J. Microbiol. Biotechnol. (2014), 24(10), 1445–1454
http://dx.doi.org/10.4014/jmb.1404.04024

jmb

Development of a Single Nucleotide Polymorphism DNA Microarray for the Detection and Genotyping of the SARS Coronavirus

Xi Guo[1,2†], Peng Geng[3†], Quan Wang[3], Boyang Cao[1,2], and Bin Liu[1,2*]

[1]*TEDA Institute of Biological Sciences and Biotechnology, Nankai University, TEDA, Tianjin 300457, P. R. China*
[2]*Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, TEDA, Tianjin 300457, P. R. China*
[3]*School of Basic Medical Sciences, Tianjin Medical University, Heping District, Tianjin 300070, P. R. China*

Received: April 14, 2014
Revised: June 16, 2014
Accepted: June 16, 2014

First published online
June 20, 2014

*Corresponding author
Phone: +86-22-66229591;
Fax: +86-22-66229584;
E-mail: liubin1981@nankai.edu.cn

†These authors contributed equally to this work.

pISSN 1017-7825, eISSN 1738-8872

Severe acute respiratory syndrome (SARS), a disease that spread widely in the world during late 2002 to 2004, severely threatened public health. Although there have been no reported infections since 2004, the extremely pathogenic SARS coronavirus (SARS-CoV), as the causative agent of SARS, has recently been identified in animals, showing the potential for the re-emergence of this disease. Previous studies showed that 27 single nucleotide polymorphism (SNP) mutations among the spike (S) gene of this virus are correlated closely with the SARS pathogenicity and epidemicity. We have developed a SNP DNA microarray in order to detect and genotype these SNPs, and to obtain related information on the pathogenicity and epidemicity of a given strain. The microarray was hybridized with PCR products amplified from cDNAs obtained from different SARS-CoV strains. We were able to detect 24 SNPs and determine the type of a given strain. The hybridization profile showed that 19 samples were detected and genotyped correctly by using our microarray, with 100% accuracy. Our microarray provides a novel method for the detection and epidemiological surveillance of SARS-CoV.

Keywords: SARS, SARS-CoV, S gene, SNP, DNA microarray

Introduction

The causative agent of the outbreak of the atypical pneumonia known as severe acute respiratory syndrome (SARS) has been identified as a novel coronavirus (CoV) [4, 16]. Before the global outbreak of SARS, SARS-CoV-like viruses had been detected in game animals that are commonly sold in Chinese markets, as the virus likely replicates in animal reservoirs [5, 11]. Despite the fact that there have been no reported infections since April 2004 [15], the extremely pathogenic virus has recently been identified in animals [1, 18, 27], which demonstrates the potential for the re-emergence of SARS.

Serological tests for SARS-CoV were not generally useful or practical since seroconversion often occurs 2 to 3 weeks after infection [14]. On the other hand, molecular tests were recommended based on the orf1b, polymerase (Pol), and nucleocapsid (N) genes for early SARS diagnosis, due to the high rapidity and accuracy of the test [9, 17, 19, 20, 26]. The spike protein (S protein), one of the major structural proteins of SARS-CoV, is essential for virus interactions with host cell receptors and the subsequent fusion of the viral envelope with the host cell membrane to allow viral entry and infection [2, 3]. The S protein also contains important epitopes that elicit the production of neutralizing antibodies in the host species. Furthermore, mutations in the spike gene (S gene) dramatically affect virulence, pathogenesis, and virus tropism [13].

Kan *et al*. [8] discovered 27 novel single nucleotide polymorphisms (SNPs) located in the S gene of the SARS-CoV, that are related to virus phylogenicity, as well as the strain-dependent clinical and epidemiological records available for this virus. Phylogenetic analysis of these SNPs demonstrated that viruses replicating in palm civets had evolved the ability to infect humans. As such, these SNPs describe microevolutionary events that are related to the

epidemicity of the SARS-CoV. Based on these SNPs, the viruses have been qualitatively classified into four groups: the prototype group with no SNPs, the low pathogenicity group containing 2–7 SNPs, the high pathogenicity group containing 17–22 SNPs, and the epidemic group containing 25–27 SNPs [8].

Oligonucleotide array technology has been successfully applied to gene discovery, the study of gene expression patterns, the detection of mutations and polymorphisms, and the mapping of genomic clones [21, 22]. With the potential to detect thousands of SNPs in parallel, oligonucleotide arrays offer a cost-effective approach for high-throughput polymorphism analysis [24, 25, 28]. A previously developed DNA microarray system targeting six SNPs within the entire SARS-CoV genome was established for viral detection and genotyping [12].

In this work, an allele-specific oligonucleotide (ASO) hybridization-based microarray was developed with the ability to screen 24 SNPs simultaneously, and was used for the detection of S gene polymorphisms in 19 cDNAs obtained from SARS-CoV clinical isolates. Since these mutations are closely related to the pathogenicity and epidemicity of the SARS-CoV, this technique can provide useful information for tracking and monitoring the phases of SARS epidemics, and provide critical information regarding clinical treatment and disease control.

