

Genomic Sequencing of the Severe Acute Respiratory Syndrome-Coronavirus

Stephen S. C. Chim, Rossa W. K. Chiu, and Y. M. Dennis Lo

Summary

The polymerase chain reaction (PCR), which can exponentially replicate a target DNA sequence, has formed the basis for the sensitive and direct examination of clinical samples for evidence of infection. During the epidemic of severe acute respiratory syndrome (SARS) in 2003, PCR not only offered a rapid way to diagnose SARS-coronavirus (SARS-CoV) infection, but also made the molecular analysis of its genomic sequence possible. Sequence variations were observed in the SAR-CoV obtained from different patients in this epidemic. These unique viral genetic signatures can be applied as a powerful molecular tool in tracing the route of transmission and in studying the genome evolution of SARS-CoV. To extract this wealth of information from the limited primary clinical specimens of SARS patients, we were presented with the challenge of efficiently amplifying fragments of the SARS-CoV genome for analysis. In this chapter, we will discuss how we managed to accomplish this task with our optimized protocols on reverse-transcription, nested PCR amplification, and DNA cycle sequencing. We will also discuss the sequence variations that typified some strains of SARS-CoV in the different phases during this epidemic. PCR amplification of the viral sequence and genomic sequencing of these critical sequence variations of re-emerging SARS-CoV strains would give us quick insights into the virus.

Key Words: SARS coronavirus; viral RNA extraction; reverse-transcription PCR; sequencing; genomic sequence variation.

1. Introduction

Severe acute respiratory syndrome-coronavirus (SARS-CoV), the etiologic agent of SARS (*I-3*), is a virus that was unknown to us before the SARS epidemic. The concerted efforts of researchers have promptly elucidated its genetic code. The genome of SARS-CoV is a 29,727-nucleotide, polyadenylated RNA. The genomic organization is typical of coronaviruses, having the characteristic gene order (5'-*polymerase* [*Orflab*], *spike* [*S*], *envelope* [*E*], *mem-*

From: *Methods in Molecular Biology*, vol. 336: *Clinical Applications of PCR*
Edited by: Y. M. D. Lo, R. W. K. Chiu, and K. C. A. Chan © Humana Press Inc., Totowa, NJ

brane [M], and *nucleocapsid* [N]-3') and short untranslated regions at both termini (4,5).

With this sequence information, rapid PCR-based molecular diagnostic tests of SARS-CoV infection were designed (1,6–10). Besides offering molecular diagnosis and quantitative measurement of viral load, PCR-based technologies have also been exploited to amplify the genomic fragments of SARS-CoV for sequence analysis. The high sensitivity and specificity of PCR has made this genomic sequence analysis possible even for uncultured clinical specimens. Unlike the conventional microbiological methods, PCR-based technologies may not require viral culture, which could introduce culture-derived artifacts in the genomic sequence. The specific PCR primers selectively amplify SARS-CoV sequences from the background of other nucleic acid sequences contributed by the patient or other microbes. Moreover, the PCR-based method is versatile in terms of the type of clinical specimens. In our hands, we have successfully analyzed the SARS-CoV genome directly from uncultured samples of serum, nasopharyngeal aspirate, and stools (11). This obviates any concern about the poor or even unsuccessful viral culture of the precious clinical specimens. The risk in handling large-volume and hazardous viral culture could also be avoided.

Genomic sequence variations were observed in the SARS-CoV obtained from different patients in this epidemic. Based on these sequence variations, most of the isolates are typified by two groups: isolates obtained from patients who were epidemiologically linked to the Metropole Hotel in Hong Kong, and those who were not (3,12,13). For example, there are seven sequence variations that can distinguish isolate CUHK-Su10, which is linked to the Metropole Hotel, from isolate CUHK-W1, which is not linked to this hotel case cluster (Table 1). Among them, four variations at nucleotide positions 17564, 21721, 22222, and 27827 (according to the Tor2 sequence in GenBank, accession no. AY274119 [5]) were suggested by The Chinese SARS molecular epidemiology consortium (14) as part of a haplotype configuration that marks the different phases of a tri-phasic SARS epidemic in Guangdong Province of China. CUHK-W1 carried a haplotype G:A:C:C that typified the middle phase. Notably, the same haplotype was observed in CUHK-L2, which was one of the earliest confirmed case of SARS in Hong Kong, having been documented even before any report of the hotel case cluster (15). CUHK-Su10 carried a haplotype T:T:T:T that typified the late phase, marked by the hotel case cluster that spread the virus to many other parts of the world.

Genomic sequence variations in SARS-CoV have also revealed the route of infection from within communities and across cities. For instance, compared with isolate CUHK-Su10, two mutations, T3852C and C11493T, first appeared

Table 1
Comparison of the Sequences of Two Strains of Severe Acute Respiratory Syndrome (SARS)-Coronavirus Isolated From Patients in Hong Kong at the Beginning of the Epidemic^a

Nucleotide position	CUHK-Su10	CUHK-W1
9404	T	C
9479	T	C
17564 ^b	T	G
19064	A	G
21721 ^b	G	A
22222 ^b	T	C
27827 ^b	T	C

^aSequence variations at seven positions between the two viral strains (CUHK-Su10 and CUHK-W1) are indicated. The nucleotide positions are numbered according to the sequence of GenBank accession number AY274119.

^bPart of the haplotype suggested by The Chinese SARS molecular epidemiology consortium for distinguishing the early, middle, and late phase of the SARS epidemic in 2003.

in isolates CUHK-AG01, CUHK-AG02, CUHK-AG03 (GenBank accession numbers AY345986, AY345987, AY345988) obtained from patients involved in the Amoy Gardens outbreak in Hong Kong (**11**). Later, these two genetic fingerprints appeared in 10 completely sequenced Taiwanese isolates (**16**).

