may require relatively large amounts of starting material of relatively high quality, which can limit the use of these systems. Moreover, the inclusion of non-desirable template materials can also interfere in one or more downstream processes in conventional systems and methods. For example, if an investigation is being conducted that focuses on one or more organisms that may be associated with another organism (e.g., bacteria associated with a human), the sampling of the target organism may result in template contamination from the host organism.

[0068] In particular, in some aspects, obtaining samples of pathogenic or commensal bacteria from, on, or within a human may also result in the collection of human tissue. As such, when isolating the template, human nucleic acids may contaminate the bacterial template. Some embodiments of the invention are configured such that the contaminating template (e.g., from a human) would not interfere with downstream processes, including sequencing. For example, some embodiments of the invention operate such that only a limited amount of starting template (e.g., 500 femtograms or greater) can be used. Moreover, some embodiments are also configured such that the starting material (e.g., template contaminated with foreign nucleic acids) can still produce the required amplicons for sequencing in the presence of more than a 1,000-fold excess of contaminating template with no discernible inhibition of the multiplex PCR.

[0069] In certain aspects, the present invention provides an assay that works with as little as about 1pg, about 900fg, about 800fg, about 700fg, about 600fg, about 500fg, about 400fg, about 300fg, about 200fg, or about 100fg of genomic DNA.

[0070] In other aspects, the present invention provides methods that can be used on complex sample types. For example, a 6-plex assay described herein can be used to analyze tuberculosis sputum samples.

[0071] The methods described herein allow for new targets to easily be added due to Universal indexing oligonucleotides. Moreover, in some embodiments

the multiplex reactions are optimized with SYBR qPCR using universal tail gene-specific primers, for example.

[0072] In yet other embodiments, several multiplex reactions from the same sample can be pooled and indexed prior to sequencing. The current invention also allows for flexible scale and read lengths, on multiple platforms.

[0073] In some embodiments, the methods described herein are applied for quantitative analysis including the detection of gene copy number and targeted RNAseq.

[0074] In certain aspects, a synthetic plasmid is used as an internal positive PCR sequencing control.

[0075] The methods and assays of the present invention have various applications. Among these applications is an assay for *Mycobacterium tuberculosis* (Mtb). In one embodiment, this Mtb assay comprises a six-plex antibiotic resistance gene SNP variant sequencing assay.

[0076] The methods and assays described herein can also be used to analyze and sequence DNA from *Burkholderia pseudomallei* with, for example, 7-, 6- and 5-plex species specific amplicon multiplex assays.

[0077] In certain embodiments, the present invention comprises a 16S rRNA and/or an Internal Transcribed Spacer (ITS) amplicon assay. These assays can be used in a variety of clinical sample types and with samples containing the bacterial pathogen, *Leptospira*.

[0078] In some embodiments, the present invention comprises a major histocompatibility complex (MHC) amplicon assay. In one embodiment, the MHC amplicon assay is used with samples from Prairie Dogs.

[0079] In other aspects, the methods and assays of the present invention comprise species specific amplicon assays in several clinically important pathogens and/or antibiotic resistance gene multiplex assays covering a variety of important genes.

EXAMPLES

[0080] Herein below are set forth various embodiments of the present invention. It is anticipated that suitable modifications can be made thereto which will nonetheless remain within the scope of the invention. The invention shall therefore

only be construed in accordance with the claims below and not any particular embodiment or example.

Example 1. Universal Multiplex Indexing Strategy for ILLUMINA® Next Generation Amplicon Sequencing Process

[0081] Referring to **FIGS. 1** and **2**, embodiments of the invention can be used as a universal multiplex indexing strategy for the ILLUMINA® Next Generation amplicon sequencing process. As used in this example, a multiplex PCR is performed with a total of five amplicons per reaction contemplated. In particular, each amplicon is considered to be associated with a different marker (e.g., antibiotic resistance genes). As such, to amplify each marker, the PCR reaction mixture is provided with five unique forward primers and five unique reverse primers (i.e., each primer designed to bind to a marker-specific sequence). Moreover, as shown in **FIG. 1**, the forward primers comprise a first universal tail sequence (UT1) at the 5' end of the respective primers, with the universal tail sequence being identical for each of the forward primers in the reaction mixture. Similarly, the reverse primers comprise a second universal tail sequence (UT2) at the 5' end of the respective primers, with the second universal tail sequences being identical for each of the reverse primers in the mixture, but different than the sequence of UT1. The sequences associated with UT1 and UT2 (and other potential sequences for universal tails, UT3 and UT4) are recited in **FIG. 2**. Moreover, as illustrated in **FIG. 3**, the multiplex PCR system using the universal tail sequences and marker-specific primers produces high-quality amplicons when the amount of input template is as low as 500 femtograms. Even this de minimis amount of starting material provides a sufficient concentration of amplicon for downstream processes.

[0082] After completion of the initial multiplex PCR, the assay mixture comprises five amplicons that include the desired marker-specific amplicon with UT1 and UT2 integrated into the ends of the amplicons, as illustrated in **FIG. 1**. Thereafter, the five resulting amplicons are extended using platform-specific primers that recognize UT1 and UT2 for adding the indexes to each amplicon. In particular, as illustrated in **FIG. 2**, index-extension primers were designed that recognize at least one of UT1 and UT2. The index-extension primers also include Next Generation sequencing-specific elements, including the index and the read primer sequences that are used during the sequencing process. This extension process

occurs during a simplified PCR reaction that uses an abbreviated cycle process, such as 5-10 total cycles of amplification. In particular, **FIG. 2** also includes a list of potential index – extension primers that have been designed to recognize at least one of the universal tail sequences.

[0083] After completion of the extension process, the resulting modified amplicons are pooled and sequenced. For example, as shown in **FIG. 1**, samples from multiple multiplex PCR reactions can be pooled together for a single sequencing event. In particular, each multiplex PCR assay mixture may contain the template from a single organism such that each of the amplicons provides data regarding that particular organism. In addition, the same multiplex PCR reaction can be performed on multiple templates from multiple samples in multiple reaction vessels at the same time such that amplicons of the same five markers are amplified from each respective template. However, in order to differentiate between the sources of the amplicons upon pooling, during the extension step, unique indexes can be integrated into amplicons that originate from different organisms. For example, if ten different clinical isolates of bacteria are being tested with the amplification of five markers for each of the ten isolates, then a different index is integrated into the five amplicons that originate from a given isolate (i.e., each of the amplicons originating from the isolate will have the same index, which will differ from the index associated with the remaining nine isolates). As a result, the samples can be pooled for sequencing during a single sequencing run and distinguished based on the index sequence during analysis of the data. In particular, after sequencing, the resulting data can be demultiplexed and aligned to a reference sequence (e.g., a wild type sequence and/or other alleles for each of the respective markers). As a result, the aligned sequences can be analyzed for the presence or absence of markers, variant signatures associated with the markers, differential marker presence in the sample, which includes the capability of analyzing gene expression, and an estimate of allele frequencies of various alleles of the markers in the pooled samples.

Example 2. Universal Tail Amplicon Sequencing Assay for Use with Samples Containing *Mycobacterium tuberculosis*

[0084] The following protocol was used to prepare and sequence amplicons from samples containing *Mycobacterium tuberculosis*.

