Journal of Microbiology, Immunology and Infection 55 (2022) 870—879

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jmii.com

Original Article

Rapid identification of *Streptococcus pneumoniae* serotypes by *cpsB* gene-based sequetyping combined with multiplex PCR

Meng-Lan Zhou [a,b,1], Zi-Ran Wang [a,b,1], Yan-Bing Li [a,b,1], Timothy Kudinha [c,d], Jian Wang [e], Yao Wang [a,b], Meng Xiao [a,b], Ying-Chun Xu [a,b,*], Zheng-Yin Liu [f,], Po-Ren Hsueh [g,h,i,j,***]**

[a] *Department of Clinical Laboratory, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China*
[b] *Beijing Key Laboratory for Mechanisms Research and Precision Diagnosis of Invasive Fungal Diseases, Beijing, China*
[c] *Charles Sturt University, Leeds Parade, Orange, NSW, Australia*
[d] *Regional and Rural, NSW Health Pathology, Orange Hospital, NSW 2800, Australia*
[e] *Beijing Center for Disease Prevention and Control, Beijing, China*
[f] *Department of Infectious Disease, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China*
[g] *Departments of Laboratory Medicine and Internal Medicine, China Medical University Hospital, Taichung, Taiwan*
[h] *School of Medicine, China Medical University, Taichung, Taiwan*
[i] *Ph.D Programme for Aging, School of Medicine, China Medical University, Taichung, Taiwan*
[j] *Departments of Laboratory Medicine and Internal Medicine, National Taiwan University Hospital, National Taiwan University College of Medicine, Taipei, Taiwan*

Received 22 April 2021; received in revised form 7 November 2021; accepted 20 November 2021
Available online 29 November 2021

KEYWORDS
Streptococcus pneumoniae;

Abstract *Background/purpose: Streptococcus pneumoniae* is an important human pathogen that causes invasive infections in adults and children. Accurate serotyping is important to study its epidemiological distribution and to assess vaccine efficacy.

* Corresponding author. Department of Clinical Laboratory, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Number 1, Shuaifuyuan Road, Dongcheng District, Beijing, China. Fax: +86 10 69159766.
** Corresponding author. Department of Infectious Disease, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Number 1, Shuaifuyuan Road, Dongcheng District, Beijing, China.
*** Corresponding author. Departments of Laboratory Medicine and Internal Medicine, National Taiwan University Hospital, Number 7, Chung-Shan South Road, Taipei, 100, Taiwan.
E-mail addresses: xycpumch@139.com (Y.-C. Xu), zhengyinl@hotmail.com (Z.-Y. Liu), hsporen@gmail.com (P.-R. Hsueh).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jmii.2021.11.004

Serotype;
Quellung reaction;
cpsB sequetyping;
Multiplex PCR

Methods: Invasive *S. pneumoniae* isolates (n = 300) from 27 teaching hospitals in China were studied. The Quellung reaction was used as the gold standard to identify the *S. pneumoniae* serotypes. Subsequently, multiplex PCR and *cpsB* gene-based sequetyping methods were used to identify the serotypes.
Results: Based on the Quellung reaction, 299 *S. pneumoniae* isolates were accurately identified to the serotype level and 40 different serotypes were detected. Only one strain was non-typeable, and five most common serotypes were identified: 23F (43, 14.3%), 19A (41, 13.7%), 19F (41, 13.7%), 3 (31, 10.3%), and 14 (27, 9.0%). Overall, the multiplex PCR method identified 73.3 and 20.7% of the isolates to the serotype and cluster levels, respectively, with 1.7% of the isolates misidentified. In contrast, the *cpsB* sequetyping method identified 59.0 and 30.3% of the isolates to the serotype and cluster levels, respectively, and 7% were misidentified.
Conclusions: The *cpsB* gene sequetyping method combined with multiplex PCR, can greatly improve the accuracy and efficiency of serotyping, besides reducing the associated costs.

Introduction

Streptococcus pneumoniae, also known as pneumococcus, is a gram-positive extracellular pathogen. It is mainly transmitted through the respiratory tract into the nasal cavity where it colonizes the epithelial cells of the nasal mucosa. When human immunity declines, *S. pneumoniae* can cause non-invasive pneumococcal disease (NIPD), such as otitis media and sinusitis, which can then spread to the lungs and cause pneumococcal pneumonia.[1,2] *S. pneumoniae* also can cause serious invasive pneumococcal diseases (IPD), such as sepsis and meningitis, with high morbidity and mortality rates of 18 and 33%, respectively, especially in developing countries.[3] A recent study has shown that COVID-19 has a poor prognosis when patients are co-infected with *S. pneumoniae*.[4] Because of the organism's recognized clinical pathogenicity, the rapid identification and differentiation from closely related commensal species is essential. In our previous study, we systematically compared the ability of different MALDI-TOF systems to distinguish species within the viridans group streptococci (VGS) and proposed an algorithm based on MALDI-TOF analysis combined with 16S rRNA and *gyrB* gene sequencing to identify common VGS species.[5] Based on this identification algorithm we could differentiate *S. pneumoniae* from other VGS species rapidly and efficiently. However, simply identifying *S. pneumoniae* is not sufficient, as different *S. pneumoniae* serotypes are associated with geographical prevalence, attack rate, age distribution, tendency to cause outbreaks, and propensity to acquire antimicrobial resistance genes.[6] The capsular polysaccharide (CPS) of *S. pneumoniae* is the most important factor involved in its pathogenesis, which facilitates the organism evade host immune killing and complement-mediated modulation of phagocytosis *in vivo*. The CPS contributes to bacterial survival in the host and promotes the development of IPD. The CPS is also the basis of *S. pneumoniae* serotyping.[7,8] To date, more than 90 *S. pneumoniae* serotypes have been identified based on differences in pneumococcal capsular polysaccharides.[9] The administration of vaccines covering common serotypes is an effective measure to prevent *S. pneumoniae* infection.[10] However, the proportion of IPD caused by uncovered serotypes of vaccines is gradually increasing, commonly referred to as "serotype replacement" in *S. pneumoniae*.[9] Therefore, accurate identification of *S. pneumoniae* serotypes and monitoring their distribution are of great significance for the control of invasive infections caused by this organism.

