## **Genomic analysis of invasive and non-invasive disease-causing *Streptococcus pneumoniae* among children between 2014-2023 in Suzhou, China**

## 

LiLi Huang1†, Alannah C. King2†, Yue Liu3†, Harry C. H. Hung2, Stephen D Bentley2, Yue Jiang4, PanPan Lv4, Xuebin Xu3\*, Stephanie W. Lo2,5,6\*, Mingliang Chen4\*

## 

1 Laboratory department, Children's Hospital of Soochow University, Suzhou, Jiangsu, China

2 Parasites and Microbes, The Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

3 Department of Microbiology, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China

4 Research and Translational Laboratory of Acute Injury and Secondary Infection, and Department of Laboratory Medicine, Minhang Hospital, Fudan University, Shanghai, China

5 Milner Centre for Evolution, Department of Life Sciences, University of Bath, Bath, UK

6 The Great Ormond Street Institute of Child Health, University College London, London, UK

## 

† These authors have contributed equally to this work.

\* **Correspondence:**

Mingliang Chen; Stephanie W. Lo; Xuebin Xu

cmlbright@126.com; sl28@sanger.ac.uk; xuxuebin@scdc.sh.cn.

**Data Summary**

## Genome sequences are deposited in the European Nucleotide Archive (ENA), the accession number, and a phylogenetic snapshot is available in<https://microreact.org/project/gps2-china>. Metadata of the pneumococcal isolates in this study is submitted as a supplementary file. The authors confirm all supporting data, code, and protocols have been provided within the article or through supplementary data files.

## **Abstract**

## In 2017, 553,000 clinical *Streptococcus pneumoniae* cases in children were reported in China, yet the pneumococcal conjugate vaccine (PCV), which targets the pneumococcal capsule, is not included in the Chinese National Immunisation Program (NIP) for children. Therefore, the PCV uptake rate is very low. To investigate the *S. pneumoniae* population over the past 10 years in China, we collected 418 *S. pneumoniae* isolates from children with pneumococcal diseases in Suzhou, China, 2014-2023, and whole-genome sequenced them. A total of 27 serotypes expressed by 36 Global Pneumococcal Sequence Clusters (GPSCs) that encompassed 72 sequence types (STs) were identified, with serotype 19F (38.3%, n = 160) and GPSC1 (60.8%, n = 254) as the predominant serotype and lineage, respectively. We found that the majority (64.8%, n = 271) of samples represented serotypes that are covered by the GSK 10-valent pneumococcal conjugate vaccine (PCV10) formulation, and that even more were covered by the SII PCV10 formulation (89.2%, n = 373). Almost all (94.3%, n = 394) samples represented serotypes that are included within the 13-valent PCV (PCV13) vaccine formulation. This suggests that the inclusion of the PCV vaccine in the NIP would lead to significant benefits for child health. Also, we observed that there were no significant differences in the serotypes or lineages seen in cases of invasive and non-invasive pneumococcal disease. Additionally, we investigated the prevalence of antimicrobial resistance (AMR) within the population and found that 99.8% (n = 417) of isolates were predicted to be resistant to at least one antibiotic tested. This again supports the need to increase PCV uptake to both prevent infections with antibiotic-resistant *S. pneumoniae*, and to reduce the number of infections in general, consequently lowering the consumption of antibiotics. In summary, the PCV13 vaccine could potentially cover over 90% of invasive and non-invasive *S. pneumoniae* isolates in Suzhou. Therefore, increasing the uptake of PCV vaccines by including PCV13 in the NIP would lead to significant benefits for child health.

## 

## **Impact statement**

## Despite high levels of pneumococcal disease, the PCV vaccine is not part of the Chinese NIP. This study suggests that including PCV is likely to reduce the prevalence of the pneumococcal serotypes associated with disease in Suzhou. Additionally, this study raises awareness of the exceptionally high levels of antimicrobial resistance (AMR) within the Suzhou pneumococcal population, which represents a significant public health risk. The introduction of PCV in NIP, alongside with implementation antibiotic stewardship, has a strong potential to reduce AMR.

## **Data Summary**

## Genome sequences are deposited in the European Nucleotide Archive (ENA), the accession number, and a phylogenetic snapshot is available in<https://microreact.org/project/gps2-china>. Data can also be found and downloaded through Monocle (<https://data-viewer.monocle.sanger.ac.uk/project/gps>). Metadata of the pneumococcal isolates in this study is submitted as a supplementary file. The authors confirm all supporting data, code, and protocols have been provided within the article or through supplementary data files.

