Development of a microarray-based method for simultaneous detection and serotyping of *Streptococcus pneumoniae* from culture negative serum samples

Avid Hussain Shaik*, Vandana Govindan, Geetha Nagaraj, Kadahalli Lingegowda Ravikumar

Central Research Laboratory, Kempegowda Institute of Medical Sciences, Bangalore, India

ARTICLE INFO

ABSTRACT

Streptococcus pneumoniae identification and serotyping are vital for disease management, surveillance, vaccine development, and to monitor the emerging serovars. Despite the potential benefits of conventional typing method based on culture isolate, it lacks the ability to type directly from the clinical samples. To address the challenge, a novel microarray method was developed to identify and serotype *S. pneumoniae* in culture positive and culture negative serum specimens. The custom pneumococcal microarray chip was designed using sure design software and evaluated with 90 reference strains, culture positive quantitative real time multiplex polymerase chain reaction (qmPCR) positive (*n* = 8) and culture negative qmPCR positive (*n* = 96) serum samples. To selectively amplify *S. pneumoniae* DNA from serum, three different methods: (a) Capsular polysaccharide PCR, (b) Whole genome amplification, and (c) Microbial DNA enrichment were assessed. An in-house developed excel-based software was used to quantify the signals from each serotype. With a combination of Microbial DNA enrichment, the custom pneumococcal microarray chip identified all the reference strains and serum samples serogroup/type information accurately. Of the 104 serum samples, 65 were uniquely identified and 39 were assigned with a combination of their homologous types. A 100% concordance was observed with the results of qmPCR and PCRSeqTyping methods with an additional advantage of multiple serotype detection. Our results signify the ability of the microarray technology to identify and detect *S. pneumoniae* serogroup/type from culture-negative serum specimens. The test is of use even in patients with prior antibiotic treatment and can be used in surveillance and vaccine impact studies.

INTRODUCTION

Pneumonia is a common illness that affects almost 450 million people worldwide. It is a major cause of death among all age groups, resulting in 1.4 million deaths every year [1]. Pneumococcal diseases range from mild respiratory tract mucosal infections, such as otitis media and sinusitis, to more severe diseases, such as pneumonia, septicemia, and meningitis [2]. Globally, pneumonia alone accounted for 16% of deaths in children under the age of 5 in 2016 [3]. Although the global disease burden among children due to pneumococcal pneumoniae is well characterized in children <5 years with an estimated 13.9 million cases [4], it is not well documented in adults, thereby limiting treatment options for prevention.

The incidence of invasive pneumococcal disease (IPD) varies across different countries and is largely underestimated since the diagnosis of this condition, which is usually performed by culturing whole blood or cerebrospinal fluid (CSF), requires the presence of viable pathogens in the clinical samples [5]. The clinical utility of blood culture, the gold standard for the identification and diagnosis of pneumococcal pneumonia is limited to less than 10% of the patients who are positive for *S. pneumoniae*. The low detection rate in culture dependent methods could be due to the presence of pneumonia without bacteremia, autolysis of *S. pneumoniae* during the stationary growth phase, use of antibiotics prior to culture, or due to the inadequacy of samples. In the absence of blood culture, the detection of *S. pneumoniae* teichoic acid antigen in urine (Binax) offers a better alternative. However, this method exhibits

Corresponding Author
Avid Hussain Shaik, Central Research Laboratory, Kempegowda Institute of Medical Sciences, Bangalore, India. E-mail: avidhs@gmail.com

low specificity in children [6]. A better sensitivity and specificity obtained in a relatively shorter time have provided molecular assays, such as Pulse Field Gel Electrophoresis, Restriction fragment length polymorphism, sequential multiplex polymerase chain reaction (PCR), and quantitative multiplex PCR (qmPCR) as additional alternatives to culture or enzyme immunoassay methods [7–11].

The pathogenicity of pneumococci has been linked to various virulence factors such as Capsular polysaccharide (CPS), pneumolysin, autolysin, pneumococcal surface proteins A and C, hyaluronate lyase, pneumococcal surface antigen A, choline binding protein A, and neuraminidase enzymes [12]. CPS, the primary virulence factor, inhibits complement activity and phagocytosis. Based on the composition of polysaccharide capsules, *S. pneumoniae* is characterized into more than 90 serotypes. Genes required for the synthesis of capsule are within the CPS operon, which ranges from 10 (serotype 3) to 30 kb (serotype 38) [13].

The phenotypic serotyping methods, such as capsular swelling reaction (Quellung test) and latex agglutination tests, are laborious, expensive, and are associated with subjective interpretation bias [14]. Genotyping methods, such as sequential multiplex PCR, qmPCR, sequencing, and microarrays that assess genome variations, are inherently beneficial for the detection and typing of infectious agents [9,10,15]. Microarray technology can be adopted for accurate identification, multiple serotype detection, and relative abundance estimation of *S. pneumoniae* serotypes. Target-probe hybridization in microarray reduces amplification bias and use of multiple probes increases the specificity in detecting serotypes and reduces false positives. These improvements have allowed the transition of microarrays from strictly research settings to clinical diagnostic applications and have led to the development of commercially available detection platforms [16–21].

A barrier to the approach of detecting pneumococcal infection from clinical samples is the overwhelming ratio of human to pathogen DNA in samples with low abundance of pathogen, which is typical for most clinical specimens [22–24]. In the present study, a custom array was designed for the detection and serotyping of *S. pneumoniae* in suspected pneumonia cases using serum samples. Three different techniques, namely, CPS PCR, whole genome amplification (WGA), and microbial DNA enrichment were evaluated. Simultaneous identification and serotyping of *S. pneumoniae* was achieved using microarray from culture in negative serum samples.

MATERIALS AND METHODS

Reference strains, media, and serotyping

A total of 90 *S. pneumoniae* reference strains were sourced from Staten Serum Institute (SSI), Denmark. Strains stored in skimmed milk, tryptone, glucose, and glycerol media at −80°C, were cultured on 5% sheep blood agar (Chromogen, Hyderabad) for 18–24 hours at 37°C with 5% CO_2. Quellung test was performed using the Pneumotest kit (SSI, Denmark) with 12 pool antisera following manufacturer's instructions. The pneumotest kit detects only 55 serogroups/types and the non-vaccine groups/types are reported as Non-typeable (NT).