At present, the Quellung reaction remains the gold standard for *S. pneumoniae* serotyping, which is based on the antigenic specificity of different CPSs. The capsular-specific antibody binds to the corresponding capsular polysaccharide, resulting in capsular swelling, which can be observed under a microscope. Although this method has high sensitivity and specificity, it requires specific antisera, which are very expensive and have a longer delivery period, limiting its use in less economically developed areas. Meanwhile, the use of molecular and immunological typing techniques has gained prominence. In 2012, the United States Centers for Disease Control and Prevention (CDC) published a protocol for the detection of *S. pneumoniae* serotypes using multiplex PCR technology.[11] This method is economical, fast, and flexible, with eight reactions in total and can detect more than 40 serotypes. Many researchers have developed customized combinations based on this method as an alternative to the capsular swelling test.[12—14] The CPS regulatory gene *cpsA* is relatively conserved across all serotypes, while the presence of serotype-associated polymorphisms in *cpsB* is the basis for a molecular biology-based approach to serotyping.[7] The DNA sequencing method based on the *cpsB* gene established by Leung et al.[15] could putatively amplify 84 serotypes and differentiate 46 serotypes. This method is simple and cost-effective and can potentially identify more serotypes in the future. However, the method is not sufficiently accurate and comprehensive as the data are derived from the uncurated GenBank database. In 2016, Jin et al.[16] improved the *cpsB* gene sequence database corresponding to different serotypes. In that study, combining *cpsB* sequetyping and selective multiplex

PCR resulted in 98.9% (191/193) of the isolates being accurately identified to the serotype/cluster level. The *cpsB* sequetyping database developed by Jin et al. has more serotype sequences, leading to more accurate serotype prediction, compared to that developed by Leung et al. However, this database has not been evaluated by other researchers.

In this study, a total of 300 invasive *S. pneumoniae* isolates from 27 teaching hospitals in China (2010—2015) were studied. Using the Quellung reaction as a gold standard, we evaluated and compared the accuracy of multiplexed PCR and *cpsB* gene-based sequetyping in *S. pneumoniae* serotyping.

Materials and methods

Bacterial strain collection

A total of 300 invasive *S. pneumoniae* isolates were obtained from 27 teaching hospitals in China between January 2010 and October 2015. Only one strain per patient was included in this study. The isolates were identified by morphology, optochin tests, and MALDI-TOF-MS analysis. All isolates were confirmed as *S. pneumoniae* by Vitek MS (bioMérieux, France) and Bruker Biotyper (Daltonics, Germany). The single identification result showed a confidence value of 99.9% for Vitek MS and a confidence score of $\geq$2.0 for Bruker Biotyper.

Latex agglutination test

The latex agglutination test was performed using the *S. pneumoniae* detection emulsion kit (National Serum Institute, Copenhagen, Denmark) according to the manufacturer's instructions. The main groups were distinguished by a simple checkerboard typing system.[17] A bacterial suspension of 0.5 McFarland was prepared using fresh cultured pure colonies with normal saline, and 10 μL of bacterial droplets was added to clean slides. The same amount of latex reagent was then added and mixed evenly, and agglutination was immediately observed visually. The Quellung reaction was then performed to further determine the serotypes.

Quellung reaction

A 0.5 McFarland bacterial suspension was prepared for each isolate, and 1.5 μL of the bacterial suspension was added to a clean glass slide. An equal volume of antiserum (National Serum Institute in Copenhagen, Denmark) was added and mixed with the bacterial suspension. The bacterial size, capsular swelling, and aggregation were observed using an oil immersion microscope.

DNA extraction

The isolates were cultured on blood agar plates (Columbia II agar base supplemented with 5% horse blood) and incubated overnight at 35 °C in a 5% CO_2 atmosphere. DNA was extracted using the AxyGenamp DNA Mini Extraction Kit (Axygen, USA) according to the manufacturer's instructions. The final pure DNA was stored at −20 °C until use.

Multiplex PCR

Sequential eight-step multiplex PCR was performed according to the protocol[18] recommended by the US CDC to identify the serotypes based on the specific amplicon length size. Ten microliters of each reaction mixture were separated by 1.5% agarose gel electrophoresis and visualized with ethidium bromide staining and UV illumination. The primers used for multiplex PCR serotyping are listed in Table S1.

cpsB gene-based sequetyping

The *S. pneumoniae* capsule-specific gene *cpsB* was amplified using PCR. The primer pairs used were as follows: 5′-GCA ATG CCA GAC AGT AAC CTC TAT-3′ and 5′-CCT GCC TGC AAG TCT TGA TT-3'. The PCR product was sequenced and aligned with the GenBank database (http://blast.ncbi.nlm.nih.gov). If the obtained sequences matched with any one or more sequences in GenBank at $\geq$98%, the *cpsB* genotype and relevant serotype of the isolate were determined by referring to the *cpsB* genotype-serotype database established by Jin et al.[16] Otherwise, amplification, sequencing, and BLASTn were performed again to eliminate detection errors. If the matching consistency of the repeated test results was still less than 98%, the sequence was considered a likely new *cpsB* gene sequence type, which needs to be further confirmed by PCR or the Quellung reaction.

Statistical analysis

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated using MedCalc (version 19.6.4, MedCalc Software Ltd, Belgium) software.

Results

Clinical data

A total of 300 non-duplicate invasive *S. pneumoniae* isolates were collected from 27 teaching hospitals in 20 cities and 13 provinces in China, during 2010—2015. Of the 300 patients, the majority (197/300; 65.7%) were males. The median age of the patients was 49 years (range, 1—91 years). Children less than 18 years of age dominated, accounting for 56.7% (170/300) of the total population, among which children aged 5 years or less were the most represented, accounting for 47% (141/300) of the isolates (Table 1). Adults aged 18—65 years accounted for 30.7% (92/300), while those aged 65 years and above accounted for 12.7% (38/300) of the isolates. The most common specimen type was blood, accounting for 72.7% (218/300), followed by cerebrospinal fluid (19.0%, 57/300) and pleural effusion (n = 17, 5.7%). Other specimen types, such as ascetic fluid (n = 4), joint tissue

Table 1 Characteristics of 300 patients with *Streptococcus pneumoniae* isolation by patient sex, age, and specimen types.