Interestingly, toward the end of the epidemic, another type of fingerprint was found by PCR-based method. A variant of the SARS-CoV with a 386-nucleotide deletion was reported in a cluster of patients that seem to be epidemiologically related (**17**). Most of the cases were part of a documented outbreak in the North District Hospital in Hong Kong.

We have illustrated that sequence variations among different isolates have a remarkable epidemiological correlation. Thus, PCR amplification followed by sequencing is a powerful tool in tracing the route of transmission. The sequence information may provide objective support to epidemiological investigations. Moreover, in the event that SARS re-emerges, one could quickly gain important insight into the origin and evolution status of the SARS-CoV simply by sequencing the critical sequence variations of the genome, as exemplified here. However, to extract this wealth of information from the limited primary clinical SARS specimens, we need very sensitive and efficient protocols to efficiently amplify fragments of the SARS-CoV genome for analysis.

2. Materials

2.1. RNA Extraction

1. QIAamp viral RNA Mini Kit (Qiagen, Hilden, Germany).
2. Absolute ethanol.

2.2. Reverse-Transcription

1. Superscript III RNase H⁻ reverse transcriptase (Invitrogen, Carlsbad, CA).
2. Random hexamers (Applied Biosystems, Foster City, CA).
3. 5X First-strand synthesis buffer: 250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM MgCl₂ (Invitrogen).
4. RNasin RNase inhibitor (Promega, Madison, WI).
5. dNTP (Invitrogen).
6. 0.1 M Dithiothreitol (DTT).
7. RNase-free water (Promega).

2.3. PCR Amplification

1. Advantage cDNA Polymerase mix and buffer (BD Biosciences Clontech, Palo Alto, CA).
2. dNTP.
3. PCR primers.
4. Distilled water.

2.4. Genomic Sequencing

1. BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems).
2. ABI Prism 3100 Genetic Analyzer (Applied Biosystems).
3. DYEnamic ET Dye Terminator Kit (GE Healthcare-Biosciences, Little Chalfont, UK).
4. MegaBACE 1000 Sequencing System (GE Healthcare-Biosciences).

2.5. Sequence Analysis and Comparison

SeqScape software (Applied Biosystems).

3. Methods

3.1. Precautions Against Potential Contamination

Genomic sequencing involves PCR amplification, which produces numerous copies of the target DNA, and cycle sequencing, which requires the pipetting and manipulation of PCR products. These steps could easily contaminate the laboratory environment with amplified products. Such contamination problems would affect the interpretation of sequencing results, and adversely affect the performance of diagnostic tests designed to detect the same viral sequences. Hence, extreme care should be taken to avoid contamination. We suggest the following precautions:

1. Perform RNA extraction, PCR amplification, and genome sequencing in different laboratories, or at least in separate and dedicated compartments of the same laboratory.
2. Transfer reagents and samples only with aerosol-resistant pipet tips.
3. Prepare the PCR reagent master mix in a hood dedicated for this purpose. A set of clean gloves and dedicated lab gown should be worn in this area. Illuminate the hood with ultraviolet before and after use.
4. Any steps that involve the handling of cDNA, primary and secondary PCR products (including addition of DNA templates in assembling the PCR), electrophoresis, and cycle sequencing should be performed in a dedicated area far away from any PCR reagents. A separate lab gown and set of gloves should be worn in this area.
5. Discard all pipet tips that contacted DNA with extreme care. Use a double bag for disposal.
6. Include multiple negative PCR controls in each amplification to monitor for environmental contamination.

3.2. RNA Extraction

1. Prepare AVL lysis buffer and AW1 and AW2 wash buffers according to manufacturer's (Qiagen) instructions (*see Note 1*).
2. In a biosafety level 2 (or above) containment laboratory, lyse 0.28 mL (1 vol) of viral culture by adding 1.12 mL (4 vol) of AVL buffer, mixing and incubating at room temperature for 10 min. Direct clinical samples, e.g., serum, nasopharyngeal aspirates, and stools, can also be used (*see Note 2*).
3. Add 1.12 mL of absolute ethanol to the mixture. Pulse-vortex for 15 s.
4. Load the mixture to QIAamp spin column and wash the column according to the manufacturer's instructions.
5. Add 60 μ L of RNase-free water onto the membrane and incubate for 1 min at room temperature. Centrifuge the spin column for 1 min at 6000g.
6. Quantify a small aliquot of the extracted viral RNA yield by real-time quantitative reverse-transcription (RT)-PCR (9) (*see Note 3*).
7. Store the extracted RNA at -80°C .

3.3. Reverse-Transcription

1. Prewarm two thermocycler blocks with heated lid at 72 and 25 $^{\circ}\text{C}$, respectively.
2. Mix 1 μ L (50 pmol) random hexamer with 10 μ L RNA in a 0.5-mL tube. Denature at 72 $^{\circ}\text{C}$ for 10 min (*see Note 4*).
3. During this period, assemble the reaction mix in another tube on ice according to **Table 2** using SuperScript III RNase H⁻ Reverse Transcriptase (*see Note 5*).
4. After denaturation, snap-cool the RNA-primer mixture on ice for 1 min. Briefly spin the tubes. Add the reaction mix prepared in **step 2** to the RNA-primer mixture to make up a total reaction volume of 20 μ L. Mix by pipetting gently up and down.

