

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 January 2011 (27.01.2011)

PCT

(10) International Publication Number
WO 2011/011431 A1

- (51) **International Patent Classification:**
C12P 19/34 (2006.01) C12Q 1/68 (2006.01)
- (21) **International Application Number:**
PCT/US2010/042630
- (22) **International Filing Date:**
20 July 2010 (20.07.2010)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/227,635 22 July 2009 (22.07.2009) US
- (71) **Applicant (for all designated States except US):** **IBIS BIOSCIENCES, INC.** [US/US]; 1891 Rutherford Road, Carlsbad, California 92008-7326 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **SAMPATH, Rangarajan** [US/US]; 10817 Cherry Hill Drive, San Diego, California 92130 (US). **LI, Feng** [US/US]; 5651 Glenstone Way, San Diego, California 92121 (US).
- (74) **Agents:** **MUELLER, Lisa, V.** et al.; Polsinelli Shughart PC, 161 N. Clark Street, Suite 4200, Chicago, Illinois 60601 (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, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
- with international search report (Art. 21(3))
 - with sequence listing part of description (Rule 5.2(a))



WO 2011/011431 A1

(54) **Title:** COMPOSITIONS FOR USE IN IDENTIFICATION OF CALICIVIRUSES

(57) **Abstract:** The present invention relates generally to the detection and identification of caliciviruses and provides methods, compositions and kits useful for this purpose when combined, for example, with molecular mass or base composition analysis.

COMPOSITIONS FOR USE IN IDENTIFICATION OF CALICIVIRUSES

CROSS-REFERENCE TO RELATED APPLICATIONS

This claims priority to U.S. Provisional Patent Application No. 61/227,635,
5 filed on July 22, 2009, the contents of which are hereby incorporated by reference.

FIELD OF THE INVENTION

The present invention relates generally to the detection, identification and
characterization of caliciviruses, and provides methods, compositions, systems and kits
10 useful for this purpose when combined, for example, with molecular mass or base
composition analysis.

BACKGROUND OF THE INVENTION

The family Caliciviridae includes small, non-enveloped, icosahedral viruses with
15 a positive-sense, single-stranded, polyadenylated RNA genome (of about 7.5 to 8.5 kb).
This family consists of four genera; Norovirus, Sapovirus, Lagovirus, and Vesivirus (see
Farkas et al, J. Virol. **2008**, 82, 5408-5416 and references therein). The recent genomic
characterization of unique bovine enteropathogenic caliciviruses (Newbury agent-1 and
Nebraska) revealed that these viruses represent a distinct fifth genus with the proposed
20 name Becovirus or Nabovirus. Caliciviruses cause a wide spectrum of diseases in
animals, including respiratory infections, vesicular lesions, gastroenteritis, and
hemorrhagic disease. Noroviruses and Sapoviruses are important etiologic agents of acute
gastroenteritis in humans and therefore are also referred to as human Caliciviruses.
Viruses genetically and antigenically closely related to human Caliciviruses have also
25 been isolated from animals which has raised a concern about calicivirus gastroenteritis as
a zoonotic disease and the role of animals as reservoirs for human Caliciviruses.

Since there is no effective tissue culture system or animal model available for
human Caliciviruses, animal Caliciviruses are often used as surrogates to model human
Caliciviruses stability in the environment, replication, and pathogenesis.

30 Improved methods of diagnosing and characterizing Calicivirus infections as well
as identifying newly emergent strains of Caliciviruses are needed.

SUMMARY OF THE INVENTION

The present invention relates generally to the detection and identification of caliciviruses, and provides methods, compositions and kits useful for this purpose when combined, for example, with molecular mass or base composition analysis. However, the compositions find use in a variety of biological sample analysis techniques and are not limited to processes that employ or require molecular mass or base composition analysis. For example, primer pair compositions described herein may be used in a variety of research, surveillance, and diagnostic approaches that utilize one or more primers, including a variety of approaches that employ the polymerase chain reaction. In addition, the methods may be used to characterize a previously unknown calicivirus such as newly emerging strains which develop as a result of rapid evolution under selection pressure.

To further illustrate, in certain embodiments, the invention provides for the rapid detection and characterization of caliciviruses. The primer pairs described herein, for example, may be used to detect individual sub-species characteristics or strains of known caliciviruses.

In one aspect, a purified oligonucleotide primer pair composition is provided for identifying a known calicivirus or characterizing a previously unknown calicivirus. Among the advantages provided by the primer pair composition is the capability to hybridize to portions of calicivirus nucleic acid which are conserved among caliciviruses. This advantage allows nucleic acid from various caliciviruses to be amplified without the specific knowledge of the identity of any of the caliciviruses in a given sample. For example, it is desirable that a newly emergent calicivirus strain containing one or more SNPs, deletions or insertions be detected. In this case, the skilled person will recognize that SNPs, deletions or insertions occurring within the amplification products produced by the primer pair composition contain base composition information which would in most cases distinguish the newly emergent calicivirus strain from known caliciviruses. Selection of primer hybridization coordinates as well as the sequence of the primers themselves is a result of addressing a number of potential problems which may conspire to result in poor yields of amplification products or poorly resolvable amplification

products. Extensive testing and redesign is often required as part of the validation process to ensure that the primer pair compositions operate as intended.

5 The primer pair composition includes a forward primer and a reverse primer, each configured to hybridize to nucleic acid of two or more different caliciviruses in a nucleic acid amplification reaction which produces an amplification product between about 29 to about 200 nucleobases in length. The amplification product includes portions corresponding to a forward primer hybridization region, a reverse primer hybridization region and an intervening region having a base composition which varies among amplification products produced from nucleic acid of the two or more different
10 caliciviruses. The base composition of the intervening region provides a means for identifying the previously known calicivirus or characterizing the previously unknown calicivirus at the species level or the subspecies level. The subspecies level may represent a strain of a known species having one or more single nucleotide polymorphisms (SNPs), insertions or deletions for example.

15 The primer pair is configured to hybridize with nucleic acid of caliciviruses. In some embodiments, the nucleic acid of the caliciviruses is DNA which is obtained by performing a reverse transcriptase reaction on the native RNA of the calicivirus.

In some embodiments, each member of the primer pair has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting
20 of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3 wherein, with respect to pairs of sequence identifiers (X:Y) for primer pairs, the convention as defined herein is that the sequence identifier to the left of the colon (X:) represents the forward primer and the sequence identifier to the right of the colon (:Y) represents the reverse primer.

25 In some embodiments, the forward and reverse primers are about 14 to about 40 nucleobases in length. This range encompasses 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 35, 37, 38, 39 and 40 nucleobases. The forward and/or the reverse primer may include modifications such as having a non-templated thymidine residue on the 5'-end, at least one molecular mass modifying tag, at least one
30 modified nucleobase such as 5-propynyluracil or 5-propynylcytosine, a mass-modified nucleobase such as 5-iodo-cytosine, and a universal nucleobase such as inosine. Such

modifications are introduced with the aim of improving aspects of the amplification reaction such as minimizing 5'-adenylation catalyzed by polymerase enzymes, changing the mass of the amplification product to improve resolution of mass spectrum peaks, improving the affinity of the primer for the calicivirus nucleic acid, and improving the
5 range of hybridization of the primers across conserved regions of several different caliciviruses.

Another aspect of the invention is an isolated amplification product for identification of a known calicivirus or characterizing a previously unknown calicivirus at the species or subspecies level. The isolated amplification product is produced by
10 amplifying nucleic acid of a calicivirus in a reaction mixture comprising a forward primer and a reverse primer, each configured to hybridize to nucleic acid of two or more different caliciviruses. The amplification product has a length of about 29 to about 200 nucleobases and comprises portions corresponding to a forward primer hybridization region, a reverse primer hybridization region and an intervening region having a base
15 composition which varies among amplification products produced from nucleic acid of the two or more different caliciviruses. The expected lower limit of 29 nucleobases is based on the possibility of having an amplification product consisting of a forward 14-mer primer hybridization region, a single intervening nucleotide residue and reverse 14-mer primer hybridization region ($14+1+14 = 29$). The base composition of the
20 intervening region provides a means for identifying the previously known calicivirus or characterizing the previously unknown calicivirus. The amplification product is isolated from the reaction mixture and may be analyzed by a variety of analytical methods, preferably mass spectrometry.

In some embodiments, the step of isolating the amplification product is performed
25 using an anion exchange resin linked to a magnetic bead.

In some embodiments, the amplification product is produced using a primer pair wherein each member of the primer pair has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID
NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.

In some embodiments, the forward and reverse primers used to obtain the
30 inventive amplification products are about 14 to about 40 nucleobases in length. This

range encompasses 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 35, 37, 38, 39 and 40 nucleobases. The forward and/or the reverse primer may include modifications such as having a non-templated thymidine residue on the 5'-end, at least one molecular mass modifying tag, at least one modified nucleobase such as 5-propynyluracil or 5-propynylcytosine, a mass-modified nucleobase such as 5-iodo-cytosine, and a universal nucleobase such as inosine.

In another aspect, a method is provided for identifying a known calicivirus or characterizing a previously unknown calicivirus in a sample. The method includes the steps of:

- 10 (a) obtaining an amplification product by amplifying one or more nucleic acids of one or more caliciviruses in the sample using the primer pair composition described above;
- (b) measuring the molecular mass of one or both strands of the amplification product;
- 15 (c) comparing the molecular mass to a plurality of database-stored molecular masses of strands of amplification products of known caliciviruses; and
- d) identifying a match between the molecular mass and at least one of the database-stored molecular masses of amplification products, thereby identifying the known calicivirus or, alternatively, failing to identify a match between the molecular mass and at least one of the database-stored molecular masses, thereby characterizing a previously unknown calicivirus. In some embodiments of this method, each member of the primer pair has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3. In some embodiments, the molecular mass is determined by mass spectrometry.

In another aspect, a method is provided for identifying a known calicivirus or characterizing a previously unknown calicivirus in a sample. The method includes the steps of:

- 30 (a) obtaining an amplification product by amplifying one or more nucleic acids of one or more caliciviruses in the sample using the using the primer pair composition described above;

(b) measuring the molecular mass of one or both strands of the amplification product;

(c) determining the base composition of the amplification product from the molecular mass;

5 (d) comparing the base composition to a plurality of database-stored base compositions of strands of amplification products of known caliciviruses; and

(e) identifying a match between the base composition and at least one of the database-stored molecular masses of amplification products, thereby identifying the known calicivirus or, alternatively, failing to identify a match between the base
10 composition and at least one of the database-stored base compositions, thereby characterizing a previously unknown calicivirus. In some embodiments of this method, each member of the primer pair has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3. In some embodiments, the molecular mass is
15 determined by mass spectrometry.

In some embodiments, step (e) identifies the calicivirus as a member of a plurality of caliciviruses and the method further comprises repeating steps (a) to (e) using one or more additional primer pairs as defined in claim 1, wherein one or more repetitions of step (e) with the one or more additional primer pairs identifies the calicivirus or
20 characterizes the calicivirus as a unique calicivirus. In this particular embodiment, each member of the one or more additional primer pairs has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.

Another aspect of the invention is a kit comprising one or more purified primer
25 pairs for identifying a known calicivirus or characterizing a previously unknown calicivirus in a nucleic acid sample. Each member of the one or more primer pairs has at least 70% sequence identity with a corresponding member of one or more primer pairs selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3. The kit may include additional components such as a reverse
30 transcriptase, a polymerase and deoxynucleotide triphosphates which may be ¹³C-enriched for altering the molecular mass of the amplification products.

Another aspect of the invention is a system which includes the following components:

(a) a mass spectrometer configured to detect one or more molecular masses of the amplification products described above;

5 (b) a database of known molecular masses and/or known base compositions of amplification products of known caliciviruses; and

(b) a controller operably connected to the mass spectrometer and to the database. The controller is configured to match the molecular mass of the amplification product with a measured or calculated molecular mass of a corresponding amplification
10 product of a known calicivirus.

In some embodiments of the system described above, the database of known molecular masses and/or known base compositions of amplification products of known caliciviruses includes amplification products defined by one or more primer pairs wherein each member of the one or more primer pairs has at least 70% sequence identity with a
15 corresponding member of a corresponding primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3

In some embodiments, the compositions, methods, kits and systems described above are configured for identifying one or more calicivirus strains selected from the group consisting of Bovine Calicivirus, Calicivirus isolate 2117, Calicivirus isolate TCG,
20 Calicivirus pig/AB104/CAN, Calicivirus pig/AB90/CAN, Calicivirus pig/F15-10/CAN, Calicivirus strain NB, Canine Calicivirus, Cetacean Calicivirus, Feline Calicivirus, Mink Calicivirus, Newbury agent 1, Primate Calicivirus, Rabbit Calicivirus, Reptile Calicivirus, San Miguel sea lion virus, Skunk Calicivirus, Steller Sea Lion Calicivirus, Tulane Virus, Vesicular Exanthema of Swine Virus, VESV-like Calicivirus, and Walrus Calicivirus.

25

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing summary and detailed description is better understood when read in conjunction with the accompanying drawings which are included by way of example and not by way of limitation.

30 Figure 1 shows a process diagram illustrating one embodiment of the primer pair selection process.

Figure 2 shows a process diagram illustrating one embodiment of the primer pair validation process. Criteria include but are not limited to, the ability to amplify nucleic acid of caliciviruses, the ability to exclude amplification of extraneous nucleic acids and dimerization of primers, analytical limits of detection of 100 or fewer genomic
5 copies/reaction, and the ability to differentiate caliciviruses from each other.

Figure 3 shows a process diagram illustrating an embodiment of the calibration method.

Figure 4 shows a block diagram showing a representative system.

Figure 5 shows a mass spectrum obtained using primer pair VIR4962.

10 Figure 6 shows a mass spectrum obtained using primer pair VIR4962.

Figure 7 shows a mass spectrum obtained using primer pair VIR4961.

Figure 8 shows a mass spectrum obtained using primer pair VIR4961.

DETAILED DESCRIPTION OF EMBODIMENTS

15 It is to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. Further, unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains.

20 In describing and claiming the present invention, the following terminology and grammatical variants will be used in accordance with the definitions set forth below.

As used herein, the term “about” means encompassing plus or minus 10%. For example, about 200 nucleotides refers to a range encompassing between 180 and 220 nucleotides.

25 As used herein, the term “amplicon” or “bioagent identifying amplicon” refers to a nucleic acid segment deduced from hybridization of primer pairs to a known nucleic acid sequence. The deduction of an amplicon is well within the capabilities of a person skilled in the art. An amplicon may, for example, be deduced on page containing the known nucleic acid sequence and the sequences of the primers or using *in silico* methods
30 such as electronic PCR which are known to the skilled person. The skilled person will also readily recognize that the amplicon contains primer hybridization portions and an

intervening portion between the two primer hybridization portions. One important objective is to define many bioagent identifying amplicons using as few primer pairs as possible. Another important objective is to provide a primer pair which is specific for a specific calicivirus strain.

5 As used herein, the term “amplicon” or “bioagent identifying amplicon” is distinct from the term “amplification product” in that the term “amplification product” refers to the actual biomolecule produced in an actual amplification reaction. With respect to these definitions, an amplification product “corresponds” to an amplicon. This means that an amplicon may be present in a database even prior to a corresponding amplification
10 product ever being produced in an amplification reaction. An amplification product which corresponds to an amplicon must be produced by the same primers used to deduce the amplicon. The skilled person will recognize that if an amplicon residing in a database is in the form of a DNA sequence, an RNA sequence may be readily deduced from it, or *vice versa*. Thus, in the case of RNA viruses for example, a DNA sequence of an
15 amplicon may be deduced from the native RNA sequence for any given primer pair.

 The amplification products are typically double stranded DNA; however, it may be RNA and/or DNA:RNA. In some embodiments, the amplification product comprises DNA complementary to the RNA of caliciviruses. In some embodiments, the
20 amplification product comprises sequences of conserved regions/primer pairs and intervening variable region. As discussed herein, primer pairs are configured to generate amplification products from nucleic acid of caliciviruses. As such, the base composition of any given amplification product includes the base composition of each primer of the primer pair, the complement of each primer the primer pair and the intervening variable
25 region from the bioagent that was amplified to generate the amplification product. One skilled in the art understands that the incorporation of the designed primer pair sequences into an amplification product may replace the native sequences at the primer binding site, and complement thereof. In certain embodiments, after amplification of the target region using the primers the resultant amplification product having the primer sequences are used to generate the molecular mass data. Generally, the amplification product further
30 comprises a length that is compatible with mass spectrometry analysis. The amplification

products corresponding to bioagent identifying amplicons have base compositions that are preferably unique to the identity of a bioagent such as a calicivirus.

Amplicons and amplification products typically comprise from about 29 to about 200 consecutive nucleobases (*i.e.*, from about 29 to about 200 linked nucleosides). One of ordinary skill in the art will appreciate that this range expressly embodies compounds of 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, and 200 nucleobases in length. One of ordinary skill in the art will further appreciate that the above range is not an absolute limit to the length of an amplicon and amplification product, but instead represents a preferred length range. Lengths of amplification products falling outside of this range are also included herein so long as the amplification product is amenable to experimental determination of its molecular mass and/or its base composition as herein described.

The term "amplifying" or "amplification" in the context of nucleic acids refers to the production of multiple copies of a polynucleotide, or a portion of the polynucleotide, typically starting from a small amount of the polynucleotide (*e.g.*, a single polynucleotide molecule), where the amplification products or amplicons are generally detectable. Amplification of polynucleotides encompasses a variety of chemical and enzymatic processes. The generation of multiple DNA copies from one or a few copies of a target or template DNA molecule during a polymerase chain reaction (PCR) or a ligase chain reaction (LCR) are forms of amplification. Amplification is not limited to the strict duplication of the starting molecule. For example, the generation of multiple cDNA molecules from a limited amount of RNA in a sample using reverse transcription (RT)-PCR is a form of amplification. Furthermore, the generation of multiple RNA molecules

from a single DNA molecule during the process of transcription is also a form of amplification.

As used herein, the term “base composition” refers to the number of each residue in an amplicon, amplification product or other nucleic acid, without consideration for the linear arrangement of these residues in the strand(s). The residues may comprise, 5 adenosine (A), guanosine (G), cytidine, (C), (deoxy)thymidine (T), uracil (U), inosine (I), nitroindoles such as 5-nitroindole or 3-nitropyrrole, dP or dK (Hill F et al. Polymerase recognition of synthetic oligodeoxyribonucleotides incorporating degenerate pyrimidine and purine bases - *Proc. Natl. Acad. Sci. U. S. A.* **1998**, 95, 4258-63), an acyclic 10 nucleoside analog containing 5-nitroindazole (Van Aerschot *et al.*, *Nucleosides and Nucleotides*, 1995, 14, 1053-1056), or any of the following purine analogs: 1-(2-deoxy-beta-D-ribofuranosyl)-imidazole-4-carboxamide, 2,6-diaminopurine, 5-propynyluracil, 5-propynylcytosine, phenoxazines, including G-clamp, 5-propynyl deoxy-cytidine, deoxy-thymidine nucleotides, 5-propynylcytidine, 5-propynyluridine and mass tag modified 15 versions thereof, including 7-deaza-2'-deoxyadenosine-5'-triphosphate, 5-iodo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxycytidine-5'-triphosphate, 5-iodo-2'-deoxycytidine-5'-triphosphate, 5-hydroxy-2'-deoxyuridine-5'-triphosphate, 4-thiothymidine-5'-triphosphate, 5-aza-2'-deoxyuridine-5'-triphosphate, 5-fluoro-2'-deoxyuridine-5'-triphosphate, O6-methyl-2'-deoxyguanosine-5'- 20 triphosphate, N2-methyl-2'-deoxyguanosine-5'-triphosphate, 8-oxo-2'-deoxyguanosine-5'-triphosphate or thiothymidine-5'-triphosphate. In some embodiments, the mass-modified nucleobase comprises ^{15}N or ^{13}C or both ^{15}N and ^{13}C . In some embodiments, the non-natural nucleosides used herein include 5-propynyluracil, 5-propynylcytosine and inosine. Herein the base composition is notated as $A_wG_xC_yT_z$, wherein w, x, y and z are 25 each independently a whole number representing the number of the nucleoside residues in an amplicon and wherein T (thymidine) may be replaced by uracil (U) if desired, by simply using uridine triphosphates in the amplification reaction.

