ELSEVIER

Journal of Clinical Virology 38 (2007) 19–26

JOURNAL OF CLINICAL VIROLOGY

www.elsevier.com/locate/jcv

Characterizing 56 complete SARS-CoV S-gene sequences from Hong Kong

Julian W. Tang [a], Jo L.K. Cheung [a], Ida M.T. Chu [a], Margaret Ip [a], Mamie Hui [a], Malik Peiris [b], Paul K.S. Chan [a,c,*]

[a] *Department of Microbiology, School of Public Health, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong SAR, China*

[b] *Department of Microbiology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong SAR, China.*

[c] *Centre for Emerging Infectious Diseases, School of Public Health, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong SAR, China*

Received 11 May 2006; received in revised form 18 September 2006; accepted 6 October 2006

Abstract

Background: The spike glycoprotein (S) gene of the severe acute respiratory syndrome-associated coronavirus (SARS-CoV) has been useful in analyzing the molecular epidemiology of the 2003 SARS outbreaks.

Objectives: To characterize complete SARS-CoV S-gene sequences from Hong Kong.

Study design: Fifty-six SARS-CoV S-gene sequences, obtained from patients who presented with SARS to the Prince of Wales Hospital during March–May 2003, were analysed using a maximum likelihood (ML) approach, together with 138 other (both human and animal) S-gene sequences downloaded from GenBank.

Results: The maximum-likelihood (ML) trees showed little evolution occurring within these 56 sequences. Analysis with the other sequences, showed three distinct SARS clusters, closely correlated to previously defined early, middle and late phases of the 2003 international SARS outbreaks. In addition, two new single nucleotide variations (SNVs), T21615A and T21901A, were discovered, not previously reported elsewhere.

Conclusions: The ML approach to the reconstruction of tree phylogenies is known to be superior to the more popular, less computationally and time-demanding neighbour-joining (NJ) approach. The ML analysis in this study confirms the previously reported SARS epidemiology analysed mostly using the NJ approach. The two new SNVs reported here are most likely due to the tissue-culture passaging of the clinical samples.

Keywords: SARS-CoV; S-gene; Phylogenetics; Maximum-likelihood; Neighbour-joining; Single nucleotide variation

1. Introduction

Since the severe acute respiratory syndrome (SARS) epidemic of 2003 around the world, many researchers have attempted to determine the natural reservoir of the SARS-associated coronavirus (SARS-CoV). Studies on the possible animal source of SARS-CoV have mainly focused on the Himalyan palm civet (*Paguma larvata*), though other animals (e.g. the raccoon dog, *Nyctereutes procyonoides*) have been shown to carry coronaviruses closely related to the SARS-CoV. Of note, these related animal coronaviruses possess a 29 base-pair sequence (position 27,869–27,897) in the

Abbreviations: SARS-CoV, severe acute respiratory syndrome-associated coronavirus; S-gene, spike glycoprotein gene; CS, clinical sample; TC, tissue culture; RT-PCR, reverse transcription polymerase chain reaction; SNV, single nucleotide variation; ML, maximum likelihood; NJ, neighbour-joining; CUHK, Chinese University of Hong Kong

* Corresponding author at: Department of Microbiology, School of Public Health, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China. Tel.: +852 2632 3333; fax: +852 2647 3227.

E-mail address: paulkschan@cuhk.edu.hk (P.K.S. Chan).

1386-6532/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jcv.2006.10.001

putative open reading frame (ORF) 11 (Marra et al., 2003), that is absent in most human SARS-CoV isolates (Guan et al., 2003; Song et al., 2005). A more recent study on SARS-CoV-like coronaviruses from palm civets and raccoon dogs from various live markets in China identified a series of single nucleotide variations (SNVs) in the SARS-CoV spike (S) glycoprotein gene (Kan et al., 2005). The authors suggested that these SNVs marked the transmission of SARS-CoV-like viruses from these animals into humans during various phases of the SARS epidemic. Most recently, SARS-like CoV have been found in bats though the S-gene homology with SARS-CoV is low, around 80%, and it is still uncertain as to whether this is the true natural reservoir for SARS-CoV (Lau et al., 2005; Li et al., 2005; Poon et al., 2005a,b).

