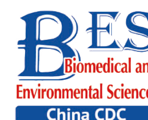
Original Article



Establishment and Modification of Ninety-seven Pneumococcal Serotyping Assays Based on Quantitative Real-time Polymerase Chain Reaction*

CHE Jie[1], CHEN Bo Han[1], XU Li[1], GAO Yuan[1], YUE Meng Meng[1,2], CHEN Zi Man[1], ZHANG Mao Jun[1], and SHAO Zhu Jun[1,2,#]

1. National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 2. School of Public Health, Nanjing Medical University, Nanjing 210029, Jiangsu China

Abstract

Objective To establish and modify quantitative real-time polymerase chain reaction (qPCR)-based serotyping assays to distinguish 97 pneumococcal serotypes.

Methods A database of capsular polysaccharide (*cps*) loci sequences was generated, covering 97 pneumococcal serotypes. Bioinformatics analyses were performed to identify the *cps* loci structure and target genes related to different pneumococcal serotypes with specific SNPs. A total of 27 novel qPCR serotyping assay primers and probes were established based on qPCR, while 27 recombinant plasmids containing serotype-specific DNA sequence fragments were constructed as reference target sequences to examine the specificity and sensitivity of the qPCR assay. A panel of pneumococcal reference strains was employed to evaluate the capability of pneumococcal serotyping.

Results A total of 97 pneumococcal serotyping assays based on qPCR were established and modified, which included 64 serotypes previously reported as well as an additional 33 serotypes. Twenty-seven novel qPCR serotyping target sequences were implemented in the pneumococcal qPCR serotyping system. A total of 97 pneumococcal serotypes, which included 52 individual serotypes and 45 serotypes belonging to 20 serogroups, could not be identified as individual serotypes. The sensitivity of qPCR assays based on 27 target sequences was 1–100 copies/μL. The specificity of the qPCR assays was 100%, which were tested by a panel of 90 serotypes of the pneumococcal reference strains.

Conclusion A total of 27 novel qPCR assays were established and modified to analyze 97 pneumococcal serotypes.

Key words: *Streptococcus pneumoniae*; Serotyping; *cps* loci; Quantitative real-time PCR (qPCR)

Biomed Environ Sci, 2023; 36(9): 787-799 *doi: 10.3967/bes2023.078* *ISSN: 0895-3988*

www.besjournal.com (full text) *CN: 11-2816/Q* *Copyright ©2023 by China CDC*

INTRODUCTION

Streptococcus pneumoniae (or pneumococcus) is a respiratory pathogen that affects humans. More than two serotypes of pneumococcus can coexist at the site of the naso-oropharynx. Pneumococcus is a bacterium that has been linked to a wide range of pneumococcal diseases (PD), ranging from non-invasive PD (NIPD), such as sinusitis, otitis media,

*This work was supported by a grant from Beijing Municipal Natural Science Foundation [L212011] and National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention [131031102000210003 & 102393230020020000002].

#Correspondence should be addressed to SHAO Zhu Jun, E-mail: shaozhujun@icdc.cn

Biographical note of the first author: CHE Jie, female, born in 1989, Master, majoring in prevention and control of infectious diseases.

and pneumonia, to serious invasive PD (IPD), particularly in children of age < 5 years and individuals aged > 65 years. Pneumococcus is the primary pathogen of community-acquired pneumonia, lower respiratory tract infections, and childhood pneumonia[1-3]. Capsular polysaccharide (CPS) is the most critical virulence factor of pneumococci, and > 100 serotypes of it have been identified and reported based on the biochemical and antigenic properties of the pneumococcal CPSs[4,5]. The pneumococcal conjugate vaccine (PCV) is highly effective in preventing IPD caused by vaccine serotypes (VTs). PCV vaccination can reduce the carriage rate as well as the transmission route of pneumococci among populations, which is referred to as herd immunity[6-9]. As the carriage rate of pneumococci VTs decreases, other serotypes (such as non-VTs [NVTs]) become prevalent, which is referred to as "serotype replacement". A multicenter study conducted in the United States revealed a 66% decrease in the overall (all serotypes) numbers of IPD cases among children aged < 2 years from the baseline (1994–2000) until 2002. However, the number of IPD cases involving NVT increased by 66% from the baseline (NVTs only)[10]. In the United States, dramatic reductions (45% and 94%, respectively) have been recorded in overall and PCV7-type IPD after PCV7 introduction in children aged < 5 years between 1998–1999 and 2007. However, the rate of IPD caused by NVTs increased by 128%[8]. Between 2001 and 2017, the cases of major NVTs in France increased from 10.5% to 41.8% for the overall IPD cases, with the numbers of cases increasing from 5.3% to 58.3% in children aged < 2 years and serotype 24F becoming the most predominant serotype of IPD in children aged < 2 years[11,12]. The protective effect of PCV is significantly compromised by an increase in NVTs resulting from "serotype replacement"[8,10,11]. "Serotype replacement" advocates the creation and application of next-generation pneumococcal vaccines and implies that serotype surveillance is essential for preventing and controlling PD.

Serotype-based serological approaches (or phenotypic methods) that rely on serotype-specific antisera or monoclonal antibodies (mAbs), molecular technologies based on *cps* loci such as sequential multiplex PCR (mPCR), multiplex qPCR (mqPCR), and whole-genome sequencing are already being used for bacterial serotyping in the age of genome sequencing, which includes pneumococcal serotyping[13-26]. Based on the reaction between antibodies against the pneumococcal capsule, the Quellung reaction is considered the gold standard for pneumococcal serotyping[13,14]. However, due to the high costs, experience dependency, and isolation of pure colonies, this method cannot be readily applied to most laboratories and is considered more appropriate for quality control in reference laboratories[13-16]. Owing to its practicality and instrument independency, the latex agglutination reaction has been suggested as the most practicable technique for pneumococcal serotyping[17,18]. Briefly, equal quantities of latex reagent and bacterial culture are mixed, and the resultant agglutination reaction is read within 5–10 s. However, there are numerous pneumococcal serotypes. According to the chessboard scheme, a single isolate should be typed by pool antisera first and then be further typed by type and factor antisera. However, due to the high costs of the pneumococcal latex agglutination kit (only 14 pools antisera kit up to $10,000, 75 tests/bottle), not all laboratories can afford it. Due to its simplicity, affordability, and lack of dependency on culture-based techniques, the PCR-based pneumococcal serotyping approach is frequently employed for pneumococcal surveillance[19-21]. Sequential mPCR and sequential mqPCR methodologies have been developed for the most prevalent serotypes in Active Bacterial Core Surveillance (ABCs) to track the dynamics of pneumococcal serotypes. The 64 serotypes identified using the sequential mqPCR so far include 34 individual serotypes and 13 minor serogroups[21]. Based on the prevalence of IPD serotypes across different countries and regions, similar strategies were studied, optimized, and applied in regions such as Latin America, Africa, and Asia[23,24]. However, previously reported PCR- or qPCR-based pneumococcal serotyping methods cover only limited serotypes, many of which have not yet been reported.