Characteristics	No. of isolates	Percentage (%) of isolates
Patient sex		
Male	197	65.7
Female	103	34.2
Patient age, years		
<5	141	47.0
5–18	29	9.7
18–65	92	30.7
>65	38	12.7
Specimen type		
Blood	218	72.7
Cerebrospinal fluid	57	19.0
Pleural effusion	17	5.7
Others[a]	8	2.7

[a] Includes ascetic fluid (n = 4), joint tissue (n = 2), empyema (n = 1), and lung tissue (n = 1).

Table 2 Serotype distribution among 300 invasive *Streptococcus pneumoniae* isolates from different sources as determined by the Quellung reaction.

Serotype	No. (%) of isolates	Serotype	No. (%) of isolates	Serotype	No. (%) of isolates
23F	43 (14.3)	15B	4 (1.3)	6C	2 (0.7)
19A	41 (13.7)	23A	4 (1.3)	7C	2 (0.7)
19F	41 (13.7)	13	3 (1.0)	7F	2 (0.7)
3	31 (10.3)	29	3 (1.0)	9A	2 (0.7)
14	27 (9.0)	10A	3 (1.0)	9N	2 (0.7)
6A	12 (4.0)	15A	3 (1.0)	2	1 (0.3)
6B	11 (3.7)	24F	3 (1.0)	4	1 (0.3)
20	7 (2.3)	28F	3 (1.0)	17	1 (0.3)
34	7 (2.3)	9V	3 (1.0)	17A	1 (0.3)
15C	6 (2.0)	5	2 (0.7)	18C	1 (0.3)
1	5 (1.7)	15F	2 (0.7)	22F	1 (0.3)
8	4 (1.3)	25A	2 (0.7)	25F	1 (0.3)
11A	4 (1.3)	28A	2 (0.7)	NT	1 (0.3)
12F	4 (1.3)	33B	2 (0.7)		

(n = 2), empyema (n = 1), and lung tissue (n = 1), accounted for less than 2.0% (Table 1). All 300 *S. pneumoniae* isolates were isolated from sterile specimens and were considered invasive *S. pneumoniae* isolates.

Serotype distribution

The *S. pneumoniae* isolates were serotyped using the latex agglutination test and the Quellung reaction as the gold standard. The results were used as a reference standard for assessing the performance of multiplex PCR serotyping and *cpsB* gene sequetyping. Overall, 299 of the 300 isolates were accurately identified to serotype level by the latex agglutination test and the Quellung reaction, and only one strain did not agglutinate with any of the antisera and thus was deemed non-typeable. A total of 40 serotypes were identified, including 23F (n = 43, 14.3%), 19F (n = 41, 13.7%), 19A (n = 41, 13.7%), 3 (n = 31, 10.3%), 14 (n = 27, 9.0%), 6A (n = 12, 4.0%), 6B (n = 11, 3.7%), 20 (n = 7, 2.3%), 34 (n = 7, 2.3%) and 15C (n = 6, 2.0%), etc (Table 2).

Multiplex PCR

The 300 isolates were serotyped according to the eight-step PCR typing method recommended by the US CDC, and the results are shown in Table S2. Overall, 220 (73.3%) isolates were accurately identified to the serotype level, and 62 (20.7%) isolates could only be identified to the group/cluster level due to the limitations of the multiplex PCR primers. Incorrect serotype results were obtained in 5 (1.7%) isolates. Notably, due to the genetic similarity of the different serotype *S. pneumoniae* isolates, some isolates' serotypes were identified as just a combination of serotypes, e.g., 7C/7B/40. Specifically, one isolate of serotype 15B was misidentified as 15A/15F, one of serotype 15C as 7C/7B/40, one of serotype 17 as 34, and two of serotype 29 as 35B. Additionally, 13 isolates (4.3%) were non-typeable after all eight reactions, including 3 of serotype 28F, 1 of serotype 29, 3 of serotype 24F, 2 of serotype 28A, 2 of serotype 33B, 1 of serotype 17A, and 1 as non-typeable.

cpsB gene-based sequetyping

Among the 300 *S. pneumoniae* isolates, no *cpsB* gene was detected in four (1.3%) isolates, and the remaining 296 isolates were successfully amplified and sequenced. As shown in Table S3, based on the *cpsB* genotype-serotype database, 177 (59.0%) isolates were accurately identified to the serotype level, 91 (30.3%) isolates to the group/cluster level due to their similarity in the *cpsB* gene, and 21 (7.0%) isolates were incorrectly serotyped. There is a high degree of homology in the *cpsB* genes of different serotype isolates, in some isolates, their serotypes can only be identified as a combined form of serotypes, for example, 13-20A-20B or 6C-6D. Among the 21 misidentified serotype isolates, there was 100% agreement between the *cpsB* gene sequence of 10 of these isolates and the reference sequence in the genotype-serotype database used. The alignment of the *cpsB* gene sequence between the reference sequence in the genotype-serotype gene bank and the remaining 11 isolates ranged from 98.2 to 99.9%, with 1–13 bp base differences (Table 3).

Additionally, 11 (3.7%) isolates could not be genotyped based on the existing *cpsB* genotype-serotype database. The *cpsB* gene was not amplified in four of the isolates, including two strains of serotype 25A, one of serotype 25F, and one non-typeable (Table 4). Furthermore, in seven of the isolates, the *cpsB* gene was amplified and subsequently sequenced, but no corresponding gene sequences of any serotype matched any of the sequences in the genotype-serotype database. Three of these isolates were of serotype 11A and two each of serotypes 9V and 9A. The *cpsB* gene sequences of the three serotype 11A isolates were identical to the reference sequence MF140335 (serotype 11A according to GenBank), and the *cpsB* gene sequences of the

Table 3 Summary of serotype misidentification of *Streptococcus pneumoniae* isolates by *cpsB* sequetyping.

