International Journal of Infectious Diseases 160 (2025) 108042

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

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

Potential application of MALDI-TOF MS to identify *Streptococcus parapneumoniae*, an emerging pathogen previously misidentified as *Streptococcus pneumoniae*

Yan Zhao[1,†], Yuanchao Ma[1,2,†], Hwei Huih Lee[1,2], Wing Ho Li[1], Antonio Hung-Yee Ngan[3], Jonathan Hon-Kwan Chen[3], Simon Yung-Chun So[3], Vincent Chi-Chung Cheng[3], Kwok-Yung Yuen[2,4,5,6], Man Lung Yeung[2,4,5,6,†], Jade Lee-Lee Teng[1,*]

[1] *Faculty of Dentistry, The University of Hong Kong, Hong Kong Special Administrative Region, China*
[2] *Department of Microbiology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China*
[3] *Department of Microbiology, Queen Mary Hospital, Pokfulam, Hong Kong Special Administrative Region, China*
[4] *Centre for Virology, Vaccinology and Therapeutics, Hong Kong Science and Technology Park, Hong Kong Special Administrative Region, China*
[5] *State Key Laboratory of Emerging Infectious Diseases, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China*
[6] *Carol Yu Centre for Infection, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China*

ARTICLE INFO

Article history:
Received 29 July 2025
Revised 29 August 2025
Accepted 30 August 2025

Keywords:
Streptococcus parapneumoniae
Streptococcus pneumoniae
Misidentification
MALDI-TOF MS
Multidrug-resistant

ABSTRACT

Objectives: Accurate identification of *Streptococcus* species is critical for clinical management and epidemiology. Misidentification of *Streptococcus parapneumoniae* as *Streptococcus pneumoniae* can hinder diagnosis and affect treatment outcomes.
Methods: From 385 archived *S. pneumoniae* isolates, species-specific polymerase chain reaction (PCR) was used to identify potential *S. parapneumoniae*. Confirmatory species determination, virulence, and antimicrobial resistance profile analyses were performed through whole-genome sequencing (WGS), phylogenomic, and comparative genomic analyses. Matrix-associated laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) spectral analysis aimed to identify biomarkers for *S. parapneumoniae*.
Results: Three *S. parapneumoniae* strains, representing a novel species first identified in Japan in 2024, were isolated from patients with respiratory infections in Hong Kong. WGS showed >99% average nucleotide identity (ANI) with *S. parapneumoniae* SP4011^T, distinct from *S. pneumoniae* (<94%). These strains possessed virulence factors similar to *S. pneumoniae*, suggesting pathogenic potential. All isolates exhibited multidrug and levofloxacin resistance, unlike local *S. pneumoniae* strains. MALDI-TOF MS identified two peaks (*m/z* 6,399 and *m/z* 6,960) unique to *S. parapneumoniae*.
Conclusions: The multidrug resistance of *S. parapneumoniae* complicates antimicrobial resistance surveillance data and empirical treatment accuracy for *S. pneumoniae*. The identified discriminatory peaks offer promising tools for accurate species identification. The prevalence of *S. parapneumoniae* is likely underestimated; expanded surveillance is warranted to determine its true distribution and clinical significance.

Introduction

The genus *Streptococcus* encompasses a diverse group of bacteria, with 244 recognized species https://lpsn.dsmz.de/search?word=streptococcus as of July 16, 2025; List of Prokaryotic names with Standing in Nomenclature [1]. Many of these species are significant human pathogens, contributing to substantial morbidity and mortality worldwide. They are found in almost every location in the human body and are the dominant species in the human oral cavity and upper respiratory tract [2]. Traditionally, streptococci have been classified based on their hemolytic properties on blood agar, categorized as α-hemolytic, β-hemolytic, or non-

* Corresponding author.
E-mail address: llteng@hku.hk (J.L.-L. Teng).
† Yan Zhao, Yuanchao Ma and Man Lung Yeung contributed equally to this study.

https://doi.org/10.1016/j.ijid.2025.108042

hemolytic depending on the type of hemolysis that the bacterium generates on blood agar. The β-hemolytic streptococci were further classified using Lancefield grouping, although it should be noted that certain α-hemolytic and non-hemolytic streptococci also reacted with specific Lancefield antisera. Historically, the oral streptococci were referred to as viridans streptococci (*Streptococcus viridans*) due to their high propensity to display partial hemolysis when cultured on blood agar plates, resulting in a green coloration surrounding the colonies. Although the viridans streptococci or α-hemolytic designations are still used today to classify the oral streptococci, this term is not fully accurate, as some isolates can display β-hemolysis or completely lack any hemolytic activity.

To overcome the limitations of traditional methods, newer methods have been developed for the classification and identification of *Streptococcus* species. Amplification and sequencing of universal gene targets, such as the 16S rRNA gene, have been extensively studied [3,4]. However, some studies have raised concerns regarding the resolution and taxonomic grouping capabilities of the 16S rRNA gene in certain circumstances [5,6]. Recently, a more robust genome-based phylogenetic approach has been used to categorize streptococci into nine distinct groups*:* bovis, gordonii, mitis, mutans, pluranimalium, pyogenic, salivarius, sorbrinus, and suis [7]. The mitis group, the largest among those found in the oral cavity with 20 species, presents challenges in species differentiation, leading to the reclassification of strains and the emergence of new species descriptions [8]. Another promising and newer approach for streptococci identification is matrix-associated laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). This method analyzes the protein mass spectra of bacterial isolates, providing a unique fingerprint for each species. However, current databases used in MALDI-TOF MS still face limitations in distinguishing *Streptococcus pneumoniae* from other closely related species of the *Streptococcus mitis* group [9]. This differentiation is crucial as *S. pneumoniae* is often highly pathogenic, and early identification can facilitate prompt antimicrobial therapy, potentially improving clinical outcomes. *Streptococcus parapneumoniae*, a newly discovered species in Japan in 2024, serves as an example of misidentification, initially mistaken as *S. pneumoniae* in a case of acute pyelonephritis with bacteremia [10]. This species shares morphological and biochemical similarities with *S. pneumoniae*, posing challenges for accurate diagnosis. We hypothesized that some *S. parapneumoniae* were previously misidentified as *S. pneumoniae*. Therefore, we conducted targeted screenings for *S. parapneumoniae* using polymerase chain reaction (PCR) with specific primers designed for this species in 385 archived isolates previously reported as *S. pneumoniae*. Subsequently, we present the discovery and detailed characterization of three additional clinical isolates of *S. parapneumoniae* in Hong Kong, originally misidentified as *S. pneumoniae*. Through comprehensive genome sequencing analysis, we confirmed the true identities of these isolates as *S. parapneumoniae*. Furthermore, we propose the utilization of a MALDI-TOF-based method for precise differentiation between *S. parapneumoniae* and *S. pneumoniae*.

