Journal of Microbiology, Immunology and Infection 58 (2025) 325–332

Contents lists available at ScienceDirect

Journal of Microbiology, Immunology and Infection

journal homepage: www.e-jmii.com

An insight of *Streptococcus pneumoniae* serotype 3 genomic profile in Indonesia

Ratna Fathma Sari [a,c,**], Fadilah Fadilah [b], Yustinus Maladan [c], Rosantia Sarassari [c], Miftahuddin Majid Khoeri [d], Kuntjoro Harimurti [e], Lindawati Alimsardjono [f], Dodi Safari [c,*]

[a] *Biomedical Sciences, Faculty of Medicine, University of Indonesia, Jl. Salemba Raya No. 6, Jakarta Pusat, DKI Jakarta, 10430, Indonesia*
[b] *Medical Chemistry Department, Faculty of Medicine, University of Indonesia, Jl. Salemba Raya No. 6, Jakarta Pusat, DKI Jakarta, 10430, Indonesia*
[c] *Eijkman Research Center for Molecular Biology, National Research and Innovation Agency, Jl. Raya Bogor KM 46, Cibinong, Jawa Barat, Indonesia*
[d] *Infectious Disease and Immunology Research Center, Indonesia Medical and Education Research Institute (IMERI), Faculty of Medicine, University of Indonesia, Indonesia*
[e] *Department of Internal Medicine, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia*
[f] *Department of Clinical Microbiology, Dr. Soetomo General Academic Hospital, Surabaya, East Java, Indonesia*

ARTICLE INFO

Keywords:
S. pneumoniae serotype 3
Genome analysis
Pneumococcal conjugate vaccine
Indonesia

ABSTRACT

Background: Streptococcus pneumoniae Serotype 3 (SPN3) remains a significant cause of morbidity and mortality worldwide despite of pneumococcal conjugate vaccine (PCV) implementation. We explored genomic profile of SPN3 from children and adult groups to understand population structure and evolution dynamics of SPN3 in Indonesia.

Methods: We undertook whole genome sequencing (WGS) from 19 isolates of SPN3 in Indonesia between 2017 and 2021 prior to PCV introduction. This study assessed sequence types (STs), global pneumococcal sequence cluster (GPSC), genome prediction of antimicrobial resistance (AMR) profile, pangenome analysis, phylogenetic tree, and genome comparative of capsular polysaccharide (*cps*) locus.

Results: We identified ST451-GPSC234 (n = 5) and ST180-GPSC12 (n = 4), ST458-GPSC51 (n = 2), ST3805-GPSC12 (n = 2), ST4909-GPSC363 (n = 2), ST700-GPSC10 (n = 1), ST5292-GPSC309 (n = 1), ST505-GPSC12 (n = 1), and ST4233 (n = 1). Genome prediction of AMR discover isolates were resistant to tetracycline (n = 5); co-resistant of chloramphenicol and tetracycline (n = 2); co-trimoxazole and tetracycline (n = 1). We observed SPN3 possess closed pangenome characteristic, indicates more stable genetic repertoire. We found 5 absent genes in *cps* locus including *cpsABCD* and *tnp* in ST700-GPSC10 lineage.

Conclusions: SPN3 has potential genomic profile to enhance the ability of this strain to endure selective pressure such as PCV introduction.

1. Introduction

Streptococcus pneumoniae serotype 3 (SPN3) is a major cause of invasive pneumococcal disease (IPD) and remains found in carriage despite its inclusion in the 13-valent pneumococcal conjugate vaccine (PCV13).[1] Observational studies of IPD in children and community-associated pneumonia in adults have demonstrated vary effectiveness of PCV13 against this serotype.[2,3] In Indonesia, SPN3 accounts for 4.9 % of pneumococcal carriage in children under <5 years of age and found as the highest prevalence across South East Asian countries.[4] The prevalence rate of SPN3 in Indonesian adults remains limited. However, SPN3 was identified as the most acquired serotype during Hajj in Indonesian pilgrims and predominant serotype in adult patients with community-acquired pneumonia in Jakarta, Indonesia.[5,6] SPN3 exhibits a distinctive genetic, phenotypic, and epidemiologic profile with higher prevalence in the adult population and lower colonization rate among children.[7]

SPN3 is primarily distinguished by its high virulence, high invasiveness, high mortality, and susceptible to clinical antibiotics.[8] The most commonly known characteristic of SPN3 is a thicker capsule that is less tightly attached to the cell surface, which may contribute to immune evasion among vaccinated individuals and severe clinical

* Corresponding author.
** Corresponding author. Biomedical Sciences, Faculty of Medicine, University of Indonesia, Jl. Salemba Raya No. 6, Jakarta Pusat, DKI Jakarta, 10430, Indonesia.
E-mail address: dodi004@brin.go.id (D. Safari).

https://doi.org/10.1016/j.jmii.2025.01.007
Received 10 September 2024; Received in revised form 17 January 2025; Accepted 23 January 2025
Available online 11 February 2025

manifestations.[9] SPN3 has different capsular polysaccharide (CPS) biosynthesis from other serotypes via synthase-dependent pathway, leading to profuse production and release of capsule which overwhelms the protective capacity of antibody that is elicited by the vaccine.[10,11]

SPN3 lineages exhibit significant geographic variation and a propensity for higher segregation and recombination than other serotypes. The distribution of SPN3 ST and GPSC was globally assessed under Global Pneumococcal Sequencing Project.[12] Genomic analysis has revealed that SPN3 refined the population structure into three clades: 1α, 1β, and II.[13] A genomic study in England and Wales shows that clade II has expanded since 2014 and represents 50 % of IPD cases.[14] Furthermore, a recent study in Malawi also demonstrates that ST700-GPSC10 lineage are less susceptible to opsonophagocytic killing which could enhance vaccine escape and clonal expansion after PCV13 introduction.[15]

Geographical differences and high birth rate possess challenges for implementing vaccination programs in Indonesia. However, introduction of PCV13 with 2 + 1 schedule among children has begun from 2022 in Indonesia. Despite its inclusion in PCV13, genomic analysis of SPN3 before and after the introduction of PCV is crucial for understanding the evolution and adaptation of this serotype in response to vaccine introduction. Genome analysis may reveal population structure dynamics and the emergence of antibiotic-resistant subpopulations, which have implications for vaccine design and public health strategies.[16] Hence, this study aims to analyze genomic profile of collected isolates SPN3 in Indonesia prior to PCV introduction from 2017 to 2021 including circulating ST, GPSC, genomic AMR prediction, constructing pan-genome analysis, phylogenetic insight, and genome comparison of capsular polysaccharide (*cps*) locus.

