International Journal of Infectious Diseases 86 (2019) 57–64

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Genetic diversity and antigenicity analysis of *Streptococcus pneumoniae* pneumolysin isolated from children with pneumococcal infection

Chenxi Han[a], Minmin Zhang[b,*]

[a] Queen Mary School, Nanchang University, Nanchang, Jiangxi 330019, China
[b] Clinical Microbiology Laboratory, Shanghai Center for Clinical Laboratory, Shanghai, 200126, China

ARTICLE INFO

Article history:
Received 19 April 2019
Received in revised form 13 June 2019
Accepted 21 June 2019
Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:
Streptococcus pneumoniae
Pneumolysin
Genetic diversity
Antigenicity
Pneumococcal vaccine

ABSTRACT

Background: Pneumolysin (Ply), as a major virulence factor of *Streptococcus pneumoniae*, has attracted increased attention for its potential value in the development of next-generation protein-based pneumococcal vaccines. This study aimed to analyze the genetic and antigenic diversity that can influence the immunogenicity of vaccines.
Methods: A total of 96 pneumococcal isolate samples were obtained from children of 1–35 months old with invasive pneumococcal diseases in Shanghai Children's Medical Center (Shanghai, China). After DNA amplification by PCR and Sanger sequencing, Ply DNA sequences were analyzed by bioinformatics tools, including ClustalX, BioEdit and MEGA7.
Results: Two alleles, allele 1 and 2, and 10 subtypes, of which were 6 novel subtypes, were identified. Nucleotide and amino acid sequence identity among these pneumococcal isolates were >99%. Subtypes with the same amino acid sequence were more closely evolutionarily related in the phylogenetic tree. Only minor differences in the B-cell epitopes were identified in the antigenicity plots of alleles 1 and 2. The most common serotype was serotype 19A.
Conclusions: The sequence diversity of Ply is limited although some allelic variations are detected. Different alleles exhibit similar antigenic patterns. Development of Ply-based vaccines may be a promising method to combat pneumococcal infection in the future.

Introduction

Streptococcus pneumoniae (*S. pneumoniae* or pneumococcus) is a globally widespread, Gram-positive, α-hemolytic bacterium (Haanperä et al., 2007). The encapsulated *S. pneumoniae* strains, which are more virulent, have been categorized into 97 different serotypes based on their capsular polysaccharide specificity (Anderson and Feldman, 2017; Belanger et al., 2004). The single circular genomic DNA structure of *S. pneumoniae* has a length of ~2 million base pairs and carries ~2,043 coding and 73 noncoding genes (Hoskins et al., 2001). *S. pneumoniae* is an opportunistic pathogen in humans, as it is considered part of the normal flora in the upper respiratory tract; thus, it is characterized by asymptomatic colonization (Briles et al., 2000). The cell wall, capsule, polysaccharides and pneumococcal proteins, including the pneumococcal surface protein A, pneumococcal surface antigen A, pneumolysin (Ply), autolysin (LytA), and hyaluronate lyase, are important determinants of the full pneumococcal virulence of *S. pneumoniae* (AlonsoDeVelasco et al., 1995; Jedrzejas, 2001).

Ply, an intracellular protein released during pneumococcal autolysis mainly caused by LytA, is a major virulence factor produced by all 97 known serotypes of *S. pneumoniae* (Anderson and Feldman, 2017). LytA is responsible for cell wall disruption and the release of intracellular products, such as Ply (del Mar García-Suárez et al., 2006). The Ply gene is linear and contains ~1,400 base pairs (Yun et al., 2015). Structurally, Ply protein consists of 471 amino acids with four domains (Andrew et al., 1997; Lawrence et al., 2015). Domains 1 and 3 have a role in conformational changes during Ply oligomerization. Domain 2 connects domains 1 and 4, and domain 4 is required for host cell membrane recognition and binding (Lawrence et al., 2015). Ply promotes the colonization of the upper respiratory tract in early pneumococcal infection, the migration of pneumococci into other tissues and their subsequent

* Corresponding author at: Clinical Microbiology Laboratory, Shanghai Center for Clinical Laboratory, 528 Hongshan Road, Pudong District, Shanghai 200126, China.
E-mail addresses: 664936084@qq.com (C. Han), mmzh118@163.com (M. Zhang)

https://doi.org/10.1016/j.ijid.2019.06.025

invasion, and host-to-host transmission (Zafar et al., 2017). Furthermore, Ply can exert a range of lytic and sub-lytic toxic effects on host tissues, depending on its concentration at the infectious sites. A high Ply concentration targets cholesterol-containing membranes, whereby the Trp-rich loop of domain 4 binds to cholesterol to form a Ply-cholesterol complex, leading to the membrane insertion of domain 3 before oligomerization (Kelly and Jedrzejas, 2000). Oligomerization and pore formation are subsequently induced, and Ply monomers move to the host cell membrane, with conformational changes in all four domains. Pores are formed by 30–50 Ply monomers in the lipid bilayer. These transmembrane pores, with a diameter of 35–40 nm, are sufficiently large to be seen by electron microscopy and are observed as arcs, rings and double arcs (Andrew et al., 1997; Jedrzejas, 2001; Mitchell and Mitchell, 2010; Rubins and Janoff, 1998). The cells subsequently undergo lysis; low concentrations of Ply induce sub-lytic effects. Moreover, Ply can enhance the penetration of pneumococci through the blood-brain barrier, causing pneumococcal meningitis (Kelly and Jedrzejas, 2000; Rubins and Janoff, 1998).