Base compositions of amplification products which include modified nucleosides are similarly notated to indicate the number of the natural and modified nucleosides in an 30 amplification product. Base compositions are determined from a molecular mass measurement of an amplification product, as described below. The base composition for

any given amplification product is then compared to a database of base compositions which typically includes base compositions calculated from sequences of amplicons deduced from a given primer pair and the known hybridization coordinates of the primers of the primer pair on the specific nucleic acid of a specific calicivirus. A match between the base composition of the amplification product and a single database amplicon entry reveals the identity of the bioagent. Alternatively, if a match between the base composition of the amplification product and the base compositions of individual amplicons in the database is not obtained, the conclusion may be drawn that the amplification product was obtained from nucleic acid of a previously uncharacterized calicivirus which may contain one or more SNPs, deletions, insertions or other sequence variations within the intervening variable region between the two primer hybridization sites. This is useful information which characterizes the previously uncharacterized calicivirus. It is useful to then incorporate the base composition of the previously uncharacterized calicivirus into the base composition database.

As used herein, a “base composition probability cloud” is a representation of the diversity in base composition resulting from a variation in sequence that occurs among different isolates of a given species, family or genus. Base composition calculations for a plurality of amplicons are mapped on a pseudo four-dimensional plot. Related members in a family, genus or species typically cluster within this plot, forming a base composition probability cloud.

As used herein, the term “base composition signature” refers to the base composition generated by any one particular amplicon.

As used herein, a “bioagent” means any biological organism or component thereof or a sample containing a biological organism or component thereof, including microorganisms or infectious substances, or any naturally occurring, bioengineered or synthesized component of any such microorganism or infectious substance or any nucleic acid derived from any such microorganism or infectious substance. Those of ordinary skill in the art will understand fully what is meant by the term bioagent given the instant disclosure. Still, a non-exhaustive list of bioagents includes: cells, cell lines, human clinical samples, mammalian blood samples, cell cultures, bacterial cells, viruses, viroids, fungi, protists, parasites, *Rickettsiae*, protozoa, animals, mammals or humans. Samples

may be alive, non-replicating or dead or in a vegetative state (for example, vegetative bacteria or spores). Preferably, the bioagent is a calicivirus.

As used herein, a "bioagent division" is defined as group of bioagents above the species level and includes but is not limited to, orders, families, genus, classes, clades, 5 genera or other such groupings of bioagents above the species level.

As used herein, "broad range survey primers" are primers designed to identify an unknown bioagent as a member of a particular biological division (*e.g.*, an order, family, class, clade, or genus). However, in some cases the broad range survey primers are also able to identify unknown bioagents at the species or sub-species level. As used herein, 10 "division-wide primers" are primers designed to identify a bioagent at the species level and "drill-down" primers are primers designed to identify a bioagent at the sub-species level. As used herein, the "sub-species" level of identification includes, but is not limited to, strains, subtypes, variants, and isolates. Drill-down primers are not always required for identification at the sub-species level because broad range survey primers may, in 15 some cases provide sufficient identification resolution to accomplishing this identification objective.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T- 20 C-A-5'." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in 25 amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

The term "conserved region" in the context of nucleic acids refers to a nucleobase sequence (*e.g.*, a subsequence of a nucleic acid, etc.) that is the same or similar in two or more different regions or segments of a given nucleic acid molecule (*e.g.*, an 30 intramolecular conserved region), or that is the same or similar in two or more different nucleic acid molecules (*e.g.*, an intermolecular conserved region). To illustrate, a

conserved region may be present in two or more different taxonomic ranks (*e.g.*, two or more different genera, two or more different species, two or more different subspecies, and the like) or in two or more different nucleic acid molecules from the same organism. To further illustrate, in certain embodiments, nucleic acids comprising at least one
5 conserved region typically have between about 70%-100%, between about 80-100%, between about 90-100%, between about 95-100%, or between about 99-100% sequence identity in that conserved region. A conserved region may also be selected or identified functionally as a region that permits generation of amplification products via primer extension through hybridization of a completely or partially complementary primer to the
10 conserved region for each of the target sequences to which conserved region is conserved.

The term “correlates” refers to establishing a relationship between two or more things. In certain embodiments, for example, detected molecular masses of one or more amplification products indicate the presence or identity of a given bioagent in a sample.
15 In some embodiments, base compositions are calculated or otherwise determined from the detected molecular masses of amplicons, which base compositions indicate the presence or identity of a given bioagent in a sample.

As used herein, in some embodiments, the term “database” is used to refer to a collection of molecular mass and/or base composition data. The molecular mass and/or
20 base composition data in the database is indexed to bioagents and to primer pairs. The base composition data reported in the database comprises the number of each nucleotide residue in an amplicon defined by each primer pair. The database can also be populated by empirical data determined from amplification products. In this aspect of populating the database, a primer pair is used to generate an amplification product. The molecular
25 mass of the amplification product is determined using a mass spectrometer and the base composition is calculated therefrom without sequencing *i.e.*, without determining the linear sequence of nucleobases comprising the amplification product. It is important to note that amplicon base composition entries in the database are typically derived from sequencing data (*i.e.*, known sequence information), but the base composition of the
30 amplification product being analyzed is determined without sequencing the amplification product. An entry in the database is made to associate correlate the base composition

with the identity of the bioagent and the primer pair used. The database may also be populated using other databases comprising bioagent information. For example, using the GenBank database it is possible to perform electronic PCR using an electronic representation of a primer pair. This *in silico* method may provide the base composition for any or all selected bioagent(s) stored in the GenBank database. The information may then be used to populate the base composition database as described above. A base composition database can be *in silico*, a written table, a reference book, a spreadsheet or any form generally amenable to access by data controllers. Preferably, it is *in silico* on computer readable media.

10 The term “detect”, “detecting” or “detection” refers to an act of determining the existence or presence of one or more bioagents in a sample.

As used herein, the term “etiology” refers to the causes or origins, of diseases or abnormal physiological conditions.

As used herein, the term "gene" refers to a nucleic acid (*e.g.*, DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (*e.g.*, rRNA, tRNA). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (*e.g.*, enzymatic activity, ligand binding, signal transduction, immunogenicity, etc.) of the full-length sequence or fragment thereof are retained.

20 As used herein, the term "heterologous gene" refers to a gene that is not in its natural environment. For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (*e.g.*, mutated, added in multiple copies, linked to non-native regulatory sequences, etc). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are typically joined to nucleic acid sequences that are not found naturally associated with the gene sequences in the chromosome or are associated with portions of the chromosome not found in nature (*e.g.*, genes expressed in loci where the gene is not normally expressed).

25 The terms "homology," "homologous" and “sequence identity” refer to a degree of identity. There may be partial homology or complete homology. A partially homologous sequence is one that is less than 100% identical to another sequence.

Determination of sequence identity is described in the following example: a primer 20 nucleobases in length which is otherwise identical to another 20 nucleobase primer but having two non-identical residues has 18 of 20 identical residues ($18/20 = 0.9$ or 90% sequence identity). In another example, a primer 15 nucleobases in length having all residues identical to a 15 nucleobase segment of a primer 20 nucleobases in length would have $15/20 = 0.75$ or 75% sequence identity with the 20 nucleobase primer. In context of the present invention, sequence identity is meant to be properly determined when the query sequence and the subject sequence are both described and aligned in the 5' to 3' direction. Sequence alignment algorithms such as BLAST, will return results in two different alignment orientations. In the Plus/Plus orientation, both the query sequence and the subject sequence are aligned in the 5' to 3' direction. On the other hand, in the Plus/Minus orientation, the query sequence is in the 5' to 3' direction while the subject sequence is in the 3' to 5' direction. It should be understood that with respect to the primers of the present invention, sequence identity is properly determined when the alignment is designated as Plus/Plus. Sequence identity may also encompass alternate or "modified" nucleobases that perform in a functionally similar manner to the regular nucleobases adenine, thymine, guanine and cytosine with respect to hybridization and primer extension in amplification reactions. In a non-limiting example, if the 5-propynyl pyrimidines propyne C and/or propyne T replace one or more C or T residues in one primer which is otherwise identical to another primer in sequence and length, the two primers will have 100% sequence identity with each other. In another non-limiting example, Inosine (I) may be used as a replacement for G or T and effectively hybridize to C, A or U (uracil). Thus, if inosine replaces one or more G or T residues in one primer which is otherwise identical to another primer in sequence and length, the two primers will have 100% sequence identity with each other. Other such modified or universal bases may exist which would perform in a functionally similar manner for hybridization and amplification reactions and will be understood to fall within this definition of sequence identity.

As used herein, "housekeeping gene" refers to a gene encoding a protein or RNA involved in basic functions required for survival and reproduction of a bioagent. Housekeeping genes include, but are not limited to, genes encoding RNA or proteins

involved in translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like.

As used herein, the term "hybridization" or "hybridize" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is influenced by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the melting temperature (T_m) of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized." An extensive guide to nucleic hybridization may be found in Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," Elsevier (1993), which is incorporated by reference.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced (*e.g.*, in the presence of nucleotides and an inducing agent such as a biocatalyst (*e.g.*, a DNA polymerase or the like) and at a suitable temperature and pH). The primer is typically single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is generally first treated to separate its strands before being used to prepare extension products. In some embodiments, the primer is an oligodeoxyribonucleotide. The primer is sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, "primers" or "primer pairs," in some embodiments, are oligonucleotides that are designed to bind to conserved sequence regions of one or more bioagent nucleic acids to generate bioagent identifying amplicons. In some embodiments, the bound primers flank an intervening variable region between the

conserved binding sequences. Upon amplification, the primer pairs yield amplification products that provide base composition variability between the two or more bioagents. The variability of the base compositions allows for the identification of one or more individual bioagents from, *e.g.*, two or more bioagents based on the base composition distinctions. In some embodiments, the primer pairs are also configured to generate amplification products amenable to molecular mass analysis. Further, the sequences of the primer members of the primer pairs are not necessarily fully complementary to the conserved region of the reference bioagent. For example, in some embodiments, the sequences are designed to be “best fit” amongst a plurality of bioagents at these conserved binding sequences. Therefore, the primer members of the primer pairs have substantial complementarity with the conserved regions of the bioagents, including the reference bioagent.

In some embodiments of the invention, the oligonucleotide primer pairs described herein can be purified. As used herein, “purified oligonucleotide primer pair,” “purified primer pair,” or “purified” means an oligonucleotide primer pair that is chemically-synthesized to have a specific sequence and a specific number of linked nucleosides. This term is meant to explicitly exclude nucleotides that are generated at random to yield a mixture of several compounds of the same length each with randomly generated sequence. As used herein, the term “purified” or “to purify” refers to the removal of one or more components (*e.g.*, contaminants) from a sample.

As used herein, the term “molecular mass” refers to the mass of a compound as determined using mass spectrometry, for example, ESI-MS. Herein, the compound is preferably a nucleic acid. In some embodiments, the nucleic acid is a double stranded nucleic acid (*e.g.*, a double stranded DNA nucleic acid). In some embodiments, the nucleic acid is an amplification product. When the nucleic acid is double-stranded the molecular mass may be determined for either strand or, preferably both strands. In one embodiment, the strands may be separated before introduction into the mass spectrometer, or the strands may be separated by the mass spectrometer itself (for example, electro-spray ionization will separate the hybridized strands). The molecular mass of each strand is measured by the mass spectrometer.

As used herein, the term "nucleic acid molecule" refers to any nucleic acid containing molecule, including but not limited to, DNA or RNA. The term encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4 acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxyl-methyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, dihydrouracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudo-uracil, 1-methylguanine, 1-methylinosine, 2,2-dimethyl-guanine, 2-methyladenine, 2-methylguanine, 3-methyl-cytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxy-amino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N- isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

As used herein, the term "nucleobase" is used as a term for describing the length of a given segment of nucleic acid and is synonymous with other terms in use in the art including "nucleotide," "deoxynucleotide," "nucleotide residue," and "deoxynucleotide residue." As is used herein, a nucleobase includes natural and modified nucleotide residues, as described herein.

An "oligonucleotide" refers to a nucleic acid that includes at least two nucleic acid monomer units (*e.g.*, nucleotides), typically more than three monomer units, and more typically greater than ten monomer units. The exact size of an oligonucleotide generally depends on various factors, including the ultimate function or use of the oligonucleotide. To further illustrate, oligonucleotides are typically less than 200 residues long (*e.g.*, between 15 and 100), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a "24-mer". Typically, the nucleoside monomers are linked by phosphodiester bonds or analogs thereof, including phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoranilidate, phosphoramidate, and

the like, including associated counterions, *e.g.*, H⁺, NH₄⁺, Na⁺, and the like, if such counterions are present. Further, oligonucleotides are typically single-stranded.

Oligonucleotides are optionally prepared by any suitable method, including, but not limited to, isolation of an existing or natural sequence, DNA replication or amplification, reverse transcription, cloning and restriction digestion of appropriate sequences, or direct chemical synthesis by a method such as the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68:90-99; the phosphodiester method of Brown *et al.* (1979) *Meth. Enzymol.* 68:109-151; the diethylphosphoramidite method of Beaucage *et al.* (1981) *Tetrahedron Lett.* 22:1859-1862; the triester method of Matteucci *et al.* (1981) *J. Am. Chem. Soc.* 103:3185-3191; automated synthesis methods; or the solid support method of U.S. Pat. No. 4,458,066, entitled "PROCESS FOR PREPARING POLYNUCLEOTIDES," issued Jul. 3, 1984 to Caruthers *et al.*, or other methods known to those skilled in the art. All of these references are incorporated by reference in entirety.

As used herein a "sample" refers to anything capable of being analyzed by the methods provided herein. In some embodiments, the sample comprises or is suspected one or more nucleic acids capable of analysis by the methods. Preferably, the samples comprise nucleic acids (*e.g.*, DNA, RNA, cDNAs, etc.) from one or more caliciviruses. Samples can include, for example, urine, feces, rectal swabs, blood, serum/plasma, cerebrospinal fluid (CSF), pleural/synovial/ocular fluids, blood culture bottles, culture isolates, and the like. In some embodiments, the samples are "mixture" samples, which comprise nucleic acids from more than one subject or individual. In some embodiments, the methods provided herein comprise purifying the sample or purifying the nucleic acid(s) from the sample. In some embodiments, the sample is purified nucleic acid. Essentially any sample preparation technique can be utilized to prepare samples for further analysis. In some embodiments, for example, commercially available kits, such as the Ambion TNA kit is optionally utilized.

A "sequence" of a biopolymer refers to the order and identity of monomer units (*e.g.*, nucleotides, etc.) in the biopolymer. The sequence (*e.g.*, base sequence) of a nucleic acid is typically read in the 5' to 3' direction.

As is used herein, the term “single primer pair identification” means that one or more bioagents can be identified using a single primer pair. A base composition signature for an amplicon may singly identify one or more bioagents.

5 As used herein, a “sub-species characteristic” is a genetic characteristic that provides the means to distinguish two members of the same bioagent species. For example, one viral strain may be distinguished from another viral strain of the same species by possessing a genetic change (*e.g.*, for example, a nucleotide deletion, addition or substitution) in one of the bacterial genes.

10 As used herein, in some embodiments the term “substantial complementarity” means that a primer member of a primer pair comprises between about 70%-100%, or between about 80-100%, or between about 90-100%, or between about 95-100%, or between about 99-100% complementarity with the conserved hybridization sequence of a nucleic acid from a given bioagent. Similarly, the primer pairs provided herein may comprise between about 70%-100%, or between about 80-100%, or between about 90-15 100%, or between about 95-100% identity, or between about 99-100% sequence identity with the primer pairs disclosed in Table 1. These ranges of complementarity and identity are inclusive of all whole or partial numbers embraced within the recited range numbers. For example, and not limitation, 75.667%, 82%, 91.2435% and 97% complementarity or sequence identity are all numbers that fall within the above recited range of 70% to 20 100%, therefore forming a part of this description. In some embodiments, any oligonucleotide primer pair may have one or both primers with less than 70% sequence homology with a corresponding member of any of the primer pairs of Table 1 if the primer pair has the capability of producing an amplification product corresponding to a calicivirus amplicon.

25 A “system” in the context of analytical instrumentation refers to a group of objects and/or devices that form a network for performing a desired objective.

As used herein, “triangulation identification” means the use of more than one primer pair to generate corresponding amplification products for identification of a bioagent. The more than one primer pair can be used in individual wells or vessels or in a 30 multiplex PCR assay. Alternatively, PCR reactions may be carried out in single wells or vessels comprising a different primer pair in each well or vessel. Following amplification

the amplification products are pooled into a single well or container which is then subjected to molecular mass analysis. The combination of pooled amplification products can be chosen such that the expected ranges of molecular masses of individual amplification products are not overlapping and thus will not complicate identification of signals. Triangulation is a process of elimination, wherein a first primer pair identifies that an unknown bioagent may be one of a group of bioagents. Subsequent primer pairs are used in triangulation identification to further refine the identity of the bioagent, for example, at the species or sub-species level amongst the subset of possibilities generated with the earlier primer pair. Triangulation identification is complete when the identity of the bioagent at the desired level of identification is determined. The triangulation identification process may also be used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen *et al.*, *J Appl Microbiol.*, 1999, 87, 270-278) in the absence of the expected compositions from the *B. anthracis* genome would suggest a genetic engineering event.

As used herein, the term “unknown bioagent” can mean, for example: (i) a bioagent whose existence is not known (for example, the SARS coronavirus was unknown prior to April 2003) and/or (ii) a bioagent whose existence is known (such as the well known bacterial species *Staphylococcus aureus* for example) but which is not known to be in a sample to be analyzed. For example, if the method for identification of coronaviruses disclosed in commonly owned U.S. Patent Serial No. 10/829,826 (incorporated herein by reference in its entirety) was to be employed prior to April 2003 to identify the SARS coronavirus in a clinical sample, both meanings of “unknown” bioagent are applicable since the SARS coronavirus was unknown to science prior to April, 2003 and since it was not known what bioagent (in this case a coronavirus) was present in the sample. On the other hand, if the method of U.S. Patent Serial No. 10/829,826 was to be employed subsequent to April 2003 to identify the SARS coronavirus in a clinical sample, the second meaning (ii) of “unknown” bioagent would apply because the SARS coronavirus became known to science subsequent to April 2003 because it was not known what bioagent was present in the sample.