2. Methods

2.1. Sample collection

Sample collections were conducted from 2017 to 2021 (n = 19) (Table 1). We obtained a total of 19 isolates from 14 carriage isolates and 5 clinical isolates from IPD patient. We collected isolates from nasopharyngeal swabs (NP swab) in children (n = 7) and adult above 18 years old (n = 7). We also obtained isolates from cerebrospinal fluid (CSF) (n = 1), sputum (n = 1), blood (n = 2), and bronchoalveolar lavage (BAL) (n = 1) in children and adult group.[5,6,17–19] A total of 19 isolates were confirmed as SPN3 by multiplex polymerase chain reaction (PCR) and Quellung reaction in the previous study.[5,18–20]

2.2. DNA extraction and sequencing

Archived isolates were cultured on 8 % sheep blood agar and incubated in 5 % of CO_2 for 18–24 h. On the following day, one loop of

Table 1
An overview of genomic profile SPN3.

No.	Sample ID	SRR code	Age (year)	Specimen Type	Region	Year of collection	GPSC	MLST	Allelic Profile						
									aroE	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*
1	SU0209A	SRR 25316461	55	Nasopharyngeal (NP) swab	Surabaya, East Java	2015	10	700	12	11	2	6	6	22	14
2	MA0208A	SRR 25316481	61	NP swab	Makassar, South Sulawesi	2015	12	180	7	15	2	10	6	1	22
3	MA020B	SRR 25316493	61	NP swab	Makassar, South Sulawesi	2015	309	5292	2	32	165	171	6	112	145
4	MA0126B	SRR 25316486	29	NP swab	Makassar, South Sulawesi	2015	51	458	2	32	9	47	6	21	17
5	MA0189B	SRR 25316483	63	NP swab	Makassar, South Sulawesi	2015	51	458	2	32	9	47	6	21	17
6	ME0080B	SRR 25316479	75	NP swab	Medan, North Sumatra	2015	12	505	46	8	2	10	6	1	22
7	S975	SRR 25316466	No data	Sputum	Jakarta	2019	234	451	1	423	28	1	9	1	17
8	12798236	SRR 25316518	67	Blood	Surabaya, East Java	2021	363	4909	2	12	19	1	6	1	22
9	12819835	SRR 25316517	32	Cerebrospinal fluid	Surabaya, East Java	2021	363	4909	2	12	19	1	6	1	22
10	2372563	SRR 25316544	78	Bronchoalveolar lavage	Jakarta	2021	12	180	7	15	2	10	6	1	22
11	RMD 139	SRR 25316469	63	NP swab	Manado, North Sulawesi	2020	12	180	7	15	2	10	6	1	22
12	2171 GK	SRR 29733325	4	NP swab	Gunung Kidul, Yogyakarta	2017	234	451	1	423	28	1	9	1	603
13	2402 GK	SRR 29733324	4	NP swab	Gunung Kidul, Yogyakarta	2017	234	451	1	423	28	1	9	1	603
14	2393 GK	SRR 29733323	2	NP swab	Gunung Kidul, Yogyakarta	2017	234	451	1	423	28	1	9	1	17
15	2311 GK	SRR 29733322	4	NP swab	Gunung Kidul, Yogyakarta	2017	12	180	7	15	2	10	6	1	*novel*
16	RA 3041	SRR 29733319	15	NP swab	Jakarta	2021	234	451	1	423	28	1	9	1	17
17	RA 1033	SRR 29733321	14	NP swab	Jakarta	2021	12	3805	7	203	40	1	6	1	22
18	RA 2033	SRR 29733320	13	NP swab	Jakarta	2022	12	3805	7	203	40	1	6	1	22
19	10592	SRR 25316550	<1	Blood	Medan, North Sumatra	2021	*Not assigned*	4233	15	16	1	16	6	6	14

overnight culture was inoculated into 5 mL of brain heart infusion and 1 mL of rabbit serum for 5 h as part of pre-lysis step. DNA extraction of *S. pneumoniae* was carried out using DNeasy Blood and Tissue® kit (Qiagen, Carlsbad, CA, USA) with pre-treatment of mutanolysin and lysozyme. Genomic DNA extracts was sequenced at PT Indolab Utama, Jakarta, Indonesia/Macrogen Co., Ltd., Singapore. Library preparation was performed using TruSeq Nano DNA kit on Illumina NovaSeq6000 platform according to the manufacture instructions with 2 x 150 bp and 100x coverage (Illumina, NE, USA).[21]

2.3. Identification of ST, GPSC, and genomic AMR prediction

Raw sequencing data were analyzed using ASA^{3}P bioinformatics pipeline (https://github.com/oschwengers/asap). ASA^{3}P pipeline were used for quality control (FastQC), trimming (Trimmomatic) and construct *de novo* genome assembly (SPAdes).[22] Genome assembly file output from ASA^{3}P were submitted to Pathogenwatch (https://pathogen.watch/) to obtain general overview of genomic profile SPN3 including species name, taxonomy ID, genome length (bp), GC content, serotype, sequence type (ST), GPSC, and genomic prediction of antimicrobial resistance (AMR). Pathogenwatch retrieved database from PubMLST for ST assignment. The outcomes of the assembly will be submitted concurrently to PubMLST (https://pubmlst.org/) to acquire an official ST assignment in cases Pathogenwatch identifies a new combination of existing allelic profile. Genomic AMR prediction were retrieved database from PAARSNP AMR Library 1313 version 0.0.16. Database of AMR genomic prediction in Pathogenwatch were included chloramphenicol, clindamycin, erythromycin, fluoroquinolones, kanamycin, linezolid, tetracycline, trimethoprim, sulfamethoxazole, and co-trimoxazole.

2.4. Pangenome analysis

Pangenome analysis is a comprehensive approach to understanding the genetic composition of SPN3 by combining the genetic information from multiple STs. A total of 19 isolates from 11 different ST were used to construct pangenome. The pangenome analysis was conducted by Anvi'O v.7.1 following pangenome workflow (https://merenlab.org/2016/11/08/pangenomics-v2/).[23]

2.5. Phylogenetic tree

We performed BacWGSTdb v.2.0 (http://bacdb.cn/BacWGSTdb/index.php) in whole SNP analysis to generate.nwk files output for phylogenetic tree visualization. Metadata of SPN3 and.nwk files were submitted to Microreact (https://microreact.org/) for visualization.