Currently, there is no national surveillance system for PD or IPD in China[27], and most information on pneumococcal serotype prevalence comes from individuals or regional hospital IPD patients[26,27]. Moreover, the serotype prevalence reported to guide the development of pneumococcal serotype strategies may be biased. Therefore, there is a need for systematic analysis of the *cps* loci in different pneumococcal serotypes for rapid and easy typing with a high serotype coverage.

In this study, we downloaded the reported *cps* loci sequences of pneumococcus and analyzed the structure and sequence of different serotypes by bioinformatics technology. A qPCR-based pneumococcal serotyping assay was developed and modified to cover 97 serotypes, which included 64

serotypes that have been previously reported and an additional 33 serotypes. A total of 27 novel serotyping target sequences were added to the previous serotyping system, which contained 46 serotyping target sequences. In addition, a total of 97 pneumococcal serotypes were typed into 52 individual serotypes and 20 serogroups. Here, we expounded on the process of sequence analysis and serotyping assay establishment and proved their feasibility in typing serotypes caused by *S. pneumoniae* using pure strain cultures. This approach is expected to facilitate the monitoring of the prevalence trend of pneumococcal serotypes and guide the development of pneumococcal vaccines and immunization strategies.

METHODS

Strains and DNA Extraction

A panel of pneumococcal reference strains containing 90 pneumococcal serotypes was applied to evaluate the capability of pneumococcal serotyping (Supplementary Table S1, available in www.besjournal.com). All pneumococcal reference strain serotypes are already known. These strains are stored at the Department of Respiratory Infectious Diseases, National Institute of Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention.

An overnight culture was prepared in Columbia agar plate supplemented with sheep blood (PB0123A, Thermo Fisher Oxoid, Basingstoke, UK) and incubated under 5% CO_2 at 37 °C. The genomic DNA of the pneumococcal reference strains was extracted using the "Wizard" Genomic DNA Purification Kit (A1120, Promega Corporation, Madison, USA). In order to extract the DNA better, the culture products were pre-treated as follows: overnight colonies from the Columbia agar plate supplemented with sheep blood were collected and resuspended in 200 µL of the TE buffer with 10-µL inoculation loops. Then, a 100-µL mixture of lysozyme (RT401, TIANGEN, Beijing, China) and mutanolysin (M9901, Sigma-Aldrich, Darmstadt, Germany) was added and digested at 37 °C for 1 h. The lysozyme and mutanolysin reaction concentrations were 40 mg/mL and 75 U/mL, respectively. All extracted DNA was stored at −20 °C until further use.

Reference Sequence and Bioinformatic Methods

The sequences of *cps* loci were downloaded from GeneBank (https://www.ncbi.nlm.nih.gov/). A database of *cps* loci sequences was generated, which covered 97 pneumococcal serotypes (Supplementary Table S2, available in www.besjournal.com). Based on the database, multiple sequence alignment was performed using the MAFFT v7.505. The maximum likelihood phylogenetic tree was generated with Fasttree 2.1.11[28], which contained 94 serotypes with a complete sequence of *cps* loci in the reference database. The tree is annotated as graphics using the iTol program[29]. Combined with the *cps* loci sequence annotation information, the gene structure diagrams were drawn with the RStudio 4.2.1. For potentially typed targets, multiple sequence alignment was performed with DNAMAN v9.0.

Real-time PCR Primers, Fluorescence Probes for Novel Serotyping Targets, and Real-time PCR Assay

Based on previously reported qPCR-based pneumococcal serotyping assays[21], novel serotyping targets sequence were selected to identify more individual serotypes or divide the strains into smaller serogroups. The primers and probes were assessed and selected by Primer Express 3.0.1. The GeneBank primer blast confirmed the sequence specificity of the primers and probes, and the probes with appropriate reporting dyes (FAM) and quenchers were synthesized at Sangon Biotech (Shanghai) and DIA-UP (Beijing). The target genes, primers, and probes sequences, as well as the optimal assay concentrations selected for serotyping, are shown in Supplementary Table S3, available in www.besjournal.com.

The qPCR reactions were performed in a 20-µL solution, with each reaction mixture containing the following: 2× concentration probe qPCR mix (RR391A, TaKaRa, Beijing, China), forward primer, reverse primer, probe, sterilized water, and 2-µL DNA template. The optimal reaction concentrations of each primer and probe in each reaction are shown in Supplementary Table S3. The LightCycler 480 II (Roche Diagnostics, Switzerland) and Quant Gene9660 (BIOER, China) were used to amplify and detect the target genes. Each qPCR run included a negative control and an external positive control. The amplification conditions were 95 °C for 600 s, followed by 40 quantitative cycles of 95 °C for 15 s and 60 °C for 60 s.

The Recombinant Plasmid Carrying Target Gene Sequence and A Standard Curve Plot of Recombinant Plasmids

The pUC57 cloning vectors (Sangon Biotech,

Shanghai) were used to generate recombinant reference plasmids containing serotype-specific DNA sequence fragments for quantification. A unique recombinant reference plasmid was constructed for each target gene. A total of 27 recombinant reference plasmids were generated for this study's 27 novel serotyping assays. Several plasmid copies were calculated using the value for the plasmid molecular weight and the following formula:

$$\text{Recombinant reference plasmid (copies/μL)} = \frac{\textit{Avogadros number} \times \textit{Recombinant reference plasmid concentration}\ (\text{ng/μL}) \times 10^{-9}}{660 \times \textit{Number of recombinant reference plasmid base pairs}}$$

The DNase/RNase-free water (RT121, TIANGEN, Beijing, China) was used to generate recombinant reference plasmid serial dilutions, and 10 concentration gradients from 10^0 to 10^9 copies/μL of recombinant reference plasmids were generated. The concentration-gradient generation was performed in triplicate for standard curve construction and sensitivity assessment of the serotyping assay.

Performance of This Pneumococcal Serotyping Assay

The experimental strains without the target gene were used as negative strains. Genomic DNA extracted from pneumococcal reference strains and 27 recombinant plasmids were used to evaluate the specificity of the qPCR assay.

A mixture of two serotypes in a simulation sample was used to evaluate the ability of multiple serotypes co-existing in the same sample of this pneumococcal serotyping assay. Genomic DNA (1,000 copies/μL) of serotype 18F and serotype 29 (fluorescent signal marker: FAM) and genomic DNA (1,000 copies/μL) of serotype 16A and serotype 17F (different fluorescent signal markers: FAM and HEX) were mixed at different ratios of 0%, 5%, 25%, 50%, 75%, 95%, and 100%, respectively. Three replicate analyses were performed using the pneumococcal serotyping assay established in this study. To evaluate the ability to detect the rare serotypes in the presence of multiple serotypes.

Statistical Analyses and Graphing

LightCycler®480 software (version 1.5.0), Quant Gene 9600 software and WPS office (version 11.1.012763) were used to calculate the standard curve, efficiency (%), and R^2, respectively. R Studio (version 4.2.1) was used for graphics drawing, which included the gene structure of pneumococcal *cps* loci and serotyping interpretation of pneumococcal cps loci, as annotated with Adobe Illustrator (version 25.2.1).

