Journal of Applied Microbiology, 2026, **137**(2), lxag040
https://doi.org/10.1093/jambio/lxag040
Published: 6 February 2026
Research article

JOURNAL OF APPLIED MICROBIOLOGY OXFORD

Comparative whole-genome sequencing of seven invasive *Streptococcus pneumoniae* isolates from Malaysia reveals genetic diversity, recombination events, and global lineage linkages

Nurul Diana Dzaraly[1,2], AbdulRahman Muthanna[2], James John[3], Siti Norbaya Masri[4], Zarizal Suhaili[5], Nurshahira Sulaiman[2], Nor Iza A. Rahman[6], Tuan Suhaila Tuan Soh[7], Fatimah Haslina Abdullah[8], Sangita Biswas[1], Mazen M. Jamil Al-Obaidi[9], Mohd Nasir Mohd Desa [2,*]

[1]Department of Preclinical Sciences, Faculty of Dentistry, MAHSA University, 42610 Jenjarom, Selangor, Malaysia
[2]Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[3]Department of Medical Laboratory Technology and Biomedical Engineering, School of Allied Health Sciences, Sathyabama Institute of Science and Technology, Chennai 600119, India
[4]Department of Medical Microbiology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[5]Faculty of Bioresources and Food Industry, Universiti Sultan Zainal Abidin, Besut Campus, 22200 Besut, Terengganu, Malaysia
[6]Faculty of Medicine, Universiti Sultan Zainal Abidin, 20400 Kuala Terengganu, Terengganu, Malaysia
[7]Department of Pathology, Sungai Buloh Hospital, Ministry of Health Malaysia, Jalan Hospital, 47000 Sungai Buloh, Selangor, Malaysia
[8]Department of Pathology, Sultanah Nur Zahirah Hospital, Ministry of Health Malaysia, Jalan Sultan Mahmud, 20400 Kuala Terengganu, Terengganu, Malaysia
[9]Science Department (Biology Unit), College of Education, University of Technology and Applied Sciences, Rustaq 329, Sultanate of Oman
*Corresponding author. Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia. E-mail: mnasir@upm.edu.my

Abstract

Background ***Streptococcus pneumoniae* remains a major global health threat, causing diseases ranging from mild respiratory infections to severe conditions like pneumonia, sepsis, and meningitis. Although pneumococcal conjugate vaccines (PCVs), including PCV7, PCV10, and PCV13 have significantly reduced disease burden, especially in children, *S. pneumoniae* continues to exhibit high serotype and genetic diversity. Whole-genome sequencing (WGS) analysis offers high-resolution insights into clonal lineages and multidrug-resistant strains. However, genomic data on Malaysian isolates remain limited.**

Methods **This study characterized the whole genome features and comparative profiles of seven invasive *S. pneumoniae* isolates from two tertiary hospitals in Malaysia. WGS analyses described serotype, sequence type (ST), antimicrobial resistance determinant genes, pan-genome structure, and recombination events.**

Results **The average genome size was ∼2.12 Mbp, with 1 988–2 205 coding sequences. WGS-based MLST identified five sequence types (ST236, ST320, ST386, ST671, ST695), with ST236 linked to serotypes 19A and 19F related to PMEN clones Taiwan^{19F}-14 and CC271. Core genome analysis with 35 global reference strains revealed three major clades. Notably, isolates TSP95, SSP45, and SSP46 clustered closely with strains from South Korea, suggesting a long-term persistence of ST320 over a decade. Recombination analysis identified both shared and isolate-specific events, forming distinct phylogenetic clusters. Extensive shared recombination was observed in several isolates, while others displayed isolate-specific events, indicating ongoing genetic diversification.**

Conclusion **These findings underscore the critical role of recombination in shaping pneumococcal population structure, evolution, and adaptation.**

Impact statement

This study shows that recombination plays a major role in shaping the genetic diversity and evolutionary dynamics of invasive *Streptococcus pneumoniae* isolates in Malaysia, highlighting their global lineage connections and providing important insights into pneumococcal adaptation and epidemiology.

Keywords *Streptococcus pneumoniae*, whole-genome sequencing, genetic diversity, recombination events, global lineage

Received: 20 October 2025. **Revised:** 15 January 2026. **Accepted:** 25 January 2026

Introduction

Streptococcus pneumoniae (pneumococcus) is a major global pathogen responsible for a wide range of diseases, such as otitis media, sinusitis, to severe invasive infections such as pneumonia, sepsis, and meningitis (Chavanet 2012). These infections contribute substantially to morbidity and mortality, especially among young children, elderly, and immunocompromised individuals (Wahl et al. 2018). More than 100 pneumococcal serotypes have been identified, and differences in the antigenic properties of their capsular polysaccharides influence virulence, prevalence, and antimicrobial resistance patterns (Geno et al. 2015, Ganaie et al. 2020). The introduction of pneumococcal conjugate vaccines (PCVs) has markedly reduced disease burden. PCV7, introduced in 2000, targeted seven serotypes: 4, 6B, 9V, 14, 18C, 19F, and 23F (Ansaldi et al. 2011). This was followed by PCV10 (adding serotypes 1, 5, and 7F) in 2008 and PCV13 (adding serotypes 3, 6A, and 19A) in 2009 (Diez-Domingo et al. 2013). These vaccines have significantly decreased pneumonia-related hospitalizations, particularly in children (Alicino et al. 2017).

According to the Department of Statistics Malaysia, pneumonia was the two leading causes of death in 2019, accounting for 12.2% of all deaths, following ischaemic heart disease (15.0%) (Department of Statistics Malaysia 2020). Finding from a local study indicates that *S. pneumoniae* is a major contributor to severe childhood pneumonia and invasive infections (Arushothy et al. 2019). Malaysia continues to experience a high burden and mortality rate of pneumonia, with evidence indicating that early cessation of breastfeeding, exposure to second-hand smoking, and limited access to pharmaceutical interventions are associated with poorer outcomes among children with pneumonia (Ooi et al. 2019). To mitigate the burden of pneumococcal infections among children, the 10-valent pneumococcal conjugate vaccine (PCV10) was incorporated into the National Immunization Programme (NIP) in Malaysia in late 2020 (Lister et al. 2023) with expected reductions in serious paediatric pneumococcal infections, including pneumonia, meningitis, and sepsis. Despite the introduction of the PCV, post-implementation surveillance in Malaysia is inadequate, limiting the ability to predict future disease burden, monitor serotype shifts, and assess herd immunity (Gunasegaran et al. 2025).

Meanwhile, multilocus sequence typing (MLST) has long served as a standard method for analysing pneumococcal epidemiology and evolution through sequence type (ST) and clonal complex (CC) assignments (Enright and Spratt 1998). Whole-genome sequencing (WGS), however, offers a higher resolution for determining genetic relatedness and inferring phenotypic traits such as virulence and antimicrobial resistance (Ben Zakour et al. 2012). Therefore, WGS has greatly improved the ability to study the evolutionary relationships among closely related strains, evaluate vaccine effects, analyse genomic diversity, and track the emergence and spread of multidrug-resistant clones (Jindal et al. 2018, Almeida et al. 2021, Rodriguez-Ruiz et al. 2024).