Target amplification

[0085] Primer sets targeting *Mycobacterium tuberculosis* gene regions were used (see **Table 1**). The target amplicons were between 180-400 bp long. The size of the amplicons was dependent on the positions of the SNPs and the constraints of the length of the sequencing read. Along with organism specific sequence, each primer has a universal tail sequence (bold sequences in **Table 1**). All forward primers have one universal tail sequence while all the reverse primers have a second universal tail sequence. In Gene Specific multiplex PCR reactions, all target amplicons are synthesized with the universal tail sequence added to the amplicons. The PCR parameters are as follows: initial denaturation at 98°C for 1 min, twenty-five cycles of denaturation at 98°C for 10 sec, annealing at 60°C for 15 sec, and extension at 72°C for 20 sec and a final extension at 72°C for 2 min. A single 25 uL PCR reaction containing 2 uL of DNA, 12.5 uL of Q5[®] Hot Start High-Fidelity 2X Master Mix (New England Biolabs[®] Inc.), 4 uL of a primer mix, 1.5 uL of molecular grade H₂O, and 5uL of 5M Betaine solution (Sigma-Aldrich[®]). After PCR the reaction is cleaned up using a 1X Agencourt[®] AMPure[®] XP bead (Beckman Coulter) clean up with elution in 25 uL of a 10mM Tris-HCl 0.05% Tween 20 solution.

Index Extension – Library Creation

[0086] A second PCR adds a specific index sequence, based on the Kozarewa and Turner 8 bp indexing scheme (see Kozarewa I, Turner D. 96-Plex Molecular Barcoding for the Illumina Genome Analyzer. In: Kwon YM, Ricke SC, editors. High-Throughput Next Generation Sequencing: Humana Press; 2011. p. 279-98.), to the amplicons using the universal tail sequences on either end of the amplicon. At the end of the Index Extension PCR there is a sequencer ready amplicon library. The PCR parameters are as follows: initial denaturation at 98°C for 2 min, six cycles of denaturation at 98°C for 30 sec, annealing at 60°C for 20 sec, and extension at 72°C for 30 sec and a final extension at 72°C for 5 min. A single

50uL PCR reaction containing 2 uL of template, 25uL of 2x KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 2 uL of the 10 uM common universal tail primer, 2uL of 10uM specific index universal tail primer, 9 uL of molecular grade H₂O, and 10uL of 5M Betaine solution (Sigma-Aldrich[®]). After PCR the reaction was cleaned using a 0.8x Agencourt[®] AMPure[®] XP bead (Beckman Coulter) clean up with elution in 40 uL of a 10 mM Tris-HCl 0.05% Tween 20 solution to remove primer dimers. All completed libraries were run on the Bioanalyzer[®] 2100 (Agilent Technologies[®]) for confirmation of target amplification.

Pooling and Sequencing

[0087] By adding sample specific index sequences to the amplicons, pools of several samples are made for sequencing. Each individual library is quantified using KAPA Library Quantification Kit - Illumina/ABI Prism[®] (KAPA Biosystems) qPCR, and pooled in equal molar concentrations. At least 25% of each sequencing run was filled with whole genome or PhiX control samples to ensure base diversity and reduce complications with sequencing. For the validation a single sequencing pool was sequenced on the Illumina[®] MiSeq platform using 2x300bp version 3 sequencing chemistry (Illumina[®]). If PhiX control was not used for base diversity (e.g., at 25% of a sequencing run), then it was spiked in a low concentrations (i.e., 1-5%) in each run for error rate examination. Novel Read 1, Read 2, and indexing sequencing primers were used for sequencing (see SEQ ID NOs: 6, 7, and 8 in **Figure 2**). A representative chromatogram depicting an analysis of a multiplex amplification reaction after amplicon yield optimization is shown in **Figure 4**.

Table 1. *M. tuberculosis* specific primers with universal tail sequences. All oligos are with standard de-salting. The universal tail sequences are highlighted in bold with the forward primer sequence differing from the reverse primer sequence.

	SEQ ID	
Forward Primer	NO:	Sequence
gyrAv2fUT1	45	ACCCA ACTGAATGGAGCGGGTGCTCTATGCAATGTTTCGAT
eisv2fUT1	46	ACCCA ACTGAATGGAGCCGTCACCGCAGATCCATGTAC
rpoBv2fUT1	47	ACCCA ACTGAATGGAGCCGATCACACCGCAGACGTT
katGv2fUT1	48	ACCCA ACTGAATGGAGCCCATGAACGACGTCGAAACAG
inhAv2fUT1	49	ACCCA ACTGAATGGAGCCCTCGCTGCCAGAAAGG
rrsv2fUT1	50	ACCCA ACTGAATGGAGCCTAGTAATCGCAGATCAGCAACG
Reverse Primer	NO:	Sequence
gyrAv2rUT2	51	ACGCA CTTGACTTGTCTTCGGGCTTCGGTGTACCTCATC
eisv2rUT2	52	ACGCA CTTGACTTGTCTTCGGTCGCTGATTCTCGCAGTG
rpoBv2rUT2	53	ACGCA CTTGACTTGTCTTCGTTTCGATCGGGCACATCC
katGv2rUT2	54	ACGCA CTTGACTTGTCTTCGCTCTTCGTCAGCTCCCCTC
inhAv2rUT2	55	ACGCA CTTGACTTGTCTTCGTCACATTTCGACGCCAAACAG
rrsv2rUT2	56	ACGCA CTTGACTTGTCTTCGCCTACGCCACCAGTT

Example 3. Universal Tail Amplicon Sequencing Assay for Use with Samples Containing *Burkholderia pseudomallei* and *Burkholderia mallei*

[0088] The protocol outlined in Example 2 was used to prepare and sequence amplicons from samples containing *Burkholderia pseudomallei* and *Burkholderia mallei*.

[0089] The primers specific to *Burkholderia pseudomallei* and *Burkholderia mallei* are outlined in **Table 2**.

[0090] **Figure 5** shows consistent PCR amplification and Illumina[®] index extension (peaks on chromatogram) for 12 diverse *Burkholderia pseudomallei* and *Burkholderia mallei* isolate DNA samples, using the oligos listed in **Table 2** in a multiplex fashion.

[0091] Twelve genomic DNA samples were amplified using the Bp/Bm UT multiplex amplicon assay. Arrows in the Bp_406e panel indicate peaks corresponding to each amplicon. Optimal amplification occurred in all 12 samples

Sizing standards failed during the run, but relative peak sizes and migration rates indicate successful amplification in all samples. Sequencing resulted in ~4 fold variance in read counts among all amplicons and recovery of correct sequence.

Table 2. *Burkholderia pseudomallei* and *Burkholderia mallei* specific primers with universal tail sequences.

