

## METHODS FOR IDENTIFICATION OF CORONAVIRUSES

### CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority to: 1) U.S. application Serial No. 60/466,009 filed April 26, 2003; 2) U.S. application Serial No. 60/467,768 filed May 2, 2003; 3) U.S. application Serial No. 60/468,743 filed May 7, 2003 and 4) U.S. application Serial No. 60/542,510 filed February 6, 2004, each of which is incorporated herein by reference in its entirety.

10

### FIELD OF THE INVENTION

The present invention relates generally to the field of genetic identification and quantitation of coronaviruses and provides methods, compositions and kits useful for this purpose when combined with molecular mass analysis.

15

### BACKGROUND OF THE INVENTION

Coronaviruses, a genus in the family Coronaviridae, are large, enveloped RNA viruses that cause highly prevalent diseases in humans and domestic animals. Coronavirus particles are irregularly-shaped, 60-220 nm in diameter, with an outer envelope bearing distinctive, "club-shaped" peplomers. This "crown-like" appearance gives the family its name. Coronaviruses have the largest genomes of all RNA viruses and replicate by a unique mechanism which results in a high frequency of recombination. Virions mature by budding at intracellular membranes and infection with some coronaviruses induces cell fusion.

Most human coronaviruses (HCoV) do not grow in cultured cells, therefore relatively little is known about them, but two strains (229E and OC43) grow in some cell lines and have been used as a model. Replication is slow compared to other enveloped viruses. Viral entry

occurs via endocytosis and membrane fusion (probably mediated by E2) and replication occurs in the cytoplasm.

Initially, the 5' 20kb of the (+)sense genome is translated to produce a viral polymerase, which is believed to produce a full-length (-)sense strand which, in turn, is used as a template to produce mRNA as a "nested set" of transcripts, all with an identical 5' non-translated leader sequence of 72 nucleotides and coincident 3' polyadenylated ends. Each mRNA is monocistronic, the genes at the 5' end being translated from the longest mRNA. These unusual cytoplasmic structures are produced not by splicing (post-transcriptional modification) but by the polymerase during transcription.

10 Coronaviruses infect a variety of mammals and birds. The exact number of human isolates is not known as many cannot be grown in culture. In humans, they cause: respiratory infections (common), including Severe Acute Respiratory Syndrome (SARS), and enteric infections.

Coronaviruses are transmitted by aerosols of respiratory secretions, by the fecal-oral route, and by mechanical transmission. Most virus growth occurs in epithelial cells. Occasionally the liver, kidneys, heart or eyes may be infected, as well as other cell types such as macrophages. In cold-type respiratory infections, growth appears to be localized to the epithelium of the upper respiratory tract, but there is currently no adequate animal model for the human respiratory coronaviruses. Clinically, most infections cause a mild, self-limited disease (classical "cold" or upset stomach), but there may be rare neurological complications. Coronavirus infection is very common and occurs worldwide. The incidence of infection is strongly seasonal, with the greatest incidence in children in winter. Adult infections are less common. The number of coronavirus serotypes and the extent of antigenic variation are unknown. Re-infections appear to occur throughout life, implying multiple serotypes (at least four are known) and/or antigenic variation, hence the prospects for immunization appear bleak.

25 SARS (Severe Acute Respiratory Syndrome) is a newly-recognized type of viral pneumonia, with symptoms including fever, a dry cough, dyspnea (shortness of breath), headache, and hypoxemia (low blood oxygen concentration). Typical laboratory findings include lymphopenia (reduced lymphocyte numbers) and mildly elevated aminotransferase levels (indicating liver damage). Death may result from progressive respiratory failure due to alveolar damage.

The outbreak is believed to have originated in February 2003 in the Guangdong

province of China. After initial reports that a paramyxovirus was responsible, researchers now believe SARS to causally-linked with a type of novel coronavirus with some unusual properties. For example, the SARS virus can be grown in Vero cells (a primate fibroblast cell line) - a novel property for HCoV, most of which cannot be cultivated. In these cells, virus  
5 infection results in a cytopathic effect, and budding of coronavirus-like particles from the endoplasmic reticulum within infected cells.

Amplification of short regions of the polymerase gene, (the most strongly conserved part of the coronavirus genome) by reverse transcriptase polymerase chain reaction (RT-PCR) and nucleotide sequencing revealed that the currently evaluated examples of the SARS virus are  
10 of a novel coronavirus which has not previously been present in human populations.

Different isolates of coronaviruses that have been causally linked to SARS have been independently sequenced by BCCA Genome Sciences Center, Vancouver, Canada; the Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences / Beijing Genomics Institute, Chinese Academy of Sciences, Beijing, China; the Centers for Disease Control and  
15 Prevention (CDC), Atlanta; the Chinese University of Hong Kong; and the University of Hong Kong. As new SARS-linked coronavirus samples are obtained and sequenced, and as the initial SARS coronaviruses mutate, other coronavirus sequences causally-linked to SARS will emerge.

While the SARS epidemic is still at the early stages, Ruan et al have identified a number of variations in existing SARS CoV isolates that suggest the emergence of new  
20 genotypes (Y. Ruan et al., *Lancet*, May 9, (2003)). This phenomenon is likely to continue if SARS CoV passes through the human population and will have a detrimental impact on detection and treatment. Additional primers that flank regions of high variability could be valuable in epidemiological tracking of strain variants. Moreover, as loci important to virulence become identified, primers that flank these locations could provide valuable information.

25 Diagnostic tests are now available, but all have limitations as tools for bringing this outbreak quickly under control. An ELISA test detects antibodies reliably but only from about day 20 after the onset of clinical symptoms. It therefore cannot be used to detect cases at an early stage prior to spread of the infection to others. The second test, an immunofluorescence assay (IFA), detects antibodies reliably as of day 10 of infection. It shares the defect of the ELISA test  
30 in that test subjects have become infective prior to IFA-based diagnosis. Moreover, the IFA test is a demanding and comparatively slow test that requires the growth of virus in cell culture. The third test is a polymerase chain reaction (PCR) molecular test for detection of SARS virus

genetic material is useful in the early stages of infection but undesirably produces false-negatives. Thus the PCR test may fail to detect persons who actually carry the virus, even in conjunction with clinical diagnostic evaluation, creating a dangerous sense of false security in the face of a potential epidemic of a virus that is known to spread easily in close person-to-  
5 person contact (WHO. Severe acute respiratory syndrome (SARS). *Wkly Epidemiol. Rec.* 2003, 78, 121-122).

Nucleic acid tests for infectious diseases are largely based upon amplifications using primers and probes designed to detect specific bioagents. Because prior knowledge of nucleic acid sequence information is required to develop these tests they are not able to identify  
10 unanticipated, newly emergent, or previously unknown infectious bioagents. Thus, the initial discovery of infectious bioagents still relies largely on culture and microscopy, which were as important in the recent identification of the SARS coronavirus as they were in the discovery of the human immunodeficiency virus two decades ago.

An alternative to single-agent tests is to do broad-range consensus priming of a gene  
15 target conserved across groups of bioagents. Broad-range priming has the potential to generate amplification products across entire genera, families, or, as with bacteria, an entire domain of life. This strategy has been successfully employed using consensus 16S ribosomal RNA primers for determining bacterial diversity, both in environmental samples (T. M. Schmidt, T. M., DeLong, E. F., Pace, N. R. *J. Bact.* 173, 4371-4378 (1991)) and in natural human flora (Kroes,  
20 I., Lepp, P. W., Relman, D. A. *Proc Nat Acad Sci (USA)* 96, 14547-14552 (1999)). The drawback of this approach for unknown bioagent detection and epidemiology is that analysis of the PCR products requires the cloning and sequencing of hundreds to thousands of colonies per sample, which is impractical to perform rapidly or on a large number of samples.

Consensus priming has also been described for detection of several viral families,  
25 including coronaviruses (Stephensen, C. B., Casebolt, D. B. Gangopadhyay, N. N. *Vir. Res.* 60, 181-189 (1999)), enteroviruses (M. S. Oberste, K. Maher, M. A. Pallansch, *J. Virol.* 76, 1244-51 (2002); M. S. Oberste, W. A. Nix, K. Maher, M. A. Pallansch, *J. Clin. Virol.* 26, 375-7 (2003); M. S. Oberste, W. A. Nix, D. R. Kilpatrick, M. R. Flemister, M. A. Pallansch, *Virus Res.* 91, 241-8(2003)), retnoid viruses(D. H. Mack, J. J. Sninsky, *Proc. Natl. Acad. Sci. U. S. A.* 85, 6977-  
30 81 (1988); W. Seifarth et al., *AIDS Res. Hum. Retroviruses* 16, 721-729 (2000); L. A. Donehower, R. C. Bohannon, R. J. Ford, R. A. Gibbs, *J. Vir. Methods* 28, 33-46 (1990)), and adenoviruses (M. Echavarria, M. Forman, J. Ticehurst, S. Dumler, P. Charache, *J. Clin. Micro.*

36, 3323-3326 (1998)). However, as with bacteria, there is no adequate analytical method other than sequencing to identify the viral bioagent present. Methods of identification of bioagents are described in U.S. Patent application Serial Nos: 09/798,007, filed March 3, 2001; 10/405,756, filed March 31, 2003; 10/660,122, filed September 11, 2003; and 10/728,486, filed 5 December 5, 2003, all of which are commonly owned and incorporated herein by reference in entirety as essential material.

Mass spectrometry provides detailed information about the molecules being analyzed, including high mass accuracy. It is also a process that can be easily automated. However, high-resolution MS alone fails to perform against unknown or bioengineered agents, or in 10 environments where there is a high background level of bioagents ("cluttered" background). Low-resolution MS can fail to detect some known agents, if their spectral lines are sufficiently weak or sufficiently close to those from other living organisms in the sample. DNA chips with specific probes can only determine the presence or absence of specifically anticipated organisms. Because there are hundreds of thousands of species of benign bacteria, some very similar in 15 sequence to threat organisms, even arrays with 10,000 probes lack the breadth needed to detect a particular organism.

Antibodies face more severe diversity limitations than arrays. If antibodies are designed against highly conserved targets to increase diversity, the false alarm problem will dominate, again because threat organisms are very similar to benign ones. Antibodies are only capable of 20 detecting known agents in relatively uncluttered environments.

Several groups have described detection of PCR products using high resolution electrospray ionization -Fourier transform- ion cyclotron resonance mass spectrometry (ESI-FT-ICR MS). Accurate measurement of exact mass combined with knowledge of the number of at least one nucleotide allowed calculation of the total base composition for PCR duplex products 25 of approximately 100 base pairs. (Aaserud et al., *J. Am. Soc. Mass Spec.* 7:1266-1269, 1996; Muddiman et al., *Anal. Chem.* 69:1543-1549, 1997; Wunschel et al., *Anal. Chem.* 70:1203-1207, 1998; Muddiman et al., *Rev. Anal. Chem.* 17:1-68, 1998). Electrospray ionization-Fourier transform-ion cyclotron resistance (ESI-FT-ICR) MS may be used to determine the mass of double-stranded, 500 base-pair PCR products via the average molecular mass (Hurst et al., *Rapid* 30 *Commun. Mass Spec.* 10:377-382, 1996). The use of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for characterization of PCR products has been described. (Muddiman et al., *Rapid Commun. Mass Spec.* 13:1201-1204, 1999). However, the

degradation of DNAs over about 75 nucleotides observed with MALDI limited the utility of this method.

U.S. Patent No. 5,849,492 describes a method for retrieval of phylogenetically informative DNA sequences which comprise searching for a highly divergent segment of  
5 genomic DNA surrounded by two highly conserved segments, designing the universal primers for PCR amplification of the highly divergent region, amplifying the genomic DNA by PCR technique using universal primers, and then sequencing the gene to determine the identity of the organism.

U.S. Patent No. 5,965,363 discloses methods for screening nucleic acids for  
10 polymorphisms by analyzing amplified target nucleic acids using mass spectrometric techniques and to procedures for improving mass resolution and mass accuracy of these methods.

WO 99/14375 describes methods, PCR primers and kits for use in analyzing preselected DNA tandem nucleotide repeat alleles by mass spectrometry.

WO 98/12355 discloses methods of determining the mass of a target nucleic acid by  
15 mass spectrometric analysis, by cleaving the target nucleic acid to reduce its length, making the target single-stranded and using MS to determine the mass of the single-stranded shortened target. Also disclosed are methods of preparing a double-stranded target nucleic acid for MS analysis comprising amplification of the target nucleic acid, binding one of the strands to a solid support, releasing the second strand and then releasing the first strand which is then analyzed by  
20 MS. Kits for target nucleic acid preparation are also provided.

PCT WO97/33000 discloses methods for detecting mutations in a target nucleic acid by nonrandomly fragmenting the target into a set of single-stranded nonrandom length fragments and determining their masses by MS.

U.S. Patent No. 5,605,798 describes a fast and highly accurate mass spectrometer-based  
25 process for detecting the presence of a particular nucleic acid in a biological sample for diagnostic purposes.

WO 98/21066 describes processes for determining the sequence of a particular target nucleic acid by mass spectrometry. Processes for detecting a target nucleic acid present in a biological sample by PCR amplification and mass spectrometry detection are disclosed, as are  
30 methods for detecting a target nucleic acid in a sample by amplifying the target with primers that contain restriction sites and tags, extending and cleaving the amplified nucleic acid, and detecting the presence of extended product, wherein the presence of a DNA fragment of a mass

different from wild-type is indicative of a mutation. Methods of sequencing a nucleic acid via mass spectrometry methods are also described.

WO 97/37041, WO 99/31278 and US Patent No. 5,547,835 describe methods of sequencing nucleic acids using mass spectrometry. US Patent Nos. 5,622,824, 5,872,003 and 5 5,691,141 describe methods, systems and kits for exonuclease-mediated mass spectrometric sequencing.

The present invention provides a novel approach for rapid, sensitive, and high-throughput identification of coronaviruses and includes the capability of identification of coronaviruses not yet observed and characterized. The methods described can be applied to 10 additional viral families to cover a broad range of potential newly emerging viruses, or to bacterial, protozoal or fungal pathogens for epidemic disease surveillance in the future.

## SUMMARY OF THE INVENTION

The present invention is directed to, *inter alia*, methods of identification of one or more 15 unknown coronaviruses in a sample by obtaining coronavirus RNA from the sample, obtaining corresponding DNA from the RNA, amplifying the DNA with one or more pairs of oligonucleotide primers that bind to conserved regions of a coronavirus genome which are flanked a variable region of the coronavirus genome, determining the molecular masses or base compositions of the one or more amplification products and comparing the molecular masses or 20 base compositions with calculated or experimentally determined molecular masses or base compositions, wherein one or more matches identifies the unknown coronavirus.