Despite these advances, pneumococcal genomic studies in Malaysia remain scarce, with one comparative analysis of 10 pneumococcal genomes published to date (Jindal et al. 2018). We had previously applied MLST for epidemiological investigation involving pneumococcal isolates from a few major tertiary hospitals in Malaysia but, at the genotypic level, limited to the circulating STs. Throughout our studies, an interesting observation was the long persistency of a few clonal groups such as ST236 and ST320 (Nathan et al. 2014, Dzaraly et al. 2020, Dzaraly et al. 2021, Dzaraly et al. 2023, Rahman et al. 2025). Expanded analysis using WGS would enable comprehensive genetic characterization of the local pneumococcal isolates, including insights into strain evolution, epidemiological patterns, and other associated genetic features. This study therefore aims to describe the whole-genome and comparative genomics of seven representative invasive pneumococcal isolates, involving the most and less common circulating STs in the Malaysian context.

Material and methods

Ethics approval

Acquisition of pneumococcal isolates from the hospital settings was previously obtained through the Medical Research and Ethics Committee of the Ministry of Health Malaysia (National Medical Research Register approval no. NMRR 17–1025-35696).

Bacterial identification

The seven clinical isolates were available in our laboratory's culture collection; acquired in previous studies from different patients between October 2017 and December 2019 at two major tertiary hospitals located in highly populated cities on the west (SSP) and east (TSP) coasts of peninsular Malaysia (Dzaraly et al. 2021, Dzaraly et al. 2023). The selection criteria for isolates included in the current WGS analysis focused on microbiological aspects of global interest; invasiveness (based on the site of isolation), ST, and previously determined antibiotic resistance profiles. The chosen isolates represented both common (ST236, ST320) and less common (ST386, ST671, ST695) pneumococcal STs previously identified in the Malaysian setting across the two hospitals.

DNA isolation and whole-genome sequencing

Genomic DNA was extracted from overnight *S. pneumoniae* cultures using cetyltrimethylammonium bromide (CTAB) method as previously described by Minas et al. (2011). The extracted DNA quality was evaluated using a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA).

Genome assembly and annotation

WGS of the representative *S. pneumoniae* isolates was conducted by a commercial laboratory services provider (Apical Scientific, Malaysia) using a HiSeq 2500 Genome Analyzer platform. Genome sequencing was performed with a minimum coverage depth of 150× with paired-ends reads. The quality of the generated reads from high-throughput NGS was assessed using FastQC v0.11.8 R2. Adapter and quality trimming were performed using Trimmomatic v0.39 with a sliding window of 4: Q20, removal of bases $<$ Q20, and a minimum read-length cutoff of 50 bp. Q20 and Q30 distributions were verified through FastQC reports. After removal adopter sequences, reads of each isolates were assembled *de novo* using St. Petersburg genome assembly; SPAdes v3.11.1, and Algorithmic Biology Lab (Bankevich et al. 2012). The draft genome assemblies were annotated using the Prokka program, which uti-

lizes the curated database from NCBI Prokaryotic Genome Annotation Pipeline (PGAP), to predict genes, especially proteins and coding sequences (CDS). Assembly quality was assessed using QUAST v5.0.2, yielding N50 values between 29 412 bp and 66 871 bp. Contamination screening with CheckM v1.1.3 confirmed >98.5% completeness and <1.5% contamination for all assemblies. To investigate the relatedness among the selected genomes, Average Nucleotide Identity (ANI) values were calculated as a measure of whole-genome similarity (Jain et al. 2018).

Genome characterization

MLST was performed by mapping short reads against the seven loci of *S. pneumoniae* on the MLST web server database (http://pubmlst.org/; Larsen et al. 2012), using the algorithm SRST2 v0.1.8 scheme. To identify the antibiotic resistance genes among the seven isolates, the draft genome contigs were submitted to the ResFinder web server (https://cge.cbs.dtu.dk/services/ResFinder; Bortolaia et al. 2020) and to the Comprehensive Antibiotic Resistance Database (CARD) web server using Resistance Gene Identifier (RGI) software (https://card.mcmaster.ca.analyze/rgi; Alcock et al. 2020). The antimicrobial resistance genes were carefully examined and chosen based on their hit value (sequence identity and coverage recommended by the database). Serotypes were identified using the SeroBA pipeline, which incorporates a comprehensive WGS-based serotyping database. SeroBA enables high-throughput serotyping that is fast, scalable, and resource-efficient, capable of detecting up to 92 *S. pneumoniae* serotypes with 100% specificity (Epping et al. 2018); it is a Python3-based, open-source tool that predicts pneumococcal serotypes from raw WGS reads with 98% accuracy concordance using a k-mer-based method (https://github.com/sanger-pathogens/seroba).

Pan- and core-genome reconstruction

Pan- and core-genome analyses were performed using Roary v3.13.0 with default parameters, specifying 95% BLASTp identity for gene clustering. Core genes were defined as presence in ≥ 99% of genomes. The 95% identity threshold is consistent with pneumococcal pan-genome studies and helps ensure accurate ortholog clustering. The total conserved, unique, and new genes were also calculated. The core genomic alignments obtained from the pneumococcal isolates were utilized to deduce the maximum-likelihood tree. FastTree was employed with 100 bootstraps and the GTR time-reversible model (Price et al. 2010). The resulting phylogenetic tree was then redrawn using the iTOL v4 (interactive tree of life) online application (Letunic and Bork 2019) with additional 35 pneumococcal genomes available in the GenBank, all included in the analysis (Supplementary S1).

Recombination detection

To detect recombination events in the seven isolates with an accurate phylogenetic analysis, the Gubbins software version 1.4.6 (Page et al. 2015) was used. Recombination refers to the exchange of genetic material between different DNA molecules, which can obscure the true evolutionary relationships among bacterial strains if it is not properly accounted for. Gubbins is specifically designed to address this by identifying loci with unusually high densities of base substitutions hallmark of recombination. These recombined regions are then excluded from the analysis to avoid distortions in the phylogenetic inference. Gubbins was performed to iteratively refine the core genome alignment by identifying and removing the regions affected by recombination. Gubbins v2.4.1 was run for five iterations using the GTR + Gamma model and default SNP-density filters. The recombination-filtered alignment was used for phylogenetic inference in RAxML v8.2.12 using the GTRGAMMA model with 1000 bootstrap replicates to ensure robust branch support. The resulting phylogenetic tree was therefore constructed solely on point mutations unaffected by recombination. This approach provides a more accurate evolutionary relationship among the isolates. After identifying and removing the recombination regions, the filtered data was used for Maximum Likelihood (ML) phylogenetic analysis with RAxML (version 8.2.12; Stamatakis 2014), which is a widely used software for phylogenetic tree construction. The core genome was aligned using the progressive Mauve algorithm, which creates multiple genome alignments by identifying regions of sequence conservation across the bacterial isolates. Mauve was used exclusively for whole-genome alignment and synteny inspection, while all recombination detection, mapping of recombinant regions, and visualization of red/blue blocks were generated solely from Gubbins v2.4.1 outputs. This separation ensures methodological clarity and avoids confusion between alignment and recombination inference tools. By excluding recombination regions with Gubbins and focusing on conserved, non-recombined sequences, a robust phylogenetic tree was constructed to better reflect the evolutionary relationships among the selected invasive pneumococcal isolates. For clarity, in the Gubbins recombination plot, red blocks represent shared recombination events occurring across multiple isolates, whereas blue blocks denote recombination events unique to a single genome.

