Available online at www.sciencedirect.com
ScienceDirect
Journal of Microbiological Methods 68 (2007) 128–136
Journal of Microbiological Methods
www.elsevier.com/locate/jmicmeth

Development of a DNA microarray to identify the *Streptococcus pneumoniae* serotypes contained in the 23-valent pneumococcal polysaccharide vaccine and closely related serotypes

Quan Wang [a,b], Min Wang [a,b], Fanrong Kong [c], Gwendolyn L. Gilbert [c], Boyang Cao [a,b], Lei Wang [a,b,d,*], Lu Feng [a,b,d,*]

[a] *TEDA School of Biological Sciences and Biotechnology, Nankai University, 23# HongDa Street, TEDA, Tianjin 300457, China*
[b] *Tianjin Key Laboratory for Microbial Functional Genomics, TEDA College, Nankai University, 23# HongDa Street, TEDA, Tianjin 300457, China*
[c] *Centre for Infectious Diseases and Microbiology, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, 2145 New South Wales, Australia*
[d] *Tianjin Biochip Corporation, 23# HongDa Street, TEDA, Tianjin 300457, China*

Received 12 April 2006; received in revised form 4 July 2006; accepted 5 July 2006
Available online 14 August 2006

Abstract

Streptococcus pneumoniae is a major worldwide human pathogen. This investigation has developed a reliable and accurate DNA microarray method for inter-species differentiation of *S. pneumoniae* and intra-species differentiation of the 23 groups of *S. pneumoniae* including serotypes represented in the 23-valent pneumococcal vaccine and the other 20 closely related serotypes. In addition to 16S rDNA probes, serotype- or serogroup-specific probes targeting the capsular polysaccharide synthesis (*cps*) genes, *wzy* or *capA* were generated. We adopted a two-step multiplex PCR to improve the sensitivity of detection to a level of 10^5 cfu/ml in pure culture or 50 ng DNA. A total of 169 isolates (from China, Australia, Canada and New Zealand) including 147 belonging to 23-valent vaccine and closely related serotypes of *S. pneumoniae*, 11 belonging to other serotypes and 11 of different species commonly isolated from respiratory tract were tested to verify the method. The DNA microarray method developed provides a sensitive means to rapidly identify the members of the most common *S. pneumoniae* serotypes in patients and to monitor their distribution in different patient groups and geographic locations. Such information is needed for disease surveillance and to monitor vaccine efficacy.

Keywords: *Streptococcus pneumoniae*; 23-valent; Microarray; Detection; Multiplex PCR

1. Introduction

Streptococcus pneumoniae is a major worldwide human pathogen associated with community-acquired pneumonia, septicemia, meningitis, and otitis media (Brown and Lerner, 1998). Among the 90 capsular serotypes of *S. pneumoniae* identified thus far, 23 are responsible for up to 90% of invasive disease in some geographic areas (Klein, 1981) and are therefore included in the 23-valent polysaccharide pneumococcal vaccine. These serotypes are 1 to 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F and 33F (Smit et al., 1977).

The serotypes of *S. pneumoniae* most commonly isolated from patients with invasive pneumococcal disease vary in different age groups and geographic locations (Klein, 1981; Sniadack et al., 1995). Traditional serotyping requires the use of a large panel of expensive anti-sera, and it is subjective, cross-reactions are common and a significant number of isolates are non-typeable (Barker et al., 1999; Henrichsen, 1999). On the other hand, molecular typing has many advantages and several PCR-based methods have been developed recently (Brito et al., 2003; Lawrence et al., 2003). However, it is difficult to quantify

* Corresponding authors. Lu Feng, is to be contacted at Tel.: +86 22 66229592; fax: +86 22 66229596. Lei Wang, Tel.: +86 22 66229588; fax: +86 22 66229596.
E-mail addresses: wanglei@nankai.edu.cn (L. Wang), fenglu63@nankai.edu.cn (L. Feng).

0167-7012/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2006.07.001

PCR products and differentiate those of similar size, in a multiplex PCR. DNA microarray is a rapid, reliable and sensitive method for the molecular typing of pathogenic microorganisms, which overcomes some of the disadvantages of multiplex PCR alone (Chizhikov et al., 2001; Wilson et al., 2002).

The polysaccharide capsules of Gram-positive bacteria, external to the cell wall, provide resistance to phagocytosis. In *S. pneumoniae* the capsule is the most important virulence factor and, therefore, the target of pneumococcal vaccines (AlonsoDeVelasco et al., 1995). Genes responsible for capsular synthesis are normally located in a gene cluster (the *cps* gene cluster), which maps between *dexB* and *aliA* in the pneumococcal chromosome (Llull et al., 1999). There are four relatively conserved genes — *cpsA* (*wzg*)–*cpsB* (*wzh*)–*cpsC* (*wzd*)–*cpsD* (*wze*) at the 5′ end of the *cps* gene cluster (Jiang et al., 2001), also included in the *cps* gene cluster are nucleotide sugar biosynthesis genes, sugar transferase genes and, except in serotype 3, capsule assembly genes *wzx* (polysaccharide flippase gene) and *wzy* (polysaccharide polymerase gene). The *cps* gene clusters of all 90 *S. pneumoniae* serotypes have been sequenced by Sanger Institute (http://www.sanger.ac.uk/Projects/S_pneumoniae/CPS/), and *wzx* and *wzy* of the 89 serotypes (all except serotype 3) have been annotated and analyzed in our previous report (Kong et al., 2005). The assembly genes were highly variable and contain serotype- or serogroup-specific regions (Wang and Reeves, 1998). For serotype 3, serotype-specific regions of *capA*, *capB* and *capC*, in the *cps* gene cluster, can be used in place of *wzx* or *wzy* (Garcia et al., 1999). In this study, we used these serotype- or serogroup-specific sequences as probes in a microarray, and developed a method based on the microarray to identify 23-valent pneumococcal vaccine, and closely related *S. pneumoniae* serotypes.