## 

## **1. Introduction**

*Streptococcus pneumoniae (*pneumococcus*)* is a leading vaccine-preventable cause of childhood diseases. This disease can be classed as invasive pneumococcal disease (IPD), where pneumococcal bacteria can be cultured from a usually sterile site [4], and non-invasive pneumococcal disease, such as otitis media or pneumonia. It was estimated that nearly 9 million pneumococcal disease cases occurred in children aged under five years old worldwide in 2015, and that the pneumococcus was responsible for 317,000 deaths [1]. In China, *S. pneumoniae* is the predominant bacterium responsible for bacterial infections in children; The Infectious Disease Surveillance of Pediatrics (ISPED) data shows that it is the leading bacterium isolated from lower respiratory tract samples, accounting for 22.5% of clinical cases [2]. In 2017, China reported 553,000 clinical pneumococcal cases in children, with 218,200 severe cases, including 8,000 deaths [3]. Therefore, *S. pneumoniae* remains a significant cause of invasive diseases in children under five years old. Despite preventive vaccines, pneumococcal diseases continue to be a substantial health burden, with China reporting thousands of severe cases and deaths annually.

Pneumococcal conjugate vaccines (PCVs) have proved to be effective in reducing childhood invasive pneumococcal disease (IPD) in many countries [5–10]. However, despite PCV13 being licensed for use in 2016, PCV vaccinations are not part of the Chinese National Immunization Program (NIP), and the administration of PCV vaccines largely relies on personal payment and voluntary vaccination. Domestic PCV vaccines cost approximately 600 RMB (~$83 USD) per dose, whilst imported vaccines cost around 700 RMB (~$96 USD) per dose. The vaccination schedule typically follows a "3+1" regimen of three primary doses, plus one booster dose. Completing the full vaccination course can cost between 2,400 to 2,800 RMB (~$330-385 USD), making it a significant financial burden for many families [13]. Due to the high cost of PCV vaccines and a lack of public awareness regarding their necessity and importance, the PCV vaccination rate in China remains low. Even in economically developed areas like Shanghai, the PCV13 vaccination rate is only 10.2%, whilst in less developed Western regions the vaccination rate is below 1% [2]. There are three main PCV vaccine formulations available on the market. Two are ten-valent vaccines: PCV10 (GSK) contains the serotypes 1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, and 23F, whilst PCV10 (SII) contains the serotypes 1, 5, 6A, 6B, 7F, 9V, 14, 19A, 19F, and 23F. Finally, the 12-valent PCV13 includes all serotypes in both PCV10 formulations alongside serotype 3.

The city of Suzhou, located in Jiangsu province in eastern China, has a population exceeding 10 million and is one of the most economically developed cities in the country. Jiangsu province has a very low PCV vaccination rate of 1.7% [2]. The Children's Hospital of Soochow University is a prominent paediatric tertiary hospital in the Suzhou region, featuring 1,500 beds. Annually, the hospital manages over 3 million outpatients and admits more than 80,000 inpatients.

Antibiotic consumption in China is very high due to a variety of factors including patient pressure, financial incentives, and a lack of knowledge amongst physicians when prescribing antibiotics, and this has led to generally very high rates of antimicrobial resistance (AMR) [11]. Across China, *S. pneumoniae* has been seen to exhibit significant resistance to multiple antibiotics, particularly clindamycin, erythromycin, and tetracycline, with resistance rates exceeding 94% [12]. More than 50% of isolates from children's respiratory samples are non-susceptible to penicillin [13–15]. Studies in Suzhou children show that the non-susceptibility rates to common first-line treatments for respiratory infections, such as penicillin, amoxicillin, and cefotaxime, are 9.5%, 27.7%, and 27.2%, respectively [13–15].

In this study, we undertook whole-genome sequencing on a collection of 418 disease-causing pneumococcal isolates collected from young children in Suzhou, China, between 2014-2023 as part of the Global Pneumococcal Sequencing (GPS) project [17]. We aimed to update knowledge on the current lineages, serotypes, and antibiotic resistance of *S. pneumoniae* circulating in the city of Suzhou, Jiangsu province, China. The data generated will provide evidence to support future pneumococcal disease prevention and treatment strategies, including the inclusion of the PCV13 vaccine in the NIP.

## **2. Materials and methods**

## **2.1 Bacterial isolates**

All patients with infections caused by *S. pneumoniae* at the Children's Hospital of Soochow University from 2014 to 2023 were enrolled in this study. All isolates in this study were cultured from clinical specimens obtained from children under 16 years old. Pus and tissue were plated on Columbia blood agar plates. Blood and cerebrospinal fluid specimens were cultured using the BD BACTEC™ FX blood culture system, and positive samples were transferred to Columbia blood agar plates. The plates were incubated with 5% CO2 in 37°C for 18-24 hours. The colonies that showed α-haemolysis on the plates were further identified using Optochin disk sensitivity. For an ambiguous result on Optochin, bile soluble test was performed to confirm the result. All the suspected isolates were also tested using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS; bioMérieux). All the pneumococcal isolates were stored in 10% skim milk at -80°C.

