

Streptococcus pneumoniae serotype 33G: genetic, serological, and structural analysis of a new capsule type

Sam Manna,^{1,2,3} Joel P. Werren,⁴ Belinda D. Ortika,¹ Barbara Bellich,⁵ Casey L. Pell,¹ Elissavet Nikolaou,^{1,3} Ilche Gjuroski,⁶ Stephanie Lo,⁷ Jason Hinds,^{8,9} Odgerel Tundev,¹⁰ Eileen M. Dunne,¹¹ Bradford D. Gessner,¹¹ Stephen D. Bentley,⁷ Fiona M. Russell,^{1,2} E. Kim Mulholland,^{1,12} Tuya Mungun,¹⁰ Claire von Mollendorf,^{1,2} Paul V. Licciardi,^{1,2} Paola Cescutti,⁵ Neil Ravenscroft,¹³ Markus Hilty,⁴ Catherine Satzke^{1,2,3}

AUTHOR AFFILIATIONS See affiliation list on p. 15.

ABSTRACT *Streptococcus pneumoniae* (the pneumococcus) is a human pathogen responsible for a spectrum of diseases such as pneumonia, sepsis, and meningitis. The capsule is the major pneumococcal virulence factor and is encoded by the capsular polysaccharide (*cps*) locus, a recombination hotspot that has resulted in over 100 distinct capsular polysaccharide types (serotypes) identified to date. Recently, 33X (also known as 10X) was proposed as a putative novel serotype, but the capsule structure had not been elucidated. Here, we provide an in-depth investigation of 33X, demonstrating it is a new pneumococcal capsular serotype. In this study, we screened 12,850 nasopharyngeal swabs from both healthy children and pneumonia patients (adults and children) in Mongolia collected between 2015 and 2022. We identified 20 pneumococcal 33X isolates. Using whole genome sequencing, we found that the 33X *cps* locus is a chimera of genes from pneumococcal serogroups 35, 10, and 33, as well as other *Streptococcal* species. Serotyping of 33X pneumococci by the Quellung reaction revealed a unique serological profile, typing as both 10B and 33B. Competitive ELISAs confirmed that antibodies that were generated in mice directed against 33X were inhibited by 33X pneumococci but not 10B or 33B. Lastly, the elucidation of the 33X capsule structure revealed that the polysaccharide is distinct from other serotypes, consisting of an O-acetylated hexasaccharide repeat unit of $\rightarrow 5$ - β -Gal β -(1 \rightarrow 3)- β -Glc β -(1 \rightarrow 5)- β -Gal β 2Ac-(1 \rightarrow 3)- β -Gal β NAc-(1 \rightarrow 3)- α -Gal β -(1 \rightarrow 4)-Rib-ol-(5 \rightarrow P \rightarrow). Therefore, 33X meets the requisite genetic, serological, and biochemical criteria to be designated as a new serotype, which we have named 33G.

IMPORTANCE *Streptococcus pneumoniae* (the pneumococcus) is a bacterial pathogen with the greatest burden of disease in Asia and Africa. The pneumococcal capsular polysaccharide has biological relevance as a major virulence factor as well as public health importance as it is the target for currently licensed vaccines. These vaccines have limited valency, covering up to 23 of the >100 known capsular types (serotypes) with higher valency vaccines in development. Here, we have characterized a new pneumococcal serotype, which we have named 33G. We detected serotype 33G in nasopharyngeal swabs ($n = 20$) from children and adults hospitalized with pneumonia, as well as healthy children in Mongolia. We show that the genetic, serological, and biochemical properties of 33G differ from existing serotypes, satisfying the criteria to be designated as a new serotype. Future studies should focus on the geographical distribution of 33G and any changes in prevalence following vaccine introduction.

KEYWORDS *Streptococcus pneumoniae*, capsule, serotypes, recombination, vaccines

Editor Christopher N. LaRock, Emory University School of Medicine, Atlanta, Georgia, USA

Address correspondence to Sam Manna, sam.manna@mcri.edu.au.

