JOURNAL OF CLINICAL MICROBIOLOGY, July 2004, p. 3291–3294 Vol. 42, No. 7
0095-1137/04/$08.00+0 DOI: 10.1128/JCM.42.7.3291–3294.2004

Identification of the Severe Acute Respiratory Syndrome Coronavirus by Simultaneous Multigene DNA Sequencing

T. Vinayagamoorthy,* Kirk Mulatz, and Roger Hodkinson

Bio-ID Diagnostic Inc., Saskatoon, Saskatchewan, Canada S7N 4N1

Received 9 January 2004/Returned for modification 24 March 2004/Accepted 2 April 2004

The recent severe acute respiratory syndrome (SARS) outbreak resulted in calls for an accurate diagnostic test that can be used not only for routine testing but also for generating nucleotide sequences to monitor the epidemic. Although the identity of the SARS coronavirus (SARS-CoV) genome was confirmed by DNA sequencing, it is impractical to sequence the entire 29-kb SARS-CoV genome on a routine basis. Therefore, alternative assay methods such as the enzyme-linked immunosorbent assay and PCR have been pursued for routine testing, primarily to resolve probable cases. We report here a modification of standard DNA sequencing technology for accurate identification of SARS-CoV in routine testing. Instead of requiring the sequencing of the whole SARS-CoV genome, our modification enables the simultaneous sequencing of three regions of the SARS-CoV genome, the spike protein-encoding gene (35 nucleotides), gene M (43 nucleotides), and gene N (45 nucleotides), in a single electropherogram. Comparing these nucleotide sequences to DNA databank entries (National Institutes of Health) conclusively identified them as SARS-CoV sequences.

Although the entire genomic sequence of severe acute respiratory syndrome coronavirus (SARS-CoV) is known, sequencing the entire 29-kb CoV genome is impractical for routine testing. Accurate identification of SARS-CoV does not, however, require the determination of the nucleotide sequence of the entire viral genome; it is only necessary to identify those short nucleotide sequences that are unique to SARS-CoV (signature sequences). In this report, we present a new assay to simultaneously generate and analyze those signature sequences characteristic of SARS-CoV.

This new assay is a modification of the standard chain termination DNA sequencing technology (4) and principally involves the capability to simultaneously generate short nucleotide sequences (5) from more than one region of the SARS-CoV viral genome. In this report we describe a multiplex sequencing method that involves the simultaneous amplification of three regions of the SARS-CoV genome, the spike protein-encoding gene, gene M, and gene N (6), by reverse transcriptase-multiple PCR (RT-PCR). This is followed by the use of modified sequencing primers for simultaneously sequencing of the 3′ end of the all three amplicons (Fig. 1). This modification is such that the longest truncated molecule generated from the gene encoding the spike protein has a slightly lower molecular weight than the shortest one generated from gene M and the longest truncated molecule generated from gene M has a slightly lower molecular weight than the shortest one generated from the gene N.

Plasmid clones. This assay was performed using cDNA plasmid clones. Three purified plasmid DNA clones (gene M, gene N, and the spike protein-encoding gene) were obtained from the National Microbiology Laboratory, Winnipeg, Canada.

Transformation. Top10 *Escherichia coli* cells (Invitrogen, Carlsbad, Calif.) were grown overnight at 37°C in 5 ml of Luria-Bertani broth. A total of 250 μl of the overnight culture was transferred to 5 ml of fresh Luria-Bertani (LB) broth and incubated for an additional 2 h at 37°C. The tubes were left on ice for 10 min. The culture (1.5 ml) was centrifuged in a 1.5-ml microcentrifuge tube at 500 × *g* for 10 min, and the supernatant was discarded. The cells were resuspended in 1.5 ml of ice-cold 100 mM $MgCl_2$ and left on ice for 20 min. The tubes were centrifuged at 500 × *g* for 12 min, and the supernatant was discarded. Ice-cold 100 mM $CaCl_2$ (250 μl) was used to resuspend the cells. Plasmid DNA (20 ng) was added to the resuspended cells, and the tube was left on ice for 2 min. This was followed by a 3-min heat shock at 42°C and 3 min on ice. LB broth (1,250 μl) was added, and the transformed cells were incubated at 37°C for 30 min. A total of 100 μl of the culture was plated on an LB agar plate with 50 μg of ampicillin/ml and incubated overnight at 37°C.