[a]Serotype	No. of isolates	Inconsistency			
		cpsB sequetyping results	[b]Reference sequence	[c]Similarity (%)	[d]Base difference
23F	1	23A	CR931683	100	None
14	1	13-20A-20B	JQ653094	100	None
6A	1	28F-28A	CR931693	98.9	8 bp
	1	6C-6D	JN660130	100	None
6B	3	6A	KT164779	100	None
20	4	9N	JN660134	99.9	1 bp
29	2	35B—35C	JN660114	100	None
24F	3	13-20A-20B	JQ653094	98.2	13 bp
33B	2	33D	CR931701	100	None
15F	1	23F	KT164778	98.2	13 bp
22F	1	17F—33C	JN660137	99.7	2 bp
17A	1	41F	CR931714	99.9	1 bp

[a] Serotype: as determined by the gold standard (Quellung reaction).
[b] Reference sequence: reference sequence in GenBank database.
[c] Similarity: Similarity between *cpsB* amplicon nucleotide sequence of *S. pneumoniae* isolate in this study and reference sequence in GenBank database.
[d] Base difference: Base difference between *cpsB* amplicon nucleotide sequence of *S. pneumoniae* isolate in this study and reference sequence in GenBank database.

Table 4 Summary of non-typeable *Streptococcus pneumoniae* isolates by *cpsB* sequetyping.

[a]Serotype	No. of isolates	% of isolates			
		cpsB sequetyping result	[b]Reference sequence	[c]Similarity (%)	[d]Base difference
11A	3	NT	MF140335	100	none
9V	2	NT	LR216062	100	none
9A	2	NT	LR216062	100	none
25A	2	NT	none	none	none
25F	1	NT	none	none	none
NT	1	NT	none	none	none

[a] Serotype: as determined by the gold standard (Quellung reaction).
[b] Reference sequence: reference sequence in GenBank database.
[c] Similarity: Similarity between *cpsB* amplicon nucleotide sequence of *S. pneumoniae* isolate in this study and reference sequence in GenBank database.
[d] Base difference: Base difference between *cpsB* amplicon nucleotide sequence of *S. pneumoniae* isolate in this study and reference sequence in GenBank database.

four strains of serotype 9V and 9A strains were identical to the reference sequence LR216062 (serotype 9V according to GenBank), but neither MF140335 nor LR216062 was included in the existing *cpsB* gene sequence type database (Table 4).

Comparison of multiplex PCR and *cpsB* sequetyping

Using the Quellung reaction as the gold standard, the sensitivity, specificity, PPV, NPV, and accuracy of the two methods are shown in Table 5. Overall, multiplex PCR showed 100% sensitivity and 100% specificity for the majority of the serotypes identified, except for serotypes 6A/B/C (0%) and 15A/B/C/F (0%). However, at the serogroup level, sensitivities of 100 and 93.3% were achieved for serogroups 6 and 15, respectively (Table 5).

Compared to the gold standard, the *cpsB* gene sequetyping method had 100% identification accuracy for serotypes 19A, 3, and 1. However, for serotypes 19F, 6B, 20, 34, and 15C, *cpsB* gene sequetyping could only identify these to the serogroup level. For example, due to the high similarity in the *cpsB* gene sequences of serotypes 19F and 19A, serotype 19F can only be identified as serotype 19A-19F. Interestingly, *cpsB* gene sequetyping could differentiate serotype 6A among the serogroup 6 strains, with an 83.3% (95% confidence interval [CI]: 51.6—97.9%) sensitivity and a 99.0% (95% CI: 97.0—99.8%) specificity rate, while multiplex PCR could only identify to the serogroup 6A-6B level. The discrepancies in the sensitivity and specificity rates of some serotypes occurred due to different serotype combinations (e.g., serotypes 13-20A-20B, serotypes 17A-34) sharing the same *cpsB* gene sequence in the *cpsB* genotype-serotype database. For instance, all seven strains of serotype 34 were identified as serotypes 17A-34 through *cpsB* gene sequetyping. At the single serotype level, *cpsB* gene sequetyping exhibited a 0% (95% CI: 0—41.0%) sensitivity and 100.0% (95% CI: 98.8—100.0%) specificity for serotype 34.

Table 5 Comparison of serotyping accuracy between multiplex PCR and *cpsB* gene sequetyping.

[a]Serotype	No. of isolates	% of isolates				*cpsB* sequetyping			
		Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
23F	43	100.0 (91.8–100.0)	100.0 (98.6–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	97.7 (87.7–99.9)	99.6 (97.9–99.9)	97.7 (85.6–99.7)	99.6 (97.4–99.9)
19	82	100.0 (95.6–100.0)	100.0 (98.3–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	100.0 (95.6–100.0)	100.0 (98.3–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)
19F	41	100.0 (91.4–100.0)	100.0 (98.6–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	0 (0–8.6)	100.0 (98.6–100.0)	NA	86.3 (86.3–86.3)
19A	41	100.0 (91.4–100.0)	100.0 (98.6–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	100.0 (91.4–100.0)	100.0 (98.6–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)
3	31	100.0 (88.8–100.0)	100.0 (98.6–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	100.0 (88.8–100.0)	100.0 (98.6–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)
14	27	100.0 (87.2–100.0)	100.0 (98.7–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	96.3 (81.0–99.9)	100.0 (98.7–100.0)	100.0 (100.0–100.0)	99.6 (97.6–100.0)
6	25	100.0 (86.3–100.0)	100.0 (98.7–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	96.0 (79.7–99.9)	100.0 (98.7–100.0)	100.0 (100.0–100.0)	99.6 (97.6–100.0)
6A	12	0 (0–26.5)	100.0 (98.7–100.0)	NA	96.0 (96.0–96.0)	83.3 (51.6–97.9)	99.0 (97.0–99.8)	77.0 (51.3–91.4)	99.3 (97.6–99.8)
6B	11	0 (0–28.5)	100.0 (98.7–100.0)	NA	96.3 (96.3–96.3)	0 (0–28.5)	100.0 (98.7–100.0)	NA	96.3 (96.3–96.3)
6C	2	0 (0–84.2)	100.0 (98.8–100.0)	NA	99.3 (99.3–99.3)	0 (0–84.2)	100.0 (98.8–100.0)	NA	93.3 (93.3–93.3)
15	15	93.3 (68.1–99.8)	100.0 (98.7–100.0)	100.0 (100.0–100.0)	99.7 (97.7–100.0)	93.3 (68.1–99.8)	100.0 (98.7–100.0)	100.0 (100.0–100.0)	99.7 (97.7–100.0)
15A	3	0 (0–70.8)	100.0 (98.8–100.0)	NA	99.0 (99.0–99.0)	0 (0–70.8)	100.0 (98.8–100.0)	NA	99.0 (99.0–99.0)
15B	4	0 (0–60.2)	100.0 (98.8–100.0)	NA	98.7 (96.6–99.6)	0 (0–60.2)	100.0 (98.8–100.0)	NA	98.7 (96.6–99.6)
15C	6	0 (0–45.9)	100.0 (98.8–100.0)	NA	98.0 (98.0–98.0)	0 (0–45.9)	100.0 (98.8–100.0)	NA	98.0 (98.0–98.0)
15F	2	0 (0–84.2)	100.0 (98.8–100.0)	NA	99.3 (97.6–99.9)	50.0 (1.3–98.7)	100.0 (98.8–100.0)	NA	99.7 (98.2–100.0)
20	7	100.0 (59.0–100.0)	100.0 (98.8–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	0 (0–41.0)	100.0 (98.8–100.0)	NA	97.7 (97.7–97.7)
[b]20	7					42.9 (9.9–81.6)	97.6 (95.1–99.0)	30.0 (12.2–56.9)	98.6 (97.4–99.3)
34	7	100.0 (59.0–100.0)	99.7 (98.1–100.0)	87.5 (49.7–98.0)	100.0 (100.0–100.0)	0 (0–41.0)	100.0 (98.8–100.0)	NA	97.7 (97.7–97.7)
[c]34	7					100.0 (59.0–100.0)	99.7 (98.1–100.0)	87.5 (49.7–98.0)	100.0 (100.0–100.0)
1	5	100.0 (47.8–100.0)	100.0 (98.8–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)	100.0 (47.8–100.0)	100.0 (98.8–100.0)	100.0 (100.0–100.0)	100.0 (100.0–100.0)