Serum specimens

Culture negative qmPCR positive ($N = 96$) and culture positive qmPCR positive ($N = 8$) *S. pneumoniae* serum samples from children <5 years with IPD were sourced from Central Research Laboratory, KIMS Hospital. The samples were collected during 2014–2015 from Bangalore, India (Table 2). The serum samples were previously subjected for qmPCR assay [25,26] and PCRSeqTyping [26,27] method for detection of pneumococcal infection and serotype, respectively. The qmPCR assay developed and validated by Ganaie et al. [25] uses four oligo primer-probe sets, namely, Pneumolysin (ply), Autolysin (lytA), Pneumococcal surface adhesin-A (psaA), Spn9802 gene fragment for the detection of *S. pneumoniae* from serum samples. The PCRSeqTyping assay involves PCR amplification and sequencing of the cpsB gene of *S. pneumoniae* as detailed in Nagaraj et al. [27].

Ethics statement

The study was conducted according to the guidelines and declaration of Helsinki. Written informed consent was obtained from each patient or legal guardian as applicable.

DNA extraction

Genomic DNA was extracted from standard strains and serum (volume 400–600 µl) specimens using QiaAmp DNA mini kit protocol for Gram-positive bacteria (Qiagen, Germany) following the manufacturer's instructions. The extracted DNA was assessed using Nanodrop (Thermo fisher scientific, USA) and stored at −20°C until further analysis.

Custom array designing

The custom pneumococcal microarray chip was designed by sure design software (Agilent technologies, USA) using the published sequences of *S. pneumoniae* CPS gene cluster (http://www.sanger. ac.uk/Projects/S. /CPS/) and printed on the Agilent Sure Print 8 × 15K format array (Agilent Technologies, USA). The custom array comprises of 15,167 oligonucleotide probes with multiple probes per serotype (16–575 probes) targeting 90 *S. pneumoniae* serotypes. The probes (57–60 mer) were T_m (melting temperature) matched and have a predicted performance score based on T_m, GC content, a hairpin ΔG, sequence complexity, and metrics to measure homology with the rest of the reference genome. Additionally, the array has 87 internal positive control probes (targeting Endogenous "species-specific" sequences and predicted to have high signal due to multiple copies found in the targeted genome), 82 bright corner control probes (used for orientation purposes and are placed in the corners of the array with a different pattern for each corner), 80 negative control probes (designed to measure background noise. These probes form a hairpin and do not hybridize well with labeled sample of any species) and 49 dark corner probes (used for orientation purposes in the array corners). The custom array also contains species-specific probes for *Streptococcus pyogenes*, *Streptococcus mitis*, *Streptococcus oralis*, *Moraxella catarrhalis*,

Neisseria meningitidis, Haemophilus influenzae, Pseudomonas aeruginosa, Klebsiella pneumoniae, Staphylococcus aureus, Legionella pneumophila, Chlamydophila pneumoniae, and *Mycoplasma pneumoniae,* which reside in human nasopharynx along with *S. pneumoniae.* The list of probes, targeted CPS genes and their sequences, will be provided upon request.

Agilent microarray protocol

DNA labeling and purification

The microarray experiment was conducted by following Agilent microarray protocol (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis v 7.2) using SureTag Complete DNA Labeling Kit (Agilent p/n 5190–4240). Briefly, 2.5 μl of random primer was added to each reaction tube containing 13 μl of gDNA (180–220 ng) to make a total volume of 15.5 μl. Sample tubes were incubated on a heat block at 95°C for 3 minutes and then cooled on ice for 5 minutes. Following the incubation, the samples were centrifuged at 6,000 g for 1 minute. DNA was then labeled by adding 9.5 μl of labeling mix containing 2.5 μl of 10× dNTP mix, 0.5 μl of Exo-Klenow enzyme, and 1.5 μl of Cy3-dUTP or Cy5-dUTP. Sample tubes were incubated in a thermocycler at 37°C for 2 hours followed by 65°C for 10 minutes. Labeled DNA was then purified using Agilent purification Columns. The concentration and specific activity of purified, labeled DNA was determined by Nanodrop.

Probe hybridization

Twenty-nine microliter of the hybridization master mix was added to purified, labeled DNA (3–5 μg) making it to a total volume of 45 μl. The mixture was then incubated at 95°C for 3 minutes, 37°C for 30 minutes and loaded onto the microarray slide and hybridized at 67°C for 16 hours.

Washing and image acquisition

The hybridized array was subjected for washing following the manufacture instructions. The array image was acquired with Agilent Sure scanner (Agilent Technologies, USA) using Scan control software v9.0, by selecting default protocol "AgilentHD_CGH." The microarray scanner is a sophisticated laser-induced fluorescence scanner designed to measure the fluorescence intensity of labeled sample nucleic acid (DNA and RNA) bound to microarrays.

Data analysis

The signal and background intensities of each spot were extracted from array image using feature extraction software v.3.0. Data analysis was performed using in-house developed Microsoft Excel software. The software assigns "Detected or Compromised" status to probes based on the background subtracted signal intensity (gProcessed Signal) levels. The following cut-offs were used for quality control testing: "Detected" = Signal intensities well above background intensities (>50), "Compromised" = Signal intensities below background intensities (<50). The samples were blinded throughout microarray testing and data analysis.

Microarray analysis of reference strains and serum samples

The genomic DNA from reference strains ($N = 90$) and Genomic DNA from serum ($N = 8$, PIDOPS 1-8) were subjected for microarray analysis by following Agilent microarray protocol.

Microarray analysis of serum samples

Selective microbial DNA enhancement from serum samples

The serum DNA was subjected for three different methods, namely, CPS PCR, WGA, and Microbial DNA enrichment to explore the possibility of selective amplification or enrichment of microbial DNA from host DNA.