**2.2 Genome sequencing and analysis**

The pneumococcal isolates were sequenced on an Illumina HiSeq platform to produce paired-end reads of 150 base pairs in length and raw data were deposited to the European Nucleotide Archive (ENA) (**Supplementary Data 1**). Whole genome sequencing (WGS) data was processed as previously described [16]. Briefly, we inferred serotype and resistance for 16 antibiotics, including penicillin (PEN), amoxicillin (AMO), meropenem (MER), cefotaxime (TAX), ceftriaxone (CFT), cefuroxime (CFX), erythromycin (ERY), clindamycin (CLI), quinupristin-dalfopristin (SYN), linezolid (LZO), cotrimoxazole (COT), tetracycline (TET), levofloxacin (LFX), chloramphenicol (CHL), rifampin (RIF), doxycycline (DOX), and vancomycin (VAN) from the genomic data using SeroBA (version 1.0.0) [17] and a resistance detection pipeline developed by CDC, respectively [18–22]. Vaccine serotypes (VT) were defined by serotypes in the PCV13 (Pfizer), including 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F, 23F, and other serotypes were grouped as non-vaccine serotypes (NVT). The predicted MIC values were interpreted according to Clinical Laboratory Standards Institute (CLSI) breakpoints [23]. Penicillin resistance was defined as MIC of ≥0.12 µg/ml, according to meningitis breakpoints. Our previous study showed high concordance between phenotypic and genotypic results of serotypes and antibiotic resistance [23]. Among these antibiotics, the twelve commonly used antibiotics amoxicillin, ceftriaxone, cefotaxime, cefuroxime, meropenem, penicillin, chloramphenicol, clindamycin, cotrimoxazole, doxycycline, erythromycin, and tetracycline were used for resistant pattern analysis. Pneumococcal lineage was defined by assigning a Global Pneumococcal Sequence Cluster (GPSC) to each isolate using PopPUNK (version 2.6.0) [24] and v6 of the PopPUNK reference database of pneumococcal isolates [25]. In addition, multilocus sequence type (MLST) was derived from the genome data using MLSTcheck (version 2.1.1706216) [26]. Phylogenetic analysis was performed on all isolates by constructing a maximum likelihood tree using FastTree (version 2.1.10) [27] based on single nucleotide polymorphisms (SNPs) extracted from an alignment generated by mapping reads to the reference genome of *S. pneumoniae* ATCC 700669 (NCBI accession number FM211187) using Smalt (version 0.7.4; https://github.com/rcallahan/smalt). The metadata (Supplementary Data 2) and analysis results can be interactively visualised online using the Microreact tool at<https://microreact.org/project/gps2-china>.

**2.3 Statistical analysis**

For the purposes of comparing cases of IPD and non-IPD, the clinical manifestations “otitis media”, “eye discharge”, “pneumonia”, “urinary tract infection (UTI)”, “bronchitis” were considered non-IPD. “other” and “joint fluid/pus” were classed as unknown. The clinical manifestations “meningitis”, “sepsis”, “bacteremia”, and “pneumonia, bacteremia” were categorised as IPD. Where we analysed vaccine types (VT) compared to non-vaccine types (NVTs), VTs were the serotypes present in the PCV13 formulation: 19F, 19A, 23F, 6B, 14, 3, 6A, 18C, 9V, 4, 1, 5, and 7F. This encompasses serotypes in either of the PCV10 formulations.

Differences between the IPD samples and non-IPD samples were calculated using Fisher’s Exact test using a dataset that only included samples from pre-2018, when the samplings strategy changed. Before carrying out any Fisher’s Exact test, we calculated the number of samples that we need to achieve an 80% statistical power with a significant level of p-value <0.05 using the R package pwr which contains functions for basic power calculation [28]. Where we did not have enough samples to achieve 80% power, Fisher’s Exact Test was not performed. When comparing the differences in serotype and GPSC distribution between IPD and non-IPD cases, Fisher’s Exact Test was run in R with a simulated p value and 500,000 bootstraps due to the large number of variables. Multiple testing was adjusted using the Benjamin-Hochberg false discovery rate of 5%.

To analyse differences in AMR between IPD and non-IPD isolates, the dataset was filtered to only contain samples from between 2014-2017, as the sampling strategy changed in 2018 to focus on IPD cases. Fisher’s Exact test was performed and multiple testing was adjusted using the Benjamin-Hochberg false discovery rate of 5%. Differences were considered significant if p < 0.05.

The R script used for this anlaysis can be found at <https://github.com/ak2022/GPS2_China>.

## 

## **3. Results and Discussion**

## **3.1 Data Overview**

A total of 418 *S. pneumoniae* isolates were analysed in this study. 95.0% (n = 397/418) of samples were from children aged five years or under. The remaining samples (n = 21) were from children between the ages of 6 and 15 years old. The clinical manifestations observed were otitis media (n = 244), meningitis (n = 71), sepsis (n = 55), pneumonia (n = 18), bacteraemia (n = 12), bronchitis (n = 7), UTI (n = 1), and other (n = 9). There was one recorded case of both pneumonia and bacteraemia. Among the 418 isolates, 139 isolates were identified as invasive pneumococcal disease (IPD) strains, 271 isolates as non-IPD strains, and 8 isolates were categorized as unknown.

