

IDENTIFICATION AND CHARACTERIZATION OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS SUBGENOMIC RNAs

Snawar Hussain, Ji'an Pan, Jing Xu, Yalin Yang, Yu Chen, Yu Peng,
Ying Wu, Zhaoyang Li, Ying Zhu, Po Tien, and Deyin Guo*

1. INTRODUCTION

Severe acute respiratory syndrome (SARS) is an atypical form of pneumonia that was first recognized in Guangdong Province, China, in November 2002, and its causative agent was identified as novel coronavirus (SARS-CoV).¹⁻³ Coronaviruses are the largest RNA viruses, containing a single-stranded, plus-sense RNA ranging from 27 kb to 31.5 kb in size. The two large open reading frames (ORFs) (1a and 1b) at the 5'-end of the genome encode the viral replicase and are translated directly from the genomic RNA, while 1b is expressed by -1 ribosomal frameshifting.⁴ The 3'-one third of the genome comprises the genes encoding structural proteins S, E, M, and N and a number of auxiliary proteins of unknown function. These proteins are translated through 6-9 nested and 3'-coterminal subgenomic RNAs (sgRNAs).

The genomes of many SARS-CoV isolates have been sequenced, and consist of approximately 29,700 nucleotides.⁵⁻⁷ Fourteen ORFs have been identified, of which 12 are located in the 3'-proximal one-third of the genome.^{4,5} The exact mechanisms of expression of the 3'-proximal ORFs are unknown but on the analogy with other coronaviruses, these ORFs are predicted to be expressed through a set of sgRNAs.⁶ Identification of SARS-CoV sgRNAs in infected cells and characterization of molecular details of the leader-body fusion in the sgRNAs will help elucidate the regulatory mechanism of SARS-CoV transcription and replication. This knowledge will be useful in the development of antiviral therapeutic agents and vaccine for treatment and prevention of this newly emerged disease.

*Wuhan University, Wuhan, People's Republic of China.

2. IDENTIFICATION OF NOVEL SUBGENOMIC RNAS

Northern blot, RT-PCR, and DNA sequencing revealed the existence of ten subgenomic RNAs including two novel subgenomic RNAs named 2-1 and 3-1. The leader-body fusion site (ACGAgC) of subgenomic RNA 2-1 has one nucleotide mismatch (lowercase) with SARS-CoV leader core sequence (CS-L) ACGAAC and is located inside the S gene, 384 nucleotides downstream from the authentic CS (ACGAAC) for mRNA 2/S. The second novel subgenomic RNA (3-1) corresponded to the 3b ORF that was predicted to be expressed from mRNA 3. The leader-body fusion site (AaGAAC) for subgenomic mRNA 3-1 is 10 nucleotides upstream of AUG start codon of ORF 3b and has a mismatch (lowercase) with the SARS-CoV leader core sequence (ACGAAC).

To determine whether these new subgenomic RNAs are functional messages, the 5'-end (containing leader sequence plus downstream 150 nucleotides in case of sgRNA 2-1 and 400 nucleotides in case of sgRNA 3-1) were fused with green fluorescent protein (GFP) gene, and sgRNA codon usage was indirectly determined by the expression of the reporter gene. Strong fluorescence was observed in cells transfected with the construct in which the ORF2b was fused in-frame with GFP (Figure 1A). Western blot with anti-GFP monoclonal antibody confirmed the existence of a 32 kDa band corresponding to the fusion protein. A 27 kDa band corresponding to the wild-type GFP resulting from downstream AUG by leaky scanning was also detected in cells transfected with in-frame and out-frame construct (Figure 1B, upper panel).

Cells transfected with the ORF3b in-frame construct displayed weak fluorescence while no fluorescence was observed in cells transfected with the ORF3b out-frame construct (Figure 1A). Western blot analysis confirmed the existence of a 42, kDa band corresponding to 3b-GFP fusion protein (Figure 1B, lower panel). These results indicated that sgRNA 2-1 and 3-1 could function in the environment of the cell.