RESULTS

Genetic Basis of Pneumococcal cps loci

The phylogenetic tree and gene structure of 94 serotype pneumococcal *cps* loci sequences were constructed (Figure 1). All serotypes (except serotype 3) have polymerase (*wzy*) and flippase (*wzx*). The SNP-based phylogenetic tree was constructed for *wzy* and *wzx* to discriminate the capacity as pneumococcal serotyping targets (Supplementary Figure S1, available in www.besjournal.com). In addition to the previously reported qPCR-based pneumococcal serotyping targets using *wzy*, 17 novel serotyping target fragments by *wzy*, 7B/7C/40, 7B/7C/24F/24B/40, 16A, 17A, 19B/19C, 25F/25A/38, 27, 29, 32F/32A, 33C, 33B/33D, 35A/35C/42, 36, 41A/41F, 43, 45, and 47A, were added for pneumococcal serotyping. For serotypes that could not be distinguished by *wzy* and *wzx*, potential serotyping targets were selected based on the presence or absence of serotype-specific genes and sequence identity in different serotype *cps* loci. A total of 10 novel serotyping targets were added for 10B(*wcrG*), 10A/10C/34/35F/43(*wcrC*), 11F/15A/15B/15C/15F(*wchJ*), 12A/12B/46(*wciI*), 18B/18C/18F(*wciX*), 18F(*wcxM*), 19C(*wchU*), 18A/28A(*wciU*), 35F/34(*wcrO*), and 41F(*wcrX*). Twenty-seven novel qPCR serotyping target sequences were considered in addition to the pneumococcal qPCR serotyping system (Supplementary Table S3).

Establish and Modify 97 Pneumococcal Serotyping Assays

A total of 27 novel qPCR serotyping target sequences were supplemented in the pneumococcal qPCR serotyping system. Combined with previously reported 46 serotyping assays, there were 73 serotyping target sequences for pneumococcal serotyping. A total of 97 pneumococcal serotypes could be typed, of which 52 were identified as

individual serotypes, while the remaining 45 serotypes belonged to 20 serogroups (Figure 2, Supplementary Table S3). As shown in Figure 2, a pneumococcal serotyping process covering 97 serotypes is formed, and the strains could be typed by the serotyping assays corresponding to the serotype/serogroup. Of

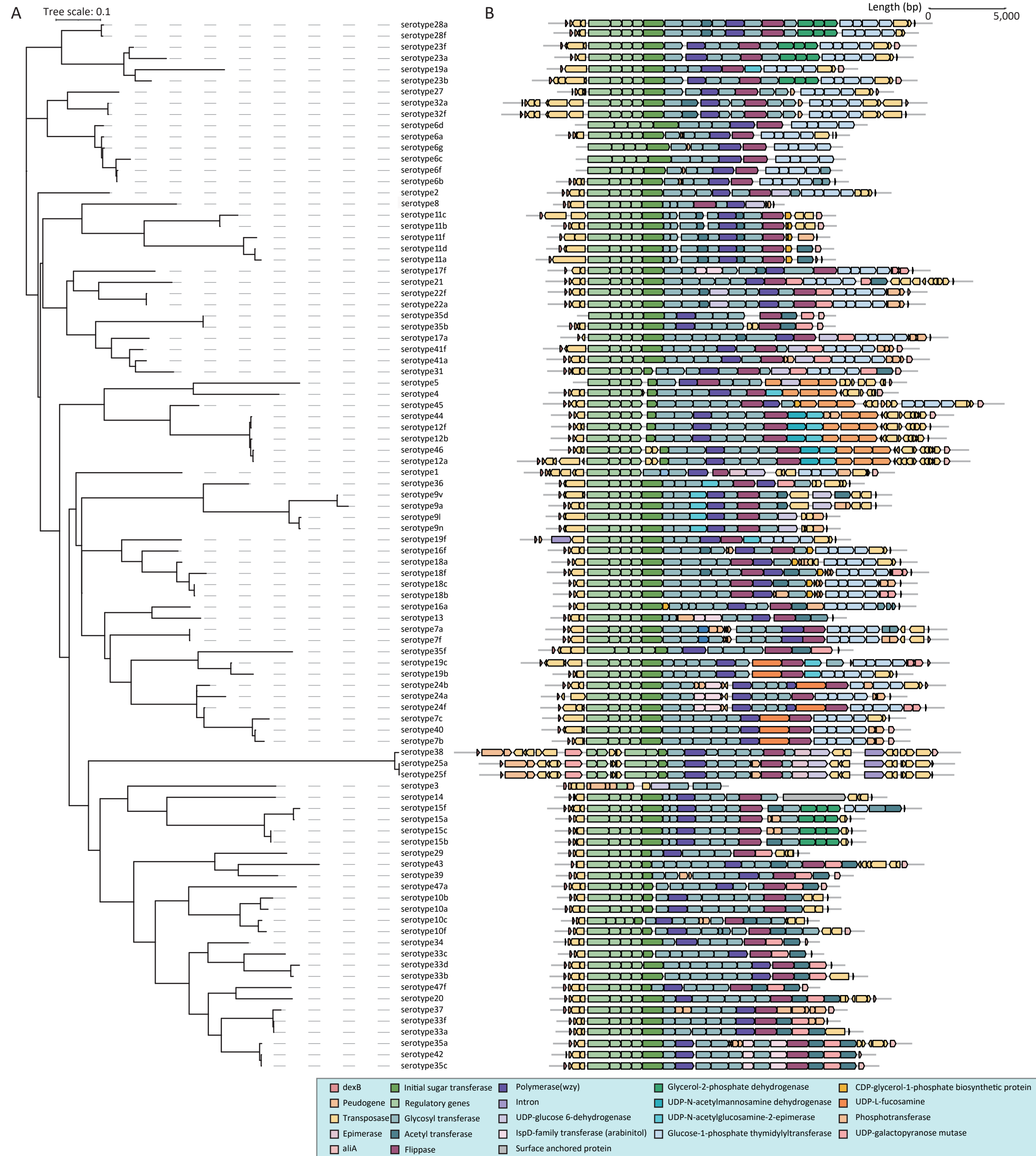


Figure 1. Maximum likelihood phylogenetic tree and the gene structure of pneumococcal *cps* loci. (A) Phylogenetic tree (Maximum likelihood phylogeny) of 94 serotype pneumococcal *cps* loci sequence; (B) The gene structure of 94 serotype pneumococcal *cps* loci sequence. All serotypes contained the regulatory region genes *wzg*, *wzh*, *wzd*, and *wze* (*cpsA-D*, light green) and serotype-specific region genes, beginning with an initial sugar transferase (emerald green). The gene structure of *S. pneumoniae cps* loci is depicted in alignment by *cpsA* (*wzg*). The arrows represent the direction of genes, while the colors represent different gene functions.

the 97 serotypes, 64 serotypes could be typed by only one reaction, of which 38 were identified as individual serotypes and 26 as 12 serogroups. A total of 16 novel typing assays were designed in this study. Twenty-three serotypes could be typed by 2 reactions, 10 by 3 reactions, and only 10C required identification by 4 reactions.