Materials and Methods

cDNA Samples

Nineteen cDNAs from SARS-CoV isolates, which had been sequenced with clear epidemiological record, and five cDNAs from other members of Coronaviridae were used in our study (Table 1).

Table 1. cDNAs used in our study.

No.	Name	Source	Location and date of sampling	Type
1	A002G	Civet	Restaurant, Guangdong, Jan. 2004	SARS-CoV, prototype group
2	A001G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
3	A013G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
4	B012G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
5	B029G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
6	B033G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
7	B040G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
8	C013G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
9	C014G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
10	C017G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
11	C019G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
12	C025G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
13	C028G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
14	C029G	Civet	Animal market, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
15	civet007G	Civet	Restaurant, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
16	civet029G	Civet	Restaurant, Guangdong, Jan. 2004	SARS-CoV, low pathogenicity group
17	Lung[a]	Patient		SARS-CoV, epidemic group
18	BJ01	Patient	Beijing, Mar. 2003	SARS-CoV, epidemic group
19	Sin2500	Patient	Singapore, Jul. 2003	SARS-CoV, epidemic group
20	g41921[b]			Avian infectious bronchitis virus (IBV)
21	g42265[b]			Murine hepatitis virus (MHV)
22	g21584[b]			Bovine coronavirus (BCV)
23	g26854[b]			Human coronavirus (HCV)
24	g26855[b]			Human coronavirus (HCV)

[a]An isolate that is given away by the laboratory of the Chinese Center for Disease Control and Prevention, Beijing, China.

[b]Obtained from Tianjin Medical University.

PCR Amplification

PCR1–4 amplification was performed in a total volume of 50 μl, and consisted of 10 ng of SARS-CoV genomic cDNA, 1× PCR buffer (50 μmol/l KCl, 2.5 μmol/l $MgCl_2$, 10 μmol/l Tris-HCl, pH 8.3), 200 μmol/l of each deoxynucleoside triphosphate (dNTP), 2.5 units of *Taq* DNA polymerase (TaKaRa Biotechnology (Dalian) Co. Ltd., China), and 0.5 μmol/l of each S gene primer (Table 2). Reaction parameters were as follows: 94°C for 5 min; 35 cycles of 94°C for 30 sec, 47°C for 1 min, and 72°C for 1 min; followed by a final extension at 72°C for 5 min. PCR products were visualized by agarose gel electrophoresis to examine the amplified DNA. After amplification, PCR products 1–4 were mixed with an equal amount, purified with the Microcon Centrifugal Filter Devices kit (Millipore Corporation, MA, USA), and eluted in 30 μl of autoclaved water and quantified using a spectrophotometer.

ASO Probe Design and Synthesis

Two pairs of probes were designed for each SNP site; one pair for the prototype group containing no SNPs and the other for the epidemic group containing 25–27 SNPs. Each pair consists of two strands, the sense (+) strand and the antisense (-) strand. SNPs are scored by discrimination between perfect match (PM) probes and single basepair mismatch (MM) probes in hybridization assays with DNA targets. With the appropriate selection of oligonucleotide probes and hybridization conditions, it is possible to discriminate between a target DNA sequence and a variant differing in sequence by a single base, and all the oligonucleotides were designed to position the targeted base as close to the center as possible [6]. All probes were synthesized by AuGCT Biotechnology Corporation (Beijing, China). A nonspecific spacer sequence of 15 dT residues was also included at the 5′ end of each oligonucleotide and all probes were synthesized with a 5′ amino group that facilitates covalent immobilization on the aldehyde glass slide.

Preparation of Microarrays

The probes were dissolved in 50% dimethyl sulfoxide (DMSO) at a concentration of 1.0 μg/μl and printed onto aldehyde-group-modified glass slides (CEL-Corporation) using a Spotarray72 instrument (Perkin-Elmer Corporation, CA, USA). Each probe was spotted in triplicate to eliminate irregular data arising from physical defects in the glass slides. Then the slides were dried for 1 day at room temperature, cross-linked by UV cross-linking (UVP Corporation), and stored at room temperature in the dark.

Labeling

Considering the length of target DNAs, the random primer N4, which could label the target DNAs uniformly and obtain shorter DNA fragments for hybridizing with the probes more easily, was used for labeling as follows: 2 μg of purified PCR products and 2 μmol/l random primer N4 (TaKaRa Biotechnology (Dalian) Co. Ltd., China) were mixed (molar ratio is 1:10), and then denatured at 98°C for 10 min and cooled on ice immediately for 15 min. The mixture was then incubated with 5 units of Klenow DNA polymerase (Jingmei Biotech Co., Ltd., China), 100 μmol/l dATP, dTTP, dGTP, 40 μmol/l dCTP, and 10 μmol/l fluorescent Cy5-dCTP at 37°C for 1 h.