[a] Serotype: as determined by the gold standard (Quellung reaction).
[b] 20: Three isolates of serotype 20 were identified as serogroup 13-20A-20B by the sequetyping method.
[c] 34: Seven isolates of serotype 34 were identified as serogroup 17A-34 by the sequetyping method.
PPV, positive predictive value; NPV, negative predictive value; NA, not available.

Table 6 Comparison between sequential MULTIPLEX PCR serotyping and *cpsB* gene sequetyping.

Parameters	Multiplex PCR	*cpsB* sequetyping
Percentage of serotype level identified	73.30%	59.0%
Percentage of serogroup level identified	20.70%	30.30%
Percentage of false identification	1.70%	7.00%
Percentage of not be identified	4.30%	3.70%
Identification accuracy	96%—100%	86.33—100%
Operation	complex, 8-steps PCR	simple and convenient,1-step PCR
Reagent cost	high	low
Labor cost	high	low
Room for improvement	The primers were recombined according to the distribution of common serotypes in China.	To update and improve the *cpsB* genotype-serotype database

However, when the combination of serotypes 17A-34 was considered, *cpsB* gene sequetyping exhibited 100.0% (95% CI: 59.0—100.0%) sensitivity and a 99.7% (95% CI: 98.1—100.0%) specificity.

Discussion

In this study, a total of 300 invasive *S. pneumoniae* isolates collected from 27 teaching hospitals in 20 cities and 13 provinces in China (2010—2015) were studied. To our knowledge, this is the largest multicenter study of invasive *S. pneumoniae* isolates in China, with the largest number of participating hospitals. Based on the Quellung reaction results, 299 isolates were accurately identified to the serotype level, with a total of 40 different serotypes detected. One strain was a non-typeable strain, and the five most common serotypes were 23F (43, 14.3%), 19A (41, 13.7%), 19F (41, 13.7%), 3 (31, 10.3%), and 14 (27, 9.0%). Our serotype distribution data are similar to those of several other studies from China.[19,20] In contrast to our study, an 11-year study from India showed that the predominant serotypes were 1, 3, 5, 19F, 8, 14, 23F, 4, 19A, and 6B.[21] Notably, only one non-typeable strain was observed in our study. However, in Taiwan, there is an increasing trend of respiratory and ophthalmological mucosal infections caused by non-typeable strains in newborns and young infants.[22] This implies that we need to continuously monitor the changes in non-typeable strains in the future.

The performance of multiplex PCR and *cpsB* gene sequetyping in the identification of serotypes was evaluated using the Quellung reaction as the gold standard for serotype testing. The multiplex PCR method correctly identified 73.3 and 20.7% of the isolates to the serotype and serogroup levels, respectively. The overall serogroup identification accuracy was 94%, which is lower than that in previous studies (99%).[23] Five isolates were incorrectly serotyped, while 13 isolates could not be typed even after sequential eight-step PCR reactions. The multiplex PCR assay is reliable and expeditious, with 75% of our isolates assigned to a serotype after only three PCRs.[6] The main disadvantage of PCR-based serotyping as per the findings of this study is the inability to discriminate closely related serotypes, such as 6A-6B, 15A-15B—15C—15F, and 11A-11D. Moreover, some primer pairs exhibited cross-reactivity with other serogroups such as 7C—7B-40, 38-25F-25A, and 12F-12A-44—46. There were five instances where discrepancies between the sequential multiplex PCR assay and the Quellung reaction results were observed. In previous studies,[23] technical errors were the most common cause of discordant results. However, in our case, we carefully double-checked all discordant results to rule out any technical errors. We speculate that the discrepancies may be due to genetic variation in the strains and cross-reactivity between the primers. The eight-step sequential PCR reaction proposed by the CDC is primarily based on predominant *S. pneumoniae* serotypes in the U.S. However, this has changed over time and differs by country and even region.[24] The primers used in the US and other countries are not the best choices for use in Asian countries.[14] Therefore, in the future, there must be flexibility in the selection and combination of primer sets, based on the characteristics of the main serotype distribution in each country.