Methods

Patient and strains

Clinical specimens were collected and processed following standard protocols outlined in a previous study [11]. The specimens were cultured on sheep blood agar at 37°C with 5% CO_2 to obtain all three isolates. Patients' clinical data were obtained by retrieving and analyzing clinical records provided by the hospital.

The type strains of *S. pneumoniae* NCTC 7465^T and *S. mitis* NCTC 12261^T were sourced from the Biological Resource Center of Institut Pasteur, France. *S. pseudopneumoniae* PW931 was a clinical isolate.

Microbiological methods and phenotypic characterizations

Standard protocols outlined in a previous study were followed for bacterial cultures and phenotypic identification [12]. The API system (20 STREP) (BioMérieux, France) was utilized to identify the bacterial isolate in this study. The direct transfer method was employed to conduct MALDI-TOF MS, with modifications, using the Microflex LT system with MALDI Biotyper 4.1.80 with the MBT IVD Library DB Revision G (Bruker Daltonik) [13]. Antibiotic susceptibility testing was carried out using the E-test method for penicillin and the Kirby-Bauer disk diffusion method for the other antibiotics. The results were interpreted based on the Clinical and Laboratory Standards Institute's guidelines updated in 2024 [14].

16S. rRNA gene sequencing, sequence identity analyses, and phylogenetic analyses

Bacterial DNA extraction, PCR, and sequencing of the 16S rRNA gene for the three case isolates of *S. parapneumoniae* were conducted following the methods described in a previous publication [15]. The primer pair 5′-GAGTTGCGAACGGGTGAG-3′ and 5′-CTTGTTACGACTTCACCC-3′ were used for the 16S rRNA gene (1,534 bp) PCR and DNA sequencing. To construct a 16S rRNA phylogenetic tree, the obtained DNA sequences, along with those of closely related species available in the mitis group, were subjected to pairwise alignment using MEGA 11 (version 11.0.11). The substitution models were tested, and a phylogenetic tree was constructed using the maximum likelihood method in MEGA 11 (version 11.0.11) [16]. The phylogenetic analyses included 883 nucleotide positions of the 16S rRNA sequences.

Screening of S. parapneumoniae *in bacterial isolates initially identified as* S. pneumoniae

PCR screening method was employed to detect the presence of *S. parapneumoniae* in archived bacterial isolates previously identified as *S. pneumoniae*. A total of 385 archived bacterial isolates, originally identified as *S. pneumoniae*, were subjected to PCR screening using primers 5′-AGGTGTTTTAAATAATGCGTC-3′ and 5′-AATATTATCAAACTCTGGAC-3′ targeting a unique 560-bp region of a hypothetical protein gene specific to *S. parapneumoniae* but not found in other closely related *Streptococcus* species.

Genome sequencing and hybrid genome assembly

The genomes of the three case isolates were determined using Illumina paired-end short read sequencing. For PW5782, an additional Oxford Nanopore long-read sequencing (ONT) was performed. Genomic DNA extraction and the Illumina DNA library preparation were performed as described previously [15]. The Illumina DNA library was sequenced on a NovaSeq 6000 instrument (run type: PE151 bp). The ONT long-read library was prepared using SQK-RAD004 rapid sequencing kit (Oxford Nanopore Technologies, Oxford, UK) following the manufacturer's instructions and sequenced on a MinION sequencer. Unicycler v.0.4.8 was employed to assemble the Illumina and Oxford Nanopore MinION reads to obtain the draft genomes.

Genome sequence analyses

Intergenomic distances, represented by average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH), were calculated between the proposed three isolates and the previously pub-

lished *S. parapneumoniae* SP4011^T and reference strains of the closest related species. ANI and dDDH were performed using the web service available at https://www.ezbiocloud.net/tools/ani [17] and Genome-to-Genome Distance Calculator v3.0 (http://ggdc.dsmz.de/distcalc2.php), respectively. In addition to the three isolates, which were sequenced completely as part of this study, the remaining complete genome sequences of six *Streptococcus* species were downloaded from the National Center for Biotechnology Information (NCBI) databases (Supplementary Table S2). Genome characteristics, including the number of tRNA genes, rRNA genes, and coding sequences, were predicted using Prokaryotic Genome Annotation Pipeline (PGAP) provided by the NCBI [18]. The tRNA gene counts were further validated with ARAGORN 1.2.41 [19]. The G+C content was calculated directly from the genome sequence.

Phylogenomic characterization

To determine the phylogenetic position of all four isolates of *S. parapneumoniae* among the current nine taxonomic groups within the genus *Streptococcus* [20], a multigene-based phylogenomic approach was employed. This approach involved the concatenation of nucleotide sequences from 92 bacterial core genes (Supplementary Table S4). The alignment of the concatenated 92 core genes from 65 *Streptococcus* genomes and one *Lactococcus* genome was generated using the up-to-date bacterial core gene (UBCG) pipeline (https://www.ezbiocloud.net/tools/ubcg) with default parameters, as described by Na *et al.* [21]. The Neighbor-joining tree was constructed using MEGA 11 (version 11.0.11) [16].

MALDI-TOF MS analysis

Bacterial isolates were processed according to the manufacturer's recommended direct transfer and ethanol–formic acid extraction protocols for MALDI-TOF MS [22]. A single isolated colony was inoculated directly onto the Main Spectrum Profile (MSP96) target plate spot by the direct transfer method. The target plate was analyzed by the Bruker Microflex LT system. The protein profile of each spot with *m/z* values of 2,000–20,000 generated was analyzed by the MALDI Biotyper V.3.1 using the latest MBT IVD Library (10,964 spectra for bacteria, 416 spectra for *Streptococcus*). Only identification score >2.0 indicated a reliable species-level identification.

Three *S. parapneumoniae* isolates, the *S. pneumoniae* type strain NCTC 7465^T, and 20 clinical S. pneumoniae isolates (with species identities confirmed by genome sequencing) were subjected to spectral acquisition. The spectra were acquired using a Bruker Microflex LT, equipped with a 60 Hz nitrogen laser operating in positive linear mode over a mass range of 1,960-22,000 Daltons, with a FlexControl^{TM} 3.4 software. The raw spectral data were automatically processed with MBT Compass Explorer software (version 4.1.80), which included smoothing, baseline subtraction, and intensity normalization to improve spectral quality. Spectra were then matched against reference spectra stored in the database; the library included a total of 5,627 *Streptococcus* reference spectra, with existing MSPs for *S. pneumoniae*. To further improve spectral quality, FlexAnalysis was used to remove flat-line spectra and anomalous peaks [23].