2. Materials and methods

2.1. Bacterial isolates

All isolates used in this study, including 158 of *S. pneumoniae* and 11 of other bacterial species, are shown in Table 1. All the *S. pneumoniae* isolates had been well-characterized and were provided by the Centre for Infectious Diseases and Microbiology (CIDM), Sydney, Australia. The other bacterial species were identified in the First Center Hospital of Tianjin using the in-house procedures based on traditional biochemical methods and confirmed by 16S rDNA sequencing using an ABI 3730 automated DNA sequencer.

2.2. Genomic DNA extraction

S. pneumoniae isolates were cultured on fresh rabbit blood agar at 37 °C in the presence of 5% CO_2 for 24 h. Genomic DNA was extracted using a genomic DNA extraction kit (Beijing Tianweishidai Company, Beijing, China). DNA template used for the sensitivity test of detection from pure culture was extracted as follows: a sweep of culture was sampled with a disposable loop and resuspended in 0.2 ml digestion buffer (10 mM Tris/HCl [pH 8.0], 0.45% Triton X-100 and 0.45% Tween 20) in a 1.5 ml tube. The tube was heated at 100 °C for 10 min, cooled on ice for 10 min, and centrifuged for 2 min at 14,000 rpm to pellet the cell debris. A total of 5 μl of each supernatant containing extracted DNA was used as template for PCR.

2.3. Oligonucleotide primers and probes

Sequences of the targeted genes were retrieved from the Sanger Institute (http://www.sanger.ac.uk/Projects/S_pneumoniae/CPS/) and GenBank. The oligonucleotide primers and probes were listed in Table 2. Two primers and three or four probes were designed based on the targeted genes for each of the 23 *S. pneumoniae* groups as defined in Table 2, except groups 18 and 19, which share the same reverse primer, using program PRIMEGENS (Xu et al., 2002) and OligoArray 2.0 (Rouillard et al., 2003), respectively. The primer pair used to amplify 16S rDNA sequences was reported previously (Zheng et al., 2005). Probes based on 16S rDNA sequences specific to *S. pneumoniae* and all bacteria, respectively, were used as positive controls. To enhance hybridization yield, each probe was 5′-amino-modified and 10 poly Ts were added at the 5′ end. A probe containing 40 poly Ts was used as the negative control.

2.4. DNA array preparation

A 384-well plate was prepared by adding 10 μl of oligonucleotide probes (1 μg/μl in 50% DMSO) into each well. The probes were spotted on to aldehyde group modified glass slides (CEL Corporation, USA) using the SpotArray 7.2 (Perkin–Elmer Corporation, CA, USA). Each probe was spotted in duplicates on the slide. Printed slides were dried for 24 h at room temperature, crosslinked by Ultraviolet Crosslinker (UVP Corporation, CA, USA) and stored at room temperature. The slides were scanned at 532 nm to check the quality of the spotting. Each slide consisted of four microarrays framed with a 12 μl Geneframe (Beijing Capital Biochip Corporation, Beijing, China), which constituted individual reaction chambers. A schematic diagram of the probe positions on the microarray is shown in Fig. 1.

2.5. Two-step multiplex PCR and labeling of the targeted genes

The two-step multiplex PCR, in which the first step was to amplify the targeted genes and the second to label the PCR products generated, was carried out in two batches; one targeting groups 1 to 4, 6 to 10, 16, 23 and the other targeting groups 5, 11 to 15 and 17 to 22. For the first step PCR, the PCR mixture contained 1 × PCR buffer (KCl 50 mM, Tris-HCl 10 mM, [pH 8.3]); 0.5 mM $MgCl_2$; 100 μM concentration of dNTP; 0.03 U *Taq* DNA polymerase; 0.1 μM each of two primers based on 16S rDNA sequence; 0.2 μM each of primers based on each of targeted genes and 50 to 100 ng DNA template in a final volume of 50 μl. The PCR reaction was performed as follows: 94 °C for 5 min; 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min, 36 cycles; 72 °C for 5 min. The PCR products were purified with the MICROCON Centrifugal Filter Devices kit (MILLIPORE Corporation, MA, USA). For the second step PCR,

Table 1
Bacterial stains used in this study

Bacteria	Serotype	Number of strains of each geographic origin	Total number
S. pneumoniae	1	1[a,*], 1[b], 1[c,*], 4[d]	7
	2	1[a], 1[b], 1[d]	3
	3	1[a,*], 1[c], 5[d,**]	7
	4	2[a], 1[c,*], 3[d,*]	6
	5	1[a], 1[b*], 2[d]	4
	6A	1[c]	1
	6B	1[a], 4[d]	5
	7A	1[b]	1
	7F	1[a], 1[b], 1[c], 4[d]	7
	8	1[a], 5[d,*]	6
	9L	1[e,*]	1
	9N	1[b], 2[d,*]	3
	9A	1[e]	1
	9V	1[a], 1[b], 2[d], 1[e]	5
	10A	1[l], 1[b], 1[c], 3[d]	6
	10B	1[e]	1
	10C	1[e]	1
	10F	1[a]	1
	11A	1[a], 1[c], 3[d]	5
	11D	1[f]	1
	11F	1[f]	1
	12A	1[e]	1
	12B	1[f]	1
	12F	1[a], 1[b], 1[c], 3[d]	6
	44	1[f]	1
	46	1[e]	1
	14	1[a,*], 1[c], 2[d]	4
	15B	1[a], 1[b]	2
	15C	1[a]	1
	17F	1[a], 1[b], 1[c], 2[d]	5
	18A	1[d,*]	1
	18B	1[d]	1
	18C	2[a,*], 1[b,*], 1[c], 7[d]	11
	18F	1[e]	1
	19A	1[a], 1[b], 1[c], 3[d]	6
	19F	1[a], 1[c], 3[d]	5
	20	1[a], 1[b], 1[c,*], 4[d]	7
	22A	1[b]	1
	22F	1[a], 1[b], 1[c], 3[d]	6
	23F	1[a], 1[c,*], 3[d]	5
	33A	1[e]	1
	33F	1[a], 1[b], 2[d], 2[e]	6
	37	1[e]	1
	7C	1[b]	1
	13	1[b,*]	1
	15A	1[b]	1
	16F	1[e,*]	1
	19C	1[e,*]	1
	23A	1[d]	1
	24A	1[e,*]	1
	38	1[b]	1
	33C	1[e,*]	1
	31	1[b]	1
	36	1[e]	1
S. salivarius	–	1[g]	1
H. influenzae	–	1[h]	1
M. catarrhalis	–	1[i]	1
P. aeruginosa	–	1[j]	1
K. pneumoniae	–	1[h,*]	1
S. aureus	–	1[h]	1
P. maltophilia	–	1[i]	1
S. marcescens	–	1[j]	1
P. vulgaris	–	1[i]	1
L. pneumophila	–	1[k]	1
S. epidermidis	–	1[h]	1