Hybridization

The labeling mixtures were dried in a vacuum and resuspended in 25 μl of hybridization solution (6× SSC (1× SSC is 150 mmol/l NaCl and 15 mmol/l sodium citrate), 5× Denhardt's solution, 0.1 μg/l salmon sperm DNA, and 10% formamide). The hybridization mixture was denatured at 98°C for 10 min and transferred to the hybridization area on the glass slide immediately. The slide was incubated at 36°C overnight in a hybridization chamber under a cover slip for 16 h. After incubation, the slide was rinsed and washed sequentially in washing solution A (1× SSC, 0.2% sodium dodecyl sulfate), washing solution B (0.2× SSC), and washing solution C (75% ethanol) for 2 min each, and then dried at room temperature.

Fluorescence Detection

The slides were scanned with a laser at 635 nm for Cy5 labeling for each isolate, and a laser at 532 nm for Cy3 coupled with oligonucleotide poly$(dT)_{15}$ used for positional reference and printing control using the 4100A biochip scanner (Axon Corporation, CA, USA) with the following parameters: photomultiplier tube gain of 600 and a pixel size of 5 μm. For each spot, pixel intensities within the spot image were summed. The average value of pixel intensities for each spot was calculated and the local background level was subtracted from the sum of the signal intensity.

Table 2. Primers used for the amplification of the SARS-CoV spike gene.

PCR product no.	Regions	Forward primer/Reverse primer (5′–3′)	Product length (bp)
1	nt21607-22751	GGGTTTACTATCCTGATGA/ GACACAACCCATGAAATC	1,145
2	nt24111-25155	GCTGCTCTTCAAATACCT/ AACTAGTCATGCAACAAAG	1,045
3	nt23147-24053	CCGTGATGTTTCTGATTT/ GGCAGCAATCATATCATC	907
4	nt22658-23172	CAAGGGAGATGATGTAAGA/ AATCAGTGAAATCAGAAAC	515

Results

PCR Amplification and Labeling

PCR products 1–4 were obtained by using four pairs of primers. PCR product 1 was 1,145 bp in length covering eight SNPs (nt21907, nt22172, nt22219, nt22222, nt22517, nt22522, nt22570, and nt22721); PCR product 2 was 1,045 bp in length covering two SNPs (nt24730 and nt24978); PCR product 3 was 907 bp in length covering eight SNPs (nt23316, nt23317, nt23330, nt23485, nt23719, nt23785, nt23823, and nt23952); and PCR product 4 was 515 bp in length covering seven SNPs (nt22874, nt22875, nt22906, nt22927, nt22928, nt22930, and nt22951). A DNA smear was generated by 2% agarose gel electrophoresis with a range of 500–700 bp after labeling and was ready for hybridization.

Design of ASO Probes

The length of the ASO probe is the most important factor affecting the stability of the duplex and sequence discrimination. Generally, short hybridization sequences have the advantage of providing greater discrimination between PM probes and MM probes, but at a cost of reduced overall stability, whereas longer hybridization sequences form more stable duplexes, but suffer from reduced discrimination [6]. SNP nt24978 was selected to optimize the probe length for our study. A series of ASO probes targeting nt24978 consisting of 12mers (PM-ATTCAAGAAGAA/MM-ATTCAAAAAGAAAT), 14mers (PM-CATTCAAGAAGAAA/MM-ATTCAAAAAGAAAT), 16mers (PM-ACATTCAAGAAGAAAT/MM-CATTCAAAAAGAAATT), 18mers (PM-AACATTCAAGAAGAAATT/MM-ACATTCAAAAAGAAATTG), and 20mers (PM-CAACATTCAAGAAGAAATTG/MM-AACATTCAAAAAGAAATTGA) were designed and tested. Fig. 1A shows the positions of the above probes spotted on the glass slide. After hybridization with isolate civet007G and calculating the signal intensities for each probe, we found that the fluorescence intensities increased with the length of the probes (Figs. 1B and 1C). There was nearly no hybridization signal when the probe was too short (12mers and 14mers), whereas nonspecific hybridization signals were observed when the probe was too long (20mers). Consequently, the 16–18mer probes were chosen for all subsequent work, except for some 14–15mer probes that were chosen owing to their high GC contents.

Among the 27 SNPs in the S gene, four are synonymous and 23 are nonsynonymous. Probes targeting two synonymous SNPs (nt22517 and nt22874) were inevitably designed, since both were too close to two other nonsynonymous SNPs (nt22522 and nt22875). In total, 19 groups of 104 ASO probes were designed for 25 SNPs (Table 3). Considering the locations of these SNPs, probes in groups 1, 2, 5, 7, 9, 11, 12, 13, 14, 15, 16, 17, 18, and 19 were designed to target only single SNPs, and each group contained two pairs of probes. Probes in groups 3, 4, 6, and 10 targeted two SNPs each, and each group contained four pairs of probes. Finally, probes in group 8 targeted three SNPs, and this group contained eight pairs of probes.

In addition, we also designed a pair of probes that targeted a conserved region of the S gene as a positive control (P1 and P2) and a probe containing poly(dT)$_{15}$ as the negative control (N). Another probe containing poly(dT)$_{15}$ labeled with the fluorescent Cy3-dUTP at the 3′ end was designed and used for positional reference and printing control. A schematic diagram of their relative positions on the microarray is shown in Fig. 2.

