A New Nucleotide-composition Based Fingerprint of SARS-CoV with Visualization Analysis[a]

Meng Wang[1,#], Jin-Song Yao[2,#], Zhen-De Huang[2,3,#], Zhi-Jie Xu[1], Guo-Ping Liu[1], Hai-Yong Zhao[2], Xue-Yong Wang[2], Jie Yang[1], Yi-Sheng Zhu[2] and Kuo-Chen Chou[1,2,3,4,*]

[1]Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030, China, [2]Department of Biomedical Engineering, Shanghai Jiaotong University, Shanghai 200030, China, [3]Bioinformatics Research Centre, Donghua University, Shanghai 200050, China, [4]Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA

Abstract: It has been observed by conducting an extensive analysis of the two-dimensional cellular automata images of known SARS-CoV genome sequences that the V-shaped cross-lines only exist in some special locations, and hence can be used as a fingerprint to identify the SARS sequences. Such a discovery can be used to rapidly and reliably diagnose SARS coronavirus for both basic research in laboratories and practical application in clinics.

Key Words: Cellular automata images, SARS, RNA genomic, Visualization analysis, V-shaped cross-line pattern, parallel slash-line pattern.

I. INTRODUCTION

A previously unrecognized coronavirus, called SARS-coronavirus (SARS-CoV), is the leading hypothesis for the cause of SARS. Inhibiting the replication of SARS-CoV is a promising avenue to find drugs against SARS [1][2][3][4]. The coronaviruses are members of a family of enveloped viruses that replicate in the cytoplasm of animal host cell. The sequence of entire genome of SARS-CoV can be obtained from the isolates in the worldwide laboratories. These nucleotide sequences vary at only a few positions. The SARS virus contains 11 predicted open reading frames that encode 23 putative mature proteins with known and unknown functions. Most of the non-structural proteins seem to be encoded in the first half of the genome including nsp1, nsp2 and nsp9 RNA-dependent RNA polymerase, whereas most of the structural proteins such as spike, membrane, envelop, and nucleocapsid are located in the second half of the genome. Sequence conservation seems to be restricted to the middle part of the genome, between bases 14,000 and 21,000, where the RNA-dependence polymerase and several uncharacterized proteins are located. The haemagglutinin esterase, which is common in the coronavirus genome, is missing in the SARS-CoV, suggesting that some of the non-structural genes are dispensable in coronavirus. The mutations were in the following open reading frames: orf1a polyprotein, orf1a RNA-polymerase, orf1ab, spike glycoprotein, membrane nucleocapsid, and several uncharacterized putative proteins [5]. It can be found that SARS-CoV genome has the same frame of the structural proteins with other coronaviruses, because almost all the structural proteins existing in previously known coronavirus have been identified in SARS-CoV in the same order [6]. So, on the basis of comparative genomics, the homology search, phylogenetic analysis, and multi-sequence alignment were used to provide clues for studying the functions and characteristic of the structure of SARS proteins [7].

Coronaviruses typically have narrow host ranges and are fastidious in cell culture. There are three groups of coronaviruses: group1 and group2 contain mammalian viruses, whereas group3 contains only avian viruses. Within each group coronaviruses are classified into distinct species by host range, antigenic relationships, and genomic organization. However, phylogenetic analyses of the predicted viral proteins indicated that SARS-CoVs did not closely resemble any of the three previously known groups of coronaviruses, but formed a distinct group within the genus coronavirus. No evidence for recombination was detected between SARS-CoV and other coronaviruses, so the SARS-CoV can be considered as “group4” coronavirus [8,9].

After the SARS CoV is etiologically linked to the outbreak of SARS, several diagnostic assay methods have been established. The standard method of SARS checkup is the virus culture. However, owing to the rigor culture conditions, isolating SARS-CoVs is suitable only for some special laboratories and false negative consequences might occur. Since the initial completion of the SARS genome sequence, a few RT-PCR protocols have been developed from several independent laboratories. Although the PCR technique can rapidly detect the SARS-CoV RNA in samples, it requires the inclusion of both negative and positive controls, and hence a clinical oriented PCR diagnostic assay with high efficiency and reliability is not yet available. The ELISA and IFA methods on detecting SARS-CoV have the characteristics of convenience, rapidness, and trustiness, but they can not be used in forepart diagnosis of SARS: specific antibody can only be checked after 10 days with infection [10]. Therefore, the cases of SARS were so far mainly detected by the patient’s clinic

*Address correspondence to this author at the Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030, China or Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA; E-mail: kchou@san.rr.com

# Meng Wang, Jin-Song Yao, and Zhen-De Huang have made equivalent contribution to this study.

symptoms, microscopy, histochemical assay, chest X-ray assay and CT check-up, contact history with pathogen *et al*.