Neil Ravenscroft, Markus Hilty, and Catherine Satzke contributed equally to this article. Order of names was decided randomly.

C.V.M., E.K.M., T.M., B.D.G., and C.S. are investigators on a clinical research collaboration with Pfizer on PCV vaccination in Mongolia from which the isolates used in this study are derived. Salary support was received through the institutions. C.S. and E.K.M. are investigators on a Merck Investigator Studies Program grant funded by MSD for a study unrelated to this work. S.M., C.V.M., and C.S. have received honoraria [Pfizer (C.S., S.M., and C.V.M.) and MSD (C.S.)] for presentations at symposia or attendance at expert advisory meetings unrelated to this study. M.H. is an investigator on a research grant funded by Pfizer for a study unrelated to this work. E.M.D. and B.D.G. are employed by Pfizer and may hold Pfizer stock or stock options.

Received 4 October 2023

Accepted 9 November 2023

Published 7 December 2023

Copyright © 2023 Manna et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Streptococcus pneumoniae (the pneumococcus) is a common member of the upper respiratory tract microbiota of children. The pneumococcus is also an important human pathogen responsible for a range of diseases such as otitis media, pneumonia, sepsis, and meningitis (1). The polysaccharide capsule is a major pneumococcal virulence determinant. The capsule promotes survival within the host by facilitating the evasion of phagocytosis, complement, and mucus entrapment (2). Pneumococci exhibit a high level of diversity in capsular polysaccharide structure, with over 100 capsular types (referred to as “serotypes”) identified to date (3). Serotyping is the basis for the classification of pneumococci. The pneumococcal capsule also has important public health implications as it is the antigenic target for licensed vaccines. Accurate serotyping data are essential in evaluating vaccine impact in carriage and disease, as well as for disease surveillance programs more broadly.

To date, the genomes of over 50,000 pneumococci have been sequenced (4, 5). Sequence analysis has led to the identification of genetic variants of existing serotypes (6–10) as well as new serotypes (3, 11–15). Recently, there has been identification of serotype variants (6, 16) and new serotypes (15) within serogroup 33. The putative new serotype 33X [also known as 10X (17)], relevant to this study, was identified in Thailand ($n = 5$ isolates) (18) and South Africa ($n = 2$ isolates) (16). Although this 33X has been reported previously, little is known about the 33X capsule including whether it is indeed a new serotype or a molecular variant of an existing serotype as the capsule structure has not been investigated. In this study, we identify 33X pneumococci in Mongolia and show the 33X capsule possesses genetic, serological, and structural properties that differ from known serotypes. We name this new serotype 33G.

RESULTS

Isolation of 33X pneumococci from Mongolia

In Mongolia, the 13-valent Pneumococcal Conjugate Vaccine (PCV13) was introduced into the pediatric immunization program in a 2 + 1 schedule (2, 4, and 9 months of age) in a phased manner across districts of the capital city, Ulaanbaatar, from 2016 to 2018. In 2016 and 2017, three of the districts (Songinokhairkhan, Sukhbaatar, and Bayanzurkh) had a catch-up campaign for children aged 3–23 months (two doses given 2 months apart), followed by nationwide introduction with high coverage (19, 20). Pneumococcal surveillance included the collection of nasopharyngeal swabs from healthy children (cross-sectional surveys in 2015, 2017, and 2022) (21) as well as children (2015–2021) and adults (2019–2022) hospitalized with pneumonia (20, 22, 23) as part of studies assessing the direct and indirect impact of PCV13. Nasopharyngeal swabs ($n = 12,850$) were screened for pneumococci using quantitative PCR (qPCR) targeting the *lytA* gene (24). DNA was extracted from culture-positive samples for molecular serotyping by DNA microarray (25). Overall, 4,800 (37.6%) of the swabs contained pneumococci, and 20 (0.4%) of these contained the putative novel serotype 33X (designated as “35A/10B-like” by DNA microarray). This serotype was detected from 2018 onward, 2 years after the phased introduction of PCV13 commenced (2018 = 3, 2019 = 6, 2020 = 7, 2021 = 0, and 2022 = 4 isolates; Table S1). The 20 swabs containing 33X pneumococci were from both healthy carriers and patients hospitalized with pneumonia, with no obvious patterns of age, district, or vaccination status (Table 1).

