

Genomic epidemiology of *Streptococcus pneumoniae* serotype 16F lineages

Jolynne Mokaya^{1,*}, Kate C. Mellor¹, Gemma G. R. Murray^{1,2}, Akuzike Kalizang'oma^{3,4}, Cebile Lekhuleni^{5,6}, Heather J. Zar⁷, Mark P. Nicol⁸, Lesley McGee⁹, Stephen D. Bentley¹, Stephanie W. Lo^{1,10} and Felix Dube^{11,12}

Abstract

Due to the emergence of non-vaccine serotypes in vaccinated populations, *Streptococcus pneumoniae* remains a major global health challenge despite advances in vaccine development. Serotype 16F is among the predominant non-vaccine serotypes identified among vaccinated infants in South Africa (SA). To characterize lineages and antimicrobial resistance in 16F isolates obtained from South Africa and place the local findings in a global context, we analysed 10923 *S. pneumoniae* carriage isolates obtained from infants recruited as part of a broader SA birth cohort. We inferred serotype, resistance profile for penicillin, chloramphenicol, cotrimoxazole, erythromycin and tetracycline, and global pneumococcal sequence clusters (GPSCs) from genomic data. To ensure global representation, we also included *S. pneumoniae* carriage and disease isolates from the Global Pneumococcal Sequencing (GPS) project database ($n=19607$, collected from 49 countries across 5 continents, 1995–2018, accessed 17 March 2022). Nine per cent (934/10923) of isolates obtained from infants in the Drakenstein community in SA and 2%(419/19607) of genomes in the GPS dataset were serotype 16F. Serotype 16F isolates were from 28 different lineages of *S. pneumoniae*, with GPSC33 and GPSC46 having the highest proportion of serotype 16F isolates at 26% (346/1353) and 53% (716/1353), respectively. Serotype 16F isolates were identified globally, but most isolates were collected from Africa. GPSC33 was associated with carriage [OR (95% CI) 0.24 (0.09–0.66); $P=0.003$], while GPSC46 was associated with disease [OR (95% CI) 19.9 (2.56–906.50); $P=0.0004$]. Ten per cent (37/346) and 15% (53/346) of isolates within GPSC33 had genes associated with resistance to penicillin and co-trimoxazole, respectively, and 18% (128/716) of isolates within GPSC46 had genes associated with resistance to co-trimoxazole. Resistant isolates formed genetic clusters, which may suggest emerging resistant lineages. Serotype 16F lineages were common in southern Africa. Some of these lineages were associated with disease and resistance to penicillin and cotrimoxazole. We recommend continuous genomic surveillance to determine the long-term impact of serotype 16F lineages on vaccine efficacy and antimicrobial therapy globally. Investing in vaccine strategies that offer protection over a wide range of serotypes/lineages remains essential. This paper contains data hosted by Microreact.

Received 25 August 2023; Accepted 09 October 2023; Published 02 November 2023

Author affiliations: ¹Parasites and Microbes, Wellcome Sanger Institute, Hinxton, UK; ²Department of Genetics, Evolution and Environment, University College London, London, UK; ³NIHR Mucosal Pathogens Research Unit, Research Department of Infection, Division of Infection and Immunity, University College London, London, UK; ⁴Malawi-Liverpool-Wellcome Research Programme, Blantyre, Malawi; ⁵Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases, a division of the National Health Laboratory Service, Johannesburg, South Africa; ⁶School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; ⁷Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and SA-MRC unit on Child and Adolescent Health, University of Cape Town, Cape Town, South Africa; ⁸Marshall Centre, School of Biomedical Sciences, University of Western Australia, School of Biomedical Sciences, Perth, ACT, Australia; ⁹Centers for Disease Control and Prevention, Atlanta, GA, USA; ¹⁰Milner Centre for Evolution, Life Sciences Department, University of Bath, Bath, UK; ¹¹Department of Molecular and Cell Biology and Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa; ¹²School of Medicine, University of Lusaka, Lusaka, Zambia.

*Correspondence: Jolynne Mokaya, jm61@sanger.ac.uk

Keywords: antimicrobial resistance; lineages; pneumococci; serotype; virulence.

Abbreviations: CDC, centers for disease control and prevention; DCHS, Drakenstein Child Health Study; EPI, Expanded Programme on Immunization; GPSCs, global pneumococcal sequence clusters; HREC, human research ethics committee; IPD, invasive pneumococcal disease; LRTI, lower respiratory tract infection; MLST, multilocus sequence typing; NP, nasopharyngeal; NVTs, non-vaccine serotypes; PCV, pneumococcal conjugate vaccines; SA, South Africa.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Four supplementary tables are available with the online version of this article.

001123 © 2023 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Impact Statement

This study shows that serotype 16F lineages are predominant in southern Africa and are associated with disease and antimicrobial resistance. Although serotype 16F has been included in the newer formulation of the upcoming vaccine formulations of PCV21 and IVT-25, continuous surveillance to determine the long-term impact of serotype 16F lineages on vaccines and antimicrobial therapy remains essential.

DATA SUMMARY

The sequencing reads for the genomes analysed have been deposited in the European Nucleotide Archive and the accession numbers for each isolate are listed in Table S2, available in the online version of this article. A phylogenetic tree of serotype 16F pneumococcal genomes and associated metadata are available for download and visualization on the Microreact website: phylogenies of serotype 16F, GPSC33 and GPSC46 are available at serotype-16F, GPSC33 and GPSC46, respectively. The metadata are available as Table S3_(GPSC33 and GPSC46).