Virulence factors identification

Protein coding regions of four isolates of *S. parapneumoniae* were predicted, and automatic functional annotations were performed using Prokka 1.14.0 [24]. Virulence genes were identified by comparing the predicted sequences against the Virulence Factor Database (VFDB) [25]. Virulence factors of five *S. pneumoniae* strains were retrieved directly from the VFDB. Comparative analysis of virulence factors between *S. parapneumoniae* and *S. pneumoniae* was conducted based on functional annotations and sequence homology of the protein sequences. Only virulence factors present in all four *S. parapneumoniae* isolates and the five *S. pneumoniae* strains were included in further analyses. The protein sequences of virulence factors from *S. parapneumoniae* SP4011^T were analyzed using tBLASTn against the genomes of the five *S. pneumoniae* strains.

Results

Screening of S. parapneumoniae *in archived bacterial isolates identified as* S. pneumoniae

A total of 385 bacterial isolates initially identified as *S. pneumoniae* and collected between 2016 and 2024 were screened for *S. parapneumoniae* using PCR with specific primers. Among these isolates, three samples (0.78%) tested positive for *S. parapneumoniae*. Sequence analysis of the positive samples revealed that they were 100% identical and showed 100% nucleotide identity to a hypothetical protein gene sequence of *S. parapneumoniae* SP4011^T.

Patient information

The three putative *S. parapneumoniae* isolates recovered from a 66-year-old man, an 86-year-old woman, and a 63-year-old woman, who all came to Hong Kong public hospitals for treatment with respiratory infection. The first two cases were in 2016, and the third case was in 2023. Following 24 hours of incubation, a Gram-positive, aerobic, non-sporulating coccus was isolated from the sputum of the above three patients, referred to as PW5782, PW7223, and PW7224. Upon further review of the clinical history, there were no indications of nosocomial infections. All three patients developed community-acquired pneumonia prior to hospital admission.

MALDI-TOF MS identification and phenotypic characterizations

MALDI-TOF MS identified the three putative *S. parapneumoniae* isolates (PW5782, PW7223, and PW7224) as *S. pneumoniae*, with a score of 2.17, 2.10, and 2.16, respectively. All three isolates grew on sheep blood agar as α-hemolytic and grey colonies of 0.5-1 mm in diameter after 24 hours of incubation at 37°C in an aerobic environment (Supplementary Figure S1). Growth enhancement was ob-

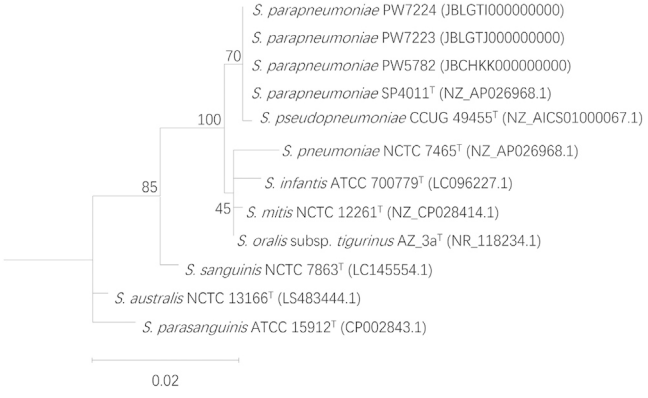
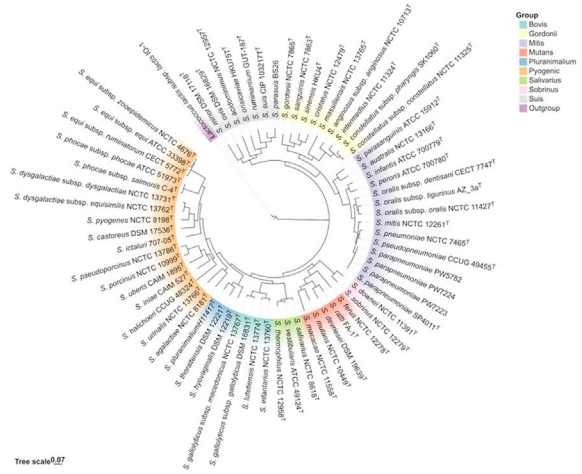
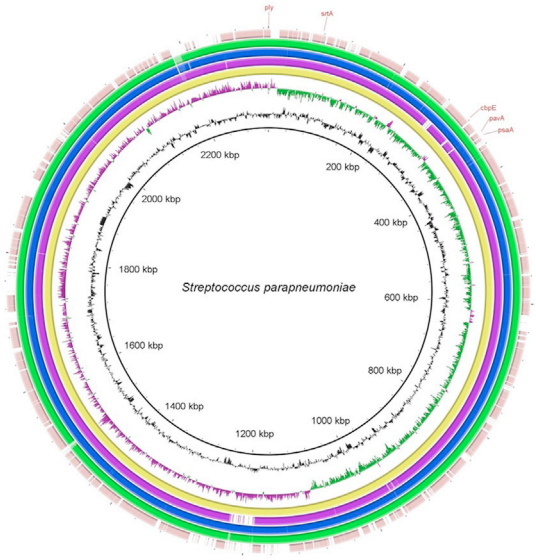


Figure 1. 16S rRNA gene-based phylogenetic tree of *Streptococcus parapneumoniae* and related *Streptococcus spp*. The tree was inferred from the 16S rRNA gene using the maximum-likelihood method with Kimura's two parameter correction and general time reversible models (K2+G+I), with *Lactococcus lactis* subsp. *lactis* IO-1 as the outgroup. The scale bar indicates the estimated number of substitutions per base. Numbers at nodes indicate levels of bootstrap support calculated from 1000 replicates. Names and nucleotide accession numbers are given as cited in GenBank.

A)



B)



C)

Class	Virulence factors	Gene symbol	Locus tag	Protein identities (%)
Adherence	Choline binding proteins	*cbpE*	AT689_RS09170	90-95
	Pneumococcal adherence and virulence factor A	*pavA*	AT689_RS08990	97
Exotoxin	Pneumolysin	*ply*	AT689_RS04370	99
Exoenzyme	Autolysin	*lytA*	AT689_RS04305	86
	Serine protease	*htrA*	Rv1223	98
	Sortase A	*srtA*	AT689_RS07830	98-99
Nutritional/metabolic factor	Pneumococcal surface antigen A	*psaA*	AT689_RS05905	97

[a] *S. pneumoniae* CGSP14 (CP001033.1); *S. pneumoniae* D39 (CP000410.2); *S. pneumoniae* Hungary19A-6 (CP000936.1); *S. pneumoniae* R6 (NC_003098.1); *S. pneumoniae* TIGR4 (CP155539.1).