As used herein, the term “variable region” is used to describe the intervening region between primer hybridization sites as described herein. The variable region possesses distinct base compositions between at least two bioagents, such that at least one bioagent can be identified at, for example, the family, genus, species or sub-species level.

5 The degree of variability between the at least two bioagents need only be sufficient to allow for identification using mass spectrometry analysis, as described herein.

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

10 Provided herein are methods, compositions, kits, and related systems for the detection and identification of caliciviruses using bioagent identifying amplicons. The primer pairs described herein, for example, may be used to detect any known member of the caliciviruses or to characterize previously uncharacterized caliciviruses or newly emergent strains of caliciviruses.

15 In some embodiments, primers are selected to hybridize to conserved sequence regions of nucleic acids of caliciviruses and which flank variable sequence regions to define a bioagent identifying amplicon. Amplification products corresponding to the amplicon are amenable to molecular mass determination. In some embodiments, the molecular mass is converted to a base composition, which indicates the number of each

20 nucleotide in the amplification product. Systems employing software and hardware useful in converting molecular mass data into base composition information are available from, for example, Ibis Biosciences, Inc. (Carlsbad, CA.), for example the Ibis T5000 Biosensor System, and are described in U.S. Patent Application No. 10/754,415, filed January 9, 2004, incorporated by reference herein in its entirety. In some embodiments,

25 the molecular mass or corresponding base composition of one or more different amplification products is queried against a database of molecular masses or base compositions indexed to bioagents and to the primer pair used to define the amplicon. A match of the measured base composition to a database entry base composition associates the sample bioagent to an indexed bioagent in the database. Thus, the identity of the

30 unknown bioagent is determined. No prior knowledge of the unknown bioagent is necessary to identify the unknown bioagent. In some instances, the measured base

composition associates with more than one database entry base composition. Thus, a second/subsequent primer pair is generally used to generate a second/subsequent amplification product, and its measured base composition is similarly compared to the database to determine its identity in triangulation identification. Furthermore, the methods and other aspects of the invention can be applied to rapid parallel multiplex analyses, the results of which can be employed in a triangulation identification strategy. Thus, in some embodiments, the present invention provides rapid throughput and does not require nucleic acid sequencing or knowledge of the linear sequences of nucleobases of the amplification product for bioagent detection and identification.

Particular embodiments of the mass-spectrum based detection methods are described in the following patents, patent applications and scientific publications, all of which are herein incorporated by reference as if fully set forth herein: US patent numbers 7,108,974; 7,217,510; 7,226,739; 7,255,992; 7,312,036; 7,339,051; US patent publication numbers 2003/0027135; 2003/0167133; 2003/0167134; 2003/0175695; 2003/0175696; 2003/0175697; 2003/0187588; 2003/0187593; 2003/0190605; 2003/0225529; 2003/0228571; 2004/0110169; 2004/0117129; 2004/0121309; 2004/0121310; 2004/0121311; 2004/0121312; 2004/0121313; 2004/0121314; 2004/0121315; 2004/0121329; 2004/0121335; 2004/0121340; 2004/0122598; 2004/0122857; 2004/0161770; 2004/0185438; 2004/0202997; 2004/0209260; 2004/0219517; 2004/0253583; 2004/0253619; 2005/0027459; 2005/0123952; 2005/0130196; 2005/0142581; 2005/0164215; 2005/0266397; 2005/0270191; 2006/0014154; 2006/0121520; 2006/0205040; 2006/0240412; 2006/0259249; 2006/0275749; 2006/0275788; 2007/0087336; 2007/0087337; 2007/0087338; 2007/0087339; 2007/0087340; 2007/0087341; 2007/0184434; 2007/0218467; 2007/0218467; 2007/0218489; 2007/0224614; 2007/0238116; 2007/0243544; 2007/0248969; 2007/0264661; 2008/0160512; 2008/0311558; 2009/0004643; 2009/0047665; 2009/0125245; WO2002/070664; WO2003/001976; WO2003/100035; WO2004/009849; WO2004/052175; WO2004/053076; WO2004/053141; WO2004/053164; WO2004/060278; WO2004/093644; WO 2004/101809; WO2004/111187; WO2005/023083; WO2005/023986; WO2005/024046; WO2005/033271; WO2005/036369; WO2005/086634; WO2005/089128;

WO2005/091971; WO2005/092059; WO2005/094421; WO2005/098047;
WO2005/116263; WO2005/117270; WO2006/019784; WO2006/034294;
WO2006/071241; WO2006/094238; WO2006/116127; WO2006/135400;
WO2007/014045; WO2007/047778; WO2007/086904; WO2007/100397;
5 WO2007/118222; WO2008/104002; WO2008/116182; WO2008/118809;
WO2008/127839; WO2008/143627; WO2008/151023; WO2009/017902;
WO2009/023358; WO2009/038840; Ecker *et al.*, Ibis T5000: a universal biosensor
approach for microbiology. *Nat Rev Microbiol.* 2008 Jun 3.; Ecker *et al.*, The Microbial
Rosetta Stone Database: A compilation of global and emerging infectious
10 microorganisms and bioterrorist threat agents. *BMC Microbiology.* 2005. 5(1): 19.; Ecker
et al., The Ibis T5000 Universal Biosensor: An Automated Platform for Pathogen
Identification and Strain Typing. *JALA.* 2006. 6(11): 341-351.; Ecker *et al.*, The
Microbial Rosetta Stone Database: A common structure for microbial biosecurity threat
agents. *J Forensic Sci.* 2005. 50(6): 1380-5.; Ecker *et al.*, Identification of *Acinetobacter*
15 species and genotyping of *Acinetobacter baumannii* by multilocus PCR and mass
spectrometry. *J Clin Microbiol.* 2006 Aug;44(8):2921-32.; Ecker *et al.*, Rapid
identification and strain-typing of respiratory pathogens for epidemic surveillance. *Proc.*
Natl. Acad. Sci. U S A. 2005 May 31;102(22):8012-7. Epub 2005 May 23.; Wortmann *et*
al., Genotypic evolution of *Acinetobacter baumannii* strains in an outbreak associated
20 with war trauma. *Infect Control Hosp Epidemiol.* 2008 Jun;29(6):553-555.; Hannis *et al.*,
High-resolution genotyping of *Campylobacter* species by use of PCR and high-
throughput mass spectrometry. *J Clin Microbiol.* 2008 Apr;46(4):1220-5.; Blyn *et al.*,
Rapid detection and molecular serotyping of adenovirus by use of PCR followed by
electrospray ionization mass spectrometry. *J Clin Microbiol.* 2008 Feb;46(2):644-51.;
25 Eshoo *et al.*, Direct broad-range detection of alphaviruses in mosquito extracts. *Virology.*
2007 Nov 25;368(2):286-95.; Sampath *et al.*, Global surveillance of emerging Influenza
virus genotypes by mass spectrometry. *PLoS ONE.* 2007 May 30;2(5):e489.; Sampath *et*
al., Rapid identification of emerging infectious agents using PCR and electrospray
ionization mass spectrometry. *Ann. N. Y. Acad. Sci.* 2007 Apr;1102:109-20.; Hujer *et al.*,
30 Analysis of antibiotic resistance genes in multidrug-resistant *Acinetobacter* sp. isolates
from military and civilian patients treated at the Walter Reed Army Medical Center.

Antimicrob Agents Chemother. 2006 Dec;50(12):4114-23.; Hall *et al.*, Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: a novel tool for the identification and differentiation of humans. *Anal Biochem.* 2005 Sep 1;344(1):53-69.; Sampath *et al.*, Rapid identification of emerging pathogens: coronavirus. *Emerg Infect Dis.* 2005 Mar;11(3):373-9.; Jiang Y, Hofstadler SA. A highly efficient and automated method of purifying and desalting PCR products for analysis by electrospray ionization mass spectrometry. *Anal Biochem.* 2003. 316: 50-57.; Jiang *et al.*, Mitochondrial DNA mutation detection by electrospray mass spectrometry. *Clin Chem.* 2006. 53(2): 195-203. Epub Dec 7.; Russell *et al.*, Transmission dynamics and prospective environmental sampling of adenovirus in a military recruit setting. *J Infect Dis.* 2006. 194(7): 877-85. Epub 2006 Aug 25.; Hofstadler *et al.*, Detection of microbial agents using broad-range PCR with detection by mass spectrometry: The TIGER concept. Chapter in *Encyclopedia of Rapid Microbiological Methods.* 2006.; Hofstadler *et al.*, Selective ion filtering by digital thresholding: A method to unwind complex ESI-mass spectra and eliminate signals from low molecular weight chemical noise. *Anal Chem.* 2006. 78(2): 372-378.; Hofstadler *et al.*, TIGER: The Universal Biosensor. *Int J Mass Spectrom.* 2005. 242(1): 23-41.; Van Ert *et al.*, Mass spectrometry provides accurate characterization of two genetic marker types in *Bacillus anthracis*. *Biotechniques.* 2004. 37(4): 642-4, 646, 648.; Sampath *et al.*, Forum on Microbial Threats: Learning from SARS: Preparing for the Next Disease Outbreak -- Workshop Summary. (ed. Knobler SE, Mahmoud A, Lemon S.) The National Academies Press, Washington, D.C. 2004. 181-185.

In certain embodiments, amplification products amenable to molecular mass determination produced by the primers described herein are either of a length, size or mass compatible with a particular mode of molecular mass determination, or compatible with a means of providing a fragmentation pattern in order to obtain fragments of a length compatible with a particular mode of molecular mass determination. Such means of providing a fragmentation pattern of an amplification product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, sonication or other means of fragmentation. Thus, in some embodiments, amplification products are larger than 200 nucleobases and are amenable to molecular mass determination following restriction

digestion. Methods of using restriction enzymes and cleavage primers are well known to those with ordinary skill in the art.

In some embodiments, amplification products corresponding to bioagent identifying amplicons are obtained using the polymerase chain reaction (PCR). Other amplification methods may be used such as ligase chain reaction (LCR), low-stringency
5 single primer PCR, and multiple strand displacement amplification (MDA). (Michael, SF., *Biotechniques* (1994), 16:411-412 and Dean *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* (2002), 99, 5261-5266).

One embodiment of a process flow diagram used for primer selection is depicted
10 in Figure 1. Likewise a process flow diagram for a primer validation process is shown in Figure 2. For each group of organisms, candidate target sequences are identified (200) from which nucleotide sequence alignments are created (210) and analyzed (220). Primers are then configured by selecting priming regions (230) to facilitate the selection of candidate primer pairs (240). Initially, the primer pair sequence is typically a “best fit”
15 amongst the aligned sequences, such that the primer pair sequence may or may not be fully complementary to the hybridization region on any one of the bioagents in the alignment. Thus, best fit primer pair sequences are those with sufficient complementarity with two or more bioagents to hybridize with the two or more bioagents and generate an amplification product. The primer pairs are then subjected to *in silico* analysis by
20 electronic PCR (ePCR) (300) wherein bioagent identifying amplicons are obtained from sequence databases such as GenBank or other sequence collections (310) and tested for specificity *in silico* (320). Bioagent identifying amplicons obtained from ePCR of GenBank sequences (310) may also be analyzed by a probability model which predicts the capability of a given amplicon to identify unknown bioagents. Preferably, the base
25 compositions of amplicons with favorable probability scores are then stored in a base composition database (325). Alternatively, base compositions of the bioagent identifying amplicons obtained from the primers and GenBank sequences are directly entered into the base composition database (330). Candidate primer pairs (240) are validated by *in vitro* amplification by a method such as PCR analysis (400) of nucleic acid from a
30 collection of organisms (410). Amplification products thus obtained are analyzed to confirm the sensitivity, specificity and reproducibility of the primers that define the

amplicons (420). If the results of the analysis are not satisfactory, a given primer may be redesigned by lengthening or shortening the primer or changing one or more of the nucleobases of the primer. Such changes may include simple substitution of a nucleobase for one of the remaining three standard nucleobases or by substitution with a modified
5 nucleobase or a universal nucleobase. The skilled person will recognize that the possible solutions to the problem of primer pair redesign is very large and that arriving at any given primer sequence either at the initial "best fit" step or in a subsequent redesign step thus requires significant inventive ingenuity in recognizing why the original primer does not function to a sufficient extent and in choosing a solution to the problem. Much more
10 than routine experimentation is thus required.

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

The primers typically are employed as compositions for use in methods for identification of caliciviruses as follows: a primer pair composition is contacted with nucleic acid such as, for example, DNA obtained from the RNA of the calicivirus via reverse transcription by known methods. The nucleic acid is then amplified by a nucleic
20 acid amplification technique, such as PCR for example, to obtain an amplification product that corresponds to a bioagent identifying amplicon. The molecular mass of the strands of the double-stranded amplification product is determined by a molecular mass measurement technique such as mass spectrometry, for example. Preferably the two strands of the double-stranded amplification product are separated during the ionization
25 process. However, they may be separated prior to mass spectrometry measurement. In some embodiments, the mass spectrometer is electrospray Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS) or electrospray time of flight mass spectrometry (ESI-TOF-MS). A list of possible base compositions may be generated for the molecular mass value obtained for each strand, and the choice of the
30 base composition from the list is facilitated by matching the base composition of one strand with a complementary base composition of the other strand. A measured

molecular mass or base composition calculated therefrom is then compared with a database of molecular masses or base compositions indexed to primer pairs and to known bioagents. A match between the measured molecular mass or base composition of the amplification product and the database-stored molecular mass or base composition for that indexed primer pair correlates the measured molecular mass or base composition with an indexed bioagent, thus identifying the unknown bioagent. In some embodiments, the primer pair used is at least one of the primer pairs of Table 1. In some embodiments, the method is repeated using a different primer pair to resolve possible ambiguities in the identification process or to improve the confidence level for the identification assignment (triangulation identification). In some embodiments, for example, where the unknown is a previously uncharacterized bioagent, the molecular mass or base composition from an amplification product generated from the previously uncharacterized bioagent is matched with one or more best match molecular masses or base compositions from a database to predict a family, genus, species, sub-type, etc. of the previously uncharacterized bioagent. Such information may assist further characterization of the this previously uncharacterized bioagent or provide a physician treating a patient infected by the unknown with a therapeutic agent best calculated to treat the patient.

In certain embodiments, caliciviruses are detected with the systems and methods of the present invention in combination with other bioagents, including other viruses, bacteria, fungi, or other bioagents. In particular embodiments, a primer pair panel is employed which includes primer pairs designed for production of amplification products of nucleic acid of caliciviruses. Other primer pairs may be included for production of amplification products of other viruses for example, in a wide viral survey. Such panels may be specific for a particular type of bioagent, or specific for a specific type of test (*e.g.*, for testing the safety of blood, one may include commonly present viral pathogens such as HCV, HIV, and bacteria that can be contracted via a blood transfusion).

In some embodiments, an amplification product may be produced using only a single primer (either the forward or reverse primer of any given primer pair), provided an appropriate amplification method is chosen, such as, for example, low stringency single primer PCR (LSSP-PCR).

In some embodiments, the oligonucleotide primers are broad range survey primers which hybridize to conserved regions of nucleic acid. The broad range primer may identify the unknown bioagent depending on which bioagent is in the sample. In other cases, the molecular mass or base composition of an amplicon does not provide sufficient resolution to identify the unknown bioagent as any one bioagent at or below the species level. These cases generally benefit from further analysis of one or more amplification products generated from at least one additional broad range survey primer pair, or from at least one additional division-wide primer pair, or from at least one additional drill-down primer pair. Identification of sub-species characteristics may be required, for example, to determine a clinical treatment of patient, or in rapidly responding to an outbreak of a new species, strain, sub-type, etc. of pathogen to prevent an epidemic or pandemic.

One with ordinary skill in the art of design of amplification primers will recognize that a given primer need not hybridize with 100% complementarity in order to effectively prime the synthesis of a complementary nucleic acid strand in an amplification reaction. Primer pair sequences may be a “best fit” amongst the aligned bioagent sequences, thus they need not be fully complementary to the hybridization region of any one of the bioagents in the alignment. Moreover, a primer may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (*e.g.*, for example, a loop structure or a hairpin structure). The primers may comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% sequence identity with any of the primers listed in Table 1. Thus, in some embodiments, an extent of variation of 70% to 100%, or any range falling within, of the sequence identity is possible relative to the specific primer sequences disclosed herein. To illustrate, determination of sequence identity is described in the following example: a primer 20 nucleobases in length which is identical to another 20 nucleobase primer having two non-identical residues has 18 of 20 identical residues ($18/20 = 0.9$ or 90% sequence identity). In another example, a primer 15 nucleobases in length having all residues identical to a 15 nucleobase segment of primer 20 nucleobases in length would have $15/20 = 0.75$ or 75% sequence identity with the 20 nucleobase primer. Percent identity need not be a whole number, for example when a 28 nucleobase primer is

completely identical to a 28 nucleobase portion of a 31 nucleobase primer, the 31 nucleobase primer is 90.3% identical to the 28 nucleobase primer ($28/31 = 0.9032$).

Percent homology, sequence identity or complementarity, can be determined by, for example, the Gap program (Wisconsin Sequence Analysis Package, Version 8 for
5 Unix, Genetics Computer Group, University Research Park, Madison WI), using default settings, which uses the algorithm of Smith and Waterman (*Adv. Appl. Math.*, 1981, 2, 482-489). In some embodiments, complementarity of primers with respect to the conserved priming regions of viral nucleic acid, is between about 70% and about 80%. In other embodiments, homology, sequence identity or complementarity, is between about
10 80% and about 90%. In yet other embodiments, homology, sequence identity or complementarity, is at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or is 100%.

In some embodiments, the primers described herein comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least
15 95%, at least 96%, at least 98%, or at least 99%, or 100% (or any range falling within) sequence identity with the primer sequences specifically disclosed herein.

In some embodiments, the oligonucleotide primers are 14 to 40 nucleobases in length (14 to 40 linked nucleotide residues). These embodiments comprise oligonucleotide primers 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
20 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nucleobases in length.

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

Primers may contain one or more universal bases. Because any variation (due to codon wobble in the third position) in the conserved regions among species is likely to
30 occur in the third position of a DNA (or RNA) triplet, oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind

to more than one nucleotide, referred to herein as a “universal nucleobase.” For example, under this “wobble” base pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal nucleobases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes *et al.*, *Nucleosides and Nucleotides*, 1995, 14, 1001-1003), the degenerate nucleotides dP or dK, an acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot *et al.*, *Nucleosides and Nucleotides.*, 1995, 14, 1053-1056) or the purine analog 1-(2-deoxy-beta-D-ribofuranosyl)-imidazole-4-carboxamide (Sala *et al.*, *Nucl Acids Res.*, 1996, 24, 3302-3306).

10 In some embodiments, to compensate for weaker binding by the wobble base, oligonucleotide primers are configured such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, 5-propynyluracil which binds to adenine and 5-
15 propynylcytosine and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is incorporated herein by reference in its entirety. Propynylated primers are described in U.S. Publication No. 2003/0170682 incorporated herein by reference in its entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and
20 6,005,096, each of which is incorporated herein by reference in its entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

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

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

In some embodiments, the primers contain mass-modifying tags. Reducing the total number of possible base compositions of a nucleic acid of specific molecular weight provides a means of avoiding a possible source of ambiguity in the determination of base composition of amplification products. Addition of mass-modifying tags to certain
10 nucleobases of a given primer will result in simplification of *de novo* determination of base composition of a given amplification product from its molecular mass.