Genetic analysis of the 33X *cps* locus

Using these isolates from Mongolia, we investigated 33X in more detail. A pneumococcal 33X isolate was purified from each swab (Table 1), and whole-genome sequencing was conducted. We determined genetic lineage using pubMLST (26, 27) (for Multi-Locus Sequence Type) and popPUNK (28) (for Global Pneumococcal Sequencing Cluster). The 33X pneumococci belonged to two genetic lineages ST2754 (GPSC230, $n = 3$) and ST6318 (GPSC687, $n = 17$).

The *cps* locus of the 33X isolates from Mongolia has over 99.7% DNA sequence identity to the 33X *cps* sequences first reported by van Tonder et al. (18) as well as by

TABLE 1 Participants from which 33X pneumococci were isolated from nasopharyngeal swabs

Participant status	Participant age	No. PCV13 doses ^a	District	Year	Isolate name
Pneumonia	2 months	1	Songinokhairkhan	2018	PMP1486
Pneumonia	9 months	2	Songinokhairkhan	2018	PMP1488
Pneumonia	14 months	3	Songinokhairkhan	2018	PMP1489
Pneumonia	4 months	2	Chingeltei	2019	PMP1521
Pneumonia	2 months	0	Songinokhairkhan	2019	PMP1596
Pneumonia	21 months	2	Bayanzurkh	2019	PMP1597
Pneumonia	7 months	2	Bayanzurkh	2019	PMP1598
Pneumonia	28 years	0	Sukhbaatar	2019	PMP1609
Pneumonia	68 years	0	Sukhbaatar	2019	PMP1614
Pneumonia	11 months	3	Sukhbaatar	2020	PMP1592
Pneumonia	17 months	3	Bayanzurkh	2020	PMP1593
Pneumonia	34 months	2	Chingeltei	2020	PMP1594
Pneumonia	4 months	1	Chingeltei	2020	PMP1595
Pneumonia	34 years	0	Chingeltei	2020	PMP1612
Pneumonia	63 years	0	Songinokhairkhan	2020	PMP1613
Pneumonia	31 years	0	Bayanzurkh	2020	PMP1616
Healthy	5 weeks	0	Sukhbaatar	2022	PMP1619
Healthy	17 months	3	Songinokhairkhan	2022	PMP1620
Healthy	15 months	3	Sukhbaatar	2022	PMP1618
Healthy	16 months	3	Songinokhairkhan	2022	PMP1617

^aAdults were not eligible for PCV13 vaccination.

Mostowy et al. (17) where it was referred to as 10X. These *cps* sequences were previously identified in isolates from Thailand (ST5123 and GPSC293) (18) and South Africa (ST5178 and GPSC241) (16). We examined the 33X *cps* locus of the 20 isolates from Mongolia using DNA sequence alignment and genome annotation by RAST (29), finding they have the same mosaic structure as the previously described 33X *cps* locus, composed of genes with high DNA identity to sequences from serogroups 35 (*wzg*, *wzh*, *wzd*, *wze*, *wchA*, and *wciB*; Fig. 1A), 10 (partial *wzy* sequences, *wcrB*, *wcrC*, *wcrD*, and *wciF*), and 33 (*wzx*, *wciG*, and *glf*; Fig. 1B).