UT Forward Oligo Name	SEQ ID NO:	UT Forward Oligo Sequence
UT2-BpCEN322640-f2	57	ACGCACTTGACTTGTCTTCCGCGGACAGCATCGATTACGTGAATC
UT1-BpCEN1761486-f1	58	ACCCA ACTGAATGGAGCGACCTGCAGCAGGTATTCGACATTATCGTTC
UT2-BpCEN1722622-f1	59	ACGCACTTGACTTGTCTTCCAACGGGCGAGTTTGCAACGGAATC
UT1-BpCEN1565214-f1	60	ACCCA ACTGAATGGAGCCTGACCGAACGATGGCTGGAGATACATGC
UT2-BpCEN894337-f1	61	ACGCACTTGACTTGTCTTCCGAAAATAATTTTCGGCCGGCGCAC
UT2-BpCEN1235988-f1	62	ACGCACTTGACTTGTCTTCGCGCTGCCCGTTTCACCACTGG
UT Reverse Oligo Name		UT Reverse Oligo Sequence
UT1-BpCEN32264-r2	63	ACCCA ACTGAATGGAGCCCGCCGAATCCGATGCTCAATTC
UT2-BpCEN1761486-r1	64	ACGCACTTGACTTGTCTTCAGCTTCGCATACAGCACTTCCGCCAG
UT1-BpCEN1722622-r1	65	ACCCA ACTGAATGGAGCGCCGGCTTGGCTTCGTCTTGTC
UT2-BpCEN1565214-r1	66	ACGCACTTGACTTGTCTTCCAAATGGGAAGCGAGCTCCCTTCCGA
UT1-BpCEN894337-r1	67	ACCCA ACTGAATGGAGCCGACAGGCATCGGGCGACTACTACCAG
UT1-BpCEN1235988-r1	68	ACCCA ACTGAATGGAGCCGTGACGCCGTCGGGAAAGATCATC

Example 4. Internal Positive PCR and Sequencing Control for Universal Tail Amplicon Sequencing Assay

[0092] The schematic in **Figure 7** depicts a plasmid construct containing a synthetic 1000 bp sequence (shown in black) around which two different amplicons have been designed (IPPSC-1, IPPSC-2.) This sequence displays no significant BLAST similarity to any sequences in the NCBI nucleotide database. This synthetic plasmid DNA template can be used in conjunction with the Universal Tail target specific oligo sequences shown in **Table 3** below to act as an internal positive PCR and sequencing control, to indicate both successful PCR amplification and subsequent amplicon sequencing in any Universal Tail Amplicon Sequencing Assay.

Table 3. Primers for internal positive PCR and sequencing control.

UT Forward Oligo Name	SEQ ID NO:	UT Forward Oligo Sequence
UT1-IPSC-f1	69	ACCCAAGTGAATGGAGCGGGCGGACGAAAACCCTTGAGCACAG
UT1-IPSC-f2	70	ACCCAAGTGAATGGAGCGCGGACGCGGTTGAGGCAAAGTGATAC
UT Reverse Oligo Name		UT Reverse Oligo Sequence
UT2-IPSC-r1	71	ACGCACTTGACTTGTCTTCGCCGGGATGCCTTACCTAGACGCAATGA
UT2-IPSC-r2	72	ACGCACTTGACTTGTCTTCGAGTTCGTCGTTAAGCGTGACAGTC

Example 5. Quantitative Analysis with the Universal Tail Amplicon Sequencing Assay

[0093] The Universal Tail Amplicon Sequencing Assay can also be used to perform quantitative analysis to determine gene copy number, for example. In addition, the assay can be used to perform targeted RNA-Seq.

[0094] To demonstrate the assay's potential for quantitative analysis a quantitative DNA sequencing (qDNA-Seq) experiment was performed with a mock variable copy number template. The results in **Figure 6A** demonstrate the assay's capability to detect variations in gene copy number.

[0095] Briefly, individual Mtb amplicon targets that comprise the Mtb multiplex (see **Table 1**) were mixed together at different concentrations across four orders of magnitude and then used as template for the Mtb multiplex Universal Tail Amplicon Sequencing Assay. Gene specific qPCR on the artificial mixture pre-(x axis) and post-(left y axis) amplification shows strong correlation ($R^2=0.99$),

indicating the relative variable concentrations of each target is maintained after the multiplex amplification. The multiplex amplification was then sequenced on a Illumina[®] MiSeq platform. The right y axis shows the sequence read counts that aligned to each of the target sequences, \log_2 transformed, compared to the pre-amplified starting template, indicating that the relative sequence read counts also strongly correlate ($R^2=0.98$) to the relative concentration of the starting template target concentrations.

[0096] The amplification comparison of transcriptome RNA-Seq analysis with UT amplicon sequencing RNA-Seq analysis using samples containing *Burkholderia pseudomallei* confirmed the utility of the Universal Tail Amplicon Sequencing Assay in targeted RNA-Seq (see **Figure 6B**). Briefly, the *Burkholderia* multiplex (see **Table 2**) was used in a multiplex reverse transcriptase reaction to generate amplicon sequence from *Burkholderia* RNA samples from a previous RNAseq study. Normalized read counts from the targets from the original study were compared to the UT amplicon sequence data. While not as highly correlated as the mock mixture experiment (**Figure 6A**), the relative read counts were still highly correlated ($R^2=0.89$). These studies indicate the potential for quantitative analysis using the Universal Tail Amplicon Sequencing methodology.

[0097] It should be understood from the foregoing that, while particular embodiments have been illustrated and described, various modifications can be made thereto without departing from the spirit and scope of the invention as will be apparent to those skilled in the art. Such changes and modifications are within the scope and teachings of this invention as defined in the claims appended hereto.

[0098] Unless defined otherwise, all technical and scientific terms herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials, similar or equivalent to those described herein, can be used in the practice or testing of the present invention, the preferred methods and materials are described herein. All publications, patents, and patent publications cited are incorporated by reference herein in their entirety for all purposes.

[0099] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

CLAIMS

What is claimed is:

1. A method for preparing a sample for sequencing, the method comprising the steps of:
 - a) receiving a sample; and
 - b) amplifying at least a first marker within the sample, wherein amplifying at least the first marker comprises mixing the sample with a first oligonucleotide comprising a first universal tail sequence and a second oligonucleotide comprising a second universal tail sequence, wherein the first universal tail sequence and the second oligonucleotide sequence are different.
2. The method of claim 1, wherein the first universal tail sequence and the second universal tail sequence independently comprise a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, and SEQ ID NO: 4.
3. The method of claim 1 or 2, further comprising:
 - c) amplifying a second marker by mixing the sample with a third oligonucleotide comprising the first universal tail sequence and a fourth oligonucleotide comprising the second universal tail sequence.
4. The method of claim 3, wherein the steps of amplifying the first marker and the second marker occur in a multiplex polymerase chain reaction.
5. The method of claim 3, wherein amplifying the first marker and the second marker result in a first amplicon and a second amplicon.
6. The method of claim 5, further comprising adding an index to the first amplicon and the second amplicon using at least one indexing oligonucleotide, wherein the at least one indexing oligonucleotide comprises a

complementary sequence that recognizes at least one of the first universal tail sequence and the second universal tail sequence.

7. The method of claim 6, wherein the at least one indexing oligonucleotide comprises a sequence selected from the group consisting of SEQ ID NOs: 9-44.
8. The method according to any one of the preceding claims, wherein the sample comprises a pathogenic organism.
9. The method of claim 8, wherein the pathogenic organism is selected from the group consisting of *Mycobacterium tuberculosis*, *Burkholderia pseudomallei*, *Burkholderia mallei*, and *Leptospira* sp.
10. The method according to any one of the preceding claims, wherein the first marker comprises an antibiotic resistance gene.
11. The method according to any one of the preceding claims, wherein the sample comprises castor beans or derivatives thereof.
12. The method according to any one of the preceding claims, wherein the sample comprises tissue from an animal.
13. The method of claim 12, wherein the animal is a human.
14. The method according to any one of the preceding claims, further comprising sequencing the amplified first marker and/or second marker.
15. The method of claim 14, wherein the sequencing comprises amplifying the first marker and/or second marker with a sequencing primer.
16. The method of claim 15, wherein the sequencing primer comprises a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