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

In some embodiments, the molecular mass of a given amplification product of nucleic acid of a calicivirus is determined by mass spectrometry. Mass spectrometry is intrinsically a parallel detection scheme without the need for radioactive or fluorescent labels, because an amplification product is identified by its molecular mass. The current
25 state of the art in mass spectrometry is such that less than femtomole quantities of material can be analyzed to provide information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons.

30 In some embodiments, intact molecular ions are generated from amplification products using one of a variety of ionization techniques to convert the sample to the gas

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

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

In some embodiments, assignment of previously unobserved base compositions (also known as “true unknown base compositions”) to a given phylogeny can be accomplished via the use of pattern classifier model algorithms. Base compositions, like sequences, may vary slightly from strain to strain within species, for example. In some embodiments, the pattern classifier model is the mutational probability model. In other embodiments, the pattern classifier is the polytope model. A polytope model is the mutational probability model that incorporates both the restrictions among strains and position dependence of a given nucleobase within a triplet. In certain embodiments, a polytope pattern classifier is used to classify a test or unknown organism according to its amplicon base composition.

In some embodiments, it is possible to manage this diversity by building “base composition probability clouds” around the composition constraints for each species. A “pseudo four-dimensional plot” may be used to visualize the concept of base composition probability clouds. Optimal primer design typically involves an optimal choice of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual bioagents. Areas where clouds overlap generally indicate regions that may result in a misclassification, a problem which is overcome by a

triangulation identification process using bioagent identifying amplicons not affected by overlap of base composition probability clouds.

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

Provided herein is bioagent classifying information at a level sufficient to identify a given bioagent. Furthermore, the process of determining a previously unknown base composition for a given bioagent (for example, in a case where sequence information is unavailable) has utility by providing additional bioagent indexing information with which to populate base composition databases. The process of future bioagent identification is thus improved as additional base composition signature indexes become available in base composition databases.

In some embodiments, the identity and quantity of an unknown bioagent may be determined using the process illustrated in Figure 3. Primers (500) and a known quantity of a calibration polynucleotide (505) are added to a sample containing nucleic acid of an unknown bioagent. The total nucleic acid in the sample is then subjected to an amplification reaction (510) to obtain amplification products. The molecular masses of the amplification products are determined (515) from which are obtained molecular mass and abundance data. The molecular mass of the amplification product corresponding to a bioagent identifying amplicon (520) provides for its identification (525) and the molecular mass of the calibration amplicon obtained from the calibration polynucleotide (530) provides for quantification of the amplification product of the bioagent identifying amplicon (535). The abundance data of the bioagent identifying amplicon is recorded (540) and the abundance data for the calibration data is recorded (545), both of which are

used in a calculation (550) which determines the quantity of unknown bioagent in the sample.

In certain embodiments, a sample comprising an unknown bioagent is contacted with a primer pair which amplifies the nucleic acid from the bioagent, and a known quantity of a polynucleotide that comprises a calibration sequence. The amplification reaction then produces two amplification products which correspond to a bioagent identifying amplicon and a calibration amplicon. The amplification products corresponding to the bioagent identifying amplicon and the calibration amplicon are distinguishable by molecular mass while being amplified at essentially the same rate. Effecting differential molecular masses can be accomplished by choosing as a calibration sequence, a representative bioagent identifying amplicon (from a specific species of bioagent) and performing, for example, a 2-8 nucleobase deletion or insertion within the variable region between the two priming sites. The amplified sample containing the bioagent identifying amplicon and the calibration amplicon is then subjected to molecular mass analysis by mass spectrometry, for example. The resulting molecular mass analysis of the nucleic acid of the bioagent and of the calibration sequence provides molecular mass data and abundance data for the nucleic acid of the bioagent and of the calibration sequence. The molecular mass data obtained for the nucleic acid of the bioagent enables identification of the unknown bioagent by base composition analysis. The abundance data enables calculation of the quantity of the bioagent, based on the knowledge of the quantity of calibration polynucleotide contacted with the sample.

In some embodiments, construction of a standard curve in which the amount of calibration or calibrant polynucleotide spiked into the sample is varied provides additional resolution and improved confidence for the determination of the quantity of bioagent in the sample. Alternatively, the calibration polynucleotide can be amplified in its own reaction vessel or vessels under the same conditions as the bioagent. A standard curve may be prepared therefrom, and the relative abundance of the bioagent determined by methods such as linear regression. In some embodiments, multiplex amplification is performed where multiple amplification products corresponding to multiple bioagent identifying amplicons are obtained with multiple primer pairs which also amplify the corresponding standard calibration sequences. In this or other embodiments, the standard

calibration sequences are optionally included within a single construct (preferably a vector) which functions as the calibration polynucleotide.

In some embodiments, the calibrant polynucleotide is also used as an internal positive control to confirm that amplification conditions and subsequent analysis steps are successful in producing a measurable amplification product. Even in the absence of
5 copies of the genome of a bioagent, the calibration polynucleotide gives rise to an amplification product corresponding to a calibration amplicon. Failure to produce a measurable amplification product corresponding to a calibration amplicon indicates a failure of amplification or subsequent analysis step such as amplicon purification or
10 molecular mass determination. Reaching a conclusion that such failures have occurred is, in itself, a useful event. In other related embodiments, a separate internal positive control polynucleotide may be used. The same strategy used to prepare the calibration polynucleotide may be employed but with an insertion or deletion which is different from the insertion or deletion used in preparation of the internal positive control
15 polynucleotide.

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

In some embodiments, a calibration sequence is inserted into a vector which then functions as the calibration polynucleotide. In some embodiments, more than one
20 calibration sequence is inserted into the vector that functions as the calibration polynucleotide. Such a calibration polynucleotide is herein termed a “combination calibration polynucleotide.” It should be recognized that the calibration method should not be limited to the embodiments described herein. The calibration method can be applied for determination of the quantity of any amplification product corresponding to a
25 bioagent identifying amplicon when an appropriate standard calibrant polynucleotide sequence and/or an appropriate internal positive control polynucleotide are designed and used.

In certain embodiments, primer pairs are configured to produce amplification products corresponding to bioagent identifying amplicons within more conserved regions
30 of nucleic acid of caliciviruses, while others produce amplification products corresponding to bioagent identifying amplicons within regions that are may evolve more

quickly. Primer pairs that define bioagent identifying amplicons in a conserved region with low probability that the region will evolve past the point of primer recognition are useful, *e.g.*, as a broad range survey-type primer. Primer pairs that define a bioagent identifying amplicon corresponding to an evolving genomic region are useful, *e.g.*, for
5 distinguishing emerging bioagent strain variants.

The primer pairs described herein provide methods for identifying diseases caused by known or emerging calicivirus strains. Base composition analysis eliminates the need for prior knowledge of the sequences of these strains for generation of hybridization probes. Thus, in another embodiment, there is provided a method for determining the
10 etiology of a particular disease when the process of identification of is carried out in a clinical setting, and even when a new strain is involved. This is possible because the methods may not be confounded by naturally occurring evolutionary variations.

Another embodiment provides a means of tracking the spread of any calicivirus strain when a plurality of samples obtained from different geographical locations are
15 analyzed by methods described above in an epidemiological setting. For example, a plurality of samples from a plurality of different locations may be analyzed with primers which define bioagent identifying amplicons, a subset of which identifies a specific strain. The corresponding locations of the members of the strain-containing subset indicate the spread of the specific strain to the corresponding locations.

Also provided are kits for carrying out the methods described herein. In some
20 embodiments, the kit may comprise a sufficient quantity of one or more primer pairs to perform an amplification reaction on a target polynucleotide from a bioagent which corresponds to a bioagent identifying amplicon. In some embodiments, the kit may comprise from one to twenty primer pairs, from one to ten primer pairs, from one to eight
25 pairs, from one to five primer pairs, from one to three primer pairs, or from one to two primer pairs. In some embodiments, the kit may comprise one or more primer pairs recited in Table 1.

In some embodiments, the kit may also comprise a sufficient quantity of reverse transcriptase, a DNA polymerase, suitable nucleoside triphosphates (including any of
30 those described above), a DNA ligase, and/or reaction buffer, or any combination thereof, for the amplification processes described above. A kit may further include instructions

pertinent for the particular embodiment of the kit, such instructions describing the primer pairs and amplification conditions for operation of the method. In some embodiments, the kit further comprises instructions for analysis, interpretation and dissemination of data acquired by the kit. In other embodiments, instructions for the operation, analysis, interpretation and dissemination of the data of the kit are provided on computer readable media. A kit may also comprise amplification reaction containers such as microcentrifuge tubes, microtiter plates, and the like. A kit may also comprise reagents or other materials for isolating bioagent nucleic acid or amplification products, including, for example, detergents, solvents, or ion exchange resins which may be linked to magnetic beads. A kit may also comprise a table of measured or calculated molecular masses and/or base compositions of bioagents using the primer pairs of the kit.

The invention also provides systems that can be used to perform various assays relating to detection, identification or characterization of caliciviruses. In certain embodiments, systems include mass spectrometers configured to detect molecular masses of amplification products produced using purified oligonucleotide primer pairs described herein. Other detectors that are optionally adapted for use in the systems of the invention are described further below. In some embodiments, systems also include controllers operably connected to mass spectrometers and/or other system components. In some of these embodiments, controllers are configured to correlate the molecular masses of the amplification products with the molecular masses of bioagent identifying amplicons of bioagents to effect detection, identification or characterization. In some embodiments, controllers are configured to determine base compositions of the amplification products from the molecular masses of the amplification products. As described herein, the base compositions generally correspond to calicivirus strain identities. In certain embodiments, controllers include (or are operably connected to) databases of known molecular masses and/or known base compositions of amplification products of known strains of caliciviruses produced with the primer pairs described herein. Controllers are described further below.

In some embodiments, systems include one or more of the primer pairs described herein. In certain embodiments, the oligonucleotides are arrayed on solid supports, whereas in others, they are provided in one or more containers, *e.g.*, for assays performed

in solution. In certain embodiments, the systems also include at least one detector or detection component (*e.g.*, a spectrometer) that is configured to detect detectable signals produced in the container or on the support. In addition, the systems also optionally include at least one thermal modulator (*e.g.*, a thermal cycling device) operably
5 connected to the containers or solid supports to modulate temperature in the containers or on the solid supports, and/or at least one fluid transfer component (*e.g.*, an automated pipettor) that transfers fluid to and/or from the containers or solid supports, *e.g.*, for performing one or more assays (*e.g.*, nucleic acid amplification, real-time amplicon detection, etc.) in the containers or on the solid supports.

10 Detectors are typically structured to detect detectable signals produced, *e.g.*, in or proximal to another component of the given assay system (*e.g.*, in a container and/or on a solid support). Suitable signal detectors that are optionally utilized, or adapted for use, herein detect, *e.g.*, fluorescence, phosphorescence, radioactivity, absorbance, refractive index, luminescence, or mass. Detectors optionally monitor one or a plurality of signals
15 from upstream and/or downstream of the performance of, *e.g.*, a given assay step. For example, detectors optionally monitor a plurality of optical signals, which correspond in position to "real-time" results. Example detectors or sensors include photomultiplier tubes, CCD arrays, optical sensors, temperature sensors, pressure sensors, pH sensors, conductivity sensors, or scanning detectors. Detectors are also described in, *e.g.*, Skoog
20 *et al.*, Principles of Instrumental Analysis, 5th Ed., Harcourt Brace College Publishers (1998), Currell, Analytical Instrumentation: Performance Characteristics and Quality, John Wiley & Sons, Inc. (2000), Sharma *et al.*, Introduction to Fluorescence Spectroscopy, John Wiley & Sons, Inc. (1999), Valeur, Molecular Fluorescence: Principles and Applications, John Wiley & Sons, Inc. (2002), and Gore,
25 Spectrophotometry and Spectrofluorimetry: A Practical Approach, 2nd Ed., Oxford University Press (2000), which are each incorporated by reference.

As mentioned above, the systems of the invention also typically include controllers that are operably connected to one or more components (*e.g.*, detectors, databases, thermal modulators, fluid transfer components, robotic material handling
30 devices, and the like) of the given system to control operation of the components. More specifically, controllers are generally included either as separate or integral system

components that are utilized, *e.g.*, to receive data from detectors (*e.g.*, molecular masses, etc.), to effect and/or regulate temperature in the containers, or to effect and/or regulate fluid flow to or from selected containers. Controllers and/or other system components are optionally coupled to an appropriately programmed processor, computer, digital
5 device, information appliance, or other logic device (*e.g.*, including an analog to digital or digital to analog converter as needed), which functions to instruct the operation of these instruments in accordance with preprogrammed or user input instructions, receive data and information from these instruments, and interpret, manipulate and report this information to the user. Suitable controllers are generally known in the art and are
10 available from various commercial sources.

Any controller or computer optionally includes a monitor, which is often a cathode ray tube ("CRT") display, a flat panel display (*e.g.*, active matrix liquid crystal display or liquid crystal display), or others. Computer circuitry is often placed in a box, which includes numerous integrated circuit chips, such as a microprocessor, memory,
15 interface circuits, and others. The box also optionally includes a hard disk drive, a floppy disk drive, a high capacity removable drive such as a writeable CD-ROM, and other common peripheral elements. Inputting devices such as a keyboard or mouse optionally provide for input from a user. These components are illustrated further below.

The computer typically includes appropriate software for receiving user
20 instructions, either in the form of user input into a set of parameter fields, *e.g.*, in a graphic user interface (GUI), or in the form of preprogrammed instructions, *e.g.*, preprogrammed for a variety of different specific operations. The software then converts these instructions to appropriate language for instructing the operation of one or more controllers to carry out the desired operation. The computer then receives the data from,
25 *e.g.*, sensors/detectors included within the system, and interprets the data, either provides it in a user understood format, or uses that data to initiate further controller instructions, in accordance with the programming.

Figure 4 is a schematic showing a representative system that includes a logic device in which various aspects of the present invention may be embodied. As will be
30 understood by practitioners in the art from the teachings provided herein, aspects of the invention are optionally implemented in hardware and/or software. In some

embodiments, different aspects of the invention are implemented in either client-side logic or server-side logic. As will be understood in the art, the invention or components thereof may be embodied in a media program component (*e.g.*, a fixed media component) containing logic instructions and/or data that, when loaded into an appropriately
5 configured computing device, cause that device to perform as desired. As will also be understood in the art, a fixed medium containing logic instructions may be delivered to a viewer on a fixed media for physically loading into a viewer's computer or a fixed media containing logic instructions may reside on a remote server that a viewer accesses through a communication medium in order to download a program component.

10 More specifically, Figure 4 schematically illustrates computer 1000 to which mass spectrometer 1002 (*e.g.*, an ESI-TOF mass spectrometer, etc.), fluid transfer component 1004 (*e.g.*, an automated mass spectrometer sample injection needle or the like), and database 1008 are operably connected. Optionally, one or more of these components are operably connected to computer 1000 via a server (not shown in Figure 4). During
15 operation, fluid transfer component 1004 typically transfers reaction mixtures or components thereof (*e.g.*, aliquots comprising amplicons) from multi-well container 1006 to mass spectrometer 1002. Mass spectrometer 1002 then detects molecular masses of the amplicons. Computer 1000 then typically receives this molecular mass data, calculates base compositions from this data, and compares it with entries in database
20 1008 to identify strains of caliciviruses in a given sample. It will be apparent to one of skill in the art that one or more components of the system schematically depicted in Figure 4 are optionally fabricated integral with one another (*e.g.*, in the same housing).

While the present invention has been described with specificity in accordance with certain of its embodiments, the following examples serve only to illustrate the
25 invention and are not intended to limit the same. In order that the invention disclosed herein may be more efficiently understood, examples are provided below. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting the invention in any manner.

30 EXAMPLES

Example 1: Selection of Design and Validation of Primers that Define Bioagent Identifying Amplicons for Caliciviruses

For design of primers that define calicivirus identifying amplicons, a series of calicivirus genome segment sequences were obtained, aligned and scanned for regions where pairs of PCR primers amplify products of about 29 to about 200 nucleobases in length and distinguish individual strains from each other by their molecular masses or base compositions. A typical process shown in Figure 1 is employed for this type of analysis. Primer pair validation is carried out according to some or all of the steps shown in Figure 2.

A database of expected base compositions for each primer region is generated using an *in silico* PCR search algorithm, such as (ePCR). An existing RNA structure search algorithm, (Macke et al. *Nucl. Acids Res.*, 2001, 29, 4724-4735, incorporated herein by reference in its entirety) has been modified to include PCR parameters such as hybridization conditions, mismatches, and thermodynamic calculations (SantaLucia, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, 95, 1460-1465, which is incorporated herein by reference in its entirety). This also provides information on primer specificity of the selected primer pairs.

Tables 1 to 3 provide information about the primers selected according to the processes described above. These tables may be conveniently cross-referenced according to the primer pair number listed in the leftmost column. Table 1 lists the sequences of the forward and reverse primers for each of the primer pairs.

Table 1: Sequences of Primer Pairs Designed for Identification of Caliciviruses

Primer Pair Number	Primer Direction	Primer Sequence	SEQ ID NO:
VIR4959	Forward	TGATGAACCAGTTGAGTTTCTCAAGCG	11
VIR4959	Reverse	TCAGCTTTGATGTAGTAGAATTGCCGA	7
VIR4960	Forward	TGGAACATCTTCGACTCGATGGACCT	13
VIR4960	Reverse	TCGGTCCGGGTCGGTTTCAG	8
VIR4961	Forward	TATTGTGTGACTACTCAAAGTGGGACTC	4
VIR4961	Reverse	TGCACAGGCCGAATCAACGATTGG	12
VIR4962	Forward	TGGTGTGTGATGTGAATTGGAATATCTACGA	14
VIR4962	Reverse	TGTACACTCCGTCATCACCATAGGTGAA	15
VIR4963	Forward	TAAGTGGGACTCAACTCAACCTCCAAATGT	1
VIR4963	Reverse	TGTAGCAACAGCAGAATCAACTATAGGA	16

VIR4964	Forward	TACCCGCCAATCAGCATGTGGTAAC	2
VIR4964	Reverse	TGACGACGAAGAGCCCAGGCC	9
VIR4965	Forward	TTCAATGGTGTGGAGGCGCGG	17
VIR4965	Reverse	TCAAACTTCGAACACATCACAGTGTAG	5
VIR4966	Forward	TGACGAGGAGTACGACGAGTGGATGAA	10
VIR4966	Reverse	TTGCCCATCGCGGCCCTGTG	18
VIR4967	Forward	TCAGCAGCACTCAAGGATGAGCC	6
VIR4967	Reverse	TAGCAAGCGCCTCTTCCC	3

Table 2 provides primer pair names constructed of notations which indicate information about the primers and their hybridization coordinates with respect to a reference sequence. The primer pair name “VESIVIRUSRDRP_NC002551-20-5665_5241_5351” of primer pair number VIR4959 indicates that the primers of this primer pair are designed to amplify a vesivirus (VESIVIRUS..) genome segment within the RNA-dependent RNA polymerase gene (..RDRP..). The reference sequence used in naming the primer pair is that of GenBank Accession No. NC_002551 (Vesicular Exanthema of Swine Virus). An extraction of residues 20 to 5241 was taken from the sequence of this GenBank Accession number. A reference amplicon formed by a theoretical amplification of this sequence extraction with the forward and reverse primers of VIR4959 define a calicivirus identifying amplicon 111 nucleobases in length corresponding to positions 5241 to 5351 of the extraction of residues 20 to 5665 of the genomic sequence of NC_002551. Thus, with this explanation of the coding of the primer pair names and the additional coding information provided in Table 3 a person skilled in the art will understand the coordinates of the amplicons with respect to the reference sequences indicated. The skilled person will also recognize that while the primer pairs are named with respect to a reference sequence, they are capable of hybridizing to nucleic acid of additional caliciviruses for amplification of segments corresponding to additional calicivirus amplicons.