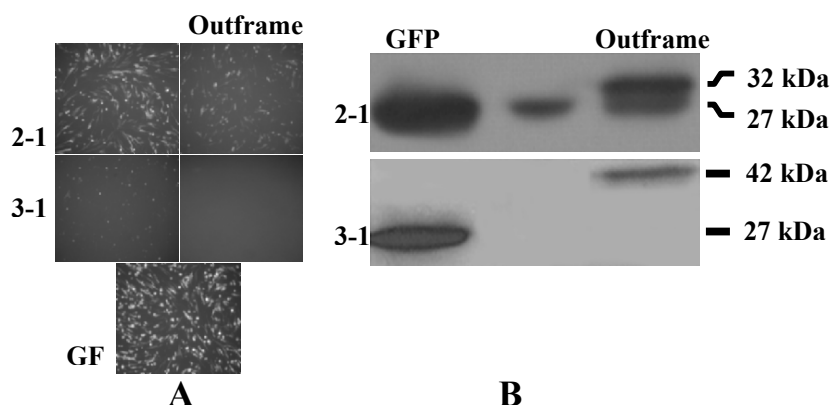


Figure 1. Translatability of novel subgenomic RNAs 2-1 and 3-1. (A) The ORF 2b (sgRNA 2-1) and 3b (sgRNA 3-1) were fused in-frame and out of frame at the 5'-end of GFP gene. The expression of GFP-fusion protein was qualitatively assessed by fluorescent microscopy (magnification: X200). (B) Proteins were extracted from transfected cells and separated by 12% SDS-PAGE. GFP-Fusion proteins were detected by anti-GFP monoclonal antibody.

3. EXPRESSION PROFILE OF SARS-COV SUBGENOMIC RNAs

The 5'-ends (including leader sequence and 200-400 nucleotides of the 5'-end of body sequence) of all 10 (2/S, 2-1, 3, 3-1, 4/E, 5/M, 6, 7, 8, 9/N) subgenomic RNAs were amplified by RT-PCR and fused to the 5'-end of the GFP reporter gene. The AUG initiator codon usage of subgenomic RNAs was indirectly assessed by level of expression of reporter gene.

Strong fluorescence was observed in the cells transfected with fusion constructs containing the 5'-ends of subgenomic RNAs 2-1, 3, 4/E, 5/M, 6 and 9/N, whereas relatively weak fluorescence was observed in cells transfected with fusion constructs containing 5'-ends of subgenomic RNAs 2, 3-1, 7 and 8 (Figure 2A).