The present invention is also directed to methods of tracking the spread of a specific coronavirus comprising: obtaining a plurality of samples containing a specific coronavirus from a plurality of different locations, identifying the specific coronavirus in a subset of the plurality 25 of samples using the method described in the paragraph above, wherein the corresponding locations of the members of the subset indicate the spread of the specific coronavirus to the corresponding locations.

The present invention is also directed to pairs of primers wherein each member of each pair has at least 70% sequence identity with the sequence of the corresponding member of any 30 one of the following intelligent primer pair sequences: SEQ ID NOs: 5:6, 7:8, 9:8, 9:10, 11:8, 11:10 or 9:10. The present invention is also directed to individual primers within each of the primer pairs described herein.

The present invention is also directed to bioagent identifying amplicons for identification of a coronavirus comprising an isolated polynucleotide of about 45 to about 150 nucleobases in length produced by the process of amplification of nucleic acid from a coronavirus genome with a primer pair wherein each primer of the pair is of a length of about 12 to about 35 nucleobases and the bioagent identifying amplicon provides identifying information about the coronavirus.

The present invention is also directed to methods for simultaneous determination of the identity and quantity of an unknown coronavirus in a sample comprising: contacting a sample with a pair of primers and a known quantity of a calibration polynucleotide comprising a calibration sequence, simultaneously amplifying nucleic acid from the unknown coronavirus with the pair of primers and amplifying nucleic acid from the calibration polynucleotide in the sample with the pair of primers to obtain a first amplification product comprising a bioagent identifying amplicon and a second amplification product comprising a calibration amplicon, subjecting the sample to molecular mass analysis wherein the result of the mass analysis comprises molecular mass and abundance data for the bioagent identifying amplicon and the calibration amplicon, and distinguishing the bioagent identifying amplicon from the calibration amplicon based on molecular mass wherein the molecular mass of the bioagent identifying amplicon identifies the coronavirus and comparison of bioagent identifying amplicon abundance data and calibration amplicon abundance data indicates the quantity of coronavirus in the sample.

The present invention is also directed to isolated polynucleotides for determining the quantity of a bioagent in a sample comprising SEQ ID NOs: 102, and 103 as well as vectors comprising of SEQ ID NOs: 102, 103 and 104.

The present invention is also directed to kits comprising one or more pairs of primers, or individual primers, wherein each member of each pair has at least 70% sequence identity with the sequence of the corresponding member of any one of the following intelligent primer pair sequences: SEQ ID NOs: 5:6, 7:8, 9:8, 9:10, 11:8, 11:10 or 9:10.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1H and Figure 2 are consensus diagrams that show examples of conserved regions of 16S rRNA (Fig. 1A, 1A-2, 1A-3, 1A-4 and 1A-5), 23S rRNA (3'-half, Fig. 1B-1, 1B-2, 1C-1, 1C-2, and 1D; 5'-half, Fig. 1E-F), 23S rRNA Domain I (Fig. 1G), 23S rRNA Domain IV (Fig. 1H) and 16S rRNA Domain III (Fig. 2). DNA segments encoding these regions are

suitable for use as templates for generation of bioagent identifying amplicons. Lines with arrows are examples of regions (in corresponding DNA) to which intelligent primer pairs for PCR are designed. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. Bases in capital letters are greater than 95% conserved; bases in lower case letters are 90-95% conserved, filled circles are 80-90% conserved; and open circles are less than 80% conserved. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. The nucleotide sequence of the 16S rRNA consensus sequence is SEQ ID NO:3 and the nucleotide sequence of the 23S rRNA consensus sequence is SEQ ID NO:4.

10        Figure 2 shows a typical primer amplified region from the 16S rRNA Domain III shown in Figure 1A-1.

Figure 3 is a schematic diagram showing conserved regions in RNase P. Bases in capital letters are greater than 90% conserved; bases in lower case letters are 80-90% conserved; filled circles designate bases which are 70-80% conserved; and open circles designate bases that are  
15 less than 70% conserved.

Figure 4 is a schematic diagram of base composition signature determination using nucleotide analog "tags" to determine base composition signatures.

Figure 5 shows the deconvoluted mass spectra of a *Bacillus anthracis* region with and without the mass tag phosphorothioate A (A\*). The two spectra differ in that the measured  
20 molecular weight of the mass tag-containing sequence is greater than the unmodified sequence.

Figure 6 is a process diagram illustrating the primer selection process. For each group of organisms, candidate target sequences are identified (200) from which nucleotide alignments are created (210) and analyzed (220). Primers are then designed by selecting appropriate priming regions (230) which then makes possible the selection of candidate primer pairs (240). The  
25 primer pairs are then subjected to *in silico* analysis by electronic PCR (ePCR) (300) wherein bioagent identifying amplicons are obtained from sequence databases such as GenBank or other sequence collections (310) and checked for specificity *in silico* (320). Bioagent identifying amplicons obtained from GenBank sequences (310) can also be analyzed by a probability model which predicts the capability of a given amplicon to identify unknown bioagents such that the  
30 base compositions of amplicons with favorable probability scores are then stored in a base composition database (325). Alternatively, base compositions of the bioagent identifying amplicons obtained from the primers and GenBank sequences can be directly entered into the

base composition database (330). Candidate primer pairs (240) are validated by *in vitro* amplification by a method such as PCR analysis (400) of nucleic acid from a collection of organisms (410). Amplification products thus obtained are analyzed to confirm the sensitivity, specificity and reproducibility of the primers used to obtain the amplification products (420).

5 Figure 7 indicates two coronavirus genomic regions used to generate bioagent identifying amplicons for coronaviruses. The two primer pair regions are mapped onto the SARS CoV (TOR2, NC\_004718.3: SEQ ID NO: 85) genome coordinates. The RdRp amplicon corresponds to positions 15,132-15,218 and the nsp11 amplicon corresponds to positions 19,098-19,234 of the same sequence. Multiple sequence alignments of additional available coronavirus  
10 sequences corresponding to these two amplified regions are shown. Coronavirus isolates that were sequenced as part of this work are shown with an asterisk. All other sequences were obtained from GenBank. nsp11 sequences for PHEV and TCoV are not available and are shown with a dashed line. Forward and reverse primer regions are highlighted. 5' ends of all primers were designed with a thymidine (T) nucleotide which acts to minimize the addition of non-  
15 templated A residues during PCR. For each specific coronavirus, the positions that are mismatched compared to the primer sequence are shown. The complement of the actual primer is shown for the reverse primers. The region between the forward and reverse primers for each virus varies among different coronaviruses (not shown) and provides the signature for resolving them by molecular mass or base composition. Primer positions chemically modified with  
20 propyne groups are highlighted. The structures of 5-propynyldeoxycytidine and 5-propynyldeoxyuridine nucleotides used in primers are shown.

Figure 8 is an ESI-FTICR mass spectrum measurement of the PCR amplicon from the SARS coronavirus obtained using the propynylated RdRp primer pairs. The electrospray ionization conditions separate the sense and antisense strands of the PCR products (500).  
25 Multiple charge states are observed across the  $m/z$  range shown (510) from which is obtained an expanded view of the isotope envelope of the  $(M-27H+)^{27-}$  species (520). The derived molecular masses (530) for the sense amplicon strand is 27298.518 from which is calculated (540) the unambiguous base composition of  $A_{27}G_{19}C_{14}T_{28}$  for the sense strand (seen also in Table 4). This base composition corresponds to the base composition calculated based on the  
30 bioagent identifying amplicon of the genomic sequence of the SARS coronavirus.

Figure 9 shows a series of mass spectra which was used to identify three human coronaviruses in a mixture. The deconvoluted (neutral mass) mass spectra obtained for the RdRp

primer for the three human coronaviruses, HCoV-229E, HCoV-OC43 and SARS CoV, that were tested individually and in a mixture are shown. Forward and reverse amplicons are shown with the measured monoisotopic masses for each strand.

Figure 10 indicates three dimensional plots of experimentally determined base compositions (solid cones) and calculated base compositions (spheres) for bioagent identifying amplicons of coronaviruses obtained with a RdRp primer set and a nsp11 primer set. The position of the base composition indicates the intersection of A, C and T content in each base composition, with G content illustrated in a “pseudo-fourth” dimension indicated by the angle of rotation of the cone. For the RdRp primer set, the experimentally-determined base compositions were in agreement with the calculated base compositions except for the canine coronavirus isolate (CcoV) which was found to have a composition that differed from that calculated based on available sequence information by a single T to C substitution (indicated by a dashed line).

Figure 11 shows a mass spectrum containing mass peaks corresponding to the PCR internal standard calibrant amplicon at a dilution corresponding to a nucleic acid copy number of  $3 \times 10^4$  and a SARS coronavirus bioagent identifying amplicon. The comparison of peak abundances of the two peaks enabled the determination that 1 PFU of SARS coronavirus = 300 copies of nucleic acid.

## DESCRIPTION OF EMBODIMENTS

### A. Introduction

The present invention provides methods for detection and identification of bioagents in an unbiased manner using “bioagent identifying amplicons.” “Intelligent primers” are selected to hybridize to conserved sequence regions of nucleic acids derived from a bioagent and which bracket variable sequence regions to yield a bioagent identifying amplicon which can be amplified and which is amenable to molecular mass determination. The molecular mass then provides a means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass or corresponding “base composition signature” (BCS) of the amplification product is then matched against a database of molecular masses or base composition signatures. Furthermore, the method can be applied to rapid parallel “multiplex” analyses, the results of which can be employed in a triangulation identification strategy. The present method provides rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent detection and identification.

## B. Bioagents

In the context of this invention, a “bioagent” is any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus. Examples of bioagents include, but are not limited, to cells, including but not limited to human clinical samples, 5 bacterial cells and other pathogens), viruses, fungi, protists, parasites, and pathogenicity markers (including but not limited to: pathogenicity islands, antibiotic resistance genes, virulence factors, toxin genes and other bioregulating compounds). Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores) and may be encapsulated or bioengineered. In the context of this invention, a “pathogen” is a bioagent which causes a disease 10 or disorder.

Despite enormous biological diversity, all forms of life on earth share sets of essential, common features in their genomes. Bacteria, for example have highly conserved sequences in a variety of locations on their genomes. Most notable is the universally conserved region of the ribosome. There are also conserved elements in other non-coding RNAs, including RNase P 15 (Figure 3) and the signal recognition particle (SRP) among others. Bacteria have a common set of absolutely required genes. About 250 genes are present in all bacterial species (*Proc. Natl. Acad. Sci. U.S.A.*, **1996**, 93, 10268; *Science*, **1995**, 270, 397), including tiny genomes like *Mycoplasma*, *Ureaplasma* and *Rickettsia*. These genes encode proteins involved in translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid 20 metabolism, lipid metabolism, energy generation, uptake, secretion and the like. Examples of these proteins are DNA polymerase III beta, elongation factor TU, heat shock protein groEL, RNA polymerase beta, phosphoglycerate kinase, NADH dehydrogenase, DNA ligase, DNA topoisomerase and elongation factor G. Operons can also be targeted using the present method. One example of an operon is the bfp operon from enteropathogenic *E. coli*. Multiple core 25 chromosomal genes can be used to classify bacteria at a genus or genus species level to determine if an organism has threat potential. The methods can also be used to detect pathogenicity markers (plasmid or chromosomal) and antibiotic resistance genes to confirm the threat potential of an organism and to direct countermeasures.

## 30 C. Selection of “Bioagent Identifying Amplicons”

Since genetic data provide the underlying basis for identification of bioagents by the methods of the present invention, it is necessary to select segments of nucleic acids which ideally

provide enough variability to distinguish each individual bioagent and whose molecular mass is amenable to molecular mass determination. In one embodiment of the present invention, at least one polynucleotide segment is amplified to facilitate detection and analysis in the process of identifying the bioagent. Thus, the nucleic acid segments which provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as "bioagent identifying amplicons." The term "amplicon" as used herein, refers to a segment of a polynucleotide which is amplified in an amplification reaction.

In one embodiment, bioagent identifying amplicons are from about 45 nucleobases to about 150 nucleobases in length.

Pre-bioagent identifying amplicons are amplicons which may greatly exceed about 45 to about 150 nucleobases in length and which contain sites for cleavage (by restriction endonucleases, for example) to yield bioagent identifying amplicons which are fragments of a given pre-bioagent identifying amplicon and which are amenable to molecular mass analysis.

As used herein, "intelligent primers" are primers that are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. By the term "highly conserved," it is meant that the sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus design of intelligent primers requires selection of a variable region with appropriate variability to resolve the identity of a given bioagent. Bioagent identifying amplicons are ideally specific to the identity of the bioagent. A plurality of bioagent identifying amplicons selected in parallel for distinct bioagents which contain the same conserved sequences for hybridization of the same pair of intelligent primers are herein defined as "correlative bioagent identifying amplicons."

In one embodiment, the bioagent identifying amplicon is a portion of a ribosomal RNA (rRNA) gene sequence. With the complete sequences of many of the smallest microbial genomes now available, it is possible to identify a set of genes that defines "minimal life" and identify composition signatures that uniquely identify each gene and organism. Genes that encode core life functions such as DNA replication, transcription, ribosome structure, translation, and

transport are distributed broadly in the bacterial genome and are suitable regions for selection of bioagent identifying amplicons. Ribosomal RNA (rRNA) genes comprise regions that provide useful base composition signatures. Like many genes involved in core life functions, rRNA genes contain sequences that are extraordinarily conserved across bacterial domains interspersed with regions of high variability that are more specific to each species. The variable regions can be utilized to build a database of base composition signatures. The strategy involves creating a structure-based alignment of sequences of the small (16S) and the large (23S) subunits of the rRNA genes. For example, there are currently over 13,000 sequences in the ribosomal RNA database that has been created and maintained by Robin Gutell, University of Texas at Austin, and is publicly available on the Institute for Cellular and Molecular Biology web page on the world wide web of the Internet at, for example, "rna.icmb.utexas.edu/." There is also a publicly available rRNA database created and maintained by the University of Antwerp, Belgium on the world wide web of the Internet at, for example, "rna.uia.ac.be."

These databases have been analyzed to determine regions that are useful as bioagent identifying amplicons. The characteristics of such regions include: a) between about 80 and 100%, or greater than about 95% identity among species of the particular bioagent of interest, of upstream and downstream nucleotide sequences which serve as sequence amplification primer sites; b) an intervening variable region which exhibits no greater than about 5% identity among species; and c) a separation of between about 30 and 1000 nucleotides, or no more than about 50-250 nucleotides, or no more than about 60-100 nucleotides, between the conserved regions.

As a non-limiting example, for identification of *Bacillus* species, the conserved sequence regions of the chosen bioagent identifying amplicon must be highly conserved among all *Bacillus* species while the variable region of the bioagent identifying amplicon is sufficiently variable such that the molecular masses of the amplification products of all species of *Bacillus* are distinguishable.