* One strain was selected for the double-blind test.
** Two strains were selected for the double-blind test.
[a] Isolated from Queensland, Australia.
[b] Isolated from Canada.
[c] Isolated from normally sterile sites by CIDM during the period from January 1999 to June 2001.
[d] Isolated from New Zealand.
[e] From National Institute for the Control of Pharmaceutical and Biological Products, Beijing, China.
[f] From Statens Serum Institut, Denmark.
[g] From China Veterinary Culture Collection Center (CVCC), China.
[h] From National Center for Medical Culture Collection (CMCC), China.
[i] From American Type Culture Collection (ATCC).
[j] From Institute of Microbiology, Chinese Academy of Sciences (IMAS), China.
[k] From Academy of Military Medical Sciences (AMMS), China.

the same PCR mixture was used except that 0.3 μl of 25 nM Cy3-dUTP was included and only the reverse primers were used, and 20 μl of the purified PCR products from the first step was used as the template. The PCR conditions were the same as for the first step PCR except that 41 cycles were performed. 2 μl of each PCR product was run on an agarose gel to examine the amplified DNA.

2.6. Hybridization process

The labeled PCR products were dried for about 3 h at 65 °C, diluted in 11 μl hybridization buffer (25% formamide, 0.1% SDS, 6 × SSPE) and then hybridized to the prepared microarray at 40 °C for 16 h. After the hybridization, the Geneframe was removed and the slide was washed with washing buffer A (1 × SSC and 0.1% SDS) for 3 min, followed by a wash with 0.05 × SSC for 3 min, and finally a wash with 95% ethanol for 90 s. The slides were dried under a gentle airstream before they were scanned.

2.7. Fluorescence scanning and automated data analysis

The hybridized microarray was scanned with a laser at 532 nm using the GenePix personal 4100A (Axon Instruments, CA, USA) and the signals were calculated with GenePix Pro 6.0. The data were analyzed and the results were reported using a program developed in-house.

3. Results

3.1. Twenty-three groups of S. pneumoniae

We compared the *cps* gene clusters of the 23-valent vaccine serotypes with those of the other serotypes, and found that these 23 serotypes were indistinguishable from other 20 serotypes,

most of which belonged to the same serogroup (Table 2). Therefore, we divided these 23-valent vaccine serotypes and 20 closely related serotypes (a total of 43 serotypes) as 23 groups (Table 2). This microarray was used to identify the most common *S. pneumoniae* serotypes in patients including 23-valent vaccine serotypes and 20 closely related serotypes.

3.2. Targeted genes

At the beginning of this study, we used both *wzx* and *wzy* as the targeted genes for each of 23 groups except group 3, for which *capA*, *capB*, *capC* were used. After hybridization to all the probes designed in these regions, we found probes of *wzy* and *capA* were more specific (data not shown). Our previous study also showed that *wzy* of *S. pneumoniae* was more heterogeneous than *wzx* (Kong et al., 2005). Therefore, we use *wzy* as the targeted gene in this study for all 23 groups except group 3, for which *capA* was used.

3.3. Multiplex PCR

The multiplex PCR primers were designed based on the serotype- or serogroup-specific DNA sequences and bacterial 16S rDNA. *S. pneumoniae* isolates belonging to any of the 23 groups produced two bands of the amplified DNA-16S rDNA and the targeted sequence in both steps of the multiplex PCR reaction (Fig. 2), and *S. pneumoniae* isolates belonging to other serotypes produced only one band of 16S rDNA sequence. Therefore the primers were proved to be specific for their respective targeted gene segments. As few as 10^6 cfu/ml strains could be detected using the multiplex PCR, and for some strains, it could be 10^5 cfu/ml (Supplementary Fig. 1).

3.4. Specificity of the probes

We designed and screened 238 oligonucleotide probes including specific probes based on *wzx* and *wzy* of 23 groups except *capA*, *capB* and *capC* of group 3 and probes for positive and negative controls. Finally, 93 probes (including specific probes based on *wzy* and *capA*, positive and negative controls) were selected for use in this study (Table 2). A total of 147 isolates belonging to 23 groups (among isolates from China, Australia, Canada and New Zealand (Table 1)) were tested against the probes on the microarray and all consistently hybridized to three or four of their corresponding probes and could be successfully differentiated into the appropriate group of serotypes (Fig. 3). In contrast, none of the isolates belonging to one of 11 other serotypes of *S. pneumoniae* bound to the serotype/serogroup-specific probes on the microarray. Three bacterial species — *Streptococcus salivarius*, *Moraxella catarrhalis* and *Staphylococcus epidermidis* — which are frequently isolated from normal respiratory flora and eight — *Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Staphylococcus aureus*, *Pseudomonas maltophilia*, *Serratia marcescens*, *Proteus vulgaris*, *Legionella pneumophila* (Table 1), which are potential respiratory pathogens, failed to hybridize to the *S. pneumoniae* probes.

3.5. A double-blind test

A double-blind test was performed to verify the specificity of the microarray. A total of 23 clinical isolates were randomly selected from all the strains we stored (Table 1) to hybridize to the microarray without knowledge of their identity at the time of testing. After the hybridizations, we found two isolates belonged to the *S. pneumoniae* group 1, three belonged to group 3, two belonged to group 4, one belonged to group 5, one belonged to group 8, two belonged to group 9, one belonged to group 14, three belonged to group 17, one belonged to group 20 and one belonged to group 22. We also found 5 isolates of *S. pneumoniae* not belonging to any of the targeted 23 group, and one isolate only hybridized to the positive control probes based on the conserved region of bacterial 16S rDNA. All the results were consistent with conventional serotyping method.