GenBank and genome repository

Bio Project accession number was registered, followed by the Biosample registration of each isolate. This was followed by the submission and deposition of the sequences via the Genome Submission website (http://submit.ncbi.nlm.nih.gov/subs/genome/) (Table 1).

Results

Genomes assembly, general features, and annotation

The genome sequences of the respective isolates had been deposited in GenBank with accession numbers as listed in Table 1. Sequencing produced a mean coverage ranging from 158× to 221× across the isolates (TSP28: 165×, TSP95: 172×, TSP102: 221×, TSP106: 158×, SSP32: 189×, SSP45: 174×, SSP46: 181×). Genomes sequences for all the seven isolates produced 74 to 177 contig sequences. The seven invasive isolates had assembled genomes with an average length of approximately 2.11 Mbp, with an average G + C content of 39.5%. The range of CDS content spanned from 1988 to 2205, with 47 tRNA and 5 rRNA. Species identity of all seven isolates was confirmed by whole genome average nucleotide identity (ANI) analysis, which demonstrated 98%–99% ANI values with *S. pneumoniae* reference genomes.

Table 1 Characteristics and genomic features of the seven pneumococcal isolates.

Isolate ID	Site of isolation	Years	Serotype	MLST	rRNA	No of contigs	GC content (%)	No of CDS	Genome size (Mbp)	GenBank accession no	Locus tag	Biosample No
TSP28	Blood	2017	19A	ST236	5	102	39.8%	2009	2.06	JAFLMT000000000	J0S79	SAMN18117809
TSP95	Blood	2019	19A	ST320	5	168	39.4%	1998	2.07	JAFLMS000000000	J0S80	SAMN18117810
TSP102	Blood	2019	6E	ST386	5	166	39.3%	2186	2.22	JAHPIS000000000	KSW98	SAMN19791188
TSP106	Blood	2019	14	ST671	5	114	39.6%	2106	2.14	JAFEML000000000	JL103	SAMN17393262
SSP32	Blood	2018	19A	ST695	8	177	39.8%	2205	2.22	JAFEMM000000000	JTE97	SAMN17393263
SSP45	Blood	2019	19F	ST320	5	74	40.3%	2002	2.05	JAHPIT000000000	KSW95	SAMN19791189
SSP46	Blood	2019	19F	ST320	5	99	38.7%	1988	2.04	JAIOTU000000000	K7C87	SAMN20968556

TSP; isolates from east coast peninsular Malaysia, SSP; isolates from west coast peninsular Malaysia.

Identification of sequence types and serotypes

A WGS-based MLST search using the Center for Genomic Epidemiology web tool identified five different STs: three isolates were ST320, and one each was ST236, ST386, ST695, and ST671. WGS-serotyping identified three serotypes 19A, two serotypes 19F, and one serotype 6E and 14 each (Table 1). All these were vaccine serotypes, except for one isolate of non-vaccine serotype 6E.

Identification of antimicrobial resistance genes

As shown in Table 2, all seven isolates carried genes associated with fluoroquinolone resistance; *patA*, *patB*, and *pmrA*. The isolates also harboured various MLS family-associated resistance genes, carrying *erm(B)* (4/7; 57.1%), *mef(A)* (5/7; 71.4%), *msr(D)* (5/7; 71.4%), and *rlmA(II)* (7/7; 100%). The *tet(M)* (71.4%) was found in five isolates, while alterations in *pbp1a* (1/7; 14.2%) and *pbp2x* (3/7; 42.9%) were detected in a few isolates.

Genome analysis

The genomes of each isolate were analysed using Roary tool to estimate the number of shared genes among the included genomes to explore the genetic diversity and, the core and accessory genes composition of all the strains ($n = 42$; 7 invasive isolates of this study and 35 pneumococcal isolates/genomes available in GenBank). Table 3 shows the distribution of genes, whereby 1157 (24.7%) were classified as core genes and 3524 (75.3%) as accessory genes.

The phylogenetic tree was used to infer and compare the relationship between the seven genomes of this study against a diverse set of 35 pneumococcal genomes derived from the GenBank (Fig. 1). Clade I grouped strains from South Korea and Taiwan; TSP95, SSP45, and SSP46 clustered closely within this clade, associated with subclades ST320-19A strains. TSP28 of ST236-19A was also grouped in clade I suggesting its potentially related lineage with strains from South Korea and Taiwan. Meanwhile, TSP102 in this analysis was distinct with its own lineage that branched away in a separate direction of clade II. SSP32 was also grouped in this diverse clade and potentially linked to lineage associated with strains from Denmark. TSP106 was clustered in clade III, which was quite distantly clustered with strains WSP8 and WSP7 from Malaysia.

Recombination analysis

The recombination events were analysed in the core genome alignment to understand their role in the genetic diversity of the seven invasive pneumococcal isolates. The findings revealed one dominant lineage with several distinct lineages among the isolates, each marked by specific recombination events (Fig. 2). Based on the count of distinct recombinant DNA segments ('blocks') that Gubbins assigns to each tip in the tree, the number of recombination events per isolate was quantified, whereby TSP28 (54 events), TSP95 (61 events), SSP45 (67 events), and SSP46 (42 events) exhibited the highest levels of recombination, whereas TSP102 (12 events), SSP32 (18 events), and TSP106 (9

Table 2 Distribution of antimicrobial resistance determinant genes among genomes of the seven pneumococcal isolates.

Antibiotic class	Gene	TSP28	TSP95	TSP102	TSP106	SSP32	SSP45	SSP46
				Antibiotic efflux				
Fluoroquinolone	*pmrA*	/	/	/	/	/	/	/
	patA	/	/	/	/	/	/	/
	patB	/	/	/	/	/	/	/
				Antibiotic target protection				
Tetracycline	*tet(M)*	/	/	/	x	x	/	/
				Antibiotic target detection				
Macrolide	*erm(B)*	x	/	/	x	x	/	/
	mef(A)	/	/	x	/	x	/	/
Lincosamide	*msr(D)*	/	/	x	/	x	/	/
Streptogramin	*rlmA(II)*	/	/	/	/	/	/	/
				Antibiotic target detection				
Cephalosporin	*pb1a*	x	x	x	x	/	x	x
Cephamycin	*pbp2x*	x	x	/	/	/	x	x

'/' indicates presence of gene and 'X' indicate absence of gene.
MLS family (Macrolide, Lincosamide, and Streptogramin).

Table 3 Core genome analysis showing the number of core and accessory genes of the 42 pneumococcal genomes.

Gene types	Average covering	Number of genes
Core genes	(99% $\leq$ strains $\leq$ 100%)	1157
Soft core genes	(95% $\leq$ strains $<$ 99%)	306
Shell genes	(15% $\leq$ strains $<$ 95%)	1089
Cloud genes	(0% $\leq$ strains $<$ 15%)	2129
Total genes	(0% $\leq$ strains $\leq$ 100%)	4681

Soft core, shell, and cloud genes are categorized as accessory genes ($n = 3524$).