Bioagent identifying amplicons amenable to molecular mass determination are either of a length, size or mass compatible with the particular mode of molecular mass determination or compatible with a means of providing a predictable fragmentation pattern in order to obtain predictable fragments of a length compatible with the particular mode of molecular mass determination. Such means of providing a predictable fragmentation pattern of an amplification product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, for example.

Identification of bioagents can be accomplished at different levels using intelligent primers suited to resolution of each individual level of identification. "Broad range survey" intelligent primers are designed with the objective of identifying a bioagent as a member of a particular division of bioagents. A "bioagent division" is defined as group of bioagents above the  
5 species level and includes but is not limited to: orders, families, classes, clades, genera or other such groupings of bioagents above the species level. As a non-limiting example, members of the *Bacillus/Clostridia* group or gamma-proteobacteria group may be identified as such by employing broad range survey intelligent primers such as primers which target 16S or 23S ribosomal RNA.

10 In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species level. One main advantage of the detection methods of the present invention is that the broad range survey intelligent primers need not be specific for a particular bacterial species, or even genus, such as *Bacillus* or *Streptomyces*. Instead, the primers recognize highly conserved regions across hundreds of bacterial species including, but not  
15 limited to, the species described herein. Thus, the same broad range survey intelligent primer pair can be used to identify any desired bacterium because it will bind to the conserved regions that flank a variable region specific to a single species, or common to several bacterial species, allowing unbiased nucleic acid amplification of the intervening sequence and determination of its molecular weight and base composition. For example, the 16S\_971-1062, 16S\_1228-1310 and  
20 16S\_1100-1188 regions are 98-99% conserved in about 900 species of bacteria (16S=16S rRNA, numbers indicate nucleotide position). In one embodiment of the present invention, primers used in the present method bind to one or more of these regions or portions thereof.

Due to their overall conservation, the flanking rRNA primer sequences serve as good intelligent primer binding sites to amplify the nucleic acid region of interest for most, if not all,  
25 bacterial species. The intervening region between the sets of primers varies in length and/or composition, and thus provides a unique base composition signature. Examples of intelligent primers that amplify regions of the 16S and 23S rRNA are shown in Figures 1A-1H. A typical primer amplified region in 16S rRNA is shown in Figure 2. The arrows represent primers that bind to highly conserved regions of the DNA encoding these regions, which flank a variable  
30 region in 16S rRNA domain III. The amplified region corresponds to the stem-loop structure under "1100-1188." It is advantageous to design the broad range survey intelligent primers to minimize the number of primers required for the analysis, and to allow detection of multiple

members of a bioagent division using a single pair of primers. The advantage of using broad range survey intelligent primers is that once a bioagent is broadly identified, the process of further identification at species and sub-species levels is facilitated by directing the choice of additional intelligent primers.

5       “Division-wide” intelligent primers are designed with an objective of identifying a bioagent at the species level. As a non-limiting example, a *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis* can be distinguished from each other using division-wide intelligent primers. Division-wide intelligent primers are not always required for identification at the species level because broad range survey intelligent primers may provide sufficient identification  
10 resolution to accomplishing this identification objective.

      “Drill-down” intelligent primers are designed with an objective of identifying a sub-species characteristic of a bioagent. A “sub-species characteristic” is defined as a property imparted to a bioagent at the sub-species level of identification as a result of the presence or absence of a particular segment of nucleic acid. Such sub-species characteristics include, but are  
15 not limited to, strains, sub-types, pathogenicity markers such as antibiotic resistance genes, pathogenicity islands, toxin genes and virulence factors. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of pathogen infections.

#### **D. Selection and Optimization of Intelligent Primers**

20       A representative process flow diagram used for primer selection and validation process is outlined in Figure 6. Many of the important pathogens, including the organisms of greatest concern as biological weapons agents, have been completely sequenced. This effort has greatly facilitated the design of primers and probes for the detection of bacteria. Partial or full-length sequences from over 225 bacterial genomes have been obtained and sequence alignments have  
25 been generated for essential genes that are conserved either broadly across all organisms or within members of specific, related phylogenetic groups. In bacteria, for instance, alignments have been generated from over 170 housekeeping genes that are present in almost all major bacterial divisions. These genes have been used for identification of broad diagnostic primers. PCR primer selection and optimization has been largely automated. A number of genes, in  
30 addition to 16S rRNA, are targets of “broad-range” primers, thus increasing the redundancy of detection and classification, while minimizing potential missed identifications. Many of these genes are, expectedly, essential to information processing, with more than half being associated

with the translational machinery, such as elongation factors, ribosomal proteins and tRNA synthetases. Other classes of conserved protein-encoding genes include transcription-associated genes such as RNA polymerases and genes associated with DNA replication such as DNA gyrase and DNA polymerase. This combination of broad-range priming with clade- and species-specific priming has been used very successfully in several applications of the technology, including environmental surveillance for biowarfare threat agents and clinical sample analysis for medically important pathogens.

Ideally, intelligent primer hybridization sites are highly conserved in order to facilitate the hybridization of the primer. In cases where primer hybridization is less efficient due to lower levels of conservation of sequence, intelligent primers can be chemically modified to improve the efficiency of hybridization.

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

In another embodiment of the invention, to compensate for the somewhat weaker binding by the "wobble" base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, 5-propynyluracil which binds to adenine and 5-propynylcytosine and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is commonly owned and incorporated herein by reference in its entirety. Propynylated primers are claimed in U.S. Serial No. 10/294,203 which is also commonly owned and

incorporated herein by reference in entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096, each of which is incorporated herein by reference in its entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

5 In other embodiments, non-template primer tags are used to increase the melting temperature ( $T_m$ ) of a primer-template duplex in order to improve amplification efficiency. A non-template tag is designed to hybridize to at least three consecutive A or T nucleotide residues on a primer which are 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. The extra hydrogen bond in a G-C pair  
10 relative to a A-T pair confers increased stability of the primer-template duplex and improves amplification efficiency.

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.

15 In other embodiments, a primer contains a modified internucleoside linkage such as a phosphorothioate linkage, for example.

#### **E. Characterization of Bioagent Identifying Amplicons**

A theoretically ideal bioagent detector would identify, quantify, and report the complete  
20 nucleic acid sequence of every bioagent that reached the sensor. The complete sequence of the nucleic acid component of a pathogen would provide all relevant information about the threat, including its identity and the presence of drug-resistance or pathogenicity markers. This ideal has not yet been achieved. However, the present invention provides a straightforward strategy for obtaining information with the same practical value based on analysis of bioagent identifying  
25 amplicons by molecular mass determination.

In some cases, a molecular mass of a given bioagent identifying amplicon alone does not provide enough resolution to unambiguously identify a given bioagent. For example, the molecular mass of the bioagent identifying amplicon obtained using the intelligent primer pair “16S\_971” would be 55622 Da for both *E. coli* and *Salmonella typhimurium*. However, if  
30 additional intelligent primers are employed to analyze additional bioagent identifying amplicons, a “triangulation identification” process is enabled. For example, the “16S\_1100” intelligent primer pair yields molecular masses of 55009 and 55005 Da for *E. coli* and *Salmonella*

*typhimurium*, respectively. Furthermore, the “23S\_855” intelligent primer pair yields molecular masses of 42656 and 42698 Da for *E. coli* and *Salmonella typhimurium*, respectively. In this basic example, the second and third intelligent primer pairs provided the additional “fingerprinting” capability or resolution to distinguish between the two bioagents.

5 In another embodiment, the triangulation identification process is pursued by measuring signals from a plurality of bioagent identifying amplicons selected within multiple core genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. In this process, after identification of multiple core genes, alignments are created from nucleic acid sequence databases. The  
10 alignments are then analyzed for regions of conservation and variation, and bioagent identifying amplicons are selected to distinguish bioagents based on specific genomic differences. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen *et al.*, *J. Appl. Microbiol.*, 1999, 87, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

15 The triangulation identification process can be pursued by characterization of bioagent identifying amplicons in a massively parallel fashion using the polymerase chain reaction (PCR), such as multiplex PCR, and mass spectrometric (MS) methods. Sufficient quantities of nucleic acids should be present for detection of bioagents by MS. A wide variety of techniques for preparing large amounts of purified nucleic acids or fragments thereof are well known to those of  
20 ordinary skill in the art. PCR requires one or more pairs of oligonucleotide primers that bind to regions which flank the target sequence(s) to be amplified. These primers prime synthesis of a different strand of DNA, with synthesis occurring in the direction of one primer towards the other primer. The primers, DNA to be amplified, a thermostable DNA polymerase (e.g. *Taq* polymerase), the four deoxynucleotide triphosphates, and a buffer are combined to initiate DNA  
25 synthesis. The solution is denatured by heating, then cooled to allow annealing of newly added primer, followed by another round of DNA synthesis. This process is typically repeated for about 30 cycles, resulting in amplification of the target sequence.

Although the use of PCR is suitable, other nucleic acid amplification techniques may also be used, including ligase chain reaction (LCR) and strand displacement amplification  
30 (SDA). The high-resolution MS technique allows separation of bioagent spectral lines from background spectral lines in highly cluttered environments.

In another embodiment, the detection scheme for the PCR products generated from the

bioagent(s) incorporates at least three features. First, the technique simultaneously detects and differentiates multiple (generally about 6-10) PCR products. Second, the technique provides a molecular mass that uniquely identifies the bioagent from the possible primer sites. Finally, the detection technique is rapid, allowing multiple PCR reactions to be run in parallel.

5

#### **F. Mass Spectrometric Characterization of Bioagent Identifying Amplicons**

Mass spectrometry (MS)-based detection of PCR products provides a means for determination of BCS which has several advantages. MS is intrinsically a parallel detection scheme without the need for radioactive or fluorescent labels, since every amplification product is identified by its molecular mass. The current state of the art in mass spectrometry is such that less than femtomole quantities of material can be readily analyzed to afford information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons. Intact molecular ions can be generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but are not limited to, electrospray ionization (ES), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). For example, MALDI of nucleic acids, along with examples of matrices for use in MALDI of nucleic acids, are described in WO 98/54751.

In some embodiments, large DNAs and RNAs, or large amplification products therefrom, can be digested with restriction endonucleases prior to ionization. Thus, for example, an amplification product that was 10 kDa could be digested with a series of restriction endonucleases to produce a panel of, for example, 100 Da fragments. Restriction endonucleases and their sites of action are well known to the skilled artisan. In this manner, mass spectrometry can be performed for the purposes of restriction mapping.

Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a single mass spectrum affords an estimate of molecular mass of the bioagent. 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 in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), ion trap, quadrupole, magnetic sector, time of flight (TOF), Q-TOF, and triple quadrupole.

In general, the mass spectrometric techniques which can be used in the present invention include, but are not limited to, tandem mass spectrometry, infrared multiphoton dissociation and pyrolytic gas chromatography mass spectrometry (PGC-MS). In one embodiment of the invention, the bioagent detection system operates continually in bioagent detection mode using pyrolytic GC-MS without PCR for rapid detection of increases in biomass (for example, increases in fecal contamination of drinking water or of germ warfare agents). To achieve minimal latency, a continuous sample stream flows directly into the PGC-MS combustion chamber. When an increase in biomass is detected, a PCR process is automatically initiated. Bioagent presence produces elevated levels of large molecular fragments from, for example, about 100-7,000 Da which are observed in the PGC-MS spectrum. The observed mass spectrum is compared to a threshold level and when levels of biomass are determined to exceed a predetermined threshold, the bioagent classification process described hereinabove (combining PCR and MS, such as FT-ICR MS) is initiated. Optionally, alarms or other processes (halting ventilation flow, physical isolation) are also initiated by this detected biomass level.

The accurate measurement of molecular mass for large DNAs is limited by the adduction of cations from the PCR reaction to each strand, resolution of the isotopic peaks from natural abundance  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes, and assignment of the charge state for any ion. The cations are removed by in-line dialysis using a flow-through chip that brings the solution containing the PCR products into contact with a solution containing ammonium acetate in the presence of an electric field gradient orthogonal to the flow. The latter two problems are addressed by operating with a resolving power of  $>100,000$  and by incorporating isotopically depleted nucleotide triphosphates into the DNA. The resolving power of the instrument is also a consideration. At a resolving power of 10,000, the modeled signal from the  $(\text{M}-14\text{H}^+)^{14-}$  charge state of an 84mer PCR product is poorly characterized and assignment of the charge state or exact mass is impossible. At a resolving power of 33,000, the peaks from the individual isotopic components are visible. At a resolving power of 100,000, the isotopic peaks are resolved to the baseline and assignment of the charge state for the ion is straightforward. The  $(^{13}\text{C}, ^{15}\text{N})$ -depleted triphosphates are obtained, for example, by growing microorganisms on depleted media and harvesting the nucleotides (Batey *et al.*, *Nucl. Acids Res.*, **1992**, *20*, 4515-4523).

While mass measurements of intact nucleic acid regions are believed to be adequate to determine most bioagents, tandem mass spectrometry ( $MS^n$ ) techniques may provide more definitive information pertaining to molecular identity or sequence. Tandem MS involves the coupled use of two or more stages of mass analysis where both the separation and detection steps are based on mass spectrometry. The first stage is used to select an ion or component of a sample from which further structural information is to be obtained. The selected ion is then fragmented using, e.g., blackbody irradiation, infrared multiphoton dissociation, or collisional activation. For example, ions generated by electrospray ionization (ESI) can be fragmented using IR multiphoton dissociation. This activation leads to dissociation of glycosidic bonds and the phosphate backbone, producing two series of fragment ions, called the *w*-series (having an intact 3' terminus and a 5' phosphate following internal cleavage) and the  $\alpha$ -Base series (having an intact 5' terminus and a 3' furan).

The second stage of mass analysis is then used to detect and measure the mass of these resulting fragments of product ions. Such ion selection followed by fragmentation routines can be performed multiple times so as to essentially completely dissect the molecular sequence of a sample.

If there are two or more targets of similar molecular mass, or if a single amplification reaction results in a product which has the same mass as two or more bioagent reference standards, they can be distinguished by using mass-modifying "tags." In this embodiment of the invention, a nucleotide analog or "tag" is incorporated during amplification (e.g., a 5-(trifluoromethyl) deoxythymidine triphosphate) which has a different molecular weight than the unmodified base so as to improve distinction of masses. Such tags are described in, for example, PCT WO97/33000, which is incorporated herein by reference in its entirety. This further limits the number of possible base compositions consistent with any mass. For example, 5-(trifluoromethyl)deoxythymidine triphosphate can be used in place of dTTP in a separate nucleic acid amplification reaction. Measurement of the mass shift between a conventional amplification product and the tagged product is used to quantitate the number of thymidine nucleotides in each of the single strands. Because the strands are complementary, the number of adenosine nucleotides in each strand is also determined.

In another amplification reaction, the number of G and C residues in each strand is determined using, for example, the cytosine analog 5-methylcytosine (5-meC) or 5-propynylcytosine. The combination of the A/T reaction and G/C reaction, followed by molecular

weight determination, provides a unique base composition. This method is summarized in Figure 4 and Table 1.