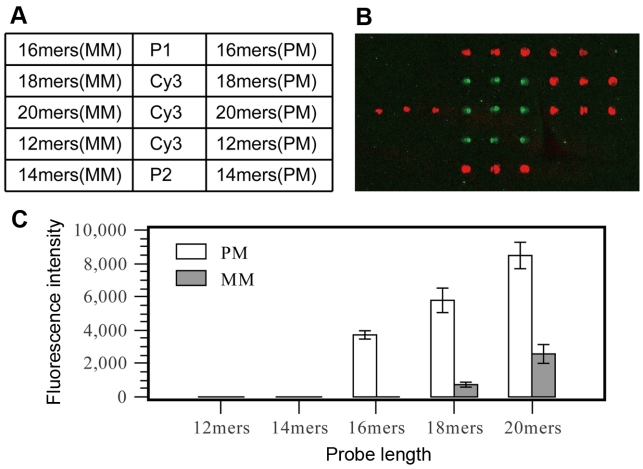

Fig. 1. Signal intensities of the probes with different lengths for SNP nt24978 detection.

(**A**) Layout of the probes with different lengths. P1 and P2 represent the positive control; Cy3 represents the oligonucleotide poly(dT)$_{15}$ labeled with the fluorescent Cy3-dUTP at the 3′ end for positional reference and printing control. (**B**) Fluorescence images obtained from civet007G. (**C**) Signal intensity ratios of PM to MM. Error bars represent the standard deviation.

Probe Specificity

Six isolates, three of BJ01, Sin2500, and Lung representing the epidemic group and the rest of civet007G, A001G, and B029G representing the low pathogenicity group, were selected for the verification of our probes. The scanned fluorescence images of BJ01 and civet007G are shown in

Table 3. Probes used in our microarray.

Group	Number	Probe sequence (5′-3′)	SNP position	Type
1	S001	CGTTTG***G***CAACCCT	nt21721	epi[a]
	S002	AGGGTTG***C***CAAACG	nt21721	epi
	S003	TACGTTTG***A***CAACCCT	nt21721	pro[b], low[c], high[d]
	S004	AGGGTTG***T***CAAACGTA	nt21721	pro, low, high
2	S005	TTTCTTTG***C***TGTTTCTA	nt21907	high, epi
	S006	AGAAACA***G***CAAAGAAA	nt21907	high, epi
	S007	TTTCTTTG***T***TGTTTCTAA	nt21907	pro, low,
	S008	TTAGAAACA***A***CAAAGAAA	nt21907	pro, low,
3	S009	CTCAAG***A***CA***T***TTGGG	nt22219, nt22222	epi
	S010	CCCAA***A***TG***T***CTTGAG	nt22219, nt22222	epi
	S011	TCAAG***G***CA***T***TTGGG	nt22219, nt22222	
	S012	CCAA***A***TG***C***CTTGAA	nt22219, nt22222	
	S013	CTCAAG***G***CA***C***TTGG	nt22219, nt22222	pro
	S014	GCCCCAA***G***TG***C***CTTGAG	nt22219, nt22222	pro
	S015	TCAAG***A***CA***C***TTGGGG	nt22219, nt22222	low, high
	S016	CCCCAA***G***TG***T***CTTGA	nt22219, nt22222	low, high
4	S017	GGAGAG***A***AAAAA***A***AATTTC	nt22517, nt22522	epi
	S018	GAAATT***T***TTTTT***T***CTCTCC	nt22517, nt22522	epi
	S019	GAGAG***G***AAAAA***A***AATTTC	nt22517, nt22522	epi
	S020	GAAATT***T***TTTT***C***CTCTC	nt22517, nt22522	epi
	S021	GAGAG***G***AAAA***G***AATTTC	nt22517, nt22522	pro, low, high
	S022	GAAATT***C***TTTT***C***CTCTC	nt22517, nt22522	pro, low, high
	S023	GGAGAG***A***AAAA***G***AATTT	nt22517, nt22522	pro, low, high
	S024	AAATT***C***TTTT***T***CTCTCC	nt22517, nt22522	pro, low, high
5	S025	CTCAACAT***T***TTTTTCAAC	nt22570	high, epi
	S026	GTTGAAAAAA***A***ATGTTGAG	nt22570	high, epi
	S027	CTCAACAT***C***TTTTTCAAC	nt22570	pro, low, high
	S028	GTTGAAAAA***G***ATGTTGAG	nt22570	pro, low, high
6	S029	CTTTCTC***C******C***CTGATGG	nt22874, nt22875	low, high, epi
	S030	CCATCAG***G******G***GAGAAAG	nt22874, nt22875	low, high, epi
	S031	CTTTCTC***T******C***CTGATGG	nt22874, nt22875	low, high, epi
	S032	CCATCAG***G******A***GAGAAAG	nt22874, nt22875	low, high, epi
	S033	CCTTTCTC***T******T***CTGATGG	nt22874, nt22875	pro, low
	S034	CCATCAG***A******A***GAGAAAGG	nt22874, nt22875	pro, low
	S035	CTTTCTC***C******T***CTGATGG	nt22874, nt22875	pro, low
	S036	CCATCAG***A******G***GAGAAAG	nt22874, nt22875	pro, low
7	S037	CACCTGCTC***T***TAATTG	nt22906	high, epi
	S038	CAATTA***A***GAGCAGGTG	nt22906	high, epi
	S039	CACCTGCTC***C***TAATTG	nt22906	pro, low
	S040	CAATTA***G***GAGCAGGTG	nt22906	pro, low
8	S041	CCATTAA***A******T***G***A***TTATGG	nt22927, nt22928, nt22930	high, epi
	S042	CCATAA***T***C***A******T***TAATGG	nt22927, nt22928, nt22930	high, epi
	S043	CCATTAA***G******T***G***A***TTATGG	nt22927, nt22928, nt22930	
	S044	CCATAA***T***C***A******C***TTAATGG	nt22927, nt22928, nt22930	