CPS PCR:

DNA from reference strains ($N = 6$) and culture negative qmPCR positive Serum DNA ($N = 24$, PIDOPS1-24) were subjected to PCR amplification of complete CPS region (10–30 kb) of *S. pneumoniae* using specific primers pair [20] targeting *dexB* and *aliA* genes CPS1 (TTGCCAATGAAGAGCAAGACTTGACAGTAG) and CPS2 (CAATAATGTCACGCCCGCAAGGGCAAGT). The CPS PCR reaction was performed on ABI 9700 thermal cycler in 50 μl volume using 5 μl Long Range PCR Master mix (Qiagen), Primer (concentration 400 nM) and input DNA (100–200 ng) by following thermal program: initial denaturation at 93°C for 3 minutes, 35 cycles of 93°C 15 seconds, 62°C 30 seconds, and 68°C 20 minutes (extension 1 minute/kb). PCR products were analyzed with 1.0% agarose gel electrophoresis.

WGA:

Culture negative qmPCR positive serum DNA ($N = 24$, PIDOPS1-24) and DNA from reference strains ($N = 6$) were subjected to WGA using Genome Plex amplification kit (Rubicon Genomics/Sigma-Aldrich, Poole, UK) according to the manufacturer's instructions. In the first step, DNA samples were incubated at 95°C for 4 minutes in 10× fragmentation buffer. In the second step, the samples were further incubated with the Library Preparation Buffer and Library Stabilization Solution at 95°C for 2 minutes. One microliter of Library Preparation Enzyme was added to the mix and then incubated at 16°C for 20 minutes, 24°C for 20 minutes, 37°C for 20 minutes, and 75°C for 5 minutes. In the third step, the resulting sample was amplified using WG DNA polymerase, by following thermal conditions—denaturation at 95°C for 3 minutes and 14 cycles of 94°C for 15 seconds, 65°C for 5 minutes. Whole genome amplified products were purified using QiaQuick PCR purification columns (Qiagen, Germany) and quantified using NanoDrop.

Enrichment of microbial DNA:

Culture negative qmPCR positive ($N = 96$, PIDOPS1-96) culture positive qmPCR positive serum DNA ($N = 6$, PIDOPS 97-104) and six reference strains (positive controls) were enriched using NEBNext Microbiome DNA Enrichment Kit (New England BioLabs) according to manufacturer instructions. Briefly, 80 μl of MBD2-Fc-bound magnetic beads were prepared and incubated the mixture with rotation for 10 minutes at room temperature (RT). The bound MBD2-Fc magnetic beads were then collected by

magnet and washed twice with 1 ml ice-cold 1× bind/wash buffer before being resuspended in ice-cold 1× bind/wash buffer. MBD-Fc Protein beads prepared as described above was mixed with 500 ng of serum DNA (contains bacterial and host DNA) and incubated for 15 minutes at RT with rotation. The incubated mixture was placed on a magnetic rack until the supernatant was clear and beads were collected on the wall of the tube. The supernatant, containing enriched microbial DNA, was carefully removed with a pipette and then purified by following Agencourt AMPureXP bead cleanup protocol (Beckman Coulter Inc. #A63881). The captured host DNA from the MBD2-Fc-bound magnetic bead pellet was eluted by following manufacturer protocol.

Validation of enriched DNA by quantitative PCR (qPCR):

Un-enriched (#3) and enriched DNA (#3) samples in triplicates were subjected for qPCR reaction with following primers/probes aiming conserved region of 16SrRNA gene and Human GAPDH housekeeping gene. GAPDHFP-GAAGGTGAAGGTCGGAGT, GAPDHRP-GAAGATGGTGATGGGATTTC, GAPDH probe Quasar670-CTCAAGTTGGAAACCACGAGTAAGAGTGAT GAA-BHQ2, 16SrRNAFP-CCATGAAGTCGGAATCGCTAG, 16SrRNARP-GCTTGACGGGCGGTGT, 16SrRNA probe FAM-GCCACGGTGAATACGTTCCCGGGCCTTGTAC-BHQ1. qPCR reaction was performed on Rotor Gene Q (Qiagen) instrument in 25 µl volumes using 2× Rotor gene multiplex master mix (Qiagen), primer/probe mix (concentrations 400 nM/200 nM), and template DNA (50–100 ng) by following program: 5 minutes at 95°C, 30 seconds at 95°C, 30 seconds at 60°C, for 40 cycles. The results were analyzed using Rotor Gene Q (Qiagen) software.

Multiple serotype detection:

The custom array was validated for detection of multiple serotypes by pooling enriched DNA samples ($N = 16$) with different serotypes in 1:1 ratio (200 ng of each enriched DNA) and subjected for Agilent microarray protocol as described previously (Table 2).

RESULTS

Reference strains results

The custom pneumococcal microarray identified the reference strains to their respective serogroup/type accurately. Overall 61 (67.7%) serotypes were uniquely identified and 29 (32.2%) were identified along with homologous serotypes. Serotypes reported as NT by Pneumotest kit were resolved to serogroup/type by Microarray (Fig. 1).

Microarray results for serum samples

There was no signal observed for culture negative qmPCR positive serum DNA (PIDOPS1-8), which was directly hybridized on custom pneumococcal chip and could be attributed to host DNA dominance. On the other hand, reference strains DNA samples showed strong signals and led to accurate identification of serotype. There was no signal observed for species-specific probes, indicating absence of other respiratory pathogens (listed in custom array designing section) in the tested serum samples.

CPS PCR

The 24-culture negative qmPCR positive Serum DNA samples did not exhibit any amplification, but the six reference strains DNA samples (positive control) had a strong amplification indicating poor quality of DNA from the Serum.