2.6. Genome comparative of cps locus

Comparative genome of *cps* locus SPN3 was performed by Geneviewer (https://github.com/nvelden/geneviewer) and R package for plotting gene clusters. Genome annotation was performed by Prokka (https://github.com/tseemann/prokka). The *cps* locus of SPN3 references were retrieved from NCBI (GenBank accession no. AP017971) with excision between *dexB* and *pgm* genes using Magnifying Genomes/MaGe (https://mage.genoscope.cns.fr/).

3. Results

The genome sequences of 19 isolates were used in this analysis (Table 1). Based on Pathogenwatch analysis result, all of samples were confirmed as SPN3 with taxonomy ID 1313. Genome length of all isolates was ranging from 1,979,077—2,080,630 bp with the average of GC content 39.6 %. In this study, we discover a diverse distribution of ST and GPSC from children and adult groups. In our study, there were no dominant STs found due to slightly different prevalence among sample. We found ST451-GPSC234 (n = 3), ST180-GPSC12 (n = 3) and similarly followed by ST458-GPSC51 (n = 2), ST3805-GPSC12 (n = 2), ST*2a38-GPSC234 (n = 2), and ST4909-GPSC363 (n = 2) (10.5 %). Moreover, we also found ST700-GPSC10 (n = 1), ST5292-GPSC309 (n = 1), ST505-GPSC12 (n = 1), and ST*f4fb-GPSC12 (n = 1). Remarkably, we also discover ST*225a which remains not assigned to any GPSC (n = 1). ST*2a38, ST*f4fb, and ST*225a was marked by asterisk symbol in Pathogenwatch result indicates novel or a combination of existing alleles resulting in a new ST, thus we conducted further submission on PubMLST to obtain an official ST assignment. We found that ST *2a38 has the nearest match with ST451 (85.7 % loci matched), whereas ST *f4fb has nearest match with ST180 (85.7 % loci matched), and ST *225a has nearest match with ST 4233 (74.1 % loci matched). Hence, this study found a total of ST 451-GPSC234 (n = 5), ST180-GPSC12 (n = 4), and ST 4233-not assigned GPSC (n = 1).

A phylogenetic tree was created to depicts the lines of evolutionary lineages of various ST (Fig. 1). In this study, we discover three major roots with small distance differences in general. Based on the result, isolate ST4909 from the first root shares common ancestor. Additionally, we found ST180, ST505, and ST3805 from the second root shares common recent ancestor. Furthermore, ST451, ST458, ST5292, and ST700 shares common recent ancestor.

We discover various genomic prediction of AMR profile among samples despite of no inferred resistant isolate was predominantly found in this study (n = 11). However, most of isolate were found resistant to tetracycline (n = 5); followed by co-resistant of chloramphenicol and tetracycline (n = 2); co-trimoxazole and tetracycline (n = 1).

In comparison between adult and children group, we found non-resistant isolates (n = 7), tetracycline resistant (n = 4), tetracycline and co-trimoxazole resistant (n = 1), tetracycline and chloramphenicol resistant (n = 1) in adult group. Conversely, we found non-resistant isolates (n = 4), and tetracycline resistant isolate (n = 4) in children group. We also discover AMR-associated genes and mobile genetic element (MGE) that confer resistance to antibiotics in our study. In general, most of isolate possess *tetM* gene and constituting Tn916 since tetracycline resistant were predominantly found in this study. On the other side, isolate that possess *cat* and *tetM* genes were constitute Tn5253. Moreover, we discover co-trimoxazole resistant was carried by *folA* and *folP* genes. We provide Antimicrobial Susceptibility Testing (AST) data from the previous study as supplementary information (Tables 2a and 2b) using disc diffusion and Minimum Inhibitory Concentration (MIC) method.[5,19,20,24]

We use AST data in the previous study to validate genomic prediction of AMR profile in this study. The genomic prediction AMR result was in line with AST result, except 2 samples (RA 1033 and RA 2033) which demonstrate resistant to tetracycline in genomic AMR result, whereas the AST result shows sensitive result to tetracycline (Tables 2a and 2b).

Pangenome analysis involves 11 ST from SPN3 isolates (Fig. 2). These genomes were found to consist of a single circularized contigs with an average length of 2,023,379 bp. The pangenome analysis delineated 4912 gene families which consist of 4028 core genomes, 884 accessory genomes, and 700 singletone. The core genome refers to the set of genes that are present in all ST of SPN3. These core genomes are essential for the survival and function of the bacteria, whereas the accessory genome consists of genes that are not present in all ST but are found in some or many STs. Accessory genomes are often involved in specific functions such as virulence, antibiotic resistance, or adaptation to different environments. In this study, we found that the number proportion of core genome families is higher than accessory genome families, which seems to possess a closed pangenome characteristics. Closed pangenome typically has a large core genome and small accessory genome with fewer unique genes.

We confirmed this finding after performing *de novo* assemblies of the whole-genome sequences to reconstruct the *cps* locus. We analyze the region between *dexB* and *pgm* genes according to Chiba et al., (GenBank accession no. AP017971). Genome comparative of SPN3 *cps* locus is