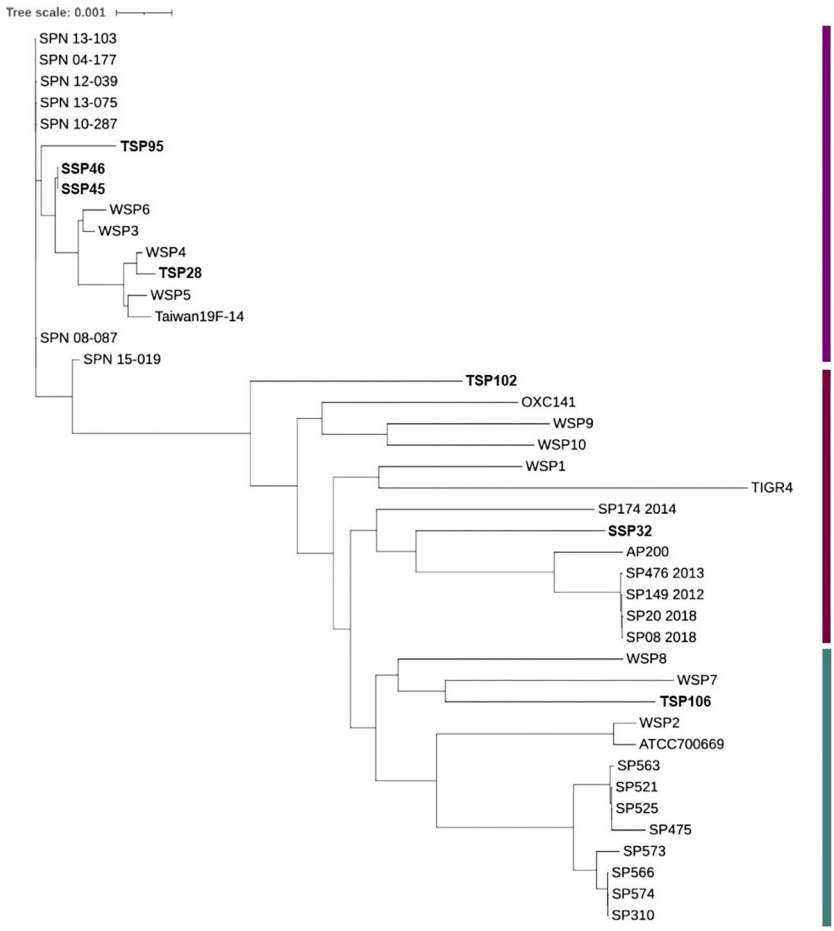
events) showed considerably fewer events. These numerical differences support lineage-specific variation in recombination intensity. The recombined gene functions listed in Table 4 indicate that recombination may influence virulence, capsule variation, and resistance–related traits in these isolates. The recombination events were more frequently observed across the remaining four isolates: TSP28, TSP95, SSP45, and SSP46. Isolates with the highest recombination counts (primarily ST320 serotype 19A/19F) also carried a higher number of antimicrobials–resistance determinants, as reflected by the multiple resistance–associated genes within their recombinant regions (Table 4). Several recombinant blocks mapped close to the cps region in the Gubbins plot (Fig. 2), the CDS listed in Table 4 themselves are not canonical cps genes but neighbouring metabolic or surface–associated genes. The genes located within these recombinant tracts include loci involved in capsule biosynthesis, surface–exposed and metabolic proteins, and several resistance–associated genes, suggesting that recombination can modulate serotype, virulence–associated functions, and antimicrobial resistance potential in these isolates.

Discussion

This study performed genomic comparison of seven invasive clinical isolates of *S. pneumoniae* in Malaysia, giving the average genome length of about 2.11 Mbp. The isolates had an average gene content of 2070, within the range of 1988 to 2205 CDS. The genomes were comparable in size to the reference strain *S. pneumoniae* TIGR4 (accession: GCA_000006885.1), which was 2.16 Mbp. The reference strain had 2153 CDS and shared similar gene content with the isolates in this study. The SPAdes *de novo* assemblies yielded an average of 128 contigs, with a range of 74–177. Among the predicted genes, 1157 (24.7%) were identified as part of the core genome across all the seven isolates of this study. These core genes accounted for 79.7% of the total translated genome, while the remaining 3524 (75.3%) were classified as accessory genes. The size of the bacterial core genome can differ across various isolate collections, primarily depending on how the core genome is defined and the parameters applied during pan-genome analysis. In most cases, the core genome consists of genes responsible for essential functions, including protein synthesis, DNA metabolism, and key cellular processes (Donkor et al. 2012).

ST320, a globally disseminating multidrug-resistant clone, accounted for three strains among the seven invasive pneumococcal isolates exhibiting serotypes 19A and 19F. The global prevalence of ST320-19A, not covered by the earlier PCV7 and PCV10 vaccines, has risen significantly due to serotype replacement. The emergence of the lineage is attributed to capsular gene exchange between ST236 strains of the CC271 clonal complex (Mott et al. 2019, Zeng et al. 2023). ST320 also demonstrated a colonization advantage over its ancestral PMEN Taiwan^{19F}-14 strain (Hsieh et al. 2013). This highlights both the ongoing challenge of serotype replacement and the persistency of specific clonal lineages carrying their infection capability and antibiotic resistance.

The antibiotic resistance genes were detected by the WGS database using the pipelines ResFinder and CARD for a comprehensive antibiotic resistance profile. The analysis showed that isolates TSP32, TSP102, and TSP106 appeared to carry altered *pbp1a* and *pbp2x* genes, while the earlier study by Dzaraly et al. (2021) using E-test method phenotypically showed no penicillin resistance in these isolates. Alteration of the transpeptidase domains of penicillin-binding proteins (PBPs) in β-lactam–



Serotype	ST	Source	Location
19A	320	Blood	South Korea
19A	320	Blood	South Korea
19A	320	Blood	South Korea
19A	320	Blood	South Korea
19A	320	Blood	South Korea
19A	**320**	**Blood**	**Malaysia**
19F	**320**	**Blood**	**Malaysia**
19F	**320**	**Blood**	**Malaysia**
14	1464	Nasopharyngeal	Malaysia
19F	271	Nasopharyngeal	Malaysia
14	5278	Nasopharyngeal	Malaysia
19A	**236**	**Blood**	**Malaysia**
23F	5278	Eye	Malaysia
19F	236	-	Taiwan
19A	320	Ear	South Korea
19A	320	Blood	South Korea
6E	**386**	**Blood**	**Malaysia**
3	180	-	UK
18	4901	Blood	Malaysia
8	6022	Blood	Malaysia
NT	-	Nasopharyngeal	Malaysia
4	205	Blood	Norway
8	404	Blood	Denmark
19A	**695**	**Blood**	**Malaysia**
11A	62	CSF	Italy
8	53	Blood	Denmark
8	53	Blood	Denmark
8	53	Blood	Denmark
8	53	Blood	Denmark
18	12838	Nasopharyngeal	Malaysia
1	51	Blood	Malaysia
14	**671**	**Blood**	**Malaysia**
31	81	Nasopharyngeal	Malaysia
23F	81	-	Spain
24F	15253	Blood	Lebanon
24F	15253	Blood	Lebanon
24F	15253	Pleural fluid	Lebanon
24F	15253	CSF	Lebanon
24F	14184	Blood	Lebanon
24F	11618	Blood	Lebanon
24F	11618	Blood	Lebanon
24F	11618	Blood	Lebanon

Figure 1 Core genome phylogenetic tree as visualized by iTOL Version 4 (interactive tree of life) based on seven isolates from this study (labelled with TSP; east coast peninsular Malaysia and SSP; west coast peninsular Malaysia) and 35 reference genomes. The distribution of serotype, sequence type (ST), sources, and location are shown in the right columns. Three major clusters (clade I—clade III) were observed with clade I representing genomes of isolates from Malaysia, Taiwan, and South Korea. The reference genomes were retrieved from GenBank comprising various origins; SPN (South Korea), WSP (Malaysia), TIGR4 (Norway), Taiwan19F-14 (Taiwan), OXC141 (United Kingdom), AP200 (Italy), ATCC700669 (Spain), SP (Denmark), and SP (Lebanon).

resistant pneumococcal isolates may lead to reduced susceptibility to cephalosporins, including penicillins but remain multifactorial to be demonstrated *in vitro*. (Zhou et al. 2022).