In addition to the genes described above, the 33X *cps* locus contains another four putative open reading frames. *Orf1* exhibits partial similarity to part of the *wcrG* and *wzx* genes from the serotype 10B reference *cps* locus (Genbank accession [CR931650](#)), possibly representing 10B remnant sequences following a recombination event at this site. Interestingly, *orf2*, *orf3*, and *orf4* do not show sequence similarity to any pneumococcal *cps* genes. Rather, they show similarity to a region in the *cps* locus from *Streptococcus oralis* subsp. *dentisani* strain Y5914-11 (Fig. 1C), a blood isolate from a patient with infective endocarditis from Denmark (31). Between *wciB* and *wcrB* are two sequences with similarity to part of the *wzy* gene, which encodes the polymerase from the serotype 10B reference *cps* locus. However, the deletions apparent in these sequences suggest these would not encode a functional enzyme. *Orf2* is predicted to encode a Wzy homolog, with the predicted amino acid sequence exhibiting high identity to O-antigen ligase family proteins from *S. oralis* (97.2%, Genbank accession [WP_204956359](#)) and *Streptococcus intermedius* (97.6%, Genbank accession [MBF1713530](#)). Among pneumococcal Wzy proteins, ORF2 exhibits the highest identity to Wzy from serotype 37 (41.7%, Genbank accession [CAI34388](#)). Lastly, ORF3 and ORF4 are partial amino acid sequences that match portions of the *S. oralis* acyltransferase (86.5% and 74.7%, to *S. oralis* acyltransferase Genbank accession [WP_084971822](#) for ORF3 and ORF4, respectively). However, the *orf3-orf4* region in the 33X *cps* locus contains multiple frameshift mutations, suggesting it is unlikely to encode a functional acyltransferase (Fig. S1).