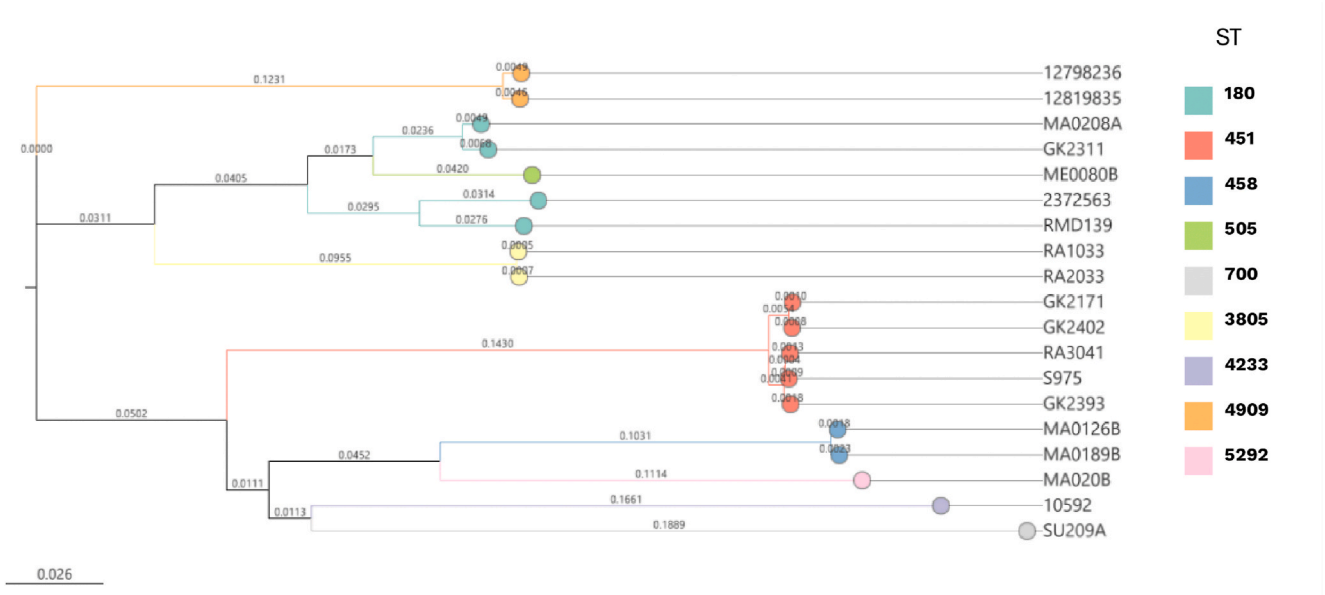

Fig. 1. Phylogenetic Tree of SPN3 illustrating genetic relationships between STs. Alignments of each individual were generated using BacWGSTdb v.2.0 to infer maximum likelihood tree from this ST alignment. Visualization of phylogenetic tree was constructed by Microreact.

Table 2a
Genomic AMR profile and AST profile using disc diffusion method.

Sample ID	Genomic AMR Profile	Antimicrobial susceptibility testing (AST) Profile				
		Chlorampenicol	Erythromycin	Clindamycin	Tetracycline	Trimethoprim/ Sulfamethoxazole
SU0209A	Tetracycline *tetM8*, co- trimoxazole *folA* 1I00L, *folP_aa* insert 57-70	S	S	S	I	R
MA0208A	–	S	S	S	S	S
MA020B	–	S	S	S	S	S
MA0126B	Chlorampenicol *catP* 194, Tetracycline *tetM4*	R	S	S	R	S
MA0189B	Chlorampenicol *catP* 194, Tetracycline *tetM4*	R	S	S	R	S
ME0080B	–	S	S	S	S	S
S975	–	S	S	S	S	S
RMD 139	–	S	S	S	S	S
RA 3041	–	S	S	S	S	S
RA 1033	Tetracycline *tet* 32	S	S	S	S	S
RA 2033	Tetracycline *tet* 32	S	S	S	S	S
2372563	Tetracycline *tet12*	N/A	N/A	N/A	N/A	N/A

S = Sensitive, I = Intermediate, R= Resistant, N/A= Not available.

Table 2b
Genomic AMR profile and AST profile using MIC method.

Sample ID	Genomic AMR profile	Antimicrobial susceptibility testing (AST) Profile																	
		PEN	LVX	MEM	AZM	TET	ETP	ERY	CXM	AMC	SXT	AXO	LZD	VAN	CTX	CLI	DAP	FEP	CHL
10592	Tetracycline *tetM*_12	≤0.03	=1	≤0.25	≤0.25	=8	≤0.5	≤0.25	≤0.5	≤2/1	≤0.5/9.5	≤0.12	=0.5	≤0.5	≤0.12	≤0.12	=0.06	≤0.5	≤1
12798236	-	≤0.03	=1	≤0.25	≤0.25	≤1	≤0.5	≤0.25	≤0.5	≤2/1	≤0.5/9.5	≤0.12	=1	≤0.5	≤0.12	≤0.12	=0.12	≤0.5	=2
12819835	-	=0.25	=1	≤0.25	≤0.25	≤1	≤0.5	≤0.25	≤0.5	≤2/1	≤0.5/9.5	≤0.12	=0.5	≤0.5	≤0.12	≤0.12	=0.12	≤0.5	≤1
2171 GK	-	≤0.03	=1	≤0.25	≤0.25	≤1	≤0.5	≤0.25	≤0.5	≤2/1	≤0.5/9.5	≤0.12	=1	≤0.5	≤0.12	≤0.12	=0.12	≤0.5	=2
2402 GK	-	≤0.03	=1	≤0.25	≤0.25	≤1	≤0.5	≤0.25	≤0.5	≤2/1	≤0.5/9.5	≤0.12	=0.5	≤0.5	≤0.12	≤0.12	=0.12	≤0.5	=2
2393 GK	-	≤0.03	=1	≤0.25	≤0.25	≤1	≤0.5	≤0.25	≤0.5	≤2/1	≤0.5/9.5	≤0.12	=0.5	≤0.5	≤0.12	≤0.12	=0.12	≤0.5	=2
2311 GK	Tetracycline *tetM*_32	≤0.03	=1	≤0.25	≤0.25	>8	≤0.5	≤0.25	≤0.5	≤2/1	≤0.5/9.5	≤0.12	=1	≤0.5	≤0.12	≤0.12	=0.12	≤0.5	≤1

Sensitive Resistant

PEN = Penicillin; LVX = Levofloxacin; MEM = Meropenem; AZM = Azithromycin; TET = Tetracycline; ETP = Ertapenem; ERY = Erythromycin; CXM = Cefuroxime; AMC = Amoxicillin/Clavulanic Acid 2:1; SXT = trimethoprim/sulfamethoxazole; AXO = Ceftriaxone; LZD = Linezolid; VAN = Vancomycin; CTX = Ceftaxime; CLI = Clindamycin; FEP = Cefepime CHL = chloramphenicol