In this paper, the nucleotide-composition based fingerprint of SARS-CoV is discovered with the cellular automata (CA) method [11,12].

II. THE FINGERPRINT OF SARS-COV

Analyzing the available 96 SARS-CoVs and other 25 coronaviruses (Tables **1**, **2**) (all downloaded from GenBank: http://www.ncbi.nlm.nih.gov) with our method that will be introduced in the following section, the characteristic of SARS-CoV is discovered. From about 3040 to 5439nt in the SARS-CoV sequence near 5-terminal [13,14] where the V-shaped cross-lines appear (Fig. **1**), the occurrence numbers of repeated character 'A' (i.e., 'AA', 'AAA', 'AAAA') are obviously larger than those of repeated character 'T' (i.e., 'TT', 'TTT', 'TTTT') as shown in Table **1**. However, for the 25 non-SARS-CoVs, almost all the corresponding occurrence numbers of 'AA', 'AAA', 'AAAA' are less than those of 'TT', 'TTT', 'TTTT' in the same region (Table **2**) as reflected by the feature of many parallel slash-lines (Fig. **2**). Thus, such a unique feature of V-shaped cross-lines can be used to define the fingerprint of SARS-CoVs, in contrast to non-SARS-CoVs. To illustrate their differences, the ratios of different

Table 1. The Statistical Results of SARS-CoVs for the Repeated Character 'A', 'T', 'G', 'C'. Here 2n Denotes 'AA',3n Denotes 'AAA', and so on.

Name	A				T				G				C			
	2n	3n	4n	5n	2n	3n	4n	5n	2n	3n	4n	5n	2n	3n	4n	5n
bj01	162	48	13	3	153	37	6	0	96	7	0	0	68	7	1	0
bj02	167	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
bj03	162	47	13	3	154	37	6	0	97	7	0	0	69	8	1	0
bj04	163	48	13	3	154	37	6	0	97	7	0	0	68	7	1	0
gz01	162	48	13	3	152	36	5	0	96	7	0	0	69	7	1	0
zj01	162	48	13	3	157	38	7	0	93	7	0	0	68	7	1	0
hku39849	162	48	13	3	157	38	7	0	93	7	0	0	69	7	1	0
cuhkw1	162	48	13	3	157	38	7	0	93	7	0	0	68	7	1	0
cuhksu10	162	48	13	3	157	38	7	0	93	7	0	0	68	7	1	0
sin2500	161	48	13	3	157	38	7	0	93	7	0	0	70	7	1	0
sin2677	162	48	13	3	157	38	7	0	93	7	0	0	68	7	1	0
sin2679	164	48	13	3	155	37	6	0	97	7	0	0	68	7	1	0
sin2748	162	48	13	3	155	38	6	0	94	7	0	0	68	7	1	0
sin2774	162	48	13	3	157	38	7	0	93	7	0	0	68	7	1	0
tw1	162	48	13	3	155	38	6	0	94	7	0	0	68	7	1	0
urbani	162	48	13	3	157	38	7	0	93	7	0	0	68	7	1	0
tor2	162	48	13	3	155	38	6	0	94	7	0	0	68	7	1	0
gz50	166	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
sz16	166	49	12	3	151	37	6	0	99	5	0	0	69	8	1	0
sz3	166	49	12	3	151	37	6	0	99	5	0	0	69	8	1	0
fra	166	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
gd01	166	49	12	3	151	36	5	0	101	5	0	0	69	8	1	0
twc	166	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
twc2	166	49	12	3	152	37	6	0	99	5	0	0	70	8	1	0
twc3	166	49	12	3	152	37	6	0	99	5	0	0	70	8	1	0

(Table 1. Contd….)