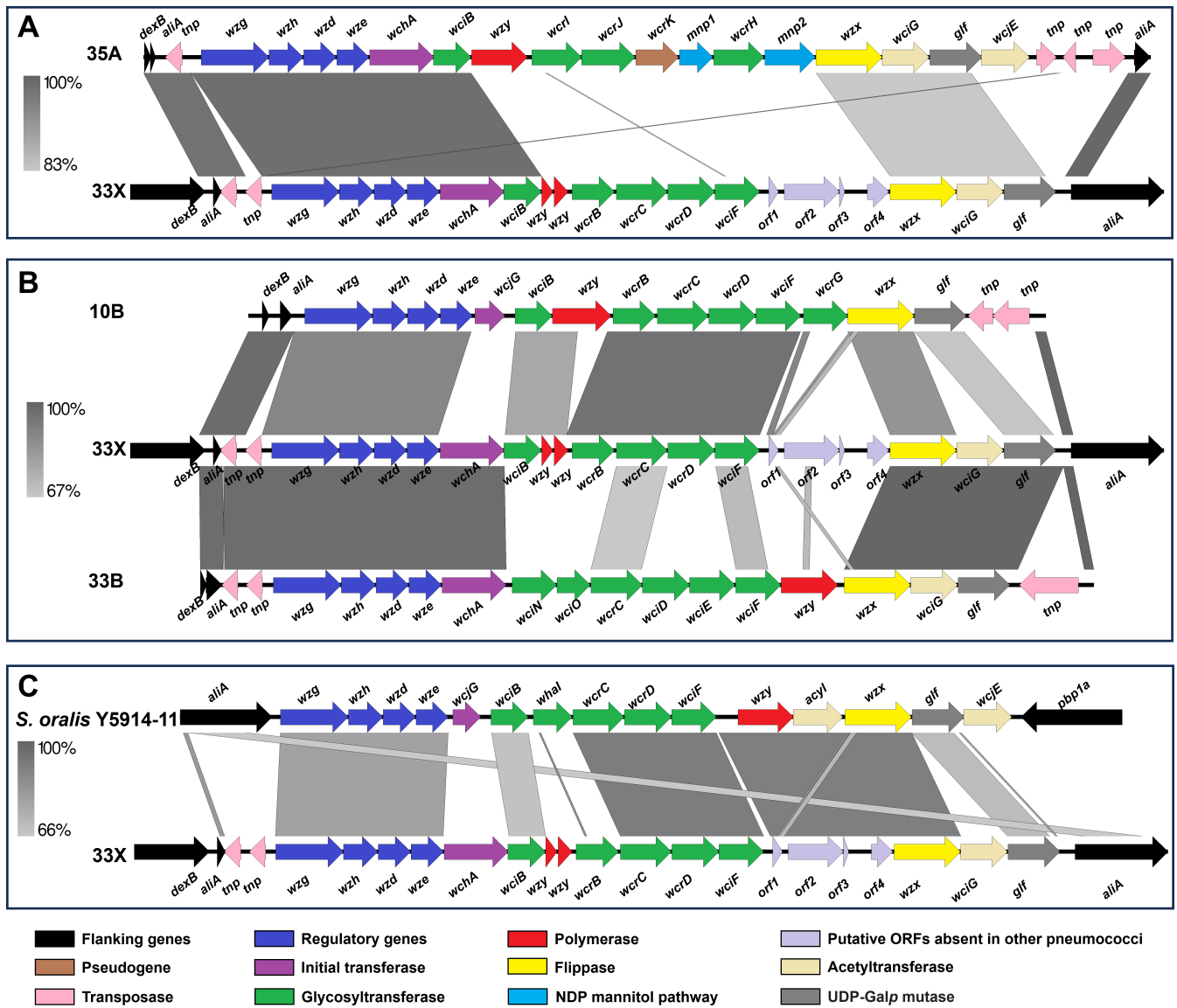


FIG 1 Schematic representation of the 33X *cps* locus with the 35A (A), 10B and 33B (B) *cps* reference sequences (30) and *Streptococcus oralis* subsp. *dentisani* strain Y5914-11 *cps* locus (Genbank accession no. [NZ_NCUW01000028](https://www.ncbi.nlm.nih.gov/nuccore/NZ_NCUW01000028)) (31). Images were generated using Easyfig version 2.2.5 (32).

To determine the implications of this genetic variation for molecular approaches for serotyping, we tested the 33X isolates in five different whole genome sequencing-based serotyping tools [PneumoCaT (33), seroBA (34), SeroCall (35), PneumoKITy (36), and PfaSTer (37)] as well as DNA microarray (as described above) and the TaqMan Array Card, a qPCR-based assay (38). Most tools were unable to designate a serotype to 33X pneumococci (Table S2). Tools that report the closest “hit” (PneumoCaT, PneumoKITy, and PfaSTer) were matches to serotypes in serogroup 35. As described earlier, DNA microarray designates 33X as 35A/10B-like due to the detection of *cps* genes from serotypes 35A and 10B. The TaqMan Array Card mistyped these isolates as 10B. The seroBA tool was the only one to detect 33X (designated as 10X in this tool) given the *cps* locus from the isolates previously identified in Thailand and South Africa is included in the seroBA database.

Serological properties of 33X pneumococci

To investigate the serological properties of the 33X capsule, the isolates from Mongolia were examined using the Quellung reaction, the gold standard for pneumococcal serotyping (39, 40) using commercially available typing sera from the Statens Serum Institut (SSI). All isolates reacted with both group 10 and 33, but not group 35, sera. When tested against the relevant factor sera, the 33X pneumococci reacted with 10b, 10d, and 33f factor sera. Thus, the 33X pneumococcal isolates serotyped as both 10B and 33B by Quellung, despite being confirmed as pure cultures by whole genome sequencing and the 10B and 33B control strains reacting as expected (Table 2).