Guan et al. (2004) described the genetic variation of the first 2149 bp of the S-gene, mainly focusing on the S1 (amino) region. This study compared local SARS-CoV S1 gene sequences collected from 137 Hong Kong SARS patients from February to April 2003, with 27 other sequences then available from GenBank, using neighbour-joining (NJ) phylogenetic analysis. The authors concluded that the international SARS epidemics were caused by closely related SARS-CoVs. However, some strains isolated from Hong Kong in the early phase of epidemic, were distinct from the international outbreak strain, and may have represented transitory strains of SARS-CoV (Chim et al., 2004). Furthermore, analysis of SARS-CoV strains subsequently isolated from China suggested that this virus may have been circulating there, unidentified, for some time before the international epidemic occurred (Chinese SARS Molecular Epidemiology Consortium, CSMEC, 2004; Zhao et al., 2004).

In order to further characterize the evolution of SARS-CoV, S-gene sequences of isolates from patients in Hong Kong were analysed using a maximum likelihood (ML) method, and any new SNVs documented.

2. Materials and methods

2.1. Patients, sample collection and processing

Patients with laboratory-confirmed SARS-CoV infection admitted to the New Territories East Cluster Hospitals in Hong Kong were included (Chan et al., 2004). For each patient, a stored original clinical sample (CS), or its primary isolate from tissue culture (TC) when the original sample has been exhausted, was retrieved for this study. The nucleotide sequence of the whole S-gene was obtained by direct sequencing. Briefly, the extracted RNA was amplified with two sets of overlapping primers using Superscript™ III One-Step RT-PCR System with Platinum *Taq* DNA polymerase (Invitrogen, Life Technology, Carlsbad, CA). Purified PCR products were sequenced using six internal sequencing primers using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Their sequences were edited and aligned using the Seqscape v2.1.1 software (Applied Biosystems, Foster City, CA).

All 56 Hong Kong S-gene sequences obtained in this study were deposited on the National Center for Biotechnology Information Genome Database (GenBank, accession numbers: DQ412574–DQ412629).

2.2. SARS-CoV S-gene phylogenetic analysis

The S-gene sequences were aligned using the Clustal X's multiple alignment algorithm, then manually checked and edited using Bioedit v7.0.4.1 (Tippmann, 2004). After alignment and editing, the total length of the S-gene sequences was 3759 bp. A maximum likelihood (ML) tree were then constructed using the program PAUP* v4.0b10 (Swofford, 2001) under an optimum model of evolution as selected by the program Modeltest v3.7 (Posada and Crandall, 1998). The robustness of the trees' topology was statistically assessed by bootstrap analysis, with a minimum of 1000 rounds of replication. To put these 56 CUHK S-gene sequences in context, 138 other S-gene sequences were downloaded from the GenBank and these 194 S-gene sequences were drawn, using the same methods above. Both trees were rooted against a civet cat sequence (CIV007, AY572034), and plotted using NJPlot (Perrière and Gouy, 1996).

3. Results

3.1. Patients, sample collection and processing

A total of 56 SARS patients were included in this study (mean age 54.8, S.D. 21.9 years, 21 males). These patients all became ill during March–May 2003. Samples for analysis were collected after a mean of 5.9 (S.D. 4.8) days after illness onset. Table 1 shows the specimen type and numbers used in this study, 11 were original clinical samples (CS) and 45 were primary tissue culture isolates (TC).

3.2. SARS-CoV S-gene phylogenetic analysis

The ML trees in Figs. 1 and 2 were both constructed using a Kimura three-parameters model of base substitution with unequal base frequencies, a proportion of invariant sites and

Table 1
Types of specimens used for S-gene sequencing

Specimen type	No. of specimens tested	Clinical sample (CS)	Tissue culture isolate (TC)
Respiratory[a]	44	4	40
Stool	5	5	0
Urine	2	2	0
Tissue[b]	5	0	5
Total	56	11	45

[a] Includes nasopharyngeal aspirate, throat and nasal swabs.
[b] Includes four lung and one terminal ileum biopsy samples.