Name	A				T				G				C			
	2n	**3n**	**4n**	**5n**	**2n**	**3n**	**4n**	**5n**	**2n**	**3n**	**4n**	**5n**	**2n**	**3n**	**4n**	**5n**
zmy1	165	48	12	3	152	37	6	0	100	8	0	0	70	8	1	1
twy	166	49	12	3	152	37	6	0	99	5	0	0	70	8	1	0
tws	166	49	12	3	152	37	6	0	99	5	0	0	70	8	1	0
twk	165	49	12	3	153	38	7	0	99	5	0	0	71	9	1	0
twj	166	49	12	3	152	37	6	0	100	5	0	0	69	8	1	0
twh	165	49	12	3	153	38	7	0	99	5	0	0	71	9	1	0
tc3	166	49	12	3	152	37	6	0	99	5	0	0	71	8	1	0
tc2	166	49	12	3	152	37	6	0	99	5	0	0	71	8	1	0
tc1	166	49	12	3	152	37	6	0	99	5	0	0	71	8	1	0
hsr1	166	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
frankfurt1	166	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
as	165	50	12	3	153	37	6	0	99	5	0	0	68	8	1	0
cuhk-ag01	165	50	12	3	153	37	6	0	99	5	0	0	69	8	1	0
cuhk-ag02	165	50	12	3	153	37	6	0	99	5	0	0	69	8	1	0
cuhk-ag03	165	50	12	3	153	37	6	0	99	5	0	0	69	8	1	0
gd69	166	49	12	3	152	37	6	0	100	5	0	0	69	8	1	0
pumc01	166	50	12	3	152	37	6	0	99	5	0	0	68	8	1	0
pumc02	166	50	12	3	153	37	6	0	99	5	0	0	68	8	1	0
pumc03	166	50	12	3	153	37	6	0	99	5	0	0	68	8	1	0
sino1-11	166	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
sino3-11	166	49	12	3	152	37	6	0	99	5	0	0	69	8	1	0
sod	165	50	12	3	153	37	6	0	99	5	0	0	68	8	1	0
gz02	162	48	12	3	154	38	8	0	99	5	0	0	72	9	1	0
zs-c	166	49	12	3	152	37	6	0	100	5	0	0	69	8	1	0
lc5	166	49	12	3	152	37	6	0	100	5	0	0	68	8	1	0
lc4	164	48	12	3	154	38	7	0	99	5	0	0	71	9	1	0
lc3	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
lc2	164	50	12	3	153	37	6	0	100	5	0	0	68	8	1	0
lc1	166	50	12	3	153	37	6	0	99	5	0	0	68	8	1	0
zs-a	165	50	12	3	153	37	6	0	99	5	0	0	69	8	1	0
zs-b	165	49	12	3	152	38	6	0	99	5	0	0	71	9	1	0
hsz-cc	165	49	12	3	153	37	6	0	99	5	0	0	70	9	1	0
hsz-bc	167	51	13	3	154	37	6	0	99	5	0	0	68	8	1	0
hgz8l2	167	51	13	3	153	37	6	0	99	5	0	0	68	8	1	0
hzs2-c	165	48	12	3	154	38	7	0	98	5	0	0	71	9	1	0

(Table 1. Contd….)