To investigate the immunological properties of 33X isolates, we conducted competition ELISAs. To obtain sera, we infected 5-day-old C57BL/6 mice with a 33X pneumococcal strain (PMP1486) to establish nasopharyngeal colonization and collected serum 7 weeks post-infection. The serum was pre-incubated with pneumococci expressing serotype 33X, 10B, 33B, or 14 capsules as competitors before conducting the 33X ELISA. Serotype 14 was used as a negative control, as it was not expected to be recognized by capsular antibodies raised against other serotypes based on previous competition ELISA work (41). This allowed us to establish background inhibition levels in the assay. Pre-incubation of serum with 33X pneumococci resulted in greater inhibition (76.6%) compared with pre-incubating with 10B (36.4%), 33B (33.9%), or 14 (negative control, 28%), suggesting that the immunological properties of 33X differ to 10B and 33B (Fig. 2).

TABLE 2 Quellung reaction profile of 33X pneumococci (all strains) compared with 10B (PMP738) and 33B (PMP790) SSI strains with “+” and “–” denoting a positive and negative reaction with the typing sera, respectively

	Sera	33X	10B	33B
First pool	A	–	–	–
	B	–	–	–
	C	–	–	–
	D	–	–	–
	E	+	+	+
	F	–	–	–
	G	–	–	–
	H	–	–	–
	I	–	–	–
Second pool	P	–	–	–
	Q	–	–	–
	R	–	–	–
	S	+	+	–
	T	+	–	+
Group	10	+	+	–
	12	–	–	–
	21	–	–	–
	35	–	–	–
	33	+	–	+
	39	–	–	–
Factors – Group 10	10b	+	+	–
	10d	+ (weak) ^a	+	–
	10f	–	–	–
Factors – Group 33	33b	–	–	–
	33e	–	–	–
	33f	+	–	+
	6a	–	–	–
	20b	–	–	–

^aReaction is weak with one 1 µL loopful of 10d factor sera. A stronger reaction was obtained using two 1 µL loopfuls.

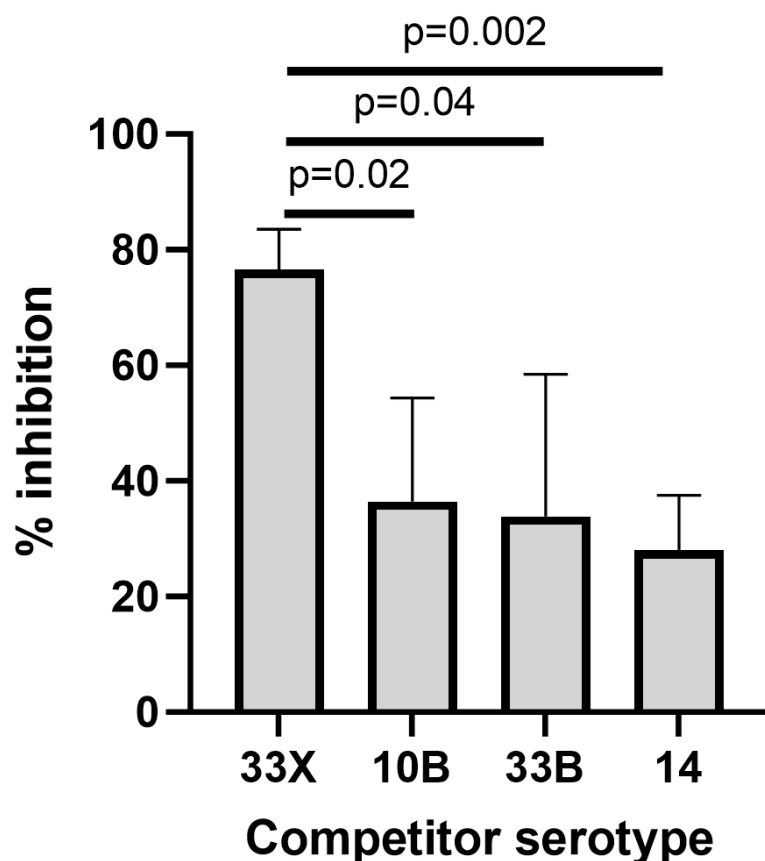


FIG 2 Competition ELISA with serum from mice colonized with 33X pneumococci pre-incubated with serotypes 33X, 10B, 33B, or 14 as a competing capsular antigen source. Data presented are mean + SD and analyzed by unpaired t test.