Whole genome amplification

The 24-culture negative qmPCR positive Serum DNA samples were subjected for WGA, but there were no signals from these
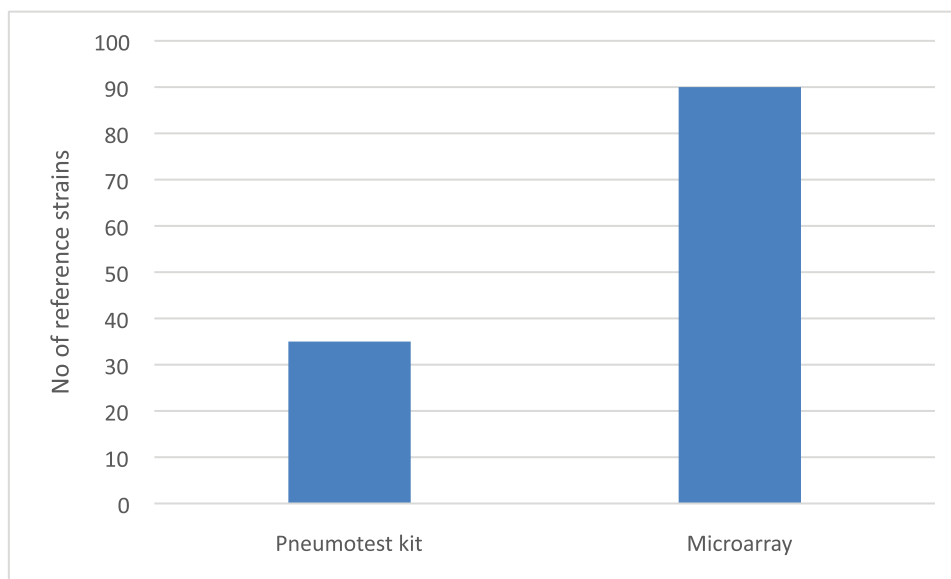

Figure 1: Reference strains (SSI) results by pneumotest kit and Microarray: Reference strains not typed by Pneumotest kit (Quellung): 35/90 no's- 13,21,27,31,34,36,3,38,39,40,42,43,44,45,46,48,16A,16F,24A,24B,24F,25A,25F,28A,28F,32A,32F,35A,35B,35C,35F,41A,41F,47A,47.

samples when they were subjected for microarray. This could be due to fact that the WGA step amplified both bacterial DNA and Host DNA equally and did not help in the removal of host DNA.

Enrichment of microbial DNA

There was drastic decrease of concentration in culture negative qmPCR positive ($N = 96$) and culture positive qmPCR positive ($N = 8$) serum DNA samples, indicating the removal of host DNA (Fig. 2).

Validation of enriched DNA by qPCR

The unenriched samples showed very little changes in the C_t (cycle threshold) value for the 16S rDNA target, whereas there was a decrease of ~8–9 C_t values between the unenriched and enriched samples for GAPDH target indicating the removal of host DNA (Table 1).

Enriched serum sample processing with Agilent microarray protocol

The enriched serum samples (#104) hybridized on custom pneumococcal chip showed strong signals and the serotype was identified. Microarray results and serotyping results by PCRSeqTyping showed 100% concordance without any discrepancies (Table 2).

Multiple serotype detection

The serotypes of pooled serum samples (Table 2) were identified accurately confirming the ability of custom pneumococcal chip for detection of multiple serotypes.

DISCUSSION

The capacity to detect pneumococcal serotypes is limited largely in the developing world. On the other hand, a large number of deaths occur in these countries, which calls for an epidemiological monitoring of pneumococcus in the population [28]. Identifying serotype of *S. pneumoniae* strains provides an invaluable tool to correlate the association of serotype with carriage or invasive disease and to monitor the impact of vaccine introduction [29]. The current gold standard, culture followed by a capsular reaction test, is particularly challenging, as it requires experienced technicians and is limited to a few laboratories. In addition, this methodology cannot identify the carriage of multiple serotypes, emerging serovars, replacement strains, and vaccine escape mutants [30]. Several culture-independent molecular methods can address many of these issues and has the potential to provide additional information. They have shown to increase the detection rates in both colonization and invasive disease studies [25,31,32]. Conventional PCR-based methods to detect and type pneumococci has been described with limitations due to the difficulty in quantifying PCR products, detecting multiple serotypes in a single step, differentiating amplicons of similar size and erroneous results in presence of genetic variations in the primer binding region [33–35]. Besides general applicability with high discriminatory power, the molecular typing assays are economical, detect and type more than one serotype from the clinical specimen and its implementation in clinical laboratories with ease and accuracy. One such molecular approach to multiplex detection and characterization is the microarray technology [18,20,21,27].
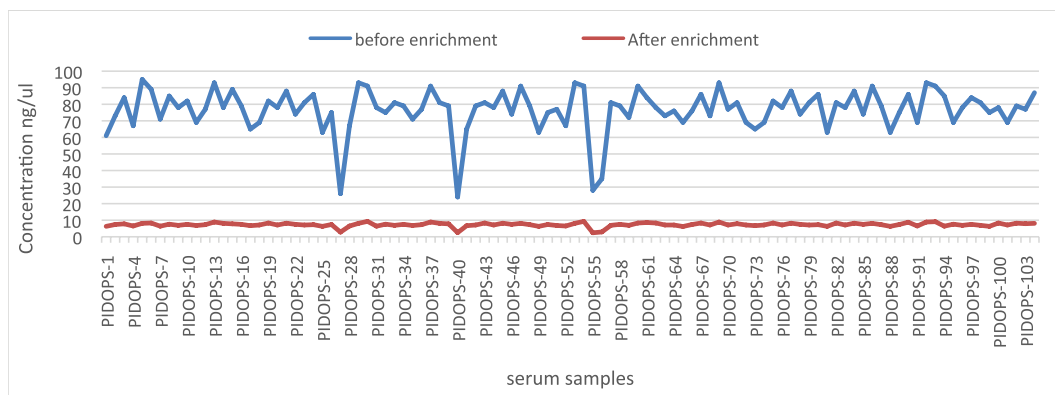
Figure 2: Serum DNA samples concentrations before and after Microbial DNA enrichment method.

Table 1. qPCR validation results for 16S rDNA and GAPDH PCR targets.