INTRODUCTION

Childhood morbidity and mortality caused by *Streptococcus pneumoniae* remain a major global health challenge despite advances in antimicrobial treatment and vaccine development. In 2015, 294 000 HIV-uninfected infants and 23 300 HIV-infected infants were estimated to have died from pneumococcal disease globally [1]. Pneumococcal conjugate vaccines (PCV) are the current recommended formulation for children and they target different serotypes that cause most invasive pneumococcal disease (IPD) [2]. However, pneumococcal disease persists, in part due to the increase of disease caused by non-vaccine serotypes, a phenomenon called serotype replacement [3]. In serotype replacement, the reduction of vaccine serotypes by PCV or/and antimicrobials creates room for the expansion of non-vaccine serotypes [3, 4]. These non-vaccine serotypes are also sometimes resistant to antimicrobial treatments, leading to an increased risk to public health [2].

The capsule of *S. pneumoniae* has been classified into 104 serotypes based on the reaction of a set of antisera against the capsular antigen [5]. While serotype classification helps to identify virulent strains and subsequently inform vaccine development, it provides little information about the genetic shifts in *S. pneumoniae* strains, given that the *cps* locus (which encodes for the capsular polysaccharides) only accounts for 0.2% of the genome [6]. Multilocus sequence typing (MLST) was previously the gold standard for characterizing bacterial isolates, based on the genetic sequences of seven housekeeping genes, however, recombination in some of these genes and limited resolution inhibits its utility to infer relationships between strains [7, 8]. Whole-genome analysis offers the opportunity to classify strains that share an evolutionary history into lineages, allowing for inference of relationships between strains across a species [7]. Furthermore, housekeeping genes represent a very small proportion of the genome, and are not necessarily representative of the relationship across the rest of the genome due to recombination within these genes [7]. The Global Pneumococcal Sequencing (GPS) project, whose aim is to provide an international understanding of *S. pneumoniae* population structure and its impact on vaccine and treatment strategies, has therefore classified *S. pneumoniae* strains into >900 lineages [also known as global pneumococcal sequencing clusters (GPSCs)], and each of these lineages can contain a single or multiple serotypes [4, 7, 9].

In South Africa, where PCV13 is part of the expanded programme on immunization (EPI), serotype 16F has been reported to be among the predominant non-vaccine serotypes (NVTs) among fully vaccinated infants [10] and has been shown to contribute to invasive disease among children below the age of 3 years after the introduction of PCV13 [4]. Using whole-genome sequencing, we characterized the major *S. pneumoniae* lineages containing serotype 16F in isolates from children in the Western Cape of South Africa and assess the public health relevance of serotype 16F lineages by describing their association with disease and antimicrobial susceptibility.

METHODS**Study setting**

We analysed 10923 *S. pneumoniae* carriage isolates obtained from 1020 of 1143 infants enrolled from May 2012–September 2015 as part of a population-based, longitudinal prospective birth cohort study [the Drakenstein Child Health Study (DCHS)] in the Western Cape Province in South Africa [11]. Both the parent study (401/2009) and this study (188/2017) received ethical approval from the Faculty of Health Sciences, Human Research Ethics Committee (HREC) of the University of Cape Town, South Africa. This study took place at two primary healthcare clinics located 2 km apart, i.e. TC Newman and Mbekweni. All infants received routine immunization including PCV13 as part of the national immunization programme, administered at 6 weeks, 14 weeks and 9 months of age through a 2+1 vaccine schedule. Nasopharyngeal (NP) swab collection was performed every 2 weeks for the first year of life as well as at 6, 12, 18 and 24 months, and whenever infants presented with pneumonia or lower respiratory tract infection (LRTI). Details on sample collection, transportation, culture and storage have been described previously [10]. For global context, we included *S. pneumoniae*

isolates of serotype 16F and lineages of interest included in the GPS project database ($n=19\,607$) from 49 countries across 5 continents, 1995–2018 (last accessed on 17 March 2022) [12].

Definition of carriage and disease and vaccine and non-vaccine serotypes

We define carriage isolates as those collected from healthy individuals and disease isolates as those from sputum or other specimen (i.e. blood, joint fluid, aspirates, etc.) from individuals with disease (including septicaemia, bacteraemia, pneumonia, cellulitis, meningitis, otitis media, bronchitis, osteomyelitis, septic arthritis, sinusitis, empyema, abscess, surgical site infection, sepsis, encephalitis, lower respiratory tract infection, conjunctivitis, bronchitis, peritonitis, pericarditis). In this paper, the vaccine serotypes are those included in PCV13, PCV15 and/or PCV20 (Table S1).

Sequencing and bioinformatics analysis

Single-colony picks of presumptive *S. pneumoniae* were inoculated onto Columbia blood agar base with 2% agar, 5% horse blood (BA) plates and incubated at 37°C in 5% CO₂ overnight. DNA was extracted and quantified as described previously [10]. DNA was sequenced on an Illumina HiSeq platform at Wellcome Sanger Institute, generating ≥ 100 bp paired-end reads. The reads were assessed for quality, assembled, annotated and mapped as previously described [13]. We inferred serotype using seroBA v1.0.2 and resistance profiles for amoxicillin, cefoxitin, ceftazidime, penicillin, chloramphenicol, clindamycin, cotrimoxazole, doxycycline, erythromycin, levofloxacin, meropenem, rifampicin, tetracycline and vancomycin from the genomic data using the Centers for Disease Control and Prevention (CDC) antimicrobial detection tool [14]. Multidrug resistance was defined as predicted resistance to ≥ 3 antimicrobial classes. PopPUNK v1.1.6 was used to assign GPSC to the genomes [15]. We mapped reads to GPSC-specific reference [GPSC 46 (accession number ERS628712) and GPSC33 accession number ERS566825)] sequences using BWA v 0.7.17 to create an alignment and then assessed for recombination within each GPSC using Gubbins v2.4.1. Phylogenetic trees for recombination-stripped alignments for each GPSC were generated using RAxML v 8.2.8. To investigate the emergence of serotype 16F lineages, we generated time trees using BactDating with a mixed gamma, relaxed clock model and visualized these trees using FigTree v1.4.4 [16]. Using Gubbins output, we calculated and compared recombination rates of different lineages as follows: sum of the recombination base substitutions across all branches divided by the sum of the point mutations across all branches.