17. A method of assessing the presence of at least one antibiotic resistance gene within a population of organisms, the method comprising the steps of:
- a) receiving a sample of the population of organisms;
 - b) extracting a template from the sample;
 - c) performing a multiplex polymerase chain reaction assay comprising the steps of:
 - i) amplifying a first marker from the template, wherein amplifying the first marker comprises mixing the template with a first oligonucleotide comprising a first universal tail sequence and a second oligonucleotide comprising a second universal tail sequence, wherein the first universal tail sequence and the second oligonucleotide sequence are different;
 - ii) amplifying a second marker by mixing the template with a third oligonucleotide comprising the first universal tail sequence and a fourth oligonucleotide comprising the second universal tail sequence;wherein amplification of the first marker and the second marker result in the production of a first amplicon and a second amplicon.
18. The method of claim 17, wherein the first universal tail sequence and the second universal tail sequence independently comprise a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, and SEQ ID NO: 4.
19. The method of claim 17 or 18, wherein the first marker and the second marker are antibiotic resistance genes.
20. The method according to any one of claims 17 to 19, further comprising adding an index to the first amplicon and the second amplicon using at least one indexing oligonucleotide, wherein the at least one indexing

oligonucleotide comprises a complementary sequence that recognizes at least one of the first universal tail sequence and the second universal tail sequence.

21. The method of claim 20, wherein the at least one indexing oligonucleotide comprises a sequence selected from the group consisting of SEQ ID NOs: 9-44.
22. The method according to any one of claims 17 to 21, wherein the population of organisms comprises a pathogenic organism.
23. The method of claim 22, wherein the pathogenic organism comprises a bacterium.
24. The method of claim 23, wherein the bacterium is selected from the group consisting of *Mycobacterium tuberculosis*, *Burkholderia pseudomallei*, *Burkholderia mallei*, and *Leptospira* sp.
25. The method according to any one of claims 17 to 24, further comprising sequencing the first amplicon and the second amplicon.
26. The method of claim 25, wherein the sequencing comprises amplifying the first amplicon and the second amplicon with a sequencing primer.
27. The method of claim 26, wherein the sequencing primer comprises a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.
28. The method according to any one of claims 17 to 27, wherein the template is DNA.
29. The method of claim 28, wherein the DNA is single-stranded DNA.

30. The method according to any one of claims 17 to 27, wherein the template is RNA.
31. A method of determining gene copy number and/or quantifying gene expression in a sample, the method comprising the steps of:
- a) receiving a sample;
 - b) amplifying at least a first marker within the sample to produce a first amplicon, wherein amplifying at least the first marker comprises mixing the sample with a first oligonucleotide comprising a first universal tail sequence and a second oligonucleotide comprising a second universal tail sequence, wherein the first universal tail sequence and the second oligonucleotide sequence are different;
 - c) adding an index to the first amplicon using at least one indexing oligonucleotide, wherein the at least one indexing oligonucleotide comprises a complementary sequence that recognizes at least one of the first universal tail sequence and the second universal tail sequence;
 - d) sequencing the first amplicon to produce a number of sequencing reads; and
 - e) determining gene copy number and/or quantifying gene expression from differential target sequencing read counts, wherein the sequencing read counts of the sample are compared with sequencing read counts of a reference sample.
32. The method of claim 31, wherein the reference sample is from a wild-type organism or a healthy subject.
33. The method of claim 31, wherein the reference sample is an internal standard comprising the first marker.