Although the incidence of pneumococcal infection has been dramatically reduced by availability of antibiotics and pneumococcal capsular vaccines, *S. pneumoniae* still creates a heavy clinical burden globally, and causes ~14.5 million infections worldwide (Blumental et al., 2015). It is critical to develop a more potent protein-based vaccine. Studies have shown that the investigational vaccines containing Ply toxoid and pneumococcal histidine triad protein D had comparable reactogenicity, compared with some existing vaccines, such as 13-valent pneumococcal conjugate vaccines (PCV13) (Odutola et al., 2017, 2019; Prymula et al., 2017).

In the past, Ply was regarded as an evolutionarily stable protein. However, over the past several years, accumulating evidence has revealed that Ply variants exist, with 20 allele types that cause amino acid changes reported by 2015 (Jefferies et al., 2010; Yun et al., 2015). It has been revealed that variations in Ply accumulate over time, influencing its immunogenicity and biological activities. For instance, non-hemolytic Ply was found to be produced in certain serotype 1 colonies (Jefferies et al., 2007). Variations in Ply may also result in the production of more allele types (Jefferies et al., 2010; Yun et al., 2015). Therefore, the diversity of the Ply amino acid sequence is worthy of further analysis.

However, thus far, few studies have been conducted to assess the sequence variations of the Ply genes in strains isolated from children with pneumococcal disease in China. Thus, in this project, pneumococcal isolates from children with pneumococcal infection were obtained from the Department of Clinical Laboratory of Shanghai Children's Medical Center (Shanghai, China) for analysis of the genetic diversity and antigenicity of Ply. The study aimed to identify the predominant allele types and determine whether novel allele types were present in these clinical isolates. The findings may contribute to the development of Ply-based vaccines against *S. pneumoniae*.

Materials and methods

Pneumococcal isolates

Samples of blood or cerebrospinal fluid were collected from children ages 1–35 months with pneumococcal infection. Among the total 96 children, there were 66 (68.75%) patients who were infants (<1 year old) and 30 (31.25%) who were toddlers (1–3 years old). Samples were used in a series of pneumococci identification tests at the Department of Clinical Laboratory (Shanghai Children's Medical Center). A total of 96 pneumococcal isolates were confirmed by pneumococcal growth in aerobic blood culture bottles (Becton, Dickinson and Company), α-hemolytic colony growth on blood agar plates (Oxoid Limited; Thermo Fisher Scientific, Inc.), positive results in optochin susceptibility tests and Gram-positive staining of encapsulated diplococcus. A confirmed colony of each pneumococcal isolate was collected and suspended in 250 μl double-distilled water. All bacterial suspensions were stored at −20 °C for extraction of genomic DNA.

Genomic DNA extraction

Pneumococcal genomic DNA was extracted by the heat-inactivation method. Tubes containing 250 μl bacterial suspensions were boiled in a water bath at 100 °C for 15 min to lyse the pneumococcal cells. After cooling at room temperature for 5 min, the tubes were centrifuged at 4,024.8 × *g* for 12 min. Subsequently, 150 μl supernatant fluid containing pneumococcal genomic DNA of each pneumococcal sample was collected and stored at −20 °C prior to PCR amplification.

PCR and sequence analysis

The complete sequences of Ply and LytA genes were amplified using forward and reverse primers designed based on the reference sequences from the GenBank database (ncbi.nlm.nih.gov/; Ply-accession no. M17717.1; LytA-accession no. AM113494.1; Table 1). The reagents and conditions required for PCR are listed in Tables S1 and S2. The PCR products were separated by electrophoresis on a 1% agarose gel at 150 mA/120 V for 30 min. GoldView dye (2 μl; Solarbio Science & Technology Co., Ltd.) was added to the gel for visualization of the bands by UV illumination. Ply PCR products (20 μl) were sent to GenScript for Sanger sequencing.

Allele type determination and identity analysis

The Ply sequence (GenBank accession no. CP000410.2) from the D39 strain, the first sequenced Ply gene, was designated as allele 1 and served as a reference sequence. All Ply DNA sequences from the pneumococcal isolates in this study were aligned with the reference sequence using ClustalX (European Bioinformatics Institute). The aligned DNA sequences were translated into amino acid sequences. The different allele types were determined based on the variations of the amino acid sequences. Subtypes were defined as variations in nucleotide sequences without changes in amino acid sequences. According to the alignment results, allele types and allelic subtypes were determined and numbered by comparison with previously reported alleles (Jefferies et al., 2010; Yun et al., 2015). A sequence identity matrix was constructed by pairwise comparison analysis using BioEdit (Borland Software Corporation) to analyze the identities among the amino acid and nucleotide sequences, revealing the degree of homogeneity of the different sequences.

Phylogenetic analysis

A phylogenetic tree was constructed using the complete DNA sequences of the Ply genes obtained from the pneumococcal

Table 1
Primers designed for PCR and sequencing of pneumolysin and autolysin genes.