Descriptive and statistical analysis

We described the proportion of isolates which were serotype 16F using the Drakenstein and GPS datasets. We reported the total number of isolates and serotype 16F isolates within each lineage. We determined the association between serotype 16F lineages and carriage or disease status. This analysis was limited to GPS data from South Africa to control for data collection practices, which varied amongst locations. We restricted data to isolates from children <7 years as pneumococcal disease is most common in this population. For each lineage, we compared the proportions of pneumococcal disease and carriage isolates to those of other lineages, reporting an odds ratio with 95% confidence interval (CI) by Fisher's exact test. We calculated the frequency of antimicrobial resistance (AMR) for different classes of antimicrobials in each GPSC as follows: (total number of isolates with predicted antimicrobial resistance within a specific lineage/total number of isolates within that lineage) multiplied by 100.

RESULTS

The characteristics of isolates obtained from infants recruited from the Drakenstein community and isolates obtained from the GPS project are summarized in Table 1.

Table 1. Characteristics of *Streptococcus pneumoniae* obtained from the Drakenstein community and GPS dataset

	n (%)	
	Drakenstein community	GPS
Total no. of isolates	10923 (100)	19607 (100)
Clinical manifestation		
Carriage	10626 (97.3)	6345 (32.4)
Disease*	297 (2.7)	12435 (63.4)
Other†	–	827 (4.2)

*Isolates obtained from sputum collected at the time of lower respiratory tract infection from children participating in the Drakenstein Child Health Study were classified as disease isolates. (–) indicates no data.

†This includes isolates whose disease or carriage status is unknown in the Global Pneumococcal Sequence (GPS) database.

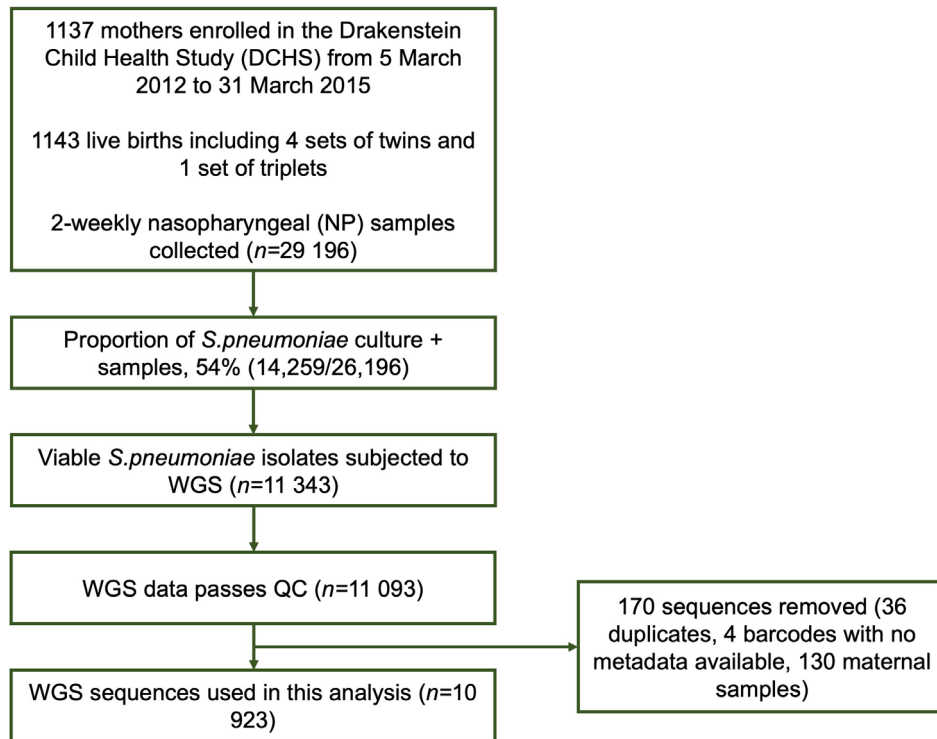


Fig. 1. Flow chart showing how samples and sequences were included from the Drakenstein community.

Proportion of serotype 16F

A total of 10923 *S. pneumoniae* genomes, comprising 61 serotypes, were isolated from nasopharyngeal swabs taken from the 1020 infants enrolled in the DCHS (Fig. 1). Serotype 16F was the most common serotype, accounting for 9% (934/10923) of isolates (Fig. 2a). In the GPS dataset, there were 19607 *S. pneumoniae* genomes, comprising 98 serotypes and serotype 16F had a prevalence of 2% (419/19607) (Fig. 2b).

Global distribution and lineages associated with serotype 16F

We considered the overall geographical distribution of 1353 serotype 16F isolates from the DK and GPS datasets. Serotype 16F was identified in six continents, with the majority of 16F isolates collected from Africa [92% (1248/1353)]. The proportion of serotype 16F isolates from other continents were as follows: 2% (59/1353) were from Asia, 1.8% (25/1353) were from North America, 1.4% (19/1353) were from South America and 0.15% (2/1353) were from Europe (Fig. 3). Serotype 16F was present in 28 distinct lineages, with GPSC33 and GPSC46 having the highest proportion of 16F strains, at 26% (346/1353) and 53% (716/1353), respectively (Fig. 3). Geographical structure was observed within the African continent, with GPSC103 and GPSC274 more commonly detected in West Africa, GPSC46, GPSC47 and GPSC207 in southern Africa, GPSC268 in East Africa, and GPSC33 and GPSC114 on the south and east coasts of Africa. In North America, the predominant lineages were GPSC135 and GPSC165. GPSC18, GPSC156 and GPSC104 were predominant in Latin America, Europe and Asia, respectively.