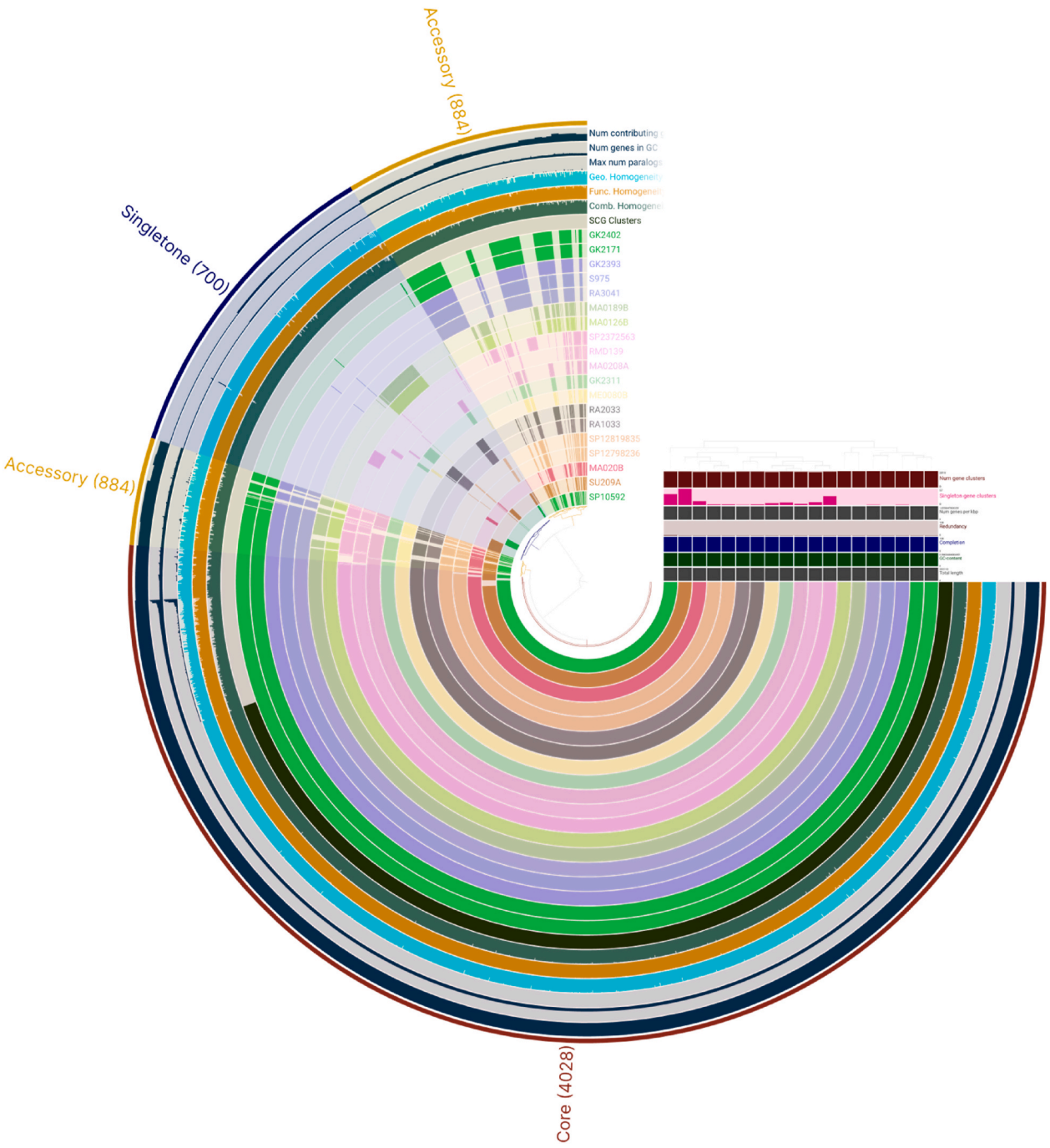

Fig. 2. Pangenome map of SPN3. In the outer side, core genomes (4,028) represent in red color, accessory genome in golden brown (884), and singletone in blue (700). Each rail in the inner side represents different samples and colored based on STs (green = ST451, light purple = ST451, sage green = ST458, pink = ST180, light green = ST180, light brown = ST505, black = ST3805, light orange = ST 4909, magenta = ST5292, brown = ST700, dark green = ST4233).

demonstrated in Fig. 3. We found that all ST has essential genes for SPN3 capsular biosynthesis (*ugd*, *wchE/cap3B*, and *galU*). We identified the absence of *dexB* and *aliB* gene in ST458 and ST5292. Furthermore, we found the absence of *pgm* gene in ST451, ST180, ST3805, and ST458. Notably, we observed multiple absent genes in ST700 including *cpsA/wzg*, *cpsB/wzh*, *cpsC/wzd*, *cpsD/wze*, and *tnp*.

4. Discussion

SPN3 is a significant cause of pneumonia, bacteremia, and meningitis in developing countries. Recent study shows that IPD in children and adults has been associated with SPN3, although the frequency of SPN3 infections varies by country and age group.[8,9] Serotype 3 infections are characterized by severe clinical manifestations, such as bacteremia-induced septic shock, meningitis, and pneumonia with mortality rate ranging from 30 to 47 %.[9]

A study in United States found that ST451 has been associated with IPD cases. Moreover, a study from Columbia discover that ST180 was highly prevalent in IPD isolates and exhibit macrolides resistance.[25] Conversely, this study found ST180-GPSC12 isolate from BAL specimen were predicted as tetracycline resistant which carried by Tn916. In worldwide study, the Global Pneumococcal Sequencing Project demonstrates that GPSC12 being the sole cluster that encompassed strains expressing serotype 3 and ST180 that were both susceptible and