Sensitivity and Specificity of This Pneumococcal Serotyping Assay

A total of 27 novel recombinant reference plasmids with specific fragments were constructed in this study. Real-time PCR assays performed on these plasmids gave a limit of detection (LOD), the limit of quantitation (LOQ), and the unique linear dynamic range for each target gene and standard curve parameters are listed in Table 1. The parameter of the standard curve revealed a good relationship between the logarithmic value of recombinant plasmid diluted concentration and the Cq (quantification cycle) value. An excellent LOQs in a linear dynamic range of all recombinant plasmids, $10^1/10^2$–$10^1/10^9$ (Supplementary Figure S2, available in www.besjournal.com). The LODs ranged from 1 to 100 copies/reaction, indicating good detectability in this study.

All 27 recombinant reference plasmids and

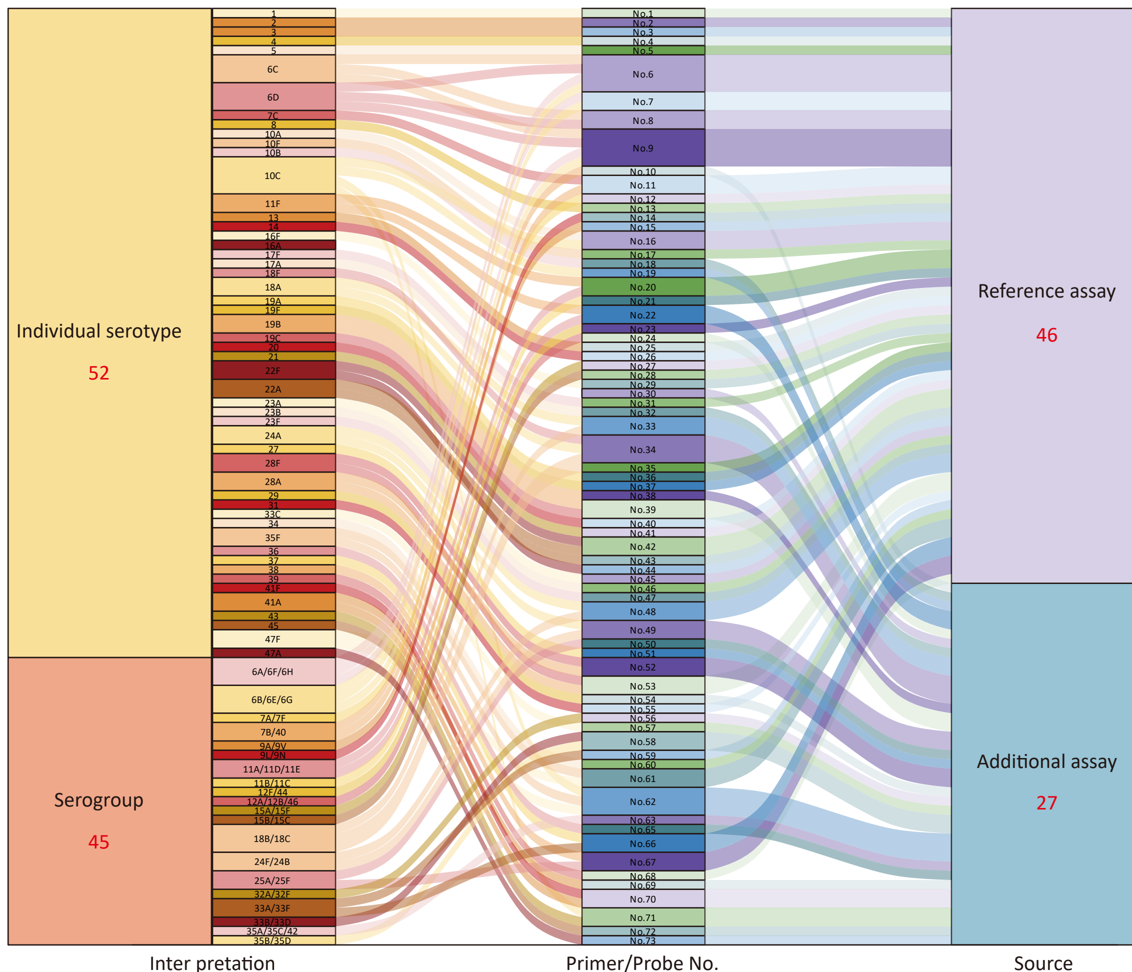


Figure 2. Primers/probes and serotyping interpretation of pneumococcus. (1) Interpretation: All 97 serotypes could be identified into 72 individual serotypes or small serogroups, of which 52 were identified as individual serotypes, and 45 serotypes belonged to 20 serogroups. (2) Primer/Probe No.: The primers/probes number of pneumococcal serotyping targets sequences. A total of 73 serotyping assays were used for 97 pneumococcal serotyping in this study. (3) Source: The source of primers/probes, 46 assays were previously reported (light purple)[20,21], and 27 additional assays were designed in this study (light blue). The flow from "Interpretation" to "Primer/Probe No." represents the primers/probes assays that were required to identify this serotype or serogroup.