Antimicrobial resistance across serotype 16F lineages

We assessed the presence of AMR genes in serotype 16F lineages (Fig. 4). Ten per cent (37/346) or 15% (53/346) of isolates within GPSC33 had genes associated with resistance to penicillin or co-trimoxazole, respectively, while 18% (128/716) of isolates within GPSC46 had genes associated with resistance to co-trimoxazole. All isolates ($n=30$) within the GPSC103 lineage had genes associated with resistance to cotrimoxazole, tetracycline and doxycycline.

Lineage analysis of GPSC33 and GPSC46

We explored the global phylogeny of GPSC33 and GPSC46 given that they are the predominant lineages of serotype 16F (Fig. 5). GPSC33 lineage consists almost exclusively (99%, 346/350) of serotype 16F. All isolates within this lineage were from southern Africa [South Africa: 88% (307/350)] and Malawi [12% (43/350)]. Similarly, serotype 16F predominates [99%, (716/723)] in the GPSC46 lineage. Most of isolates within this lineage were from South Africa [98% (711/723)], with few isolates collected from elsewhere [Belarus

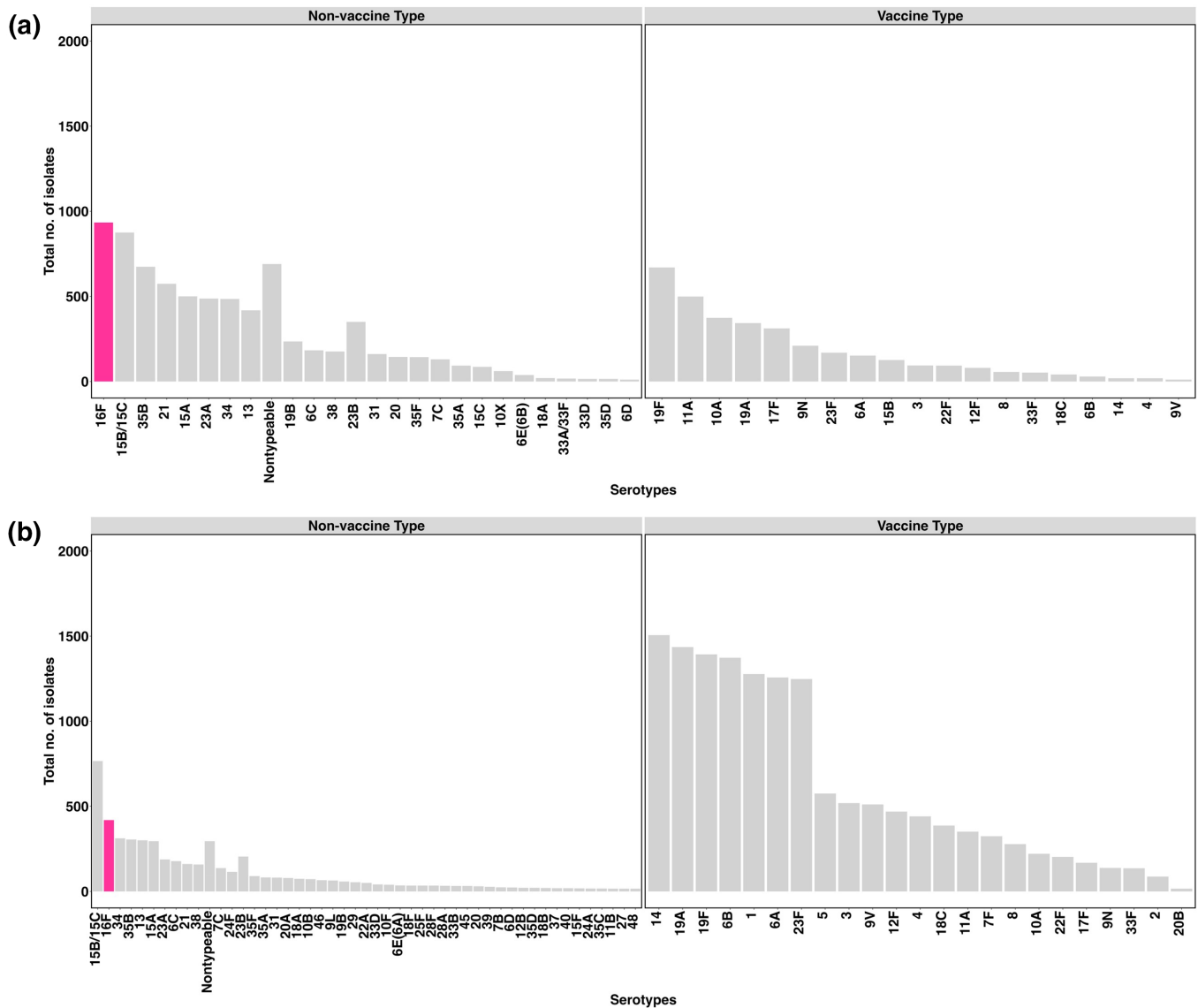


Fig. 2. Distribution of *Streptococcus pneumoniae* serotypes within the (a) Drakenstein community and (b) GPS datasets. Bars indicate the number of isolates of each serotype. The pink bar represents serotype 16F and other serotypes are shown with grey bars. Serotypes are classified as non-vaccine type [not included in pneumococcal vaccines (PCV)] and vaccine type (included in PCV13, PCV15 and/or PCV20, i.e. 1, 2, 3, 4, 5, 6A/B, 7F, 8, 9N/V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A/F, 20B, 22F, 23F, 33F). Serotypes with fewer than 10 isolates are not shown in these bar plots.

($n=2$), Israel ($n=3$), Papua New Guinea ($n=1$), Russia ($n=3$) and the USA ($n=3$)). In GPSC33 there were two distinct clusters of isolates with penicillin resistance determinants and one cluster of isolates with cotrimoxazole resistance determinants; and in GPSC46, there was one cluster of isolates with cotrimoxazole resistance determinants (Fig. 4).