Name	A				T				G				C			
	2n	3n	4n	5n	2n	3n	4n	5n	2n	3n	4n	5n	2n	3n	4n	5n
hzs2-fc	166	50	12	3	153	37	6	0	99	5	0	0	68	8	1	0
hzs2-e	167	51	13	3	154	37	6	0	98	5	0	0	69	8	1	0
hzs2-d	165	49	12	3	153	38	7	0	99	5	0	0	71	9	1	0
hzs2-fb	165	49	12	3	152	38	6	0	99	5	0	0	70	9	1	0
hsz-cb	166	49	12	3	153	37	6	0	100	5	0	0	68	8	1	0
hsz-bb	166	49	12	3	153	37	6	0	100	5	0	0	68	8	1	0
LLJ-2004	166	49	12	3	152	37	6	0	100	5	0	0	68	8	1	0
Sin842	165	49	12	3	153	38	6	0	99	5	0	0	70	9	1	0
Sin845	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
Sin846	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
Sin847	165	49	12	3	152	38	7	0	99	5	0	0	71	9	1	0
Sin848	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
Sin849	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
Sin850	165	49	12	3	153	38	7	0	99	5	0	0	71	9	1	0
Sin852	164	48	12	3	154	38	7	0	99	5	0	0	71	9	1	0
Sin3408	165	49	13	3	155	38	8	0	99	5	0	0	70	8	1	0
Sin3408L	165	49	12	3	154	37	6	0	99	5	0	0	70	9	1	0
Sin3725V	166	49	12	3	152	37	6	0	99	5	0	0	70	8	1	0
Sin3765V	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
SinP1	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
SinP2	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
SinP3	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
SinP4	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
SinP5	165	49	12	3	152	37	6	0	99	5	0	0	70	9	1	0
TJF	164	48	12	3	154	38	7	0	99	5	0	0	71	9	1	0
TW2	164	48	12	3	154	38	7	0	100	5	0	0	71	9	1	0
TW3	164	48	12	3	154	38	7	0	100	5	0	0	71	9	1	0
TW4	164	48	12	3	154	38	7	0	100	5	0	0	71	9	1	0
TW5	164	48	12	3	154	38	7	0	99	5	0	0	71	9	1	0
TW6	164	48	12	3	154	38	7	0	99	5	0	0	72	9	1	0
TW7	164	48	12	3	154	38	7	0	99	5	0	0	72	9	1	0
TW8	164	48	12	3	154	38	7	0	99	5	0	0	72	9	1	0
TW9	164	48	12	3	154	38	7	0	99	5	0	0	72	9	1	0
TW10	164	48	12	3	154	38	7	0	99	5	0	0	72	9	1	0
TW11	163	48	12	3	154	38	7	0	99	5	0	0	72	9	1	0
WHU	165	48	12	3	153	38	7	0	99	5	0	0	71	9	1	0

Table 2. The Statistical Results of non-SARS-CoVs for the Repeated Character 'A', 'T', 'G', 'C'. Here 2n Denotes 'AA', 3n Denotes 'AAA', and so on.

Name	A				T				G				C			
	2n	3n	4n	5n	2n	3n	4n	5n	2n	3n	4n	5n	2n	3n	4n	5n
AF029248_murine4	138	38	7	3	180	58	20	5	117	19	2	0	62	10	1	0
AF029248_Murine	138	38	7	3	180	58	20	5	117	19	2	0	62	10	1	0
AF201929_ Murine	130	35	4	1	192	65	25	5	116	16	1	0	63	12	0	0
AF207902_murine2	131	35	4	1	192	65	25	5	116	16	1	0	63	12	0	0
AF208066_Murine	133	35	5	2	195	64	25	6	116	15	1	0	63	12	0	0
AF208067_murine1	138	38	7	3	177	57	19	5	119	19	2	0	62	10	1	0
AF250295_bovine1	153	51	4	1	246	77	33	5	95	10	3	2	38	7	0	0
AF304460_259E	185	54	16	5	250	68	21	6	77	10	0	0	48	5	0	0
AF353511_porcine1	131	34	7	3	212	52	25	6	113	11	1	0	79	8	0	0
AF391541_ Bovine	153	52	4	1	238	77	30	5	91	9	1	1	39	6	0	0
AF391542_ Bovine	154	52	5	1	240	77	30	5	92	9	1	1	39	6	0	0
AJ311317_avain1	193	57	15	2	207	52	24	6	84	14	3	0	55	5	2	0
AY319651_ Avian	182	54	17	6	210	54	16	2	86	11	3	0	62	6	1	0
AY391777_HCoV	154	47	5	1	250	79	32	6	90	9	1	1	39	6	1	0
AY514485_Avian	182	54	14	3	198	48	23	7	85	12	2	0	53	7	0	0
AY700211_Murine	138	38	7	3	180	58	20	5	117	19	2	0	63	10	1	0
D13096_avain	193	57	15	2	207	52	24	6	84	14	3	0	55	5	2	0
NC_001451_ Avian	193	57	15	2	207	52	24	6	84	14	3	0	55	5	2	0
NC_001846_murine	138	38	7	3	180	58	20	5	117	19	2	0	62	10	1	0
NC_002645_259E	185	54	16	5	250	68	21	6	77	10	0	0	48	5	0	0
NC_003045_ Bovine	153	52	4	1	238	77	30	5	91	9	1	1	39	6	0	0
NC_003436_porcine	131	34	7	3	212	52	25	6	113	11	1	0	79	8	0	0
NC_005147_HCoV	154	47	5	1	250	79	32	6	90	9	1	1	39	6	1	0
NC_005831_NL63	170	52	19	8	299	85	42	10	79	10	0	0	39	3	0	0
U00735_bovine	154	51	4	1	246	78	32	5	94	11	3	2	39	6	1	0