We observed a noticeable decline in the number of collected isolates per year in the dataset since 2018, with a greater proportion of isolates being from IPD cases compared to non-IPD (**Figure S1**). This reduction is attributed to the hospital’s increased focus on invasive pneumococcal disease cases in recent years. As a result, from 2018 onwards, the hospital has primarily collected isolates from blood and cerebrospinal fluid, leading to a significant drop in the overall number of isolates, and a greater proportion of recent isolates being from IPD cases (**Figure S1**).

## **3.2 Serotypes and PCV Coverage**

Serotypes predicted from WGS data revealed 27 serotypes expressed by 36 pneumococcal lineages (or Global Pneumococcal Sequence Clusters, GPSC) that encompassed 72 sequence types (STs) (**Figure 1**). The five most prevalent serotypes were serotype 19F (38.3%, n = 160), 19A (23.7%, n = 99), 23F (8.4%, n = 35), 6B (7.4%, n = 31), and 14 (7.2%, n = 30), and together they accounted for 85% of the whole collection. All of these major serotypes are covered by PCV10 (SII) and PCV13, however serotype 19A is not covered by PCV10 (GSK). Across the whole dataset, we found PCV10 (GSK) to have 64.8% coverage, PCV10 (SII) to have 89.2%, and PCV13 to have 94.6%. The large difference in coverage between PCV10 (GSK) and PCV10 (SII)/PCV13 is due to the lack of serotype 19A. PCV13 having the highest coverage underscores the limited impact of the PCV13 vaccine in Suzhou currently; Although the PCV13 vaccine was made available privately within China in 2016 [29], the low vaccination rate of 6.5% (2016-2018) [30] is leading to the vaccination having very little impact on the pneumococcal population in Suzhou, Jiangsu.

Although PCV13 showed the highest coverage of serotypes in the dataset, it was not 100%. The most prevalent non-PCV13 serotypes were serotype 23A (n = 6, four expressed by GPSC5 and two expressed by GPSC10), followed by serotype 6D (n = 3, all GPSC856). Overall, only 5.7% (n = 24) of samples within the dataset were non-vaccine serotypes. It is important to monitor these serotypes and lineages, as they are likely to persist even in a population with high PCV13 coverage.

## **3.3 Pneumococcal lineages**

Among the 418 isolates, the five most predominant pneumococcal lineages were GPSC1 (60.8%, n = 254), GPSC4 (6.5%, n = 27), GPSC23 (3.3%, n = 14), GPSC16 (3.1%, n = 13), and GPSC12 (2.9%, n = 12), which together accounted for 77% (320/418) of the collection (**Figure 1**). All major lineages express serotypes found within PCV13.

There were four lineages expressing a combination of PCV13 vaccine serotypes (VT) and non-vaccine serotypes (NVT); GPSC5 (n = 9), GPSC10 (n = 6), GPSC107 (n = 6), and GPSC230 (n = 4). The largest was GPSC5, which expressed NVT serotype 23A (44.4%, n = 4) alongside VTs 19F (22.2%, n = 2), and 23F (33.3%, n = 3). Within the GPS1 dataset (n = 26,306, <https://www.pneumogen.net/gps/>, last accessed in May 2024), GPSC5 and GPSC10 were seen to be global-spreading lineages. However, GPSC107 has only been identified in China (n = 6) and Thailand (n = 21), and the vast majority (98.3%, n = 57/58) of GPSC230 isolates have been identified from Asian countries such as Nepal (44.8%, n = 26), China (15.5%, n = 9), India (15.5%, n = 9), and Bangladesh (13.7%, n = 8).

Eight GPSC lineages expressed only NVT serotypes. These were GPSC152, GPSC856, GPSC69, GPSC158, GPSC186, GPSC215, GPSC224, and GPSC45. Of these, only GPSC152, GPSC856, and GPSC69 were represented by more than a single sample. GPSC152 expressed serotypes 15C (66.7%, n = 2), and 15B (33.3%, n = 1). According to the GPS1 database, GPSC152 is only seen within China. Specifically, the GPSC152 samples with a region are from Hong Kong (76.0%, n = 19), the Chongqing municipality (16.0%, n = 4), the Beijing municipality (4.0%, n = 1), and Gansu province (4.0%, n = 1). Two GPSC152 samples have no regional data. All these regions are geographically separated from each other, suggesting a wider existence of GPSC152 pneumococci across China, and that GPSC152 may represent a China-specific lineage. Within the GPS1 dataset, the samples were mostly serotype 15B or 15C (92.6%, n = 25), with two members expressing serotype 13 (7.4%, n = 2). This reflects what is seen in the Suzhou dataset. GPSC856 expressed only serotype 6D in all samples (n = 3). There was only one other sample of GPSC856 in the GPS1 database, identified in the Chongqing municipality of China and expressing NVT serotype 6D, in line with the Suzhou dataset. It is possible that GPSC856 also reflects a China-specific lineage, however the small number of samples makes this difficult to confirm. In the Suzhou dataset, GPSC69 only expressed serotype 15A (n = 2). In the GPS1 dataset, GPSC69 members were found in Peru (33.3%, n = 15), China (22.2%, n = 10), and Nepal (22.2%, n = 10) expressing mostly serotype 15A, with a single isolate from Peru expressing serotype 19A. This suggests that GPSC69 is also a wide-spreading lineage.