Compared to GPSC46, the time to the most recent common ancestor (tMRCA) for GPSC33 is recent; serotype 16F isolates within GPSC33 likely originated in 1908 (95% HPD 1898–1917), compared with 1896 (95% HPD 1795–1926) for GPSC46. For isolates with resistance that formed clusters – which suggests emerging resistant lineages – the estimated tMRCA for the cluster with penicillin resistance in GPSC33 is 1937 (95% HPD 1927–1946) and that for the cluster with cotrimoxazole resistance in GPSC46 is 1922 (95% HPD 1821–1988). The tMRCA of lineages is between the 1920s and the 1930s, coinciding with the discovery of penicillin and cotrimoxazole, but prior to their widespread use [17, 18]. The tMRCA confidence intervals for the cluster with resistance to cotrimoxazole are wide, and antimicrobial use may have contributed to the emergence of this cluster. We compared the genetic variation through recombination (a process in which exogenous DNA is acquired and incorporated into the genome [19]) of GPSC33 and GPSC46. GPSC33 had a higher recombination ratio (i.e. the ratio of the number of recombination events to point mutations on a branch) at 8.2 compared to 4.9 for GPSC46, which could be why there are more clusters with AMR in this lineage.

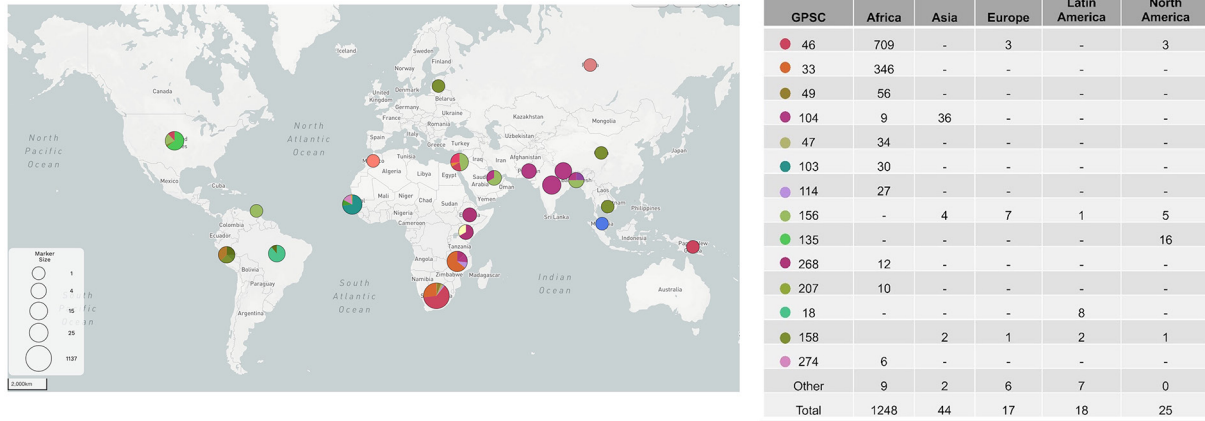


Fig. 3. Global distribution of serotype 16F lineages. The pie charts show the proportions of serotype 16F lineages in different countries. GPSC, global pneumococcal sequencing clusters. Available on the Microreact serotype-16F . Table on the right summarizes the number of serotype 16F isolates within each lineage across different continents. GPSCs with fewer than five isolates (GPSC80, 313, 348, 386, 428, 434, 477, 485, 505, 542, 568, 583, 676, 846) are not shown on the table. One pneumococcal isolate from Oceania belonging to GPSC46 is not shown on the table.

We compared the association of serotype 16F isolates to other non-serotype 16F isolates within GPSC33 and GPSC46 to carriage and disease. This analysis included GPS project isolates from children <7 years old from South Africa. Isolates from the Drakenstein study were excluded due to the focus on carriage sampling. Compared to other serotypes within GPSC33 and GPSC46 lineages, serotype 16F isolates were associated with carriage in GPSC33 [OR (95% CI) 0.24 (0.09–0.66); *P*=0.003], and with disease in GPSC46 [OR (95% CI) 19.9 (2.56–906.50); *P*=0.0004] (Table 2).

DISCUSSION

Serotype replacement from non-vaccine serotypes continues to threaten the effectiveness of current pneumococcal vaccine interventions [11–15]. However, pneumococcal surveillance, including the use of genomics, has been able to inform vaccine strategies [16]. We show that serotype 16F remains a dominant non-vaccine type in both carriage and invasive disease in South Africa. We further contextualized our findings in a global context using the GPS Project Database [7], and we highlight that serotype 16F is prevalent in Africa. Serotype 16F lineages, that include GPSC33 and GPSC46, have been highly successful in the region and are associated with antimicrobial resistance and pneumococcal disease. Together, these data highlight that non-vaccine serotype 16F has become