Table 2: Primer Pair Name Codes and Reference Amplicon Lengths

Primer Pair Number	Primer Pair Name	Reference Amplicon Length
VIR4959	VESIVIRUSRDRP_NC002551-20-5665_5241_5351	111
VIR4960	VESIVIRUSRDRP_NC002551-20-5665_5071_5207	137

VIR4961	VESIVIRUSRDRP_NC002551-20-5665_4786_4902	117
VIR4962	VESIVIRUSRDRP_AY343325-1-5451_4866_4936	71
VIR4963	VESIVIRUSRDRP_AY343325-1-5451_4614_4716	103
VIR4964	VESIVIRUSRDRP_NC001481-20-5311_2405_2471	67
VIR4965	VESIVIRUSRDRP_NC001481-20-5311_5212_5292	81
VIR4966	VESIVIRUSRDRP_NC006875-75-6707_2808_2909	102
VIR4967	VESIVIRUSRDRP_NC012699-11-5950_3424_3489	66

Table 3 provides names for individual primers of the indicated primer pairs. The individual primer naming convention is similar to that of the primer pairs except that the last two numbered coordinates indicate the hybridization coordinates of the individual primer with respect to the reference sequence whereas the primer pair names indicate the coordinates of the entire amplicon with respect to the reference sequence. For example, the forward primer of primer pair number VIR4959 hybridizes to residues 5241 to 5267 of an extraction consisting of residues 20 to 5665 of GenBank Accession number NC_002551. The final letter code specifies the primer direction, wherein “_F” indicates forward primer and “_R” indicates reverse primer.

Table 3: Individual Primer Names

Primer Pair Number	Primer Direction	Individual Primer Names
VIR4959	Forward	VESIVIRUSRDRP_NC002551-20-5665_5241_5267_F
VIR4959	Reverse	VESIVIRUSRDRP_NC002551-20-5665_5325_5351_R
VIR4960	Forward	VESIVIRUSRDRP_NC002551-20-5665_5071_5096_F
VIR4960	Reverse	VESIVIRUSRDRP_NC002551-20-5665_5188_5207_R
VIR4961	Forward	VESIVIRUSRDRP_NC002551-20-5665_4786_4814_F
VIR4961	Reverse	VESIVIRUSRDRP_NC002551-20-5665_4879_4902_R
VIR4962	Forward	VESIVIRUSRDRP_AY343325-1-5451_4866_4895_F
VIR4962	Reverse	VESIVIRUSRDRP_AY343325-1-5451_4909_4936_R
VIR4963	Forward	VESIVIRUSRDRP_AY343325-1-5451_4614_4643_F
VIR4963	Reverse	VESIVIRUSRDRP_AY343325-1-5451_4689_4716_R
VIR4964	Forward	VESIVIRUSRDRP_NC001481-20-5311_2405_2429_F
VIR4964	Reverse	VESIVIRUSRDRP_NC001481-20-5311_2451_2471_R
VIR4965	Forward	VESIVIRUSRDRP_NC001481-20-5311_5212_5232_F
VIR4965	Reverse	VESIVIRUSRDRP_NC001481-20-5311_5266_5292_R
VIR4966	Forward	VESIVIRUSRDRP_NC006875-75-6707_2808_2834_F
VIR4966	Reverse	VESIVIRUSRDRP_NC006875-75-6707_2890_2909_R
VIR4967	Forward	VESIVIRUSRDRP_NC012699-11-5950_3424_3446_F
VIR4967	Reverse	VESIVIRUSRDRP_NC012699-11-5950_3472_3489_R

Example 2: One-Step RT-PCR of RNA Virus Samples

RNA is isolated from virus-containing samples according to methods well known in the art. To generate DNA from the RNA viruses, a one-step RT-PCR protocol was developed. RT-PCR reactions are assembled in 50 μ L reactions in the 96 well microtiter plate format using a Packard MPII liquid handling robotic platform and MJ Dyad[®] thermocyclers (MJ research, Waltham, MA). A typical RT-PCR reaction consists of 4 units of Amplitaq Gold[®], 1.5x buffer II (Applied Biosystems, Foster City, CA), 1.5 mM MgCl₂, 0.4 M betaine, 10 mM DTT, 20 mM sorbitol, 50 ng random primers (Invitrogen, Carlsbad, CA), 1.2 units Supersasin (Ambion, Austin, TX), 100 ng polyA DNA, 2 units Superscript III (Invitrogen, Carlsbad, CA), 400 ng T4 Gene 32 Protein (Roche Applied Science, Indianapolis, IN), 800 μ M dNTP mix, and 250 nM of each primer.

The following PCR conditions are typically used to produce amplification products for mass spectrometry analysis: 60°C for 5 minutes, 40°C for 10 minutes, 55°C for 45 minutes, 95°C for 10 minutes followed by 8 cycles of 95°C for 30 seconds, 48°C for 30 seconds, and 72°C for 30 seconds, with the 48°C annealing temperature increased 0.9°C after each cycle. The PCR reaction is then continued for 37 additional cycles of 95°C for 15 seconds, 56°C for 20 seconds, and 72°C for 20 seconds. The reaction concludes with 2 minutes at 72°C.

20

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

For solution capture of nucleic acids with ion exchange resin linked to magnetic beads, 25 μ L of a 2.5 mg/mL suspension of BioClone amine-terminated supraparamagnetic beads are added to 25 to 50 μ L of a PCR (or RT-PCR) reaction containing approximately 10 pM of a typical PCR amplification product. This suspension is mixed for approximately 5 minutes by vortexing or pipetting, after which the liquid is removed after using a magnetic separator. The beads containing bound PCR amplification product are then washed three times with 50 mM ammonium bicarbonate/50% MeOH or 100 mM ammonium bicarbonate/50% MeOH, followed by three more washes with 50% MeOH. The bound PCR amplification products are eluted in

30

a solution containing 25 mM piperidine, 25 mM imidazole, 35% MeOH and peptides as mass calibration standards.

Example 4: Mass Spectrometry and Base Composition Analysis

5 The ESI-FTICR mass spectrometer is based on a Bruker Daltonics (Billerica, MA) Apex II 70e electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer that employs an actively shielded 7 Tesla superconducting magnet. The active shielding constrains the majority of the fringing magnetic field from the superconducting magnet to a relatively small volume. Thus, components that might be
10 adversely affected by stray magnetic fields, such as CRT monitors, robotic components, and other electronics, can operate in close proximity to the FTICR spectrometer. All aspects of pulse sequence control and data acquisition are performed on a 600 MHz Pentium II data station running Bruker's Xmass software under the Windows NT 4.0 operating system. Sample aliquots, typically 15 μ L, are extracted directly from 96-well
15 microtiter plates using a CTC HTS PAL autosampler (LEAP Technologies, Carrboro, NC) triggered by the FTICR data station. Samples are injected directly into a 10 μ L sample loop integrated with a fluidics handling system that supplies the 100 μ L/hr flow rate to the ESI source. Ions are formed via electrospray ionization in a modified Analytica (Branford, CT) source employing an off axis, grounded electrospray probe positioned
20 approximately 1.5 cm from the metalized terminus of a glass desolvation capillary. The atmospheric pressure end of the glass capillary is biased at 6000 V relative to the ESI needle during data acquisition. A counter-current flow of dry N₂ is employed to assist in the desolvation process. Ions are accumulated in an external ion reservoir comprised of an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode, prior to injection
25 into the trapped ion cell where they are mass analyzed. Ionization duty cycles > 99% are achieved by simultaneously accumulating ions in the external ion reservoir during ion detection. Each detection event consists of IM data points digitized over 2.3 s. To improve the signal-to-noise ratio (S/N), 32 scans are typically co-added for a total data acquisition time of 74 s.

30 The ESI-TOF mass spectrometer is based on a Bruker Daltonics MicroTOF™. Ions from the ESI source undergo orthogonal ion extraction and are focused in a

reflectron prior to detection. The TOF and FTICR are equipped with the same automated sample handling and fluidics described above. Ions are formed in the standard MicroTOF™ ESI source that is equipped with the same off-axis sprayer and glass capillary as the FTICR ESI source. Consequently, source conditions are the same as those described above. External ion accumulation is also employed to improve ionization duty cycle during data acquisition. Each detection event on the TOF is typically comprised of 75,000 data points digitized over 75 μ s.

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

Raw mass spectra are post-calibrated with an internal mass standard and deconvoluted to monoisotopic molecular masses. Unambiguous base compositions are derived from the exact mass measurements of the complementary single-stranded oligonucleotides. Quantitative results are obtained by comparing the peak heights with an internal PCR calibration standard present in every PCR well at 500 molecules per well. Calibration methods are commonly owned and disclosed in U.S. Patent Application No. 20090004643 which is incorporated herein by reference in entirety.

Example 5: *De Novo* Determination of Base Composition of Amplicons using Molecular Mass Modified Deoxynucleotide Triphosphates

Because the molecular masses of the four natural nucleobases fall within a narrow molecular mass range (A = 313.058, G = 329.052, C = 289.046, T = 304.046, values in Daltons – See, Table 4), a source of ambiguity in assignment of base composition may occur as follows: two nucleic acid strands having different base composition may have a difference of about 1 Da when the base composition difference between the two strands is G ↔ A (-15.994) combined with C ↔ T (+15.000). For example, one 99-mer nucleic acid strand having a base composition of A₂₇G₃₀C₂₁T₂₁ has a theoretical molecular mass of 30779.058 while another 99-mer nucleic acid strand having a base composition of A₂₆G₃₁C₂₂T₂₀ has a theoretical molecular mass of 30780.052 is a molecular mass difference of only 0.994 Da. A 1 Da difference in molecular mass may be within the experimental error of a molecular mass measurement and thus, the relatively narrow molecular mass range of the four natural nucleobases imposes an uncertainty factor in this type of situation. One method for removing this theoretical 1 Da uncertainty factor uses amplification of a nucleic acid with one mass-tagged nucleobase and three natural nucleobases.

Addition of significant mass to one of the 4 nucleobases (dNTPs) in an amplification reaction, or in the primers themselves, will result in a significant difference in mass of the resulting amplicon (greater than 1 Da) arising from ambiguities such as the G ↔ A combined with C ↔ T event (Table 6). Thus, the same G ↔ A (-15.994) event combined with 5-Iodo-C ↔ T (-110.900) event would result in a molecular mass difference of 126.894 Da. The molecular mass of the base composition A₂₇G₃₀5-Iodo-C₂₁T₂₁ (33422.958) compared with A₂₆G₃₁5-Iodo-C₂₂T₂₀, (33549.852) provides a theoretical molecular mass difference is +126.894. The experimental error of a molecular mass measurement is not significant with regard to this molecular mass difference. Furthermore, the only base composition consistent with a measured molecular mass of the 99-mer nucleic acid is A₂₇G₃₀5-Iodo-C₂₁T₂₁. In contrast, the analogous amplification without the mass tag has 18 possible base compositions.

Table 4: Molecular Masses of Natural Nucleobases and the Mass-Modified Nucleobase 5-Iodo-C and Molecular Mass Differences Resulting from Transitions

Nucleobase	Molecular Mass	Transition	Δ Molecular Mass
------------	----------------	------------	------------------

A	313.058	A→T	-9.012
A	313.058	A→C	-24.012
A	313.058	A→5-Iodo-C	101.888
A	313.058	A→G	15.994
T	304.046	T→A	9.012
T	304.046	T→C	-15.000
T	304.046	T→5-Iodo-C	110.900
T	304.046	T→G	25.006
C	289.046	C→A	24.012
C	289.046	C→T	15.000
C	289.046	C→G	40.006
5-Iodo-C	414.946	5-Iodo-C→A	-101.888
5-Iodo-C	414.946	5-Iodo-C→T	-110.900
5-Iodo-C	414.946	5-Iodo-C→G	-85.894
G	329.052	G→A	-15.994
G	329.052	G→T	-25.006
G	329.052	G→C	-40.006
G	329.052	G→5-Iodo-C	85.894

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

The algorithm emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-detection plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of *a priori* expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database is used to define the mass base count matched filters. The database contains the sequences of known bioagents (such as caliciviruses) and may include threat organisms as well as benign background organisms. The latter is used to estimate and subtract the spectral signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a running-sum estimate of the noise

covariance. Background signal strengths are estimated and used along with the matched filters to form signatures which are then subtracted. The maximum likelihood process is applied to this “cleaned up” data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

5 The amplitudes of all base compositions of bioagent-identifying amplicons for each primer are calibrated and a final maximum likelihood amplitude estimate per organism is made based upon the multiple single primer estimates. Models of system noise are factored into this two-stage maximum likelihood calculation. The processor reports the number of molecules of each base composition contained in the spectra. The
10 quantity of amplicon corresponding to the appropriate primer set is reported as well as the quantities of primers remaining upon completion of the amplification reaction.

 Base count blurring may be carried out as follows. Electronic PCR can be conducted on nucleotide sequences of the desired bioagents to obtain the different expected base counts that could be obtained for each primer pair. See for example,
15 Schuler, *Genome Res.* 7:541-50, 1997; or the e-PCR program available from National Center for Biotechnology Information (NCBI, NIH, Bethesda, MD). In one embodiment, one or more spreadsheets from a workbook comprising a plurality of spreadsheets may be used (*e.g.*, Microsoft Excel). First, in this example, there is a worksheet with a name similar to the workbook name; this worksheet contains the raw electronic PCR data.
20 Second, there is a worksheet that contains bioagent name and base count; there is a separate record for each strain after removing sequences that are not identified with a genus and species and removing all sequences for bioagents with less than 10 strains. Third, there is a worksheet that contains the frequency of substitutions, insertions, or deletions for this primer pair. This data is generated by first creating a pivot table from
25 the data worksheet and then executing an Excel VBA macro. The macro creates a table of differences in base counts for bioagents of the same species, but different strains.

 Application of an exemplary script, involves the user defining a threshold that specifies the fraction of the strains that are represented by the reference set of base counts for each bioagent. The reference set of base counts for each bioagent may contain as
30 many different base counts as are needed to meet or exceed the threshold. The set of reference base counts is defined by selecting the most abundant strain’s base type

composition and adding it to the reference set, and then the next most abundant strain's base type composition is added until the threshold is met or exceeded.

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

Differences between a base count and a reference composition are categorized as one, two, or more substitutions, one, two, or more insertions, one, two, or more deletions, and combinations of substitutions and insertions or deletions. The different classes of nucleobase changes and their probabilities of occurrence have been delineated in U.S. Patent Application Publication No. 2004209260, incorporated herein by reference in entirety.

20

Example 6: Identification of a Calicivirus as an Adventitious Contaminant in Bioreactor Samples

This example illustrates the results obtained in an analysis of quality control samples obtained from bioreactors used in the production of recombinant proteins. The samples were prepared for analysis by first isolating viral nucleic acid according to the methods described in Example 2. The nucleic acid was amplified according to the procedures described in Example 2 using primer pair numbers VIR4959, VIR4960, VIR4961, VIR4962 and VIR4963. The amplification products were purified according to the procedures described in Example 3. The molecular masses of the products were measured by mass spectrometry as described in Example 4. The base compositions of the products were determined according to the procedures outlined in Example 5.

30

In this example, an amplification product was obtained with primer pair number VIR4962. The amplification product produced with this primer pair exhibits two peaks in the mass spectrum (see Figures 5 and 6). The masses of the two peaks are 22180.50 amu and 21707.37 amu. It should be noted that these masses are altered relative to the masses presented in, e.g., Table 8 (e.g., Forward Strand Mass), because a mass-tagged dNTP (i.e., $^{13}\text{C}_{10}$ -deoxyguanosine triphosphate) was utilized in the amplification reaction. These molecular masses correspond to opposing strands of an amplification product with matched base compositions of $\text{A}_{19}\text{G}_{20}\text{C}_9\text{T}_{23}$ and $\text{A}_{23}\text{G}_9\text{C}_{20}\text{T}_{19}$. The base compositions were compared with a database containing base compositions of all of the known calicivirus identifying amplicons defined by primer pair number VIR4962. An example of a portion of a base composition database is shown in Tables 5 to 12 (reverse strand masses and base compositions are omitted for clarity). The portion of the database corresponding to VIR4962 is shown in Table 8.