**Table 1: Mass Tagging of G and C Residues**

| Mass tag              | Double strand sequence      | Single strand Sequence | Total mass this strand | Base info this strand | Base info other strand | Total base comp. Top strand | Total base comp. Bottom strand |
|-----------------------|-----------------------------|------------------------|------------------------|-----------------------|------------------------|-----------------------------|--------------------------------|
| T*.mass<br>(T*-T) = x | T*ACGT*ACGT*<br>AT*GCAT*GCA | T*ACGT*ACGT*           | 3x                     | 3T                    | 3A                     | 3T<br>2A<br>2C<br>2G        | 3A<br>2T<br>2G<br>2C           |
|                       |                             | AT*GCAT*GCA            | 2x                     | 2T                    | 2A                     |                             |                                |
| C*.mass<br>(C*-C) = y | TAC*GTAC*GT<br>ATGC*ATGC*A  | TAC*GTAC*GT            | 2x                     | 2C                    | 2G                     |                             |                                |
|                       |                             | ATGC*ATGC*A            | 2x                     | 2C                    | 2G                     |                             |                                |

5 The mass tag phosphorothioate A (A\*) was used to distinguish a *Bacillus anthracis* cluster. The *B. anthracis* (A<sub>14</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) had an average MW of 14072.26, and the *B. anthracis* (A<sub>1</sub>A\*<sub>13</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) had an average molecular weight of 14281.11 and the phosphorothioate A had an average molecular weight of +16.06 as determined by ESI-TOF MS. The deconvoluted spectra are shown in Figure 5.

10 In another example, assume the measured molecular masses of each strand are 30,000.115Da and 31,000.115 Da respectively, and the measured number of dT and dA residues are (30,28) and (28,30). If the molecular mass is accurate to 100 ppm, there are 7 possible combinations of dG+dC possible for each strand. However, if the measured molecular mass is accurate to 10 ppm, there are only 2 combinations of dG+dC, and at 1 ppm accuracy there is  
15 only one possible base composition for each strand.

Signals from the mass spectrometer may be input to a maximum-likelihood detection and classification algorithm such as is widely used in radar signal processing. The detection processing uses matched filtering of BCS observed in mass-basecount space and allows for detection and subtraction of signatures from known, harmless organisms, and for detection of  
20 unknown bioagent threats. Comparison of newly observed bioagents to known bioagents is also

possible, for estimation of threat level, by comparing their BCS to those of known organisms and to known forms of pathogenicity enhancement, such as insertion of antibiotic resistance genes or toxin genes.

Processing may end with a Bayesian classifier using log likelihood ratios developed  
5 from the observed signals and average background levels. The program emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of a priori expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database (e.g. GenBank) is used to  
10 define the mass basecount matched filters. The database contains known threat agents and benign background organisms. The latter is used to estimate and subtract the 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  
15 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.

#### **G. Base Composition Signatures as Indices of Bioagent Identifying Amplicons**

20 Although the molecular mass of amplification products obtained using intelligent primers provides a means for identification of bioagents, conversion of molecular mass data to a base composition signature is useful for certain analyses. As used herein, a “base composition signature” (BCS) is the exact base composition determined from the molecular mass of a bioagent identifying amplicon. In one embodiment, a BCS provides an index of a specific gene  
25 in a specific organism.

Base compositions, like sequences, vary slightly from isolate to isolate within species. It is possible to manage this diversity by building “base composition probability clouds” around the composition constraints for each species. This permits identification of organisms in a fashion similar to sequence analysis. A “pseudo four-dimensional plot” can be used to visualize the  
30 concept of base composition probability clouds. Optimal primer design requires optimal choice of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual bioagents. Areas where clouds overlap indicate regions that may result

in a misclassification, a problem which is overcome by selecting primers that provide information from different bioagent identifying amplicons, ideally maximizing the separation of base compositions. Thus, one aspect of the utility of an analysis of base composition probability clouds is that it provides a means for screening primer sets in order to avoid potential  
5 misclassifications of BCS and bioagent identity.

Another aspect of the utility of base composition probability clouds is that they provide a means for predicting the identity of a bioagent whose exact measured BCS was not previously observed and/or indexed in a BCS database due to evolutionary transitions in its nucleic acid sequence.

10 It is important to note that, in contrast to probe-based techniques, mass spectrometry determination of base composition does not require prior knowledge of the composition in order to make the measurement, only to interpret the results. In this regard, the present invention provides bioagent classifying information similar to DNA sequencing and phylogenetic analysis at a level sufficient to detect and identify a given bioagent. Furthermore, the process of  
15 determination of a previously unknown BCS for a given bioagent (for example, in a case where sequence information is unavailable) has downstream utility by providing additional bioagent indexing information with which to populate BCS databases. The process of future bioagent identification is thus greatly improved as more BCS indexes become available in the BCS databases.

20 Another embodiment of the present invention is a method of surveying bioagent samples that enables detection and identification of all bacteria for which sequence information is available using a set of twelve broad-range intelligent PCR primers. Six of the twelve primers are "broad range survey primers" herein defined as primers targeted to broad divisions of bacteria (for example, the *Bacillus/Clostridia* group or gamma-proteobacteria). The other six  
25 primers of the group of twelve primers are "division-wide" primers herein defined as primers which provide more focused coverage and higher resolution. This method enables identification of nearly 100% of known bacteria at the species level. A further example of this embodiment of the present invention is a method herein designated "survey/drill-down" wherein a subspecies characteristic for detected bioagents is obtained using additional primers. Examples of such a  
30 subspecies characteristic include but are not limited to: antibiotic resistance, pathogenicity island, virulence factor, strain type, sub-species type, and clade group. Using the survey/drill-down method, bioagent detection, confirmation and a subspecies characteristic can be provided

within hours. Moreover, the survey/drill-down method can be focused to identify bioengineering events such as the insertion of a toxin gene into a bacterial species that does not normally make the toxin.

## 5 H. Use of Bioagent Identifying Amplicons for Identification of RNA Viruses

Coronaviruses represent RNA virus examples of bioagents which can be identified by the methods of the present invention.

Examples of (-)-strand RNA viral genera include arenaviruses, bunyaviruses, and mononegavirales. Species that are members of the arenavirus genus include, but are not limited to, are sabia virus, lassa fever virus, Machupo Virus, Argentine hemorrhagic fever virus, and flexal virus. Species that are members of the bunyavirus genus include, but are not limited, to hantavirus, nairovirus, phlebovirus, hantaan virus, Congo-Crimean hemorrhagic fever, and rift valley fever. Species that are members of the monoegavirales genus include, but are not limited to, filovirus, paramyxovirus, ebola virus, Marburg, and equine morbillivirus.

Examples of (+)-strand RNA viral genera include, but are not limited to, picornaviruses, astroviruses, calciviruses, nidovirales, flaviviruses, and togaviruses. Species of the picornavirus genus include, but are not limited to, coxsackievirus, echovirus, human coxsackievirus A, human echovirus, human enterovirus, human poliovirus, hepatitis A virus, human parechovirus, and human rhinovirus. A species of the astrovirus genus, includes but is not limited to, human astrovirus. Species of the calcivirus genus include, but are not limited to, chiva virus, human calcivirus, and norwalk virus. Species of the nidovirales genus include, but are not limited to coronavirus and torovirus. Species of the flavivirus genus include, but are not limited to, Alfuy virus, Alkhurma virus, Apoi virus, Aroa virus, Bagaza virus, Banzi virus, Batu cave virus, Bouboui virus, Bukalasa bat virus, Bussliquara virus, Cacipacore virus, Carey island virus, Cowbone ridge virus, Dakar bat virus, Deer tick virus, Dengue virus type 1, Dengue virus type 2, Dengue virus type 3, Dengue virus type 4, Edge hill virus, Entebbe bat virus, Flavivirus sp., Gadgets gully virus, Hepatitis C virus, Iguape virus, Ilheus virus, Israel turkey meningoencephalitis virus, Japanese encephalities virus, Jugra virus, Jutiapa virus, Kadam virus, Kedougou virus, Kokobera virus, Koutango virus, Kunjin virus, Kyasanur forest disease virus, Langata virus, Louping III virus, Maeбан virus, Modoc virus, Montana myotic leukoencephalitis virus, Murray Valley encephalitis virus, Naranjal virus, Negishi virus, Ntaya virus, Omsk hemorrhagic fever virus, Phnom-Penh bat virus, Potiskum virus, Powassan virus, Rio bravo

virus, Rocio virus, Royal farm virus, Russian spring-summer encephalitis virus, Saboya virus, Saint Louis encephalitis virus, Sal vieja virus, San perlita virus, Saumarez reef virus, Sepik virus, Sitiawan virus, Sokuluk virus, Spondweni virus, Stratford virus, Tembusu virus, Tick-borne encephalitis virus, Tyulenly virus, Uganda 5 virus, Usutu virus, West Nile virus, and Yellow fever virus. Species of the togavirus genus include, but are not limited to, Chikugunya virus, Eastern equine encephalitis virus, Mayaro virus, O'nyong-nyong virus, Ross river virus, Venezuelan equine encephalitis virus, Rubella virus, and hepatitis E virus. The hepatitis C virus has a 5'-untranslated region of 340 nucleotides, an open reading frame encoding 9 proteins having 3010 amino acids and a 3'-untranslated region of 240 nucleotides. The 5'-UTR and 3'-UTR are 99% conserved in hepatitis C viruses.

Species of retroviruses include, but are not limited to, human immunodeficiency virus and hepatitis B virus.

In one embodiment of the present invention, the target gene is an RNA-dependent RNA polymerase or a helicase encoded by (+)-strand RNA viruses, or RNA polymerase from a (-)-strand RNA virus. (+)-strand RNA viruses are double stranded RNA and replicate by RNA-directed RNA synthesis using RNA-dependent RNA polymerase and the positive strand as a template. Helicase unwinds the RNA duplex to allow replication of the single stranded RNA. These viruses include viruses from the genera picornaviridae, togaviridae, flaviviridae, arenaviridae, cononaviridae (e.g., human respiratory virus) and Hepatitis A virus. The genes encoding these proteins comprise variable and highly conserved regions which flank the variable regions. The genes can be used to identify the species of the virus and if necessary the strain of the viral species.

In some embodiments of the present invention, RNA viruses are identified by first obtaining RNA from an RNA virus, obtaining corresponding DNA from the RNA via reverse transcription, amplifying the DNA to obtain one or more amplification products using one or more pairs of oligonucleotide primers that bind to conserved regions of the RNA viral genome, which flank a variable region of the genome, determining the molecular mass or base composition of the one or more amplification products and comparing the molecular masses or base compositions with calculated or experimentally determined molecular masses or base compositions of known RNA viruses wherein at least one match identifies the RNA virus.

In one embodiment of the present invention, the RNA virus is a coronavirus. In other embodiments, the coronavirus includes but is not limited to, a member of the following group of

coronaviruses: avian infectious bronchitis, bovine coronavirus, canine coronavirus, feline infectious peritonitis virus, human coronavirus 229E, human coronavirus OC43, murine hepatitis virus, porcine epidemic diarrhea virus, porcine hemagglutinating encephalomyelitis virus, rat sialodacryoadenitis coronavirus, SARS coronavirus, transmissible gastroenteritis virus and  
5 turkey coronavirus.

In other embodiments of the present invention, the intelligent primers produce bioagent identifying amplicons within stable and highly conserved regions of coronaviral genomes. The advantage to characterization of an amplicon in a highly conserved region is that there is a low probability that the region will evolve past the point of primer recognition, in which case, the  
10 amplification step would fail. Such a primer set is thus useful as a broad range survey-type primer. In one embodiment of the present invention, an example of a highly conserved region of coronaviruses is the gene encoding RNA-dependent RNA polymerase (RdRp). In another embodiment of the present invention, the intelligent primers produce bioagent identifying amplicons in a region which evolves more quickly than the stable region described above. The  
15 advantage of characterization bioagent identifying amplicon corresponding to an evolving genomic region is that it is useful for distinguishing emerging strain variants. In another embodiment, an example of an evolving genomic region of coronaviruses is the gene encoding nsp11.

The present invention also has significant advantages as a platform for identification of  
20 diseases caused by emerging coronaviruses. The present invention eliminates the need for prior knowledge of sequence to generate hybridization probes. Thus, in another embodiment, the present invention provides a means of determining the etiology of a coronavirus infection when the process of identification of coronaviruses is carried out in a clinical setting and, even when the coronavirus is a new species never observed before (as used herein, the term "etiology" refers  
25 to the causes or origins, of diseases or abnormal physiological conditions).

This is possible because the methods are not confounded by naturally occurring evolutionary variations (a major concern for characterization of viruses which evolve rapidly) occurring in the sequence acting as the template for production of the bioagent identifying amplicon. Measurement of molecular mass and determination of base composition is accomplished in an  
30 unbiased manner without sequence prejudice.

Another embodiment of the present invention also provides a means of tracking the spread of any species or strain of coronavirus when a plurality of samples obtained from

different locations are analyzed by the methods described above in an epidemiological setting. In one embodiment, a plurality of samples from a plurality of different locations are analyzed with primers which produce bioagent identifying amplicons, a subset of which contain a specific coronavirus. The corresponding locations of the members of the coronavirus-containing subset  
5 indicate the spread of the specific coronavirus to the corresponding locations.

In another embodiment, the present invention also provides kits for carrying out the methods described herein. In some embodiments, the kit may comprise a sufficient quantity of one or more primer pairs to perform an amplification reaction on a target polynucleotide from a bioagent to form a bioagent identifying amplicon. In some embodiments, the kit may comprise  
10 from one to fifty primer pairs, from one to twenty primer pairs, from one to ten primer pairs, or from two to five primer pairs. In some embodiments, the kit may comprise one or more primer pairs recited in Table 2. In some embodiments, the kit may comprise broad range survey primers, division wide primers, or drill-down primers, or any combination thereof. A kit may be designed so as to comprise particular primer pairs for identification of a particular bioagent. For example,  
15 a broad range survey primer kit may be used initially to identify an unknown bioagent as a coronavirus. Another kit may be used to distinguish any coronavirus from any other coronavirus. In some embodiments, any of these kits may be combined to comprise a combination of broad range survey primers and division-wide primers so as to be able to identify the species of an unknown bioagent.

20 The kit may also comprise a sufficient quantity of reverse transcriptase, a DNA polymerase, suitable nucleoside triphosphates (including any of those described above), a DNA ligase, and/or reaction buffer, or any combination thereof, for the amplification processes described above. A kit may further include instructions pertinent for the particular embodiment of the kit, such instructions describing the primer pairs and amplification conditions for  
25 operation of the method. A kit may also comprise amplification reaction containers such as microcentrifuge tubes and the like. A kit may also comprise reagents for isolating bioagent nucleic acid, including, for example, detergent. 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.