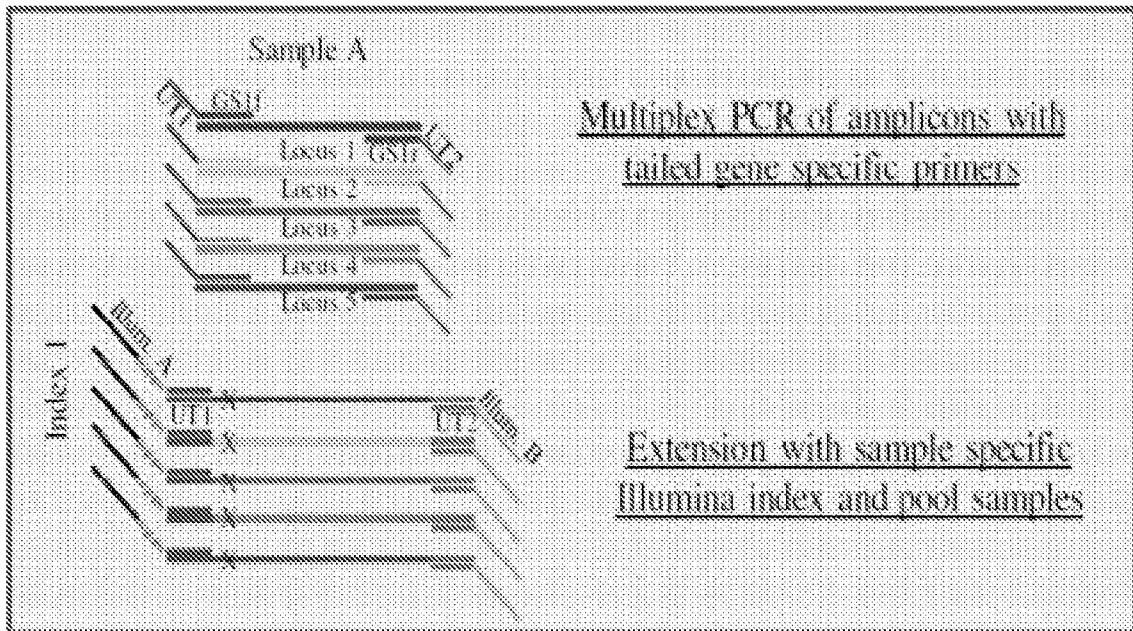


FIGURE 1A

2/15

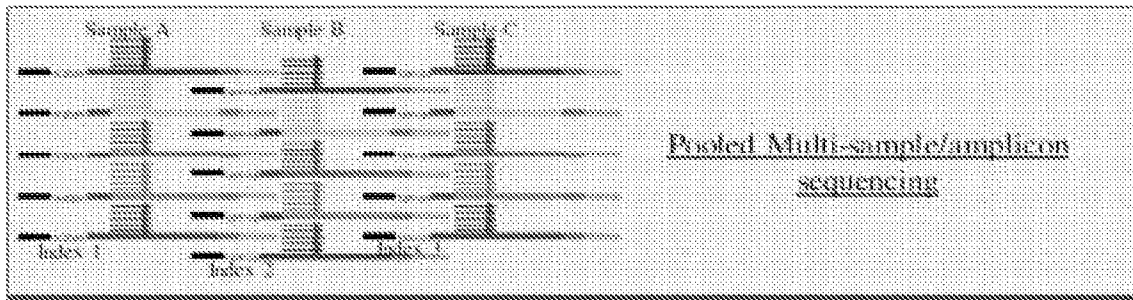


FIGURE 1B

3/15

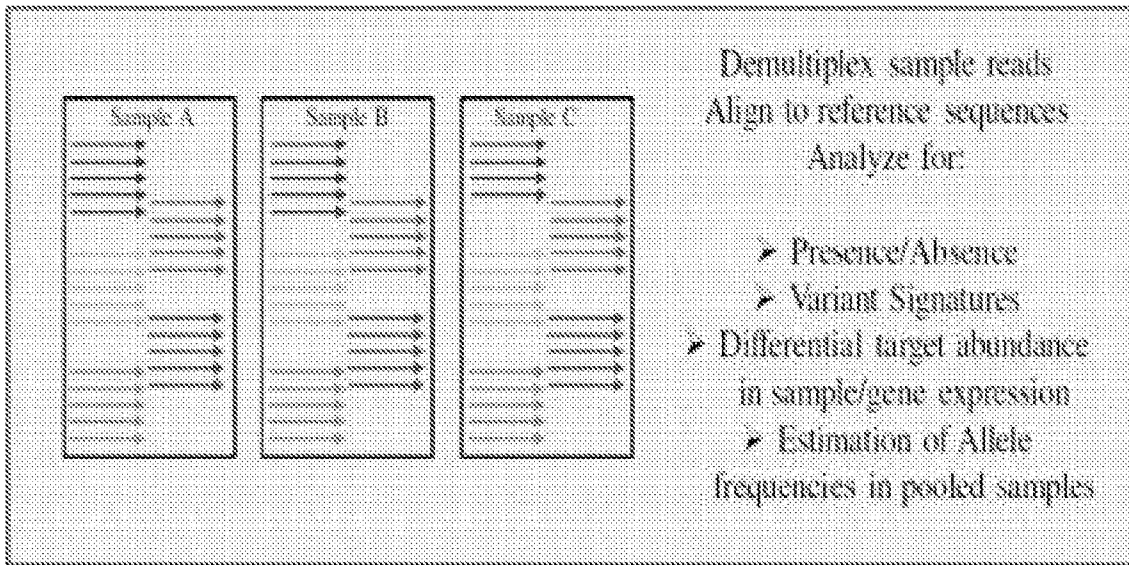


FIGURE 1C

4/15

<u>Name</u>	<u>Sequence (5'-3')</u>
UT1	ACCCA ACTGAATGGAGC (SEQ ID NO: 1)
UT2	ACGCA CTTGACTTGTCTTC (SEQ ID NO: 2)
UT3	ATCGACTGTGTTAGGTCAC (SEQ ID NO: 3)
UT4	CTGTCCTTACCTCAATCTC (SEQ ID NO: 4)

Common UT1 oligonucleotide
AATGATACGGCGACCACCGAGATCTACACTATGGT
AATTGT**ACCCA**ACTGAATGGAGC (SEQ ID NO: 5)

UT1 Read 1 seq primer
CCGAGATCTACACTATGGTAATTGT**ACCCA**ACTGA
ATGGAGC (SEQ ID NO: 6)

UT2 index read seq primer
GAAGACAAGTCAAGTCGTGGCTGACTGACT (SEQ ID NO: 7)

UT2 read 2 seq primer
AGTCAGTCAGCC**ACGCA**CTTGACTTGTCTTC (SEQ ID NO: 8)

FIGURE 2

5/15

Universal tail index oligo sequences

Indexed -UT2-1 (SEQ ID NO: 9)

CAAGCAGAAGACGGCATAACGAGATAACGCTAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-2 (SEQ ID NO: 10)

CAAGCAGAAGACGGCATAACGAGATAAACATCGAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-3 (SEQ ID NO: 11)

CAAGCAGAAGACGGCATAACGAGATAACATTGGCAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-4 (SEQ ID NO: 12)

CAAGCAGAAGACGGCATAACGAGATAACCACTGTAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-5 (SEQ ID NO: 13)

CAAGCAGAAGACGGCATAACGAGATAACGTGATAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-6 (SEQ ID NO: 14)

CAAGCAGAAGACGGCATAACGAGATCGCTGATCAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-7 (SEQ ID NO: 15)

CAAGCAGAAGACGGCATAACGAGATCAGATCTGAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-8 (SEQ ID NO: 16)

CAAGCAGAAGACGGCATAACGAGATATGCCTAAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-10 (SEQ ID NO: 17)

CAAGCAGAAGACGGCATAACGAGATAGTACAAGAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-11 (SEQ ID NO: 18)

CAAGCAGAAGACGGCATAACGAGATCATCAAGTAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-12 (SEQ ID NO: 19)

CAAGCAGAAGACGGCATAACGAGATAGTGGTCAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-13 (SEQ ID NO: 20)

CAAGCAGAAGACGGCATAACGAGATAACAACCAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-37 (SEQ ID NO: 21)

CAAGCAGAAGACGGCATAACGAGATCCGAAGTAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-38 (SEQ ID NO: 22)

CAAGCAGAAGACGGCATAACGAGATCCGTGAGAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

FIGURE 2 (Continued)

6/15

Universal tail index oligo sequences

Indexed -UT2-39 (SEQ ID NO: 23)

CAAGCAGAAGACGGCATAACGAGATCCTCCTGAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-40 (SEQ ID NO: 24)

CAAGCAGAAGACGGCATAACGAGATCGAACTTAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-41 (SEQ ID NO: 25)

CAAGCAGAAGACGGCATAACGAGATCGACTGGAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-42 (SEQ ID NO: 26)

CAAGCAGAAGACGGCATAACGAGATCGCATACAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-43 (SEQ ID NO: 27)

CAAGCAGAAGACGGCATAACGAGATCTCAATGAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-44 (SEQ ID NO: 28)

CAAGCAGAAGACGGCATAACGAGATCTGAGCCAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-45 (SEQ ID NO: 29)

CAAGCAGAAGACGGCATAACGAGATCTGGCATAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-46 (SEQ ID NO: 30)

CAAGCAGAAGACGGCATAACGAGATGAATCTGAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-47 (SEQ ID NO: 31)

CAAGCAGAAGACGGCATAACGAGATGACTAGTAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-48 (SEQ ID NO: 32)

CAAGCAGAAGACGGCATAACGAGATGAGCTGAAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-49 (SEQ ID NO: 33)

CAAGCAGAAGACGGCATAACGAGATGATAGACAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-50 (SEQ ID NO: 34)

CAAGCAGAAGACGGCATAACGAGATGCCACATAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-51 (SEQ ID NO: 35)

CAAGCAGAAGACGGCATAACGAGATGCGAGTAAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-52 (SEQ ID NO: 36)

CAAGCAGAAGACGGCATAACGAGATGCTAACGAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

FIGURE 2 (Continued)

7/15

Universal tail index oligo sequences

Indexed -UT2-53 (SEQ ID NO: 37)

CAAGCAGAAGACGGCATACGAGATGCTCGGTAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-54 (SEQ ID NO: 38)

CAAGCAGAAGACGGCATACGAGATGGAGAACAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-55 (SEQ ID NO: 39)

CAAGCAGAAGACGGCATACGAGATGGTGCGAAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-56 (SEQ ID NO: 40)

CAAGCAGAAGACGGCATACGAGATGTACGCAAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-57 (SEQ ID NO: 41)

CAAGCAGAAGACGGCATACGAGATGTCGTAGAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-58 (SEQ ID NO: 42)

CAAGCAGAAGACGGCATACGAGATGTCTGTCAAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-97 (SEQ ID NO: 43)

CAAGCAGAAGACGGCATACGAGATACTTGATGAGTCAGTCAGCCACGCACTTGACTTGTCTTC

Indexed -UT2-98 (SEQ ID NO: 44)

CAAGCAGAAGACGGCATACGAGATTGACAGACAGTCAGTCAGCCACGCACTTGACTTGTCTTC

FIGURE 2 (Continued)