Table 2
Composition of Reaction Mix for Reverse-Transcription
of Severe Acute Respiratory Syndrome-Coronavirus RNA

Component	Volume for one reaction (μL)	Final concentration
5X first strand buffer	4	1X
dNTP mix (10 mM each dATP, dGTP, dCTP, and dTTP at neutral pH)	1	0.5 mM each
0.1 M dithiothreitol	5 mM	
RNasin RNase inhibitor (40 U/ μL)	1	2 U/ μL
SuperScript III reverse transcriptase (200 U/ μL)	2	20 U/ μL
Total volume	9	

Table 3
Composition of Reaction Mix for Polymerase Chain Reaction Amplification

Component	Volume for one reaction (μL)	Final concentration
Distilled water	19.5	
Advantage PCR buffer (10X)	2.5	1X
dNTP mix (10 mM each dATP, dGTP, dCTP, and dTTP at neutral pH)	0.5	200 μM each
cDNA polymerase mix (50X)	0.5	1X
Total volume	23.0	

5. Immediately transfer the tube from ice to the prewarmed 25°C thermocycler block for a 5-min incubation. Prewarm the other thermocycler block at 55°C.
6. Transfer the tube to the prewarmed 55°C thermocycler block for a 1-h incubation.
7. Heat inactivate at 72°C for 15 min.
8. Add 1 μL (2 U) of RNase H and incubate at 37°C for 20 min to remove RNA complementary to the cDNA.
9. Dilute the product two- to fivefold with distilled water. Store at -20°C before use.

3.4. Primary PCR Amplification

1. Inside a hood dedicated for setting up PCR, assemble the PCR master mix for the 50 reactions according to **Table 3** with cDNA polymerase mix (*see Note 6*) in a

final reaction volume of 25 μL . Add 50 aliquots of 23 μL into a 96-well PCR microplate.

2. Add 5 pmol each of forward (PCR-F) and reverse (PCR-R) series of primers for each of the 50 reactions amplifying the overlapping amplicons that cover the whole SARS-CoV genome (*see Note 7*). The primer sequences are shown in **Table 4**.
3. In an area separate from the hood dedicated for PCR, add 1 μL of diluted reverse-transcribed products.
4. Commence with PCR in a thermocycler with initial denaturation at 95°C for 1 min and 35 cycles of 95°C for 0.5 min, 55°C for 0.5 min, 68°C for 1.5 min, and a final extension at 68°C for 10 min.

3.5. Secondary PCR Amplification

1. Inside a hood dedicated for setting up PCR, assemble the PCR master mix for the 50 reactions according to **Table 3** in a final reaction volume of 25 μL . Add 50 aliquots of 23 μL into a new 96-well PCR microplate.
2. Add 5 pmol each of forward (PCR-F) and reverse (BSEQ-R) series of primers for each of the 50 semi-nested PCR reactions. The primer sequences are shown in **Table 4**.
3. In an area separate from the hood dedicated for PCR, add 1 μL of the corresponding primary PCR product.
4. Commence PCR in a thermocycler with initial denaturation at 95°C for 1 min and 35 cycles of 95°C for 0.5 min, 55°C for 0.5 min, 68°C for 1.5 min, and a final extension at 68°C for 10 min.
5. Electrophorese 5 μL of the secondary PCR product in a 2% agarose gel to verify the success of the PCR amplification. Estimate the amount of PCR product by comparison to DNA marker. Only products with single band should be used for sequencing.

3.6. Cycle Sequencing

Perform sequencing reaction based on the dideoxy dye terminator method, according to manufacturers' instructions:

1. Separate from the hood dedicated for PCR, assemble the cycle sequencing reaction with ASEQ-F, BSEQ-F, ASEQ-R, and BSEQ-R series of oligonucleotides as sequencing primers for each of the amplicon, and with 2–5 ng of secondary PCR product as sequencing template (*see Note 8*).
2. Commence with cycle sequencing reaction in a thermocycler.
3. Purify the extension products with either spin column purification or ethanol precipitation. Mix or resuspend the DNA in formamide solution according to the manufacturer's instructions.
4. Denature the purified extension products at 95°C for 5 min, snap-cool on ice, and load onto the automated capillary DNA sequencer for injection.