Based on the *cpsB* genotype-serotype database, *cpsB* gene sequencing accurately identified 177 (59.0%) *S. pneumoniae* isolates to the serotype level and 91(30.3%) isolates to the serogroup level, with the overall serogroup identification accuracy was 89.3% for all isolates. Twenty-one isolates (7%) were incorrectly serotyped and 11 (3.7%) isolates could not be typed using the existing *cpsB* sequence type-serotype database, of which 4 isolates were negative for *cpsB* gene amplification. In general, *cpsB* gene sequetyping is useful for identifying most serogroups. However, the disadvantage of the sequetyping method is that certain related (antigenic cross-reaction) and unrelated (no antigenic cross-reaction) serotypes share the same *cpsB* gene sequence.[25] In the present study, 41 serotype 19F isolates were only identified to the 19F-19A serogroup level. A high degree of similarity has been reported between serotypes 19F and 19A in the capsular polysaccharide synthesis locus. Primarily, the difference in the *wzy* (*cpsI*) gene contributes to the two different serotypes.[26] Therefore, the high degree of genetic homology of the *cpsB* gene may hamper the

distinction of serotypes 19F and 19A. However, given the high prevalence of serotype 19F in China, it is reasonable to use other methods such as multiplex PCR to differentiate between serotypes 19A and 19F. Serogroup 6 includes serotypes 6A/B/C/D/E, which are difficult to identify accurately by most methods.[27] In our study, there were 25 isolates of serogroup 6, including 12 isolates of serotype 6A, 11 of serotype 6B, and two of serotype 6C. The multiplex PCR method can only identify the isolates to the 6A-6B serogroup level, while the sequetyping method can identify most of the 6A isolates with an 83.3% (95% CI: 51.6—97.9%) sensitivity and a 99.0% (95% CI: 97.0—99.8%) specificity rate. Specific PCRs are needed to accurately identify serotypes 6A/B/C/D/E when the typing result is unclear.[16] Notably, some misidentified and non-typeable strains that have a high degree of similarity with sequences in NCBI did not match with sequences in the *cpsB* genotype database. This discrepancy highlights the fact that the *cpsB* gene sequetyping database needs to be constantly updated. Otherwise, there is a risk of incomplete data and incorrect assignment of serotype designations.[28] Our findings suggest that the serotype-specific sequence database still needs to be continually improved and updated, which will ultimately improve the accuracy of the *cpsB* gene sequetyping method. As per previous findings,[29,30] the *cpsB* gene could not be amplified in the four isolates in the present study. We speculate that this may be due to nucleotide variations or the absence of amplicons in the *cpsB* gene of these strains.

Overall, in terms of identification accuracy, multiplex PCR typing is slightly more accurate than *cpsB* gene sequetyping, but the whole procedure is cumbersome, and rare serotypes may not be identified even after eight consecutive steps of PCRs. However, *cpsB* gene sequetyping is simple to perform, and serotype prediction can be performed by one-step PCR amplification and sequence comparison, along with low reagent and labor costs (Table 6). The average cost of testing per strain for the multiplex PCR method is approximately US$6, compared to US$2.50 for the *cpsB* gene sequetyping method.[16,31] However, we can combine and take full advantage of both methods, that is, using the *cpsB* gene sequetyping method for initial