Table 1. Recombinant reference plasmid parameters in this study

No.	Recombinant plasmids (containing serotype-specific target fragments)	Target gene (Accession No.)	Cloning vectors	Concentration		LOD	LOQ	Standard curve	Efficiency (%)	R^2	Instrument
				ng/μL	copies/μL						
1	7B/7C/40	*wzy* (CR931642)	pUC57	60.0	1.95×10^{10}	1.95×10^{1}	$10^{1} - 10^{9}$	y = −3.52x + 41.00	92.50	1	gene9660
2	10B	*wcrG* (CR931650)	pUC57	62.9	2.04×10^{10}	2.04×10^{1}	$10^{1} - 10^{9}$	y = −3.44x + 40.75	95.49	1	gene9660
3	10A/10C/34/35/43	*wcrC* (CR931652)	pUC57	84.5	2.76×10^{10}	2.76×10^{1}	$10^{1} - 10^{9}$	y = −3.22x + 41.28	100.24	1	Roche480
4	11F/15A/15B/15C/15F	*wchJ* (CR931657)	pUC57	71.9	2.32×10^{10}	2.32×10^{1}	$10^{1} - 10^{9}$	y = −3.54x + 40.70	91.50	0.998	gene9660
5	12A/12B/46	*wciI* (CR931658)	pUC57	56.7	1.86×10^{10}	1.86×10^{1}	$10^{1} - 10^{9}$	y = −3.56x + 41.17	90.77	1	gene9660
6	16A	*wzy* (CR931667)	pUC57	66.6	2.15×10^{10}	2.15×10^{1}	$10^{2} - 10^{9}$	y = −3.51x + 42.19	96.55	0.999	Roche480
7	17A	*wzy* (CR931669)	pUC57	66.9	2.15×10^{10}	2.15×10^{0}	$10^{2} - 10^{9}$	y = −3.49x + 41.92	96.15	0.999	Roche480
8	18B/18C/18F	*wciX* (CR931674)	pUC57	72.7	2.35×10^{10}	2.35×10^{1}	$10^{1} - 10^{9}$	y = −3.53x + 42.17	92.80	1	gene9660
9	18F	*wcxM* (CR931674)	pUC57	59.2	1.94×10^{10}	1.94×10^{1}	$10^{1} - 10^{9}$	y = −3.33x + 40.14	96.32	1	Roche480
10	19B/19C	*wzy* (CR931676)	pUC57	78.3	2.47×10^{10}	2.47×10^{1}	$10^{1} - 10^{9}$	y = −3.59x + 41.70	90.07	1	gene9660
11	19C	*wchU* (CR931677)	pUC57	55.3	1.77×10^{10}	1.77×10^{1}	$10^{1} - 10^{9}$	y = −3.43x + 39.93	97.90	0.999	Roche480
12	7B/7C/24F/24B/40	*wzy'* (CR931688)	pUC57	62.6	2.01×10^{10}	2.01×10^{1}	$10^{1} - 10^{9}$	y = −3.56x + 40.36	95.45	0.999	Roche480
13	25F/25A/38	*wzy* (CR931689)	pUC57	48.9	1.58×10^{10}	1.58×10^{0}	$10^{1} - 10^{9}$	y = −3.27x + 38.75	102.07	1	genen9660
14	27	*wzy* (CR931691)	pUC57	61.3	1.98×10^{10}	1.98×10^{0}	$10^{1} - 10^{9}$	y = −3.37x + 39.26	97.99	0.996	gene9660
15	18A/28A	*wciU* (CR931692)	pUC57	79.0	2.54×10^{10}	2.54×10^{1}	$10^{2} - 10^{9}$	y = −3.36x + 41.50	99.85	0.999	Roche480
16	29	*wzy* (CR931694)	pUC57	82.0	2.72×10^{10}	2.72×10^{1}	$10^{1} - 10^{9}$	y = −3.47x + 41.96	96.85	0.998	Roche480
17	32A/32F	*wzy* (CR931696)	pUC57	63.2	2.03×10^{10}	2.03×10^{0}	$10^{2} - 10^{9}$	y = −3.35x + 41.13	99.85	0.999	Roche480
18	33C	*wzy* (CR931700)	pUC57	61.8	1.20×10^{10}	1.20×10^{1}	$10^{1} - 10^{9}$	y = −3.40x + 39.77	96.65	1	gene9660
19	33B/33D	*wzy* (CR931699)	pUC57	64.0	2.06×10^{10}	2.06×10^{1}	$10^{2} - 10^{9}$	y = −3.28x + 39.45	100.91	1	gene9660
20	35F/34	*wcrO* (CR931707)	pUC57	75.0	2.46×10^{10}	2.46×10^{1}	$10^{2} - 10^{9}$	y = −3.24x + 39.14	101.80	0.999	Roche480
21	35A/35C/42	*wzy* (CR931706)	pUC57	97.2	3.13×10^{10}	3.13×10^{1}	$10^{2} - 10^{9}$	y = −3.30x + 39.13	96.15	0.999	Roche480
22	36	*wzy* (CR931708)	pUC57	43.3	1.39×10^{10}	1.39×10^{1}	$10^{1} - 10^{9}$	y = −3.47x + 39.89	94.00	1	gene9660
23	41A/41F	wzy (CR931713)	pUC57	38.8	1.19×10^{10}	1.19×10^{2}	$10^{2} - 10^{9}$	y = −3.79x + 43.51	91.80	0.999	Roche480
24	41F	*wcrX* (CR931714)	pUC57	59.3	1.90×10^{10}	1.90×10^{1}	$10^{1} - 10^{9}$	y = −3.29x + 41.77	100.75	0.999	Roche480
25	43	*wzy* (CR931716)	pUC57	45.1	1.43×10^{10}	1.43×10^{1}	$10^{1} - 10^{9}$	y = −3.34x + 40.95	99.36	1	gene9660
26	45	*wzy* (CR931718)	pUC57	44.5	1.38×10^{10}	1.38×10^{1}	$10^{2} - 10^{9}$	y = −3.62x + 42.66	94.90	0.999	Roche480
27	47A	wzy (CR931720)	pUC57	53.2	1.64×10^{10}	1.64×10^{1}	$10^{2} - 10^{9}$	y = −3.61x + 43.14	93.40	0.999	Roche480

Note. LOD, limit of detection; LOQ, the limit of quantitation.

reference strains demonstrated amplification and were identified as the accurate serotype/ subserogroup, while the remaining serotypes did not demonstrate any specific amplification curves.

A total of 73 pneumococcal serotyping targets sequence were used in this study (Figure 2, Supplementary Table S3, available in www. besjournal.com), and 27 novel additional serotyping assays were developed. The results of a panel of 90 serotypes of pneumococcal reference strains are shown in Table 2. All 90 serotypes of pneumococci were identified, of which 50 could be identified as individual serotypes and the remaining 40 to 20 small serogroups, respectively. The specificity of qPCR assays was 100%.

Table 2. Performance of serotyping assays in a panel of pneumococcal reference strains

NO.	Serotype (pneumococcal reference strains)	Reference assay only[20,21]		Reference assay and 27 additional assay	
		serotype	serogroup	serotype	serogroup
1	1	1		1	
2	2	2		2	
3	3	3		3	
4	4	4		4	
5	5	5		5	
6	6A		6A/6F/6H		6A/6F/6H
7	6B		6B/6E/6G		6B/6E/6G
8	6C	6C		6C	
9	6G		6B/6E/6G		6B/6E/6G
10	7F		7A/7F		7A/7F
11	7A		7A/7F		7A/7F
12	7B		7B/40		7B/40
13	7C	7C		7C	
14	8	8		8	
15	9A		9A/9V		9A/9V
16	9N		9L/9N		9L/9N
17	9V		9A/9V		9A/9V
18	9L		9L/9N		9L/9N
19	10A	10A		10A	
20	10F	10F		10F	
21	10B	Unidentified		10B	
22	10C	Unidentified		10C	
23	11F		11A/11D/11E/11F	11F	
24	11A		11A/11D/11E/11F		11A/11D/11E
25	11B		11B/11C		11B/11C
26	11C		11B/11C		11B/11C
27	11D		11A/11D/11E/11F		11A/11D/11E
28	12F		12F/44		12F/44
29	12A	Unidentified			12A/12B/46
30	12B	Unidentified			12A/12B/46
31	13	13		13	
32	14	14		14	
33	15A		15A/15F		15A/15F
34	15B		15B/15C		15B/15C
35	15C		15B/15C		15B/15C
36	15F		15A/15F		15A/15F
37	16F	16F		16F	
38	16A	Unidentified		16A	
39	17F	17F		17F	
40	17A	Unidentified		17A	