The genotypic features of erythromycin-resistant *S. pneumoniae* show geographical variation. In Asia, the *erm(B)* is the most prevalent macrolide resistance determinant gene in mainland China, Taiwan, Korea, and Sri Lanka, whereas the *mef(A)* is more dominant in Hong Kong, Singapore, Malaysia, and Thailand (Song et al. 2004). In this study, the *mef(A)* was detected in five out of the seven isolates (71.4%) followed by *erm(B)* in four isolates (57.1%), which were consistent phenotypically as previously reported by Dzaraly et al. (2021) and Dzaraly et al. (2023). The pneumococcal resistance to tetracycline is mostly due to ribosome protection, which is facilitated by the *tet(M)* and *tet(O)* mechanisms (Doherty et al. 2000, Dallas et al. 2013) as well as ribosomal mutation. In this study, the *tet(M)* was present in five out of the seven isolates. The association of macrolide and tetracycline resistance has been reported with the acquisition of *tet(M)*, possibly through the conjugative transposon Tn6002 that was frequently reported in Europe (Akdoğan Kittana et al. 2019, Kireeva and Dmitriev 2023). Additionally, both *erm(B)* and *tet(M)* were found on ICESp12ST230; integrative and conjugative element, which is a type of mobile genetic element found in *S. pneumoniae* (Lo et al. 2020).

Meanwhile, phylogenetic analysis was conducted to assess evolutionary relationships and genetic relatedness among the isolates together with the reference genome isolates derived from GenBank. The core genome analysis grouped the pneumococcal isolates into three major clades with distinct lineages, indicating that selected pneumococcal isolates in this study evolved separately along with their lineages. Notably, isolates TSP95, SSP45, and SSP46 clustered in subclade I, showed close relatedness to strains from South Korea, mainly ST320-19A isolated back in year 2004, suggesting this particular genetic group to have persisted for over a decade.

Additionally, SSP45 and SSP46, being closely positioned in the phylogenetic tree, were likely clone of a serotype 19F strain that had infected two individuals. TSP28 (ST236-19A) was classified in subclade I alongside five known strains from Taiwan and Malaysia isolated between 2011 and 2012, indicating over a decade of their global and local evolution. Other than ST236 and ST320, clade I also showed diverse sequence types (ST271, ST1464, and ST5278) with various vaccine serotypes. Isolate TSP102 (ST386-6E) evolved

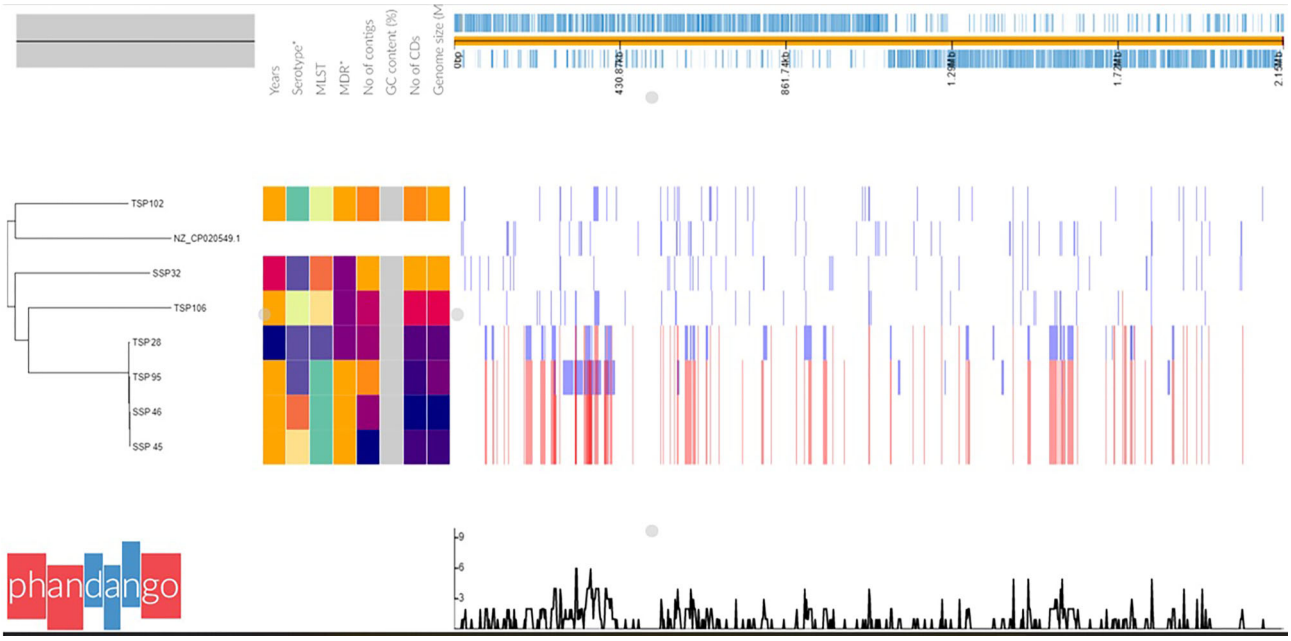


Figure 2 The recombination landscape of the seven invasive pneumococcal isolates, visualized using Gubbins and displayed in Phandango. On the left, the phylogenetic tree illustrates the genetic diversification of isolates with one dominant lineage. The heatmap next to the tree summarizes genomic features (serotype, MLST, GC content, genome size, etc.) for each isolate. The central panel shows recombination blocks across the genome, where red blocks represent shared recombination events found in multiple isolates, and blue blocks indicate isolate-specific recombination events. The ST236 and ST320 isolates (TSP28, TSP95, SSP45, and SSP46) display dense red regions, indicating high levels of shared recombination, whereas isolates TSP102, SSP32, and TSP106 show mostly blue blocks, indicating limited and unique recombination, while NZ_CPO20549.1 indicated as a reference genome in this recombinant analysis. The bottom plot shows SNP density along the genome, highlighting hotspots where recombination is more frequent. Overall, the figure demonstrates lineage-specific recombination patterns and the higher recombination load in ST320 strains.

independently within clade II, while SSP32 (ST695-19A) showed a possibly close relationship with Danish strains having serotype 8, as seen in the respective clades. Whereas TSP106 (ST671-14) of Clade III, showed close genetic and geographic ties to Malaysian strains WSP7 and WSP8 (isolated in 2011). The geographical clustering of the TSP106 lineage in this study, as well as TSP28, TSP95, SSP45 and SSP46 with reference genomes from Malaysia (Fig. 1), suggested the phylogeography of *S. pneumoniae* that is likely structured by country despite the small number of the local isolates (Nzoyikorera et al. 2021).