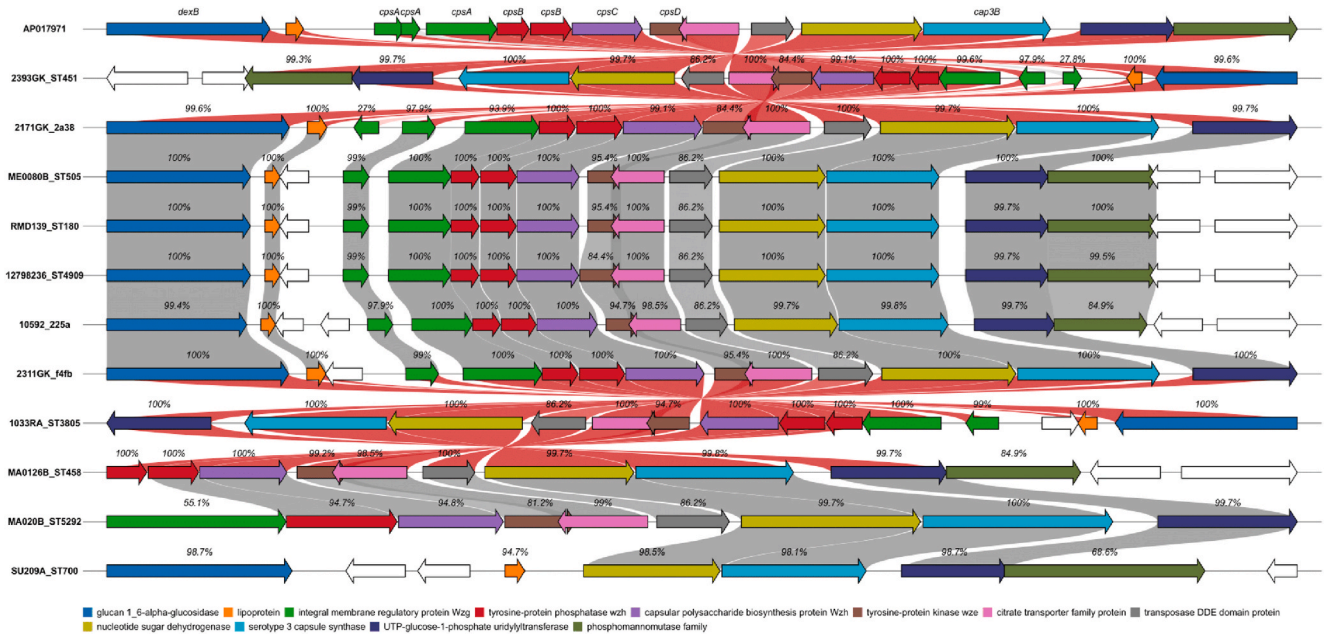

Fig. 3. Genome comparative of SPN3 *cps* locus in different ST using Geneviewer. A total of 11 isolates were used to represent diversity of ST in this study. All isolates were present in 5'→3′ direction except 2303 GK and 1033RA isolates in 3'→5′ direction.

resistant to certain treatments.[8] Hence, continuous monitoring and additional studies are necessary to determine the evolution and impact of ST distribution in SPN3 isolates.

This study shows that majority of SPN3 isolates shows no inferred resistant in predicted genomic AMR profile (57.9 %). A study in Moscow discover similar findings which demonstrate SPN3 are largely preserved susceptibility to the tested antibiotics and none of SPN3 isolates shows multidrug resistant phenotype.[26] Furthermore, a study on serotype competence level in *S. pneumoniae* demonstrate that SPN3 were less competent with low genetic heterogeneity. Large amount of capsular polysaccharide has been reported to have an inhibitory effect on transformation in *S. pneumoniae*, and therefore relatively rich amounts of capsular polysaccharide in serotype 3 may block uptake of foreign DNA, or in other words, it has limited capacity to facilitate capsular transformation through horizontal DNA transfer.[27] We thought that it plays significant role in evolution of SPN3, particularly in virulence levels and AMR rates. Our investigation was in line with a study in China which demonstrate that SPN3 has low AMR rates compared to serotype 23F, 19F, 19A, and 14A which exhibited high AMR rates due to greater genetic diversity and ability to undergo recombinational exchanges.[28]

Conversely, a study of mucoid SPN3 from children in China shows that SPN3 has high susceptibility to numerous antibiotics, especially β-lactams, but resistant to macrolides (100 %). These study also shows that 70 % of isolates are resistant to tetracyline.[29] Here, we indicated a high prevalence of tetracycline resistant (42.1 %) in various lineages including ST458-GPSC51, ST180-GPSC12, ST180-GPSC12, ST3805-GPSC12, ST700-GPSC10, and ST4233. Another finding shows that ST271 and ST700 exhibit tetracycline resistance via *tetM* gene, which mediates ribosome alteration.[15,30] Furthermore, a study on serotype distribution and AMR profiles in adult patients with community-acquired pneumonia in Jakarta, Indonesia reported that SPN3 being the most prevalent serotype with 44 % of isolates were resistant to tetracycline.[5] Therefore, the resistance pattern of SPN3 should be closely monitored since population dynamics of serotype were influenced by antibiotic use and vaccination programs.

This study also highlighted higher proportion number of core genome in pangenome analysis, which displays a closed pangenome characteristic. Closed pangenomes typically have large core genomes, which are the set of genes that are present in all members of species or genus.[31] A closed pangenome indicates genetic diversity within SPN3 is relatively stable and does not increase significantly with the addition of more genomes. Moreover, molecular typing studies have shown that low genetic diversity is associated with the short duration of nasopharyngeal carriage and its propensity to cause an invasive infection.[8] The low genetic diversity is thought to contribute to the persistence of SPN3 as a significant cause of IPD despite the introduction of PCV. These also allows pneumococcus to adapt and evolve more efficiently in response to selective pressure including vaccination.[32] Notably, a recent study of SPN3 ST700-GPSC10 lineage demonstrate clonal expansion after PCV13 introduction could enhance vaccine escape.[11]

This study also highlights ST700-GPSC10 lineage as particular concerns, since the previous study shows that SPN3 was found as most prevalent serotype before and after Hajj pilgrimage.[6] We found that SU209A isolate was resistant to tetracycline and co-trimoxazole. A recent study in Malawi also shows that ST700-GPSC10 lineage is notable for its increased antimicrobial resistance.[11] Our study shows that SU209A has the highest line (0.19) in phylogenetic tree among others, represents the most recent common ancestor of all STs. Moreover, 5 absent genes in ST700-GPSC10 were found in *cps* locus including *cpsA/wzg*, *cpsB/wzg*, *cpsC/wzh*, *cpsD/wze*, and *tnp*. Hence, we explored the involvement of these genes in SPN3 CPS biosynthesis pathway.

SPN3 produce CPS via a synthase-dependent pathway which involves a fewer gene and less complex mechanism compared to Wzx/Wzy-dependent pathway in majority serotypes of pneumococcus. CPS of SPN3 has a simple disaccharide repeat unit consist of glucose (Glc) and glucuronic acid (GlcA). Synthase-dependent pathway involves two primary genes including *ugd/cps3D*, which encodes UDP-Glc dehydrogenase for the synthesis of UDP-GlcA; and *wchE/cps3S/cap3B* which encodes for 3 synthase, a processive β-glycosyltransferase that connects the alternate Glc and GlcA components through specific glycosidic bonds.[9,33] Loss of any of these enzymatic functions leads to inability to produce SPN3 capsule, resulting non-encapsulated *S. pneumoniae* strain.[34]

This study found the absence of *pgm* genes in 4 isolates. The *pgm* gene or phosphoglucomutase (PGM) plays essential for catalyze the conversion of Glc-6-P to Glc-1-P as the precursors for SPN3 capsule production. Although *pgm* is not prominent as *ugd* and *wchE* role in SPN3 capsule production, a study shows that mutations in *pgm* gene were reduced-

capsule isolates. Insertion mutants that lacked PGM activity were avirulent in both immunologically normal (BALB/cByJ) and immunodeficient (CBA/N) mice.[34]

We also found the absence of *cpsABCD* genes in ST700-GPSC10 lineage. However, as SPN3 involved *wchE* and *ugd* in capsule production, the specific involvement of *cpsABCD* genes in SPN3 capsule production remains to be determined. Notably, the emergence of the ST700-GPSC10 lineage is thought to initially emerged from capsular switching, from serotype 19A/19F to serotype 3. This capsule switching phenomenon has been linked to obtaining a truncated *cps* locus lack of *cpsABCD* genes. Hence, it is postulated that the absence of these genes in the ST700-GPSC10 is more likely due to gene loss rather than never been present.[11] Furthermore, we suggest further study to evaluate this phenomenon in clinical manifestations using *in vivo* study to understand different impact in SPN3 strains.