Target gene	Primer name	Sequence (5′–3′)
Ply	PF1 (forward)	GCTGATTTCGCTGAACAAGTC
	PR1 (reverse)	GTCGCAAGCATTCTCCTCTC
LytA	AF1 (forward)	TTGATAAGGAGTAGAATATGG
	AR1 (reverse)	TACTGTAATCAAGCCATCTGG

Ply, pneumolysin; LytA, autolysin.

isolates to analyze evolutionary relationships using the neighbor-joining method in MEGA7 (megasoftware.net). The subtypes that belonged to an identical main branch of the tree were aligned to one clade and the bootstrap replicates were 1,000.

Physicochemical properties and structure prediction

Six online analysis tools were used to predict the structural properties of Ply. TMHMM Server v. 2.0 (cbs.dtu.dk/services/TMHMM/) was developed for the prediction of transmembrane helices in proteins. SignalP 4.1 Server (cbs.dtu.dk/services/SignalP/) was used to identify the locations of the signal peptide cleavage sites in Ply protein. The ProtParam tool (web.expasy.org/protparam/) was used to calculate the physical and chemical properties of proteins, such as the molecular weight and grand average of hydropathicity index (GRAVY). The Hphob./Kyte & Doolittle method (with hydropathicity amino acid scale) was provided by ProtScale (web.expasy.org/protscale/) to present profiles of Ply. PHD (npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html) was used as a secondary structure prediction method. Homology modeling analysis was performed to predict the tertiary structure of Ply using SWISS-MODEL (swissmodel.expasy.org).

Antigenicity analysis

Antigenicity analysis was performed based on B-cell epitopes. Surface accessibility is an important parameter to determine the presence and location of epitopes. Hence, linear B-cell epitopes were predicted using the Immune Epitope Database Analysis Resource (tools.iedb.org/main/bcell/) via the Emini surface accessibility scale (Emini et al., 1985). The epitope prediction calculations were based on the surface accessibility for each of the 20 amino acids. The accessibility parameters consisted of 20 values assigned to each of the amino-acid residues used to complete the accessibility profile. The sequences with larger scores had a higher probability of to being found on the surface of the protein.

Serotying

Serotype determination was performed with a pneumotest kit (Statens Serum Institute, Denmark) based on a quelling reaction.

Results

Ply gene amplification and DNA sequencing

The Ply genes from 88 isolates were amplified with PCR and successfully sequenced using the Sanger method. LytA genes of the same 88 isolates were also amplified successfully with PCR and gel electrophoresis. The DNA sequences of Ply genes obtained by Sanger sequencing were compared with a reference sequence (Figure S1).

Allele types and sequence diversity

A total of 10 different Ply gene DNA sequences generated in this study have been deposited in GenBank under accession nos. **MK286376**–**MK286385**. Two alleles, allele 1 (n = 24; 27.3%) and 2 (n = 64; 72.7%), were identified by comparison to the 20 known Ply alleles. A single amino acid change (D380N) was detected in allele 2 (Table 2). Allele subtypes were found to display more variations in the nucleotide sequences than in the amino acid sequence; 10 allele subtypes were identified, including 6 novel allelic subtypes (Table 3). Notably, nucleotide 195 was identified as a novel variation position, with C195T substitution in allele subtype 1.9. The novel subtype, 2.11 (n = 44; 50%), was the most common in this study (Table 3).

A heatmap revealed high identity among the alleles based on the nucleotide and amino acid sequences using pairwise comparisons (Figure 1). The amino acid sequence identities were 99.7% and 100%. Among all subtypes determined in the study, a nucleotide identity of >99% (99.3–100%) was shown.

Phylogenetic analysis

All nucleotide sequences in this study were aligned to construct a phylogenetic tree by neighbor-joining analysis using MEGA7 software (Figure 2). The splits of the branches represent the

Table 2
Allele types and the corresponding variations in the amino-acid sequences of pneumolysin.

	Amino acid position																								
Allele (n)	14	136	142	150	154	167	172	224	260	265	267	270	271	273	296	323	366	372	380	386	394	402	415	439	468
1 (24)	N	Q	N	Y	T	S	T	K	E	A	I	V	K	A	P	D	D	V	D	H	P	Q	K	V	V
2 (64)																			N						
3 (0)							I	R		S		del	del												
4 (0)																			N				ins		
5 (0)				H			I	R		S		del	del												
6 (0)						F	I	R		S		del	del												
7 (0)	D							R																	
8 (0)														D											
9 (0)											M														
10 (0)				M						S		del	del						N						
11 (0)		K																	N						
12 (0)																			N			E			
13 (0)									D										N						
14 (0)			ins																						
15 (0)									D																
16 (0)								R							S						A				I
17 (0)	D							R								N				N					
18 (0)																								I	
19 (0)																	N	I							
20 (0)								R												Y					

Wild-type D39 strain was defined as allele 1. n = 0, not detected in this study; del, single amino-acid deletion; ins, 8 amino-acid insertions in allele 4, and 291 amino-acid insertions in allele 14.

Table 3
Allele subtypes covering alleles 1 and 2, and corresponding variations in the nucleotide sequences of pneumolysin.