30 The present invention is also directed to methods of characterizing a double etiology of a subject presenting at least one symptom of SARS comprising: contacting nucleic acid from a sample from the subject with a first pair of oligonucleotide primers which hybridize to conserved

sequences of a coronavirus, wherein said conserved sequences of a coronavirus flank a variable nucleic acid sequence; contacting nucleic acid from the sample with a second pair of oligonucleotide primers which hybridize to conserved sequences of a putative secondary bioagent(s), wherein the sequences of putative secondary bioagents flank a variable sequence;

5 amplifying the variable nucleic acid sequences between the first pair of primers and the second pair of primers to produce a coronavirus amplification product and a secondary bioagent amplification product; determining the base composition signature of each of the amplification products; using the base composition signatures of each of the amplification products to identify the combination of a SARS-linked coronavirus and a secondary bioagent as a probable cause of

10 the at least one symptom of SARS. In some embodiments, the secondary bioagent correlates with increased severity of the at least one symptom of SARS. In some embodiments, the secondary bioagent correlates with increased incidence of mortality of subjects presenting the at least one symptom of SARS. In some embodiments, the at least one symptom of SARS is high fever ( $>38^{\circ}\text{C}$ ), dry cough, shortness of breath, headache, muscular stiffness, loss of appetite,

15 malaise, confusion, rash, or diarrhea, or any combination thereof. In some embodiments, the double etiology comprises a synergistic viral infection of a SARS-linked coronavirus and a secondary virus. In some embodiments, the secondary virus is adenovirus, parainfluenza virus, respiratory syncytial virus, measles virus, chicken pox virus, or influenza virus, or any combination thereof. In some embodiments, the double etiology comprises a synergistic

20 viral/bacterial infection of a SARS-linked coronavirus and a secondary bacterial agent. In some embodiments, the secondary bacterial agent is *Streptococcus pneumoniae*, *Mycoplasma pneumoniae*, or *Chlamydia trachomatis*, or any combination thereof. In some embodiments, the contacting steps are performed in parallel. In some embodiments, the contacting steps are performed simultaneously.

25 The present invention is also directed to methods of identifying the etiology of a subject presenting at least one symptom of SARS comprising: employing the method described above to rule out the presence of a SARS-linked coronavirus in a sample, wherein lack of amplification of a SARS-linked coronavirus by the first pair of primers indicates absence of a SARS-linked coronavirus, and wherein the base composition signature of the amplification product of the

30 second pair of primers identifies the secondary bioagent, thereby indicating the etiology of the at least one symptom of SARS. In some embodiments, the secondary bioagent is the cause of an acute respiratory infection. In some embodiments, the secondary bioagent is a bacterial agent

such as, for example, *Streptococcus pneumoniae*, *Mycoplasma pneumoniae* or *Chlamydia trachomatis*. In some embodiments, the secondary bioagent is a viral agent such as, for example, adenoviruses, parainfluenza, respiratory syncytial virus, measles virus, chicken pox virus, or influenza virus.

5

## EXAMPLES

### Example 1: Coronavirus Samples, Nucleic Acid Isolation and Amplification

HRT-18 and MRC5 cell lines were inoculated with HCoV-OC43 and HCoV-229E (University of Colorado and Naval Health Research Center, San Diego, CA), HCoV-229E. SARS  
10 RNA was obtained from the CDC (Atlanta, GA) as a 1 mL extract of SARS coronavirus in TRIzol extraction buffer. The SARS CoV-Tor2 strain was obtained from the University of Manitoba as a cell culture supernatant from infected Vero-E6 cells.

RNA was isolated from 250 µL of coronavirus infected cells or culture supernatant using Trizol or Trizol LS respectively (Invitrogen Inc., Carlsbad, CA) according to the  
15 manufacturer's protocol. 5 µg of sheared poly A DNA was added for the precipitation of the RNA. The pelleted nucleic acids were washed in 70% ethanol and resuspended in 100 µL DEPC-treated water containing 20 units of Suprase•In™ (Ambion, Austin, TX). The resuspended RNA was purified using the Qiagen RNAeasy mini kit according to the manufacturer's protocol. The RNA was eluted from the RNAeasy™ columns in 30 µL of DEPC-  
20 treated water and was stored at -70°C.

Purified RNA was primed for reverse transcription by mixing 10 µL of the purified RNA with 5 µL DEPC-treated water containing 500 ng random primers, 1 µg of sheared poly-A DNA and 10 units Suprase•In™. The mixture was heated to 60°C for 5 minutes and then cooled to 4°C. Following the annealing of the random primers to the RNA, 15 µL of first strand reaction  
25 mix consisting of 2x first strand buffer (Invitrogen Inc., Carlsbad, CA), 10 mM DTT, 500 µM dNTPs, and 75 units of SuperScript II was added to the RNA primer mixture. The RNA was reverse transcribed for 45 minutes at 45°C. Various dilutions of the reverse transcription reaction mixes were used directly in the PCR reactions.

All PCR reactions were performed in 50 µL using 96-well microtiter plates and MJ.  
30 Dyad thermocyclers (MJ research, Waltham, MA). The PCR reaction buffer consisted of 4 units of Amplitaq Gold, 1x buffer II (Applied Biosystems, Foster City, CA), 2.0 mM MgCl<sub>2</sub>, 0.4 M betaine, 800 µM dNTP mix, and 250 nM propyne containing PCR primers. The following PCR

conditions were used to amplify coronavirus sequences: 95°C for 10 min followed by 50 cycles of 95°C for 30 sec, 50°C for 30 sec, and 72°C for 30 sec.

20 µL of each crude PCR product was transferred into a 96-well plate. Pretreated anion exchange ZipTips™ (Millipore) were loaded onto the head of a 96-tip Evolution P3 (Perkin Elmer), and then a 20 µL aliquot of crude PCR product was loaded onto each tip by repeatedly aspirating and expelling the 20 µL aliquot. Following sample loading, aliquots of 40 mM NH<sub>4</sub>HCO<sub>3</sub> were used to wash each sample six times to remove unconsumed primers and dNTPs. This step was followed by rinses with 10 µL aliquots of a 20% MeOH solution to remove any residual polymeric material that originated from polymerase or PCR buffer. Elution of the final purified/desalted PCR products was accomplished by rinsing each tip with a 10 µL aliquot of 0.4 M NH<sub>4</sub>OH and dispensing the 10 µL eluent into a well of a 96-well plate. Prior to analysis by ESI-MS, the eluent was diluted 1:1 with a solution containing 50% MeOH and 50 mM piperidine/imidazole. A small oligonucleotide designated SH2 (CGTGCATGGCGG; SEQ ID NO:105, Synthetic Genetics, San Diego, California) was added as an internal mass standard at a final concentration of 50 nM.

## Example 2: Molecular Mass Determination

The mass spectrometer is based on a Bruker Daltonics (Billerica, MA) Apex II 70e electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer (ESI-FTICR-MS) that employs an actively shielded 7 Tesla superconducting magnet. All aspects of pulse sequence control and data acquisition were performed on a 1.1 GHz Pentium II data station running Bruker's Xmass software. 20 µL sample aliquots were extracted directly from 96-well microtiter plates using a CTC HTS PAL autosampler (LEAP Technologies, Carrboro, NC) triggered by the data station. Samples were injected directly into the ESI source at a flow rate of 75 µL/hr. Ions were formed via electrospray ionization in a modified Analytica (Branford, CT) source employing an off axis, grounded electrospray probe positioned ca. 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 N<sub>2</sub>/O<sub>2</sub> was employed to assist in the desolvation process. Ions were accumulated in an external ion reservoir comprised of an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode, prior to injection into the trapped ion cell where they were mass analyzed. Spectral acquisition was performed in the continuous duty cycle mode whereby ions were

accumulated in the hexapole ion reservoir simultaneously with ion detection in the trapped ion cell. Following a 1.2 ms transfer event, in which ions were transferred to the trapped ion cell, the ions were subjected to a 1.6 ms chirp excitation corresponding to 8000 – 500 m/z. Data was acquired over an m/z range of 500 – 5000 (1M data points over a 225K Hz bandwidth). Each spectrum was the result of co-adding 32 transients. Transients were zero-filled once prior to the magnitude mode Fourier transform and post calibration using the internal mass standard. The ICR-2LS software package (G. A. Anderson, J. E. Bruce. (Pacific Northwest National Laboratory, Richland, WA, 1995)) was used to deconvolute the mass spectra and calculate the mass of the monoisotopic species using an “averaging” fitting routine (M. W. Senko, S. C. Beu, F. W. McLafferty, Journal of the American Society for Mass Spectrometry 6, 229 (1995)) modified for DNA. Using this approach, monoisotopic molecular weights were calculated.

### Example 3: Selection of Primers

To design primers that amplify all known coronavirus species and to identify new members, alignments were carried out using all available coronavirus sequences from GenBank, including complete genomes and individual genes, and scanned for regions where pairs of PCR primers would generate bioagent identifying amplicons of a length of about 150 or fewer nucleobases. The current length limit of about 150 nucleobases is dictated by the ability of electrospray mass spectroscopy to determine the mass of a PCR amplification product with sufficient accuracy to unambiguously determine the base composition. One with ordinary skill in the art will recognize that this limit may increase subject to improvements in the art of molecular mass determination of nucleic acids.

Two target regions were selected in coronavirus orf-1b, one in the RNA-dependent RNA polymerase (RdRp) and the other in Nsp11 (Figure 7). Locations of primers within these regions were optimized for sensitivity and broad-range priming potential simultaneously by performing limiting dilutions of multiple, diverse coronaviruses. The primer pair names shown in Table 2 refer to the forward and reverse primer for a given region. Each primer was designed to include a thymidine (T) nucleotide on the 5' end to minimize addition of non-templated adenosine (A) during PCR.

Table 2 represents the collection of intelligent primers (SEQ ID NOs:5-11) designed to identify coronaviruses using the method of the present invention. The forward or reverse primer name indicates the gene region of coronavirus genome to which the primer hybridizes relative to

a reference sequence, in this case, the human coronavirus 229E sequence. The primers represented by SEQ ID NOs:5 and 6 were designed to yield an amplicon originating from a coronavirus nsp11 gene with reference to GenBank Accession No: NC\_002645 (incorporated herein as SEQ ID NO:30). The primers represented by SEQ ID NOs:7-11 were designed to yield an amplicon originating from a coronavirus RNA-dependent RNA polymerase gene with reference to GenBank Accession No: AF304460 (incorporated herein as SEQ ID NO: 31). In Table 2, @ = 5-propynyluracil (which is a chemically modified version of T); & = 5-propynylcytosine (which is a chemically modified version of C).

**Table 2: Primer Pairs for Identification of Coronaviruses**

| Forward primer name             | Forward sequence          | SEQ ID NO: | Reverse primer name              | Reverse sequence          | SEQ ID NO: |
|---------------------------------|---------------------------|------------|----------------------------------|---------------------------|------------|
| CV_NC002645_18<br>190 18215 F   | TGTTTGTTTTGAATTGTAATGTTGA | 5          | CV_NC002645_183<br>02 18326 R    | TGGAATGCATGCTTATTAACATACA | 6          |
| CV_NC002645_18<br>190 18215P F  | TGTTTG&&&GGAATTGTAATGTTGA | 5          | CV_NC002645_183<br>02 18326P R   | TGGAATGCATGC&&A&&AACATACA | 6          |
| CVPOL_AF304460<br>1737 1757 F   | TAAGTTTATGGCGGGTGGGA      | 7          | CVPOL_AF304460<br>1804 1824 R    | TTTAGGGTAGTCCCAACCCAT     | 8          |
| CVPOL_AF304460<br>1737 1755 F   | TAAGTTTATGGCGGGTGGG       | 9          | CVPOL_AF304460<br>1804 1824 2 R  | TTTAGGGTAGTCCCAACCCAT     | 8          |
| CVPOL_AF304460<br>1737 1755P F  | TAAG&&T&ATGGCGGG&GG       | 9          | CVPOL_AF304460<br>1804 1824 2P R | TTTAGGATAGT@@@AACCCAT     | 10         |
| CVPOL_AF304460<br>1737 1755 2 F | TAAGTTTATGGCGGCTGG        | 11         | CVPOL_AF304460<br>1804 1824 2 R  | TTTAGGGTAGTCCCAACCCAT     | 8          |
| CVPOL_AF304460<br>1737_1755_2P_ | TAAG&&T&ATGGCGGC&GG       | 11         | CVPOL_AF304460<br>1804 1824 2P R | TTTAGGATAGT@@@AACCCAT     | 10         |
| CVPOL_AF304460<br>1737 1755 F   | TAAGTTTATGGCGGGTGG        | 9          | CVPOL_AF304460<br>1804 1824 3 R  | TTTAGGATAGTCCCAACCCAT     | 10         |

**Example 4: Database of Bioagent Identifying Amplicons for Coronaviruses**

A database of expected molecular masses and base compositions of bioagent identifying amplicons was generated using an electronic PCR search algorithm (ePCR). An existing RNA structure search algorithm (T. Macke et al., *Nuc. Acids Res.* 29, 4724 (2001)) was modified to include PCR parameters such as hybridization conditions, mismatches, and thermodynamic calculations (J. SantaLucia, *Proc. Natl. Acad. Sci. U.S.A* 95, 1460 (1998)). ePCR was used first to check primer specificity and the selected primer pairs were searched against GenBank nucleotide sequence database for matches to the primer sequences. ePCR showed that the coronavirus primers should prime all known coronaviruses in GenBank, but should not prime bacterial, viral, or human DNA sequences. For each match, A, G, C, and T base counts of the predicted amplicon sequence were calculated and a database of coronavirus bioagent identifying amplicons was created (Table 3).

Shown in Table 3 are molecular masses and base compositions of both strands of bioagent identifying amplicons for a series of different coronaviruses obtained using primer sets

CV\_NC002645\_18190\_18215P\_F (nsp11 primer set SEQ ID NOs: 5 and 6) and VPOL\_AF304460\_1737\_1755P\_F (RdRp primer set SEQ ID NOs: 9 and 10).