Preparation of plasmid DNA. Ten bacterial colonies from ampicillin-supplemented LB plates were transferred into 50 μl of extraction buffer 1–25% Chelex (Bio-ID Diagnostic Inc., Saskatoon, Saskatchewan, Canada) in an Eppendorf tube. To this suspension 2.5 μl of lysozyme solution (50 mg/ml) was added, and the resulting mixture was incubated at 37°C for 45 min. Following incubation, the Eppendorf tube was centrifuged at 5,000 × *g* for 2 min. The supernatant was transferred into another Eppendof tube. Extraction buffer 1 (50 μl)–25% Chelex and 1 μl of proteinase K (Sigma, St. Louis, Mo.) were added. Following 30 min of incubation at 56°C, the tube was centrifuged at 5,000 × *g* for 2 min. The supernatant was then transferred into a new tube.

PCR amplification. Initially, three specific individual DNA targets were amplified using three sets of PCR primers separately and the thermocycling protocol shown in Table 1. Amplification was performed in a 50-μl volume containing 5 μl of master buffer (Bio-ID Diagnostic Inc.), 5 U of *Taq* polymerase (Roche Diagnostic, Indianapolis, Ind.), target DNA, and nu-

* Corresponding author. Mailing address: Bio-ID Diagnostic Inc., No. 1, 410 Downey Rd., LFK Biotechnology Complex, Saskatoon, Saskatchewan, Canada S7N 4N1. Phone: 306-975-9161. Fax: 306-938-0751. E-mail: moorthy@bio-id-diagnostic.com.