Continued

NO.	Serotype (pneumococcal reference strains)	Reference assay only[20,21]		Reference assay and 27 additional assay	
		serotype	serogroup	serotype	serogroup
41	18F		18A/18B/18C/18F	18F	
42	18A		18A/18B/18C/18F	18A	
43	18B		18A/18B/18C/18F		18B/18C
44	18C		18A/18B/18C/18F		18B/18C
45	19A	19A		19A	
46	19F	19F		19F	
47	19B	Unidentified		19B	
48	19C	Unidentified		19C	
49	20	20		20	
50	21	21		21	
51	22F	22F		22F	
52	22A	22A		22A	
53	23A	23A		23A	
54	23B	23B		23B	
55	23F	23F		23F	
56	24F		24F/24A/24B		24F/24B
57	24A		24F/24A/24B	24A	
58	24B		24F/24A/24B		24F/24B
59	25F	Unidentified			25A/25F
60	25A	Unidentified			25A/25F
61	27	Unidentified		27	
62	28F		28A/28F	28F	
63	28A		28A/28F	28A	
64	29	Unidentified		29	
65	31	31		31	
66	32F	Unidentified			32A/32F
67	32A	Unidentified			32A/32F
68	33F		33A/33F/37		33A/33F
69	33A		33A/33F/37		33A/33F
70	33B	Unidentified			33B/33D
71	33D	Unidentified			33B/33D
72	34	34		34	
73	35F		35F/47F	35F	
74	35A		35A/35C/42		35A/35C/42
75	35B		35B/35D		35B/35D
76	35C		35A/35C/42		35A/35C/42
77	36	Unidentified		36	
78	37	37		37	
79	38	38		38	
80	39	39		39	
81	40	Unidentified			7B/40
82	41F	Unidentified		41F	
83	41A	Unidentified		41A	
84	42		35A/35C/42		35A/35C/42
85	43	Unidentified		43	
86	44		12F/44		12F/44
87	45	Unidentified		45	
88	46	Unidentified			12A/12B/12C/46
89	47F		35F/47F	47F	
90	47A	Unidentified		47A	

Detection Ability of the Coexistence of This Pneumococcal Serotyping Assay

The detection capability of the pneumococcal serotyping assay for the coexistence of multiple serotypes has demonstrated that this assay has good multi-serotype analytical capability. A relatively rare serotype, as low as 5% of the simulation DNA samples, can be detected, regardless of whether the target is the same or from different fluorescent channels.

DISCUSSION

PD is a global public health concern with a serious burden of IPD. However, there are several problems and challenges in the research and prevention of PD[12]. There has been a persistent increase in the proportion of IPD caused by NVTs of pneumococci as well as that in the rate of naso-oropharynx carriage of NVTs after vaccine application, with a naturally higher carriage prevalence of NVTs in some regions and certain populations in the pre-PCV application[30]. As the basis for PD research and prevention, the existing pneumococcal serotyping methods have several shortcomings. Therefore, there is a need to establish a rapid, accurate, high-throughput, and high serotype coverage of culture-independent pneumococcal serotyping methods for the research and prevention of PD.

During the COVID-19 pandemic, nucleic acid detection, especially qPCR, became critical to rapidly diagnosing infectious diseases. Recently, PCR- and qPCR-based methods were widely used for pneumococcal serotyping. Seven consecutive quadruple PCR reactions designed for the 29 most common pneumococcal serotypes in ABCs were employed for serotyping 421 IPD isolates, and 54.3% (229/421) of the isolates were successfully identified to serotypes, and 40.9% (172/421) of the strains were identified to the serogroup[19]. Subsequently, mqPCR-based pneumococcal serotyping strategies were recommended for ABCs serotype surveillance[21]. This mqPCR typing method was applied for 64 serotypes, of which 34 could be typed into individual serotypes and the remaining 30 into 13 small serogroups. Pholwat et al. developed a TaqMan array card for pneumococcal serotyping, and 53 sequence-specific PCR reactions were performed to identify 74 serotypes/serogroups[31]. Sakai et al. developed 11 single-plex qPCR serotyping assays, which, when combined with past studies of qPCR-based serotyping systems, expanded the coverage of serotypes to 94 serotypes. However, this study did not clarify the correspondence between 94 serotypes of pneumococci and the serotyping assays, nor did it establish a systematic analysis and result interpretation scheme, making it difficult to promote its application[32]. The nanofluidic real-time PCR serotyping method, a high-throughput serotyping method based on qPCR, was also employed for pneumococcal serotyping[33,34]. A total of 29 primer pairs were selected for the assay to cover 50 serotypes, 17 individual serotypes, and 33 serotypes in 12 serogroups[33]. This nanofluidic method was also designed for serotyping PCV-associated serogroups 6, 18, and 22, which are pneumococci with highly similar *cps* loci[35]. However, the reported qPCR-based pneumococcal serotyping methods currently lack systematic analysis and have limited coverage of serotypes, especially for NVTs. Several countries and regions, such as U.S., Latin-Amer, and Asian counties, have established their pneumococcal serotyping schemes based on the prevalence of predominant serotypes, seeking to cover more prevalent serotypes with fewer reactions[23,24]. However, there is a lack of a national surveillance network for IPD in China, and the distribution of pneumococcal serotypes in China is not adequately known. The serotype prevalence and VTs coverage varied by age group, source, and region. In Zhongjiang County[36], where the PCV13 vaccination rate was meager in 2018–2020, the proportion of non-PCV13 type increased remarkably with an increase in age, 28.7% at 2 years of age to 58.1% at ≥ 60 years of age. Therefore, enhanced detection and surveillance of all serotypes, including VTs, and NVTs, is vital for preventing and controlling PD. We accordingly extended and modified the previously reported typing assays to establish a typing system that covers the majority of pneumococcal serotypes.

The pneumococcal *cps* loci sequences are the genetic basis for PCR-based serotyping methods. The genetic analysis of pneumococcal *cps* loci sequence has already been studied[37]. When compared with the previous analysis, this study focused on the selection and modification of serotyping target sequences and the establishment of serotyping assays. When selecting specific targets, *wzy* and *wzx* were first considered pneumococcal serotyping targets because they were required to be serotype-specific, especially *wzy*[38]. When the serogroup/serotype was hard to distinguish with *wzy* and *wzx*, the presence or absence of genes in serotype-specific regions in *cps* loci or the difference in their sequence in gene

structure analysis for further typing. With this sequential flow of serotyping, a pneumococcal serotyping process covering 97 serotypes was formed. To the best of our knowledge, this is the first time that a phylogenetic tree of the *cps* loci sequence, covering the sequences of 94 pneumococcal serotypes, was used in combination with gene structure to investigate the differences in gene composition and sequence identity in different serotypes and to select suitable serotyping targets. In addition, in this study, we added five pneumococcal serotypes *cps* loci sequences (6C, 6D, 6F, 6G, and 35D) to the previous gene structure analysis. Regrettably, serotypes 6E, 6H, and 11E could not be included in the analysis of phylogeny and gene structure because these serotype strains were not available in our laboratory, and their *cps* loci sequence could not be accessed from the public databases.