**Table 3: Database of Molecular Masses and Base Compositions for Coronavirus Bioagent Identifying Amplicons**

| Prime<br>r<br>Pair | Corona-virus                               | STRAIN             | ACCESSION   | SEQ<br>ID<br>NO: | Forward<br>Strand Exact<br>Mass | Forward<br>Strand<br>Base<br>Compo-<br>sition | Opposite<br>Strand Exact<br>Mass | Opposite<br>Strand<br>Base<br>Compo-<br>sition |
|--------------------|--------------------------------------------|--------------------|-------------|------------------|---------------------------------|-----------------------------------------------|----------------------------------|------------------------------------------------|
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | Beaudette          | Z30541.1    | 12               | 27396.5334                      | A24 G24<br>C14 T26                            | 27032.54201                      | A26 G14<br>C24 T24                             |
| nsp11              | Avian<br>infectious<br>bronchitis<br>virus | Beaudette          | Z30541.1    | 12               | 42530.9462                      | A33 G32<br>C17 T55                            | 42129.10847                      | A55 G17<br>C32 T33                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | Beaudette          | Z69629.1    | 13               | 27396.5334                      | A24 G24<br>C14 T26                            | 27032.54201                      | A26 G14<br>C24 T24                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | Beaudette          | M95169.1    | 14               | 27396.5334                      | A24 G24<br>C14 T26                            | 27032.54201                      | A26 G14<br>C24 T24                             |
| nsp11              | Avian<br>infectious<br>bronchitis<br>virus | Beaudette          | M95169.1    | 14               | 42530.9462                      | A33 G32<br>C17 T55                            | 42129.10847                      | A55 G17<br>C32 T33                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | Beaudette          | NC_001451.1 | 15               | 27396.5334                      | A24 G24<br>C14 T26                            | 27032.54201                      | A26 G14<br>C24 T24                             |
| nsp11              | Avian<br>infectious<br>bronchitis<br>virus | Beaudette          | NC_001451.1 | 15               | 42530.9462                      | A33 G32<br>C17 T55                            | 42129.10847                      | A55 G17<br>C32 T33                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | Beaudette<br>(M42) | M94356.1    | 16               | 27396.5334                      | A24 G24<br>C14 T26                            | 27032.54201                      | A26 G14<br>C24 T24                             |
| nsp11              | Avian<br>infectious<br>bronchitis<br>virus | Beaudette<br>(M42) | M94356.1    | 16               | 42530.9462                      | A33 G32<br>C17 T55                            | 42129.10847                      | A55 G17<br>C32 T33                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | Beaudette<br>CK    | AJ311317.1  | 17               | 27396.5334                      | A24 G24<br>C14 T26                            | 27032.54201                      | A26 G14<br>C24 T24                             |
| nsp11              | Avian<br>infectious<br>bronchitis<br>virus | Beaudette<br>CK    | AJ311317.1  | 17               | 42530.9462                      | A33 G32<br>C17 T55                            | 42129.10847                      | A55 G17<br>C32 T33                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | BJ                 | AY319651.1  | 18               | 27413.56129                     | A27 G23<br>C13 T25                            | 27013.52362                      | A25 G13<br>C23 T27                             |
| nsp11              | Avian<br>infectious<br>bronchitis<br>virus | BJ                 | AY319651.1  | 18               | 42502.91625                     | A31 G31<br>C16 T59                            | 42155.14792                      | A59 G16<br>C31 T31                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | LX4                | AY223860.1  | 19               | 27419.54938                     | A26 G23<br>C12 T27                            | 27006.54027                      | A27 G12<br>C23 T26                             |
| RdRp               | Avian<br>infectious<br>bronchitis<br>virus | LX4                | AY338732.1  | 20               | 27419.54938                     | A26 G23<br>C12 T27                            | 27006.54027                      | A27 G12<br>C23 T26                             |
| nsp11              | Avian<br>infectious<br>bronchitis<br>virus | LX4                | AY338732.1  | 20               | 42532.97442                     | A36 G31<br>C17 T53                            | 42126.085                        | A53 G17<br>C31 T36                             |
| RdRp               | Bovine<br>coronavirus                      | BCoV-ENT           | AF391541.1  | 21               | 27358.49663                     | A22 G22<br>C12 T32                            | 27066.59778                      | A32 G12<br>C22 T22                             |
| nsp11              | Bovine                                     | BCoV-ENT           | AF391541.1  | 21               | 42606.00337                     | A38 G32                                       | 42052.0608                       | A52 G15                                        |

|       |                                     |              |             |    |             |                    |             |                    |
|-------|-------------------------------------|--------------|-------------|----|-------------|--------------------|-------------|--------------------|
|       | coronavirus                         |              |             |    |             | C15 T52            |             | C32 T38            |
| RdRp  | Bovine coronavirus                  | BCoV-LUN     | AF391542.1  | 22 | 27358.49663 | A22 G22<br>C12 T32 | 27066.59778 | A32 G12<br>C22 T22 |
| nsp11 | Bovine coronavirus                  | BCoV-LUN     | AF391542.1  | 22 | 42606.00337 | A38 G32<br>C15 T52 | 42052.0608  | A52 G15<br>C32 T38 |
| RdRp  | Bovine coronavirus                  | Mebus        | U00735.2    | 23 | 27358.49663 | A22 G22<br>C12 T32 | 27066.59778 | A32 G12<br>C22 T22 |
| nsp11 | Bovine coronavirus                  | Mebus        | U00735.2    | 23 | 42606.00337 | A38 G32<br>C15 T52 | 42052.0608  | A52 G15<br>C32 T38 |
| RdRp  | Bovine coronavirus                  | Quebec       | AF220295.1  | 24 | 27358.49663 | A22 G22<br>C12 T32 | 27066.59778 | A32 G12<br>C22 T22 |
| nsp11 | Bovine coronavirus                  | Quebec       | AF220295.1  | 24 | 42591.0037  | A38 G32<br>C16 T51 | 42068.05572 | A51 G16<br>C32 T38 |
| RdRp  | Bovine coronavirus                  | Unknown      | AF124985.1  | 25 | 27358.49663 | A22 G22<br>C12 T32 | 27066.59778 | A32 G12<br>C22 T22 |
| RdRp  | Bovine coronavirus                  | Unknown      | NC_003045.1 | 26 | 27358.49663 | A22 G22<br>C12 T32 | 27066.59778 | A32 G12<br>C22 T22 |
| nsp11 | Bovine coronavirus                  | Unknown      | NC_003045.1 | 26 | 42606.00337 | A38 G32<br>C15 T52 | 42052.0608  | A52 G15<br>C32 T38 |
| RdRp  | Canine coronavirus                  | 1-71; VR-809 | AF124986.1  | 27 | 27486.53139 | A24 G24<br>C8 T32  | 26936.57252 | A32 G8<br>C24 T24  |
| RdRp  | Canine respiratory coronavirus      | respiratory  | AY150273.1  | 28 | 27343.49696 | A22 G22<br>C13 T31 | 27082.59269 | A31 G13<br>C22 T22 |
| RdRp  | Feline infectious peritonitis virus | UCD2         | AF124987.1  | 29 | 27472.52698 | A23 G25<br>C10 T30 | 26953.56268 | A30 G10<br>C25 T23 |
| RdRp  | Human coronavirus 229E              | 229E         | NC_002645.1 | 30 | 27450.54396 | A25 G24<br>C11 T28 | 26975.54569 | A28 G11<br>C24 T25 |
| nsp11 | Human coronavirus 229E              | 229E         | NC_002645.1 | 30 | 42462.96894 | A36 G30<br>C20 T51 | 42198.08098 | A51 G20<br>C30 T36 |
| RdRp  | Human coronavirus 229E              | 229E         | AF304460.1  | 31 | 27450.54396 | A25 G24<br>C11 T28 | 26975.54569 | A28 G11<br>C24 T25 |
| nsp11 | Human coronavirus 229E              | 229E         | AF304460.1  | 31 | 42462.96894 | A36 G30<br>C20 T51 | 42198.08098 | A51 G20<br>C30 T36 |
| RdRp  | Human coronavirus 229E              | 229E         | X69721.1    | 32 | 27450.54396 | A25 G24<br>C11 T28 | 26975.54569 | A28 G11<br>C24 T25 |
| nsp11 | Human coronavirus 229E              | 229E         | X69721.1    | 32 | 42462.96894 | A36 G30<br>C20 T51 | 42198.08098 | A51 G20<br>C30 T36 |
| RdRp  | Human coronavirus OC43              | OC43         | AF124989.1  | 33 | 27328.4973  | A22 G22<br>C14 T30 | 27098.58761 | A30 G14<br>C22 T22 |
| RdRp  | Murine hepatitis virus              | A59          | X51939.1    | 34 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| nsp11 | Murine hepatitis virus              | A59          | X51939.1    | 34 | 42599.97755 | A34 G35<br>C18 T50 | 42064.05812 | A50 G18<br>C35 T34 |
| RdRp  | Murine hepatitis virus              | JHM          | M55148.1    | 35 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| nsp11 | Murine hepatitis virus              | JHM          | M55148.1    | 35 | 42529.97207 | A34 G34<br>C21 T48 | 42136.05409 | A48 G21<br>C34 T34 |
| RdRp  | Murine hepatitis virus              | MHV-2        | AF201929.1  | 36 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| nsp11 | Murine hepatitis virus              | MHV-2        | AF201929.1  | 36 | 42576.99929 | A37 G33<br>C18 T49 | 42085.04588 | A49 G18<br>C33 T37 |
| RdRp  | Murine hepatitis virus              | MHV-A59      | NC_001846.1 | 37 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| nsp11 | Murine hepatitis virus              | MHV-A59      | NC_001846.1 | 37 | 42599.97755 | A34 G35<br>C18 T50 | 42064.05812 | A50 G18<br>C35 T34 |
| RdRp  | Murine hepatitis virus              | MHV-A59      | AF029248.1  | 38 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| nsp11 | Murine hepatitis virus              | MHV-A59      | AF029248.1  | 38 | 42599.97755 | A34 G35<br>C18 T50 | 42064.05812 | A50 G18<br>C35 T34 |
| RdRp  | Murine hepatitis virus              | ML-10        | AF208067.1  | 39 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |

|       |                                                  |             |             |    |             |                    |             |                    |
|-------|--------------------------------------------------|-------------|-------------|----|-------------|--------------------|-------------|--------------------|
| nsp11 | Murine hepatitis virus                           | ML-10       | AF208067.1  | 39 | 42599.97755 | A34 G35<br>C18 T50 | 42064.05812 | A50 G18<br>C35 T34 |
| RdRp  | Murine hepatitis virus                           | ML-11       | AF207902.1  | 40 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| nsp11 | Murine hepatitis virus                           | ML-11       | AF207902.1  | 40 | 42576.99929 | A37 G33<br>C18 T49 | 42085.04588 | A49 G18<br>C33 T37 |
| RdRp  | Murine hepatitis virus                           | Penn 97-1   | AF208066.1  | 41 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| nsp11 | Murine hepatitis virus                           | Penn 97-1   | AF208066.1  | 41 | 42576.99929 | A37 G33<br>C18 T49 | 42085.04588 | A49 G18<br>C33 T37 |
| RdRp  | Porcine epidemic diarrhea virus                  | CV777       | AF353511.1  | 42 | 27392.51468 | A23 G23<br>C12 T30 | 27033.57498 | A30 G12<br>C23 T23 |
| nsp11 | Porcine epidemic diarrhea virus                  | CV777       | AF353511.1  | 42 | 42459.90775 | A29 G33<br>C21 T54 | 42205.12316 | A54 G21<br>C33 T29 |
| RdRp  | Porcine epidemic diarrhea virus                  | CV777       | NC_003436.1 | 43 | 27392.51468 | A23 G23<br>C12 T30 | 27033.57498 | A30 G12<br>C23 T23 |
| nsp11 | Porcine epidemic diarrhea virus                  | CV777       | NC_003436.1 | 43 | 42459.90775 | A29 G33<br>C21 T54 | 42205.12316 | A54 G21<br>C33 T29 |
| RdRp  | Porcine hemagglutinating encephalomyelitis virus | VR-741      | AF124988.1  | 44 | 27343.49696 | A22 G22<br>C13 T31 | 27082.59269 | A31 G13<br>C22 T22 |
| RdRp  | Rat sialodacryoadenitis coronavirus              | 8190        | AF124990.1  | 45 | 27344.49221 | A21 G23<br>C14 T30 | 27083.58794 | A30 G14<br>C23 T21 |
| RdRp  | SARS coronavirus                                 | BJ01        | AY278488.1  | 46 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | BJ01        | AY278488.1  | 46 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                                 | BJ01        | AY278488.2  | 47 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| nsp11 | SARS coronavirus                                 | BJ02        | AY278487.3  | 48 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                                 | BJ02        | AY278487.1  | 49 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | BJ02        | AY278487.1  | 49 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | BJ03        | AY278490.1  | 50 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | BJ03        | AY278490.1  | 50 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                                 | BJ03        | AY278490.3  | 51 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| nsp11 | SARS coronavirus                                 | BJ04        | AY279354.2  | 52 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                                 | BJ04        | AY279354.1  | 53 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | BJ04        | AY279354.1  | 53 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | CUHK-Su10   | AY282752.1  | 54 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                                 | CUHK-Su10   | AY282752.1  | 54 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| nsp11 | SARS coronavirus                                 | CUHK-W1     | AY278554.2  | 55 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                                 | CUHK-W1     | AY278554.1  | 56 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | CUHK-W1     | AY278554.1  | 56 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus                                 | FRA         | AY310120.1  | 57 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                                 | FRA         | AY310120.1  | 57 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                                 | Frankfurt 1 | AY291315.1  | 58 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                                 | Frankfurt 1 | AY291315.1  | 58 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |

|       |                  |              |            |    |             |                    |             |                    |
|-------|------------------|--------------|------------|----|-------------|--------------------|-------------|--------------------|
| nsp11 | SARS coronavirus | GD01         | AY278489.2 | 59 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | GD01         | AY278489.1 | 60 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus | GZ01         | AY278489.1 | 60 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus | HKU-39849    | AY278491.2 | 61 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | HKU-39849    | AY278491.2 | 61 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| nsp11 | SARS coronavirus | HSR 1        | AY323977.2 | 62 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | HSR 1        | AY323977.1 | 63 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus | HSR 1        | AY323977.1 | 63 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus | Shanghai LY  | AY322206.1 | 64 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus | Sin2500      | AY283794.1 | 65 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Sin2500      | AY283794.1 | 65 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | Sin2677      | AY283795.1 | 66 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Sin2677      | AY283795.1 | 66 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | Sin2679      | AY283796.1 | 67 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Sin2679      | AY283796.1 | 67 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | Sin2748      | AY283797.1 | 68 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Sin2748      | AY283797.1 | 68 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| nsp11 | SARS coronavirus | Shanghai QXC | AY322198.1 | 69 | 42504.96559 | A34 G33<br>C21 T49 | 42160.06533 | A49 G21<br>C33 T34 |
| RdRp  | SARS coronavirus | Sin2774      | AY283798.1 | 69 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Sin2774      | AY283798.1 | 69 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | Taiwan TC1   | AY338174.1 | 70 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Taiwan TC1   | AY338174.1 | 70 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | Taiwan TC2   | AY338175.1 | 71 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Taiwan TC2   | AY338175.1 | 71 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | Taiwan TC3   | AY348314.1 | 72 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Taiwan TC3   | AY348314.1 | 72 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | Tor2         | AY274119.1 | 73 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus | Tor2         | AY274119.1 | 73 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| RdRp  | SARS coronavirus | Tor2         | AY274119.2 | 74 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | Tor2         | AY274119.3 | 75 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | TW1          | AY291451.1 | 76 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | TW1          | AY291451.1 | 76 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | TWC          | AY321118.1 | 77 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | TWC          | AY321118.1 | 77 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | TWC2         | AY362698.1 | 78 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | TWC2         | AY362698.1 | 78 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | TWC3         | AY362699.1 | 79 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | TWC3         | AY362699.1 | 79 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus | TWH          | AP006557.1 | 80 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus | TWH          | AP006557.1 | 80 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS             | TWJ          | AP006558.1 | 81 | 27298.53569 | A27 G19            | 27125.56347 | A28 G14            |