Nucleotide position	Allele subtype (n)																								
	1.0 (0)	1.1 (7)	1.2 (0)	1.3 (0)	1.4 (0)	1.5 (0)	1.6 (0)	1.7 (0)	1.8 (0)	1.9 (9)	1.10 (2)	1.11 (6)	2.0 (0)	2.1 (0)	2.2 (0)	2.3 (0)	2.4 (2)	2.5 (0)	2.6 (8)	2.7 (0)	2.8 (0)	2.9 (2)	2.10 (0)	2.11 (44)	2.12 (3)	2.13 (5)
195	C									T																
261	C																			T	T					
276	C						T	T	T		T	T														
351	C												T	T											T	T
363	A				G																					
459	C						A	A			A															
555	A																					G	G			
558	C						T	T	T		T	T														
588	G					A												A	A							
609	C																				T					
615	T					C	C	C	C		C	C	C	C	C	C		C	C	C				C	C	C
729	G																	A								
732	C														T											
828	A																	G								
837	G			A		A												A	A			A	A			
924	T	C																								
1038	T		C	C	C			C	C	C	C	C														
1056	G																				A					
1138	G												A	A	A	A	A	A	A	A	A	A	A	A	A	A
1212	G					A											A	A	A			A	A	A	A	A
1257	T																									
1272	A																				G					
1296	C												T												T	
1356	G																						T			
1386	T				G								G												G	
1395	A		G					G	G																	

The novel nucleotide variation at position 195 is indicated in bold font.

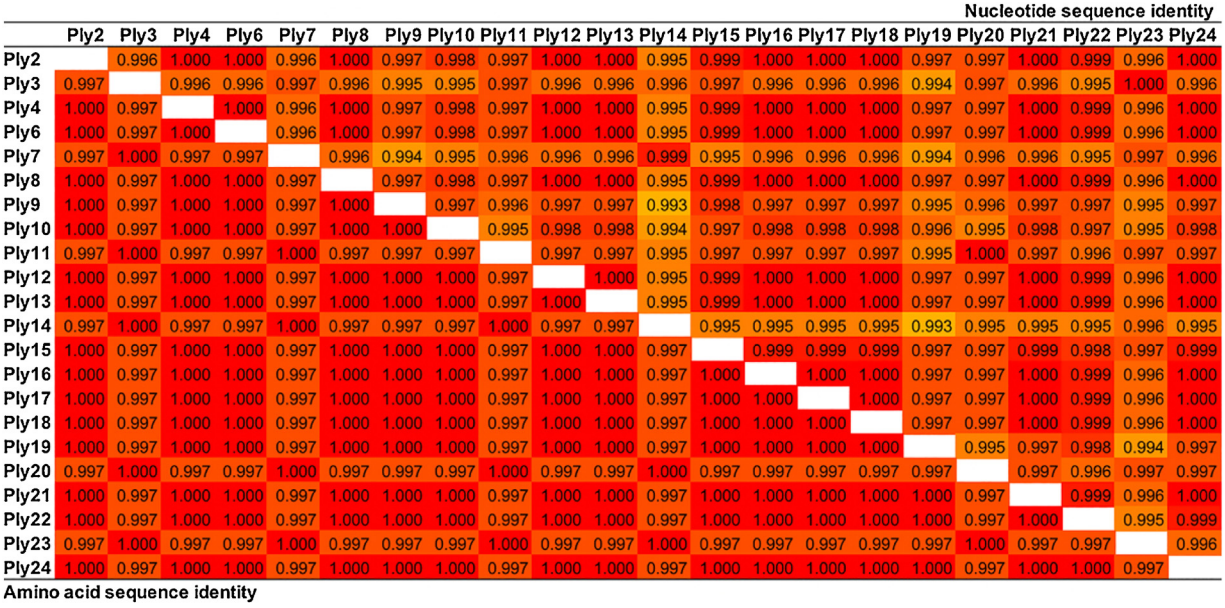


Figure 1. Amino acid sequences identity and nucleotide sequences identity in the heatmap (for 20 samples, used as a representative example). Ply X indicates the Ply sequence from sample X. Ply, pneumolysin.

evolutionary history. The subtypes that shared a main branch were assigned into a clade, indicating a closer evolutionary relationship. In this study, the phylogenetic tree representing subtypes branched into two main clades, clade A (n = 24; 27.3%) and clade B (n = 64; 72.7%). Allele 1 was placed in clade A and allele 2 clustered with clade B; however, subtype 1.5 was assigned to clade B, which was derived from allele 2.

Physicochemical properties and structure prediction

The structural analysis focused on the prediction of transmembrane helices, signal peptide cleavage sites and hydropathicity. The structural prediction of allele 1 transmembrane helices, signal peptide cleavage sites and hydropathicity is displayed in Figure S2A, B and C, respectively. The results of allele 2 were extremely similar to those of allele 1 (data not shown). The absence of a transmembrane region, an outside region and a signal peptide