3.6. Sensitivity of detection with DNA

Serial dilutions of *S. pneumoniae* strain G1834 (serotype 8) DNA containing 1 μg, 500 ng, 100 ng, 50 ng, 10 ng, 1 ng, 0 ng were used to test the sensitivity of the microarray (Supplementary Fig. 2). The results indicate that the assay generated strong positive signals at DNA levels of 10 ng or above. We selected 50 ng as the most suitable amount of DNA for the microarray detection. At this level, all *S. pneumoniae* isolates in 43 targeted serotypes could be detected correctly.

3.7. Sensitivity of detection from pure cultures

DNA extracts from 10-fold serial dilutions of *S. pneumoniae* strain G2189 (serotype 4) in the range of 10^3 to 10^8 cfu/ml, were amplified, labeled by two-step multiplex PCR, and hybridized to the microarray. In the microarray assay, all the specific probes could detect DNA extracts from the culture containing 10^5 cfu/ml or more, and some probes could detect DNA extracts from the culture containing 10^4 cfu/ml, which could not be detected by multiplex PCR (Figs. 4, 5).

4. Discussion

A microarray has been used previously to detect *S. pneumoniae* based on the targeted genes of *gyrB* and *parE* (Roth et al., 2004), but the use of a microarray to both identify and type (within a small group of serotypes, in some cases) the most common pneumococcal serotypes has not been described before. The method based on microarray is relatively rapid, sensitive and suitable for high throughput testing. Previous study showed that the microarray based on 16S rDNA sequences has some disadvantages to discriminate species. First, the specific region of 16S rDNA may be prone to frequent mutations, and a false negative result will be produced by a strain with random mutations in this region. An additional challenge is that strains of different species may share almost the same sequence identity in this probe region, thus it is not sufficiently specific to distinguish these species (Fox et al., 1992; Warsen et al., 2004).

Table 2
Oligonucleotide primers and probes used in this study

Group	Serotype	Forward and reverse primer (5′–3′)	Probe (5′–3′)
1	1	ATTAAGTGGGCAAGTTTCA	TTGTATCAGCTAGTCGTAATTTACAGATTCTGATAT
			TATACGGGATGCTGATATATCTGTATATGATACCTA
		CAATAACCAATGTTCCCAAT	ATATGGTTTTACTCTTGTGGGAGAGGGATAT
			ATTACATAATATGGGCAATTATTTCGAAGGTCG
2	2	CGTTATGGACTGGCTGATG	GTCAACGTATTGGAACTCTTAGAAATTGGGAAAAT
			TTCCTTATTGATACTGGGGTTATTGGCTTATTTAT
		GATAAATAAGCCAATAACCC	TATACTACAGCTGATTCAACAGTGACAAGTTTCC
			ACTATAAAGGCTATTATTGGGTCAGGATTTGG
3	3	GAGGTTCCGATACTGGTTG	TGATCCTAGAATTGGATCATACTATAATAACCCTAG
			AAATTGTTATTTACGAACCTACTATTGAGTGTGATA
		GTTCATACGATTCGCTACA	ATGAAATCTGATTCTGATAATTTTCGTTCTAGTGCTG
			CAGCTGTGGTGCAATCTAATAAAACAAGAAAAGATTA
4	4	CAGTTAGAGCGGCGAATA	TACCTAATTTTGATAATGATCCTAACAGGAAGTCGT
			GAATATGGGATGACATTTCTACGCACTATCCTA
		CACCACCATAGTAACCAAAG	CGTTGGTTTTAGTATACACACAATCTTTAGGGA
			AATGATCCTAACAGGAAGTCGTAAAATCCAGG
5	5	TCCTGTTGGGAGTAGTGTCT	TTATTCTCAACTTGTCTCTATATAATCGCAAGTG
			TTTGCTAAAGAAATTTCAGGACATGAATTAACAG
		TGCGGAAACGATGAGAAG	GTTTTCTCAATTATGGTTATATAGTAGGGACTGCT
			TGAATATCGACCATCTAGTCTGACTTTGGAC
6	6A, 6B	ATTATTCCAGCGACTACAC	TTTATTTGTTATAGATCCGATACGACGTAACAA
			ATGAGATGCTTGCAAAATTAGGATTTGACAATA
		ACTCTCCATTCTAAGCCTT	TCTTTCTTTTAAAATTTGTATTAGGGCGCTCC
			ATTTTATCAGTTACATCACTCGTTATATGGGAGGTTC
7	7A, 7F	GCTTATTGTATGGTCCCTCC	TTATTTGGCTATTCAACAGGAAACTATTCAACTAGT
			GGTCGGTTAAGTTCAATTGTTCAAATAATCTCAACT
		TCTGCCAAACATCTCCATA	TAATATCTATCACTTGCATAGTGGTGGTCG
			ACGTAGCTGCTTTGTCAGAGTTAAAGATTA
8	8	TATGTTTACTTTACGAGTTGGT	TTGTTGAAAATCTTGAACTTAGATGGATAGGATA
			TTTGATTCGTTTTATTCCTTCTTGATGAGTAATG
		AGACTGATAAGATAAGTAGCCA	GGGTCAATATGTTGATAAAGAAATTGTTTGGGATC
			ACGCAGACTAGAACAGCTCTACTAGTATCTATAGTA
9	9L, 9N	ATGGCGACTTTATTTGCT	CGACTTTATTTGCTTTGCTTTCTACTTTAGTTAAGC
			AGCATAAAGTAACACTGATTGATATTCTAGTTGTTC
		ATCCCAATAGACGTATGAAA	AAACGTTGTACCTTAACTATGTTGGAGGTC
			TTCCCAAAGTTCTTTTACAGCATAAAGTAACACTG
10	9A, 9V	TAAGGTAACGCTGATTGAC	GTCAGACAGTGAATCTTAACTATATTGGAGGACA
			GTTAACTCCGCACACAATACCTTACTAGATATA
		ATGTAGCCTCCATCGTAAT	TGCGTATGTTAGGTTTAAAGGAAAATATAGGAAAACC
			CGGGTATTTTATTTGTAGTGGTTCTTCTAAGTCTATTT
11	10A, 10B,	GTGCGACTACGACCCTTAT	TTGGCGTTTTAGTTTTGTTACAATGACTTATTATGT
			ACCATATGGTATTTGTTGCCTGTAGGTATC
	10C, 10F	TTGAACCCATAGATAACAG	GCTATATGTTGGAATAGTGCTGATGGTTTTATTTGT
12	11A, 11D, 11F	TAAGTGGGTCAATCGGTGTT	TTAAGAAAACTATTCTGGAAAAGAGGAAGGAAGATA
			GGTTTGTTTGGCTTTCAATTTAAACTTCCG
		ATTTATTCCAACTTCTCCC	TGTTGTAACTTTTATTGCATCTGTTGTGATGTTAG
13	12A, 12B, 12F, 44, 46	TCCTAGAAACAGCATTATTAC	AATAAGAACTGCTATGAAATTTAAGTTCCTTTTCTCATTT
		CAAATCCAAACCTTTTGTAA	TCTTTATCTTATTATGCCATGACAGCTTATATGATTAATG
			GACGTATTTCAGATTCTCAGCGTGGAATTTTATT
			AGTTCCTTTTCTCATTTATTGGAGGCCTGT
14	14	TTTGTATGGTGCTATGGAC	TAGCATTTGCAGTAGTCAATGAGATTAATTCAAAC
			TTCTTCTAGAGGTTTTGGTACTAAGCAAATCTAG
		TAGGGATTCCTCATACACTCT	CGGTTTATTCTATATACAAAGAGGCTCCAATGTA
15	15B, 15C	TATGTTCAAAGAGGCGCTAA	ATGATTGTAGCGTTTTATCCAACTATACCTCTTATA
			ATCTCTTAAAGCTATTAATGGGCTGGGAAT
		TCTGATTCCTGCTCCAAGT	TCAAGAACAGGTAGTAATGTGACACGTTTTATAG
			GTAGTATTGTTTGGAAGAAGCTTATTAGGTTGG
16	17F	TTGTATGGCTAGTTCTTCTA	TAGTACCTCTTGTCAAATACATACTTACCCCT
			ACCCATGCTCATAATAACTTTCTTCAAGTTTT
		CAAATCCTAAATATCCTCC	CCATATATTAGGGAATTTTCCTGGTTGGATTTG
			TCTACGTTGAGTTTTCTAATTGTAGTTACTGGT
17	18A, 18B, 18C, 18F	CTTTAGGCAGGGAGGACTT	TCGTTCTTTTGAGTGTATTAATTATCATACCAGTA
			TAATATTAATTCGATGGCTAGAACAGATTTATGG
		AATCCTACAAATCCTATCTCAA	TTTGAAGTACTACATTGAGATAGGATTTGTAGGAT
			TGAATATCAATGGTCTTACAGGGACAATGG