Samples	16S rDNA qPCR (average C_t value)	GAPDH qPCR (average C_t value)
PIDOPS-1 (un enriched sample)	24.31	16.19
PIDOPS-1 (enriched sample)	24.42	25.32
PIDOPS-2 (un enriched sample)	23.61	17.78
PIDOPS-2 (enriched sample)	23.49	25.32
PIDOPS-3 (un enriched sample)	24.31	17.43
PIDOPS-3 (enriched sample)	24.19	24.32

Table 2. Serotype distribution of *S. pneumoniae* in serum samples.

S. no	Culture negative qmPCR positive serum samples	Microarray results by Microbial DNA enrichment method	Sequetyping results
1	PIDOPS-1	19F	19F
2	PIDOPS-2	6A/6B	6B
3	PIDOPS-3	19A	19A
4	PIDOPS-4	18C	18C
5	PIDOPS-5	23F	23F
6	PIDOPS-6	6A/6B	6A
7	PIDOPS-7	1	1
8	PIDOPS-8	19A	19A
9	PIDOPS-9	19A	19A
10	PIDOPS-10	9A/9V	9V
11	PIDOPS-11	3	3
12	PIDOPS-12	9A/9V	9V
13	PIDOPS-13	9N/9L	9N
14	PIDOPS-14	6A/6B	6A
15	PIDOPS-15	13	13
16	PIDOPS-16	9N/9L	9N
17	Spiked sample: Mix of PIDOPS-17 and PIDOPS-18	14 and 7F	14
18	Spiked sample: Mix of PIDOPS-18 and PIDOPS-17	7F and 14	7F
19	PIDOPS-19	13	13
20	PIDOPS-20	6A/6B	6B
21	Spiked sample: Mix of PIDOPS-21 and PIDOPS-22	18C and 23F	18C
22	Spiked sample: Mix of PIDOPS-22 and PIDOPS-21	23F and 18C	23F
23	PIDOPS-23	19F	19F
24	PIDOPS-24	9N/9L	9N
25	Spiked sample: Mix of PIDOPS-25 and PIDOPS-26	18C and 1	18C
26	Spiked sample: Mix of PIDOPS-26 and PIDOPS-25	1 and 18C	1
27	PIDOPS-27	6A/6B	6B
28	Spiked sample: Mix of PIDOPS-28 and PIDOPS-29	23F and 1	23F
29	Spiked sample: Mix of PIDOPS-29 and PIDOPS-28	1 and 23F	1
30	PIDOPS-30	11A/11D/18F	11A
31	PIDOPS-31	23F	23F
32	PIDOPS-32	10F/10C	10F
33	PIDOPS-33	5	5
34	PIDOPS-34	9A/9V	9V
35	Spiked sample: Mix of PIDOPS-35 and PIDOPS-36	14 and 1	14
36	Spiked sample: Mix of PIDOPS-36 and PIDOPS-35	1 and 14	1
37	PIDOPS-37	19F	19F
38	PIDOPS-38	6A/6B	6B
39	PIDOPS-39	14	14
40	Spiked sample: Mix of PIDOPS-40 and PIDOPS-41	19F and 3	19F
41	Spiked sample: Mix of PIDOPS-41 and PIDOPS-40	3 and 19F	3
42	PIDOPS-42	19A	19A
43	PIDOPS-43	11A/11D/18F	11A
44	PIDOPS-44	9N/9L	9N
45	PIDOPS-45	6A/6B	6B
46	PIDOPS-46	6A/6B	6A
47	PIDOPS-47	9A/9V	9V
48	PIDOPS-48	20	20
49	PIDOPS-49	9N/9L	9N

S. no	Culture negative qmPCR positive serum samples	Microarray results by Microbial DNA enrichment method	Sequetyping results
50	Spiked sample: Mix of PIDOPS-50 and PIDOPS-51	1 and 14	1
51	Spiked sample: Mix of PIDOPS-51 and PIDOPS-50	14 and 1	14
52	PIDOPS-52	23F	23F
53	PIDOPS-53	14	14
54	PIDOPS-54	9N/9L	9N
55	PIDOPS-55	9A/9V	9V
56	PIDOPS-56	9N/9L	9N
57	Spiked sample: Mix of PIDOPS-57 and PIDOPS-58	5 and 19F	5
58	Spiked sample: Mix of PIDOPS-58 and PIDOPS-57	19F and 5	19F
59	PIDOPS-59	6A/6B	6B
60	PIDOPS-60	10F/10C	10F
61	PIDOPS-61	1	1
62	PIDOPS-62	1	1
63	PIDOPS-63	14	14
64	PIDOPS-64	1	1
65	PIDOPS-65	14	14
66	PIDOPS-66	9N/9L	9N
67	PIDOPS-67	3	3
68	PIDOPS-68	13	13
69	PIDOPS-69	14	14
70	PIDOPS-70	1	1
71	PIDOPS-71	18C	18C
72	PIDOPS-72	23F	23F
73	PIDOPS-73	1	1
74	PIDOPS-74	6A/6B	6B
75	PIDOPS-75	1	1
76	PIDOPS-76	1	1
77	PIDOPS-77	14	14
78	PIDOPS-78	5	5
79	PIDOPS-79	1	1
80	PIDOPS-80	11A/11D/18F	11A
81	PIDOPS-81	10F/10C	10F
82	PIDOPS-82	18C	18C
83	PIDOPS-83	6A/6B	6B
84	PIDOPS-84	19F	19F
85	PIDOPS-85	14	14
86	PIDOPS-86	23F	23F
87	PIDOPS-87	1	1
88	PIDOPS-88	9N/9L	9N
89	Spiked sample: Mix of PIDOPS-89 and PIDOPS-90	14 and 5	14
90	Spiked sample: Mix of PIDOPS-90 and PIDOPS-89	5 and 14	5
91	PIDOPS-91	6A/6B	6B
92	PIDOPS-92	6A/6B	6B
93	PIDOPS-93	6A/6B	6B
94	PIDOPS-94	19F	19F
95	PIDOPS-95	10F/10C	10F
96	PIDOPS-96	23F	23F
97	PIDOPS-97	6A/6B	6B
98	PIDOPS-98	19F	19F
99	PIDOPS-99	6A/6B	6B