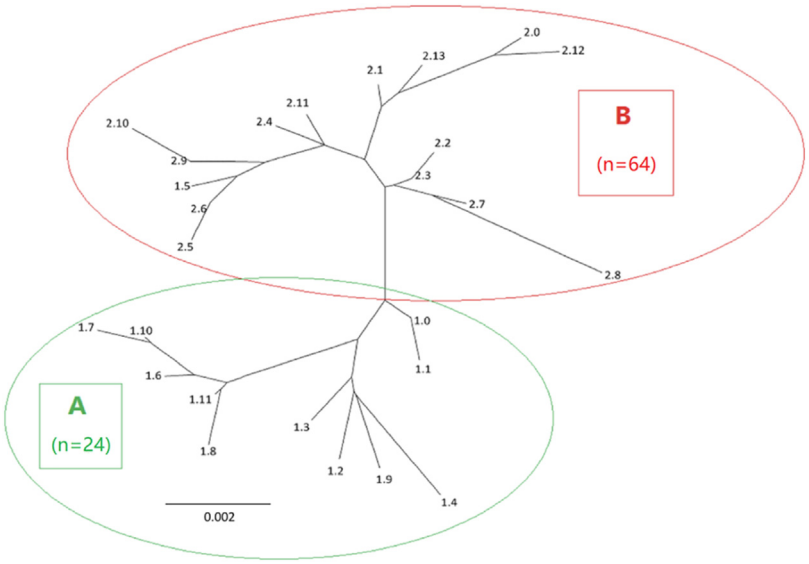


Figure 2. Phylogenetic tree of the pneumolysin alleles identified in this study. Characters A and B, and the numbers at the end of branches represent the clade types and allele subtypes, respectively. The genetic distance bar is located at the bottom, and the bootstrap replicates are 1,000.

(Figure S2A and B) indicated that the Ply is an intracellular and non-secretory protein. As observed in Figure S2C, the areas above 0 are hydrophobic, whereas areas below 0 are hydrophilic. The majority of the Ply protein was predicted to be hydrophilic. In addition, the GRAVY values of alleles 1 and 2 obtained by ProtParam were −0.42 (<0). Therefore, the ProtScale and ProtParam analysis results were consistent regarding the hydropathicity of the protein.

The secondary and tertiary structures of Ply protein were predicted by PHD (Figure S3) and SWISS-MODEL (Figure 3) using the allele 1 amino acid sequence. Ply was composed of 20.17% alpha helices, 35.03% extended strands and 44.80% random coils. As for the homology modeling, the sequence identity and similarity were high, indicating the prediction results were reliable. The properties of allele 2 were similar to those of allele 1. The relevant information and parameters detected are listed in Tables 4 and 5.

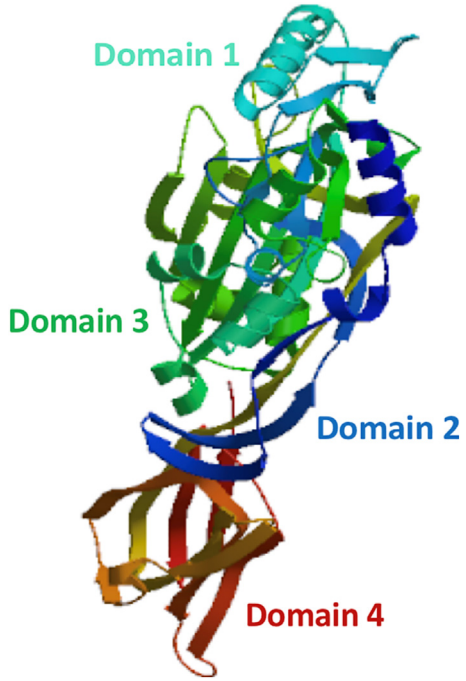


Figure 3. Prediction of the tertiary structure of allele 2. The different domains of pneumolysin are shown in four colors. The colors used for the protein structures represent the four different domains of pneumolysin. The cyan, blue, green and red areas indicate domains 1, 2, 3 and 4, respectively.

Table 4
Comparison of pneumolysin allele 1 and 2 secondary structures.

Allele	Alpha helix (Hh)	Extended strands (Ee)	Random coil (Cc)
1	20.17%	35.03%	44.80%
2	19.74%	35.46%	44.80%

Table 5
Comparison of pneumolysin allele 1 and 2 tertiary structures.

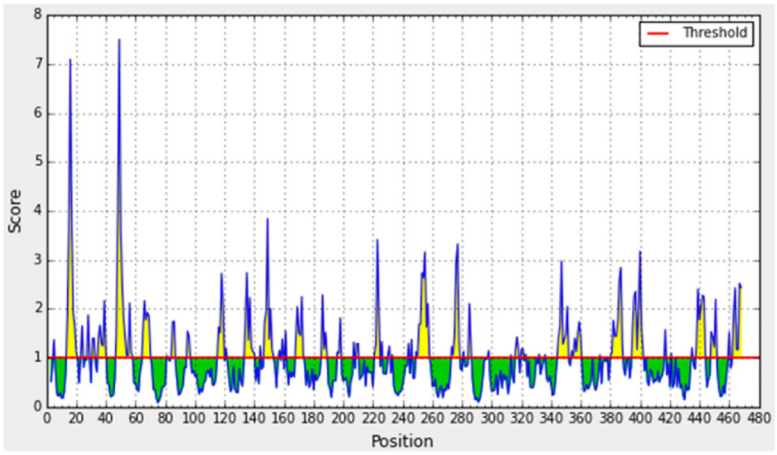
Allele	Template	Sequence identity	Sequence similarity
1	5aod. 1.A	100.00%	0.6
2	5aod. 1.A	99.79%	0.6

5aod. 1.A (SMTL ID of SWISS-MODEL): Pneumolysin template was the crystal structure of wild-type pneumolysin determined by X-ray diffraction.