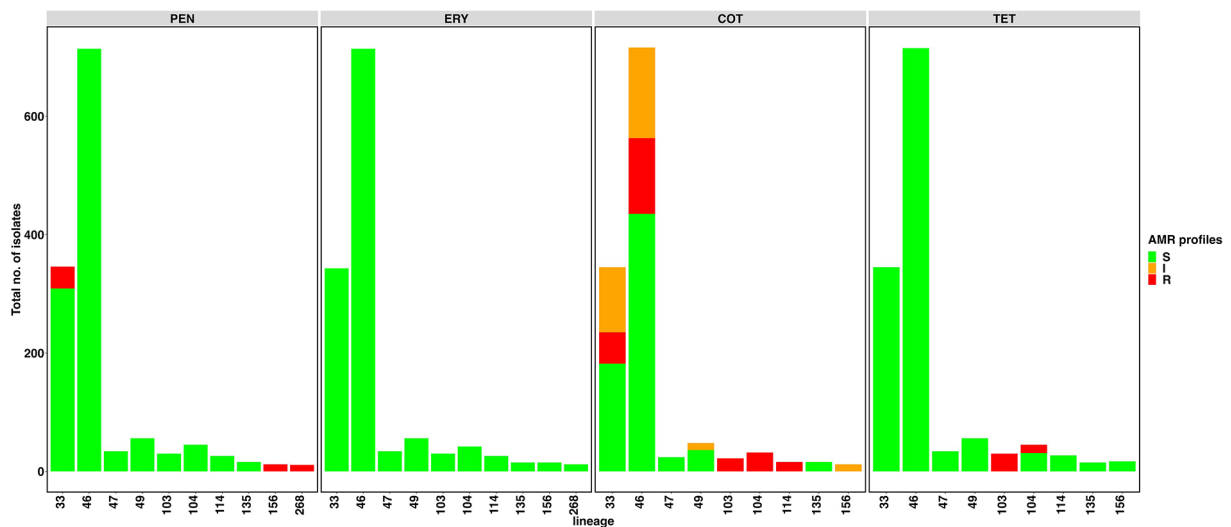


Fig. 4. Antimicrobial resistance profiles of serotype 16F lineages. PEN, penicillin; COT, cotrimoxazole; TET, tetracycline; DOX, doxycycline. S, susceptible; I, intermediate resistance; R, resistant.

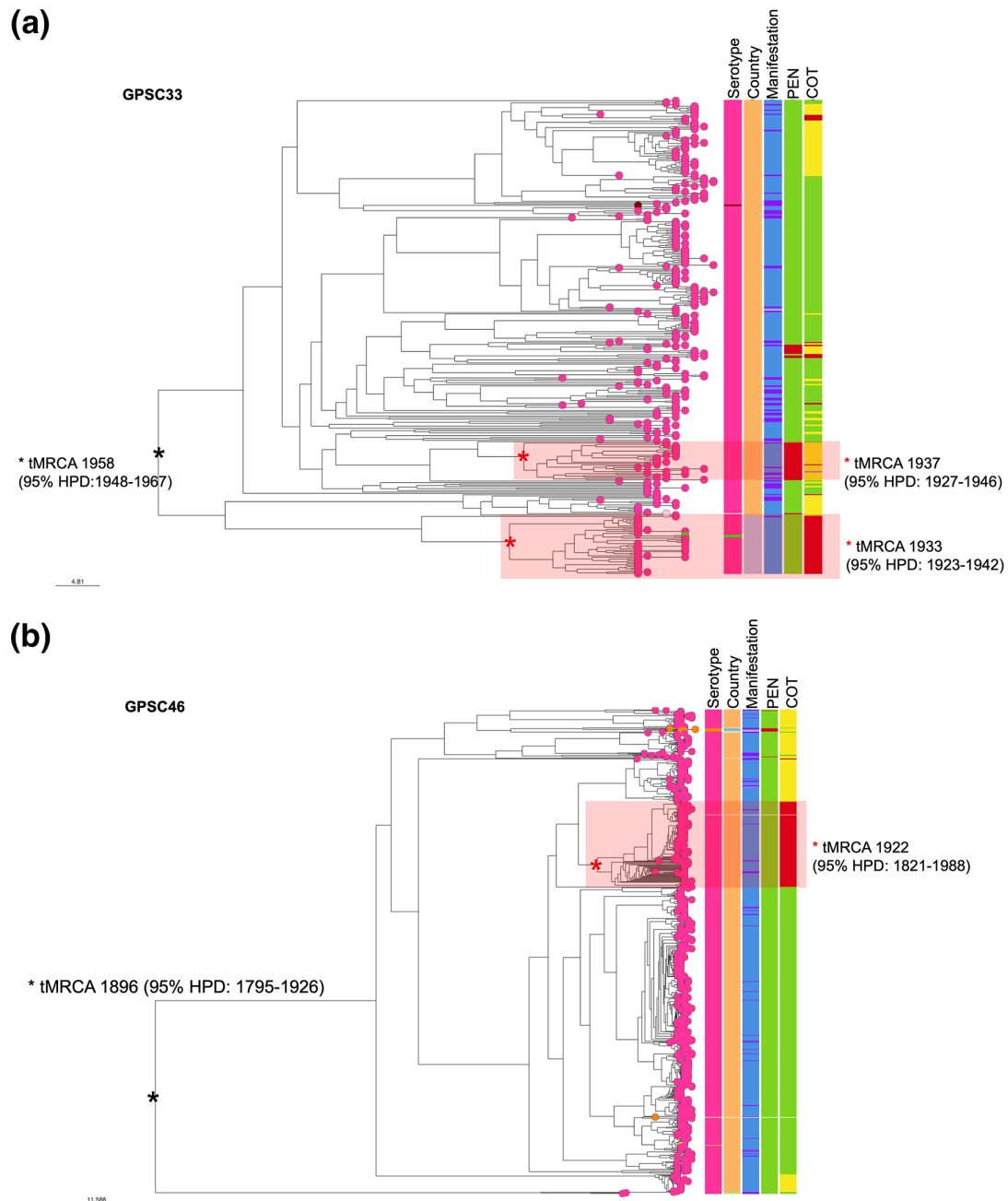


Fig. 5. Global phylogeny of (a) GPSC33 and (b) GPSC46. Serotype within each lineage is shown on the first column on the right of the tree. Countries where the isolates within each lineage were obtained are shown on the second column on the right of the tree. Clinical manifestations of individuals from whom isolates were obtained are shown on the third column. AMR profiles of PEN (penicillin) and COT (cotrimoxazole) are shown on the fourth and fifth column on the right of the tree, respectively. I, isolates with intermediate resistance; R, isolates with resistance; S, isolates that are susceptible to antimicrobials. tMRCA, time to the most recent common ancestor. The tMRCA is reported in years with 95% highest posterior density (HPD). The pink asterisk (*) corresponds to the ancestral node for isolates with antimicrobial resistance. The black asterisk (*) represents the root/ancestral node of entire phylogenetic tree. Link to access the GPSC33 phylogenetic tree is: GPSC33. Link to access the GPSC46 phylogenetic tree is: GPSC46.

increasingly important in South Africa, and therefore further surveillance is warranted to inform vaccine policy and potentially expand vaccine valency.