|       |                                     |                     |               |    |             |                    |             |                    |
|-------|-------------------------------------|---------------------|---------------|----|-------------|--------------------|-------------|--------------------|
|       | coronavirus                         |                     |               |    |             | C14 T28            |             | C19 T27            |
| nsp11 | SARS coronavirus                    | TWJ                 | AP006558.1    | 81 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                    | TWK                 | AP006559.1    | 82 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                    | TWK                 | AP006559.1    | 82 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                    | TWS                 | AP006560.1    | 83 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                    | TWS                 | AP006560.1    | 83 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                    | TWY                 | AP006561.1    | 84 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                    | TWY                 | AP006561.1    | 84 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                    | Unknown             | NC_004718.3   | 85 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                    | Unknown             | NC_004718.3   | 85 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                    | Urbani              | AY278741.1    | 86 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                    | Urbani              | AY278741.1    | 86 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                    | ZJ01                | AY297028.1    | 87 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                    | ZJ01                | AY297028.1    | 87 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | SARS coronavirus                    | ZMY 1               | AY351680.1    | 88 | 27298.53569 | A27 G19<br>C14 T28 | 27125.56347 | A28 G14<br>C19 T27 |
| nsp11 | SARS coronavirus                    | ZMY 1               | AY351680.1    | 88 | 42519.96525 | A34 G33<br>C20 T50 | 42144.07041 | A50 G20<br>C33 T34 |
| RdRp  | Transmissible gastroenteritis virus | Purdue              | AJ271965.2    | 89 | 27486.53139 | A24 G24<br>C8 T32  | 26936.57252 | A32 G8<br>C24 T24  |
| nsp11 | Transmissible gastroenteritis virus | Purdue              | AJ271965.2    | 89 | 42465.93357 | A33 G30<br>C18 T56 | 42193.12585 | A56 G18<br>C30 T33 |
| RdRp  | Transmissible gastroenteritis virus | Purdue              | NC_002306.2   | 90 | 27486.53139 | A24 G24<br>C8 T32  | 26936.57252 | A32 G8<br>C24 T24  |
| nsp11 | Transmissible gastroenteritis virus | Purdue              | NC_002306.2   | 90 | 42465.93357 | A33 G30<br>C18 T56 | 42193.12585 | A56 G18<br>C30 T33 |
| RdRp  | Transmissible gastroenteritis virus | Purdue-115          | Z34093.1      | 91 | 27486.53139 | A24 G24<br>C8 T32  | 26936.57252 | A32 G8<br>C24 T24  |
| nsp11 | Transmissible gastroenteritis virus | Purdue-115          | Z34093.1      | 91 | 42465.93357 | A33 G30<br>C18 T56 | 42193.12585 | A56 G18<br>C30 T33 |
| RdRp  | Transmissible gastroenteritis virus | Unknown             | AJ011482.1    | 92 | 27486.53139 | A24 G24<br>C8 T32  | 26936.57252 | A32 G8<br>C24 T24  |
| nsp11 | Transmissible gastroenteritis virus | Unknown             | AJ011482.1    | 92 | 42465.93357 | A33 G30<br>C18 T56 | 42193.12585 | A56 G18<br>C30 T33 |
| RdRp  | Transmissible gastroenteritis virus | Unknown             | AF124992.1    | 93 | 27486.53139 | A24 G24<br>C8 T32  | 26936.57252 | A32 G8<br>C24 T24  |
| RdRp  | Turkey coronavirus                  | ATCCVR-911          | AF124991.1    | 94 | 27402.5215  | A23 G24<br>C13 T28 | 27025.55866 | A28 G13<br>C24 T23 |
| nsp11 | Canine coronavirus                  | 1-71; VR-809        | Seq. In-house | 95 | 42475.94038 | A33 G31<br>C19 T54 | 42185.10953 | A54 G19<br>C31 T33 |
| RdRp  | Canine coronavirus                  | CCV-TN449; VR2068   | Seq. In-house | 96 | 27471.53173 | A24 G24<br>C9 T31  | 26952.56743 | A31 G9<br>C24 T24  |
| nsp11 | Canine coronavirus                  | CCV-TN449; VR-2068  | Seq. In-house | 96 | 42524.9581  | A34 G32<br>C18 T53 | 42136.09182 | A53 G18<br>C32 T34 |
| RdRp  | Feline infectious peritonitis virus | WSU 79-1683; VR-989 | Seq. In-house | 97 | 27471.53173 | A24 G24<br>C9 T31  | 26952.56743 | A31 G9<br>C24 T24  |
| nsp11 | Feline                              | WSU 79-             | Seq. In-      | 97 | 42490.94005 | A33 G31            | 42169.11462 | A55 G18            |

|       |                                     |              |               |     |             |                    |             |                    |
|-------|-------------------------------------|--------------|---------------|-----|-------------|--------------------|-------------|--------------------|
|       | infectious peritonitis virus        | 1683; VR-989 | house         |     |             | C18 T55            |             | C31 T33            |
| nsp11 | Feline infectious peritonitis virus | DF2; VR-2004 | Seq. In-house | 98  | 42500.94687 | A33 G32<br>C19 T53 | 42161.0983  | A53 G19<br>C32 T33 |
| nsp11 | Human coronavirus OC43              | OC43; NHRC   | Seq. In-house | 99  | 42580.99689 | A38 G31<br>C15 T53 | 42076.07203 | A53 G15<br>C31 T38 |
| nsp11 | Rat sialodacryoadenitis coronavirus | 8190; VR1410 | Seq. In-house | 100 |             | A34 G34<br>C20 T49 |             | A49 G20<br>C34 T34 |
| nsp11 | Murine hepatitis virus              | MHV1-VR261   | Seq. In-house | 101 | 42602.00577 | A37 G34<br>C18 T48 | 42061.03465 | A48 G18<br>C34 T37 |

Entries with “Seq. In-house” indicate that a GenBank record did not exist at the time of assembly of the database of Table 3. To verify the experimentally measured base compositions, approximately 500 base pair (bp) regions flanking each target region used in this study were 5 sequenced. The regions surrounding the target regions (615 bp for nsp11 and 454 bp for RdRp) were amplified using primers containing 5' M13 sequencing tags. Methods of sequencing are well known to those with ordinary skill in the art.

#### Example 5: Characterization of Bioagent Identifying Amplicons for Coronaviruses

10 For broad-range detection of all coronaviruses, two PCR primer target regions in orf-1b, one in the RNA-dependent RNA polymerase (RdRp) and the other in Nsp11 were identified based on the analyses described in Examples 3 and 4. Locations of primers within these regions were optimized both for sensitivity and broad-range priming potential simultaneously by performing limiting dilutions of multiple, diverse coronaviruses. Analysis of the final primer 15 pairs by ePCR of GenBank nucleotide database sequences showed that these primers would be expected to amplify all the known coronaviruses but no other viruses, bacteria, or human DNA. PCR products for each virus listed in Table 4 were generated, desalted, and analyzed by electrospray ionization Fourier transform ion cyclotron mass spectrometry (FTICR-MS) indicated in Examples 1 and 2. The spectral signals were algorithmically processed to yield base 20 composition data. Figure 8 is a schematic representation of electrospray ionization, strand separation, and the actual charge state distributions of the separated sense and antisense strands, and determination of molecular mass and base composition of the PCR products from the RdRp primer pair for the SARS coronavirus.

Due to the accuracy of FTICR-MS (mass measurement error  $\pm 1$  ppm), all detected 25 masses could be unambiguously mapped to the base compositions of sense and antisense strands. The results from analysis of 14 coronavirus isolates are shown in Table 4.

**Table 4: Experimentally Determined Molecular Masses and Base Compositions for a Selected Set of Coronaviruses**

| Corona-virus Species              | Strain               | Strand    | Experimentally Determined Monoisotopic Masses (RdRp) | Calculated Base Compositions (RdRp) | Experimentally Determined Monoisotopic Masses (nsp11) | Calculated Base Compositions (nsp11) |
|-----------------------------------|----------------------|-----------|------------------------------------------------------|-------------------------------------|-------------------------------------------------------|--------------------------------------|
| Canine                            | 1-71                 | Sense     | 27486.514                                            | A24 G24 C8 T32                      | 42475.955                                             | A33 G31 C19 T54                      |
|                                   |                      | Antisense | 26936.574                                            | A32 G8 C24 T24                      | 42185.117                                             | A54 G19 C31 T33                      |
| Canine                            | CCV-TN449            | Sense     | 27471.510                                            | A24 G24 C9 T31                      | 42474.899                                             | A34 G30 C18 T55                      |
|                                   |                      | Antisense | 26952.548                                            | A31 G9 C24 T24                      | 42184.072                                             | A55 G18 C30 T34                      |
| Feline                            | WSU 79-1683          | Sense     | 27471.517                                            | A24 G24 C9 T31                      | 42490.945                                             | A33 G31 C18 T55                      |
|                                   |                      | Antisense | 26952.556                                            | A31 G9 C24 T24                      | 42169.118                                             | A55 G18 C31 T33                      |
| Feline                            | DF2                  | Sense     | 27472.497                                            | A23 G25 C10 T30                     | 42450.904                                             | A33 G30 C19 T55                      |
|                                   |                      | Antisense | 26953.536                                            | A30 G10 C25 T23                     | 42209.081                                             | A55 G19 C30 T33                      |
| Human 229E                        | 229E                 | Sense     | 27450.532                                            | A25 G24 C11 T28                     | 42462.994                                             | A36 G30 C20 T51                      |
|                                   |                      | Antisense | 26975.545                                            | A28 G11 C24 T25                     | 42198.061                                             | A51 G20 C30 T36                      |
| Human 229E                        | 229E                 | Sense     | 27450.506                                            | A25 G24 C11 T28                     | 42462.930                                             | A36 G30 C20 T51                      |
|                                   |                      | Antisense | 26975.512                                            | A28 G11 C24 T25                     | 42198.040                                             | A51 G20 C30 T36                      |
| Bovine                            | Calf Diarrheal virus | Sense     | 27358.452                                            | A22 G22 C12 T32                     | 42606.039                                             | A38 G32 C15 T52                      |
|                                   |                      | Antisense | 27066.586                                            | A32 G12 C22 T22                     | 42052.897                                             | A52 G15 C32 T38                      |
| Human OC43                        | OC43                 | Sense     | 27328.473                                            | A22 G22 C14 T30                     | 42580.959                                             | A38 G31 C15 T53                      |
|                                   |                      | Antisense | 27098.562                                            | A30 G14 C22 T22                     | 42076.028                                             | A53 G15 C31 T38                      |
| Murine Hepatitis Virus            | MHV1                 | Sense     | 27344.491                                            | A21 G23 C14 T30                     | 42602.022                                             | A37 G34 C18 T48                      |
|                                   |                      | Antisense | 27083.564                                            | A30 G14 C23 T21                     | 42061.016                                             | A48 G18 C34 T37                      |
| Murine Hepatitis Virus            | JHM-thermostable     | Sense     | 27344.497                                            | A21 G23 C14 T30                     | 42529.960                                             | A34 G34 C21 T48                      |
|                                   |                      | Antisense | 27083.571                                            | A30 G14 C23 T21                     | 42136.047                                             | A48 G21 C34 T34                      |
| Murine Hepatitis Virus            | MHV-A59              | Sense     | 27344.503                                            | A21 G23 C14 T30                     | 42599.989                                             | A34 G35 C18 T50                      |
|                                   |                      | Antisense | 27083.572                                            | A30 G14 C23 T21                     | 42064.089                                             | A50 G18 C35 T34                      |
| Rat                               | 8190                 | Sense     | 27344.491                                            | A21 G23 C14 T30                     | 42544.967                                             | A34 G34 C20 T49                      |
|                                   |                      | Antisense | 27083.567                                            | A30 G14 C23 T21                     | 42120.041                                             | A49 G20 C34 T34                      |
| SARS                              | TOR2                 | Sense     | 27298.518                                            | A27 G19 C14 T28                     | 42519.906                                             | A34 G33 C20 T50                      |
|                                   |                      | Antisense | 27125.542                                            | A28 G14 C19 T27                     | 42144.026                                             | A50 G20 C33 T34                      |
| SARS                              | Urbani               | Sense     | 27298.518                                            | A27 G19 C14 T28                     | 42519.906                                             | A34 G33 C20 T50                      |
|                                   |                      | Antisense | 27125.542                                            | A28 G14 C19 T27                     | 42144.026                                             | A50 G20 C33 T34                      |
| Avian Infectious Bronchitis Virus | Egg-adapted          | Sense     | 27396.544                                            | A24 G24 C14 T26                     | 42530.984                                             | A33 G32 C17 T55                      |
|                                   |                      | Antisense | 27032.524                                            | A26 G14 C24 T24                     | 42129.100                                             | A55 G17 C32 T33                      |

For both primer regions, the measured signals agreed with compositions expected from the known coronavirus sequences in GenBank. Several of the isolates used in this study did not have a genome sequence record in GenBank. Nevertheless, bioagent identifying amplicons were obtained for all test viruses and their base compositions were experimentally determined. These experimentally determined base compositions were confirmed by sequencing. Thus the strategy described here enables identification of viruses without the need for prior knowledge of sequence.

To demonstrate the potential to detect multiple viruses in the same sample, as might occur during a co-infection, viral extracts from three human coronaviruses, HCoV-229E, HCoV-OC43, and SARS CoV, were pooled and the mixture was analyzed by the methods of the present invention. Signals from all three viruses were clearly detected and resolved in the mass spectrum (Figure 9), demonstrating that co-infections of more than one coronavirus species could be identified. Dynamic range for reliable multispecies detections in this system is  $10^4$  (100:1 in each

direction, data not shown). This example indicates that the etiology of a coronavirus infection can be resolved using the method of the present invention. Furthermore, the method of identification of coronaviruses can be combined with general methods of broad range and drill-down identification of bacteria to resolve a more complex etiology comprising viral and bacterial co-infections.