This data suggests that the introduction of any PCV vaccine to the NIP would have a huge impact on the population of pneumococci causing disease. However, there are lineages present in the population that either express a mix of NVTs and VTs, or only NVTs. It is likely that these populations will expand to fill the gap left after the introduction of a childhood PCV programme, and so continuous monitoring would be beneficial.

**3.4 IPD and Non-IPD Manifestations**

We investigated the difference in serotype and lineage distribution stratified by IPD cases and non-IPD cases for samples collected before 2018 (**Figure 2**). Interestingly, there were no statistically significant differences between the IPD cases and non-IPD cases from the perspective of GPSC lineage (p = 1). However, there was a statistically significant difference from the perspective of serotype (p = 0.02), with a much greater proportion of non-IPD cases being caused by serotypes 19F and 19A (**Figure 2**). Both datasets had the greatest proportion of samples from the lineage GPSC1, and the dominant serotype was 19F.

## **3.5 Antibiotic Resistance**

Finally, we investigated the prevalence of predicted antibiotic resistance based on genomic sequence within the dataset (**Table 1**). Of the 418 isolates, 99.5% (n = 416/418) were predicted to be multidrug-resistant, defined as being non-susceptible to at least three classes of antibiotics (**Figure 3**), including all samples within GPSC1, the largest lineage seen in the dataset. Only a single sample was predicted to be susceptible to all classes of antibiotics investigated.

The non-susceptibility prevalence of the 418 *S*. *pneumoniae* isolates to various antibiotics from 2014 to 2023 indicated consistently high non-susceptibility rates for many antibiotics (**Table S6**), suggesting widespread resistance. Certain antibiotics, such as clindamycin, meropenem, doxycycline, erythromycin, and tetracycline, exhibited high non-susceptibility prevalence, often nearing or reaching 100% (**Table 1**). In particular, the high non-susceptibility rate of 76.1% to meropenem reflects concerns in a recent publication about high levels of carbapenem resistance in *S. pneumoniae* [31]. Chloramphenicol was notable with very low non-susceptibility rates (3.3%). Once a common treatment for pneumococcal infections, chloramphenicol is now seldom used due to its potential for serious side effects, including severe bone marrow suppression and "gray baby syndrome”, particularly in children [32].

Earlier in this study, GPSCs containing NVTs were identified as they could expand in number after the introduction of the PCV vaccination. All four lineages that were seen to express a mix of VT and NVT serotypes (GPSC5, GPSC10, GPSc107, GPSC230) were predicted to be multidrug resistant. Of the lineages that expressed only NVTs (GPSC152, GPSC856, GPSC69, GPSC158, GPSC186, GPSC215, GPSC224, and GPSC45), all were multidrug resistant aside from GPSC224, which was predicted to harbour resistance to both doxycycline and tetracycline. All NVT members were susceptible to amoxicillin and cefotaxime. Additionally, VT isolates show significantly higher non-susceptibility prevalences across most antibiotics compared to NVT isolates (**Table S7** and **S8**). For clinical treatment, it is essential to consider the first-line drugs, such as amoxicillin, ceftriaxone, cefotaxime, and cefuroxime, which showed low or no non-susceptibility prevalences in NVT isolates, suggesting them potentially more effective for treating *S. pneumoniae* infections.

Finally, we compared antibiotic resistance profiles between IPD and non-IPD isolates for each antibiotic. We found that amoxicillin, ceftriaxone, cefotaxime, and meropenem all had significantly more resistant isolates in non-IPD cases than in IPD cases (p < 0.05). This suggests that these antibiotics will remain beneficial for treating IPD (**Table S7** and **S8**).

Selecting appropriate first-line antibiotics is crucial for clinical treatment, particularly given the high non-susceptibility rates observed in non-IPD strains. The introduction of the PCV13 vaccine to the NIP is likely to reduce *S. pneumoniae* associated antibiotic resistance (AMR) as nearly all isolates in this study are VTs. Vaccination is expected to reduce the number of diseases caused by *S. pneumoniae*, thereby decreasing the need for antibiotic prescriptions and reducing exposure to antibiotics. While amoxicillin currently appears to be a viable treatment option in IPD, its effectiveness needs to be carefully monitored to prevent the development of resistance. To address the high antibiotic resistance rates in *S. pneumoniae*, a comprehensive approach is necessary. This includes enhancing preventive measures such as vaccination and infection control practices, implementing strict monitoring and control of antibiotic use and bacterial resistance, promoting the rational use of antibiotics through antimicrobial stewardship programs and updated prescription guidelines, and developing new drugs and alternative treatments. International cooperation and data sharing are essential, alongside public awareness campaigns and patient participation, to ensure antibiotics are used responsibly and effectively. These measures collectively aim to reduce antibiotic resistance and improve public health outcomes.