Table 2 (*continued*)

Group	Serotype	Forward and reverse primer (5′–3′)	Probe (5′–3′)
18	19A	CGCCAGTTGTCATGTCTG	CTACCTTTTACTATTCTTTATTCTTATCTGGACGT
			GGTTCGATTCAGCATTTTAATCAGTATATTCAAAA
		GTTGTAAACGAGTATTGCTC	TTGTAGCTATTATGAATATTTTGGGTAATCTTGGCT
			GCTCTAAAGATTTACTGCTCTTAGTTCCATTCTT
19	19F	TTCAACGACTAGGACGCTAT	AATTCATTTAGAATTTCGGACACTAGGAGTTACTG
			ATTGGTGCCTATATTAAATATATTAGGGGAAATGGGC
		GTTGTAAACGAGTATTGCTC	TTTCGGACACTAGGAGTTACTGTAGGAAATGTTTATA
			TATCTAGGAGGCTCAATTCAGCATTTTAATCAGTAT
20	20	TATCAGGAATACGCCAATC	CGTTATTATACAGGAGCATATGATAAACAACACCAA
			ACTTCCTACTCCATCAAGAACTATTTATATGTTT
		TGTGGTACGGTAGTGTTTC	ACTATTCACAATCGTTATGTGTTTATCGGTATTTCT
			CTGGGTCTGAATTTGTATCTCGAATAATACTACATT
21	22A, 22F	AGATGCTGGTAGAACGCCT	TCCCCAATACTATTAGAAAAGTGGAGTAGTAT
			GAAGCATACACAATGATTGCGAATACTGAT
		AAAGGCACCATATAGAAGGG	TTTATCTAGGACAACAACCTCCGTTTATAGGATTC
			TGCTAGGTTCAATTATCTTATGGCTATCAACTTTAT
22	23F	ATTTCTAGCTTATCCACCTT	TTTAAGAAGGATATCATATTACTTATTGCACAGCAT
			GCTAAACATCAGATTGATAGTTTTGTATTATGGTTA
		CCCTTCGTCGTATTTCCA	TCAGCAGAAAATATGACGCTAAACATCAAAT
			CCACCTTACCGATAAAGCTAATAAATAATCTCAGTG
23	33A, 33F, 37	AGCACCTATGTGGAGTTTG	GGGGTTTATATAATAGGTTTGATTTTCCTATATGTTT
			CCATTGAACAACATATGGATAGAACTGTTAGATATAA
		GTAATACGAAGCCAACATT	CAGAAAAGTCCTATTTGGGGATTAGGTATTTCA
			CAATATTATGGGATAATAAATCTGGAACTGGTTCAG
Positive control	16S rDNA	AGAGTTTGATC(A/C)TGGCTCAG	GCTAGGTGTTAGACCCTTTCCGGGGTTTA [a]
			GAACGAGTGTGAGAGTGGAAAGTTCACACTGTG [b]
		CCGTCAATTC(A/C)TTT(A/G)AGTTT	CCAGACTCCTACGGGAGGCAGCAGTG(A)GGGAAT [c]
Negative control			TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

[a] The positive control probe for *S. pneumoniae*, which was spotted in line 1 and line 9 in the microarray (Fig. 1).
[b] The positive control probe for *S. pneumoniae*, which was spotted in line 5 and line 13 in the microarray (Fig. 1).
[c] Positive control probes for all bacteria.