Table 4
Primer Sequences

01PCR-F	CTACCCAGGAAAAGCCAACCAACCT
01ASEQ-F	AAAGCCAACCAACCTCGATC
01ASEQ-R	AAGTGCCATTTTTGAGGTGT
01BSEQ-F	TTGCCTGTCCTTCAGGTTAG
01BSEQ-R	GTCACCTAAGTCATAAGACT
01PCR-R	TGCCAAGCTCGTCACCTAAGTCATA
02PCR-F	TACCGCAATGTTCTTCTTCGTAAGA
02ASEQ-F	TTCTTCTTCGTAAGAACGGT
02ASEQ-R	GCTCGTAGCTCTTATCAGAG
02BSEQ-F	CAACTTGATTACATCGAGTC
02BSEQ-R	TTCAGTGCCACAATGTTTAC
02PCR-R	TAACTAAATTTTCAGTGCCACAATG
03PCR-F	TCTACCTTGATGGGGTGTAATCATT
03ASEQ-F	TGAAATCTAATCATTGCGAT
03ASEQ-R	GGAGATCCTCATTCAAGGTC
03BSEQ-F	AATAAGCGTGCCTACTGGGT
03BSEQ-R	AATTGATCTGATAACACCAG
03PCR-R	TGCGCGCAAAAATTGATCTGATAAC
04PCR-F	AAAGGTGCTTGGAACATTGGACAAC
04ASEQ-F	GGAACATTGTACAACAGAGA
04ASEQ-R	ATTTGAGAATCTCCCAAGCA
04BSEQ-F	GGCACTACTGTTGAAAACT
04BSEQ-R	ATGTGAATCACCTTCAAGAA
04PCR-R	GTAATGTGTCATGTGAATCACCTTC
05PCR-F	CGTCAGTGTATACGTGGCAAGGAGC
05ASEQ-F	TACGTGGCAAGGAGCAGCTG
05ASEQ-R	CAACACGTTTCAAGCTCA
05BSEQ-F	GGTGCACCAATTAAGGTGT
05BSEQ-R	ACAGGTTTTCATCAATTTCTT
05PCR-R	ACTCATGTTTACAGGTTTTCATCAAT
06PCR-F	TCATCACGTATGTATTGTTTCTTTT
06ASEQ-F	TGTATTGTTTCTTTTACCCT
06ASEQ-R	CTGCTACACCACCACCATGT
06BSEQ-F	ATTAAATGTGTTGACATCGT
06BSEQ-R	AACCGTCTGCACGCACACTT
06PCR-R	CCTGTGTACGAACCGTCTGCACGCA
07PCR-F	TCACAGGACATCTTACTTGCACCAT
07ASEQ-F	TCTTACTTGCACCATTTGTTG
07ASEQ-R	CTTACCTCTAAGCATGTTT
07BSEQ-F	ACACTGGAAGAACTAAGTT

(continued)

Table 4 (Continued)
Primer Sequences

07BSEQ-R	TACAGTTCCTAGAATCTCTT
07PCR-R	AATTCCAGGATACAGTTCCTAGAAT
08PCR-F	GCTAAGACTGCTCTTAAGAAATGCA
08ASEQ-F	CTCTTAAGAAATGCAAATCT
08ASEQ-R	CTACGGCAGGAGCTTTAAGA
08BSEQ-F	AATGAGCCGCTTGTCACAAT
08BSEQ-R	CACTTTTATAGTCTTAACCT
08PCR-R	CAGTTGTGAACACTTTTATAGTCTT
09PCR-F	CCCGTCGAGTTTCATCTTGACGGTG
09ASEQ-F	TTCATCTTGACGGTGAGGTT
09ASEQ-R	AATTGTTATCAGCCCATTTA
09BSEQ-F	AGTTTTCTTGGTAGGTACAT
09BSEQ-R	ATACATCACAGCTTCTACAC
09PCR-R	GAGTACCCATATACATCACAGCTTC
10PCR-F	TTGGAATCTGCAAAGCGAGTTCTTA
10ASEQ-F	CAAAGCGAGTTCTTAATGTG
10ASEQ-R	TGTAAGATGTTTCCTTGTAG
10BSEQ-F	GCTAAGGAGACCCTCTATCG
10BSEQ-R	AATAGCCACTACATCGCCAT
10PCR-R	GTCTATAGTCAATAGCCACTACATC
11PCR-F	TTAAATCAAATGACAGGCTTCACAA
11ASEQ-F	TGACAGGCTTCACAAAGCCA
11ASEQ-R	TGCCTACAACCTTCGGTAGTT
11BSEQ-F	TGTGAAAGTCAACAACCCAC
11BSEQ-R	AGGCATATAATTGTTAAACA
11PCR-R	TAAACACATAAGGCATATAATTGTT
12PCR-F	TATGTCAAACCATTCTTAGGACAAG
12ASEQ-F	CATTCTTAGGACAAGCAGCA
12ASEQ-R	ACGAATTAAGATACAATTCT
12BSEQ-F	GGTTCTCTAATCTGTGTAAC
12BSEQ-R	ACTAATGATAAACCACATGA
12PCR-R	TTTGTACAATACTAATGATAAACCA
13PCR-F	TTAGGTCTTTCAGCTATAATGCAGG
13ASEQ-F	CAGCTATAATGCAGGTGTTT
13ASEQ-R	ACAAATCACGAGCAACTTCA
13BSEQ-F	GGCCGTGGCTTCTGCAAGAC
13BSEQ-R	CAGAATAGGTTGGCACATCA
13PCR-R	GGTCAAGCAACAGAATAGGTTGGCA
14PCR-F	ATAGTTTTTGATGGCAAGTCCAAAT
14ASEQ-F	ATGGCAAGTCCAAATGCGAC

(continued)