repeated characters 'T' and 'A' in the given region (from 3040 to 5439nt near 5-terminal) are drawn in Fig. **3a**. It is clear that the ratios of SARS-CoVs are all lower than 1, whereas those of non-SARS-CoVs are larger than 1. The average ratios of all the SARS-CoVs and non-SARS-CoVs in the given region are shown in Fig. **3b**.

III. METHOD

The BLAST and FASTA are often used for the sequence alignment. The program CLUSTALW can produce a distance-based evolutionary tree as a multiple sequence alignment. All these methods are based on the similarity measurement of sequences that have strong co-genetic relations. However, they can not tell us what the characteristic of SARS-CoV is on earth. Since CA [11,12] (see Box **1**) is renowned with the power in treating complex systems with simple rules, it is employed to visualize the aforementioned 96 SARS-CoVs and 25 non-SARS-coronaviruses. In the current approach, we use '0-1' with length about 60kb for the initial line of the sequence, while the corresponding initial line in CA [11] is a black-white dot. In other words, a nucleotide sequence is coded as follows: A=00, T=11, G=10, and C=01.

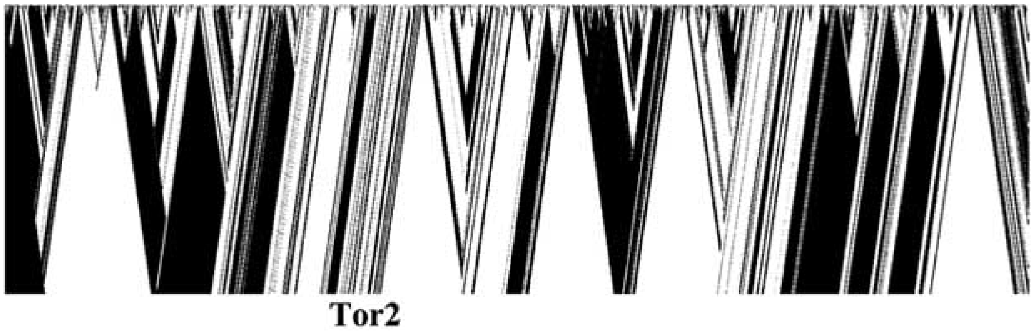


Fig. (1). Sample image obtained by applying the modified Rule 184 on SARS coronavirus: Tor2. All 96 SARS images are with a V-shaped cross-lines pattern, a token for SARS coronaviruses.

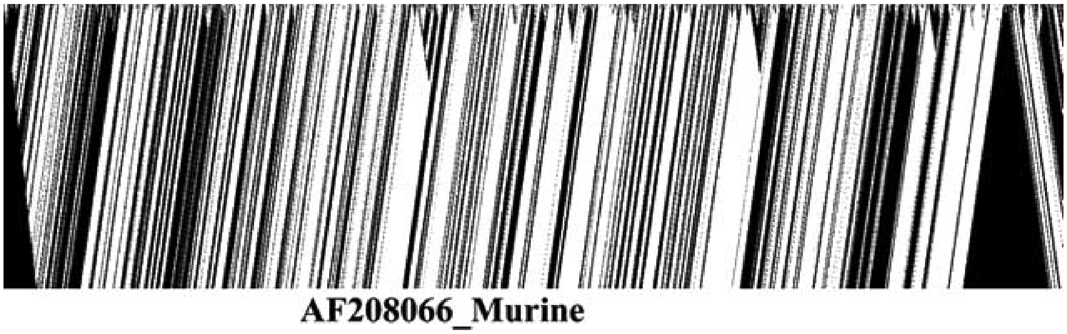


Fig. (2). Sample images obtained by applying the modified Rule 184 on the non-SARS coronavirus: AF208066_Murine. All 25 non-SARS images are with a parallel slash-lines pattern, a remarkable distinction with the SARS coronal viruses.