Evidence suggests a growing support for the theory of bacterial evolution through the process of gene loss and reduction of the genome, which has been observed in various species and strains of *Streptococcus*. The precise impact of the shortened version of the SPN3 *cps* locus on the configuration of the SPN3 capsule remains elusive and needs further investigation. However, there is a hypothesis that a shorter, yet functional SPN3 *cps* locus could potentially be easily obtained by non-SPN3 strains, including those of vaccine serotypes, thereby facilitating the switching of capsules. It is essential to conduct further research in order to fully grasp the potential ecological benefits of the shorter length of SPN3 *cps* locus.[11]

This study used short-read sequencing platform to conduct WGS. Short-read sequencing technologies may not span repetitive regions or complex genomic structure within *cps* locus leading to fragmented assemblies with multiple contigs. Unfortunately, we do not perform further analysis using PCR or Sanger sequencing as validation. However, we conducted validation in our sample (SU209A) from ST 700 with reference of ST 700 in another study (accession number OR805040) and found similarities of absent genes between both sequences.[15] Hence, this study remain suggests that WGS method remains reliable for determine genomic comparative of *cps* locus.

5. Conclusion

This study provides a comprehensive insight for genomic characteristics of SPN3 from children and adult groups from Indonesia prior to PCV intoduction. We found a diverse set of STs in this study. Genomic AMR prediction shows that most of SPN3 isolates exhibit low AMR rates with 57.9 % no inferred resistance. Our study demonstrates a closed pangenome characteristic among STs which may indicate more stable genetic repertoire. We also highlight ST700-GPSC10 for its AMR profile since it has the highest distance line in phylogenetic tree analysis and shows the multiple absence genes in *cps* locus. We suggest an extended genomic profiling of SPN3 following PCV introduction in Indonesia, paired with immunological studies to understand genetic dynamics and its implications with clinical impact.

CRediT authorship contribution statement

Ratna Fathma Sari: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Fadilah Fadilah:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Data curation, Conceptualization. **Yustinus Maladan:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation. **Rosantia Sarassari:** Writing – review & editing, Validation, Supervision, Methodology, Data curation. **Miftahuddin Majid Khoeri:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Kuntjoro Harimurti:** Writing – review & editing, Validation, Supervision, Investigation. **Lindawati Alimsardjono:** Writing – review & editing, Validation, Supervision, Data curation. **Dodi Safari:** Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Data availability statement

Whole-genome data have been deposited in National Center for Biotechnology Information (NCBI) under Bioproject accession number PRJNA995903 and PRJNA1132753. Bioproject ID PRJNA995903 were include SRR25316461, SRR25316466, SRR25316469, SRR25316479, SRR25316481, SRR25316486, SRR25316493, SRR25316517, SRR25316550, SRR25316544, SRR25316518. Bioproject ID PRJNA1132753 were include SRR29733325, SRR29733324, SRR29733323, SRR29733322, SRR29733321, SRR29733320, and SRR29733319.

Declarations of competing interest

The publication fee was supported by Merck Sharp & Dohme (MSD) Indonesia.

Acknowledgements

This research was supported by the RIIM LPDP Grant and National Research and Innovation Agency of the Republic Indonesia (BRIN) with grant number: 37/II.7/HK/2023 We also thanks to the team of The Research Organization for Health and The Eijkman Institute for Molecular Biology for their support during this research project.

References

1. Cleary DW, Jones J, Gladstone RA, et al. Changes in serotype prevalence of Streptococcus pneumoniae in Southampton, UK between 2006 and 2018. *Sci Rep*. 2022 Aug 3;12(1), 13332.
2. McLaughlin JM, Jiang Q, Gessner BD, et al. Pneumococcal conjugate vaccine against serotype 3 pneumococcal pneumonia in adults: a systematic review and pooled analysis. *Vaccine*. 2019 Oct 8;37(43):6310–6316.
3. Sings HL, De Wals P, Gessner BD, et al. Effectiveness of 13-valent pneumococcal conjugate vaccine against invasive disease caused by serotype 3 in children: a systematic review and meta-analysis of observational studies. *Clin Infect Dis*. 2019 May 30;68(12):2135–2143.
4. Daningrat WOD, Amalia H, Ayu IM, Satzke C, Safari D. Carriage of Streptococcus pneumoniae in children under five years of age prior to pneumococcal vaccine introduction in Southeast Asia: a systematic review and meta-analysis (2001–2019). *J Microbiol Immunol Infect*. 2022 Feb;55(1):6–17.
5. Amanda G, Tafroji W, Sutoyo DK, Burhan E, Haryanto B, Safari D. Serotype distribution and antimicrobial profile of Streptococcus pneumoniae isolated from adult patients with community-acquired pneumonia in Jakarta, Indonesia. *J Microbiol Immunol Infect*. 2021 Dec;54(6):1175–1178.
6. Harimurti K, Saldi SRF, Dewiasty E, et al. Streptococcus pneumoniae carriage and antibiotic susceptibility among Indonesian pilgrims during the Hajj pilgrimage in 2015. Karunasagar I. *PLoS One*. 2021 Jan 26;16(1), e0246122.
7. Lapidot R, Shea KM, Yildirim I, Cabral HJ, Pelton SI. Characteristics of serotype 3 invasive pneumococcal disease before and after universal childhood immunization with PCV13 in Massachusetts. *Pathogens*. 2020 May 21;9(5):396.
8. Alarcón ZK, Duarte C, Sanabria O, Moreno J. Streptococcus pneumoniae serotype 3 genotypes in invasive isolates from Colombia. *Biomedica*. 2021 Jun 15;41(2): 338–346.
9. Luck JN, Tettelin H, Orihuela CJ. Sugar-coated killer: serotype 3 pneumococcal disease. *Front Cell Infect Microbiol*. 2020;10 [Internet] [cited 2023 Aug 28] https://www.frontiersin.org/articles/10.3389/fcimb.2020.613287.
10. Choi EH, Zhang F, Lu YJ, Malley R. Capsular polysaccharide (CPS) release by serotype 3 pneumococcal strains reduces the protective effect of anti-type 3 CPS antibodies. *Clin Vaccine Immunol*. 2016 Feb;23(2):162–167.
11. Kalizang'oma A, Swarthout TD, Mwalukomo TS, et al. Clonal expansion of a Streptococcus pneumoniae serotype 3 capsule variant sequence type 700 with enhanced vaccine escape potential after 13-valent pneumococcal conjugate vaccine introduction. *J Infect Dis*. 2024 Mar 26, jiae040.
12. Bentley SD, Lo SW. Global genomic pathogen surveillance to inform vaccine strategies: a decade-long expedition in pneumococcal genomics. *Genome Med*. 2021 May 17;13:84.
13. Cleary DW, Jones J, Gladstone RA, et al. Changes in serotype prevalence of Streptococcus pneumoniae in Southampton, UK between 2006 and 2018. *Sci Rep*. 2022 Aug 3;12, 13332.
14. Groves N, Sheppard CL, Litt D, et al. Evolution of Streptococcus pneumoniae serotype 3 in England and Wales: a major vaccine evader. *Genes*. 2019 Oct 25;10 (11):845.