The *cps* gene cluster encodes proteins that appear to be specific for synthesis, transfer and assembly of specific polysaccharides that define serotypes. The assembly genes were highly variable and virtually serotype-specific (Wang and Reeves, 1998). In this study we used the capsular serotype-specific genes, *wzy* or *capA*, as the targeted genes, which allowed fast, reliable differentiation of *S. pneumoniae* from other *Streptococcus* species and intra-species differentiation of 23 small groups or individual pneumococcal serotypes (Fig. 3). After hybridization with 238 oligonucleotide probes, non-specific probes were discarded leaving 93 that were suitable for use.

The two-step PCR format-amplification of targeted genes, followed by labeling of the targets using only reverse primers — not only enhances amplification efficiency and sensitivity of the assay but also generates single-stranded PCR products for hybridization, without the need for denaturation. Previous experience has shown that denaturation is often incomplete and, therefore, subsequent hybridization is inefficient.

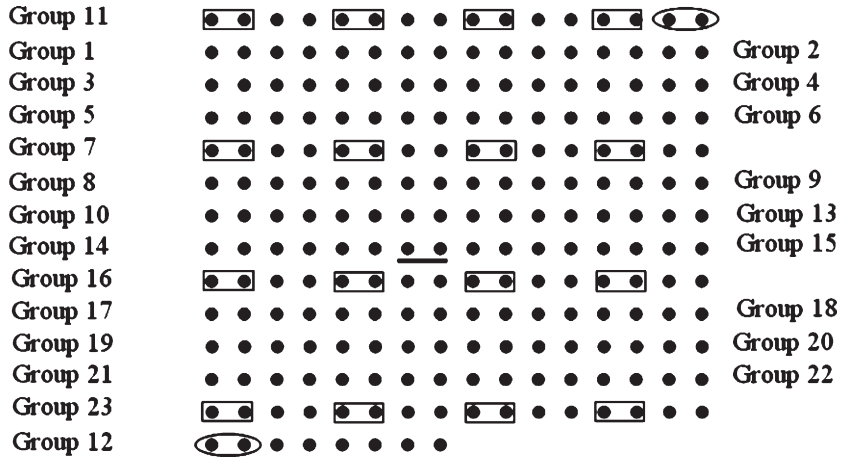

Fig. 1. A schematic diagram of the probe positions on the microarray. Each probe was spotted in duplicates on the slide. Dots in rectangles are *S. pneumoniae* specific 16S rDNA probes. Dots in ellipses are probes based on the conserved region of 16S rDNA in all bacteria. Dots underlined are negative control probes.

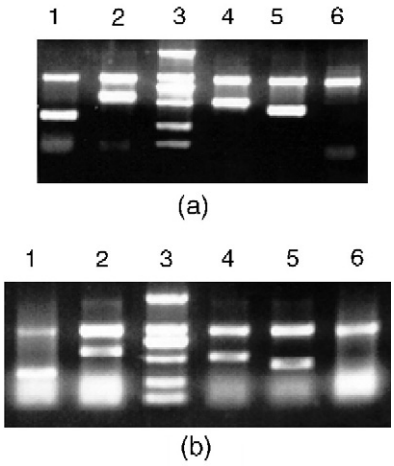

Fig. 2. Agarose gel electrophoresis of the two-step multiplex PCR products of selected *S. pneumoniae* serotypes. Lane 1: serotype 12F; lane 2: serotype 23F; lane 3: DNA markers of 100 bp, 250 bp, 500 bp, 750 bp, 1 kb, and 2 kb; lane 4: serotype 22F; lane 5: serotype 19A; lane 6: serotype 5. Two bands were produced from all PCR reactions including the PCR products of 16S rDNA (907 bp) and the targeted sequences specific to the individual serotype (there are 2 bands of 907 bp and 891 bp in lane 6). (a) After the first step. (b) After the second step.

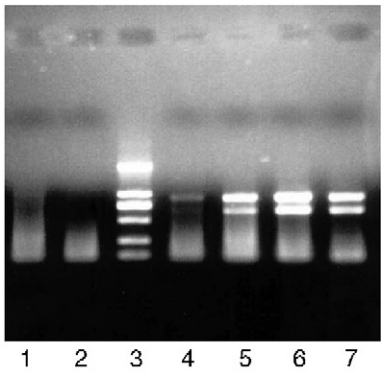

Fig. 4. Detection sensitivity of *S. pneumoniae* serotype 4 strain G2189 in pure culture by two-step multiplex PCR. Lane 1: 10^3 cfu/ml; lane 2: 10^4 cfu/ml; lane 3: DNA markers of 100 bp, 250 bp, 500 bp, 750 bp, 1 kb, and 2 kb; lane 4: 10^5 cfu/ml; lane 5: 10^6 cfu/ml; lane 6: 10^7 cfu/ml; lane 7: 10^8 cfu/ml.

We tested 169 bacterial isolates, including 147 strains belonging to the 43 targeted serotypes of *S. pneumoniae*, 11 belonging to other serotypes and 11 of different species commonly isolated from the respiratory tract. The microarray method has been proved to be specific for detection and partial serotyping of 43 *S. pneumoniae* serotypes. Those belonging to small groups of cross-reacting serotypes can be easily distinguished by limited use of anti-sera. In future, we aim to add probes targeting truly serotype-specific sequences from other genes, in order to differentiate members of these small cross-reacting groups.