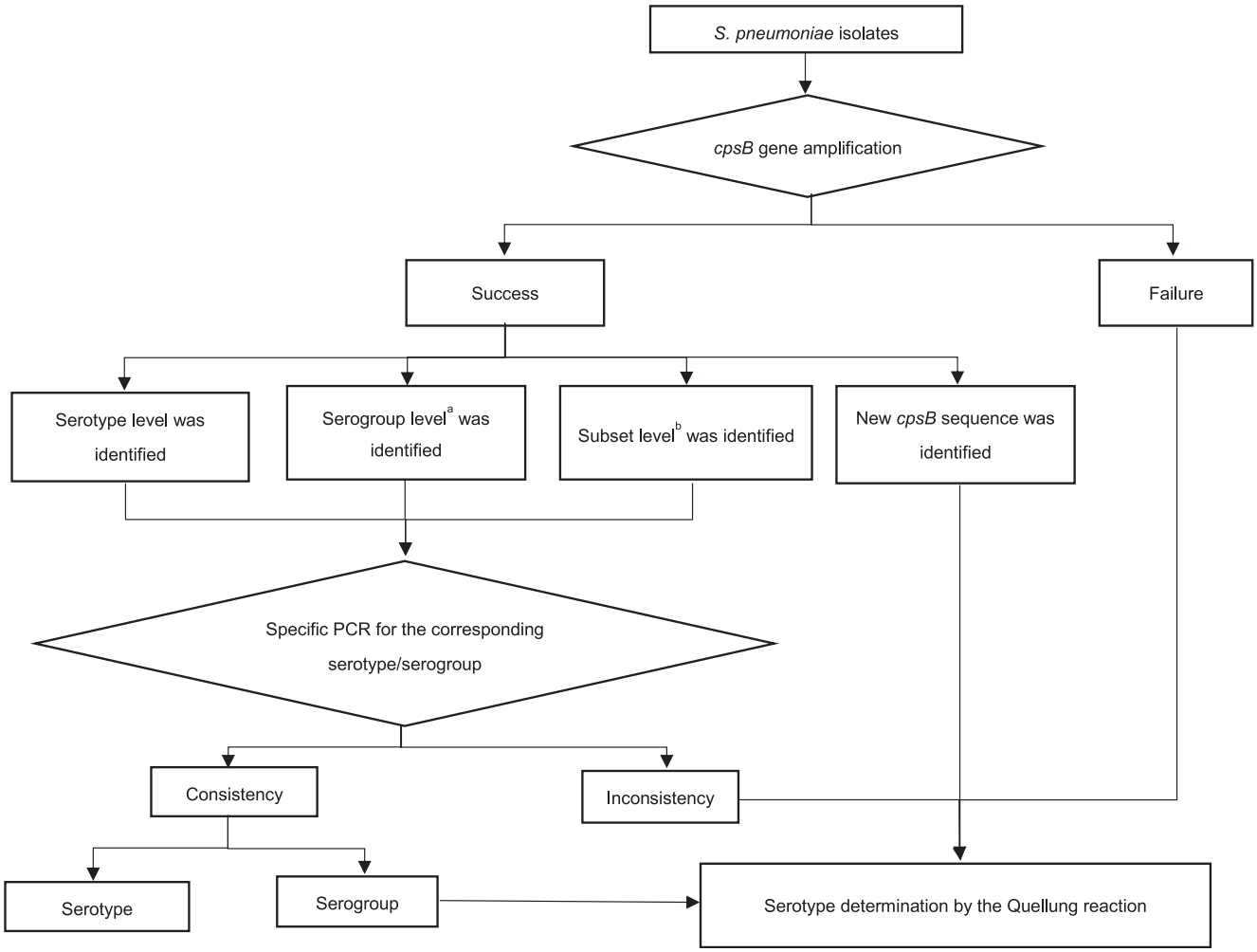

Figure 1. Algorithm for serotype identification of *Streptococcus pneumoniae* isolates. [a] Serotypes from the same serogroup sharing the same *cpsB* gene sequence in the *cpsB* genotype-serotype database. [b] Serotypes from different serogroups sharing the same *cpsB* gene sequence in the *cpsB* genotype-serotype database.

identification and then selecting the reaction of the corresponding serotype in multiplex PCR for rechecking. Although DNA-based PCR methods are fast and accurate, the Quellung reaction remains the most reliable method for serotype identification. All conflicting results between the multiplex PCR method and *cpsB* gene sequetyping, or cases where the serotype cannot be identified, should be confirmed using the Quellung reaction. Whole-genome sequencing (WGS) can be used for some non-typeable or unusual strains (Fig. 1).

Our study has several limitations. First, some discrepancies were observed between multiplex PCR/*cpsB* gene sequetyping and the Quellung reaction results. This inconsistency needs to be resolved using more precise methods, such as WGS or next-generation sequencing. Second, *S. pneumoniae* serotypes can be identified by several methods, but we have only evaluated two of them, multiplex PCR and *cpsB* sequetyping. Finally, in this study, the accuracy of *cpsB* gene sequetyping for identification to specific serotypes in our study still needs to be improved. We hope that in the future, more researchers can work together to update and improve the *cps*B genotype-serotype database so as to improve its identification accuracy.

In conclusion, our study showed that the *cpsB* gene sequetyping method is more suitable for routine primary *S. pneumoniae* serotyping screening. The *cpsB* gene sequetyping combined with multiplex PCR and the Quellung reaction can greatly improve the accuracy and efficiency of serotyping, in addition to reducing the associated costs.

Ethics statement

This study was conducted in accordance with the recommendations of the Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College Hospital Ethics Committee, and written informed consent was obtained from all subjects. The protocol was approved by the Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College Hospital Ethics Committee.

Availability of data and materials

All datasets generated for this study are included in the article/supplementary material.

Author contributions

MLZ, ZYL, PRH, and YCX conceived and designed the study. MLZ, ZRW, YW, MX, and JW performed the experiments. MLZ, ZRW, YBL, and TK performed data analysis and wrote the manuscript. All authors have read and approved the final version of the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2017YFC1601502), Special Foundation for National Science and Technology Basic Research Program of China (2019FY101200), Beijing Key Clinical Specialty for Laboratory Medicine Excellent Project (No. ZK201000) and the Graduate Innovation Fund of Peking Union Medical College (Grant No. 2017-1002-1-21).

Declaration of competing interest

The authors declare that they have no competing interests.

References

1. Lynch JP, Zhanel GG. Streptococcus pneumoniae: epidemiology, risk factors, and strategies for prevention. *Semin Respir Crit Care Med* 2009;**30**:189—209. https://doi.org/10.1055/s-0029-1202938.
2. Henriques-Normark B, Tuomanen EI. The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harb Perspect Med* 2013;**3**. https://doi.org/10.1101/cshperspect.a010215.
3. Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, et al. Global burden of childhood pneumonia and diarrhoea. *Lancet* 2013;**381**:1405—16. https://doi.org/10.1016/s0140-6736(13)60222-6.
4. Lai CC, Wang CY, Hsueh PR. Co-infections among patients with COVID-19: the need for combination therapy with non-anti-SARS-CoV-2 agents? *J Microbiol Immunol Infect* 2020;**53**:505—12. https://doi.org/10.1016/j.jmii.2020.05.013.
5. Zhou M, Yang Q, Kudinha T, Zhang L, Xiao M, Kong F, et al. Using matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) complemented with selected 16S rRNA and gyrB genes sequencing to practically identify clinical important viridans group streptococci (VGS). *Front Microbiol* 2016;**7**:1328. https://doi.org/10.3389/fmicb.2016.01328.
6. Jourdain S, Drèze PA, Vandeven J, Verhaegen J, Van Melderen L, Smeesters PR. Sequential multiplex PCR assay for determining capsular serotypes of colonizing S. pneumoniae. *BMC Infect Dis* 2011;**11**:100. https://doi.org/10.1186/1471-2334-11-100.
7. 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. https://doi.org/10.1371/journal.pgen.0020031.
8. Weinberger DM, Trzciński K, Lu YJ, Bogaert D, Brandes A, Galagan J, et al. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog* 2009;**5**:e1000476. https://doi.org/10.1371/journal.ppat.1000476.
9. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, et al. Pneumococcal capsules and their types: past, present, and future. *Clin Microbiol Rev* 2015;**28**:871—99. https://doi.org/10.1128/cmr.00024-15.
10. Huang YC, Lin CF, Ting PJ, Tang TH, Huang FL, Chao HJ, et al. Respiratory pathogens - some altered antibiotic susceptibility after implementation of pneumococcus vaccine and antibiotic control strategies. *J Microbiol Immunol Infect* 2020;**53**:682—9. https://doi.org/10.1016/j.jmii.2019.08.015.
11. 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. https://doi.org/10.1128/jcm.44.1.124-131.
12. Coskun-Ari FF, Guldemir D, Durmaz R. One-step multiplex PCR assay for detecting Streptococcus pneumoniae serogroups/types covered by 13-valent pneumococcal conjugate vaccine (PCV13).