Recombination analysis revealed both shared and isolate-specific recombination events across the seven pneumococcal isolates. Extensive red blocks in invasive isolates TSP28, TSP95, SSP45, and SSP46 represent recombinant regions; genomic segments where horizontal gene transfer or recombination occurred across multiple isolates. The intensity and density of red regions, especially in SSP45 and SSP46 suggest high recombination activity, which may affect virulence or antibiotic resistance. In contrast, blue blocks in invasive isolates like TSP102, TSP106, and SSP32 indicate isolate-specific recombination, meaning those events are unique to respective strains. These invasive isolates also exhibited fewer recombination events, suggesting more conserved genomic structures. The presence of blue blocks highlights isolate-specific recombination, supporting the idea of genetic diversification. The presence of isolate-specific recombination (blue blocks) alongside shared red regions suggests genetic diversification. These findings underscore the role of recombination in shaping pneumococcal population structure and adaptation, including potential impacts on serotype switching, virulence, and adaptation to selective pressures like vaccination or antibiotic resistance. It can be seen that the strains analysed in this study have undergone multiple recombination events. This is expected, as *S. pneumoniae* is naturally transformable and recombination events play a major role in its molecular evolution (Chaguza et al. 2015).

Streptococcus pneumoniae is a highly recombinogenic bacterium generating genetic variation at a rate far exceeding that of random mutation, enabling pneumococci to evade both host immune responses and clinical interventions, such as vaccines and antibiotics (Chaguza et al. 2015). Our findings indicate that TSP28, TPS95, SSP45, and SSP46 share a unique recombination event involving specific related genes, including sucrose glucosyltransferase or glucosyltransferase-A, hypothetical protein (HP), triose phosphate isomerase, and others. Additionally, triose phosphate isomerase *(TpiA)* is involved in a unique recombination event observed among most invasive pneumococcal isolates; TSP28, TSP95, TSP106, SSP45, and SSP46. In *S. pneumoniae*, *TpiA* is secreted extracellularly through LytA-dependent autolysis, though it also functions as an internal enzyme within the glycolytic pathway. A study mentioned that *TpiA* binds to host plasminogen, promoting its activation and facilitating the conversion of plasminogen to plasmin, which may enhance pathogen entry into host tissues (Hirayama et al. 2022). Proteomic analysis identified *TpiA* as a novel virulence factor, offering insights into its role *in vivo* (Hirayama et al. 2022). The recruitment of plasminogen is a critical mechanism used by *S. pneumoniae* for host tissue invasion during infection. *TpiA* is released from *S. pneumoniae* through autolysis, attaches to plasminogen, and promotes its activation, thereby

Table 4 Unique recombinant events among the seven pneumococcal isolates.

No	CDS	Locus_tag	Protein_id	Gene name	Product
			Unique recombination with low frequency in TSP95		
1	283 759..285 018	SPNHU17_00 336	ARD33952.1	rseP	RIP metalloprotease RseP
2	311 177..312 166	SPNHU17_00 355	ARD33972.1	Complement	Mannose-specific phosphotransferase system
3	312 399..313 418	SPNHU17_00 357	ARD33973.1	Complement	Alcohol dehydrogenase, propanol-preferring
4	334 039..337 242	SPNHU17_00 385	ARD33999.1	-	Hyaluronate lyase
5	1 421 187..1 421 245	SPNHU17_01 532	ARD35095.1	-	Phosphoglucomutase/phosphomannomutase alpha/beta/alpha domain I
6	1 428 410	SPNHU17_01 538	ARD35101.1	Complement	Transport protein
7	1 851 376..1 855 053	SPNHU17_01 994	ARD35536.1	rpoC	DNA-directed RNA polymerase, beta' subunit
	1 855 086..1 858 697	SPNHU17_01 995	ARD35537.1	rpoB	DNA-directed RNA polymerase, beta subunit
8	1 152 346–1 158 508	SPNHU17_01 232	ARD34815.1	Complement	PTS system, lactose-specific IIC component
			Unique recombination with high frequency in TSP95		
9	1 421 519	SPNHU17_01 532	ARD35095.1	-	Phosphoglucomutase/phosphomannomutase alpha/beta/alpha domain I
10	1 428 410	SPNHU17_01 538	ARD35101.1	Complement	Transport protein
			Unique recombination with high frequency in TSP106		
11	1 471 420..1 474 116	SPNHU17_01 591	ARD35150.1	Complement	Calcium-transporting ATPase 1 (Golgi Ca(2+)-ATPase)
			Shared recombination with high frequency in TSP106, TSP28, TSP95, SSP45, SSP46		
12	1 492 188–1 497 567	SPNHU17_01 616	ARD35175.1	-	
	1 492 191..1 493 663	SPNHU17_01 611	ARD35170.1	-	Hypothetical protein
	1 493 734..1 494 492	SPNHU17_01 612	ARD35171.1	tpiA	Triose-phosphate isomerase
	1 494 590..1 495 267	SPNHU17_01 613	ARD35172.1	-	DNA replication protein DnaD

Table 4 Continued

No	CDS	Locus_tag	Protein_id	Gene name	Product
	1 495 276..1 496 220	SPNHU17_01 614	ARD35173.1	metA	Homoserine O-succinyltransferase
13	314 706..316 178	SPNHU17_00 361	ARD33977.1	-	Xanthine/uracil permease family protein
14	318 165..319 487	SPNHU17_00 366	ARD33981.1	-	Dihydrofolate synthetase
15	342 250..343 440	SPNHU17_00 392	ARD34006.1	-	Unsaturated glucuronyl hydrolase
16	358 280..359 764	SPNHU17_00 409	ARD34023.1	-	Hypothetical protein
		Unique recombination with high frequency in SSP32			
17	1 795 181..1 796 623	SPNHU17_01 933	ARD35477.1	-	Sucrose phosphorylase (Sucrose glucosyltransferase) (Glucosyltransferase-A) (GTF-A)
18	1 800 012..1 802 174	SPNHU17_01 937	ARD35481.1	-	Alpha-galactosidase AgaN
		Unique recombination with high frequency in TSP95, SSP45, SSP46			
19	1 795 189–1 796 836	SPNHU17_01 933	ARD35477.1	-	sucrose phosphorylase (Sucrose glucosyltransferase) (Glucosyltransferase-A) (GTF-A)
		Shared recombination in all isolates			
20	1 447 681..1 448 775	SPNHU17_01 561	ARD35120.1	-	Cystathionine gamma-synthase (CGS) (O-succinylhomoserine(thiol)-lyase)
21	1 454 669..1 456 114	SPNHU17_01 567	ARD35126.1	-	UDP-N-acetylmuramoylalanyl-D-glutamate—2, 6-diaminopimelate ligase
22	1 450 779..1 452 737	SPNHU17_01 565	ARD35124.1	-	Oligopeptide binding lipoprotein

aiding tissue invasion by degrading the extracellular matrix (Hirayama et al. 2022).

In addition, the phylogenetic and recombination analyses collectively revealed consistent clustering patterns among the invasive *S. pneumoniae* isolates. Notably, the invasive isolates TSP28, TSP95, SSP45, and SSP46 consistently clustered together in both analyses (Figs 1 and 2). These isolates predominantly belonged to ST320 and expressed serotype 19A, which is widely associated with increased virulence, antimicrobial resistance, and post-vaccine serotype replacement globally. They also exhibited similar genomic recombination profiles, suggesting a shared evolutionary origin with strong clonal expansion. The strong concordance between phylogenetic clustering and recombination patterns highlights the importance of ongoing genomic surveillance to monitor the emergence and spread of virulent pneumococcal lineages in Malaysia and the wider Southeast Asian region through this model of analysis.