Table 4 (Continued)
Primer Sequences

14BSEQ-F	GTTGTTGATACCGATGTTGA
14BSEQ-R	AAAACAAGTACTAACAATCT
14ASEQ-R	TGACAGTTGTAACAATTTCA
14PCR-R	GCATAAGTTTAAAACAAGTACTAAC
15PCR-F	AGACTAACTTGTGCTACAAC TAGAC
15ASEQ-F	GTGCTACAAC TAGACAGGTT
15BSEQ-F	GCGTGGTGGTTCATACAAAAA
14BSEQ-R	AAAACAAGTACTAACAATCT
14ASEQ-R	TGACAGTTGTAACAATTTCA
14PCR-R	GCATAAGTTTAAAACAAGTACTAAC
15PCR-F	AGACTAACTTGTGCTACAAC TAGAC
15ASEQ-F	GTGCTACAAC TAGACAGGTT
15BSEQ-F	CGTGGTGGTTCATACAAAAA
15BSEQ-R	CTCCAGGTAAGTGTTAGGAA
15ASEQ-R	GCACAGTACCCGGTAAGCCA
15PCR-R	TAACAGAACCCTCCAGGTAAGTGTT
16PCR-F	GGTTCTATTTCTTATGGTGAGCTTC
16ASEQ-F	CTTATAGTGAGCTTCGTCCA
16ASEQ-R	ACTCACCAAAAACACGTCTG
16BSEQ-F	TTAGATGTGTCTGCTTCAGT
16BSEQ-R	AACTCCATTA AACATGACTC
16PCR-R	TACTAAATGTA ACTCCATTA AACAT
17PCR-F	ACAGCAATCTATGTATTCTGTATTT
17ASEQ-F	ATGTATTCTGTATTTCTCTG
17ASEQ-R	CTGACGGGAATGCCATTTTC
17BSEQ-F	TTTAGCAACTCAGGTGCTGA
17BSEQ-R	TTGGATACGGACAAATTTAT
17PCR-R	TTTGACCAGGTTGGATACGGACAAA
18PCR-F	ATTGGCCATTCTATGCAAAATTGTC
18ASEQ-F	CTATGCAAAATTGTCTGCTT
18ASEQ-R	CAGCATACAGCCATGCCAAA
18BSEQ-F	GAAGGTAAATTCTATGGTCC
18BSEQ-R	ACCTTGGAAGGTAACACCAG
18PCR-R	TCTTGAACTTACCTTGGAAGGTAAC
19PCR-F	GGTCGTA CTATCCTTGGTAGCACTA
19ASEQ-F	TCCTTGGTAGCACTATTTTA
19ASEQ-R	AGCTAGTGTCAGCCAATTCA
19BSEQ-F	TTACCTTCTCTTGCAACAGT
19BSEQ-R	AAATAACAATGGGTAATACT
19PCR-R	TGCCAGTAATAAATAACAATGGGTA

(continued)

Table 4 (Continued)
Primer Sequences

20PCR-F	ACCTCTAACTATTCTGGTGTCTGTTA
20ASEQ-F	ATTCTGGTGTCTGTTACGACT
20ASEQ-R	AGAGCAGTACCACAGATGTG
20BSEQ-F	AACATTAAGTTGTTGGGTAT
20BSEQ-R	ATTAGCTACAGCCTGCTCAT
20PCR-R	CAGAATCACCATTAGCTACAGCCTG
21PCR-F	CTTCAGGCTATTGCTTCAGAATTTA
21ASEQ-F	TTGCTTCAGAATTTAGTTCT
21ASEQ-R	CAACAACCATGAGTTTGGCT
21BSEQ-F	GATAATGATGCACTTAACAA
21BSEQ-R	GGCAAGTGCATTGTCATCAG
21PCR-R	TGTTATAGTAGGCAAGTGCATTGTC
22PCR-F	AATAATGAACTGAGTCCAGTAGCAC
22ASEQ-F	TGAGTCCAGTAGCACTACGA
22ASEQ-R	CTGCAAAAAGCACAGAAGGAA
22BSEQ-F	ATGGTGCTGGGCAGTTTAGC
22BSEQ-R	ACAGACTGTGTTTCTAAGTG
22PCR-R	CGCAGACGGTACAGACTGTGTTTCT
23PCR-F	GGATTCTGTGACTTGAAAGGTAAGT
23ASEQ-F	GACTTGAAAGGTAAGTACGTC
23ASEQ-R	TGTTGGTAGTTAGACATAGT
23BSEQ-F	TAAAAACTAATTGCTGTGCGC
23BSEQ-R	CTAAGTTAGCATATACGCGT
23PCR-R	ACACGCTCACCTAAGTTAGCATATA
24PCR-F	ACAATTGCTGTGATGATGATTATTT
24ASEQ-F	TGATGATGATTATTTCAATA
24ASEQ-R	AGACAAAGTCTCTCTTCCGT
24BSEQ-F	GGGCATTGGCTGCTGAGTCC
24BSEQ-R	CTGGATCAGCAGCATACACT
24PCR-R	GCATGCATAGCTGGATCAGCAGCAT
25PCR-F	GTGAGTTAGGAGTCGTACATAATCA
25ASEQ-F	AGTCGTACATAATCAGGATG
25ASEQ-R	CAATCAAAGTATTTATCAAC
25BSEQ-F	CTATCAGTGATTATGACTAT
25BSEQ-R	TTGACTTCAATAATTTCTGA
25PCR-R	GTGGCGGCTATTGACTTCAATAATT
26PCR-F	GTGCAAAGAATAGAGCTCGCACCGT
26ASEQ-F	TAGAGCTCGCACCGTAGCTG
26ASEQ-R	GATGATGTTCCACCTGGTTT
26BSEQ-F	CACACCGTTTCTACAGGTTA

(continued)