Figure 2. Comparative genomic analyses of *Streptococcus parapneumoniae* and closely related species. (a) Core genome phylogenetic tree of *S. paraneumoniae* and related *Streptococcus spp*. The tree was inferred from concatenated nucleotide sequences of 92 core genes using the UBCG pipeline, with *Lactococcus lactis* designated as the outgroup. The gene names and accession numbers are given as cited in the Database of Clusters of Orthologous Genes (COGs), listed in Supplementary Tables S4 and S5. *Streptococcus* genomes were classified into nine taxonomic groups: bovis, gordonii, mitis, mutans, pluranimalium, pyogenic, salivarius, sobrinus, and suis. The color represents different taxonomic groups. The scale bar corresponds to the average number of nucleotide substitutions per site on each branch. (b) Genomic comparison of four *S. parapneumoniae* isolates and *S. pneumoniae* type strain. The innermost circle (yellow) represents the genome of *S. parapneumoniae* SP4011^T (yellow circle), used as the reference. The outer circles show the genomes of PW5782 (purple), PW7223 (blue), PW7224 (green), and *S. pneumoniae* NCTC7465^T (pink). Virulence genes shared by both species are annotated in the figure. (c) Virulence genes common to *S. parapneumoniae* and *S. pneumoniae*[a] in the annotated genome.
UBCG, up-to-date bacterial core gene.

served for all three case isolates under 5% CO_2 conditions at 37°C. No growth was observed in the presence of 6.5% NaCl. Furthermore, similar to the typical features of *S. pneumoniae*, the three isolates demonstrated sensitivity to optochin and were bile soluble. They were also resistant to bacitracin. Detailed biochemical profiles of the isolates and *S. pneumoniae* are presented in Supplementary Table S1. The antibiotic susceptibility test results showed that all three isolates were sensitive to chloramphenicol and vancomycin, but resistant to penicillin, erythromycin, co-trimoxazole, levofloxacin, and clindamycin.

Molecular characterizations

PCR amplification of the 16S rRNA gene of the three isolates (PW5782, PW7223, and PW7224) yielded DNA products with lengths of approximately 1,500 bp. The three isolates had the same 16S rRNA gene sequence and were identical to the type strain of *S. parapneumoniae* SP4011^T. The next closest species was *S. pseudopneumoniae* CCUG 49455^T with a 99.93% nucleotide identity. They shared 99.48% nucleotide identities with *S. pneumoniae* NCTC 7465^T. The 16S rRNA phylogenetic tree of *S. parapneumoniae* and other representative species in the mitis group is shown in Figure 1.

Comparative genomic characterizations

Genome sequencing was performed to confirm the species identities of the three isolates (PW5782, PW7223, and PW7224). Using both Illumina and Nanopore reads generated four contigs for PW5782, with a total genome size of 2,273,213 bp (N50 = 1,532,916 bp, 595× coverage) and an average G+C content of 39.73%. The *de novo* assembly using Illumina reads generated 31 and 38 contigs for PW7223 and PW7224 of 2,235,223 bp and 2,260,293 bp with 39.73% and 39.77% G+C content, respectively. The contigs of all three isolates were submitted to the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) for annotation, with results provided in Supplementary Table S2. The results of phylogenetic analysis using 92 core genes are summarized in Figure 2a. An *in-silico* genome-to-genome comparison showed that all three isolates were highly related to each other (ANI of

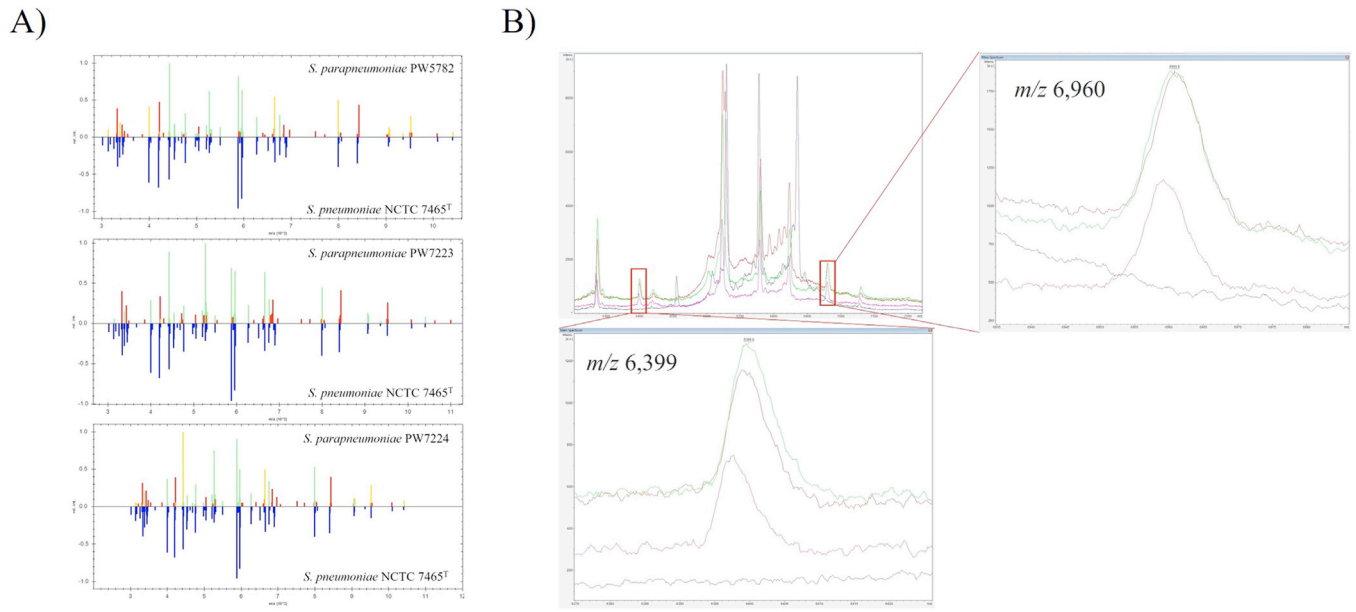


Figure 3. MALDI-TOF MS differentiation of *Streptococcus parapneumoniae* against *Streptococcus pneumoniae*. (a) Biotyper software analysis of *S. parapneumoniae* isolates (PW5782, PW7223, PW7224) compared with the reference MSP of *S. pneumoniae* NCTC 7465^T. Peak color coding indicates match quality: green (full match), yellow (partial match), and red (mismatch). Reference spectra (blue) are displayed in the lower panel. (b) Overlaid average spectra of *S. pneumoniae* NCTC 7465^T (dark blue) and *S. parapneumoniae* isolates (PW5782: purple, PW7223: red, PW7224: green) showing species-discriminative peaks. Red boxes highlight characteristic *S. parapneumoniae*-specific peaks at *m/z* 6,399 and *m/z* 6,960. Spectra were aligned and intensity-normalized using flexAnalysis software.
MALDI-TOF MS, matrix-associated laser desorption/ionization time-of-flight mass spectrometry; MSP, Main Spectrum Profile.