## **4. Conclusions**

Based on pneumococci collected from 2014 to 2023 in Suzhou, China, 27 serotypes expressed by 36 GPSC lineages were identified, with vaccine serotype 19F and multidrug-resistant GPSC1 as the predominant serotype and lineage, respectively. These study results show that the PCV13 vaccine could potentially cover over 90% of *S. pneumoniae* isolates from both invasive and non-invasive disease. An increased uptake of PCVs, as well as continued education for parents and practitioners about the benefits of the PCV13 vaccination, would lead to significant benefits in child health.

## **Funding information**

This study was co-funded by the National Natural Science Foundation of China (82272381), and the Bill & Melinda Gates Foundation (grant code OPP1034619) to the Wellcome Sanger Institute (core Wellcome grants 098051 and 206194).

## 

## **Conflict of Interest**

The authors have no conflicts of interest to report.

## 

## **Acknowledgements**

Special thanks to the researchers and clinicians who were involved in sample collection in China. We would like to thank all members of the Global Pneumococcal Sequencing Consortium for their contributions to create this rich global dataset. We are also grateful for the technical support from Wellcome Sanger Institute Pathogen Informatics team. Finally, we would like to thank Dr Lesley McGee at the CDC for her valuable comments and feedback on the manuscript. For the purposes of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

## **References**

1. Ning X, Li L, Liu J, Wang F, Tan K, Li W, et al. Invasive pneumococcal diseases in Chinese children: a multicentre hospital-based active surveillance from 2019 to 2021. Emerg Microbes Infect. 2024;13: 2332670.

2. 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. The Lancet Global Health. 2018;6: e744–e757.

3. Fu P, Xu H, Jing C, Deng J, Wang H, Hua C, et al. Bacterial Epidemiology and Antimicrobial Resistance Profiles in Children Reported by the ISPED Program in China, 2016 to 2020. Microbiol Spectr. 2021;9: e0028321.

4. Lai X, Wahl B, Yu W, Xu T, Zhang H, Garcia C, et al. National, regional, and provincial disease burden attributed to Streptococcus pneumoniae and Haemophilus influenzae type b in children in China: Modelled estimates for 2010-17. Lancet Reg Health West Pac. 2022;22: 100430.

5. Gottberg A von, von Gottberg A, de Gouveia L, Tempia S, Quan V, Meiring S, et al. Effects of Vaccination on Invasive Pneumococcal Disease in South Africa. New England Journal of Medicine. 2014. pp. 1889–1899. doi:10.1056/nejmoa1401914

6. Mackenzie GA, Hill PC, Jeffries DJ, Hossain I, Uchendu U, Ameh D, et al. Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based surveillance study. Lancet Infect Dis. 2016;16: 703–711.

7. Ladhani SN, Collins S, Djennad A, Sheppard CL, Borrow R, Fry NK, et al. Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000-17: a prospective national observational cohort study. Lancet Infect Dis. 2018;18: 441–451.

8. Moore MR, Link-Gelles R, Schaffner W, Lynfield R, Lexau C, Bennett NM, et al. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: analysis of multisite, population-based surveillance. Lancet Infect Dis. 2015;15: 301–309.

9. Weinberger R, von Kries R, van der Linden M, Rieck T, Siedler A, Falkenhorst G. Invasive pneumococcal disease in children under 16 years of age: Incomplete rebound in incidence after the maximum effect of PCV13 in 2012/13 in Germany. Vaccine. 2018;36: 572–577.

10. Turner P, Leab P, Ly S, Sao S, Miliya T, Heffelfinger JD, et al. Impact of 13-Valent Pneumococcal Conjugate Vaccine on Colonization and Invasive Disease in Cambodian Children. Clin Infect Dis. 2020;70: 1580–1588.

11. Song Y, Han Z, Song K, Zhen T. Antibiotic Consumption Trends in China: Evidence From Six-Year Surveillance Sales Records in Shandong Province. Front Pharmacol. 2020;11: 491.

12. Wang C-Y, Chen Y-H, Fang C, Zhou M-M, Xu H-M, Jing C-M, et al. Antibiotic resistance profiles and multidrug resistance patterns of Streptococcus pneumoniae in pediatrics: A multicenter retrospective study in mainland China. Medicine (Baltimore). 2019;98: e15942.

13. Zhou L, Yu S-J, Gao W, Yao K-H, Shen A-D, Yang Y-H. Serotype distribution and antibiotic resistance of 140 pneumococcal isolates from pediatric patients with upper respiratory infections in Beijing, 2010. Vaccine. 2011;29: 7704–7710.