Table 4 (Continued)
Primer Sequences

26BSEQ-R	TGCTAGCTACTAAACCTTGA
26PCR-R	AAGTTCTTAATGCTAGCTACTAAAC
27PCR-F	TGCGTAAACATTTCTCCATGATGAT
27ASEQ-F	TTTCTCCATGATGATTCTTT
27ASEQ-R	TGAAAGACATCAGCATACTC
27BSEQ-F	AAACAGATGGTACACTTATG
27BSEQ-R	GATTAACAGACAACACTAAT
27PCR-R	CAAACATAGGGATTAACAGACAACA
28PCR-F	GTGCCTGTATTAGGAGACCATTCT
28ASEQ-F	TAGGAGACCATTCTATCTT
28ASEQ-R	CCATATGACAGCTTAAATGT
28BSEQ-F	CTGGCGATTACATACTTGCC
28BSEQ-R	CATAGTGCTCTTGTGGCACT
28PCR-R	GTAATTCTCACATAGTGCTCTTGTG
29PCR-F	ACAAGTTGAATGTTGGTGATTACTT
29ASEQ-F	TGTTGGTGATTACTTTGTGT
29ASEQ-R	GAATTCACTTTGAATTTATC
29BSEQ-F	TATGTGAAAAGGCATTAATA
29BSEQ-R	CAGCAGGACAACGGCGACAA
29PCR-R	TCAACAATTTTACAGCAGGACAACGGC
30PCR-F	TAGAACCAGAATATTTTAATTCAGT
30ASEQ-F	ATATTTTAATTCAGTGTGCA
30ASEQ-R	GTAGTTTGTGTGAATATGAC
30BSEQ-F	CACAGAACGCTGTAGCTTCA
30BSEQ-R	GTATGCCTGGTATGTCAACA
30PCR-R	ATGTCCTTTGGTATGCCTGGTATGT
31PCR-F	CTGGTCTTCATCCTACACAGGCACC
31ASEQ-F	TCCTACACAGGCACCTACAC
31ASEQ-R	AGATGTTTTAACTGGTCACC
31BSEQ-F	ACTTAGTAGCTGTACCGACT
31BSEQ-R	GCTGAACATCAATCATAAAT
31PCR-R	AAGCCCCACTGCTGAACATCAATCA
32PCR-F	GCTTTTCTACTTCATCAGATACTTA
32ASEQ-F	AAGTAGTCTATGAATACGGA
32ASEQ-R	TTCCATTCTACTTCAGCCTG
32BSEQ-F	TTGTGAAGTCTGCATTGCTT
32BSEQ-R	AAAAGAAAGGCAATTGCTTT
32PCR-R	TCAGAATAGTAAAAGAAAGGCAATT
33PCR-F	GTGGTAGTTTGTATGTGAATGGGCA
33ASEQ-F	GTATGTGAATAAGCATGCAT

(continued)

Table 4 (Continued)
Primer Sequences

33ASEQ-R	GCTTCGCCGGCGTGTCCATC
33BSEQ-F	CTTATAACCTGTGGAATACA
33BSEQ-R	GTGAAGAACAAGCACTCTCA
33PCR-R	AAGACAGTAAGTGAAGAACAAGCAC
34PCR-F	AAAGAGAAGCCCCAGCACATGTATC
34ASEQ-F	CCCAGCACATGTATCTACAA
34ASEQ-R	ATGAATTCATCCATAGCGAG
34BSEQ-F	TGCTGAAACCTACTTTACT
34BSEQ-R	TGACCACTTTTGAAATCACT
34PCR-R	ATTGTAACTTGACCACTTTTGAAA
35PCR-F	AATGTGTGTGTTCTGTGATTGATCT
35ASEQ-F	TTCTGTGATTGATCTTTTAC
35ASEQ-R	TTATCAGAGCCAGCACCAAAA
35BSEQ-F	ATGTCGCAAAGTATACTCAA
35BSEQ-R	CTGTTATCTTTACAGCTATA
35PCR-R	CAAGAATGCTCTGTTATCTTTACAG
36PCR-F	ATGACTCTAAAGAAGGGTTTTTCAC
36ASEQ-F	AGAAGGGTTTTTCACTTATC
36ASEQ-R	TCCAGAAGAGAATAAATCAT
36BSEQ-F	CACTCTTGACATGAGCAAAA
36BSEQ-R	TAGTATGAAACCCTGTAACA
36PCR-R	GTATGATTAATAGTATGAAACCCTG
37PCR-F	ATCCTGATGAAATTTTTAGATCAGA
37ASEQ-F	AATTTTTAGATCAGACACTC
37ASEQ-R	TTTTCTGAAACATCAAGCGA
37BSEQ-F	AACCCATGGGTACACAGACA
37BSEQ-R	TTGTACCATTTTCATCATA
37PCR-R	GCATCTGTGATTGTACCATTTTCAT
38PCR-F	AAGACATTTGGGGCACGTCAGCTGC
38ASEQ-F	GGGCACGTCAGCTGCAGCCT
38ASEQ-R	TTCAACTTAGTGGCAGAAAC
38BSEQ-F	GAAAAAAAAATTTCTAATTGT
38BSEQ-R	GAGCAGGTGGGGTGCAAGGT
38PCR-R	TAACAATTAAGAGCAGGTGGGGTGC
39PCR-F	ATCTTAGACATGGCAAGCTTAGGCC
39ASEQ-F	TGGCAAGCTTAGGCCCTTTG
39ASEQ-R	GGTGAAATGTCTAATATTTT
39BSEQ-F	CAAAGAGATTTCAACCATTT
39BSEQ-R	TTTGGCTAGTACTACGTAAT
39PCR-R	ACAATAGATTTTTGGCTAGTACTAC

(continued)