Conclusion

This study presents a whole genome analysis of invasive pneumococcal isolates at selected major tertiary hospitals in Malaysia. We found that the invasive pneumococcal isolates were associated with clonal spread involving different serotypes such as 19F and 19A, which are mostly covered by PCVs. The predominant ST236 and ST320 of the invasive isolates were linked to clones Taiwan^{19F}-14 and also associated with a high frequency of multidrug resistance (MDR). This study highlights the genomic diversity and evolutionary dynamics of invasive *S. pneumoniae* isolates in Malaysia. The integration of serotype, ST, antibiotic resistance determinant genes, and recombination analyses reveals both clonal persistency and diversification, with evidence of close relatedness to global strains. Since our study was limited to only seven isolates, future research should include larger sample sizes and longitudinal surveillance to detect vaccine escape and emerging variants. Expanding investigations across diverse regions and time points will provide a more comprehensive understanding of pathogen evolution and vaccine effectiveness.

Acknowledgements

We would like to thank the microbiology laboratory staff at the Department of Pathology, Sultanah Nur Zahirah Hospital and Sungai Buloh Hospital, for their assistance. We would also like to thank the Director General of Health Malaysia for his permission to publish this article.

Author contributions

Nurul Diana Dzaraly (Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft), AbdulRahman Muthanna (Conceptualization, Formal Analysis, Methodology, Writing – review & editing), James John (Validation, Writing – review & editing), Siti Norbaya Masri (Methodology, Supervision, Validation, Writing – review & editing), Zarizal Suhaili (Methodology, Validation, Writing – review & editing), Nurshahira Sulaiman (Methodology, Validation, Writing – review & editing), Nor Iza A. Rahman (Supervision, Validation, Writing – review & editing), Tuan Suhaila Tuan Soh (Supervision, Validation, Writing – review & editing), Fatimah Haslina Abdullah (Validation, Writing – review & editing), Sangita Biswas (Validation, Writing – review & editing), Mazen M Jamil Al-Obaidi (Validation, Writing – review & editing), Mohd Nasir Mohd Desa (Conceptualization, Data curation, Formal Analysis, Methodology, Supervision, Validation, Writing – review & editing)

Supplementary material

Supplementary material is available at *Journal of Applied Microbiology* online.

Supplementary 1: List of isolates and reference strains (obtained from NCBI GenBank Database) used in core genome analysis.

Conflicts of interest

None declared.

Funding

This research was supported by Universiti Putra Malaysia through GP-IPS (9812000) and Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS/1/2020/SKK0/UPM/02/10).

Data availability

Data underlying this article are available in the article and its supplementary file. Sequence data are available in the NCBI genomes database as listed in Table 1.

References

Akdoğan Kittana FN, Mustak IB, Hascelik G *et al.* Erythromycin-resistant *Streptococcus pneumoniae*: phenotypes, genotypes, transposons and pneumococcal vaccine coverage rates. *J Med Microbiol* 2019;**68**:874–81. https://doi.org/10.1099/jmm.0.00099

Alcock BP, Raphenya AR, Lau TTY *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;**48**:D517–25. https://doi.org/10.1093/nar/gkz935

Alicino C, Paganino C, Orsi A *et al.* The impact of 10-valent and 13-valent pneumococcal conjugate vaccines on hospitalization for pneumonia in children: a systematic review and meta-analysis. *Vaccine* 2017;**35**:5776–85. https://doi.org/10.1016/j.vaccine.2017.09.005

Almeida SC, Lo SW, Hawkins PA *et al.* Genomic surveillance of invasive *Streptococcus pneumoniae* isolates in the period pre-PCV10 and post-PCV10 introduction in Brazil. *Microb Genom* 2021;**7**:000635. https://doi.org/10.1099/mgen.0.000635

Ansaldi F, Canepa P, De Florentiis D *et al.* Increasing incidence of *Streptococcus pneumoniae* serotype 19A and emergence of two vaccine escape recombinant ST695 strains in Liguria, Italy, 7 years after implementation of the 7-valent conjugated vaccine.

Clin Vaccine Immunol 2011;**18**:343–5. https://doi.org/10.1128/CVI.00383-10

Arushothy R, Ahmad N, Amran F *et al.* Pneumococcal serotype distribution and antibiotic susceptibility in Malaysia: a four-year study (2014–2017) on invasive paediatric isolates. *Int J Infect Dis* 2019;**80**:129–33. https://doi.org/10.1016/j.ijid.2018.12.009

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77. https://doi.org/10.1089/cmb.2012.0021

Ben Zakour NL, Venturini C, Beatson SA *et al.* Analysis of a *Streptococcus pyogenes* puerperal sepsis cluster by use of whole-genome sequencing. *J Clin Microbiol* 2012;**50**:2224–8. https://doi.org/10.1128/JCM.00675-12

Bortolaia V, Kaas RS, Ruppe E *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;**75**:3491–500. https://doi.org/10.1093/jac/dkaa345

Chaguza C, Cornick JE, Everett DB. Mechanisms and impact of genetic recombination in the evolution of *Streptococcus pneumoniae*. *Comput Struct Biotechnol J* 2015;**13**:241–7. https://doi.org/10.1016/j.csbj.2015.03.007

Chavanet P. Pneumococcus infections: is the burden still as heavy? *Médecine et Maladies Infectieuses* 2012;**42**:149–53. https://doi.org/10.1016/j.medmal.2012.02.002

Dallas SD, McGee L, Limbago B *et al.* Development of doxycycline MIC and disk diffusion interpretive breakpoints and revision of tetracycline breakpoints for *Streptococcus pneumoniae*. *J Clin Microbiol* 2013;**51**:1798–802. https://doi.org/10.1128/JCM.00125-13

Department of Statistics Malaysia. Statistics on Causes of Death, Malaysia, 2020. 2020. Available from: https://www.dosm.gov.my/portal-main/release-content/statistics-on-causes-of-death-malaysia-2020. (8 January 2025, date last accessed).

Diez-Domingo J, Gurtman A, Bernaola E *et al.* Evaluation of 13-valent pneumococcal conjugate vaccine and concomitant meningococcal group C conjugate vaccine in healthy infants and toddlers in Spain. *Vaccine* 2013;**31**:5486–94. https://doi.org/10.1016/j.vaccine.2013.06.049

Doherty N, Trzcinski K, Pickerill P *et al.* Genetic diversity of the *tet(M)* gene in tetracycline-resistant clonal lineages of *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 2000;**44**:2979–84. https://doi.org/10.1128/AAC.44.11.2979-2984.2000

Donkor ES, Stabler RA, Hinds J *et al.* Comparative phylogenomics of *Streptococcus pneumoniae* isolated from invasive disease and nasopharyngeal carriage from West Africans. *BMC Genomics* 2012;**13**:569. https://doi.org/10.1186/1471-2164-13-569

Dzaraly ND, Desa MNM, Muthanna AR *et al.* Molecular epidemiology of piliated pneumococcal isolates at a major tertiary hospital in the Klang Valley, Malaysia. *Asian Pac J Trop Med* 2023;**16**:80–8. https://doi.org/10.4103/1995-7645.370150

Dzaraly ND, Mohd Desa MN, Muthanna AR *et al.* Antimicrobial susceptibility, serotype distribution, virulence profile and molecular typing of piliated clinical isolates of pneumococci from east coast, Peninsular Malaysia. *Sci Rep* 2021;**11**:8220. https://doi.org/10.1038/s41598-021-87428-z