Table 5: Base Composition Database for Amplicons Defined by Primer Pair No. VIR4959

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
San Miguel sea lion virus	serotype 4	gi 558443	34184.6936	A33 G23 C25 T30
Primate calicivirus	NoStrain_7 0354	gi 664818	34225.7327	A37 G22 C23 T29
Bovine calicivirus	NoStrain_7 0353	gi 664816	34193.7051	A34 G23 C25 T29
Rabbit vesivirus	NoStrain_7 0358	gi 66794786	34193.7051	A34 G23 C25 T29
Rabbit vesivirus	NoStrain_7 0361	gi 118430707	34193.7051	A34 G23 C25 T29
Steller sea lion vesivirus	V810	gi 186898657	34184.6936	A33 G23 C25 T30
Steller sea lion vesivirus	V1415	gi 186898661	34171.7221	A36 G22 C26 T27
VESV-like calicivirus	Pan-1	gi 3661574	34103.686	A34 G20 C26 T31
Vesicular exanthema of swine virus	A48	gi 10314005	34240.6946	A32 G25 C24 T30
San Miguel sea lion	4	gi 608158	34184.6936	A33 G23 C25 T30

virus				
San Miguel sea lion virus	1	gi 608162	34240.6946	A32 G25 C24 T30
San Miguel sea lion virus	serotype 2	gi 664794	34194.7381	A38 G21 C24 T28
San Miguel sea lion virus	serotype 5	gi 664796	34146.7156	A36 G21 C26 T28
Vesicular exanthema of swine virus	B51	gi 113471010	34303.679	A29 G28 C23 T31
Vesicular exanthema of swine virus	C52	gi 113471013	34303.679	A29 G28 C23 T31
Vesicular exanthema of swine virus	Sec.3-54	gi 113471016	34334.6736	A28 G29 C22 T32
Vesicular exanthema of swine virus	FS 332	gi 113471025	34232.6783	A30 G26 C25 T30
San Miguel sea lion virus	serotype 7	gi 664800	34193.7051	A34 G23 C25 T29
San Miguel sea lion virus	serotype 13	gi 664802	34147.7109	A35 G22 C27 T27
San Miguel sea lion virus	serotype 14	gi 664804	34208.7048	A34 G23 C24 T30
San Miguel sea lion virus	serotype 1	gi 10141004	34226.7279	A36 G23 C24 T28
Vesicular exanthema of swine virus	A48	gi 10141008	34240.6946	A32 G25 C24 T30
Vesicular exanthema of swine virus	A48	gi 608164	34240.6946	A32 G25 C24 T30
Vesicular exanthema of swine virus	A48	gi 664808	34240.6946	A32 G25 C24 T30
Vesicular exanthema of swine virus	C52	gi 664810	34193.7051	A34 G23 C25 T29
San Miguel sea lion virus	serotype 6	gi 664798	34176.6772	A31 G24 C26 T30
San Miguel sea lion virus	2MR	gi 113471001	34194.7381	A38 G21 C24 T28
San Miguel sea lion virus	15FT	gi 113471004	34184.6936	A33 G23 C25 T30

San Miguel sea lion virus	V-31-77	gi 113471007	34184.6936	A33 G23 C25 T30
Walrus calicivirus	NoStrain_7 0357	gi 27881469	34222.6715	A30 G25 C24 T32
Walrus calicivirus	NoStrain_7 0356	gi 11992268	34222.6715	A30 G25 C24 T32
San Miguel sea lion virus	CSL-461	gi 113471022	34162.7106	A35 G22 C26 T28
Skunk calicivirus	4-1L	gi 608146	34242.7228	A35 G24 C24 T28
Skunk calicivirus	4-2S	gi 608150	34184.6936	A33 G23 C25 T30
Skunk calicivirus	7-2&3L	gi 608154	34226.7279	A36 G23 C24 T28
San Miguel sea lion virus	2012181	gi 113471019	34208.7048	A34 G23 C24 T30
Skunk calicivirus	NoStrain_7 0355	gi 664820	34242.7228	A35 G24 C24 T28
Steller sea lion vesivirus	V1415	gi 194268060	34171.7221	A36 G22 C26 T27

Table 6: Base Composition Database for Amplicons Defined by Primer Pair No. VIR4960

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
Bovine calicivirus	NoStrain_7 0353	gi 664816	42024.9413	A33 G30 C41 T33
Primate calicivirus	NoStrain_7 0354	gi 664818	41847.8888	A30 G27 C46 T34
Rabbit vesivirus	NoStrain_7 0358	gi 66794786	42024.9413	A33 G30 C41 T33
Rabbit vesivirus	NoStrain_7 0361	gi 118430707	42024.9413	A33 G30 C41 T33
San Miguel sea lion virus	serotype 4	gi 558443	41945.9242	A32 G29 C44 T32
San Miguel sea lion virus	4	gi 608158	41970.9307	A32 G30 C44 T31
San Miguel sea lion virus	serotype 2	gi 664794	41877.8881	A30 G27 C44 T36
San Miguel sea lion virus	serotype 5	gi 664796	41878.8834	A29 G28 C45 T35
San Miguel sea lion virus	serotype 6	gi 664798	41904.9228	A33 G27 C44 T33

virus				
San Miguel sea lion virus	serotype 7	gi 664800	42015.9297	A32 G30 C41 T34
San Miguel sea lion virus	1	gi 608162	41912.9014	A30 G29 C45 T33
San Miguel sea lion virus	serotype 13	gi 664802	41887.8949	A30 G28 C45 T34
San Miguel sea lion virus	serotype 14	gi 664804	41954.9358	A33 G29 C44 T31
San Miguel sea lion virus	serotype 1	gi 10141004	41895.9112	A32 G27 C44 T34
San Miguel sea lion virus	15FT	gi 113471004	41954.9358	A33 G29 C44 T31
San Miguel sea lion virus	V-31-77	gi 113471007	41990.9232	A32 G29 C41 T35
San Miguel sea lion virus	2012181	gi 113471019	42054.9406	A33 G30 C39 T35
San Miguel sea lion virus	CSL-461	gi 113471022	41872.8953	A30 G28 C46 T33
San Miguel sea lion virus	2MR	gi 113471001	41877.8881	A30 G27 C44 T36
Skunk calicivirus	4-1L	gi 608146	41895.9112	A32 G27 C44 T34
Skunk calicivirus	4-2S	gi 608150	41970.9307	A32 G30 C44 T31
Skunk calicivirus	7-2&3L	gi 608154	41895.9112	A32 G27 C44 T34
Steller sea lion vesivirus	V1415	gi 194268060	42077.9189	A30 G32 C39 T36
Steller sea lion vesivirus	V810	gi 186898657	42028.9389	A34 G28 C38 T37
Steller sea lion vesivirus	V1415	gi 186898661	42077.9189	A30 G32 C39 T36
VESV-like calicivirus	Pan-1	gi 3661574	41976.9565	A36 G27 C41 T33
Skunk calicivirus	NoStrain_70355	gi 664820	41895.9112	A32 G27 C44 T34
Vesicular exanthema of	A48	gi 10314005	42017.9202	A30 G32 C43 T32

swine virus				
Vesicular exanthema of swine virus	B51	gi 113471010	42023.946	A34 G29 C40 T34
Vesicular exanthema of swine virus	C52	gi 113471013	42023.946	A34 G29 C40 T34
Vesicular exanthema of swine virus	FS 332	gi 113471025	42038.9457	A34 G29 C39 T35
Vesicular exanthema of swine virus	A48	gi 10141008	42017.9202	A30 G32 C43 T32
Vesicular exanthema of swine virus	A48	gi 608164	42017.9202	A30 G32 C43 T32
Vesicular exanthema of swine virus	A48	gi 664808	42017.9202	A30 G32 C43 T32
Vesicular exanthema of swine virus	Sec.3-54	gi 113471016	42008.9463	A34 G29 C41 T33
Vesicular exanthema of swine virus	C52	gi 664810	42009.9416	A33 G30 C42 T32
Walrus calicivirus	NoStrain_7 0357	gi 27881469	41983.9399	A34 G28 C41 T34
Walrus calicivirus	NoStrain_7 0356	gi 11992268	41983.9399	A34 G28 C41 T34

Table 7: Base Composition Database for Amplicons Defined by Primer Pair No.

VIR4961

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
Steller sea lion vesivirus	V810	gi 186898657	35805.9744	A36 G18 C32 T31
Steller sea lion vesivirus	V1415	gi 194268060	35828.9527	A33 G20 C32 T32
Vesicular exanthema of swine virus	A48	gi 10314005	35781.9632	A35 G18 C33 T31
Vesicular exanthema of swine virus	B51	gi 113471010	35812.9578	A34 G19 C32 T32
Steller sea lion vesivirus	V1415	gi 186898661	35828.9527	A33 G20 C32 T32

Vesicular exanthema of swine virus	C52	gi 113471013	35812.9578	A34 G19 C32 T32
VESV-like calicivirus	Pan-1	gi 3661574	35790.9748	A36 G18 C33 T30
Vesicular exanthema of swine virus	Sec.3-54	gi 113471016	35836.969	A35 G19 C31 T32
Vesicular exanthema of swine virus	FS 332	gi 113471025	35781.9632	A35 G18 C33 T31
San Miguel sea lion virus	SMSV-15	gi 1929484	35766.9635	A35 G18 C34 T30
San Miguel sea lion virus	SMSV-5 Hom-1	gi 4097466	35837.9642	A34 G20 C32 T31
San Miguel sea lion virus	serotype 1	gi 10141004	35836.969	A35 G19 C31 T32
San Miguel sea lion virus	2MR	gi 113471001	35792.9653	A34 G20 C35 T28
San Miguel sea lion virus	15FT	gi 113471004	35781.9632	A35 G18 C33 T31
San Miguel sea lion virus	V-31-77	gi 113471007	35845.9806	A36 G19 C31 T31
San Miguel sea lion virus	2012181	gi 113471019	35751.9638	A35 G18 C35 T29
San Miguel sea lion virus	CSL-461	gi 113471022	35806.9697	A35 G19 C33 T30
Vesicular exanthema of swine virus	A48	gi 10141008	35781.9632	A35 G18 C33 T31
Walrus calicivirus	NoStrain_7 0357	gi 27881469	35811.9625	A35 G18 C31 T33
Walrus calicivirus	NoStrain_7 0356	gi 11992268	35811.9625	A35 G18 C31 T33
Cetacean calicivirus	Tur-1	gi 1929476	35797.9581	A34 G19 C33 T31
Primate calicivirus	Pan-1	gi 1929478	35790.9748	A36 G18 C33 T30
Rabbit vesivirus	NoStrain_7 0358	gi 66794786	35812.9578	A34 G19 C32 T32
Rabbit vesivirus	NoStrain_7 0361	gi 118430707	35812.9578	A34 G19 C32 T32
Bovine Calicivirus	Bos-2	gi 18032036	35797.9581	A34 G19 C33 T31

Reptile calicivirus	Cro-1	gi 1929480	35773.9469	A33 G19 C34 T31
San Miguel sea lion virus	serotype 4	gi 558443	35781.9632	A35 G18 C33 T31
San Miguel sea lion virus	SMSV-13	gi 1929482	35806.9697	A35 G19 C33 T30
San Miguel sea lion virus	SMSV-16	gi 1929486	35811.9625	A35 G18 C31 T33
San Miguel sea lion virus	SMSV-9	gi 1929488	35790.9748	A36 G18 C33 T30
San Miguel sea lion virus	SMSV-5	gi 1929490	35806.9697	A35 G19 C33 T30
San Miguel sea lion virus	SMSV-17	gi 1929492	35774.9798	A37 G17 C33 T30

Table 8: Base Composition Database for Amplicons Defined by Primer Pair No.

VIR4962

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
San Miguel sea lion virus	serotype 14	gi 664804	22053.649	A17 G21 C10 T23
Calicivirus isolate 2117	2117	gi 34329322	22061.6653	A19 G20 C9 T23
Canine calicivirus	No. 48	gi 3133310	22011.6524	A19 G18 C9 T25
Canine calicivirus	NoStrain_1 537	gi 18149157	22011.6524	A19 G18 C9 T25
Canine calicivirus	NoStrain_1 545	gi 27881465	22011.6524	A19 G18 C9 T25
Mink calicivirus	MCV/9/1980 /US	gi 13310423	21982.6483	A18 G19 C12 T22
Mink calicivirus	MCV/13/198 0/US	gi 13310425	21982.6483	A18 G19 C12 T22
Mink calicivirus	MCV/20/198 0/US	gi 13310427	21982.6483	A18 G19 C12 T22

5 **Table 9: Base Composition Database for Amplicons Defined by Primer Pair No.**

VIR4963

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
--------------	---------------	--------------------------	----------------------------	--

San Miguel sea lion virus	SMSV-16	gi 1929486	31494.2441	A31 G14 C26 T32
Walrus calicivirus	NoStrain_70357	gi 27881469	31494.2441	A31 G14 C26 T32
Walrus calicivirus	NoStrain_70356	gi 11992268	31494.2441	A31 G14 C26 T32
Calicivirus isolate 2117	2117	gi 34329322	31445.2264	A30 G13 C27 T33
Canine calicivirus	No. 48	gi 3133310	31524.2057	A26 G17 C26 T34
Canine calicivirus	NoStrain_1537	gi 18149157	31524.2057	A26 G17 C26 T34
Canine calicivirus	NoStrain_1545	gi 27881465	31524.2057	A26 G17 C26 T34

Table 10: Base Composition Database for Amplicons Defined by Primer Pair No. VIR4964

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
Feline calicivirus	Urbana	gi 28212283	20578.3719	A11 G17 C17 T22
Feline calicivirus	F9	gi 323877	20578.3719	A11 G17 C17 T22
Feline calicivirus	CFI/68 FIV	gi 323880	20578.3719	A11 G17 C17 T22
Feline calicivirus	Urbana	gi 845310	20578.3719	A11 G17 C17 T22
Feline calicivirus	CFI/68	gi 3056875	20578.3719	A11 G17 C17 T22
Feline calicivirus	F65	gi 5706691	20578.3719	A11 G17 C17 T22
Feline calicivirus	FCV2024	gi 21359681	20563.3722	A11 G17 C18 T21
Feline calicivirus	UTCVM-NH1	gi 49458053	20563.3722	A11 G17 C18 T21
Feline calicivirus	UTCVM-NH2	gi 49458057	20578.3719	A11 G17 C17 T22
Feline calicivirus	UTCVM-NH3	gi 49458061	20563.3722	A11 G17 C18 T21
Feline calicivirus	UTCVM-H1	gi 49458065	20563.3722	A11 G17 C18 T21
Feline calicivirus	UTCVM-H2	gi 49458069	20552.3701	A12 G15 C16 T24

Feline calicivirus	USDA	gi 49458073	20563.3722	A11 G17 C18 T21
Feline calicivirus	FCV/DD/200 6/GE	gi 90019597	20523.3661	A11 G16 C19 T21
Feline calicivirus	F4	gi 98986307	20578.3719	A11 G17 C17 T22

Table 11: Base Composition Database for Amplicons Defined by Primer Pair No.**VIR4965**

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
Feline calicivirus	FCV2024	gi 21359681	25119.137	A17 G25 C15 T24
Feline calicivirus	UTCVM-NH1	gi 49458053	25129.1438	A17 G26 C16 T22
Feline calicivirus	UTCVM-NH2	gi 49458057	25090.1329	A16 G26 C18 T21
Feline calicivirus	UTCVM-NH3	gi 49458061	25097.154	A19 G24 C16 T22
Feline calicivirus	UTCVM-H1	gi 49458065	25137.1601	A19 G25 C15 T22
Feline calicivirus	USDA	gi 49458073	25119.137	A17 G25 C15 T24
Feline calicivirus	FCV/DD/200 6/GE	gi 90019597	25144.1435	A17 G26 C15 T23
Feline calicivirus	F4	gi 98986307	25112.1536	A19 G24 C15 T23
Feline calicivirus	FCV-127	gi 115178338	25098.1492	A18 G25 C17 T21
Feline calicivirus	UTCVM-H2	gi 49458069	25130.1391	A16 G27 C17 T21
Feline calicivirus	FCV-131	gi 115178341	25137.1601	A19 G25 C15 T22
Feline calicivirus	FCV-796	gi 115178344	25114.1441	A17 G26 C17 T21
Feline calicivirus	FCV-Deuce	gi 115178347	25113.1489	A18 G25 C16 T22
Feline calicivirus	FCV- Georgie	gi 115178353	25144.1435	A17 G26 C15 T23
Feline calicivirus	FCV-Jengo	gi 115178359	25129.1438	A17 G26 C16 T22
Feline calicivirus	VS-FCV-Ari	gi 115178362	25188.1683	A18 G28 C16 T19
Feline calicivirus	FCV-Kaos	gi 115178365	25129.1438	A17 G26 C16 T22

Feline calicivirus	FCV-5	gi 115178350	25104.1373	A17 G25 C16 T23
Feline calicivirus	FCV-George Walder	gi 227859367	25129.1438	A17 G26 C16 T22
Feline calicivirus	Urbana	gi 28212283	25105.1326	A16 G26 C17 T22
Feline calicivirus	F9	gi 59260	25090.1329	A16 G26 C18 T21
Feline calicivirus	F4	gi 221264	25112.1536	A19 G24 C15 T23
Feline calicivirus	F9	gi 323877	25090.1329	A16 G26 C18 T21
Feline calicivirus	CFI/68 FIV	gi 323880	25139.1506	A17 G27 C17 T20
Feline calicivirus	Urbana	gi 845310	25105.1326	A16 G26 C17 T22
Feline calicivirus	CFI/68	gi 3056875	25139.1506	A17 G27 C17 T20
Feline calicivirus	F65	gi 5706691	25096.1587	A20 G23 C15 T23

Table 12: Base Composition Database for Amplicons Defined by Primer Pair No. VIR4966

Virus	Strain	GenBank gi Number	Forward Strand Mass	Forward Strand Base Composition
Calicivirus isolate TCG	TCG 14	gi 60418038	31752.3546	A34 G33 C22 T13
Calicivirus isolate TCG	TCG 14	gi 60677687	31752.3546	A34 G33 C22 T13
Calicivirus strain NB	NB	gi 21699776	31782.354	A34 G33 C20 T15
Newbury agent 1	NoStrain_7 0359	gi 67003918	31736.3597	A35 G32 C22 T13
Newbury agent 1	NoStrain_7 0360	gi 90403548	31736.3597	A35 G32 C22 T13
Calicivirus strain NB	NB	gi 21655150	31782.354	A34 G33 C20 T15

5 The experimentally determined base composition of $[A_{19}G_{20}C_9T_{23}]$ matches the base composition of the forward strand of the amplicon of calicivirus isolate 2117 virus (highlighted in **bold** in Table 8). As noted herein, the mass of this are altered relative to the masses presented in, e.g., Table 8 (e.g., Forward Strand Mass), because a mass-tagged

dNTP (i.e., $^{13}\text{C}_{10}$ -deoxyguanosine triphosphate) was utilized in the amplification reaction. This result identifies the unknown virus in the clinical sample as the calicivirus isolate 2117. This is a useful result because appropriate action may be taken to identify the probable source of the contamination of the bioreactor when the identity of the virus is known. When the source is identified, appropriate preventative action may be taken.

The skilled person will recognize that additional caliciviruses may be similarly identified using one or more of the primer pairs of Table 1 and the masses and base compositions of corresponding amplicons provided in Tables 5 to 12.

10 **Example 7: Identification of a Single Nucleotide Polymorphism in a Known Calicivirus Strain**

This example illustrates a case where an analysis similar to that described in Example 6 provides an amplification product produced by primer pair VIR4961 which has an experimentally-determined forward strand base composition of $\text{A}_{33}\text{G}_{17}\text{C}_{32}\text{T}_{35}$ and reverse strand base composition of $\text{A}_{35}\text{G}_{32}\text{C}_{17}\text{T}_{33}$ (see Figures 7 and 8). The forward strand base composition $\text{A}_6\text{G}_{18}\text{C}_{14}\text{T}_3$ does not appear in the database section of Table 7. It is noted however that the forward strand base composition differs from that of calicivirus isolate 2117 by a $\text{A} \rightarrow \text{T}$ single nucleotide polymorphism. This result indicates that the analysis has characterized a new calicivirus strain. It is advantageous to make a new entry in the database to reflect this new strain which has been characterized by its novel amplicon.

A future analysis of a different sample according to the methods described above with primer pair number VIR4961 which provides a forward base composition of $\text{A}_{35}\text{G}_{32}\text{C}_{17}\text{T}_{33}$ would then identify the presence of this newly discovered calicivirus strain.

This example thus illustrates characterization of a new calicivirus strain.

Various modifications of the invention, in addition to those described herein, will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. Each reference (including, but not limited to, journal articles, U.S. and non-U.S. patents, patent application publications, international patent application publications, gene bank

accession numbers, internet web sites, and the like) cited in the present application is incorporated herein by reference in its entirety.

CLAIMS

We Claim

1. A purified oligonucleotide primer pair for identifying a known calicivirus or
5 characterizing a previously unknown calicivirus, said primer pair comprising a forward primer and a reverse primer, each configured to hybridize to nucleic acid of two or more different caliciviruses in a nucleic acid amplification reaction which produces an amplification product between about 29 to about 200 nucleobases in length, said amplification product comprising portions corresponding to a forward primer hybridization region, a reverse primer hybridization
10 region and an intervening region having a base composition which varies among amplification products produced from nucleic acid of said two or more different caliciviruses, said base composition of said intervening region providing a means for identifying said previously known calicivirus or characterizing said previously unknown calicivirus.
- 15 2. The primer pair of claim 1 wherein each member of said primer pair has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.
3. The primer pair of claim 2 wherein said forward primer and said reverse primer are about
20 14 to about 40 nucleobases in length.
4. The primer pair of claim 2, wherein said base composition identifies said previously known calicivirus or characterizes said previously unknown calicivirus at the species level or the sub-species level.
25
5. The primer pair of claim 2, wherein said forward primer or said reverse primer or both further comprise a non-templated thymidine residue on the 5'-end.
6. The primer pair of claim 2, wherein said forward primer or said reverse primer or both
30 further comprise at least one molecular mass modifying tag.
7. The primer pair of claim 2, wherein said forward primer or said reverse primer or both further comprise at least one modified nucleobase.