Table 3. Continued.

Group	Number	Probe sequence (5′-3′)	SNP position	Type
	S045	CCATTAA***AA***G***A***TTATGG	nt22927, nt22928, nt22930	high
	S046	CCATAA***T***C***TT***TTAATGG	nt22927, nt22928, nt22930	high
	S047	CCATTAA***AA***G***G***TTATGG	nt22927, nt22928, nt22930	
	S048	CCATAA***CC******TT***TTAATGG	nt22927, nt22928, nt22930	
	S049	CCATTAA***AT***G***G***TTATGG	nt22927, nt22928, nt22930	low
	S050	CCATAA***CC******AT***TTAATGG	nt22927, nt22928, nt22930	low
	S051	CCATTAA***GA***GA***T***TATGG	nt22927, nt22928, nt22930	
	S052	CCATAA***T***C***TC***TTAATGGC	nt22927, nt22928, nt22930	
	S053	CCATTAA***GA***GG***T***TATGG	nt22927, nt22928, nt22930	pro, low
	S054	CCATAA***CC******TC***TTAATGG	nt22927, nt22928, nt22930	pro, low
	S055	CCATTAA***GT***G***G***TTATGG	nt22927, nt22928, nt22930	
	S056	CCATAA***CC******AC***TTAATGG	nt22927, nt22928, nt22930	
9	S057	ACCACTA***C***TGGCATTG	nt22951	high, epi
	S058	CAATGCCA***G***TAGTGGT	nt22951	high, epi
	S059	ACCACTA***G***TGGCATTG	nt22951	pro, low, high
	S060	CAATGCCA***C***TAGTGGT	nt22951	pro, low, high
10	S061	TTCTACA***GC***AATTCATG	nt23316, nt23317	high, epi
	S062	CATGAATT***GC***TGTAGAA	nt23316, nt23317	high, epi
	S063	TCTACA***GT***AATTCATGC	nt23316, nt23317	
	S064	GCATGAATT***AC***TGTAGA	nt23316, nt23317	
	S065	TTTCTACA***TC***AATTCATG	nt23316, nt23317	
	S066	CATGAATT***GA***TGTAGAAA	nt23316, nt23317	
	S067	TTCTACA***TT***AATTCATGC	nt23316, nt23317	pro, low
	S068	GCATGAATT***AA***TGTAGAA	nt23316, nt23317	pro, low
11	S069	CATGCAGA***T***CAACTCAC	nt23330	high, epi
	S070	GTGAGTTG***A***TCTGCATG	nt23330	high, epi
	S071	CATGCAGA***A***CAACTCA	nt23330	pro, low
	S072	TGAGTTG***T***TCTGCATG	nt23330	pro, low
12	S073	CAGTTTCTT***T***ATTACGTA	nt23485	high, epi
	S074	TACGTAATA***A***AGAAACTG	nt23485	high, epi
	S075	CAGTTTCTT***C***ATTACGT	nt23485	pro, low
	S076	ACGTAAT***G***AAGAAACTG	nt23485	pro, low
13	S077	GCTTTTGCA***C***ACAAC	nt23719	high, epi
	S078	GTTGT***G***TGCAAAAGC	nt23719	high, epi
	S079	GCTTTTGCA***G***ACAACT	nt23719	pro, low
	S080	AGTTGT***C***TGCAAAAGC	nt23719	pro, low
14	S081	GTGTTCG***C***TCAAGTCA	nt23785	high, epi
	S082	TGACTTGA***G***CGAACAC	nt23785	high, epi
	S083	GTGTTCG***T***TCAAGTCA	nt23785	pro, low
	S084	TGACTTGA***A***CGAACAC	nt23785	pro, low
15	S085	ACTTTGAAA***T***ATTTTGG	nt23823	epi
	S086	CCAAAAT***A***TTTCAAAGT	nt23823	epi
	S087	ACTTTGAAA***G***ATTTTGG	nt23823	pro, low, high
	S088	CCAAAAT***C***TTTCAAAGT	nt23823	pro, low, high

Table 3. Continued.