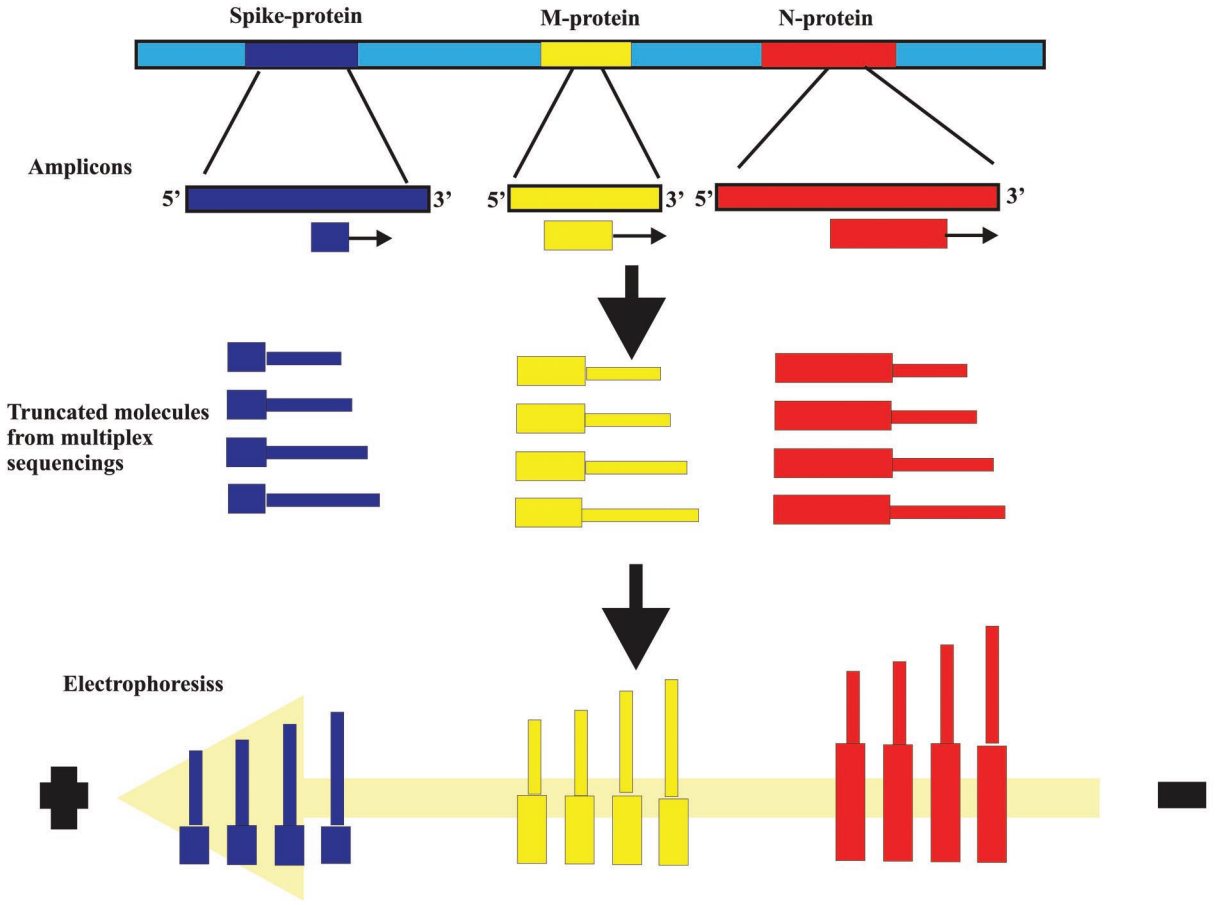

FIG. 1. Stages of the MultiGEN process and the basic scientific principles. These include simultaneous generation of three amplicons from the spike protein-encoding gene, gene M, and gene N and simultaneous sequencing of the 3′ end of each of the amplicons.

clease-free water (Gibco BRL, Gaithersburg, Md.). The amplified targets were purified using a PSI Clone HTS system (Princeton Separations Inc., Adelphia, N.J.).

Gel electrophoresis. The amplicons were visualized by loading 2 μl of 6× loading buffer, 5 μl of Tris-borate-EDTA buffer, and 5 μl of the PCR product onto a 2% agarose gel and applying 110 V for 50 min. The gel was then stained for 15 min in ethidium bromide solution (5 μg/ml in Tris-borate-EDTA buffer) and then destained in distilled water for 15 min.

Cycle sequencing. Amplicons were sequenced by cycle sequencing using an ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, Calif.), a GeneAmp 2400 thermocycler (PE Applied Biosystems), and the thermocycler profile shown in Table 1. Un-

TABLE 1. PCR and sequencing primers

Assay step and primer	Sequence
PCR[a]	
SpikeA	5′-GGGCTTCTGCTAATCTTG-3′
SpikeB	5′-CTGGGATGGCACATACGTGAC-3′
GeneMa	5′-ATTGGGTGACTGGCGGGATTG-3′
GeneMb	5′-TCACTTTCCATGAGCGGTCTG-3′
GeneNa	5′-CGAAGAGCTACCCGACGAGTT-3′
GeneNb	5′-ATGTGGTCTTTGGGTGTATTC-3′
Sequencing[b]	
Spikeseq	5′-Δ-GTCCTTCCCACAAGCAGCCCCGCAT-3′[c]
GeneMseq	5′-Δ-AACGAAGTAGCTAAGCCACATCAAG-3′
GeneNseq	5′-Δ-CTACGGCGCTAACAAAGAAGGCATC-3′

[a] PCR thermocycle profile: 1 cycle of 95°C for 3 min; 35 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 1 min; hold at 4°C.
[b] Sequencing thermocycle profile: 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min; hold at 4°C.
[c] Δ, 5′ modification to increase the primer molecular weight.