15. Kalizang'oma A, Swarthout TD, Mwalukomo TS, et al. Clonal expansion of a Streptococcus pneumoniae serotype 3 capsule variant sequence type 700 with enhanced vaccine escape potential after 13-valent pneumococcal conjugate vaccine introduction. *J Infect Dis*. 2024 Mar;26, jiae040.
16. Azarian T, Grant LR, Arnold BJ, et al. The impact of serotype-specific vaccination on phylodynamic parameters of Streptococcus pneumoniae and the pneumococcal pan-genome. *PLoS Pathog*. 2018 Apr 4;14(4), e1006966.
17. Khoeri MM, Maladan Y, Salsabila K, et al. Whole genome sequencing data of Streptococcus pneumoniae isolated from Indonesian population. *Data Brief*. 2024 Apr;53, 110251.
18. Purwanto DS, Khoeri MM, Tafroji W, et al. Nasopharyngeal carriage rate, serotype distribution, and antimicrobial profiles of Streptococcus pneumoniae among patients with acute respiratory tract infection in Manado, North Sulawesi, Indonesia. *Access Microbiol*. 2024 Mar 15;6(3), 000703.v4.
19. Safari D, Daningrat WOD, Milucky JL, et al. Nasopharyngeal carriage of Streptococcus pneumoniae among children <5 years of age in Indonesia prior to pneumococcal conjugate vaccine introduction. *PLoS One*. 2024 Jan 11;19(1), e0297041.
20. Harimurti K, Saldi SRF, Dewiasty E, et al. Streptococcus pneumoniae carriage and antibiotic susceptibility among Indonesian pilgrims during the Hajj pilgrimage in 2015. *PLoS One*. 2021 Jan 26;16(1), e0246122.
21. Khoeri MM, Maladan Y, Salsabila K, et al. Whole genome sequencing data of Streptococcus pneumoniae isolated from Indonesian population. *Data Brief*. 2024 Apr;53, 110251.
22. Schwengers O, Hoek A, Fritzenwanker M, et al. ASA3P: an automatic and scalable pipeline for the assembly, annotation and higher-level analysis of closely related bacterial isolates. *PLoS Comput Biol*. 2020 Mar 5;16(3), e1007134.
23. Eren AM, Kiefl E, Shaiber A, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol*. 2021 Jan;6(1):3–6.
24. Weinstein MP, Lewis JS. The clinical and laboratory standards Institute subcommittee on antimicrobial susceptibility testing: background, organization, functions, and processes. *J Clin Microbiol*. 2020 Feb 24;58(3), e01864, 19.
25. Gertz RE, McEllistrem MC, Boxrud DJ, et al. Clonal distribution of invasive pneumococcal isolatesfrom children and selected adults in the United States prior to7-valent conjugate VaccineIntroduction. *J Clin Microbiol*. 2003 Sep;41(9): 4194–4216.
26. Mayanskiy N, Alyabieva N, Ponomarenko O, et al. Serotypes and antibiotic resistance of non-invasive Streptococcus pneumoniae circulating in pediatric hospitals in Moscow, Russia. *Int J Infect Dis*. 2014 Mar;20:58–62.
27. Hsieh YC, Wang JT, Lee WS, Hsueh PR, Shao PL, Chang LY, et al. Serotype Competence and Penicillin Resistance in Streptococcus pneumoniae - Volume 12, Number 11—November 2006 - Emerging Infectious Diseases journal - CDC. [cited 2024 Jul 7]; Available from: https://wwwnc.cdc.gov/eid/article/12/11/06-0414_article.
28. Zhou M, Wang Z, Zhang L, et al. Serotype distribution, antimicrobial susceptibility, multilocus sequencing type and virulence of invasive Streptococcus pneumoniae in China: a six-year multicenter study. *Front Microbiol [Internet]*. 2022 Jan 13;12 [cited 2024 Jul 7] https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.798750/full.
29. Yang Y, Hua CZ, Fang C, et al. Properties of mucoid serotype 3 Streptococcus pneumoniae from children in China. *Front Cell Infect Microbiol [Internet]*; 2021 Mar 24 [cited 2024 Jul 3]; volume 12:798750. Available from: https://www.frontiersin.org/journals/cellular-and-infection-microbiology/articles/10.3389/fcimb.2021.648040/full.
30. Scherer EM, Beall B, Metcalf B. Serotype-switch variant of multidrug-resistant Streptococcus pneumoniae sequence type 271. *Emerg Infect Dis*. 2021 Jun;27(6): 1689–1692.
31. Richard GF. Eukaryotic pangenomes. In: Tettelin H, Medini D, eds. *The Pangenome [Internet]*. Cham: Springer International Publishing; 2020:253–291 [cited 2024 Jul 1] http://link.springer.com/10.1007/978-3-030-38281-0_12.
32. Cella E, Sutcliffe CG, Grant LR, et al. Streptococcus pneumoniae serotype 3 population structure in the era of conjugate vaccines, 2001–2018. *Microb Genom*. 2024 Mar 18;10(3), 001196.
33. Paton JC, Trappetti C. Streptococcus pneumoniae capsular polysaccharide. *Microbiol Spectr*. 2019 Apr 12;7(2). https://doi.org/10.1128/microbiolspec.gpp3-0019–2018.
34. Hardy GG, Magee AD, Ventura CL, Caimano MJ, Yother J. Essential role for cellular phosphoglucomutase in virulence of type 3 Streptococcus pneumoniae. *Infect Immun*. 2001 Apr;69(4):2309–2317.