Group	Number	Probe sequence (5′-3′)	SNP position	Type
16	S089	ATATGGC***G***AATGCCTA	nt23952	low, high, epi
	S090	AGGCATT***C***GCCATAT	nt23952	low, high, epi
	S091	AATATGGC***C***AATGCC	nt23952	pro
	S092	GGCATT***G***GCCATATT	nt23952	pro
17	S093	TTGTGTTTA***A***TGGCAC	nt24730	low, high, epi
	S094	GTGCCA***T***TAAACACAA	nt24730	low, high, epi
	S095	TTGTGTTTA***G***TGGCAC	nt24730	pro
	S096	GTGCCA***C***TAAACACAA	nt24730	pro
18	S097	ACATTCAA***A***AAGAAATTG	nt24978	high, epi
	S098	CAATTTCTT***T***TTGAATGT	nt24978	high, epi
	S099	AACATTCAA***G***AAGAAATT	nt24978	pro, low
	S100	AATTTCTT***C***TTGAATGTT	nt24978	pro, low
19	S101	TGGTATTAA***C***ATTACAAA	nt22172	low, high, epi
	S102	TTTGTAATGTTAATACCA	nt22172	low, high, epi
	S103	TGGTATTAA***G***ATTACAAA	nt22172	pro, low
	S104	TTTGTAAT***C***TTAATACCA	nt22172	pro, low
	P1	AAGGGAGATGATGTAAG	Positive control	
	P2	CTTACATCATCTCCCTT	Positive control	
	N	15dT	Negative control	

The italic and bold bases refer to the SNP sites.

[a]Epidemic group; [b]prototype group; [c]low pathogenicity group; [d]high pathogenicity group.

S001	S002	P1	S003	S004	S037	S038	P1	S039	S040	S073	S074	P1	S075	S076
S005	S006	P2	S007	S008	S041	S042	P2	S043	S044	S077	S078	P2	S079	S080
S009	S010	P1	S011	S012	S045	S046	P1	S047	S048	S081	S082	P1	S083	S084
S013	S014	P2	S015	S016	S049	S050	P2	S051	S052	S085	S086	P2	S087	S088
S017	S018	Cy3	S019	S020	S053	S054	Cy3	S055	S056	S089	S090	Cy3	S091	S092
S021	S022	P1	S023	S024	S057	S058	P1	S059	S060	S093	S094	P1	S095	S096
S025	S026	P2	S027	S028	S061	S062	P2	S063	S064	S097	S098	P2	S099	S100
S029	S030	P1	S031	S032	S065	S066	P1	S067	S068	S101	S102	P1	S103	S104
S033	S034	P2	S035	S036	S069	S070	P2	S071	S072	N	N	P2	N	N

Fig. 2. Layout of the surface-bound ASO probes on the microarray.

S001 to S104 represent all the ASOs covering 25 SNPs located in the S gene; P1 and P2 represent the positive controls; N represents the negative control; Cy3 represents the 3′ Cy3-dUTP-labeled oligonucleotide poly(dT)$_{15}$ for positional reference and printing control.

Figs. 3A and 3B, respectively. Since each pair of probes differs by only a single base, it is very difficult to avoid nonspecific signals. However, the nonspecific signals can be identified by their lower signal intensities compared with specific signals, rather than their disappearance [7]. We employed an intensity ratio of 3:1 as the threshold to categorize specific and nonspecific signals. For example, consider the group 15 probes S085-S088 for the detection of nt23823. When hybridized with BJ01, the ratios of signal intensities of PM (S085-086) to MM (S087-088) is approximately 33, indicating that the base at nt23823 in BJ01 is a T, whereas when the probes are hybridized with isolate civet007G, the ratios of the signal intensities of PM (S087-088) to MM (S085-086) is approximately 19,

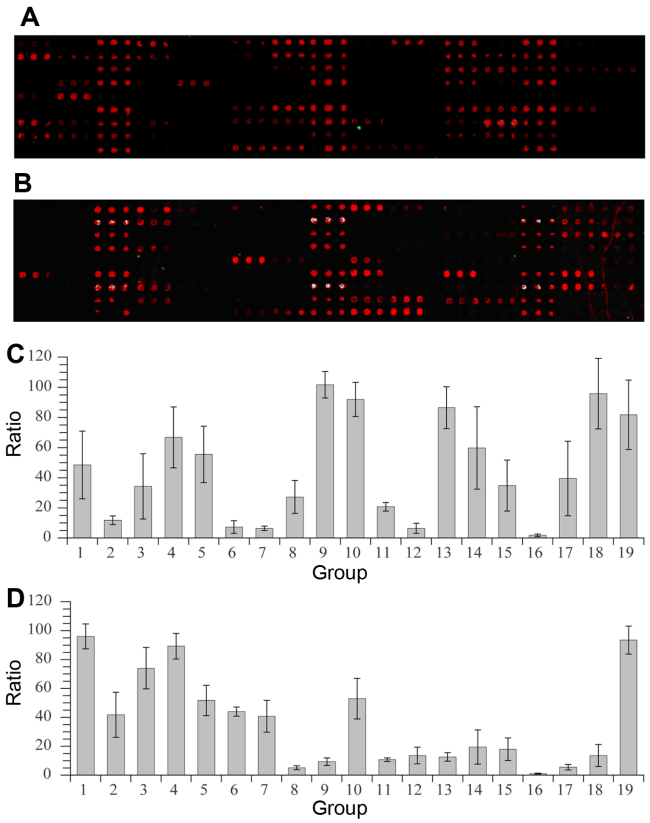

Fig. 3. Fluorescence images of the microarray and the signal intensity ratios for each pair of probes.