Another contribution of this study is the establishment and modification of a qPCR-based serotyping scheme, which proposed a series of assays covering 97 pneumococcal serotypes (containing 6E, a genetically defined entity). A total of 73 serotyping target sequences were involved, and 27 novel qPCR serotyping target sequences were added to extend the previously reported 46 qPCR serotyping assays by ABCs[21]. Moreover, 97 pneumococcal serotypes were typed into 72 serotypes or serogroups, of which 52 could be typed as individual serotypes, and the remaining 45 serotypes were identified as 20 sub-serogroups. Critically, the systematic reorganization and analysis enabled the serotyping logic and the interpretation of typing results to be more precise than that in any of the previous reports. Genetic structure analysis and the graphs of pneumococcal serotyping interpretation facilitate the reading of results for the numerous serotypes and assays. Furthermore, the covers of serotypes can be extended in this case. Moreover, on the basis of the increased serotype coverage, some strains that could be identified as serogroups were further identified as smaller serogroups and even as individual serotypes. The strains that could be identified as serogroup18A/18B/18C/18F by the previous assays could be further typed to 18A, 18F, or 18B/18C with two novel serotyping assays (i.e., *wcxM*, and *wciX*). Unlike the pneumococcal serogroup 18 serotyping scheme established by Downs et al.[35], which relies on four targets, contains *wciW* (18A/B/C), *wciX* (18B/C/F, and 18C/F), and *wcxM* (16F/18F/28AF), pneumococcal serogroup 18 serotyping in this study followed the use of *wzy* to select for serogroup18 strains before differentiating 18A, 18F, and 18B/C by *wciX* (18B/18C/18F) and *wcxM* (18F). Serogroup 18 is a Wzy-dependent serotype, and *wzy* has superiority in serotyping specified and broad application. On the other hand, this 18F *wcxM* specifically target fragment can be used as a separate identification target of serotype 18F without cross-reactivity with other serogroups. The disadvantage is that the serotyping assay of serogroup18 in this study has not separated 18B and 18C completely. Overall, the additional novel serotyping target sequence added to this study has increased the coverage of serotypes and the number of accurately identified individual serotypes. The number of serotypes covered was increased by 33, and 18 of them could be accurately identified as serotypes.

The performance of the 27 novel serotyping assays in pneumococcal serotyping applications is another key to their more widespread application. The LODs ranged from 1 copy/reaction to 10 copies/reaction for all 27 novel additional assays established in this study, except for No. 69 (41A/41F) with a strains LOD of 1.19×10^2 copies/reaction. The LODs were equal to those previously reported[21,31], demonstrating the excellent sensitivity of the typing assay in this study. A panel of 90 serotypes of pneumococcal reference was typed. All pneumococcal strains were typed correctly. The 27 novel additional serotyping assays expanded the identified serotypes from 66 to all 90 strains. The number of serotypes that could be typed into individual serotypes was increased from 38 serotypes to 50 serotypes. The analyses of reference strains revealed that the pneumococcal serotyping assays were specific and accurate for pneumococcal serotyping. The relatively rare serotypes were detected at levels as low as 5% in the coexistence simulation samples, demonstrating the ability of the method to detect multiple serotypes.

A total of 97 pneumococcal could be typed in this study. Limited by the resource of strains, 7 serotypes (including the serotypes 6D, 6E, 6F, 6H, 11E, 33C, and 35D) have not been typed by isolates. However, our pneumococcal serotyping assays could cover these serotypes by sequence analyses of pneumococcal *cps* loci. Except for the serotype 33C, the serotyping targets sequence used for serotyping the remaining 6 serotypes shared gene fragments with the existing strains. For instance, serotype 6E shares a serotyping target sequence with serotypes 6B and 6G, which could be identified as serogroup 6B/6E/6G by Primer/Probe No. 6 (6A/6B/6C/6D/6E/6F/6G/6H, *wciP*), No. 7 (6A/6B/6E/6F/

6G/6H, *wciN*), and No. 9 (6B/6D/6E/6G, *wciP*). Serotype 6D was typed with the Primer/Probe No.6 (6A/6B/6C/6D/6E/6F/6G/6H, *wciP*), No. 8 (6C/6D, *wciN*), and No. 9 (6B/6D/6E/6G, *wciP*). These serotyping target sequences were with the same for 6A/6B/6C/6D/6E/6F/6G/6H, 6C/6D, and 6B/6D/6E/6G, respectively. The specificity of No. 57 (33C, *wzy*) serotyping assay was tested by a recombinant reference plasmid containing a fragment of the target sequence for serotype 33.

The selection of serotyping targets relied on the database of *cps* loci sequences, which were downloaded from the public database or previously reported. Although the pneumococcal *cps* loci sequence is generally considered to be conserved; however, slight modifications of the *cps* loci, such as the accumulation of point mutations, insertions, or deletions of genes, may cause alterations in the serotypes. The database of pneumococcal *cps* loci sequence and the pneumococcal serotyping target sequences might need to be continuously expanded and modified to meet the needs of future pneumococcal serotyping.

The pneumococcal serotyping assays established in this study were based on single-weight qPCR reactions, and the evaluation of the integration into a continuous multiplex qPCR assay system has not been completed. Subsequent analyses will continue to optimize the test protocol with a higher throughput and serotype coverage by combining nanofluidic gene chips and MeltArray[39,40]. After obtaining the representative prevalence and carriage surveillance data, we aim to develop a serial multiplex qPCR assay protocol that meets the prevalence characteristics of pneumococcal serotypes in China based on the prevalence of pneumococcal serotypes in China. Another limitation of this study is that the evaluated strains of specificity may be insufficient, and the designed assay was not evaluated for detecting *S. pneumoniae* in human clinical samples, such as nasopharyngeal samples, sputum, blood, and CFS.

As we know that 90 strains of isolates may not be adequate, we have mentioned it as a limitation of the study.

Despite these limitations, our current findings and reports expand the current qPCR scheme and offer a practical strategy for pneumococcal serotyping.

CONCLUSION

In conclusion, a total of 27 novel qPCR assays were established and modified to analyze 97 pneumococcal serotypes.