Dzaraly ND, Muthanna AR, Mohd Desa MN *et al.* Pilus islets and the clonal spread of piliated *Streptococcus pneumoniae*: a review. *Int J Med Microbiol* 2020;**310**:151449. https://doi.org/10.1016/j.ijmm.2020.151449

Enright MC, Spratt BG. A multilocus sequence typing scheme for. *Microbiology* 1998;**144**:3049–60. https://doi.org/10.1099/00221287-144-11-3049

Epping L, van Tonder AJ, Gladstone RA *et al.* SeroBA: rapid high-throughput serotyping of *Streptococcus pneumoniae* from whole genome sequence data. *Microb Genom* 2018;**4**:1–6. https://doi.org/10.1099/mgen.0.000186

Ganaie F, Saad JS, Mcgee L *et al.* A new pneumococcal capsule type, 10D, is the 100th serotype and Has a large *cps* fragment from an oral *Streptococcus*. *mBio* 2020;**11**:e00937–20. https://doi.org/10.1128/mBio.00937-20

Geno KA, Gilbert GL, Song JY *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

Gunasegaran H, Ramzi NH, Hoong Tan AC *et al.* Prevalence of pneumococcal carriage and risk factors for pneumonia and carriage among under-5 children in Malaysia: findings from the MY-Pneumo study. *Pneumonia* 2025;**17**:24. https://doi.org/10.1186/s41479-025-00177-9

Hirayama S, Domon H, Hiyoshi T *et al.* Triosephosphate isomerase of *Streptococcus pneumoniae* is released extracellularly by autolysis and binds to host plasminogen to promote its activation. *FEBS Open Bio* 2022;**12**:1206–19. https://doi.org/10.1002/2211-5463.13396

Hsieh YC, Lin TL, Chang KY *et al.* Expansion and evolution of *Streptococcus pneumoniae* serotype 19A ST320 clone as compared to its ancestral clone, Taiwan19F-14 (ST236). *J Infect Dis* 2013;**208**:203–10. https://doi.org/10.1093/infdis/jit145

Jain C, Rodriguez-R LM, Phillippy AM *et al.* High throughput ANI analysis of 90 K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;**9**:1–8. https://doi.org/10.1038/s41467-018-07641-9

Jindal HM, Ramanathan B, Le CF *et al.* Comparative genomic analysis of ten clinical *Streptococcus pneumoniae* collected from a Malaysian hospital reveal 31 new unique drug-resistant SNPs using whole genome sequencing. *J Biomed Sci* 2018;**25**:15. https://doi.org/10.1186/s12929-018-0414-8

Kireeva A, Dmitriev A. Identification of novel mobile genetic elements associated with resistance to macrolide and lincosamide in *Streptococcus dysgalactiae* subsp. *equisimilis*. *TOMICROJ* 2023;**17**:1–7. https://doi.org/10.2174/18742858-v17-e230109-2022-16

Larsen MV, Cosentino S, Rasmussen S *et al.* Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012;**50**:1355–61. https://doi.org/10.1128/JCM.06094-11

Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;**47**:256–9. https://doi.org/10.1093/nar/gkz239

Lister AJJ, Dombay E, Cleary DW *et al.* A brief history of and future prospects for pneumococcal vaccination in Malaysia. *Pneumonia* 2023;**15**:12. https://doi.org/10.1186/s41479-023-00114-8

Lo SW, Gladstone RA, Van Tonder AJ *et al.* A mosaic tetracycline resistance gene *tet(S/M)* detected in an MDR pneumococcal CC230 lineage that underwent capsular switching in South Africa. *J Antimicrob Chemother* 2020;**74**:512–20. https://doi.org/10.1093/jac/dkz475

Minas K, Mcewan NR, Newbold CJ *et al.* Optimization of a high-throughput CTAB-based protocol for the extraction of qPCR-grade DNA from rumen fluid, plant and bacterial pure cultures. *FEMS Microbiol Lett* 2011;**325**:162–9. https://doi.org/10.1111/j.1574-6968.2011.02424.x

Mott MP, Caierão J, Cunha GR *et al.* Emergence of serotype 19A *Streptococcus pneumoniae* after PCV10 associated with a ST320 in adult population, in Porto Alegre, Brazil. *Epidemiol Infect* 2019;**147**:e93. https://doi.org/10.1017/S0950268819000013

Nathan JJ, Desa MNM, Thong KL *et al.* Genotypic characterization of *Streptococcus pneumoniae* serotype 19F in Malaysia. *Infect Genet Evol* 2014;**21**:391–4. https://doi.org/10.1016/j.meegid.2013.11.026

Nzoyikorera N, Diawara I, Fresia P *et al.* Whole genomic comparative analysis of *Streptococcus pneumoniae* serotype 1 isolates causing invasive and non-invasive infections among children under 5 years in Casablanca, Morocco. *BMC Genomics* 2021;**22**:1–9. https://doi.org/10.1186/s12864-020-07316-0

Ooi JM, Eg KP, Chinna K *et al.* Predictive risk factors for complicated pneumonia in Malaysian children. *J Paediatrics Child Health* 2019;**55**:406–10. https://doi.org/10.1111/jpc.14213

Page AJ, Cummins CA, Hunt M *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;**31**:3691–3. https://doi.org/10.1093/bioinformatics/btv421

Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**:e9490. https://doi.org/10.1371/journal.pone.0009490

Rahman NAA, Muthanna A, Desa MNM *et al.* Comparative genotyping of Malaysian clinical isolates of *Streptococcus pneumoniae* by multilocus sequence typing and multilocus variable-number tandem repeat analysis. *Malays J Med Sci* 2025;**32**:69. https://doi.org/10.21315/mjms-09-2024-677

Rodriguez-Ruiz JP, Xavier BB, Stöhr W *et al.* High-resolution genomics identifies pneumococcal diversity and persistence of vaccine types in children with community-acquired pneumonia in the UK and Ireland. *BMC Microbiol* 2024;**24**:1–15. https://doi.org/10.1186/s12866-024-03300-w

Song JH, Chang HH, Suh JY *et al.* Macrolide resistance and genotypic characterization of *Streptococcus pneumoniae* in Asian countries: a study of the Asian Network for Surveillance of Resistant Pathogens (ANSORP). *J Antimicrob Chemother* 2004;**53**:457–63. https://doi.org/10.1093/jac/dkh118

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3. https://doi.org/10.1093/bioinformatics/btu033

Wahl B, O'Brien KL, Greenbaum A *et al.* Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob Health* 2018;**6**:e744–57. https://doi.org/10.1016/S2214-109×30247-X

Zeng Y, Song Y, Cui L *et al.* Phylogenomic insights into evolutionary trajectories of multidrug resistant *S. pneumoniae* CC271 over a period of 14 years in China. *Genome Med* 2023;**15**:1–13. https://doi.org/10.1186/s13073-023-01200-8

Zhou M, Wang L, Wang Z *et al.* Molecular characterization of penicillin-binding protein2x, 2b and 1a of *Streptococcus pneumoniae* causing invasive pneumococcal diseases in China: a multicenter study. *Front Microbiol* 2022;**13**:838790. https://doi.org/10.3389/fmicb.2022.838790

Received: 20 October 2025. **Revised:** 15 January 2026. **Accepted:** 25 January 2026