8. The primer pair of claim 7, wherein said modified nucleobase is 5-propynyluracil or 5-propynylcytosine.

5 9. The primer pair of claim 7, wherein said modified nucleobase is a mass-modified nucleobase.

10. The primer pair of claim 9, wherein said mass-modified nucleobase is 5-iodo-cytosine.

10 11. The primer pair of claim 7, wherein said modified nucleobase is a universal nucleobase.

12. The primer pair of claim 11, wherein said universal nucleobase is inosine.

13. An isolated amplification product for identification of a known calicivirus or
15 characterizing a previously unknown calicivirus, said amplification product produced by a process comprising:

a) amplifying nucleic acid of a calicivirus in a reaction mixture comprising a primer pair, said primer pair comprising a forward primer and a reverse primer, each configured to hybridize to nucleic acid of two or more different caliciviruses in a nucleic acid amplification reaction, said
20 amplification product having a length of about 29 to about 200 nucleobases and comprising portions corresponding to a forward primer hybridization region, a reverse primer hybridization region and an intervening region having a base composition which varies among amplification products produced from nucleic acid of said two or more different caliciviruses, said base
25 composition of said intervening region providing a means for identifying said previously known calicivirus or characterizing said previously unknown calicivirus; and

b) isolating said amplification product from said reaction mixture.

14. The amplification product of claim 13 wherein step b) is performed using an anion exchange resin linked to a magnetic bead.

30

15. The amplification product of claim 13 wherein each member of said primer pair has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.

16. The amplification product of claim 15 wherein said forward primer and said reverse primer are about 14 to about 40 nucleobases in length.

17. The amplification product of claim 15, wherein said base composition identifies said
5 previously known calicivirus or characterizes said previously unknown calicivirus at the species level or the sub-species level.

18. The amplification product of claim 15, wherein said forward primer or said reverse primer or both further comprise a non-templated thymidine residue on the 5'-end.
10

19. The amplification product of claim 15, wherein said forward primer or said reverse primer or both further comprise at least one molecular mass modifying tag.

20. The amplification product of claim 15, wherein said forward primer or said reverse
15 primer or both further comprise at least one modified nucleobase.

21. The amplification product of claim 20, wherein said modified nucleobase is 5-propynyluracil or 5-propynylcytosine.

22. The amplification product of claim 20, wherein said modified nucleobase is a mass-modified nucleobase.
20

23. The amplification product of claim 22, wherein said mass-modified nucleobase is 5-iodo-
25 cytosine.

24. The amplification product of claim 20, wherein said modified nucleobase is a universal nucleobase.

25. The amplification product of claim 24, wherein said universal nucleobase is inosine.
30

26. The amplification product of claim 13 wherein said known calicivirus is selected from the group consisting of: Bovine calicivirus, Calicivirus isolate 2117, Calicivirus isolate TCG, Calicivirus pig/AB104/CAN, Calicivirus pig/AB90/CAN, Calicivirus pig/F15-10/CAN, Calicivirus strain NB, Canine calicivirus, Cetacean calicivirus, Feline calicivirus, Mink

calicivirus, Newbury agent 1, Primate calicivirus, Rabbit vesivirus, Reptile calicivirus, San Miguel sea lion virus, Skunk calicivirus, Steller sea lion vesivirus, Tulane virus, Vesicular exanthema of swine virus, VESV-like calicivirus, and Walrus calicivirus.

- 5 27. A method for identifying a known calicivirus or characterizing a previously unknown calicivirus in a sample, said method comprising:
- (a) obtaining an amplification product by amplifying one or more nucleic acids of one or more caliciviruses in said sample using the primer pair of claim 1;
 - (b) measuring the molecular mass of one or both strands of said amplification
10 product;
 - (c) comparing said molecular mass to a plurality of database-stored molecular masses of strands of amplification products of known caliciviruses; and
 - d) identifying a match between said molecular mass and at least one of said database-stored molecular masses of amplification products, thereby identifying said known
15 calicivirus or, alternatively, failing to identify a match between said molecular mass and at least one of said database-stored molecular masses, thereby characterizing a previously unknown calicivirus.
28. The method of claim 27 wherein each member of said primer pair has at least 70%
20 sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.
29. The method of claim 27 wherein said molecular mass is determined by mass
25 spectrometry.
30. The method of claim 27 wherein said known calicivirus is selected from the group consisting of: Bovine calicivirus, Calicivirus isolate 2117, Calicivirus isolate TCG, Calicivirus pig/AB104/CAN, Calicivirus pig/AB90/CAN, Calicivirus pig/F15-10/CAN, Calicivirus strain NB, Canine calicivirus, Cetacean calicivirus, Feline calicivirus, Mink calicivirus, Newbury agent
30 1, Primate calicivirus, Rabbit vesivirus, Reptile calicivirus, San Miguel sea lion virus, Skunk calicivirus, Steller sea lion vesivirus, Tulane virus, Vesicular exanthema of swine virus, VESV-like calicivirus, and Walrus calicivirus.

31. A method for identifying a known calicivirus or characterizing a previously unknown calicivirus in a sample, said method comprising:

(a) obtaining an amplification product by amplifying one or more nucleic acids of one or more caliciviruses in said sample using the purified primer pair of claim 1;

5 (b) measuring the molecular mass of one or both strands of said amplification product;

(c) determining the base composition of said amplification product from said molecular mass;

10 (d) comparing said base composition to a plurality of database-stored base compositions of strands of amplification products of known caliciviruses; and

(e) identifying a match between said base composition and at least one of said database-stored molecular masses of amplification products, thereby identifying said known calicivirus or, alternatively, failing to identify a match between said base composition and at least one of said database-stored base compositions, thereby characterizing a previously unknown calicivirus.

32. The method of claim 31 wherein each member of said primer pair has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.

33. The method of claim 31 wherein said molecular mass is determined by mass spectrometry.

34. The method of claim 31 wherein step (e) identifies said calicivirus as a member of a plurality of caliciviruses and said method further comprises repeating steps (a) to (e) using one or more additional primer pairs as defined in claim 1, wherein one or more repetitions of step (e) with said one or more additional primer pairs identifies or characterizes said calicivirus as a unique calicivirus.

35. The method of claim 34 wherein each member of said one or more additional primer pairs has at least 70% sequence identity with a corresponding member of a primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.

36. The method of claim 31 wherein said molecular mass is determined by mass spectrometry.

5 37. The method of claim 31 wherein said known calicivirus is selected from the group consisting of: Bovine calicivirus, Calicivirus isolate 2117, Calicivirus isolate TCG, Calicivirus pig/AB104/CAN, Calicivirus pig/AB90/CAN, Calicivirus pig/F15-10/CAN, Calicivirus strain NB, Canine calicivirus, Cetacean calicivirus, Feline calicivirus, Mink calicivirus, Newbury agent 1, Primate calicivirus, Rabbit vesivirus, Reptile calicivirus, San Miguel sea lion virus, Skunk
10 calicivirus, Steller sea lion vesivirus, Tulane virus, Vesicular exanthema of swine virus, VESV-like calicivirus, and Walrus calicivirus.

38. A kit comprising one or more purified primer pairs for identifying a known calicivirus or characterizing a previously unknown calicivirus in a nucleic acid sample, each member of said
15 one or more primer pairs having at least 70% sequence identity with a corresponding member of one or more primer pairs selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.

39. The kit of claim 38 further comprising a reverse transcriptase and a polymerase.

20

40. The kit of claim 39 further comprising deoxynucleotide triphosphates.

41. The kit of claim 40 wherein one or more of said deoxynucleotide triphosphates is ¹³C-enriched.

25

42. A system, comprising:

(a) a mass spectrometer configured to detect one or more molecular masses of an amplification product of claim 13;

(b) a database of known molecular masses and/or known base compositions of
30 amplification products of known caliciviruses; and

(b) a controller operably connected to said mass spectrometer and to said database said controller configured to match said molecular masses of said amplification product with a measured or calculated molecular mass of a corresponding amplification product of a known calicivirus.

43. The system of claim 41 wherein said database of known molecular masses and/or known base compositions of amplification products of known caliciviruses includes amplification products defined by one or more primer pairs wherein each member of said one or more primer pairs has at least 70% sequence identity with a corresponding member of a corresponding primer pair selected from the group consisting of: SEQ ID NOs: 11:7, 13:8, 4:12, 14:15, 1:16, 2:9, 17:5, 10:18, and 6:3.

44. The system of claim 41 wherein said known caliciviruses are selected from the group consisting of: Bovine calicivirus, Calicivirus isolate 2117, Calicivirus isolate TCG, Calicivirus pig/AB104/CAN, Calicivirus pig/AB90/CAN, Calicivirus pig/F15-10/CAN, Calicivirus strain NB, Canine calicivirus, Cetacean calicivirus, Feline calicivirus, Mink calicivirus, Newbury agent 1, Primate calicivirus, Rabbit vesivirus, Reptile calicivirus, San Miguel sea lion virus, Skunk calicivirus, Steller sea lion vesivirus, Tulane virus, Vesicular exanthema of swine virus, VESV-like calicivirus, and Walrus calicivirus.