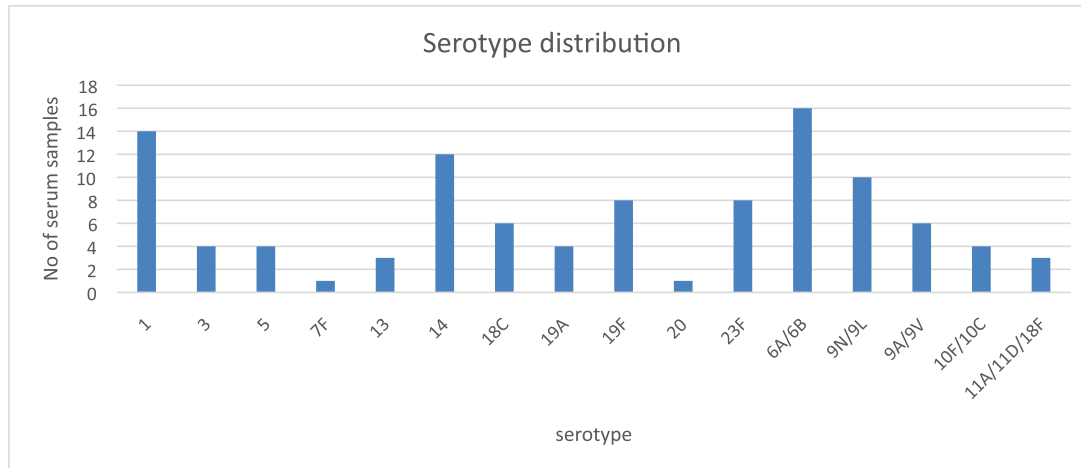| S. no	Culture negative qmPCR positive serum samples	Microarray results by Microbial DNA enrichment method	Sequetyping results
100	PIDOPS-100	9A/9V	9V
101	PIDOPS-101	9N/9L	9N
102	PIDOPS-102	18C	18C
103	PIDOPS-103	3	3
104	PIDOPS-104	14	14

Figure 3: *S. pneumoniae* serotype distribution in serum samples.

Custom microarrays have been developed to assess the genetic diversity between different *S. pneumoniae* serotypes [36] and to identify multiple bacterial pathogens from culture and clinical samples [37]. A microarray was designed targeting CPS genes, wzy or cpsA, could identify only serotypes contained in 23 valent pneumococcal polysaccharide vaccines [38]. BµG@S SP-CPS v1.4.0 array designed for targeting CPS genes could identify and type 90 serotypes from culture isolates but could not be validated on clinical specimens [18]. This array has been widely used by different groups [39–42]. Recently, a microarray developed based on glycosyltransferase (GT) gene sequences could detect only vaccine-related serotypes [19]. In an infant carriage study, the nasopharyngeal co-colonization detection rates were similar between Microarray and sweep method, but the microarray identified greatest number of serotypes [39]. In a Multi centric study, microarray had the highest sensitivity (95.8%) and high positive predictive value (93.7%) compared to other 20 alternate serotyping methods [43]. Despite the availability of different microarray-based methods for pneumococcal serotyping, there is always a room for improvement in the available methods, in terms of coverage, throughput, multiplexing capacity, and for typing directly from sterile body fluids, such as serum, CSF, and pleural fluid.

In the present study, we describe the use of high density custom pneumococcal microarray for identifying and typing 90 serotypes/serogroups from serum samples. The array (which identified all 90 reference strains accurately) were evaluated with culture negative qmPCR positive serum samples failed to produce the signals due to high amounts of Host DNA (Human DNA) and low amounts of pathogen DNA (bacterial DNA) [22–24]. To address this issue, three different methods were evaluated for selective enrichment of bacterial DNA. The CPS PCR could not amplify 20 kb CPS region from serum DNA as highly intact DNA was necessary for amplification of longer amplicons [44–46]. The WGA method amplified host and microbial DNA equally but could not produce the signal when subjected to microarray testing [47,48]. The Microbiome DNA Enrichment method successfully alienated microbial DNA from CpG-methylated human DNA while keeping the microbial diversity intact after enrichment (Table 1; Fig. 2). Other groups adopted similar strategy to enrich bacterial and malarial parasite DNA from fecal and human blood samples, respectively [49,50]. A Microarray experiment then performed using enriched samples lead to the accurate identification of all the samples tested, proving the efficiency of the enrichment method for typing (Table 2).

More than 90 *S. pneumoniae* individual serotypes were recognized. Out of these, 29 serotypes shared close homology with each other and differed in few bases (Homologous serogroups/types). They are 6A/6B, 9A/9V/9N/9L, 10A/10B/10C/10F, 11A/11B/11C/11D/18F, 15B/15C, 18C/18F, 22A/22F, 32A/32F, 33A/33B/33F and 12A/12F/44/46. In our study, the custom pneumococcal array identified accurate serotype/group for 104 serum samples and the results were in 100% concordance with PCRSeqTyping. Serotypes of 60 samples were identified uniquely and 44 were identified along with their homologous serogroup/type. 6A/6B was the most predominant homologous serogroup

followed by 9N/9L, 9A/9V, 10F/10C, and 11A/11D/18F (Fig. 3). Bentley et al. [20] descried the sequence of the capsular genes of all the 90 serotypes as reference, laying the foundation for molecular serotyping and since then new serotypes have been discovered. However, recent studies demonstrated that the CPS loci of pneumococci are different to the original reference set of 90 *cps* loci which are often used for the design of many sequence-based serotyping assays or algorithms [42,51]. It is, therefore, important to update sequence-based typing tools/algorithms with latest genetic variants for more accurate serotyping.