14. Wang Q, Zhang F-F, Zhao C-J, Chen H-B, Wang Z-W, Guo Y, et al. Antimicrobial resistance and serotype distribution of Streptococcus pneumoniae isolated from multi-centers across China, 2010 - 2011. Zhonghua Jie He He Hu Xi Za Zhi. 2013;36: 106–112.

15. Yao K-H, Wang L-B, Zhao G-M, Zheng Y-J, Deng L, Huang J-F, et al. Pneumococcal serotype distribution and antimicrobial resistance in Chinese children hospitalized for pneumonia. Vaccine. 2011;29: 2296–2301.

16. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. EBioMedicine. 2019;43: 338–346.

17. Epping L, van Tonder AJ, Gladstone RA, The Global Pneumococcal Sequencing Consortium, Bentley SD, Page AJ, et al. SeroBA: rapid high-throughput serotyping of Streptococcus pneumoniae from whole genome sequence data. Microb Genom. 2018;4. doi:10.1099/mgen.0.000186

18. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE Jr, Walker H, et al. Penicillin-Binding Protein Transpeptidase Signatures for Tracking and Predicting beta-Lactam Resistance Levels in Streptococcus pneumoniae. MBio. 2016;7. doi:10.1128/mBio.00756-16

19. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Walker H, et al. Validation of β-lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. BMC Genomics. 2017;18: 621.

20. Metcalf BJ, Gertz RE Jr, Gladstone RA, Walker H, Sherwood LK, Jackson D, et al. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. Clin Microbiol Infect. 2016;22: 60 e9-60 e29.

21. Metcalf BJ, Chochua S, Gertz RE Jr, Li Z, Walker H, Tran T, et al. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. Clin Microbiol Infect. 2016;22: 1002.e1-1002.e8.

22. Sanger\_SPN\_Scripts-Ref. Github; Available: https://github.com/BenJamesMetcalf/Sanger\_SPN\_Scripts-Ref

23. Humphries R, Bobenchik AM, Hindler JA, Schuetz AN. Overview of changes to the Clinical and Laboratory Standards Institute Performance Standards for Antimicrobial Susceptibility Testing, M100, 31st edition. J Clin Microbiol. 2021;59: e0021321.

24. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res. 2019;29: 304–316.

25. GPS :: Global Pneumococcal Sequencing Project. [cited 27 Apr 2021]. Available: https://www.pneumogen.net/gps/training\_command\_line.html

26. J. Page A, Taylor B, A. Keane J. Multilocus sequence typing by blast from de novo assemblies against PubMLST. J Open Source Softw. 2016;1: 118.

27. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010. p. e9490. doi:10.1371/journal.pone.0009490

28. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Academic Press; 2013.

29. Chen K, Zhang X, Shan W, Zhao G, Zhang T. Serotype distribution of Streptococcus pneumoniae and potential impact of pneumococcal conjugate vaccines in China: A systematic review and meta-analysis. Hum Vaccin Immunother. 2018;14: 1453–1463.

30. Zhang T, Zhang J, Shao X, Feng S, Xu X, Zheng B, et al. Effectiveness of 13-valent pneumococcal conjugate vaccine against community acquired pneumonia among children in China, an observational cohort study. Vaccine. 2021;39: 4620–4627.

31. GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990-2021: a systematic analysis with forecasts to 2050. Lancet. 2024. doi:10.1016/S0140-6736(24)01867-1

32. Laferriere CI, Marks MI. Chloramphenicol: properties and clinical use. Pediatr Infect Dis. 1982;1: 257–264.

30. Laferrière C, Marks M. Chloramphenicol: Properties and clinical use. Pediatric Infectious Disease. 1982,1(4), 257-264. https://doi.org/10.1097/00006454-198207000-00016

## *Table 1 - Antibiotic Resistance across the samples from Suzhou, China.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Antibiotic** | **Non-susceptible Isolates** | | | |
| **VT (n = 394)** | | **NVT (n = 24)** | |
| **Count** | **Percentage %** | **Count** | **Percentage %** |
| Amoxicillin | 221 | 56.1 | 0 | 0 |
| Ceftriaxone (Men. BP\*) | 341 | 86.5 | 6 | 25.0 |
| Cefotaxime (Men. BP) | 305 | 77.4 | 0 | 0 |
| Cefuroxime | 353 | 89.6 | 9 | 37.5 |
| Meropenem | 316 | 80.2 | 2 | 8.3 |
| Penicillin (Men. BP) | 361 | 91.6 | 14 | 58.3 |
| Chloramphenicol | 13 | 3.3 | 1 | 4.2 |
| Clindamycin | 389 | 98.7 | 23 | 95.8 |
| Cotrimoxazole | 344 | 87.3 | 13 | 52.2 |
| Doxycycline | 391 | 99.2 | 24 | 100 |
| Erythromycin | 392 | 99.5 | 23 | 95.8 |
| Tetracycline | 391 | 99.2 | 24 | 100 |

\* Men, meningitis; BP, breakpoint based on Clinical & Laboratory Standards Institute Guidelines.