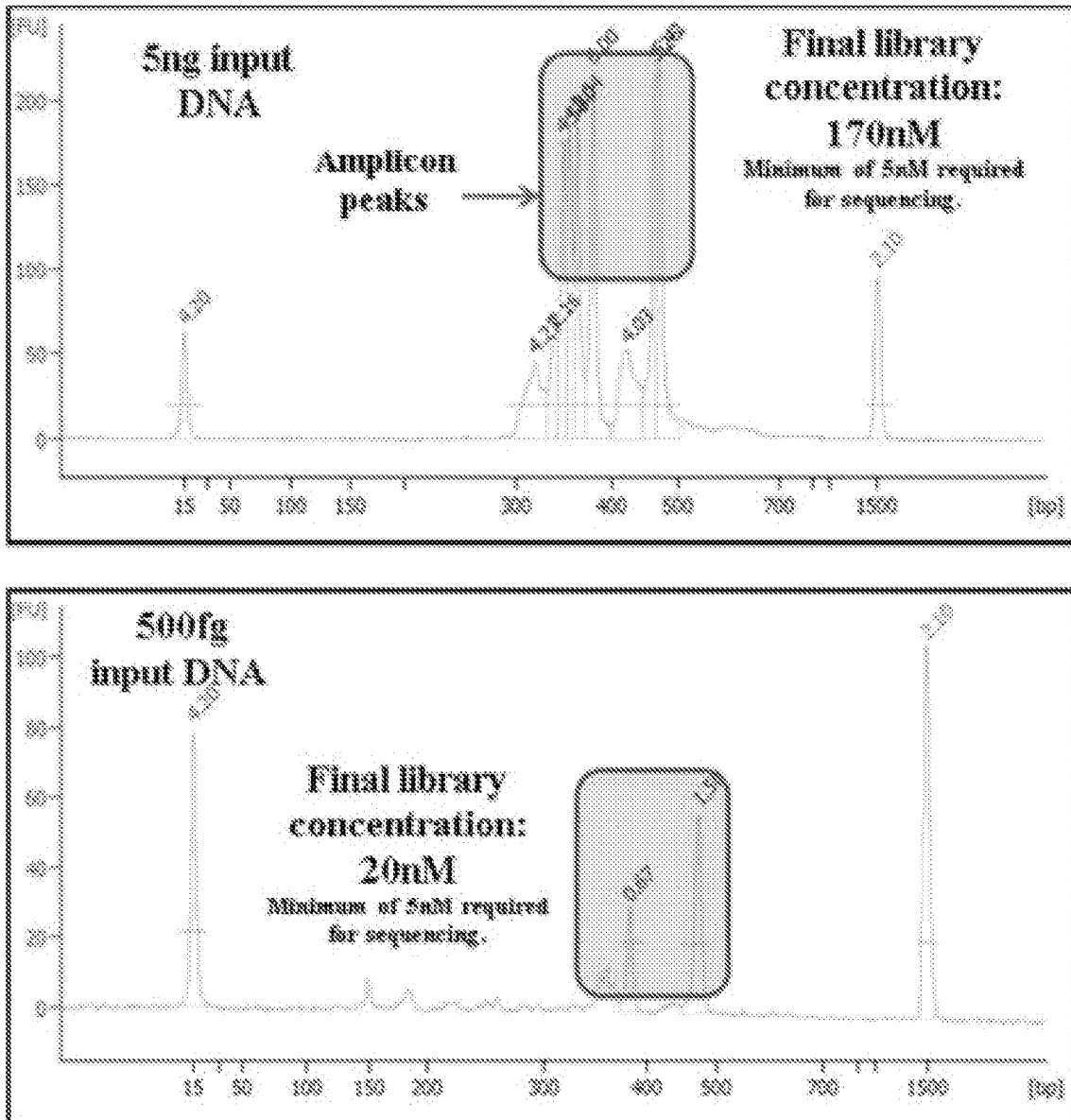
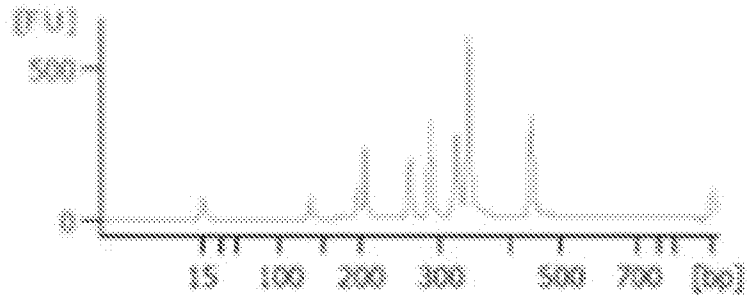