The custom array was also tested on pooled serum DNA samples with 1:1 ratio and demonstrated the capability to detect multiple serotypes. However, future studies should explore other ratios, perhaps 9:1 ratio, to know the minor serotype and evaluate complex samples, such as nasopharyngeal swabs and sputum. The current method is much more affordable, in terms of per-isolate or per-sample costs as two color hybridization approach was followed. One sample was labeled with Cy3, analyzed in Green channel, and the other one labeled with Cy5, analyzed in Red channel which enabled greater throughput (16 samples in single array) than the existing microarray methods. The limitation of our study was that the microarray method was not able to discriminate homologous serogroups/types. Recent studies, however, have shown the possibility to discriminate homologous serotypes from their respective serogroups/types in multiple steps [26]. These challenges could be addressed by including new unique serotype-specific probes to exiting custom pneumococcal chip with the availability of sequence data in public database. To the best of our knowledge, this is the first instance in which microarray was coupled with microbiome DNA enrichment method to determine *S. pneumoniae* serogroups/types from clinical specimens. Its ability to identify the serotype in culture negative serum samples with multiple serotype detection provided useful information for diagnosis, vaccine, and surveillance studies.

CONFLICT OF INTEREST

Nil.

REFERENCES

1. Ruuskanen O, Lahti E, Jennings LC, Murdoch DR. Viral pneumonia. Lancet 2011;377:1264–75.
2. Yildirim I, Shea KM, Pelton SI. Pneumococcal disease in the era of Pneumococcal conjugate vaccine. Infect Dis Clin North Am 2015;29:679–97.
3. United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). Levels & trends in child mortality: Report 2017, estimates developed by the UN Inter-agency Group for Child Mortality Estimation. United Nations Children's Fund, New York, NY, 2017.
4. Said MA, Johnson HL, Nonyane BA. Estimating the burden of pneumococcal pneumonia among adults: a systematic review and meta-analysis of diagnostic techniques. PLoS One 2013;8:e60273.
5. Azzari C, Moriondo M, Indolfi G, Massai C, Becciolini L, de Martino M, et al. Molecular detection methods and serotyping performed directly on clinical samples improve diagnostic sensitivity and reveal increased incidence of invasive disease by *Streptococcus pneumoniae* in Italian children. J Med Microbiol 2008;57:1205–12.
6. Song JY, Eun BW, Nahm MH. D The Clinical Microbiology Reviews journal is part of American Society for Microbiology (ASM) iagnosis of pneumococcal pneumonia: current pitfalls and the way forward. Infect Chemother 2013;45:351–66.
7. Burd EM. Validation of laboratory-developed molecular assays for infectious diseases. Clin Microbiol Rev 2010;23:550–76.
8. Bonofiglio L, Gardella N, Mollerach M. In: Magdeldin S, editor. Application of molecular typing methods to the study of medically relevant gram-positive Cocci, gel electrophoresis—advanced techniques. 2012.
9. Bello Gonzalez T, Rivera-Olivero IA, Sisco MC, Spadola E, Hermans PW, de Waard JH. PCR deduction of invasive and colonizing pneumococcal serotypes from Venezuela: a critical appraisal. J Infect Dev Ctries 2014;8:469–73.
10. Pai R, Gertz RE, Beall B. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. J Clin Microbiol 2006;44:124–31.
11. Batt SL, Charalambous BM, McHughes TD, Martin S, Gillespie SH. Novel PCR-restriction fragment length polymorphism method for determining serotypes or serogroups of *Streptococcus pneumoniae* isolates. J Clin Microbiol 2005;43:2656–61.
12. Jedrzejas MJ. Pneumococcal virulence factors: structure and function. Microbiol Mol Biol Rev 2001;65:187–207.
13. Cartee RT, Forsee WT, Jensen JW, Yother J. Expression of the Streptococcus pneumoniae type 3 synthase in Escherichia coli. Assembly of type 3polysaccharide on a lipid primer. J Biol Chem 2001;276:48831–9.
14. Slotved HC, Kaltoft M, Skovsted IC, Kerrn MB, Espersen F. Simple, rapid latex agglutination test for serotyping of *Pneumococci* (Pneumotest-Latex). J Clin Microbiol 2004;42:2518–22.
15. Waggoner JJ, Abeynayake J, Sahoo MK, Gresh L, Tellez Y, Gonzalez K, et al. Single- reaction, multiplex, real-time RT-PCR for the detection, quantitation, and serotyping of dengue viruses. PLoS Negl Trop Dis 2013;7:e2116.
16. Miller MB, Tang YW. Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol Rev 2009;22:611–33.
17. Ballarini A, Segata N, Huttenhower C, Jousson O. Simultaneous quantification of multiple bacteria by the BactoChip microarray designed to target species-specific marker genes. PLoS One 2013;8:e55764.
18. Hinds J, Gould KA, Witney AA, Baldry SJ, Lambertsen L, Hannage WP, et al. Molecular serotyping of *Streptococcus pneumoniae*: a microarray-based tool with enhanced utility. Poster presented at International Symposium on Pneumococci and Pneumococcal Diseases, Hyderabad, India, March 9, 2014.
19. Tomita Y, Okamoto A, Yamada K, Yagi T, Hasegawa Y, Ohta M. A new microarray system to detect *Streptococcus pneumoniae* serotypes. J Biomed Biotechnol 2011:352736.
20. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet 2006;2:e31.
21. Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehricht R, Hächler H. Fast DNA serotyping of Escherichia coli by use of an oligonucleotide microarray. J Clin Microbiol 2007;45:370–9.
22. Horz H-P, Scheer S, Vianna ME, Conrads G. New methods for selective isolation of bacterial DNA from human clinical specimens. Anaerobe 2010;16:47–53.

23. Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Pinches R, et al. An effective method to purify Plasmodium falciparum DNA directly from clinical blood samples for whole genome high-throughput sequencing. PLoS One 2011;6:e22213.

24. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, et al. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. PLoS One 2013;8:e76096.

25. Ganaie FA, Govindan V, Ravi Kumar KL. Standardisation and evaluation of a quantitative multiplex real-time PCR assay for the rapid identification of streptococcus pneumoniae. Pneumonia 2015;6:1.

26. Vandana G, Feroze AG, Geetha N, Avid H, Kumar KR. Pan India distribution of pneumococcal serotypes (PIDOPS) causing invasive pneumococcal disease and pneumonia in children between 6 weeks and 5 years and their antimicrobial resistance—Phase I. Pedia Infect Dis 2016;8:47–51.