Antigenicity analysis

B-cell epitopes have an important role in the antigenicity of Ply. On the basis of the surface accessibility scale, the IEDB Analysis Resource was used to produce an antigenic plot of allele 1 that predicts the presence and locations of the B-cell epitopes (Figure 4A). The residues with larger accessibility scores, indicated in yellow on the graphs, had a higher probability to form parts of epitopes. Of these, 14 predicted epitope peptides of allele 1 are shown in Figure 4B. The only difference between alleles 1 and 2 was that amino acid 380 in allele 2 was not included in the predicted peptide no. 12; with ELSYDHQG as the amino acid sequence of predicted peptide no. 12 in allele 2. Hydrophilicity is another important scale for predicting epitopes. The antigenicity graph (Figure 4) and hydropathicity results (Figure S2C) showed similar trends. Overall, alleles 1 and 2 exhibited very similar antigenic patterns.

A



B

Predicted peptides:

No.	Start	End	Peptide	Length
1	14	20	NYDKKKL	7
2	35	40	EGNQLP	6
3	47	58	ERKKRSLSTNTS	12
4	65	70	TNDSRL	6
5	134	140	WHQDYGQ	7
6	147	153	RMQYEKI	7
7	168	173	DFEKTG	6
8	249	258	KLETTSKSDE	10
9	273	278	APQTEW	6
10	345	351	QNSTDYV	7
11	354	360	KVTAYRN	7
12	380	388	DELSYDHQG	9
13	395	402	KAWDRNGQ	8
14	438	444	TVYEKTD	7

Figure 4. Antigenicity analysis of pneumolysin. (A) Antigenicity plot of allele 1. The y-axis depicts accessibility score for each residue, and the x-axis indicates the residue positions in the sequence. (B) List of predicted B-cell epitope peptides of allele 1.

Table 6
Serotype distribution of invasive *S. pneumoniae* strains.

Serotypes of *S. pneumoniae*	Number (%)
19A	31 (35.2%)
19F	24 (27.3%)
23F	8 (9.1%)
5	12 (13.6%)
14	13 (14.0%)
Total	88 (100%)

Serotyping

Among the 88 isolates, 5 serotypes, i.e., 19A, 19F, 23F, 5 and 14, were identified (Table 6). The serotype 19A (35.2%) was the most common serotype, but serotype 23F (9.1%) showed the opposite result.

Discussion

Currently, the 13-valent pneumococcal conjugate vaccine (PCV13) has replaced PCV7 and has become the most effective and widely used vaccine to prevent pneumococcal infection in children <2 years. In 2016, PCV13 was approved by the China Food and Drug Administration, but PCV-13 is not included in the National Immunisation Programme. The vaccination rate of PCV-13 is low, likely due to the high price of the vaccine (Yang et al., 2019). The production of PCV13, as well as PCV23, requires the purification of capsular polysaccharides from serotypes, which further increases the cost. Moreover, capsular polysaccharide-based vaccines are limited by serotypes, and, unfortunately, they do not cover all serotypes. Therefore, serotype-independent protein-based vaccines should be developed to enhance the protective capacity of vaccines. Several virulent pneumococcal proteins have been proposed as vaccine candidates (Kirkham et al., 2005).

Ply, an immunogenic protein with hemolytic activity, is found in almost all known clinical serotypes of *S. pneumoniae*. A predominant advantage of Ply-based pneumococcal vaccines is the serotype independence, however Ply sequence variations may influence the immunogenicity of the vaccine, which is developed using one specific allele of Ply. This project was designed with the aim to study the genetic and antigenic diversity of Ply using clinical isolates from children with *S. pneumoniae* infection in Shanghai.

To the best of our knowledge, this study is the first to focus on the sequence diversity and antigenicity of *S. pneumoniae* Ply isolated from children with invasive pneumococcal diseases in China. PCR was used to successfully amplify 88 Ply and LytA genes from a total of 96 samples. The existence of LytA indicated that Ply and other intracellular proteins were released and played a role in the pathogenesis and invasion of tissues. For the 8 isolates with negative PCR results, genetic variations may be present at the Ply and LytA loci. Both loci could be replaced by or integrated with foreign DNA via gene transfer mechanisms, such as transformation, transduction, conjugation or lysogenic conversion. Another reason might be that there were point mutations within the template sequences corresponding to the primers, decreasing the affinity between the primer sequence and template. Without LytA or Ply, the virulence of isolates would be weakened. The mechanisms by which the isolates with negative Ply and LytA PCR amplification invaded the patients to cause pneumococcal infection should be investigated further.

Two Ply alleles (alleles 1 and 2) were identified from the 88 clinical isolates in this study based on 20 previously reported allele types. The only difference in the amino acid sequence of the protein produced from the two alleles is a D380N substitution. The physicochemical properties of asparagine (N) and aspartic acid (D) are slightly different. The amino acid N in allele 2 is not charged, but amino acid D in allele 1 is charged. This difference could result in slight change in the conformation of the protein. According to the nucleotide sequences, 10 allele subtypes were identified, with 6 novel subtypes not reported in previously published literature. It is more likely that the gradual accumulation of nucleotide variations could alter the amino acid sequence and generate novel allele types. Among the 10 subtypes, subtype 2.11 (a novel subtype) was found in 44 isolates, which indicates that it may be subtype that is specific to and dominantly prevalent in East China. The nucleotide and amino acid sequence identities of alleles 1 and 2 were >99%, suggesting that Ply is a very evolutionarily stable protein in this area. The conserved characteristics of Ply were also reflected in the stable evolutionary relationship among the allelic subtypes. In the phylogenetic tree, the subtypes of same allelic type at the amino acid level shared a main branch of the tree. The phylogenetic clades were basically consistent with allele types.