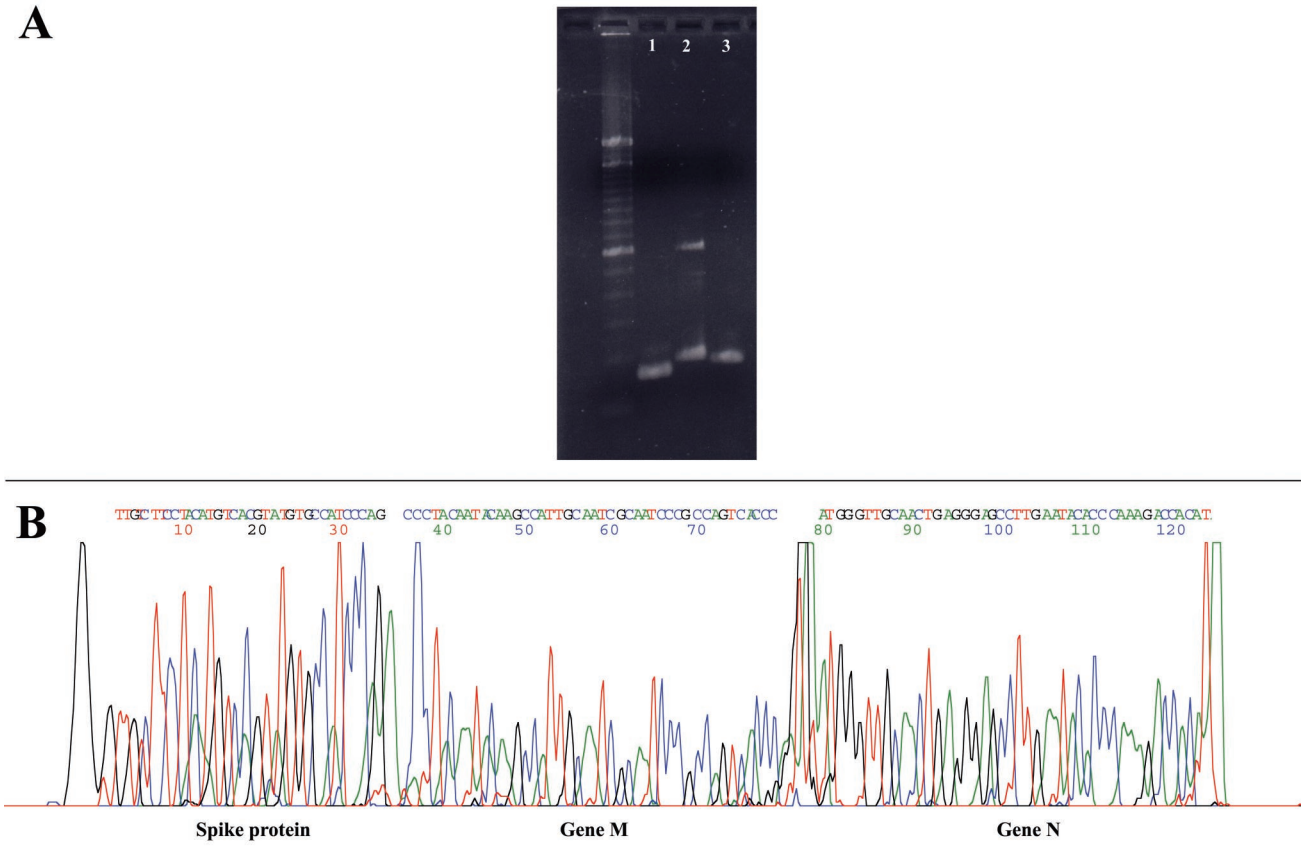

FIG. 2. (A) Agarose gel electrophoresis showing amplicons. Lane 1, spike gene amplicon (158 bp); lane 2, gene M amplicon (190 bp); lane 3, gene N amplicon (176 bp). (B) Electropherogram showing nucleotide sequences from three regions of the SARS-CoV genome: the spike protein-encoding gene (35 nucleotides), gene M (43 nucleotides), and gene N (45 nucleotides).

incorporated dye terminators were removed using Centricep chromatography columns (Princeton Separations Inc.). The samples were then dried and resuspended in 20 μl of ABI PRISM template suppression reagent. Samples were analyzed by capillary electrophoresis using an ABI PRISM 310 genetic analyzer. The 47-cm by 50-μm uncoated capillary was filled with performance-optimized polymer 6 (acrylamide-urea polymer) and heated to 50°C. A total of 20 μl of the sequencing mixture was pipetted into a 0.2-ml microcentrifuge tube provided by the manufacturer (Applied Biosystems). Samples were drawn into the capillary by electrokinetic injection at 2 kV for 50 to 200 s. The electrophoresis was carried out at 15 kV for 20 min.