We show that serotype 16F was the dominant serotype in our longitudinal carriage cohort, similar to other pneumococcal carriage surveillance findings from South Africa [17]. Previous work in the region described significant decreases in PCV7 serotypes (19F, 6B, 23F and 14) and an increase in non-vaccine serotypes (16F, 34, 35B and 11A) among children <2 years of age colonized with the

Table 2. Comparing carriage and disease between GPSC33 and GPSC46

	Carriage	Disease	OR (95% CI)
GPSC33_16F	56 (73%)	21 (27%)	0.24 (0.09–0.66) P=0.003
non-GPSC33_16F	11 (39%)	17 (61%)	
GPSC46_16F	1 (10%)	9 (90%)	19.9 (2.56–906.50) P=0.0004
non-GPSC46_16F	66 (69%)	29 (31%)	

pneumococcus [17]. Other regions, including West Africa [18] and the Middle East [19], have described the predominance of serotype 16F in carriage and disease among children and adults following pneumococcal vaccine rollout. Here we show that lineage GPSC33 was one of the dominant lineages in our cohort, and the lineage has been described as mostly prevalent on the African continent [18]. Longitudinal colonization studies have shown a typically high number of single-nucleotide polymorphisms among 16F isolates of the same GPSCs carried over multiple sampling visits, suggesting that divergent 16F strains may emerge over the course of carriage due to homologous recombination [20]. Further work is required to understand the genotypic and phenotypic traits of successful lineages such as GPSC33.

Antimicrobial selection and horizontal gene transfer could potentially facilitate the expansion of resistant serotype 16F lineages in South Africa. We identified cotrimoxazole resistance sub-lineages in both GPSC33 and GPSC46. Although cotrimoxazole is not directly used to treat diseases caused by *S. pneumoniae*, it is widely used as a prophylaxis for infants who are HIV-exposed but uninfected [21], and adults living with HIV. Antimicrobial pressure due to cotrimoxazole use has therefore been suggested to contribute to the resistance patterns observed in the region, and therefore further use of the antimicrobial may facilitate the expansion of cotrimoxazole-resistant GPSC33 and GPSC46 sub-lineages.

Although we established that GPSC33 and GPSC46 lineages were generally susceptible to first-line antimicrobials used to treat pneumonia, such as penicillin, we identified penicillin-resistant GPSC33 sub-lineages, highlighting the potential risk of this lineage expanding and limiting antimicrobial treatment options in South Africa. Penicillin-resistant serotype 16F have been described in countries such as Japan following post-PCV7 rollout [22]. Previous work on horizontal gene transfer has shown that serotypes that frequently colonize the human nasopharynx, including serotype 16F, acquire penicillin-binding protein gene fragments from *Streptococcus mitis* [23]. The authors highlighted the presence of mosaic *pbp2x* among serotype 16F GPSC33 associated with reduced susceptibility [23], and modifications at the *pbp2x* gene are known to confer reduced susceptibility to a range of beta-lactam antimicrobials [24]. Due to the high rates of recombination seen among 16F isolates in longitudinal carriage and the acquisition of mosaic *pbp* fragments from other species [23], we hypothesize that serotype 16F may develop higher levels of beta-lactam resistance with subsequent recombination in other important *pbp* genes (*pbp2b* and *pbp1a*).

Serotype 16F lineages are an important cause of invasive disease in the post-PCV13 era, particularly in countries such as South Africa. Our results show that not only is serotype 16F the predominant non-vaccine serotype carried among vaccinated infants, but serotype 16F lineages that include GPSC33 and GPSC46 cause invasive disease. In South Africa, serotype 16F has been associated with the second highest case fatality ratio after serotype 6A [25]. Serotype 16F is an important cause of invasive disease in countries such as Ethiopia [26] and Denmark [27], and the serotype has increasingly become important among cases of meningitis through serotype replacement in Israel [19]. Previous meta-analyses have shown increased case fatality rates associated with serotype 16F, along with serotypes 3, 6B, 9N, 11A, 19F and 19A [28]. Furthermore, seven studies found an increase in serious clinical outcomes attributed to serotype 16F, which is not included in the PCV13 formulation [28]. It remains unclear why serotype 16F is associated with mortality, but infection among vulnerable populations may be of concern. There is a predominance of non-vaccine serotypes of the pneumococcus in carriage among the HIV-infected children in Ghana, particularly serotype 16F [29]. Further genomic surveillance will be important to track serotypes and lineages that are associated with carriage, disease and elevated risk of serious outcomes for vaccine policy making and potentially expanding vaccine valency to include 16F.

Serotype 16F is not included in the current PCV formulations (PCV10/13/15/20) approved for use in children but is included in upcoming formulations of PCV21 and IVT-25. Therefore, continuous surveillance to determine the long-term impact of serotype 16F lineages on vaccines and antimicrobial therapy remains essential.

Funding information

This study was co-funded by the Bill and Melinda Gates Foundation (grant code OPP1034556), the Wellcome Sanger Institute (core Wellcome grants 098051 and 206194) and the US Centers for Disease Control and Prevention. This work has also been supported by H3Africa funded through the Office of Strategic Coordination/Office of the NIH Director, National Institute of Environmental Health Sciences and National Human Genome Institute of Health of the National Institutes of Health under award numbers U54HG009824 and 1U01HG006961. F.D. is supported by National Research Foundation of South Africa grants (SRUG2204224295 and SNSF22071239126).