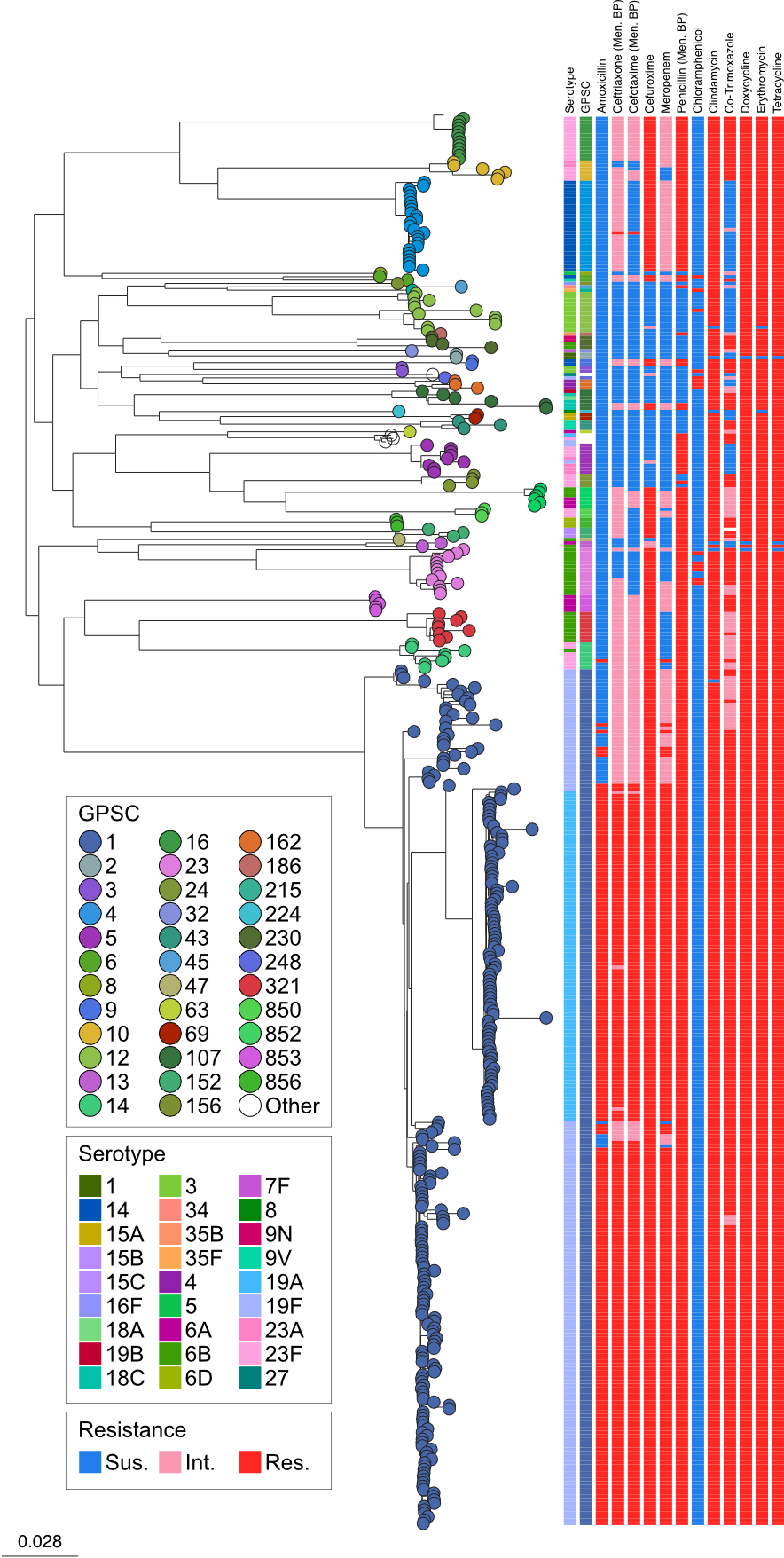
## 

## *Figure 1 - Distribution of Serotypes (A) and GPSCs (B) within the dataset, coloured by their coverage by the PCV10 (GSK), PCV10 (SII), or PCV13 vaccines.*

## 

## *Figure 2 - Distribution of GPSCs in IPD (A) and non-IPD (B) cases. There is no significant difference in the distribution of GPSCs between the two datasets, but a significant difference in serotypes. NVTs are denoted with stripes.*

## 



## *Figure 3 - Predicted antibiotic resistance of samples within the Suzhou dataset. Resistance for penicillin, ceftriaxone, and cefotaxime reported at the meningitis breakpoint.*

A graph of different colored bars

Description automatically generated

## *Figure S1 – Count of isolates collected each year, coloured by the source of the isolate. Shades of green show non-IPD collections, blue show IPD collections, and yellow show cases where it is unclear. After 2018, the strategy for collection changed to focus on IPD, leading to fewer total isolates being collected per year, and a big decrease in non-IPD samples.*