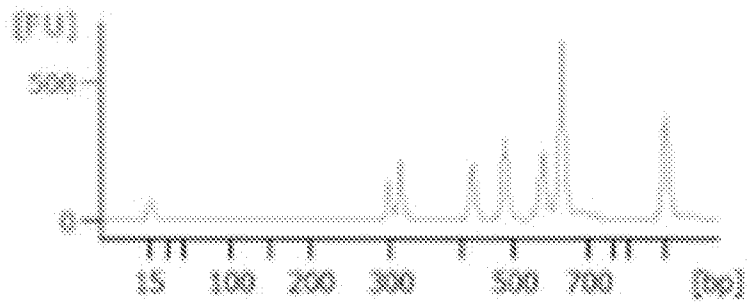
FIGURE 3

10/15

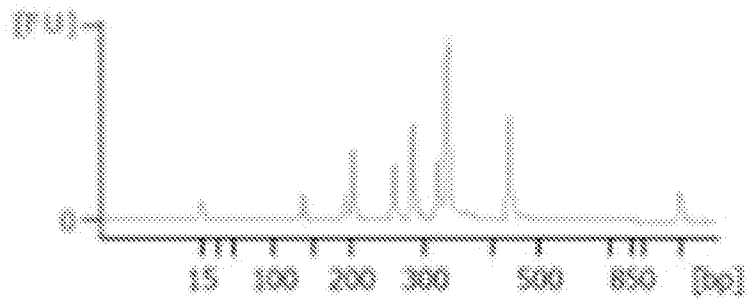
Bp_MSHR668



Bp_K96243



Bp_MSHR5848



Bp_MSHR5855

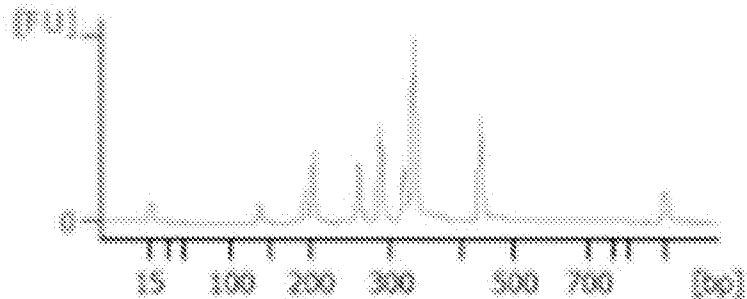


FIGURE 5

11/15

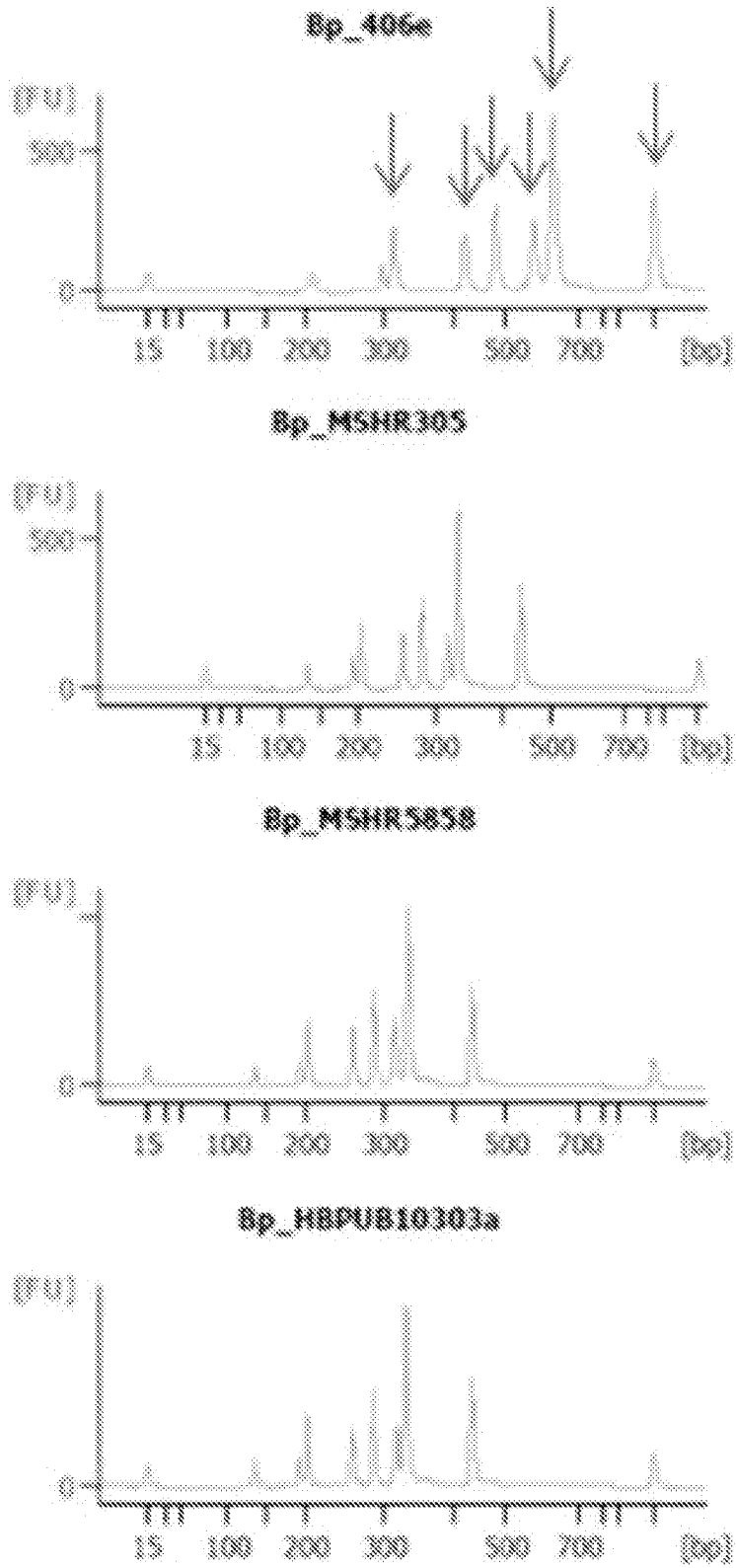
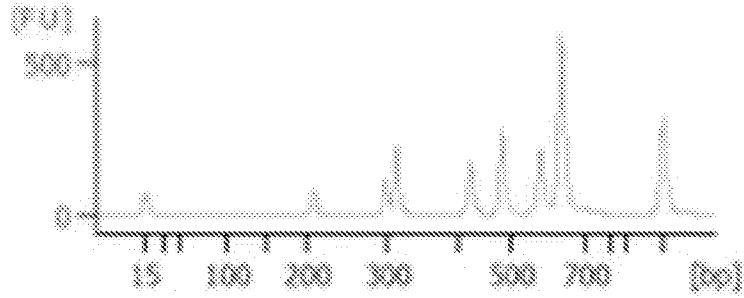


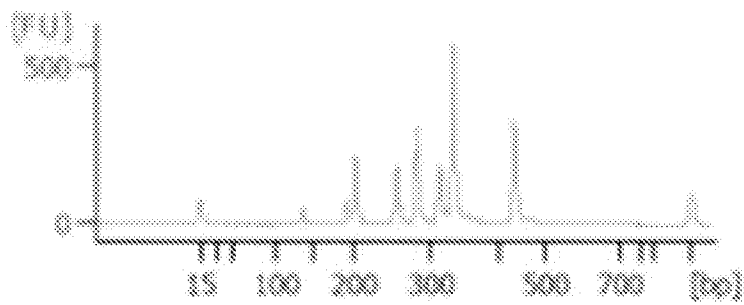
FIGURE 5 (Continued)

12/15

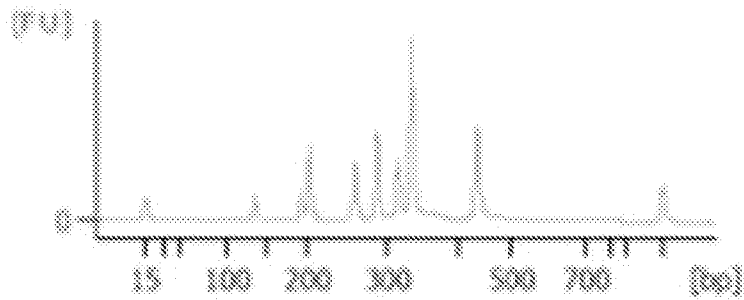
Bp_1106a



Bp_MSHR840



Bp_HBPUB10134a



Bm_2000031281

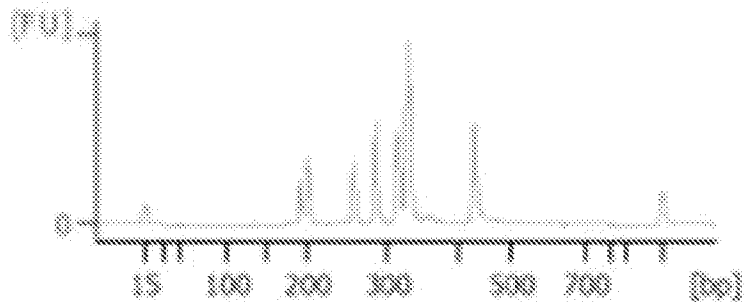


FIGURE 5 (Continued)