Table 4 (Continued)
Primer Sequences

40PCR-F	TCGACACTTCTTATGAGTGCACAT
40ASEQ-F	TTATGAGTGCACATTCTTA
40ASEQ-R	GTTGGGGTTTTGTACATTTG
40BSEQ-F	GCACACAACATAAATCGTGCA
40BSEQ-R	CACCAAATGTCCATCCAGCA
40PCR-R	GCAGGCCAGCACCAAATGTCCATC
41PCR-F	TGTTGCCACCTCTGCTCACTGATGA
41ASEQ-F	TCTGCTCACTGATGATATGA
41ASEQ-R	ATTTGTACCTCCGCCTCGAC
41BSEQ-F	ACACACTTGTTAAACAACCTT
41BSEQ-R	GGAAGTATGCTTTGCCTTCA
41PCR-R	CCTTCACGAGGGAAGTATGCTTTGC
42PCR-F	TTGTCTTCCTACATGTCACGTATGT
42ASEQ-F	ACATGTCACGTATGTGCCAT
42ASEQ-R	AAATTTTTAGCGACCTCATT
42BSEQ-F	CATCACCAGATGTTGATCTT
42BSEQ-R	ATTGATCCAAGAGTAAAAAA
42PCR-R	CTGTGCAGTAATTGATCCAAGAGTA
43PCR-F	ACTCTGAGCCAGTTCTCAAGGGTGT
43ASEQ-F	AGTTCTCAAGGGTGTCAAAT
43ASEQ-R	TTGTAGAAAATATATCAAGG
43BSEQ-F	ACTGCTGCTATTTGTTACCA
43BSEQ-R	TGGTAGTAAACTTCGGTGAA
43PCR-R	AGACTCAAGCTGGTAGTAAACTTCG
44PCR-F	CTACCAAATTGGTGGTTATTCTGAG
44ASEQ-F	GGTGGTTATTCTGAGGATAG
44ASEQ-R	GATGGCTAGTGTGACTAGCA
44BSEQ-F	TATGTACTCATTTCGTTTCGG
44BSEQ-R	AAATTGTAGTAACATAATCC
44PCR-R	TAGAATAGGCAAATTGTAGTAACAT
45PCR-F	ATTACCGTTGAGGAGCTTAAACAAC
45ASEQ-F	AGGAGCTTAAACAACCTCCTG
45ASEQ-R	CAATGACAAGTTCACCTTCC
45BSEQ-F	TCAATGTGGTCATTCAACCC
45BSEQ-R	ATTGTAACCTGGAAGTCAAC
45PCR-R	TATCTCTGCTATTGTAACCTGGAAG
46PCR-F	ACAGACCACGCCGGTAGCAACGACA
46ASEQ-F	CCGGTAGCAACGACAATATT
46ASEQ-R	GGGTGAAATGGTGAATTGCC
46BSEQ-F	GCGAGCTATATCACTATCAG

(continued)

Table 4 (Continued)
Primer Sequences

46BSEQ-R	ATTAAAACAAGGAATAGCAG
46PCR-R	AATAAGCATTATTAACAAGGAAT
47PCR-F	TCACCATTAAGAGAAAGACAGAATG
47ASEQ-F	GAGAAAGACAGAATGAATGA
47ASEQ-R	GGATCTTGACAGTTGATAGT
47BSEQ-F	CTGCTTGGCTTTGTGCTCTA
47BSEQ-R	CTTGCCATGCTGAGTGAGAG
47PCR-R	TAAGTTCCTCCTTGCCATGCTGAGT
48PCR-F	CGCAATGGGGCAAGGCCAAAACAGC
48ASEQ-F	CAAGGCCAAAACAGCGCCGA
48ASEQ-R	TTCCTTGAGGAAGTTGTAGC
48BSEQ-F	TGGGTGCAACTGAGGGAGC
48BSEQ-R	TTTTTGGCGAGGCTTTTTAG
48PCR-R	TGGCAGTACGTTTTTGGCGAGGCTT
49PCR-F	AGCAAAGTTTCTGGTAAAGGCCAAC
49ASEQ-F	CTGGTAAAGGCCAACAA
49ASEQ-R	GTGGGAATGTTTTGTATGCG
49BSEQ-F	ACTTATCATGGAGCCATTAA
49BSEQ-R	CTACTTGTGCTGTTTAGTTA
49PCR-R	TTAACTAAACCTACTTGTGCTGTTT
50PCR-F	TGGGCTATGTAAACGTTTTTCGCAAT
50ASEQ-F	AAACGTTTTTCGCAATCCGT
50PCR-R	TTACACATTAGGGCTCTTCCATATAGG

Sequences for primary (PCR-F and PCR-R) and secondary (PCR-F and BSEQ-R) PCR primers. ASEQ-F, BSEQ-F, ASEQ-R, BSEQ-R are the sequencing primers. The sequences are from 5' to 3'.

3.7. Sequence Comparison

1. Edit, align, and compare sequences using the Tor2 strain (GenBank accession number AY274119) as a reference with the software designed for this purpose, for example SeqScape (*see Note 9*).
2. Re-sequence regions that reveal nucleotide substitutions using a combination of different primer sets to ensure the quality of the sequencing data (*see Note 10*).

4. Notes

1. For RNA extraction, carrier poly(A) RNA is added to the lysis buffer to increase the yield. Because the PCR primers are specific to the SARS-CoV genome, the subsequent amplification would not be affected. However, if one wants to perform 3' rapid amplification of cDNA ends (3' RACE) or similar cloning operation that depends on oligo(dT) priming of poly(A) tail of the viral RNA, then the carrier poly(A) RNA should be avoided.