The immunogenicity of vaccines is commonly determined by the conformation of Ply based on its amino acid sequence. The allele analysis could be used to reveal conformational differences that are present among variants with different amino acid sequences. The secondary and tertiary structures of allele 1 and

2 were predicted to have high similarity. The three dimensional Ply structures predicted using SWISS-MODEL showed no difference between alleles 1 and 2, and extremely similar antigenic patterns for alleles 1 and 2 were also observed in the antigenicity plots. These findings indicate that the D380N substitution does not alter the structure of Ply or its immunogenicity. It is necessary to determine whether the D380N substitution in domain 4 of Ply influences pneumococcal virulence.

Although only two alleles were identified in this study among the 20 known alleles, the project still can be considered as a representative investigation on the genetic and antigenic diversities of Ply. Allele 1 and allele 2 are the most common Ply alleles in Shanghai and similar result were reported in studies conducted in Korea and the UK (Jefferies et al., 2010; Yun et al., 2015). It is suggested that Ply alleles 1 and 2 could be chosen as candidate molecules for the development of protein-based pneumococcal vaccines. A total of 5 serotypes of S. pneumoniae were detected in alleles 1 and 2, and they coincided with the serotypes covered by the commonly used PCV. We need more samples to determine the number of serotypes that exist in the two most common alleles. More serotypes mean the Ply protein-based vaccine based on alleles 1 and 2 sequences will be more effective. Nonetheless, the results need to be confirmed in studies using larger sample sizes, covering more geographical regions.

Another limitation is that this study only analyzed the genetic diversity and antigenicity of Ply in invasive pneumococcal isolates; however, a study in the UK discovered that the Ply alleles identified from carriage strains are similar to those among invasive isolates (Jefferies et al., 2010). In addition, the pneumococcal isolates obtained from patients with non-invasive pneumococcal diseases are also worth examining. Because the type of disease, drug resistance and degree of recovery vary among patients, the specific conditions of each patient should be considered.

In conclusion, the prevalence of different Ply alleles was determined in children <5 years old with invasive pneumococcal diseases in Shanghai, identifying alleles 1 and 2 as the most common and widely distributed alleles. Ply is highly conversed, although minor variations are present, which is critical for the potential use of a Ply-based vaccine. Ply could be used alone or as a carrier protein in conjugate vaccines. Ply is highly evolutionarily stable and can be used as a valuable antigen for the production of universal pneumococcal vaccines.

Funding

This work was supported by the Shanghai Municipal Health Commission [grant numbers 20134010, 201840005]; and the Science and Technology Department of Henan Province [grant number 181200211900].

Availability of data and materials

The datasets, 10 DNA sequences, generated during the current study are available in GenBank (accession nos. **MK286376–MK286385**).

Author contribution

CH, first author, conducted the experiments, performed statistical analysis and wrote the main manuscript text. MZ, corresponding author, was responsible for the conception and design of the study. Both authors provided final approval of the manuscript.

Ethics approval and consent to participate

This study and the protocol were approved by the Institutional Review Board of Shanghai Center for Clinical Laboratory (Shanghai, China). The Ethics Committee allowed a waiver of informed consent because this study included only the information of bacteria without any identifying information of human from whom the bacteria was obtained.

Conflict of interest

None.

Acknowledgement

The authors thank Dr. Weichun Huang (Shanghai Children's Medical Center, Shanghai, China), Shanghai Municipal Health Commission [grant numbers 20134010, 201840005], and the Science and Technology Department of Henan Province [grant number 181200211900].

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2019.06.025.

References

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

Anderson R, Feldman C. Pneumolysin as a potential therapeutic target in severe pneumococcal disease. J Infect 2017;74:527–44, doi:http://dx.doi.org/10.1016/j.jinf.2017.03.005.

Andrew PW, Mitchell TJ, Morgan PJ. Relationship of structure to function in pneumolysin. Microb Drug Resist 1997;3:11–7, doi:http://dx.doi.org/10.1089/mdr.1997.3.11.

Belanger AE, Clague MJ, Glass JI, LeBlanc DJ. Pyruvate oxidase is a determinant of Avery's rough morphology. J Bacteriol 2004;186:8164–71, doi:http://dx.doi.org/10.1128/JB.186.24.8164-8171.2004.

Blumental S, Granger-Farbos A, Moïsi JC, Soullié B, Leroy P, Njanpop-Lafourcade B-M, et al. Virulence factors of *Streptococcus pneumoniae*. Comparison between African and French invasive isolates and implication for future vaccines. PLoS One 2015;10:e0133885, doi:http://dx.doi.org/10.1371/journal.pone.0133885.

Briles DE, Hollingshead SK, Nabors GS, Paton JC, Brooks-Walter A. The potential for using protein vaccines to protect against otitis media caused by *Streptococcus pneumoniae*. Vaccine 2000;19:S87–95, doi:http://dx.doi.org/10.1016/S0264-410X(00)00285-1.

del Mar García-Suárez M, Vázquez F, Méndez FJ. *Streptococcus pneumoniae* virulence factors and their clinical impact: an update. Enferm Infecc Microbiol Clin 2006;24:512–7, doi:http://dx.doi.org/10.1157/13092469.

Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 1985;55:836–9.

Haanperä M, Jalava J, Huovinen P, Meurman O, Rantakokko-Jalava K. Identification of alpha-hemolytic streptococci by pyrosequencing the 16S rRNA gene and by use of VITEK 2. J Clin Microbiol 2007;45:762–70, doi:http://dx.doi.org/10.1128/JCM.01342-06.

Hoskins J, Alborn WE, Arnold J, Blaszczak LC, Burgett S, DeHoff BS, et al. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J Bacteriol 2001;183:5709–17, doi:http://dx.doi.org/10.1128/JB.183.19.5709-5717.2001.

Jedrzejas MJ. Pneumococcal virulence factors: structure and function. Microbiol Mol Biol Rev 2001;65:187–207, doi:http://dx.doi.org/10.1128/MMBR.65.2.187-207.2001.

Jefferies JMC, Johnston CHG, Kirkham LAS, Cowan GJM, Ross KS, Smith A, et al. Presence of nonhemolytic pneumolysin in serotypes of *Streptococcus pneumoniae* associated with disease outbreaks. J Infect Dis 2007;196:936–44, doi:http://dx.doi.org/10.1086/520091.

Jefferies JMC, Tocheva AS, Rubery H, Bennett J, Garland J, Christodoulides M, et al. Identification of novel pneumolysin alleles from paediatric carriage isolates of *Streptococcus pneumoniae*. J Med Microbiol 2010;59:808–14, doi:http://dx.doi.org/10.1099/jmm.0.018663-0.

Kelly SJ, Jedrzejas MJ. Structure and molecular mechanism of a functional form of pneumolysin: a cholesterol-dependent cytolysin from *Streptococcus pneumoniae*. J Struct Biol 2000;132:72–81, doi:http://dx.doi.org/10.1006/jsbi.2000.4308.

Kirkham LAS, Kerr AR, Douce GR, Paterson GK, Dilts DA, Liu DF, et al. Construction and immunological characterization of a novel nontoxic protective pneumolysin mutant for use in future pneumococcal vaccines. Infect Immun 2005;74:586–93, doi:http://dx.doi.org/10.1128/IAI.74.1.586-593.2006.

Lawrence SL, Feil SC, Morton CJ, Farrand AJ, Mulhern TD, Gorman MA, et al. Crystal structure of *Streptococcus pneumoniae* pneumolysin provides key insights into early steps of pore formation. Sci Rep 2015;5:14352, doi:http://dx.doi.org/10.1038/srep14352.

Mitchell AM, Mitchell TJ. *Streptococcus pneumoniae*: virulence factors and variation. Clin Microbiol Infect 2010;16:411–8, doi:http://dx.doi.org/10.1111/j.1469-0691.2010.03183.x.

Odutola A, Ota MOC, Antonio M, Ogundare EO, Saidu Y, Foster-Nyarko E, et al. Efficacy of a novel, protein-based pneumococcal vaccine against nasopharyngeal carriage of *Streptococcus pneumoniae* in infants: a phase 2, randomized, controlled, observer-blind study. Vaccine 2017;35:2531–42, doi:http://dx.doi.org/10.1016/j.vaccine.2017.03.071.

Odutola A, Ota MOC, Antonio M, Ogundare EO, Saidu Y, Owiafe PK, et al. Immunogenicity of pneumococcal conjugate vaccine formulations containing pneumococcal proteins, and immunogenicity and reactogenicity of co-administered routine vaccines — a phase II, randomised, observer-blind study in Gambian infants. Vaccine 2019;37:2586–99, doi:http://dx.doi.org/10.1016/j.vaccine.2019.03.033.

Prymula R, Szenborn L, Silfverdal S-A, Wysocki J, Albrecht P, Traskine M, et al. Safety, reactogenicity and immunogenicity of two investigational pneumococcal protein-based vaccines: results from a randomized phase II study in infants. Vaccine 2017;35:4603–11, doi:http://dx.doi.org/10.1016/j.vaccine.2017.07.008.

Rubins JB, Janoff EN. Pneumolysin: a multifunctional pneumococcal virulence factor. J Lab Clin Med 1998;131:21–7, doi:http://dx.doi.org/10.1016/S0022-2143(98)90073-7.

Yang T, Deng P, Yang L, Zhou C, Fei Y. Evaluations on coverage and safety of 13-valent pneumococcal conjugate vaccine inoculation in Pudong New Area of Shanghai, 2017 - 2018. Chin J Public Health 2019;, doi:http://dx.doi.org/10.11847/zgggws1123040.

Yun KW, Lee H, Choi EH, Lee HJ. Diversity of pneumolysin and pneumococcal histidine triad protein D of *Streptococcus pneumoniae* isolated from invasive diseases in Korean children. PLoS One 2015;10:e0134055, doi:http://dx.doi.org/10.1371/journal.pone.0134055.

Zafar MA, Wang Y, Hamaguchi S, Weiser JN. Host-to-host transmission of *Streptococcus pneumoniae* is driven by its inflammatory toxin, pneumolysin. Cell Host Microbe 2017;21:73–83, doi:http://dx.doi.org/10.1016/j.chom.2016.12.005.