Different rules have been applied to analyze the aforementioned 96+25=121 coronaviruses. But only when the Rule 184 (see Box **2**) is used, are the images of SARS-CoVs different from those of other coronaviruses most distinctively.

Using Rule 184, we ran 2,400 times and finally got a binary image with the size of about 60Kbx2.4Kb. The image was too large for analysis. In the zoomed out images (Figs. **1**, **2**), the images of SARS-CoVs are mainly with the V-shaped cross-lines, whereas those of non-SARS virus RNA sequences are mainly with parallel slash-lines, indicating that more numbers of repeated character ‘A’ than those of repeated character ‘T’ is probably the reason that leads to the V-shaped cross-lines, when the Rule 184 is used.

IV. DISCUSSION AND CONCLUSION

With the CA approach to express the primary structure of SARS-CoV RNA genome, the characteristic of this kind of novel viruses is magnified and visualized in a two-dimensional map. By analyzing the images thus generated for all the 121 coronavirus isolates obtained from the Genbank, a remarkable fingerprint for the SARS-CoV is revealed. It should be pointed out that, only in the region from 3040 to 5439nt near 5-terminal, does the V-shaped cross-lines pattern appear without exception according to our statistical analysis. In such a region the numbers of repeated character ‘A’ for all the 96 SARS-CoVs are larger than those of repeated character ‘T’; whereas for the 25 non-SARS-CoVs a completely opposite is true, as reflected by Tables 1-2 and Fig.**3**.

Owing to the limitation of the existing diagnosis methods, it is still a big problem to determine the SARS-CoV infection. However, using the VAS (Visually Analyzing Sequence) method described here, we have successfully detected the V-shape cross-lines accompanied by the ratio of repeated character ‘A’ to ‘T ’ in all the known SARS-CoV genomes. Accordingly, using the automated method proposed here can help rapidly and reliably diagnose SARS coronaviruses for both basic research in laboratories and practical application in clinics. Further work in analyzing the entire virus family and genome is under way in our laboratory.