Shown in Figure 10 is a spatial representation of the base compositions of five different coronaviruses for RdRp and nsp11 bioagent identifying amplicons. The G content of each base composition is represented by the tilt angle of the cone which indicates experimentally determined base composition. Calculated base compositions are indicated by the labeled spheres. Figure 10 indicates that the experimentally determined base compositions generally match the calculated base compositions. One exception is the canine coronavirus analysis which indicated that a single T to C substitution (single nucleotide polymorphism) exists in the amplicon.

Characterization of a bioagent identifying amplicon does not require prior knowledge of sequence. This feature is exemplified for the bioagent identifying amplicons obtained with the nsp11 primer set. No sequence was available in the nsp11 region for three of the five viral species (FIPV, CcoV and HcoV OC43). Nevertheless, base compositions of the three bioagent identifying amplicons were determined which were well within the expected bounds of base compositions of coronavirus nsp11 bioagent identifying amplicons. Thus, had the identity of these three coronaviruses been unknown and if they had been tested with the same primer sets, they would have been identified as newly discovered coronaviruses.

#### **Example 6: Quantitation of SARS Coronavirus by Internally Calibrated PCR**

SARS coronavirus was handled in a P3 facility by investigators wearing forced air respirators. Equipment and supplies were decontaminated with 10% hypochlorite bleach solution for a minimum of 30 minutes or by immersion in 10% formalin for a minimum of 12 h and virus was handled in strict accordance with specific Scripps Research Institute policy. SARS CoV was cultured on sub confluent Vero-E6 cells at 37°C, 5% CO<sub>2</sub> in complete DMEM with final concentrations of 10% fetal bovine serum (Hyclone), 292 µg/mL L-Glutamine, 100 U/mL penicillin G sodium, 100 µg/mL streptomycin sulfate (Invitrogen), and 10 mM HEPES (Invitrogen). Virus-containing medium was collected during the peak of viral cytopathic effects, 48 h after inoculation with approximately 10 PFU/cell of SARS CoV from the second passage of stock virus. Infectious virus was titered by plaque assay. Monolayers of Vero-E6 cells were

prepared at 70-80% confluence in tissue culture plates. Serial tenfold dilutions of virus were prepared in complete DMEM. Medium was aspirated from cells, replaced by 200  $\mu$ L of inoculum, and cells were incubated at 37°C, 5% CO<sub>2</sub> for 1 hour. Cells were overlaid with 2-3 mL/well of 0.7% agarose, 1x DMEM overlay containing 2% fetal bovine serum. Agarose was allowed to solidify at room temperature then cells were incubated at 37°C, 5% CO<sub>2</sub> for 72 h. Plates were decontaminated by overnight formalin immersion, agarose plugs were removed, and cells were stained with 0.1% crystal violet to highlight viral plaques.

In order to demonstrate the detection of SARS coronavirus in a clinically relevant fluid, varying dilutions of the titrated SARS virus were added to human serum. Serial tenfold dilutions (10<sup>0</sup> to 10<sup>-15</sup>) of the SARS virus were prepared in complete DMEM. 50  $\mu$ L of each dilution of the virus was added to 200  $\mu$ L human serum, mixed well and treated with 0.75 mL of Trizol Reagent LS (Invitrogen, Carlsbad, CA) at room temperature for 10 minutes. Contents were then transferred to a clean tube, which was sterilized on the outside with 10% bleach and moved to a P2 facility. RNA was extracted following protocols described above. 100% of the isolated RNA was reverse transcribed. 1/40th of the RT reaction was used per PCR reaction. RT-PCR was carried out as described in Example 2.

To determine the relationship between PFU and copies of nucleic acid target, the SARS coronavirus stock solution was analyzed using internally calibrated PCR. Synthetic DNA templates with nucleic acid sequence identical in all respects to each PCR target region from SARS CoV with the exception of 5 base deletions internal to each amplicon were cloned into a pCR-Blunt vector (Invitrogen, Carlsbad, CA). The calibrant plasmid was quantitated using OD<sub>260</sub> measurements, serially diluted (10-fold dilutions), and mixed with a fixed amount of post-reverse transcriptase cDNA preparation of the virus stock and analyzed by competitive PCR and electrospray mass spectrometry. Each PCR reaction produced two sets of amplicons, one corresponding to the calibrant DNA and the other to the SARS cDNA. Since the primer targets on the synthetic DNA calibrant and the viral cDNA were almost identical, it was assumed that similar PCR efficiencies exist for amplification of the two products. Analysis of the ratios of peak heights of the resultant mass spectra of the synthetic DNA and viral cDNA for each dilution of the calibrant were used to determine the amounts of nucleic acid copies (as measured by calibrant molecules) present per PFU, post reverse transcriptase. A PFU (plate forming unit) is defined as a quantitative measure of the number of infectious virus particles in a given sample, since each infectious virus particle can give rise to a single clear plaque on infection of a

continuous "lawn" of bacteria or a continuous sheet of cultured cells. Since all of the extracted RNA was used in the reverse transcriptase step to produce the viral cDNA, the approximate amount of nucleic acids associated with infectious virus particles in the original viral preparation was estimated.

5 To determine the relationship between PFU and copies of nucleic acid, the virus stock was analyzed using internally calibrated PCR. Synthetic DNA templates with nucleic acid sequence identical in all respects to each PCR target region from SARS CoV with the exception of 5 base deletions internal to each amplicon were cloned into a pCR-Blunt vector (Invitrogen, Carlsbad, CA). The calibrant plasmid was quantitated using OD260 measurements, serially  
10 diluted (10-fold dilutions), and mixed with a fixed amount of post-reverse transcriptase cDNA preparation of the virus stock and analyzed by competitive PCR and electrospray mass spectrometry. Each PCR reaction produced two sets of amplicons, one corresponding to the calibrant DNA and the other to the SARS cDNA. Since the amplicons generated from the synthetic DNA calibrant (calibrant amplicon) and the viral cDNA (bioagent identifying  
15 amplicon) were almost identical, it was assumed that PCR efficiencies for amplification of the two products were similar. Analysis of the ratios of peak heights of the resultant mass spectra of the synthetic DNA and viral cDNA for each dilution of the calibrant were used to determine the amounts of nucleic acid copies (as measured by calibrant molecules) present per PFU, post reverse transcriptase. Since all of the extracted RNA was used in the reverse transcriptase step to  
20 produce the viral cDNA, the approximate amount of nucleic acids associated with infectious virus particles in the original viral preparation could be estimated. Mass spectrometry analysis showed an approximate 1:1 peak abundance between the calibrant peak at the  $3 \times 10^4$  copy number dilution and the bioagent identifying amplicon peak for the RdRp primer set (Figure 11). Thus, the relationship between PFU and copies of nucleic acid was calculated to be 1 PFU = 300  
25 copies of nucleic acid.

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 of the patents, applications, printed publications, and other published documents mentioned or referred to in this  
30 specification are incorporated herein by reference in their entirety. Those skilled in the art will appreciate that numerous changes and modifications may be made to the embodiments of the invention and that such changes and modifications may be made without departing from the

spirit of the invention. It is therefore intended that the appended claims cover all such equivalent variations as fall within the true spirit and scope of the invention.

- 46 -

**WHAT IS CLAIMED IS:**

1. A method of identification of one or more unknown coronaviruses in a sample comprising:
  - a) obtaining coronavirus RNA from said sample;
  - 5 b) obtaining corresponding DNA from said RNA;
  - c) amplifying said DNA using one or more pairs of oligonucleotide primers that bind to conserved regions of a coronavirus genome, wherein said conserved regions flank a variable region of said coronavirus genome;
  - d) determining the molecular mass or base composition of the one or more amplification  
10 products obtained in step c); and
  - e) comparing said molecular mass or base composition with calculated or experimentally determined molecular masses or base compositions, wherein one or more matches identifies said unknown coronavirus.
- 15 2. The method of claim 1 wherein said conserved regions are about 80 -100% identical among different coronaviruses.
3. The method of claim 1 wherein said variable region exhibits no greater than about 5% identity among different coronaviruses.  
20
4. The method of claim 1 wherein said conserved and variable regions of a coronavirus genome are within genes encoding RNA-dependent RNA polymerase (RdRp) or nsp11.
5. The method of claim 1 wherein said coronavirus is Canine coronavirus, Feline  
25 coronavirus, Human coronavirus 229E, Human coronavirus OC43, Porcine epidemic diarrhea virus, Transmissible gastroenteritis virus, Bovine coronavirus, Murine hepatitis virus, Porcine hemagglutinating encephalomyelitis virus, Rat coronavirus, Turkey coronavirus, Avian infectious bronchitis virus, or SARS coronavirus.
- 30 6. The method of claim 1 wherein alignments of viral genome sequences are used to identify said conserved regions.

- 47 -

7. The method of claim 1 wherein said corresponding DNA is obtained from said RNA in a reverse transcriptase-catalyzed reaction.

8. The method of claim 1 wherein said amplification product is obtained using the  
5 polymerase chain reaction.

9. The method of claim 1 wherein said base composition is determined by measuring the molecular mass of both strands of said amplification product.

10 10. The method of claim 9 wherein determination of said base composition further comprises the process of strand matching.

11. A method of diagnosis of the etiology of a coronavirus infection comprising performing the method of claim 1 wherein said sample is a biological sample obtained from a  
15 human exhibiting symptoms of a coronavirus infection and the resulting identification of a specific coronavirus defines the etiology of said coronavirus infection.

12. The method of claim 11 wherein said etiology comprises infection by a plurality of coronavirus species.

20

13. The method of claim 12 wherein said plurality of coronavirus species comprises the SARS coronavirus.

14. The method of claim 12 wherein said etiology further comprises infection by a bacterial  
25 species, wherein said bacterial species is identified by determining the molecular mass of a bacterial bioagent identifying amplicon.

15. A method of tracking the spread of a specific coronavirus comprising:  
obtaining a plurality of samples containing said specific coronavirus from a plurality of  
30 different locations; and  
identifying said specific coronavirus in a subset of said plurality of samples using the

- 48 -

method of claim 1, wherein the corresponding locations of the members of said subset indicate the spread of said specific coronavirus to said corresponding locations.

16. The method of claim 15 wherein said coronavirus is Canine coronavirus, Feline coronavirus, Human coronavirus 229E, Human coronavirus OC43, Porcine epidemic diarrhea virus, Transmissible gastroenteritis virus, Bovine coronavirus, Murine hepatitis virus, Porcine hemagglutinating encephalomyelitis virus, Rat coronavirus, Turkey coronavirus, Avian infectious bronchitis virus, or SARS coronavirus.

10 17. A primer pair wherein each member of the pair has at least 70% sequence identity with the sequence of the corresponding member of any one of the following intelligent primer pair sequences: SEQ ID NOs:5:6, 7:8, 9:8, 9:10, 11:8, 11:10 or 9:10.

18. The primer pair of claim 17 further comprising at least one nucleobase analog, wherein  
15 said nucleobase analog binds with increased affinity to its complementary nucleobase compared to the native nucleobase.

19. The primer pair of claim 18 wherein said nucleobase analog comprises 5-propynyluracil, 5-propynylcytosine, 2,6-diaminopurine, phenoxazine, or G-clamp.

20

20. A bioagent identifying amplicon for identification of a coronavirus comprising an isolated polynucleotide of about 45 to about 150 nucleobases in length produced by the process of amplification of nucleic acid from a coronavirus genome with a primer pair, wherein each primer of said pair is of a length of about 12 to about 35 nucleobases, wherein the bioagent  
25 identifying amplicon provides identifying information about the coronavirus.

21. The bioagent identifying amplicon of claim 20 wherein each member of said primer pair has at least 70% sequence identity with the sequence of the corresponding member of any one of the following intelligent primer pair sequences: SEQ ID NOs:5:6, 7:8, 9:8, 9:10, 11:8, 11:10 or  
30 9:10.

- 49 -

22. The method of claim 20 wherein the sequence of said coronavirus genome is represented by any one of SEQ ID NOs:12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78,  
5 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or 101.

23. A method for simultaneous determination of the identity and quantity of an unknown coronavirus in a sample comprising:

contacting said sample with a pair of primers and a known quantity of a calibration  
10 polynucleotide comprising a calibration sequence;

simultaneously amplifying nucleic acid from said unknown coronavirus with said pair of primers and amplifying nucleic acid from said calibration polynucleotide in said sample with said pair of primers to obtain a first amplification product comprising a bioagent identifying amplicon and a second amplification product comprising a calibration amplicon;

15 subjecting said sample to molecular mass analysis wherein the result of said mass analysis comprises molecular mass and abundance data for said bioagent identifying amplicon and said calibration amplicon; and

distinguishing said bioagent identifying amplicon from said calibration amplicon based on molecular mass, wherein the molecular mass of said bioagent identifying amplicon identifies  
20 said coronavirus and comparison of bioagent identifying amplicon abundance data and calibration amplicon abundance data indicates the quantity of coronavirus in said sample.

24. The method of claim 23 wherein said calibration sequence comprises the sequence of a chosen standard sequence of a bioagent identifying amplicon with the exception of a deletion of  
25 about 2-8 consecutive nucleotide residues of said standard sequence.

25. The method of claim 23 wherein said calibration sequence comprises the sequence of a chosen standard sequence of a bioagent identifying amplicon with the exception of an insertion of about 2-8 consecutive nucleotide residues of said standard sequence.

- 50 -

26. The method of claim 23 wherein said calibration sequence has at least 80% sequence identity with a chosen standard sequence of a bioagent identifying amplicon.

27. The method of claim 23 wherein said calibration sequence has at least 90% sequence  
5 identity with a chosen standard sequence of a bioagent identifying amplicon.

28. The method of claim 23 wherein said calibration sequence has at least 95% sequence identity with a chosen standard sequence of a bioagent identifying amplicon.

10 29. The method of claim 24 or claim 25 wherein said standard sequence is a segment of a gene sequence encoding RNA-dependent RNA polymerase (RdRp) or nsp11.

30. The method of claim 23 wherein said calibration polynucleotide resides on a plasmid.

15 31. The method of claim 23 wherein said molecular mass analysis comprises FTICR or MALDI-TOF mass spectrometry.

32. An isolated polynucleotide for determining the quantity of a bioagent in a sample comprising SEQ ID NO:102.

20

33. An isolated polynucleotide for determining the quantity of a bioagent in a sample comprising SEQ ID NO:103.

34. A plasmid vector comprising the calibration polynucleotide of SEQ ID NO:102 or SEQ  
25 ID NO:103.

35. A plasmid vector comprising the calibration polynucleotide of SEQ ID NO:104.

36. A kit comprising one or more pairs of primers wherein each member of each pair has at  
30 least 70% sequence identity with the sequence of the corresponding member of any one of the following intelligent primer pair sequences: SEQ ID NOs:5:6, 7:8, 9:8, 9:10, 11:8, 11:10 or 9:10.

- 51 -

37. The kit of claim 36 further comprising a reverse transcriptase and nucleotide triphosphates.

38. The kit of claim 36 further comprising a DNA polymerase and nucleotide triphosphates.