Received: November 9, 2022;
Accepted: January 9, 2023

REFERENCES

1. Wahl B, O'Brien KL, Greenbaum A, et al. Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. Lancet Glob Health, 2018; 6, e744−57.
2. Musher DM, Thorner AR. Community-acquired pneumonia. N Engl J Med, 2014; 371, 1619−28.
3. Advisory Committee on Immunization Practices. Preventing pneumococcal disease among infants and young children. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep, 2000; 49, 1−35.
4. Geno KA, Gilbert GL, Song JY, et al. Pneumococcal capsules and their types: past, present, and future. Clin Microbiol Rev, 2015; 28, 871−99.
5. GPS. Serotypes. https://www.pneumogen.net/gps/serotypes.html. [2022-08-11].
6. Moore MR, Link-Gelles R, Schaffner W, et al. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: analysis of multisite, population-based surveillance. Lancet Infect Dis, 2015; 15, 301−9.
7. Ladhani SN, Collins S, Djennad A, et al. Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000-17: a prospective national observational cohort study. Lancet Infect Dis, 2018; 18, 441−51.
8. Pilishvili T, Lexau C, Farley MM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis, 2010; 201, 32−41.
9. Andrews N, Kent A, AMIN-Chowdhury Z, et al. Effectiveness of the seven-valent and thirteen-valent pneumococcal conjugate vaccines in England: the indirect cohort design, 2006-2018. Vaccine, 2019; 37, 4491−8.
10. Kaplan SL, Mason EO Jr, Wald ER, et al. Decrease of invasive pneumococcal infections in children among 8 children's hospitals in the United States after the introduction of the 7-valent pneumococcal conjugate vaccine. Pediatrics, 2004; 113, 443−9.
11. Levy C, Varon E, Ouldali N, et al. Changes in invasive pneumococcal disease spectrum after 13-valent pneumococcal conjugate vaccine implementation. Clin Infect Dis, 2020; 70, 446−54.
12. Ouldali N, Varon E, Levy C, et al. Invasive pneumococcal disease incidence in children and adults in France during the pneumococcal conjugate vaccine era: an interrupted time-series analysis of data from a 17-year national prospective surveillance study. Lancet Infect Dis, 2021; 21, 137−47.
13. Reasonover A, Zulz T, Bruce MG, et al. The international circumpolar surveillance interlaboratory quality control program for *Streptococcus pneumoniae*, 1999 to 2008. J Clin Microbiol, 2011; 49, 138−43.
14. Sørensen UB. Typing of pneumococci by using 12 pooled antisera. J Clin Microbiol, 1993; 31, 2097−100.
15. Lovgren M, Talbot JA, Brandileone MC, et al. Evolution of an international external quality assurance model to support

laboratory investigation of *Streptococcus pneumoniae*, developed for the SIREVA project in Latin America, from 1993 to 2005. J Clin Microbiol, 2007; 45, 3184−90.
16. Konradsen HB. Validation of serotyping of *Streptococcus pneumoniae* in Europe. Vaccine, 2005; 23, 1368−73.
17. Kuch A, Gołębiewska A, Waśko I, et al. Usefulness of Pneumotest-Latex for direct serotyping of *Streptococcus pneumoniae* Isolates in clinical samples. J Clin Microbiol, 2014; 52, 2647−9.
18. Slotved HC, Kaltoft M, Skovsted IC, et al. Simple, rapid latex agglutination test for serotyping of pneumococci (Pneumotest-Latex). J Clin Microbiol, 2004; 42, 2518−22.
19. 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.
20. Pimenta FC, Roundtree A, Soysal A, et al. Sequential triplex real-time PCR assay for detecting 21 pneumococcal capsular serotypes that account for a high global disease burden. J Clin Microbiol, 2013; 51, 647−52.
21. Velusamy S, Tran T, Mongkolrattanothai T, et al. Expanded sequential quadriplex real-time polymerase chain reaction (PCR) for identifying pneumococcal serotypes, penicillin susceptibility, and resistance markers. Diagn Microbiol Infect Dis, 2020; 97, 115037.
22. Leung MH, Bryson K, Freystatter K, et al. Sequetyping: serotyping *Streptococcus pneumoniae* by a single PCR sequencing strategy. J Clin Microbiol, 2012; 50, 2419−27.
23. National Center for Immunization and Respiratory Diseases. Multiplex conventional PCR schemes for pneumococcal serotype deduction. https://www.cdc.gov/streplab/pneumococcus/resources.html. [2022-08-21].
24. Park D, Kim SH, Bae IK, et al. Evaluation of modified sequential multiplex PCR for *Streptococcus pneumoniae* serotyping. Jpn J Infect Dis, 2019; 72, 224−7.
25. Garcia-Garcia S, Perez-Arguello A, Henares D, et al. Rapid identification, capsular typing and molecular characterization of *Streptococcus pneumoniae* by using whole genome nanopore sequencing. BMC Microbiol, 2020; 20, 347.
26. Lyu S, Hu HL, Yang YH, et al. A systematic review about *Streptococcus pneumoniae* serotype distribution in children in mainland of China before the PCV13 was licensed. Expert Rev Vaccines, 2017; 16, 997−1006.
27. Chinese Preventive Medicine Association, Vaccine and Immunology Branch of the Chinese Preventive Medicine Association. Expert consensus on immunoprophylaxis of pneumococcal disease (2020 version). Chin J Prev Med, 2020; 54, 1315−63. (In Chinese)
28. Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS One, 2010; 5, e9490.
29. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res, 2016; 44, W242−5.
30. Tvedskov ESF, Hovmand N, Benfield T, et al. Pneumococcal carriage among children in low and lower-middle-income countries: A systematic review. Int J Infect Dis, 2022; 115, 1−7.
31. Pholwat S, Sakai F, Turner P, et al. Development of a TaqMan Array Card for Pneumococcal Serotyping on Isolates and Nasopharyngeal Samples. J Clin Microbiol, 2016; 54, 1842−50.
32. Sakai F, Sonaty G, Watson D, et al. Development and characterization of a synthetic DNA, NUversa, to be used as a standard in quantitative polymerase chain reactions for molecular pneumococcal serotyping. FEMS Microbiol Lett, 2017; 364, fnx173.
33. Dhoubhadel BG, Yasunami M, Yoshida LM, et al. A novel high-throughput method for molecular serotyping and serotype-specific quantification of Streptococcus pneumoniae using a nanofluidic real-time PCR system. J Med Microbiol, 2014; 63, 528−39.
34. Kakiuchi S, Suzuki M, Dhoubhadel BG, et al. Accuracy of High-Throughput Nanofluidic PCR-Based Pneumococcal Serotyping and Quantification Assays Using Sputum Samples for Diagnosing Vaccine Serotype Pneumococcal Pneumonia: Analyses by Composite Diagnostic Standards and Bayesian Latent Class Models. J Clin Microbiol, 2018; 56, e01874−17.
35. Downs SL, Madhi SA, Van der Merwe L, et al. High-throughput nanofluidic real-time PCR to discriminate Pneumococcal Conjugate Vaccine (PCV)-associated serogroups 6, 18, and 22 to serotypes using modified oligonucleotides. Sci Rep, 2021; 11, 23728.
36. Du QQ, Zeng HL, Yuan L, et al. One cross-sectional investigation revealed that non-vaccine serotypes of Streptococcus pneumoniae could be identified more frequently in elderly Chinese people. Vaccine, 2021; 39, 3304−9.
37. Bentley SD, Aanensen DM, Mavroidi A, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet, 2006; 2, e31.
38. Kong F, Wang W, Tao J, et al. 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, 2005; 54, 351−6.
39. Che J, Lu JX, Li WG, et al. A New High-throughput Real-time PCR Assay for the Screening of Multiple Antimicrobial Resistance Genes in Broiler Fecal Samples from China. Biomed Environ Sci, 2019; 32, 881−92.
40. Huang Q, Chen D, Du C, et al. Highly multiplex PCR assays by coupling the 5'-flap endonuclease activity of Taq DNA polymerase and molecular beacon reporters. Proc Natl Acad Sci U S A. 2022; 119, e2110672119.