27. Nagaraj G, Ganaie F, Govindan V, Ravikumar KL. Development of PCRSeqTyping—a novel molecular assay for typing of Streptococcus pneumonia. Pneumonia 2017;9:8.

28. Donkor ES. Molecular typing of the pneumococcus and its application in epidemiology in sub-Saharan Africa. Front Cell Infect Microbiol 2013;3:12.

29. Vernet G, Saha S, Satzke C, Burgess DH, Alderson M, Maisonneuve JF, et al. Laboratory-based diagnosis of pneumococcal pneumonia: state of the art and unmet needs. Clin Micro Infect 2011;3:1–3.

30. Kanungo R, Rajalakshmi B. Serotype distribution & antimicrobial resistance in *Streptococcus pneumoniae* causing invasive & other infections in south India. Indian J Med Res 2001;114:127–32.

31. Molander V, Elisson C, Balaji V, Backhaus E, John J, Vargheese R, et al. Invasive pneumococcal infections in Vellore, India: clinical characteristics and distribution of serotypes. BMC Infect Dis 2013;13:532.

32. Kanungo R1, Bhaskar M, Kumar A, Badrinath S, Rajalakshmi B. Detection of pneumolysin in cerebrospinal fluid for rapid diagnosis of Pneumococcal meningitis. Indian J Med Res 2004;119:75–8.

33. A Gervaix, J Corbeil, F Raymond. A new serotyping method of *S. Pneumoniae* using an automated microarray-based assay. BMC Proc 2011;5(Suppl 6):O28.

34. Raymond F, Boucher N, Allary R, Robitaille L, Lefebvre B, Tremblay C, et al. Serotyping of *Streptococcus pneumoniae* based on capsular genes polymorphisms. PLoS One 2013;8:e76197.

35. Saha SK, Darmstadt GL, Baqui AH, Hossain B, Islam M, Foster D, et al. Identification of serotype in culture negative pneumococcal meningitis using sequential multiplex PCR: implication for surveillance and vaccine design. PLoS One 2008;3:e3576.

36. McCluskey J, Dowson CG, Mitchell TJ. The use of microarray technology for the analysis of *Streptococcus pneumoniae*. Comp Funct Genom 2002;3:366–8.

37. Roth S, Jalava J, Ruuskanen O, Ruohola A, Nikkari S. Use of an oligonucleotide array for laboratory diagnosis of bacteria responsible for acute upper respiratory infections. J Clin Microbiol 2004;42:4268–74.

38. Wang Q, Wang M, Kong F, Gilbert GL, Cao B, Wang L, et al. Development of a DNA microarray to identify the *Streptococcus pneumoniae* serotypes contained in the 23-valent pneumococcal polysaccharide vaccine and closely related serotypes. J Microbiol Methods 2007;68:128–36.

39. Turner P, Hinds J, Turner C, Jankhot A, Gould K, Bentley SD, et al. Improved detection of nasopharyngeal cocolonization by multiple Pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. J Clin Microbiol 2011;49:1784–9.

40. Kandasamy R, Gurung M, Thapa A, Ndimah S, Adhikari N, Murdoch DR, et al. Multi-serotype Pneumococcal nasopharyngeal carriage prevalence in vaccine naïve nepalese children, assessed using molecular serotyping. PLoS One 2015;10:e0114286.

41. Valente C, Hinds J, Pinto F, Brugger SD, Gould K, Mühlemann K, et al. Decrease in Pneumococcal co-colonization following vaccination with the seven-valent pneumococcal conjugate vaccine. PLoS One 2012;7:e30235.

42. Manna S, Dunne EM, Ortika BD, Pell CL, Kama M, Russell FM, et al. Discovery of a *Streptococcus pneumoniae* serotype 33F capsular polysaccharide locus that lacks *wcjE* and contains a *wcyO* pseudogene. PLoS One 2018;13:e0206622.

43. Satzke C, Dunne EM, Porter BD, Klugman KP, Mulholland EK. The PneuCarriage project: a multi-centre comparative study to identify the best serotyping methods for examining Pneumococcal carriage in vaccine evaluation studies. PLoS Med 2015;12:e1001903.

44. Farrington HL, Edwards CE, Guan X, Carr MR, Baerwaldt K, Lance RF. Mitochondrial genome sequencing and development of genetic markers for the detection of DNA of invasive bighead and silver carp (Hypophthalmichthysnobilis and H. molitrix) in environmental water samples from the United States. PLoS One 2015;10:e0117803.

45. Fouquet B, Ghosn J, Quertainmont Y, Salmon D, Rioux C, Duvivier C, et al. Identification of variants of hepatitis C virus (HCV) entry factors in patients highly exposed to HCV but remaining uninfected: an ANRS case-control study. PLoS One 2015;0:e0142698.

46. Briscoe AG, Goodacre S, Masta SE, Taylor MI, Arnedo MA, Penney D, et al. Can long-range PCR be used to amplify genetically divergent mitochondrial genomes for comparative phylogenetics? A case study within spiders (Arthropoda: Araneae). PLoS One 2013;8:e62404.

47. Barker DL, Hansen MS, Faruqi AF, Giannola D, Irsula OR, Lasken RS, et al. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. Genome Res 2004;14:901–7.

48. Gribble S, Ng BL, Prigmore E, Burford DC, Carter NP. Chromosome paints from single copies of chromosomes. Chromosome Res 2004;12:143–51.

49. Chiou KL, Bergey CM. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci Rep 2018;8:1975.

50. Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Pinches R, et al. An effective method to purify *Plasmodium falciparum* DNA directly from clinical blood samples for whole genome high-throughput sequencing. PLoS One 2011;6:e22213.

51. Van Tonder AJ, Bray JE, Quirk SJ, Haraldsson G, Jolley KA, Maiden MC, et al. Putatively novel serotypes and the potential for reduced vaccine effectiveness: capsular locus diversity revealed among 5405 pneumococcal genomes. Microb Genom 2016;2:90.