PLoS One 2012;**7**:e50406. https://doi.org/10.1371/journal.pone.0050406.
13. Veeraraghavan B, Jayaraman R, John J, Varghese R, Neeravi A, Verghese VP, et al. Customized sequential multiplex PCR for accurate and early determination of invasive pneumococcal serotypes found in India. *J Microbiol Methods* 2016;**130**: 133—5. https://doi.org/10.1016/j.mimet.2016.09.007.
14. Park D, Kim SH, Bae IK, Kim NY, Kook JK, Park YH, et al. Evaluation of modified sequential multiplex PCR for Streptococcus pneumoniae serotyping. *Jpn J Infect Dis* 2019;**72**: 224—7. https://doi.org/10.7883/yoken.
15. Leung MH, Bryson K, Freystatter K, Pichon B, Edwards G, Charalambous BM, et al. Sequetyping: serotyping Streptococcus pneumoniae by a single PCR sequencing strategy. *J Clin Microbiol* 2012;**50**:2419—27. https://doi.org/10.1128/jcm.06384-11.
16. Jin P, Wu L, Oftadeh S, Kudinha T, Kong F, Zeng Q. Using a practical molecular capsular serotype prediction strategy to investigate Streptococcus pneumoniae serotype distribution and antimicrobial resistance in Chinese local hospitalized children. *BMC Pediatr* 2016;**16**:53. https://doi.org/10.1186/s12887-016-0589-7.
17. Lalitha MK, Thomas K, Kumar RS, Steinhoff MC. Serotyping of Streptococcus pneumoniae by coagglutination with 12 pooled antisera. *J Clin Microbiol* 1999;**37**:263—5. https://doi.org/10.1128/jcm.37.1.263-265.
18. da Gloria Carvalho M, Pimenta FC, Jackson D, Roundtree A, Ahmad Y, Millar EV, et al. Revisiting pneumococcal carriage by use of broth enrichment and PCR techniques for enhanced detection of carriage and serotypes. *J Clin Microbiol* 2010;**48**: 1611—8. https://doi.org/10.1128/jcm.02243-09.
19. Zhao C, Zhang F, Chu Y, Liu Y, Cao B, Chen M, et al. Phenotypic and genotypic characteristic of invasive pneumococcal isolates from both children and adult patients from a multicenter surveillance in China 2005-2011. *PLoS One* 2013;**8**:e82361. https://doi.org/10.1371/journal.pone.
20. Shi W, Li J, Dong F, Qian S, Liu G, Xu B, et al. Serotype distribution, antibiotic resistance pattern, and multilocus sequence types of invasive Streptococcus pneumoniae isolates in two tertiary pediatric hospitals in Beijing prior to PCV13 availability. *Expert Rev Vaccines* 2019;**18**:89—94. https://doi.org/10.1080/14760584.
21. Jayaraman R, Varghese R, Kumar JL, Neeravi A, Shanmugasundaram D, Ralph R, et al. Invasive pneumococcal disease in Indian adults: 11 years' experience. *J Microbiol Immunol Infect* 2019;**52**:736—42. https://doi.org/10.1016/j.jmii.2018.03.004.
22. Chen HH, Hsu MH, Wu TL, Li HC, Chen CL, Janapatla RP, et al. Non-typeable Streptococcus pneumoniae infection in a medical center in Taiwan after wide use of pneumococcal conjugate vaccine. *J Microbiol Immunol Infect* 2020;**53**:94—8. https://doi.org/10.1016/j.jmii.
23. Richter SS, Heilmann KP, Dohrn CL, Riahi F, Diekema DJ, Doern GV. Evaluation of pneumococcal serotyping by multiplex PCR and quellung reactions. *J Clin Microbiol* 2013;**51**:4193—5. https://doi.org/10.1128/jcm.01876-13.
24. Jauneikaite E, Jefferies JM, Hibberd ML, Clarke SC. Prevalence of Streptococcus pneumoniae serotypes causing invasive and non-invasive disease in South East Asia: a review. *Vaccine* 2012; **30**:3503—14. https://doi.org/10.1016/j.vaccine.
25. Coffey TJ, Daniels M, Enright MC, Spratt BG. Serotype 14 variants of the Spanish penicillin-resistant serotype 9V clone of Streptococcus pneumoniae arose by large recombinational replacements of the cpsA-pbp1a region. *Microbiology (Read)* 1999;**145**(Pt 8):2023—31. https://doi.org/10.1099/13500872-145-8-2023.
26. Morona JK, Morona R, Paton JC. Comparative genetics of capsular polysaccharide biosynthesis in Streptococcus pneumoniae types belonging to serogroup 19. *J Bacteriol* 1999;**181**: 5355—64. https://doi.org/10.1128/jb.181.17.
27. Baek JY, Ko KS, Kim SH, Kang CI, Chung DR, Peck KR, et al. Comparison of methods for detection of Streptococcus pneumoniae serotype 6C. *Diagn Microbiol Infect Dis* 2011;**69**: 457—9. https://doi.org/10.1016/j.diagmicrobio.2010.10.014.
28. Gonzales-Siles L, Salvà-Serra F, Degerman A, Nordén R, Lindh M, Skovbjerg S, et al. Identification and capsular serotype sequetyping of Streptococcus pneumoniae strains. *J Med Microbiol* 2019;**68**:1173—88. https://doi.org/10.1099/jmm.0.001022.
29. Dube FS, van Mens SP, Robberts L, Wolter N, Nicol P, Mafofo J, et al. Comparison of a real-time multiplex PCR and sequetyping assay for pneumococcal serotyping. *PLoS One* 2015;**10**: e0137349. https://doi.org/10.1371/journal.pone.0137349.
30. Mauffrey F, Fournier É, Demczuk W, Martin I, Mulvey M, Martineau C, et al. Comparison of sequential multiplex PCR, sequetyping and whole genome sequencing for serotyping of Streptococcus pneumoniae. *PLoS One* 2017;**12**:e0189163. https://doi.org/10.1371/journal.pone.0189163.
31. Siira L, Kaijalainen T, Lambertsen L, Nahm MH, Toropainen M, Virolainen A. From Quellung to multiplex PCR, and back when needed, in pneumococcal serotyping. *J Clin Microbiol* 2012;**50**: 2727—31. https://doi.org/10.1128/jcm.00689-12.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmii.2021.11.004.