Four separate PCRs were performed; three of them used 1 μl (~200 ng) of the purified plasmid DNAs from the spike protein, M protein, and N protein as templates. In the fourth reaction, 1-μl (~200-ng) amounts from the purified plasmid DNA from spike protein, M protein, and N protein were used together. Agarose gel electrophoresis showed three amplicons of 158, 190, and 176 bp (Fig. 2A). The molecular sizes of these amplicons are so close that data are not shown. The 3′ ends of the amplicons generated from the spike protein, M protein, and N protein were sequenced simultaneously using modified sequencing primers (Bio-ID Diagnostic Inc.) sarseq1, sarseq2, and sarseq3 (Table 1). The single electropherogram produced carried three stretches of DNA sequences: 35 bases of the spike protein-encoding gene followed by 43 bases of gene M and 45 bases of gene N (Fig. 2B). A BLAST search of these nucleotide sequences identified all three sequences as unique to SARS-CoV. Although we have not used clinical samples, the use of RT-PCR to generate amplicons from RNA extracts of clinical samples has been reported previously (2).

Our multiplex sequencing technology overcomes two major limitations encountered by conventional single-RT-PCR-based identification of SARS-CoV: false negatives and false positives. As SARS-CoV is an RNA virus, it is susceptible to genetic mutations that are characteristic of RT (3). If such events lead to mutational changes within the annealing region of the primers used for RT-PCR and sequencing, the target amplicons may not be generated, thereby creating a false negative result. In our assay we target three different regions so that even when mutations do occur within two annealing regions, the third target is still amplified, thereby minimizing the possibility of a false negative result. DNA amplification-based assays often create nonspecific amplification (1). The result format of our assay is an electropherogram with nucleotide sequences whose identity can be verified and confirmed by checking the nucleotide depository (GenBank; National Institutes of Health) via an Internet BLAST search, where the results can be obtained within minutes, hence eliminating false

positive results. Further, this multiplex sequencing method allows a choice of the genomic regions to be examined and the generation of nucleotide sequences that can address epidemiologic questions of genetic drift and shift.

In summary, given the global impact of the recent SARS-CoV epidemic, we believe that it would be of paramount importance in any future outbreaks to test for the disease with a nucleotide sequence-based method, both to ensure correct identification and to provide the detailed genetic information that is crucial for understanding the molecular epidemiology of an outbreak. In this paper we report an accurate multiplex nucleotide sequencing method that meets these challenges.

We thank the National Microbiology Laboratory, Winnipeg, Canada, for providing us with plasmid clones from coronavirus and Yves Plante Biotechnology Branch, Saskatchewan Research Council, for expert review of the manuscript.

REFERENCES

1. **Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White.** 1990. PCR protocols. A guide to methods and amplifications, p. 3. Academic Press, New York, N.Y.
2. **Lau, L. T., Y. W. Fung, F. P. Wong, S. S. Lin, C. R. Wang, H. L. Li, N. Dillon, R. A. Collins, J. S. Tam, P. K. Chan, C. G. Wang, and A. C. Yu.** 2003. A real-time PCR for SARS-coronavirus incorporating target gene pre-amplification. Biochem. Biophys. Res. Commun. **312:**1290–1296.
3. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, p. 5.52. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
4. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
5. **Vinayagamoorthy, T., K. Mulatz, and R. Hodkinson.** 2003. Nucleotide sequence-based multitarget identification. J. Clin. Microbiol. **41:**3284–3292.
6. **Wang, J., J. Wen, J. Li, J. Yin, Q. Zhu, H. Wang, Y. Yang, E. Qin, B. You, W. Li, X. Li, S. Huang, R. Yang, X. Zhang, L. Yang, T. Zhang, Y. Yin, X. Cui, X. Tang, L. Wang, B. He, L. Ma, T. Lei, C. Zeng, J. Fang, J. Yu, J. Wang, H. Yang, M. B. West, A. Bhatnagar, Y. Lu, N. Xu, and S. Liu.** 2003. Assessment of immunoreactive synthetic peptides from the structural proteins of severe acute respiratory syndrome coronavirus. Clin. Chem. **49:**1989–1996.