99.74-100%) and were also most closely related to the published *S. parapneumoniae* SP4011^T (ANI of 99.76-99.97%), falling within the species boundary threshold (ANI >95%). The next closest *Streptococcus* species was *S. pseudopneumoniae* (ANI of 93.99-94.19%). These results confirmed that all three isolates in this study were classified as members of *S. parapneumoniae* species. Further comparative genomic analyses using digital DNA-DNA hybridization (dDDH) with other closest species also supported the classification of the three isolates as typical of members of *S. parapneumoniae* (Supplementary Table S3).

To elucidate the taxonomic position of the *S. parapneumoniae* isolates among the genus *Streptococcus*, a multigene-based phylogenomic analysis was performed. The tree based on the concatenated nucleotide sequences of 92 bacterial core genes showed that the four isolates of *S. parapneumoniae*, including the type strain published previously, clustered with members of the mitis group. The mitis group currently includes some medically important pathogens such as *S. pneumoniae, S. pseudopneumoniae*, and *S. mitis*, forming a distinct and well-supported phylogenetic clade. The tree was also able to recover members of the remaining eight taxonomic groups, including bovis, gordonii, mutans, pluranimalium, pyogenic, salivarius, sorbrinus, and suis (Figure 2a), as described in previous studies [11].

Virulence genes shared between S. parapneumoniae *and* S. pneumoniae

S. parapneumoniae shares many of the same virulence factors as *S. pneumoniae*, enabling it to colonize the upper respiratory tract and progress to lower respiratory tract infection (i.e., pneumonia), and possibly to invasive bacteremia, as in the case of the isolate previously reported from Japan. The genomes of all three isolates identified in this study contains homologs of several virulence genes found in *S. pneumoniae*, including phosphorylcholine esterase CbpE (*cbpE*), Rqc2 family fibronectin-binding protein PavA (*pavA*), cholesterol-dependent cytolysin pneumolysin (*ply*), N-acetylmuramoyl-L-alanine amidase LytA (*LytA*), serine protease HtrA (*htrA*), sortase SrtA (*srtA*), and metal ABC transporter substrate-binding lipoprotein/adhesin PsaA (*psaA*); these genes are known to be involved in adhesion, exotoxin, exoenzyme, or nutritional/metabolic factor (Figure 2b and c) [26–28]. The protein identity values for the matches are presented in Figure 2c.

Mass spectra analysis

MALDI-TOF MS and conventional laboratory biochemical tests could not distinguish between *S. parapneumoniae* and *S. pneumoniae*; mass spectral analysis was performed to identify any spectral peaks that could distinguish between the two species. The MSP of *S. pneumoniae* NCTC 7465^T was used as a reference, and the three *S. parapneumoniae* isolates were compared with it one by one using MBT Compass (Bruker Daltonik GmbH, Bremen, Germany). The results showed that three *S. parapneumoniae* isolates had different matching degrees with MSP of *S. pneumoniae* (Figure 3a) and obtained log scores of 2.11, 2.22, and 2.29, respectively. Comparison of MSP of the three *S. parapneumoniae* isolates and 50 *S. pneumoniae* isolates (comprising 30 isolates from the Bruker database and 20 isolates with species identity confirmed by WGS) identified two discriminatory peaks of *m/z* 6,399 and *m/z* 6,960 (Table 1, Figure 3b). These two spectra peaks were present in the three *S. parapneumoniae* isolates and absent in all 50 *S. pneumoniae* isolates (Table 1, Figure 3b).

Discussion

The recent discovery of *S. parapneumoniae* as a novel species within the mitis group underscores the ongoing challenges in differentiating closely related streptococci, particularly those with overlapping phenotypic and biochemical profiles such as *S. pneumoniae*. In this study, we identified three clinical isolates of *S. parapneumoniae* in Hong Kong, shortly after the first report of this species (*S. parapneumoniae* SP4011^T) in Japan in 2024 [10]. These isolates, alongside the Japanese type strain, were initially misidentified as *S. pneumoniae* using conventional methods, highlighting

Table 1
Characteristic MALDI-TOF masses selected as potential biomarkers for distinguishing *Streptococcus parapneumoniae* from *Streptococcus pneumoniae*. The MSPs of 50 *S. pneumoniae* isolates were compared with the MSPs of the three *S. parapneumoniae* isolates identified in this study. Only *S. parapneumoniae* strains showed peaks at *m/z* 6,399 and *m/z* 6,960, whereas none the 50 *S. pneumoniae* isolates showed peaks at these two *m/z* values.