(**A**) Fluorescence images obtained from BJ01 representing the epidemic group. (**B**) Fluorescence images obtained from civet007G representing the low pathogenicity group. (**C**) Signal intensity ratios of PM to MM probes in each group obtained from BJ01. Error bars represent the standard deviation. (**D**) Signal intensity ratios of PM to MM probes in each group obtained from civet007G. Error bars represent the standard deviation.

indicating that the base at nt23823 in civet007G is a G. All ASO probes were verified using the six isolates for four times. The intensity ratios of each pair of probes in all 19 groups for BJ01 and civet007G are shown in Figs. 3C and 3D, respectively. According to the hybridization profile and the ratios of fluorescence intensities, 24 of 25 SNPs could be detected correctly by our microarray and the results were confirmed by sequencing.

In addition, five cDNAs from other members of Coronaviridae were employed to test the specificity of our microarray. These viruses included one avian infectious bronchitis virus (IBV), one murine hepatitis virus (MHV), one bovine coronavirus (BCV), and two human coronaviruses (HCV). These isolates failed to generate any positive signals.

Detection of SARS-CoV Samples

A total of 19 cDNAs obtained from clinical isolates or specimens were obtained to validate our microarray system. The hybridization profile showed that one represented the prototype group, 15 belonged to the low pathogenicity group, three matched the epidemic group, and none fell into the high pathogenicity group. The results generated by our microarray were confirmed by sequence analysis and were in accordance with the clinical and epidemiological records of these strains, showing the accuracy is 100%.

Discussion

As shown in Figs. 3C and 3D, fluctuations among different groups of probes were observed. This may be due to the different hybridization kinetics attributed to the GC content and melting temperature among these probes. The locations of the SNPs limit our choices of designing probes with similar hybridization kinetics. The probe length and the formamide concentration in the hybridization buffer could be optimized, which might increase the stability of the microarray system. We could not detect the SNP nt23952, despite testing a series of probes. The probes S089-092 that were originally designed for nt23952 were analyzed using Primer Premier 5.0 and were found to form stable hairpins (ΔG = -2.7, ΔG refers to Gibbs free energy) and dimers among themselves (ΔG = -7.2), which might result in the failure to hybridize with target DNAs.

Techniques newly developed over these years, especially second-generation sequencing or PacBio RS sequencing, may be used to address such an important event as the re-emergence of SARS-CoV. However, they require more expensive equipments, have higher testing costs, and are especially time consuming for library construction and sequencing, which could not fulfill the need of early epidemiological investigation [10, 23]. Oligonucleotide microarrays afford a higher throughput compared with conventional Sanger sequencing-based methods, because they can be used to analyze multiple genetic regions or many SNPs in one region simultaneously. In addition, by using the methods based on the microarray, the detection work could be accomplished in one day, which is more rapid and at lower costs compared with those using new sequencing technologies.

Long *et al.* [12] previously developed a DNA microarray system targeting six SNPs distributed across the "whole genome" for the detection and genotyping of SARS-CoV, which was confirmed using 20 isolates. Comparatively, our

study focused on the 25 SNPs in the "single S gene" encoding the S protein, which is essential for viral infection and as epitopes of SARS-CoV. Hence, our microarray is more focused and relevant, and results from our assay could be more closely related to the epidemicity of this virus.

As SARS-CoV has not circulated among humans since 2004, it may be expected that new strains would have different SNPs in the spike gene after 10 years. More isolates with broad representation could help further evaluate our microarray. The microarray has its limitation in that any probe must be designed relying on a known sequence. However, once new SNPs emerged could be identified by sequencing, probes based on them can be designed and complemented to our microarray quickly and easily. Overall, our microarray provides a potential tool to identify the pathogenicity and epidemicity of a given SARS-CoV strain, in order to improve the prevention and control of SARS.

Acknowledgments

We thank the Chinese Center for Disease Control and Prevention for providing the cDNAs of SARS-CoV isolates. This work was supported by the National Key Program for Infectious Diseases of China (2013ZX10004221), the National 973 Program of China Grant (2011CB504900), the National High Technology Research and Development Program of China (863 Program) (2011AA100901-2), International Science & Technology Cooperation Program of China (2012DFG31680 and 2013DFR30640), National Natural Science Foundation of China (NSFC) Program (31371259, 31270003 and 31270016) and Research Project of Chinese Ministry of Education (NO.113015A).