qDNA-Seq on mock variable copy number template

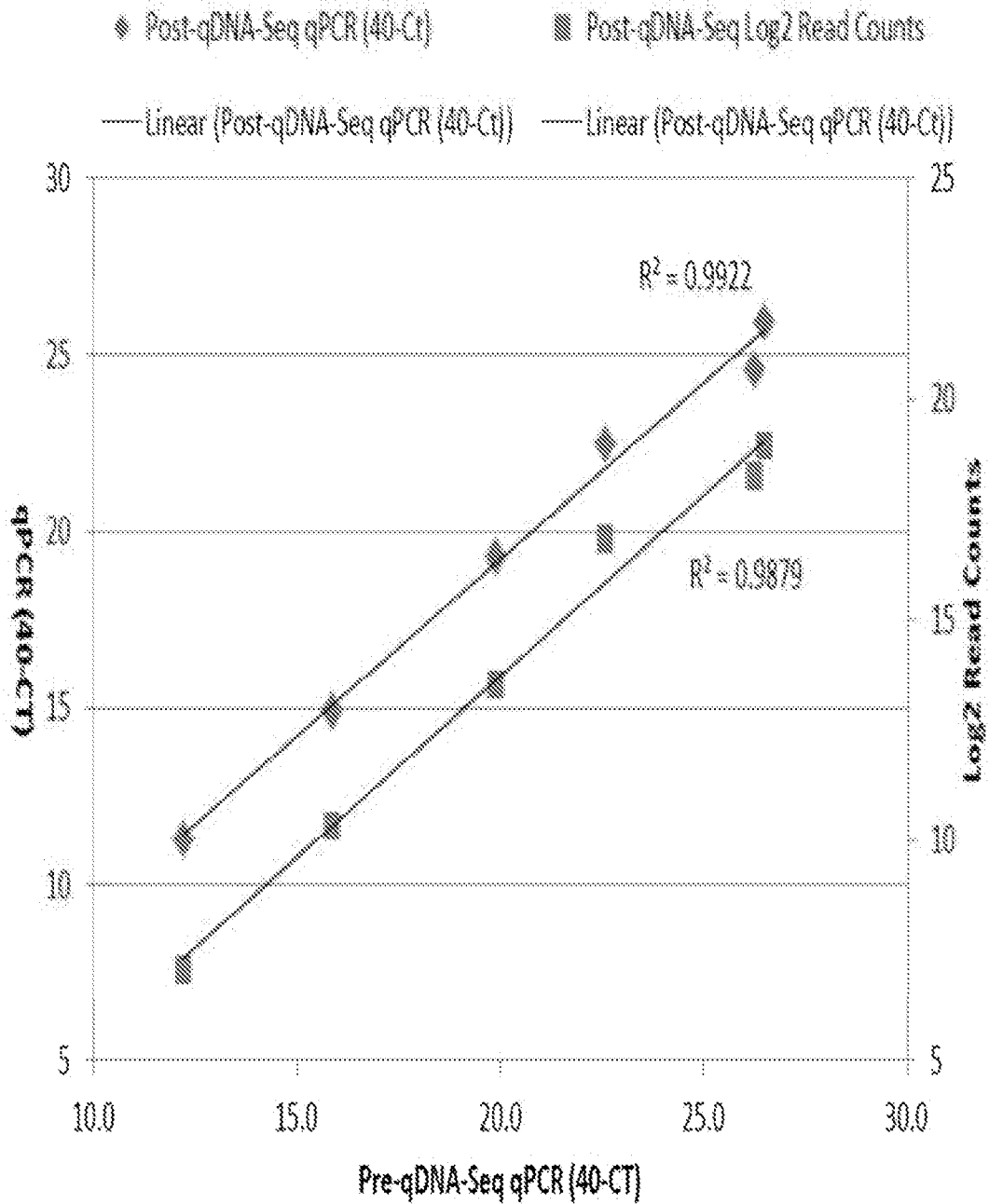


FIGURE 6A

14/15

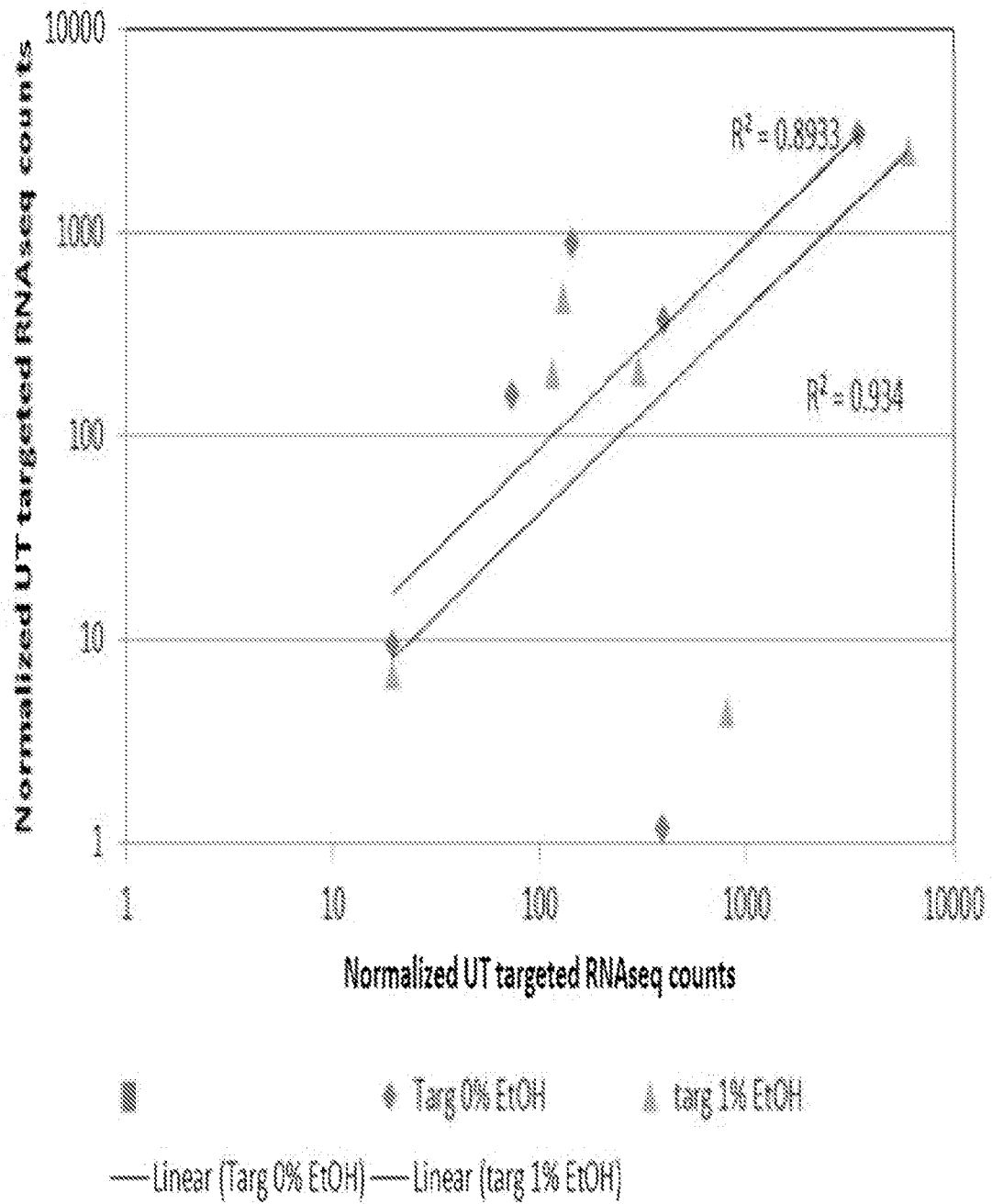


FIGURE 6B

15/15

**Internal Positive PCR and Sequencing Control for Universal Tail
Amplicon Sequencing**

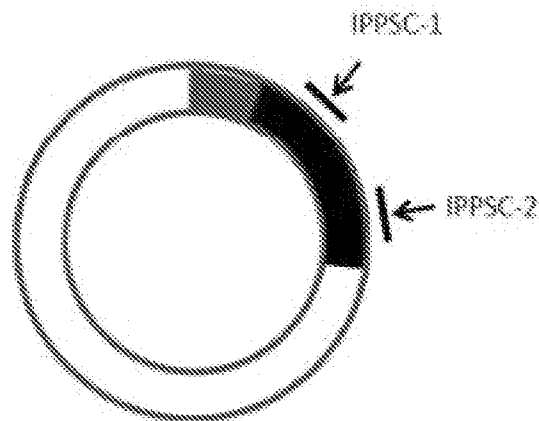


FIGURE 7