Isolates	Sources	*m/z* value [Da][a]										
		4,033	4,043	4,199	6,399	6,511	6,621	6,635	6,890	6,960	8,394	8,431
S. parapneumoniae PW5782	Clinical isolate	N	N	N	**Y**	N	N	Y	N	**Y**	N	Y
S. parapneumoniae PW7223	Clinical isolate	N	N	N	**Y**	N	N	Y	N	**Y**	Y	Y
S. parapneumoniae PW 7224	Clinical isolate	N	N	N	**Y**	N	N	Y	N	**Y**	N	Y
S. pneumoniae PW5776	Clinical isolate	N	N	N	**N**	N	N	N	N	**N**	N	N
S. pneumoniae PW5777	Clinical isolate	N	N	N	**N**	N	Y	N	N	**N**	N	N
S. pneumoniae PW5778	Clinical isolate	Y	N	N	**N**	N	Y	N	N	**N**	N	N
S. pneumoniae PW5779	Clinical isolate	N	N	N	**N**	N	N	N	N	**N**	N	N
S. pneumoniae PW5788	Clinical isolate	N	N	Y	**N**	N	N	N	N	**N**	N	N
S. pneumoniae PW5789	Clinical isolate	N	N	Y	**N**	N	N	N	N	**N**	N	N
S. pneumoniae PW5790	Clinical isolate	N	N	N	**N**	Y	N	N	N	**N**	N	Y
S. pneumoniae PW5791	Clinical isolate	N	N	N	**N**	Y	N	N	N	**N**	N	Y
S. pneumoniae PW5792	Clinical isolate	N	N	N	**N**	N	N	N	N	**N**	N	N
S. pneumoniae PW5793	Clinical isolate	N	N	N	**N**	N	N	N	N	**N**	N	N
S. pneumoniae PW5794	Clinical isolate	N	N	N	**N**	N	N	N	N	**N**	N	N
S. pneumoniae PW7242	Clinical isolate	N	N	N	**N**	N	N	N	Y	**N**	N	N
S. pneumoniae PW7243	Clinical isolate	N	Y	N	**N**	N	Y	N	N	**N**	N	N
S. pneumoniae PW7244	Clinical isolate	Y	N	N	**N**	Y	Y	N	N	**N**	N	Y
S. pneumoniae PW7245	Clinical isolate	N	N	Y	**N**	N	N	N	N	**N**	Y	N
S. pneumoniae PW7247	Clinical isolate	N	Y	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae PW7250	Clinical isolate	N	N	N	**N**	Y	N	N	N	**N**	N	Y
S. pneumoniae PW7252	Clinical isolate	N	N	Y	**N**	N	N	N	N	**N**	Y	N
S. pneumoniae PW7268	Clinical isolate	N	N	Y	**N**	N	N	N	N	**N**	Y	N
S. pneumoniae PW7269	Clinical isolate	N	N	Y	**N**	Y	N	N	N	**N**	N	N
S. pneumoniae ATCC 49619	Bruker database	Y	N	Y	**N**	Y	Y	N	Y	**N**	Y	N
S. pneumoniae bes St 29 THL	Bruker database	Y	N	N	**N**	N	N	N	N	**N**	Y	Y
S. pneumoniae DSM 11865	Bruker database	N	Y	Y	**N**	N	N	Y	Y	**N**	Y	N
S. pneumoniae DSM 11866	Bruker database	N	Y	N	**N**	N	Y	N	Y	**N**	N	N
S. pneumoniae DSM 11868	Bruker database	N	Y	Y	**N**	N	Y	N	Y	**N**	Y	N
S. pneumoniae DSM 14377	Bruker database	N	Y	N	**N**	Y	N	Y	N	**N**	N	N
S. pneumoniae DSM 14378	Bruker database	Y	N	Y	**N**	N	Y	N	N	**N**	Y	N
S. pneumoniae DSM 20566T	Bruker database	N	N	Y	**N**	Y	Y	Y	N	**N**	Y	N
S. pneumoniae NRZ 22582	Bruker database	N	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae NRZ 22589	Bruker database	Y	N	N	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae NRZ 22904	Bruker database	N	N	N	**N**	N	N	Y	Y	**N**	Y	N
S. pneumoniae NRZ 23033	Bruker database	Y	N	Y	**N**	N	Y	N	Y	**N**	Y	N
S. pneumoniae NRZ 23311	Bruker database	Y	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae NRZ 23321	Bruker database	N	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae NRZ 23346	Bruker database	N	Y	Y	**N**	N	Y	N	Y	**N**	Y	N
S. pneumoniae NRZ 28221	Bruker database	N	N	Y	**N**	N	N	Y	Y	**N**	Y	N
S. pneumoniae NRZ 29853	Bruker database	Y	N	Y	**N**	N	N	Y	Y	**N**	Y	N
S. pneumoniae NRZ 29861	Bruker database	Y	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae NRZ 30494	Bruker database	N	Y	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae NRZ 31870	Bruker database	Y	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae NRZ 49091	Bruker database	N	N	Y	**N**	N	N	Y	Y	**N**	Y	N
S. pneumoniae 15116113201 LMH	Bruker database	N	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae V17_2011100 MUZ	Bruker database	N	Y	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae V17_2011122 MUZ	Bruker database	Y	N	Y	**N**	Y	N	N	N	**N**	Y	N
S. pneumoniae V17_201192 MUZ	Bruker database	N	N	N	**N**	N	N	N	Y	**N**	N	N
S. pneumoniae V17_201193 MUZ	Bruker database	N	Y	N	**N**	N	N	Y	N	**N**	N	N
S. pneumoniae V17_201194 MUZ	Bruker database	N	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae V17_201197 MUZ	Bruker database	Y	N	Y	**N**	N	N	N	Y	**N**	Y	N
S. pneumoniae V17_201198 MUZ	Bruker database	N	N	Y	**N**	N	Y	N	Y	**N**	Y	N
S. pneumoniae V17_201199 MUZ	Bruker database	N	N	Y	**N**	N	N	N	N	**N**	Y	N

[a] N, absence of peak; Y, presence of peak at a given *m/z* value.
MALDI-TOF MS, matrix-associated laser desorption/ionization time-of-flight mass spectrometry; MSP, Main Spectrum Profile.

systemic limitations in current diagnostic workflows. MALDI-TOF MS identified the three isolates as *S. pneumoniae* with a score exceeding 2.0. Phenotypically, the isolates closely resembled *S. pneumoniae* in key tests such as α-hemolysis, optochin sensitivity, and bile solubility, further complicating their differentiation. Subsequent genomic analysis confirmed that these three isolates should be classified as *S. parapneumoniae* based on ANI and dDDH values that unequivocally placed them within the species boundary (Supplementary Table S3). Although no additional *S. parapneumoniae* strains have been reported at the time of writing, our findings suggest that clinical isolates previously identified as *S. pneumoniae* may, in fact, be *S. parapneumoniae*. The specificity of our designed hypothetical protein gene primers for detecting *S. parapneumoniae* may offer a potential rapid molecular tool for targeted screening. However, further validation is required to confirm its specificity and sensitivity using a larger collection of *S. pneumoniae* and closely related streptococci, including mitis group species.

The clinical and epidemiological implications of the present findings are significant. All three *S. parapneumoniae* isolates in Hong Kong were recovered from patients with respiratory infections (bronchitis or pneumonia), with one patient exhibiting underlying comorbidity. These cases occurred in 2016 and 2023, with no travel history reported prior to symptom onset. This temporal distribution, combined with the recent identification of *S. para-*

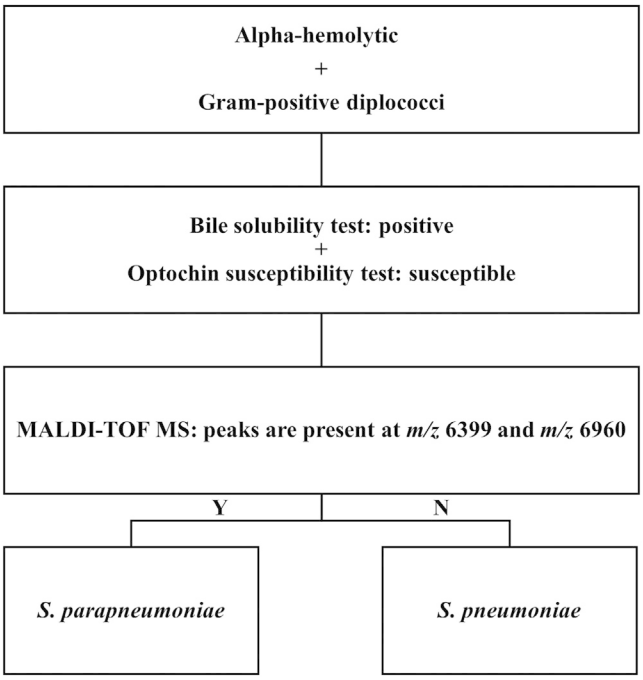


Figure 4. Proposed identification scheme for differentiating *Streptococcus parapneumoniae* from *Streptococcus pneumoniae*.
1. Characteristics of α-hemolytic and Gram-positive diplococci. 2. A positive bile solubility result along with optochin susceptibility indicates a probable identification of either S. *parapneumoniae* or *S. pneumoniae.* 3. MALDI-TOF MS analysis: detection of peaks at *m/z* 6,399 and *m/z* 6,960 confirms the isolate as *S. parapneumoniae*; otherwise, it is classified as *S. pneumoniae.*
MALDI-TOF MS, matrix-associated laser desorption/ionization time-of-flight mass spectrometry.