2. Our previous studies have shown that, with two rounds of PCR, even direct clinical samples can be sequenced. This obviates the need for viral culture, which may pose a health hazard if not handled properly. It also minimizes the possible generation of viral mutants through culturing. However, direct clinical samples of high viral titer and a sensitive PCR amplification are required.
3. Yields of viral RNA should be determined by quantitative RT-PCR, because spectrophotometric determination is prone to error as a result of low RNA quantity and interference by the carrier poly(A) RNA, which contributes to most of the RNA present.
4. A prolonged denaturation step is used to remove secondary RNA structures in the SARS-CoV genome that impede reverse-transcription. The use of random hexamer ensures an even representation of the whole RNA genome and allows more sequence information to be obtained from a limited amount of viral RNA.
5. We recommend the use of a reverse transcriptase with increased thermal stability, which facilitates reverse-transcription at a higher temperature (55°C) than normal (42°C). This unfolds some of the secondary RNA structures, and thus produces longer cDNA at higher yields.
6. We recommend the simultaneous use of two different DNA polymerases in the PCR amplification. For example, the cDNA polymerase mix that we use contains KlenTaq-1 DNA polymerase, and a second DNA polymerase with 3' to 5' proof-reading activity. The inclusion of a minor amount of a proofreading polymerase results in an error rate that is significantly lower than that for *Taq* alone (18). This advantage is obvious when one is concerned about genomic sequence variations between different viral strains. The use of a two-polymerase system also increases the efficiency and yield, and hence the sensitivity, which is important when the viral titer is suboptimal.
7. The carryover of unused PCR primers into the sequencing reaction would lead to poor sequencing results. Like the sequencing primers, these unused PCR primers would also bind nonspecifically to the sequencing template in the cycle sequencing reaction, and, hence, generate noisy sequencing traces overshadowing the intended traces. Purification of the PCR products is, thus, usually recommended prior to their use as sequencing templates. However, these methods are labor-intensive and pose extra contamination risk, as they involve additional steps of opening and handling PCR products. Notably, we have suggested an optimized PCR protocol for direct sequencing of PCR products without PCR product purification. With the low PCR primer concentrations and the optimal number of cycles, most of the PCR primers are consumed at the end of the PCR. Furthermore, a nested sequencing primer selectively extends the specific PCR product in the cycle sequencing reaction. This would suppress any nonspecific PCR product from extension. The combined effect is a neat sequencing trace.
8. The amount of PCR product used for the sequencing reaction must be optimized carefully with different sequencing systems. Although more PCR product input usually gives higher signal intensities, it may also give shorter read lengths and oversaturated signals.

9. The PCR primers target 50 700-bp amplicons that overlap with each other along the SARS-CoV genome. The sequencing primers are designed in such a way that any sequence masked over by the PCR primer binding sites and the sequencing primer peak on one amplicon are reliably backed up by the homologous sequence in the overlapping amplicon.
10. We advocate scrutinizing efforts in validating any genomic sequence variation by resequencing regions with different combination of primers and sequencing chemistry. Because variation seen in a single viral isolate could potentially be a result of sequencing artifacts, we consider only the genomic sequence variations that are shared by at least two SARS-CoV isolates.

Acknowledgments

This work is supported by a Special Grant for SARS Research (CUHK 4508/03M) from the Research Grants Council of the Hong Kong Special Administrative Region (China).

References

1. Poutanen, S. M., Low, D. E., Henry, B., et al. (2003) Identification of severe acute respiratory syndrome in Canada. *N. Engl. J. Med.* **348**, 1995–2005.
2. Lee, N., Hui, D., Wu, A., et al. (2003) A major outbreak of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **348**, 1986–1994.
3. Tsang, K. W., Ho, P. L., Ooi, G. C., et al. (2003) A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **348**, 1977–1985.
4. Rota, P. A., Oberste, M. S., Monroe, S. S., et al. (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399.
5. Marra, M. A., Jones, S. J., Astell, C. R., et al. (2003) The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404.
6. Ksiazek, T. G., Erdman, D., Goldsmith, C. S., et al. (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953–1966.
7. Drosten, C., Gunther, S., Preiser, W., et al. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976.
8. Peiris, J. S., Chu, C. M., Cheng, V. C., et al. (2003) Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**, 1767–1772.
9. Ng, E. K. O., Hui, D. S. C., Chan, K. C. A., et al. (2003) Quantitative analysis and prognostic implication of SARS coronavirus RNA in the plasma and serum of patients with severe acute respiratory syndrome. *Clin. Chem.* **49**, 1976–1980.
10. Ng, E. K. O., Ng, P. C., Hon, K. L., et al. (2003) Serial analysis of the plasma concentration of SARS coronavirus RNA in pediatric patients with severe acute respiratory syndrome. *Clin. Chem.* **49**, 2085–2088.

11. Chim, S. S. C., Tsui, S. K. W., Chan, K. C. A., et al. (2003) Genomic characterisation of the severe acute respiratory syndrome coronavirus of Amoy Gardens outbreak in Hong Kong. *Lancet* **362**, 1807–1808.
12. Tsui, S. K. W., Chim, S. S. C., and Lo, Y. M. D. (2003) Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. *N. Engl. J. Med.* **349**, 187–188.
13. Ruan, Y. J., Wei, C. L., Ee, A. L., et al. (2003) Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet* **361**, 1779–1785.
14. He, J. F., Peng, G. W., Min, J., et al. The Chinese SARS molecular epidemiology consortium (2004) Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**, 1666–1669.
15. Chim, S. S. C., Tong, Y. K., Hung, E. C. W., Chiu, R. W. K., and Lo, Y. M. D. (2004) Genomic sequencing of a SARS coronavirus isolate that predated the Metropole Hotel case cluster in Hong Kong. *Clin. Chem.* **50**, 231–233.
16. Chiu, R. W. K., Chim, S. S. C., and Lo, Y. M. D. (2003) Molecular epidemiology of SARS—from Amoy Gardens to Taiwan. *N. Engl. J. Med.* **349**, 1875–1876.
17. Chiu, R. W. K., Chim, S. S. C., Tong, Y. K., et al. (2005) Tracing SARS-coronavirus variant with large genomic deletion. *Emerg. Infect. Dis.* **11**, 168–170.
18. Barnes, W. M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91**, 2216–2220.