A recent study showed that *S. pneumoniae* should be at dose of more than 4.60 log(10) cfu to be sufficient to cause pneumonia in rabbits (Yershov et al., 2005), which is very close to the detection sensitivity (as little as 50 ng DNA or 10^5 cfu/ml in pure culture) of the microarray developed in this study. Therefore, the microarray

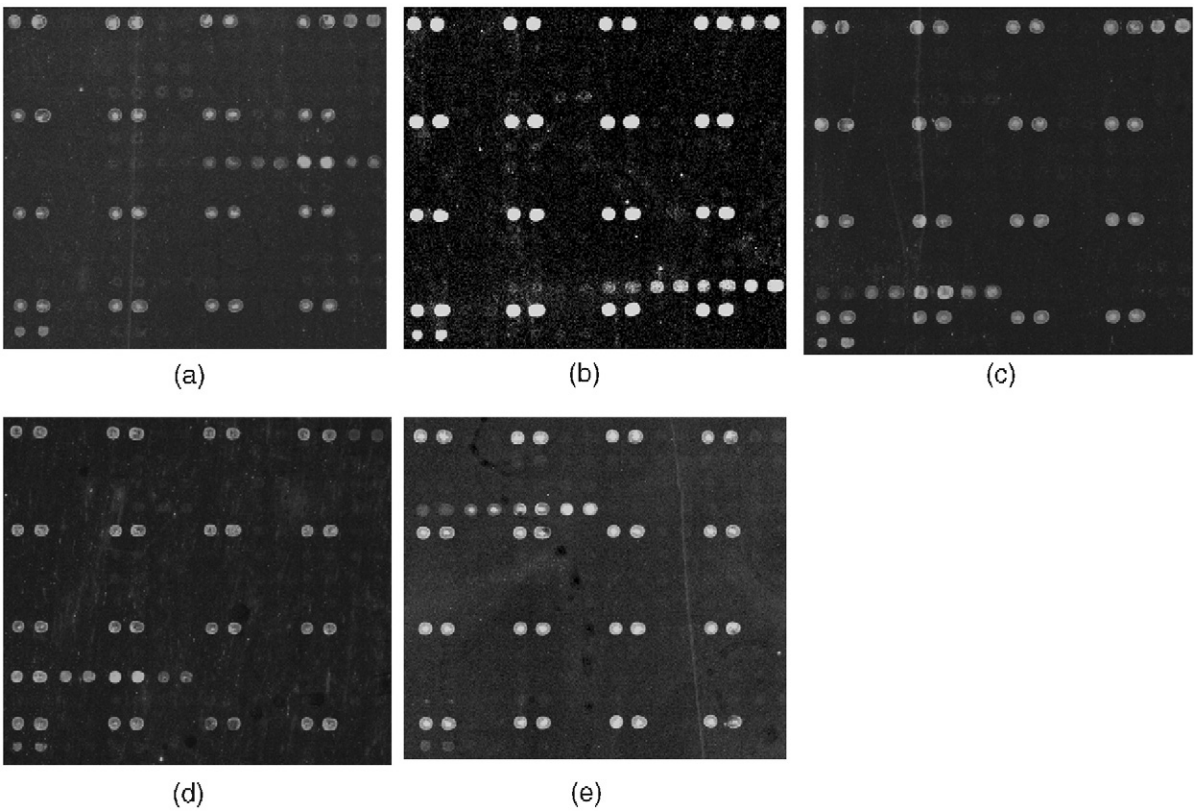

Fig. 3. Hybridization of *S. pneumoniae* to the microarray. (a) Serotype 12F; (b) serotype 23F; (c) serotype 22F; (d) serotype 19F; (e) serotype 5. All experiments were performed at least twice.

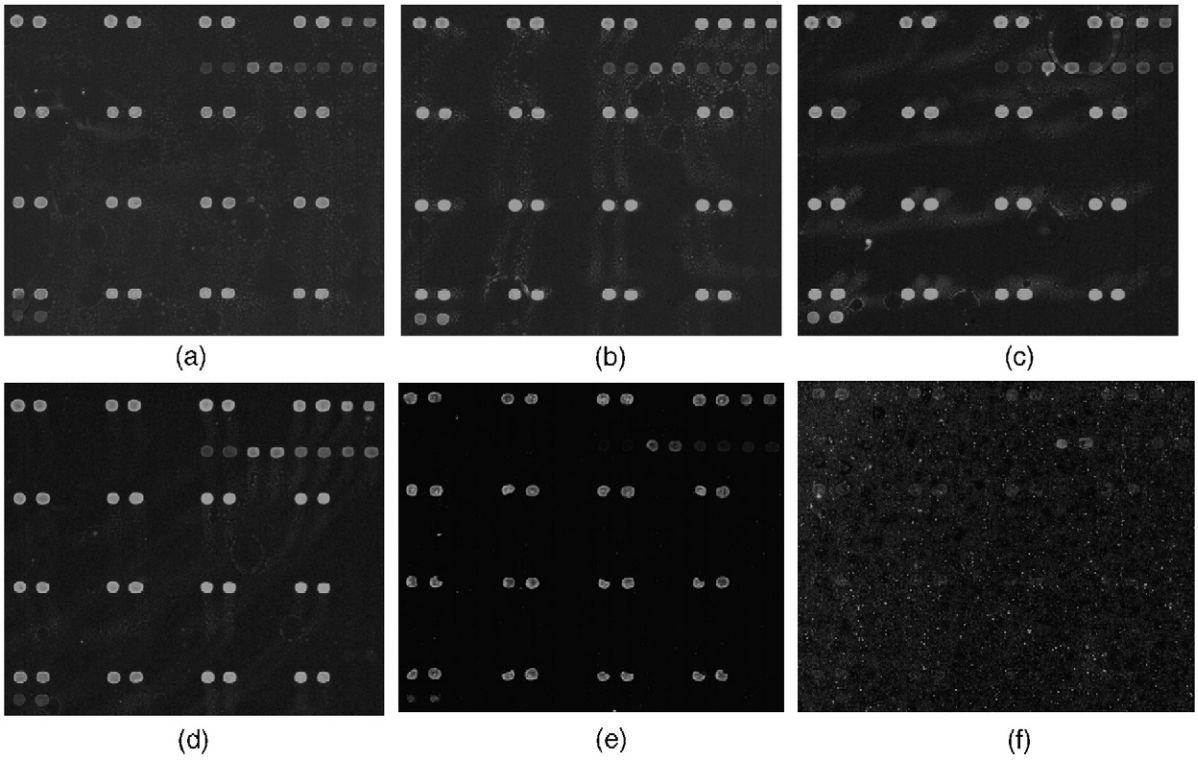

Fig. 5. Detection sensitivity of *S. pneumoniae* serotype 4 strain G2189 in pure culture by microarray. (a) 10^8 cfu/ml; (b) 10^7 cfu/ml; (c) 10^6 cfu/ml; (d) 10^5 cfu/ml; (e) 10^4 cfu/ml; (f) 10^3 cfu/ml. All experiments were performed at least twice.

should be able to detect *S. pneumoniae* directly from clinical specimens of heavily infected sputum from a patient with pneumonia, or after culture enrichment.