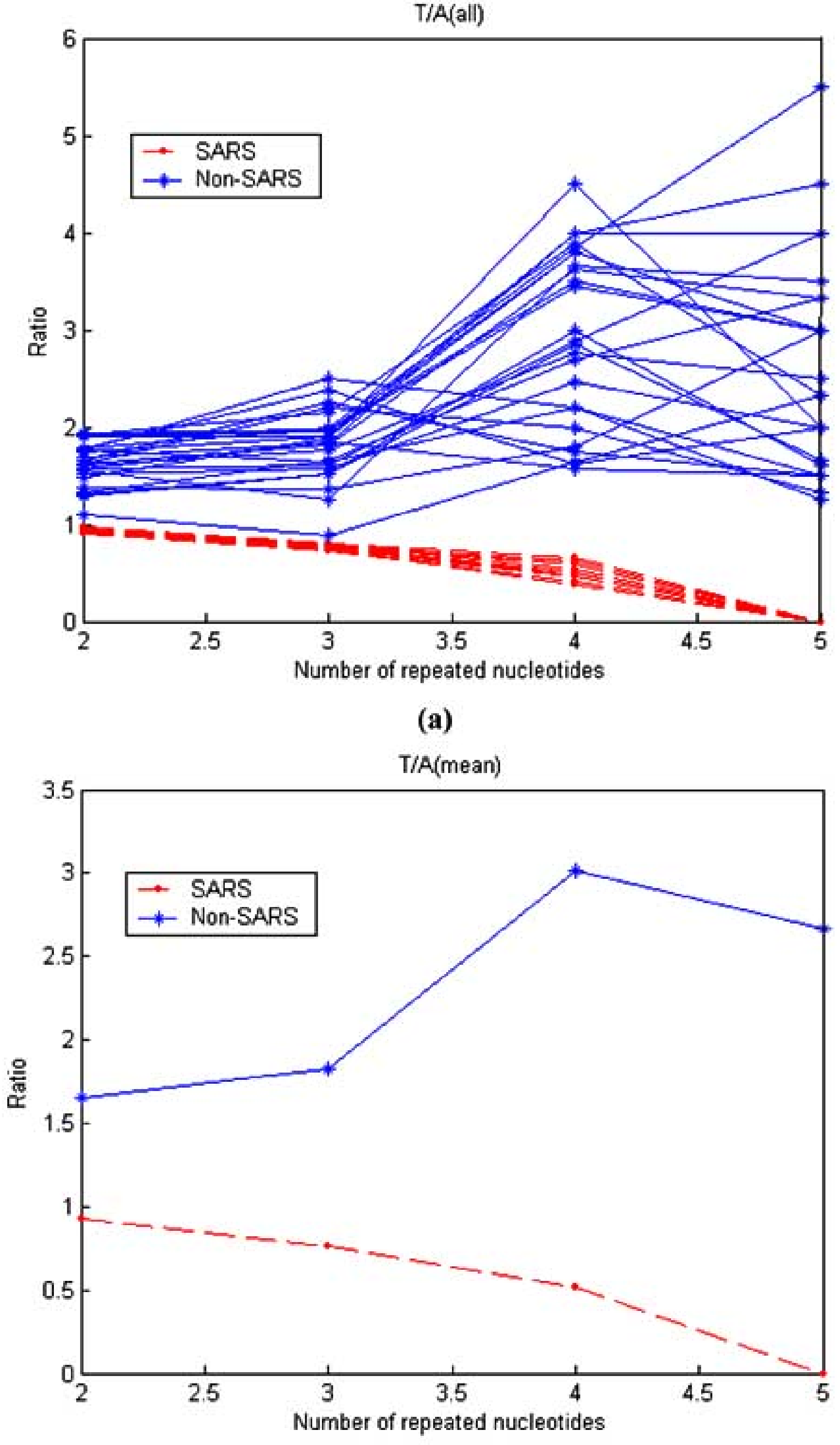
V. ACKNOWLEDGMENTS

Valuable discussions with Dr. Zhongdong Qiao are gratefully acknowledged. We are also indebted to Professor Lianrong Wang for helping us search the virus sequences, and to Dr. Gang Chen, Dr. Chunxia Lei, Mr. Mingqing Xu, Dr. Junxia Tang, and Mr. Junwei Meng for many helpful advices. We would also like to take this opportunity to express our gratitude to the two anonymous referees whose constructive comments are very helpful for strengthening the presentation of this study.

The software to generate the results reported here is available under requirement, and will appear at http://www.pami.sjtu.edu.cn/people/wm.

ABBREVIATIONS

SARS	=	Severe acute respiratory syndrome
CA	=	Cellular automata.



(a)

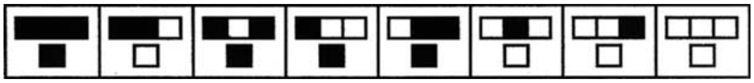
(b)

Fig. (3). The characteristic of the given region (3040~5439nt near 5-terminal) for the 96 SARS-CoVs sequences (Bj01, Bj02, …) and the 25 non-SARS-CoVs sequences (Murine, Bovine, …). (a) Each line corresponds to one virus sequence. X-axis is the number of repeated nucleotides. For example, X= 2 represents two repeated nucleotides, ‘AA’ or ‘TT’. Y-axis is the ratio of occurrence numbers of repeated character ‘T’ to those of repeated character ‘A’, denoted by R2, R3, R4 respectively. R2 = (number of TT) / (number of AA), R3 = (number of TTT) / (number of AAA), R4 = (number of TTTT) / (number of AAAA). All the lines for the SARS are below 1 (red lines); all the lines for the non-SARS are above 1 (blue lines). (b)The upper line (in blue) is the average value of all the ratios of the 25 non-SARS in the given region. The lower line (in red) is the average value of all the ratios of the 96 SARS in the same region.

CA: The cellular automata (CA) approach provides simple discrete deterministic mathematical models for physical, biological and computational systems. Although its construction is quite simple, the CA approach is capable in dealing with complicated behaviors and generating complex patterns with universal features. An "elementary" CA typically consists of a sequence of sites carrying values 0 or 1 arranged on a line, along with a simple rule for transforming that sequence into a new one. The value at each site evolves deterministically with time according to the set of rules involving the values of its nearest neighbors. A rule might go like this: If a site in a given position is flanked by the sites of its opposite value, reverse its value when assigning the corresponding site in the next line; if not, keep it the same. By automatically applying the rule on each line as it evolves (thus the term automata), the computer builds up a pattern of remarkable complexity [11,12]. In general, the sites of a cellular automaton may be arranged on any regular lattice, and each site may take on any discrete set of values. This paper concentrates on the case of "elementary" cellular automata in one dimension with binary values at each site and shows that despite their simple construction, such systems can exhibit complicated behaviors.