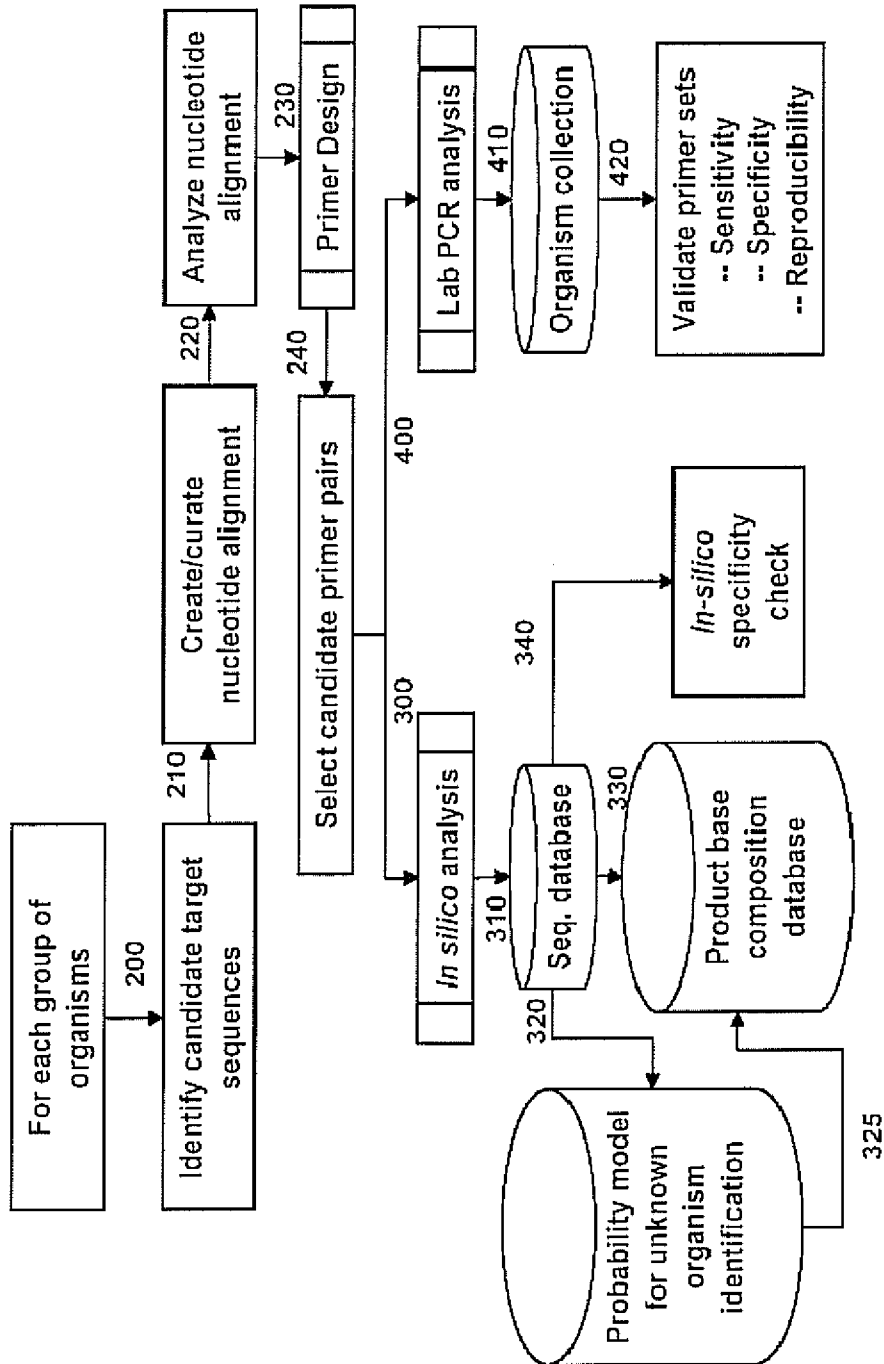


Fig. 1

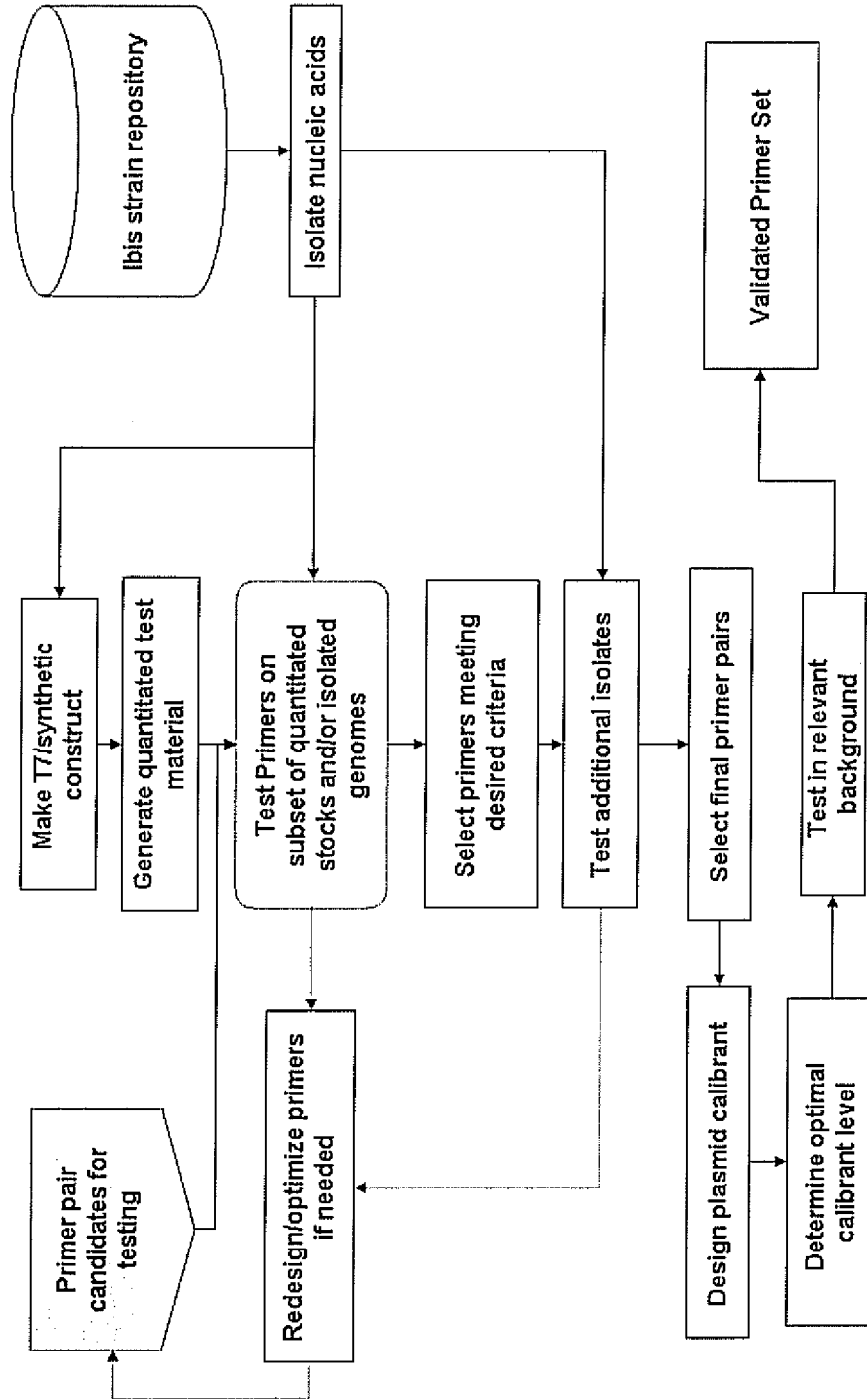


Fig. 2

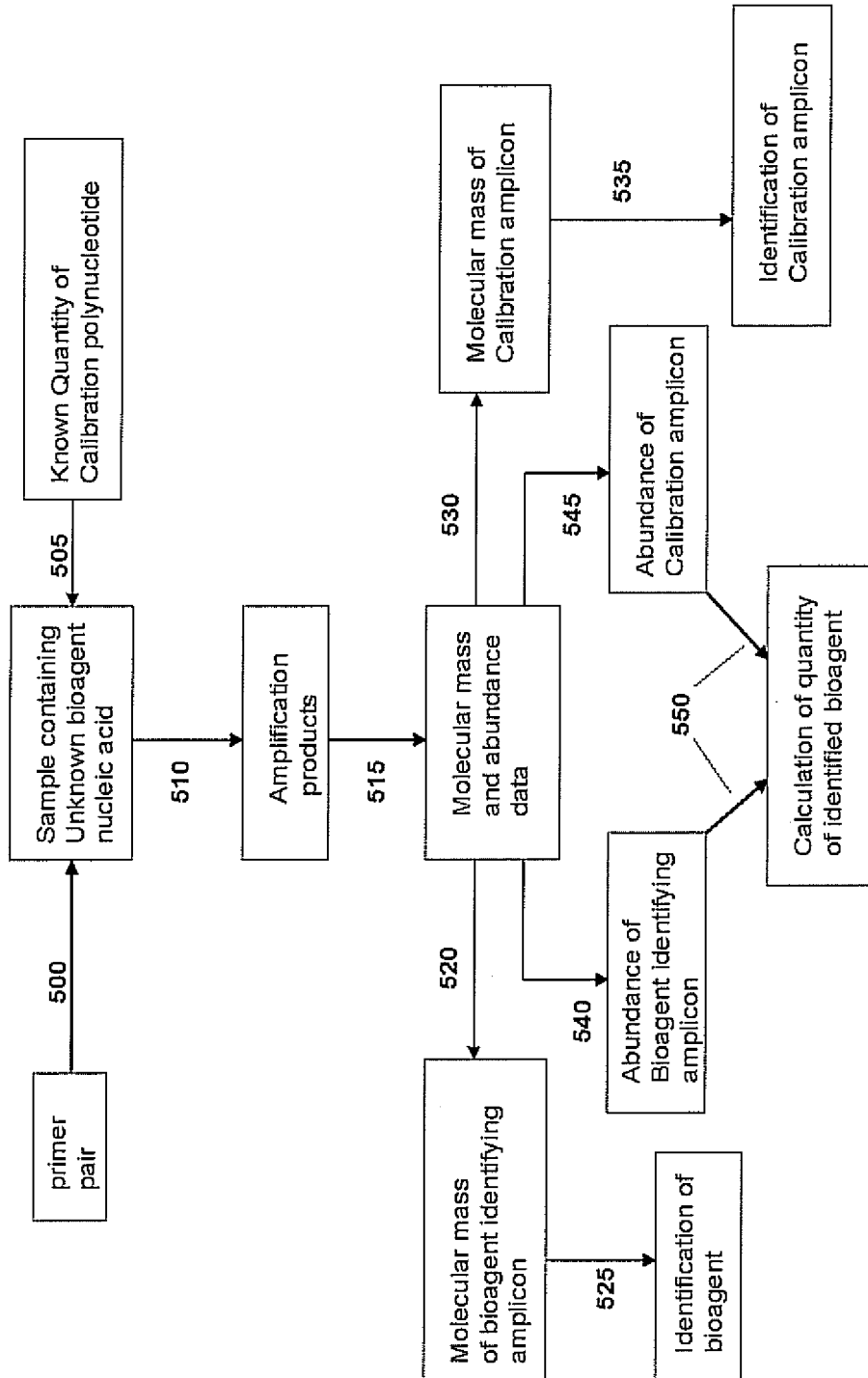


Fig. 3

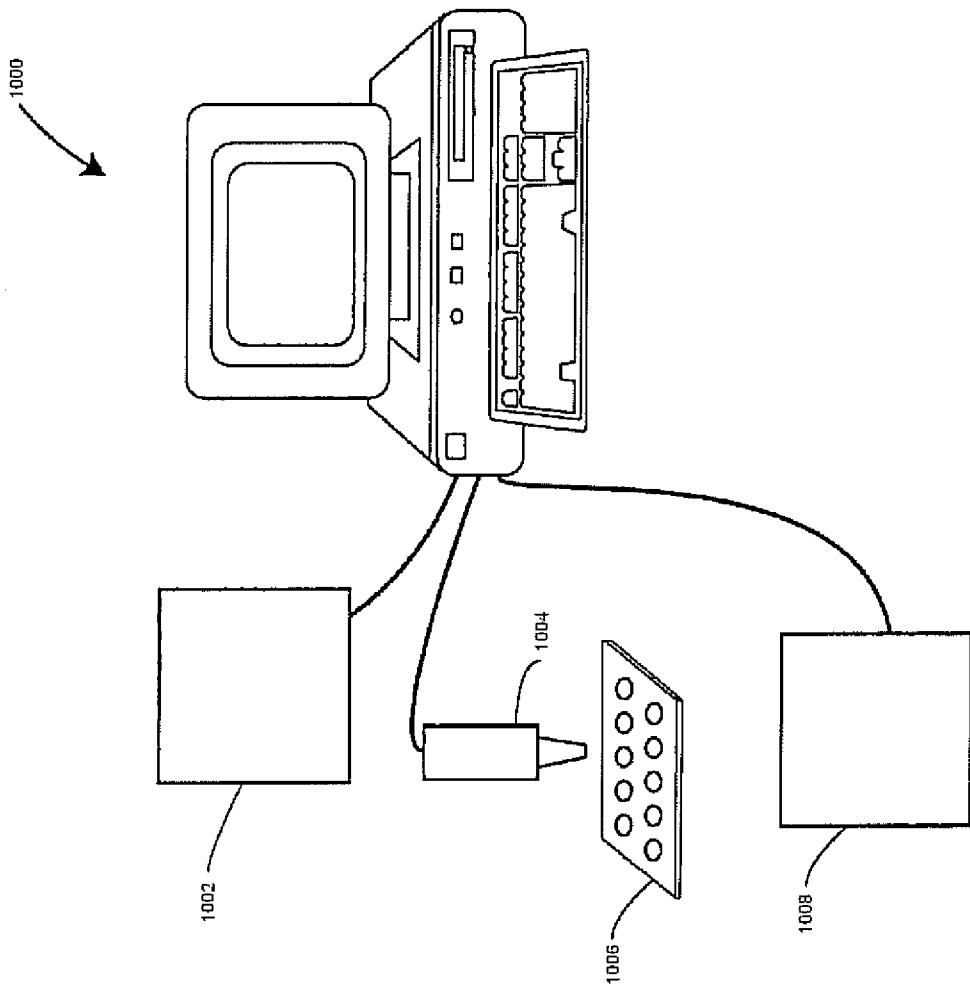


Fig. 4

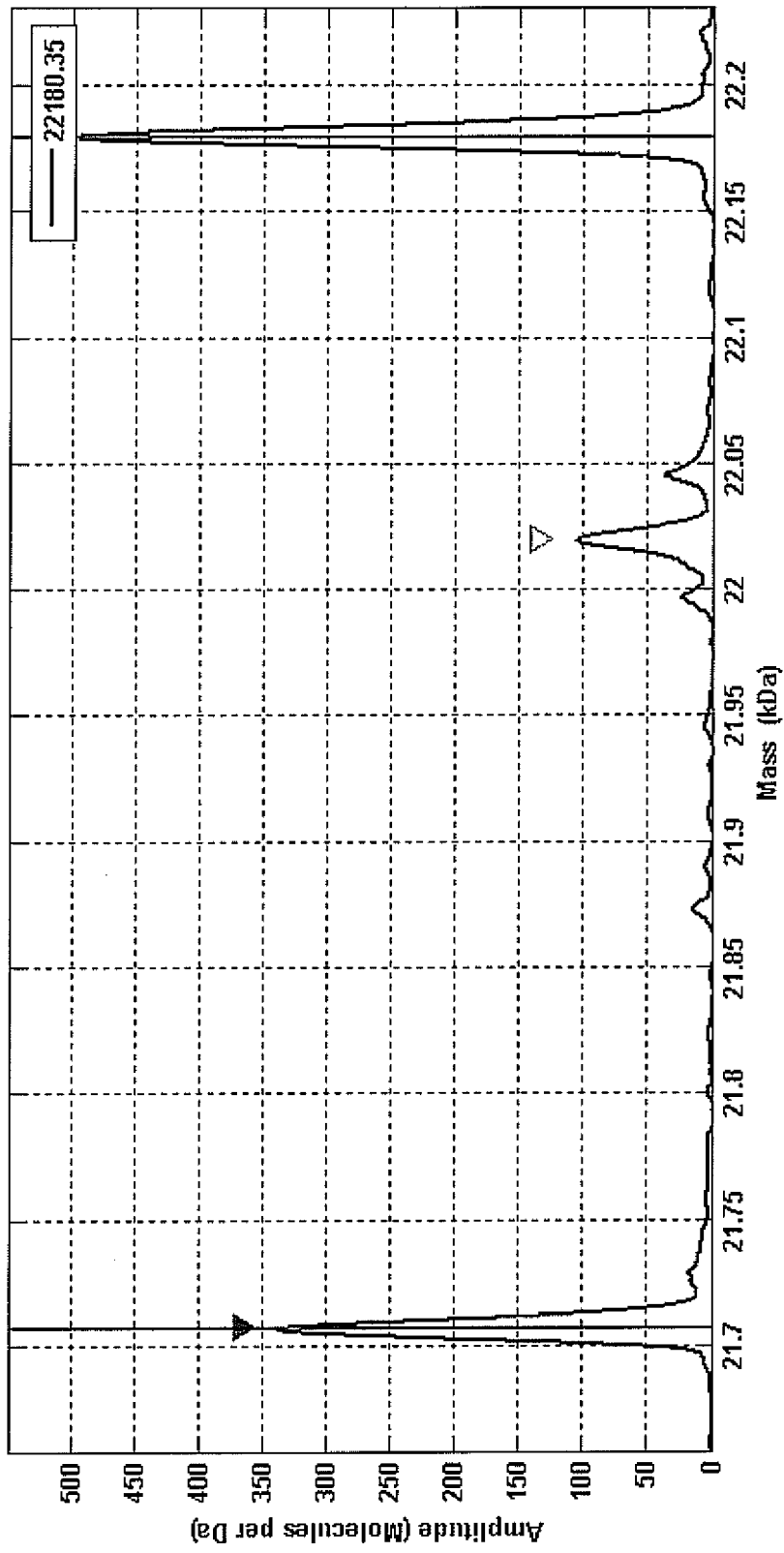


Fig. 5

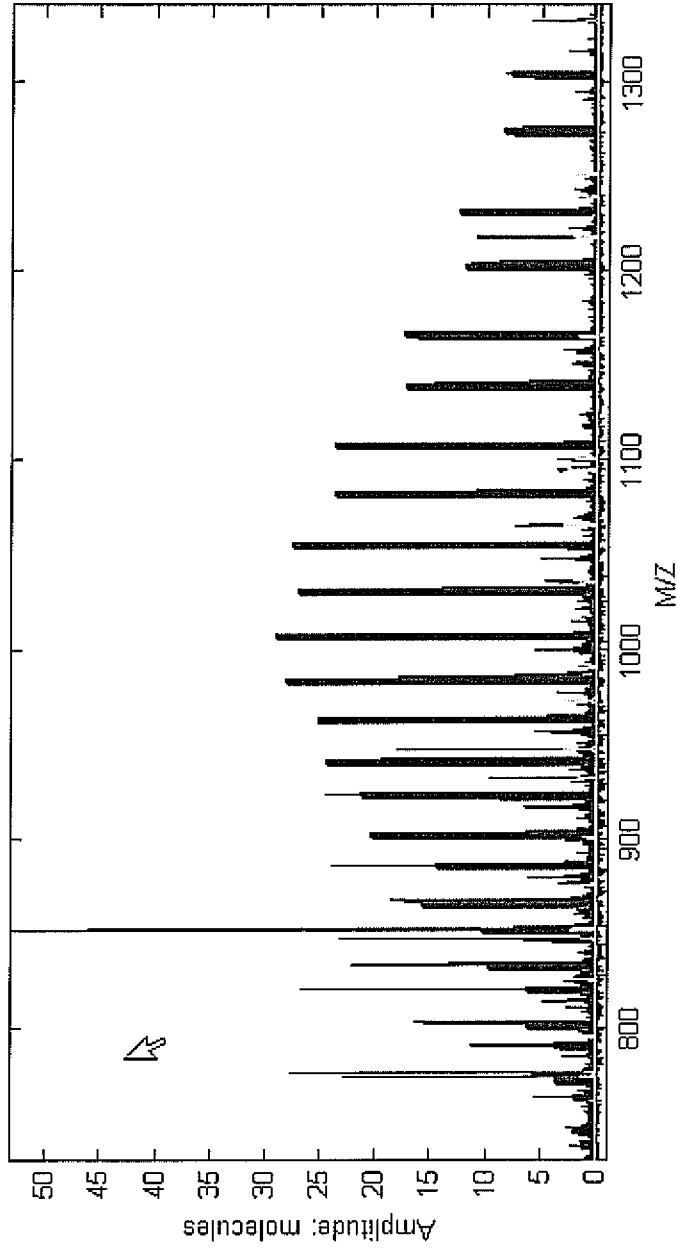


Fig. 6

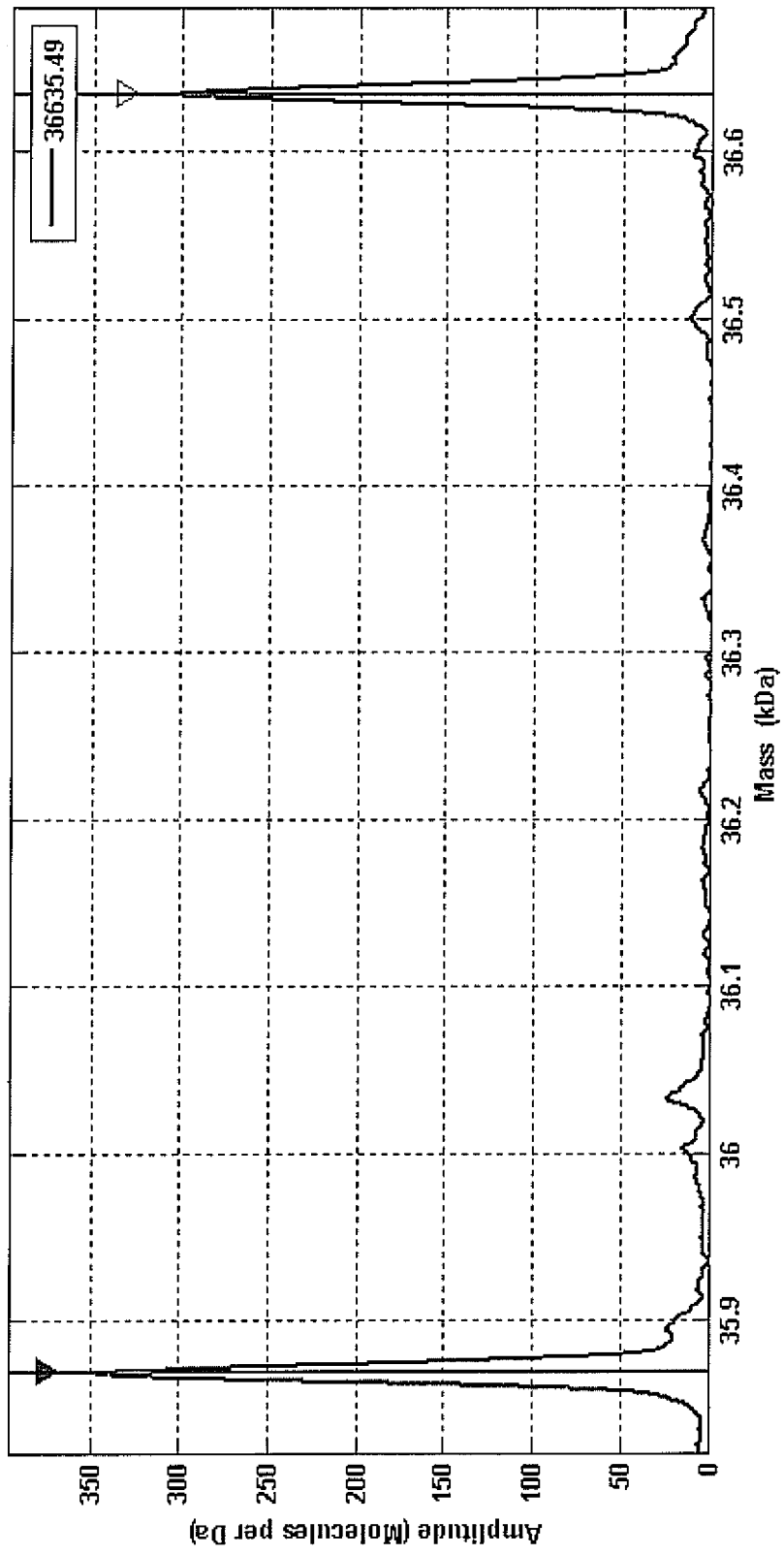


Fig. 7

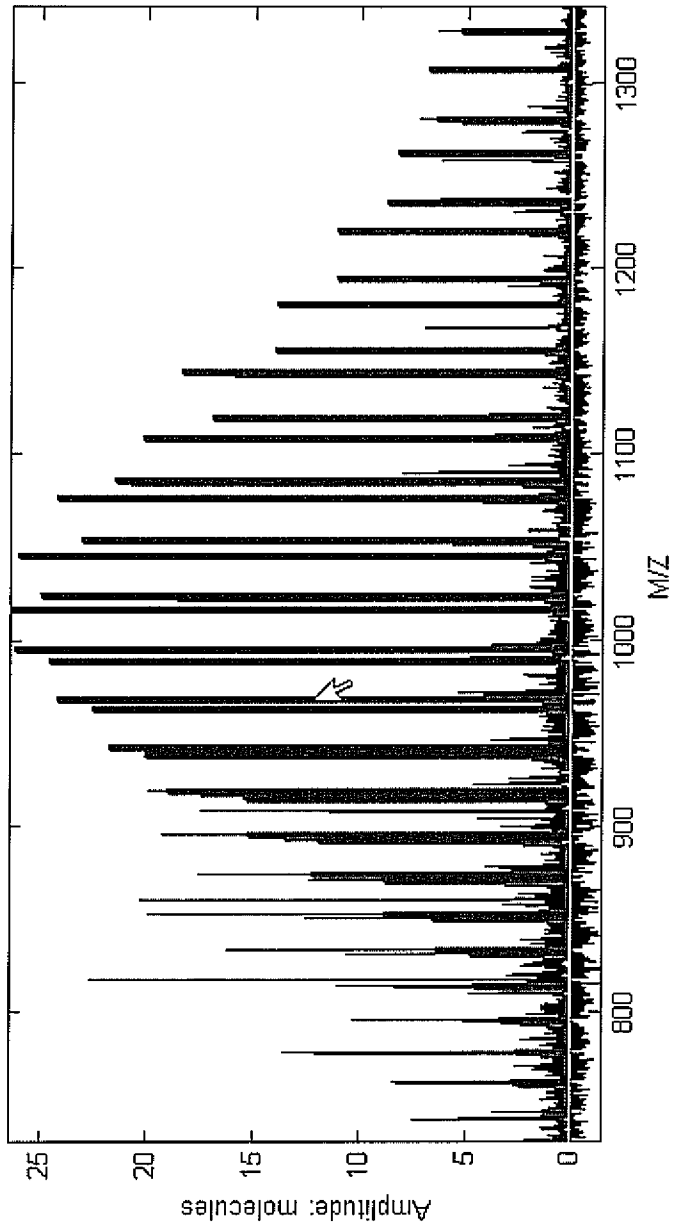


Fig. 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/42630

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12P 19/34, C12Q 1/68 (2010.01)

USPC - 536/24.33, 435/91.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8): C12P 19/34, C12Q 1/68 (2010.01)

USPC: 536/24.33, 435/91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC: 435/6

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubWest (USPT,PGPB,EPAB,JPAB), Google Scholar: 5-propynyluracil, 5-propynylcytosine, mass, tag\$2, thymidine residue\$2, isolat\$4, amplification, calicivirus\$2, nucleobase\$2, hybridiz\$5, primer\$2, forward, reverse, virus\$2, thymidine, 5', isolat\$4, length, amplification, hybridiz\$4, nucleic acid\$2, oligonuci\$6, oligonucleated, mass spec\$7, base\$2, mat

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2009/0047665 A1 (HALL et al.) 19 February 2009 (19.02.2009); abstract, para [0038], [0090], [0042], [0045], [0097], [0131], [0168], [0169]	1-44
Y	US 2003/0180323 A1 (SMITH) 25 September 2003 (25.09.2003); para [0006], [0008], [0009], [0011], [0013], [0037], [0047], [0101], [0118], [0151], [0154], [0179]	1-44
Y	US 2008/0045473 A1 (UHLMANN et al.) 21 February 2008 (21.02.2008); para [0113]	10, 23
Y	CLARKE et al., Organization and Expression of Calicivirus Genes, The Journal of Infectious Diseases, May 2000, Vol. 181, Suppl 2, Pages S309-S316; abstract	4, 17
Y	US 2007/0114376 A1 (HAGER) 24 May 2007 (24.05.2007); para [0005]	42

 Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 September 2010 (30.09.2010)

Date of mailing of the international search report

29 OCT 2010

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774