Author contributions

Conceived the study: J.M., S.W.L., F.D. Assimilated data: J.M., S.W.L., F.D. Analysed the data: J.M. Wrote the manuscript: J.M. Revised the manuscript: J.M., A.K., M.P.N., S.W.L., F.D. All authors have read and approved the manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, 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–e757.
2. Balsells E, Guillot L, Nair H, Kyaw MH. Serotype distribution of *Streptococcus pneumoniae* causing invasive disease in children in the post-PCV era: a systematic review and meta-analysis. *PLoS One* 2017;12:e0177113.
3. Hanage WP. Serotype replacement in invasive pneumococcal disease: where do we go from here? *J Infect Dis* 2007;196:1282–1284.
4. Lo SW, Gladstone RA, van Tonder AJ, Lees JA, du Plessis M, et al. Pneumococcal lineages associated with serotype replacement and antibiotic resistance in childhood invasive pneumococcal disease in the post-PCV13 era: an international whole-genome sequencing study. *Lancet Infect Dis* 2019;19:759–769.
5. Brooks LRK, Mias GI. *Streptococcus pneumoniae*'s virulence and host immunity: aging, diagnostics, and prevention. *Front Immunol* 2018;9:1366.
6. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitsch E, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2006;2:e31.
7. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine* 2019;43:338–346.
8. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, et al. Multi-locus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 1998;95:3140–3145.
9. GPS:: Global Pneumococcal Sequencing Project; (n.d.). https://www.pneumogen.net/gps/project_outline.html [accessed 6 June 2022].
10. Dube FS, Ramjith J, Gardner-Lubbe S, Nduru P, Robberts FJL, et al. Longitudinal characterization of nasopharyngeal colonization with *Streptococcus pneumoniae* in a South African birth cohort post 13-valent pneumococcal conjugate vaccine implementation. *Sci Rep* 2018;8:1–9.
11. Bar-Zeev N, Swarthout TD, Everett DB, Alaerts M, Msefula J, et al. Impact and effectiveness of 13-valent pneumococcal conjugate vaccine on population incidence of vaccine and non-vaccine serotype invasive pneumococcal disease in Blantyre, Malawi, 2006–18: prospective observational time-series and case-control studies. *Lancet Glob Health* 2021;9:e989–e998.
12. Savulescu C, Krizova P, Lepoutre A, Mereckiene J, Vestrheim DF, et al. Effect of high-valency pneumococcal conjugate vaccines on invasive pneumococcal disease in children in SplDnet countries: an observational multicentre study. *Lancet Respir Med* 2017;5:648–656.
13. Mackenzie GA, Hill PC, Jeffries DJ, Hossain I, Uchendu U, 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.
14. Ladhani SN, Collins S, Djennad A, Sheppard CL, Borrow R, 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.
15. Ben-Shimol S, Givon-Lavi N, Grisaru-Soen G, Megged O, Greenberg D, et al. Comparative incidence dynamics and serotypes of meningitis, bacteremic pneumonia and other-IPD in young children in the PCV era: insights from Israeli surveillance studies. *Vaccine* 2018;36:5477–5484.
16. Bentley SD, Lo SW. Global genomic pathogen surveillance to inform vaccine strategies: a decade-long expedition in pneumococcal genomics. *Genome Med* 2021;13:84.
17. Javaid N, Olwagen C, Nzenze S, Hawkins P, Gladstone R, et al. Population genomics of pneumococcal carriage in South Africa following the introduction of the 13-valent pneumococcal conjugate vaccine (PCV13) immunization. *Microb Genom* 2022;8:mgen000831.
18. Dayie N, Tettey EY, Newman MJ, Bannerman E, Donkor ES, et al. Pneumococcal carriage among children under five in Accra, Ghana, five years after the introduction of pneumococcal conjugate vaccine. *BMC Pediatr* 2019;19:316.
19. Phillips MT, Warren JL, Givon-Lavi N, Tothpal A, Regev-Yochay G, et al. Evaluating post-vaccine expansion patterns of pneumococcal serotypes. *Vaccine* 2020;38:7756–7763.
20. Chaguzo C, Senghore M, Bojang E, Gladstone RA, Lo SW, et al. Within-host microevolution of *Streptococcus pneumoniae* is rapid and adaptive during natural colonisation. *Nat Commun* 2020;11:3442.
21. Daniels B, Kuhn L, Spooner E, Mulol H, Goga A, et al. Cotrimoxazole guidelines for infants who are HIV-exposed but uninfected: a call for a public health and ethics approach to the evidence. *Lancet Glob Health* 2022;10:e1198–e1203.
22. Vidal JE, Ghanem EB, Wu X, Wu K, Bai G, et al. Transmission, colonization, and molecular pathogenesis of pneumococcus. *Frontiers Media SA* 2022:231.
23. Kalizang'oma A, Chaguzo C, Gori A, Davison C, Beleza S, et al. *Streptococcus pneumoniae* serotypes that frequently colonise the human nasopharynx are common recipients of penicillin-binding protein gene fragments from *Streptococcus mitis*. *Microb Genom* 2021;7:000622.
24. Dewé TCM, D'Aeth JC, Croucher NJ. Genomic epidemiology of penicillin-non-susceptible *Streptococcus pneumoniae*. *Microb Genom* 2019;5:e000305.
25. Müller A, Kleynhans J, de Gouveia L, Meiring S, Cohen C, et al. *Streptococcus pneumoniae* serotypes associated with death, South Africa, 2012–2018. *Emerg Infect Dis* 2022;28:166–179.
26. Rosenthal J-L, Wong RB. *Before and Beyond Divergence: The Politics of Economic Change in China and Europe*. Harvard University Press, 2011. p. 291.
27. Weinberger DM, Warren JL, Dalby T, Shapiro ED, Valentiner-Branth P, et al. Differences in the impact of pneumococcal serotype replacement in individuals with and without underlying medical conditions. *Clin Infect Dis* 2019;69:100–106.
28. Grabenstein JD, Musey LK. Differences in serious clinical outcomes of infection caused by specific pneumococcal serotypes among adults. *Vaccine* 2014;32:2399–2405.
29. Donkor ES, Annan JA, Badoe EV, Dayie N, Labi A-K, et al. Pneumococcal carriage among HIV infected children in Accra, Ghana. *BMC Infect Dis* 2017;17:133.