pneumoniae SP4011^T in Japan in 2024 [10], suggests the global presence of this species for nearly a decade, potentially causing both respiratory and invasive infections akin to *S. pneumoniae*. The Japanese case, involving an older adult individual with metabolic syndrome who developed bacteremia secondary to acute pyelonephritis [10], underscores the pathogenic potential of *S. parapneumoniae* across diverse patient populations, including both immunocompromised and immunocompetent individuals. Indeed, genomic comparisons have revealed shared critical virulence genes between *S. parapneumoniae* and *S. pneumoniae*, involved in adhesion (e.g., *cbpE, and pavA* genes), invasion (e.g., *lytA, htrA and srtA* genes), immune evasion (e.g., *psaA* genes), and toxin release (e.g., *ply* genes) (Figure 2b and c). These shared pathogenic mechanisms likely account for the overlapping clinical syndromes associated with *S. pneumoniae*. Notably, all four *S. parapneumoniae* isolates, including the type strain *S. parapneumoniae* SP4011^T, exhibited multidrug resistance and levofloxacin resistance, which is not common among *S. pneumoniae* isolates in Hong Kong. This resistance pattern, uncommon in local *S. pneumoniae* isolates 0-4.4% in recent surveillance [29], raises concerns that such misidentification could distort local antimicrobial resistance surveillance data of *S. pneumoniae* and lead to inappropriate empiric therapy.

The utilization of spectral peaks analysis-based identification has shown to be a promising approach for distinguishing between *S. parapneumoniae* and *S. pneumoniae*. MALDI-TOF MS, a rapid, cost-effective, and high-throughput diagnostic technology widely adopted in clinical microbiology, has significantly contributed to bacterial identification over the past decade [30]. Although MALDI-TOF MS has demonstrated success in differentiating some species within the *S. mitis* group that bear close resemblance to *S. pneumoniae* [31], concerns regarding its specificity persist [32]. For example, Yahiaoui *et al.* have highlighted the limitations of the Bruker database in accurately identifying *S. pneumoniae*, with over 50% of non-pneumococcal isolates being misidentified [32]. This lack of specificity is attributed to the low phylogenetic resolution of MALDI-TOF MS systems, which only capture a small fraction of the microbial proteome in whole-cell spectra [32]. In this study, we encountered a similar challenge of misidentification within the mitis group, despite ongoing updates to reference databases. Of the 385 pneumococcal isolates screened, three were misidentified, highlighting the current limitations of MALDI-TOF spectral libraries in effectively resolving species within the mitis group. Previous studies showed that mass spectral analysis offers improved resolution for bacterial identification [30,31], prompting us to create MSPs of the three isolates using a full extraction protocol with 30 replicates to reveal distinct profiles between *S. parapneumoniae* and *S. pneumoniae*. We further compared these three isolates with 50 additional *S. pneumoniae* isolates and identified two characteristic peaks at *m/z* 6,399 and *m/z* 6,960 that can accurately distinguish the two species (Figure 3b, Table 1). These peaks represent potential biomarkers for enhancing MALDI-TOF databases, enabling rapid and accurate differentiation between *S. parapneumoniae* and *S. pneumoniae* (Figure 4). Further research is warranted to validate the findings by expanding the study to include a broader range of strains of *S. parapneumoniae, S. pneumoniae*, and other closely related streptococci, particularly those within the mitis group.

Conclusion

In conclusion, this study identified three *S. parapneumoniae* isolates from patients with respiratory infections in Hong Kong, following the initial report of this novel species in Japan in 2024. Similar to the first strain in Japan, all isolates were initially misidentified as *S. pneumoniae*. The resistance of these isolates to levofloxacin, uncommon in local *S. pneumoniae* isolates, raises concerns about the distortion of local antimicrobial resistance surveillance data and the risk of inappropriate empiric therapy. The utilization of spectral peak analysis for distinguishing between *S. parapneumoniae* and *S. pneumoniae* shows promise to improve diagnostic accuracy in clinical microbiology. Further research is warranted to understand the global prevalence and distribution of this emerging pathogen.

Declaration of competing interest

The authors have no competing interests to declare.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

The whole-genome sequences of the three case isolates have been deposited in GenBank under the accession JBCHKK000000000, JBLGTJ000000000, JBLGTI000000000, respectively.

Ethical approval statement

The study was conducted following the approval of the Institute Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (Reference number: UW 16-365).

Author contributions

Conceptualization, JLLT; Methodology and investigation, YZ, YM, HHL, WHL, AHYN, JHKC, SYCS, JLLT; Formal analysis, YZ, YM, JLLT;

Writing - original draft, YZ, YM, JLLT; Writing - review and editing, HHL, WHL, AHYN, JHKC, SYCS, VCCC, KYY, MLY; Resources, VCCC, KYY, MLY; Supervision, JLLT. All authors have read and approved the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2025.108042.

References

[1] Parte AC. LPSN – list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* 2014;**42**:D613–16. doi:10.1093/nar/gkt1111.

[2] Lee IPA, Andam CP. Frequencies and characteristics of genome-wide recombination in Streptococcus agalactiae, Streptococcus pyogenes, and Streptococcus suis. *Sci Rep* 2022;**12**:1515. doi:10.1038/s41598-022-04995-5.

[3] GBD 2015 LRI Collaborators. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect Dis* 2017;**17**:1133–61. doi:10.1016/S1473-3099(17)30396-1.

[4] Ganaie F, Saad JS, McGee L, van Tonder AJ, Bentley SD, Lo SW, 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. doi:10.1128/mBio.00937-20.

[5] Ganaie FA, Saad JS, Lo SW, McGee L, Bentley SD, van Tonder AJ, et al. Discovery and characterization of pneumococcal serogroup 36 capsule subtypes, serotypes 36A and 36B. *J Clin Microbiol* 2023;**61**:e0002423. doi:10.1128/jcm.00024-23.

[6] Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, 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. doi:10.1016/S2214-109X(18)30247-X.

[7] Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, et al. Phylogenomics and the dynamic genome evolution of the genus Streptococcus. *Genome Biol Evol* 2014;**6**:741–53. doi:10.1093/gbe/evu048.