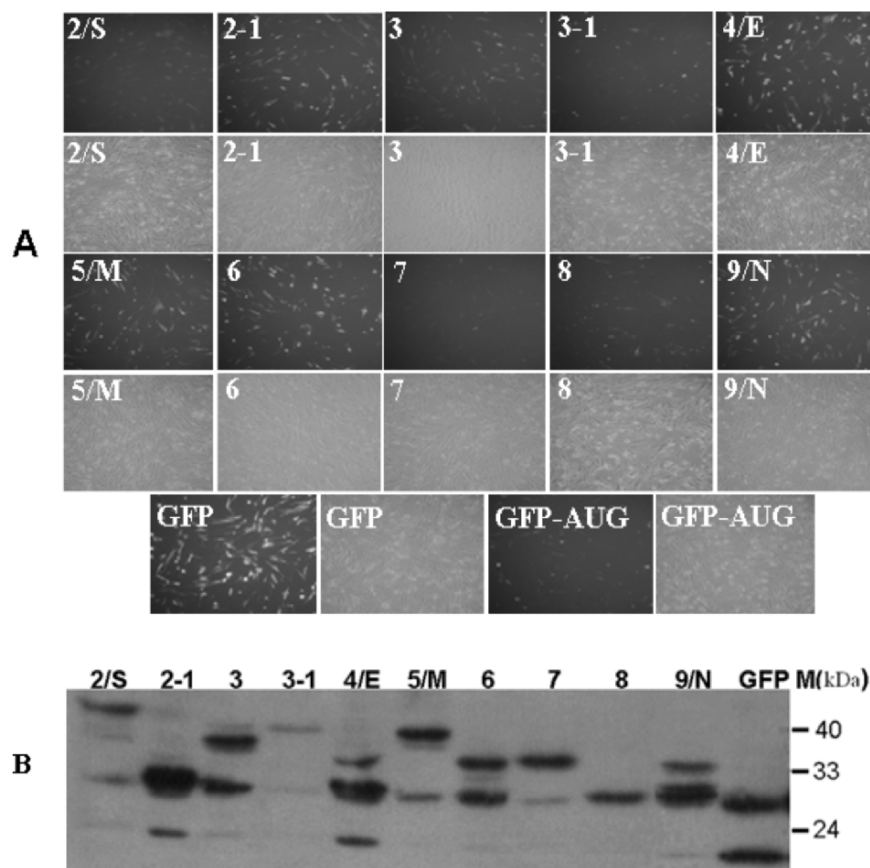


Figure 2. Expression of GFP-fusion protein in BHK cells. (A) The ORFs of all ten subgenomic were fused in-frame with GFP gene. The expression of GFP-fusion proteins was qualitatively assessed by fluorescent microscopy (magnification: X200). (B) Proteins were extracted from transfected cells and separated by 12% SDS-PAGE. GFP-Fusion proteins were detected by anti-GFP monoclonal antibody. Names of the individual subgenomic RNAs are marked on top and molecular weight (kDa) is marked on right side of image.

A single band of fusion protein was detected by Western blot in cells transfected with constructs containing the 5'-end of sgRNA 2, 2-1, 3, 3-1, 4, 6 and 7, whereas, in cells transfected with mRNA 5 and 9, beside the expected fusion protein band, a low molecular weight band was also detected that may result from a downstream AUG codon (Figure 2B).

We also detected some minor low molecular weight bands in some cases, possibly resulting from downstream AUG codons by leaky scanning. Cells transfected with subgenomic RNA8-GFP fusion construct displayed fair amount of fluorescence but in Western blot only a 27 kDa band of wild-type GFP was detected. Taken together-we showed that nine out of ten subgenomic RNAs could be functional messages *in vivo*; however, existence of proteins encoded by these subgenomic in SARS-CoV infected cells is yet to be determined.

In summary, ten subgenomic RNAs including two novel subgenomic RNAs (2-1/3-1) were identified in SARS-CoV infected cells by Northern blot, RT-PCR, and DNA sequencing. Nine out of 10 subgenomic RNAs were potentially functional messages in SARS-CoV infected cells. The initiator AUG codon of subgenomic RNA 8 (ORF8b) was inactive in our experimental set-up.

4. REFERENCES

1. R. A. Fouchier, R. Kuiken, M. Schutten, et al., Aetiology: Koch's postulates fulfilled for SARS virus, *Nature* **423**, 240 (2003).
2. T. G. Ksiazek, D. Erdman, C. S. Goldsmith, et al., A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953-1966 (2003).
3. J. S. Peiris, C. M. Chu, V. C. Cheng, et al., Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study, *Lancet* **361**, 1767-1772 (2003).
4. V. Thiel, K. A. Ivanov, A. Putics, et al., Mechanisms and enzymes involved in SARS coronavirus genome expression, *J. Gen. Virol.* **84**, 2305-2315 (2003).
5. M. A. Marra, S. J. Jones, C. R. Astell, et al., The genome sequence of the SARS-associated coronavirus, *Science* **300**, 1399-1404 (2003).
6. P. A. Rota, M. S. Oberste, S. S. Monroe, et al., Characterization of a novel coronavirus associated with severe acute respiratory syndrome, *Science* **300**, 1394-1399 (2003).
7. Y. J. Ruan, C. L. Wei, A. L. Ee, et al., Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection, *Lancet* **361**, 1779-1785 (2003).