In this paper, the initial line is the coded RNA sequence as described in the text. The rule used is '184', which will be explained in Box 2. The generated images for SARS and non-SARS are obviously different (See Fig.1 and Fig.2) after zoomed out, indicating these two kinds of sequences have different nucleotide-compositions. The readers are referred to [11,12] for the whole theoretical foundation, the details of CA as well as the concept of Rule 184.

Box 1



Rule 184: The above 8-subschemes are actually a set of modified local Rules 184 for a CA [11,12]. Here the black and white dots may be viewed as "0" and "1" , respectively. Each of the eight possible sets of values for a site and its nearest neighbors appear on the upper line, while the lower line gives the value to be taken by the central site on the next step. These rules are applied synchronously to each site at every time step. For example, if the current line has such a sequence with 3 pixels that the middle pixel is white and the other pixels are black, then the pixel corresponding to the middle point in the next row will be drawn as a black point, just as illustrated by the above 3rd subscheme. Thus, for example, the sequence 010110110 becomes -0101101- after one time step according to the rule illustrated above (the two end sites depend on unspecified values, and they are set as 0 for the computation of next time step). Rules may be interpreted as Boolean operations on the values of the three sites in each neighborhood. Although the rules are simple in themselves, and the CA approach is easy to be constructed, yet they are capable of describing complicated behaviors and potentially amenable to a precise mathematical analysis.

Box 2

REFERENCES

[1] Chou, K. C.; Wei, D. Q.; Zhong, W. Z. *Biochem Biophys Res Comm.*, **2003**, *308*, 148-151 (Erratum: ibid., **2003**, Vol. *310*, 675).

[2] Chou, K. C. *Current Medicinal Chemistry*, **2004**, *11*, 2105-2134.

[3] Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1111-1122.

[4] Du, Q. S.; Wang, S. Q.; Wei, D. Q.; Zhu, Y.; Guo, H.; Sirois, S.; Chou, K. C. *Peptides*, **2004**, *25*, 1857-1864.

[5] Ruan, Y.J.; Wei, C.L.; Ee, A.L.; Vega, V.B.; Thoreau, H.; Su, S.T.; Chia, J.M.; Ng, P.; Chiu, K.P.; Lim, L.; Zhang, T.; Peng, C.K.; Lin, E.O.; Lee, N.M.; Yee, S.L.; Ng, L.F.; Chee, R.E.; Stanton, L.W.; Long, P.M.; Liu, E.T. *The Lancet*, **2003**, *361*, 1779-1785.

[6] Ahlquist, P. *Science*, **2002**, *196*, 1270-1273.

[7] Yu, X. J. *Acta Pharmacol. Sin*, **2003**, *24*, 481-488.

[8] Drosten, C.; Gunther, S.; Preiser, W.; van der Werf, S.; Brodt, H. R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A. *N. Engl. J. Med.*, **2003**, *348*, 1967-1976.

[9] Poutanen, S.M.; Low, D.E.; Henry, B.; Finkelstein, S.; Rose, D.; Green, K.; Tellier, R.; Draker, R.; Adachi, D.; Ayers, M.; Chan, A.K.; Skowronski, D.M.; Salit, I.; Simor, A.E.; Slutsky, A.S.;

Doyle, P.W.; Krajden, M.; Petric, M.; Brunham, R.C.; McGeer, A.J. *The New England Journal of Medicine*, **2003**, *348*, 1995-2005.

[10] Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J. A.; Lim, W. *The New England Journal of Medicine*, **2003**, *348*, 1953-1966.

[11] Wolfram, S. *A New Kind of Science,* Wolfram Media Inc. : Champaign, IL, **2002.**

[12] Wolfram, S. *Nature*, **1984**, *311*, 419-424.

[13] Chou, K. C.; Zhang, C. T.; Elrod, D. W. *Journal of Protein Chemistry*, **1996**, *15*, 59-61.

[14] Zhang, C. T.; Chou, K. C. *Amino Acids*, **1996**, *10*, 253-262.

Received: 31 July, 2004 Accepted: 02 September, 2004