With 12 or 13 primer pairs in each multiplex PCR reaction, there was excellent specificity based on more than 500 reactions, but a single reaction with all 24 primer pairs would make the assay simpler, which we will try in future work. Compared to traditional serotyping methods, the microarray assay has a number of advantages, including high-throughout, speed, specificity and sensitivity. It has the potential for routine use in surveillance of serogroup/serotype distribution of *S. pneumoniae* in pneumococcal disease.

Acknowledgments

This work was supported by the NSFC Key Program (30530010), and funds from the Science and Technology Committee of Tianjin City (05YFGPGX07300) to L. W. and L. F.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mimet.2006.07.001.

References

AlonsoDeVelasco, E., Verheul, A.F., Verhoef, J., Snippe, H., 1995. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. Microbiol. Rev. 59, 591–603.

Barker, J.H., Musher, D.M., Silberman, R., Phan, H.M., Watson, D.A., 1999. Genetic relatedness among nontypeable *pneumococci* implicated in sporadic cases of conjunctivitis. J. Clin. Microbiol. 37, 4039–4041.

Brito, D.A., Ramirez, M., de Lencastre, H., 2003. Serotyping *Streptococcus pneumoniae* by multiplex PCR. J. Clin. Microbiol. 41, 2378–2384.

Brown, P.D., Lerner, S.A., 1998. Community-acquired pneumonia. Lancet 352, 1295–1302.

Chizhikov, V., Rasooly, A., Chumakov, K., Levy, D.D., 2001. Microarray analysis of microbial virulence factors. Appl. Environ. Microbiol. 67, 3258–3263.

Fox, G.E., Wisotzkey, J.D., Jurtshuk Jr., P., 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. 42, 166–170.

Garcia, E., Llull, D., Lopez, R., 1999. Functional organization of the gene cluster involved in the synthesis of the pneumococcal capsule. Int. Microbiol. 2, 169–176.

Henrichsen, J., 1999. Typing of *Streptococcus pneumoniae*: past, present, and future. Am. J. Med. 107, 50S–54S.

Jiang, S.M., Wang, L., Reeves, P.R., 2001. Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene clusters. Infect. Immun. 69, 1244–1255.

Klein, J.O., 1981. The epidemiology of pneumococcal disease in infants and children. Rev. Infect. Dis. 3, 246–253.

Kong, F., Wang, W., Tao, J., Wang, L., Wang, Q., Sabananthan, A., Gilbert, G.L., 2005. A molecular–capsular-type prediction system for 90 *Streptococcus pneumoniae* serotypes using partial *cpsA–cpsB* sequencing and *wzy*- or *wzx*-specific PCR. J. Med. Microbiol. 54, 351–356.

Lawrence, E.R., Griffiths, D.B., Martin, S.A., George, R.C., Hall, L.M., 2003. Evaluation of semiautomated multiplex PCR assay for determination of *Streptococcus pneumoniae* serotypes and serogroups. J. Clin. Microbiol. 41, 601–607.

Llull, D., Munoz, R., Lopez, R., Garcia, E., 1999. A single gene (*tts*) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. J. Exp. Med. 190, 241–251.

Roth, S.B., Jalava, J., Ruuskanen, O., Ruohola, A., Nikkari, S., 2004. Use of an oligonucleotide array for laboratory diagnosis of bacteria responsible for acute upper respiratory infections. J. Clin. Microbiol. 42, 4268–4274.

Rouillard, J.M., Zuker, M., Gulari, E., 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res. 31, 3057–3062.

Smit, P., Oberholzer, D., Hayden-Smith, S., Koornhof, H.J., Hilleman, M.R., 1977. Protective efficacy of pneumococcal polysaccharide vaccines. JAMA 238, 2613–2616.

Sniadack, D.H., Schwartz, B., Lipman, H., Bogaerts, J., Butler, J.C., Dagan, R., Echaniz-Aviles, G., Lloyd-Evans, N., Fenoll, A., Girgis, N.I., 1995. Potential interventions for the prevention of childhood pneumonia: geographic and temporal differences in serotype and serogroup distribution of sterile site pneumococcal isolates from children — implications for vaccine strategies. Pediatr. Infect. Dis. J. 14, 503–510.

Wang, L., Reeves, P.R., 1998. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. Infect. Immun. 66, 3545–3551.

Warsen, A.E., Krug, M.J., LaFrentz, S., Stanek, D.R., Loge, F.J., Call, D.R., 2004. Simultaneous discrimination between 15 fish pathogens by using 16S ribosomal DNA PCR and DNA microarrays. Appl. Environ. Microbiol. 70, 4216–4221.

Wilson, W.J., Strout, C.L., DeSantis, T.Z., Stilwell, J.L., Carrano, A.V., Andersen, G.L., 2002. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Mol. Cell. Probes 16, 119–127.

Xu, D., Li, G., Wu, L., Zhou, J., Xu, Y., 2002. PRIMEGENS: robust and efficient design of gene-specific probes for microarray analysis. Bioinformatics 18, 1432–1437.

Yershov, A.L., Jordan, B.S., Guymon, C.H., Dubick, M.A., 2005. Relationship between the inoculum dose of *Streptococcus pneumoniae* and pneumonia onset in a rabbit model. Eur. Respir. J. 25, 693–700.

Zheng, G., Zhai, J., Tang, X., Wang, J., Guo, Z., Yang, R., 2005. Development of 16S rDNA oligonucleotide microarray for identification of pathogenic bacteria. Bull. Acad. Mil. Med. Sci. 29, 13–17.