References

1. Balboni A, Battilani M, Prosperi S. 2012. The SARS-like coronaviruses: the role of bats and evolutionary relationships with SARS coronavirus. *New Microbiol*. **35:** 1-16.
2. Bosch BJ, van der Zee R, de Haan CA, Rottier PJ. 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77:** 8801-8811.
3. Colman PM, Lawrence MC. 2003. The structural biology of type I viral membrane fusion. *Nat. Rev. Mol. Cell Biol*. **4:** 309-319.
4. Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, Becker S, *et al*. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348:** 1967-1976.
5. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, *et al*. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302:** 276-278.
6. Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM. 1994. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res*. **22:** 5456-5465.
7. Guo Z, Gatterman MS, Hood L, Hansen JA, Petersdorf EW. 2002. Oligonucleotide arrays for high-throughput SNPs detection in the MHC class I genes: HLA-B as a model system. *Genome Res*. **12:** 447-457.
8. Kan B, Wang M, Jing H, Xu H, Jiang X, Yan M, *et al*. 2005. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. *J. Virol.* **79:** 11892-11900.
9. Keightley MC, Sillekens P, Schippers W, Rinaldo C, George KS. 2005. Real-time NASBA detection of SARS-associated coronavirus and comparison with real-time reverse transcription-PCR. *J. Med. Virol.* **77:** 602-608.
10. Kircher M, Kelso J. 2010. High-throughput DNA sequencing – concepts and limitations. *Bioessays* **32:** 524-536.
11. Liang G, Chen Q, Xu J, Liu Y, Lim W, Peiris JS, *et al*. 2004. Laboratory diagnosis of four recent sporadic cases of community-acquried SARS, Guangdong Province, China. *Emerg. Infect. Dis*. **10:** 1774-1781.
12. Long WH, Xiao HS, Gu XM, Zhang QH, Yang HJ, Zhao GP, *et al*. 2004. A universal microarray for detection of SARS coronavirus. *J. Virol. Methods* **121:** 57-63.
13. Lu L, Manopo I, Leung BP, Chng HH, Ling AE, Chee LL, *et al*. 2004. Immunological characterization of the spike protein of the severe acute respiratory syndrome coronavirus. *J. Clin. Microbiol.* **42:** 1570-1576.
14. Mahony JB, Richardson S. 2005. Molecular diagnosis of severe acute respiratory syndrome: the state of the art. *J. Mol. Diagn.* **7:** 551-559.
15. McBride R, Fielding BC. 2012. The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. *Viruses* **4:** 2902-2923.
16. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, *et al*. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361:** 1319-1325.
17. Peiris JS, Poon LL. 2008. Detection of SARS coronavirus in humans and animals by conventional and quantitative (real time) reverse transcription polymerase chain reactions. *Methods Mol. Biol*. **454:** 61-72.
18. Peiris JS, Poon LL. 2011. Detection of SARS coronavirus. *Methods Mol. Biol*. **665:** 369-382.
19. Poon LL, Chan KH, Wong OK, Yam WC, Yuen KY, Guan Y, *et al*. 2003. Early diagnosis of SARS coronavirus infection by real time RT-PCR. *J. Clin. Virol.* **28:** 233-238.

20. Poon LL, Leung CS, Tashiro M, Chan KH, Wong BW, Yuen KY, *et al*. 2004. Rapid detection of the severe acute respiratory syndrome (SARS) coronavirus by a loop-mediated isothermal amplification assay. *Clin. Chem.* **50:** 1050-1052.
21. Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with complementary DNA microarray. *Science* **270:** 467-470.
22. Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW. 1998. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol*. **16:** 301-306.
23. Shallom SJ, Tae H, Sarmento L, Preston D, McIver L, Franck C, *et al*. 2012. Comparison of genome diversity of *Brucella* spp. field isolates using universal bio-signature detection array and whole genome sequencing reveals limitations of current diagnostic methods. *Gene* **509:** 142-148.
24. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K. 1979. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res*. **6:** 3543-3557.
25. Wallace RB, Johnson MJ, Hirose T, Miyake T, Kawashima EH, Itakura K. 1981. The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit beta-globin DNA. *Nucleic Acids Res*. **9:** 879-894.
26. Wang B, Potter SJ, Lin Y, Cunningham AL, Dwyer DE, Su Y, *et al*. 2005. Rapid and sensitive detection of severe acute respiratory syndrome coronavirus by rolling circle amplification. *J. Clin. Microbiol*. **43:** 2339-2344.
27. Yip CW, Hon CC, Shi M, Lam TT, Chow KY, Zeng F, *et al*. 2009. Phylogenetic perspectives on the epidemiology and origins of SARS and SARS-like coronaviruses. *Infect. Genet. Evol.* **9:** 1185-1196.
28. Zhang Y, Coyne MY, Will SG, Levenson CH, Kawasaki ES. 1991. Single-base mutational analysis of cancer and genetic diseases using membrane bound modified oligonucleotides. *Nucleic Acids Res.* **19:** 3929-3933.