[8] Jensen A, Scholz CFP, Kilian M. Re-evaluation of the taxonomy of the Mitis group of the genus Streptococcus based on whole genome phylogenetic analyses, and proposed reclassification of Streptococcus dentisani as Streptococcus oralis subsp. dentisani comb. nov., Streptococcus tigurinus as Streptococcus oralis subsp. tigurinus comb. nov., and Streptococcus oligofermentans as a later synonym of Streptococcus cristatus. *Int J Syst Evol Microbiol* 2016;**66**:4803–20. doi:10.1099/ijsem.0.001433.

[9] Aupaix A, Verroken A, Rodriguez-Villalobos H. Evaluation of a new protocol for rapid identification of Streptococcus pneumoniae in blood cultures using the modified bile solubility test: Gram staining is still standing. *J Clin Microbiol* 2025;**63**:e0122224. doi:10.1128/jcm.01222-24.

[10] Katayama Y, Morita M, Chang B, Katagiri D, Ishikane M, Yamada G, et al. Identification and characterization of a novel α-haemolytic streptococci, Streptococcus parapneumoniae sp. nov., which caused bacteremia with pyelonephritis. *Int J Med Microbiol* 2024;**315**:151625. doi:10.1016/j.ijmm.2024.151625.

[11] Teng JLL, Ma Y, Chen JHK, Luo R, Foo CH, Li TT, et al. Streptococcus oriscaviae sp. nov. infection associated with Guinea Pigs. *Microbiol Spectr* 2022;**10**:e0001422. doi:10.1128/spectrum.00014-22.

[12] Wiley *Wiley.com*. [accessed 22 March 2025] https://www.wiley.com/en-us/Manual+of+Clinical+Microbiology%2C+11th+Edition-p-9781683672807 .

[13] Chen JHK, Leung H-Y, Wong CMC, Yuen KY, Cheng VC. Prevalence and characteristics of invasive Staphylococcus argenteus among patients with bacteremia in Hong Kong. *Microorganisms* 2023;**11**:2435. doi:10.3390/microorganisms11102435.

[14] Clinical Laboratory Standards Institute *M02 | Performance standards for antimicrobial disk susceptibility tests*; 2024 *https://clsi.org/shop/standards/m02/* [accessed 08 April 2025].

[15] Teng JLL, Fong JYH, Fok KMN, Lee HH, Chiu TH, Tang Y, et al. Tsukamurella asaccharolytica sp. nov., Tsukamurella conjunctivitidis sp. nov. and Tsukamurella sputi sp. nov., isolated from patients with bacteraemia, conjunctivitis and respiratory infection in Hong Kong. *Int J Syst Evol Microbiol* 2020;**70**:995–1006. doi:10.1099/ijsem.0.003861.

[16] Tamura K, Stecher G, Kumar S. MEGA 11: Molecular evolutionary genetics analysis version 11. *Mol Biol Evol* 2021;**38**:3022–7. doi:10.1093/molbev/msab120.

[17] Sánchez-Pérez S, Comas-Basté O, Duelo A, Veciana-Nogués MT, Berlanga M, Latorre-Moratalla ML, et al. Intestinal dysbiosis in patients with histamine intolerance. *Nutrients* 2022;**14**:1774. doi:10.3390/nu14091774.

[18] Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, et al. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 2021;**49**:D1020–8. doi:10.1093/nar/gkaa1105.

[19] Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 2004;**32**:11–16. doi:10.1093/nar/gkh152.

[20] Pan Y, An H, Fu T, Zhao S, Zhang C, Xiao G, et al. Characterization of Streptococcus pluranimalium from a cattle with mastitis by whole genome sequencing and functional validation. *BMC Microbiol* 2018;**18**:182. doi:10.1186/s12866-018-1327-0.

[21] Na S-I, Kim YO, Yoon S-H, Ha SM, Baek I, Chun J. UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol* 2018;**56**:280–5. doi:10.1007/s12275-018-8014-6.

[22] Freiwald A, Sauer S. Phylogenetic classification and identification of bacteria by mass spectrometry. *Nat Protoc* 2009;**4**:732–42. doi:10.1038/nprot.2009.37.

[23] Ye Z, Zhu J, Liu Y, Lu J. Rapid identification and typing of carbapenem-resistant Klebsiella pneumoniae using MALDI-TOF MS and machine learning. *Microb Biotechnol* 2025;**18**:e70184. doi:10.1111/1751-7915.70184.

[24] Tun T, Baleivanualala SC, Erdmann MB, Gimenez G, Aung HL. An extensively drug-resistant Pseudomonas aeruginosa isolate from Myanmar. *J Glob Antimicrob Resist* 2024;**36**:1–3. doi:10.1016/j.jgar.2023.11.004.

[25] Zhou S, Liu B, Zheng D, Chen L, Yang J. VFDB 2025: an integrated resource for exploring anti-virulence compounds. *Nucleic Acids Res* 2025;**53**:D871–7. doi:10.1093/nar/gkae968.

[26] Pereira JM, Xu S, Leong JM, Sousa S. The Yin and Yang of pneumolysin during pneumococcal infection. *Front Immunol* 2022;**13**:878244. doi:10.3389/fimmu.2022.878244.

[27] Maziero M, Lane D, Polard P, Bergé M. Fever-like temperature bursts promote competence development via an HtrA-dependent pathway in Streptococcus pneumoniae. *PLoS Genet* 2023;**19**:e1010946. doi:10.1371/journal.pgen.1010946.

[28] Gu K, Ding L, Wang Z, Sun Y, Sun X, Yang W, et al. Wogonin attenuates the pathogenicity of Streptococcus pneumoniae by double-target inhibition of pneumolysin and sortase A. *J Cell Mol Med* 2023;**27**:563–75. doi:10.1111/jcmm.17684.

[29] Centre for Health Protection, Department of Health. Streptococcus pneumoniae, https://www.chp.gov.hk/en/statistics/data/10/100044/6967.html; n.d. [accessed 15 May 2025].

[30] Teng JLL, Tang Y, Wong SSY, Fong JY, Zhao Z, Wong CP, et al. MALDI-TOF MS for identification of Tsukamurella species: Tsukamurella tyrosinosolvens as the predominant species associated with ocular infections. *Emerg Microbes Infect* 2018;**7**:80. doi:10.1038/s41426-018-0083-4.

[31] Basheer M, Sasic K, Carlsson A, Vestberg N, Henriques-Normark B, Blomqvist K, et al. A novel LAMP-based assay for the identification of Streptococcus pneumoniae and Streptococcus pseudopneumoniae in clinical isolates. *J Clin Microbiol* 2024;**62**:e0091224. doi:10.1128/jcm.00912-24.

[32] Yahiaoui RY, Goessens WH, Stobberingh EE, Verbon A. Differentiation between Streptococcus pneumoniae and other viridans group streptococci by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Clin Microbiol Infect* 2020;**26**:1088 .e1. doi:10.1016/j.cmi.2019.11.024.