Taken together, these data show that 33X pneumococci have genetic and immunological properties that are distinct from previously characterized pneumococcal serotypes.

Elucidation of 33X capsule repeat unit structure

We next determined the repeat unit structure of the polysaccharide purified from 33X isolates by the use of chemical analysis and nuclear magnetic resonance (NMR) spectroscopy. Acid hydrolysis of the serotype 33X polysaccharide followed by derivatization and gas chromatography-mass spectrometric (GLC-MS) analysis showed the presence of ribitol (Rib-ol), glucose (Glc), galactose (Gal), and N-acetylgalactosamine (GalNAc) in the molar ratio of 1.6:1.0:2.2:0.8. Performing the reduction step with NaBD₄ and the absence of a deuterium label in the ribitol derivative confirmed that ribitol was not derived from ribose. Composition analysis following methanolysis gave comparable results and confirmed the absence of uronic acid in the repeat unit. The determination of the glycosidic linkages was achieved by GLC-MS analysis of the partially methylated alditol acetates which identified: 4,5-Rib-ol, 3-Glcp, 4-Galp or 5-Galf, 3-Galp, and 3-GalNAcp in the molar ratio of 0.4:1.6:2.1:1.0:0.7. These data are consistent with a linear hexasaccharide repeat unit without branch points and side chains, unlike the repeat units reported for related serotypes 10B (42) and 33B (43).

The ¹H NMR spectrum of the serotype 33X polysaccharide (Fig. 3A) at 25°C (500 MHz) showed five signals and an H-COAc peak at 4.93 ppm in the anomeric region, ring

protons, and peaks from low molecular weight components (glycerol, buffer, and amino acids).

The full spectrum (Fig. S2A) contains an O-acetyl signal at 2.14 ppm and an N-acetyl signal at 2.03 ppm indicating that serotype 33X polysaccharide has an O-acetylated hexasaccharide repeating unit containing an N-acetyl sugar. The peaks from HOD (water), process residuals, and amino acids were readily removed in the diffusion-ordered spectroscopy (DOSY) experiment in Fig. 3B (and Fig. S2B) to give the polysaccharide signals together with peaks from residual cell wall polysaccharide (CWPS) with the characteristic choline resonance at 3.23 ppm. Further NMR studies were performed at 600 MHz and at higher temperatures that resulted in gradual de-O-acetylation (Fig. 3C) leading to the NMR spectrum of the polysaccharide backbone shown in Fig. 3D (and Fig. S2D). The structure of the 33X hexasaccharide repeat unit was determined on the de-O-acetylated sample by the use of an array of ^1H - ^1H homonuclear experiments (COSY, TOCSY, and NOESY) and ^1H - ^{13}C heteronuclear correlation experiments (HSQC and HMBC) described in Fig. S3 to S5. The ^1H - ^{31}P HMBC experiment (Fig. S6) showed that the phosphodiester signal at 0.34 ppm was correlated to the H5 of ribitol and the broad proton signal at 4.35 ppm, assigned to H5 of β -Gal $^{\text{I}}$ $^{\text{f}}$. The proton/carbon pairs are labeled in the HSQC spectrum (Fig. 4), the NMR data are collected in Table 3.

Finally, the linkages and sequence of sugar residues followed from the ^1H - ^{13}C HMBC inter-residue correlations are labeled in Fig. 5. Thus, NMR analysis established the structure of the hexasaccharide repeat unit of serotype 33X as $\rightarrow 5$ - β -Gal $^{\text{I}}$ $^{\text{f}}$ -(1 \rightarrow 3)- β -Glc $^{\text{p}}$ -(1 \rightarrow 5)- β -Gal $^{\text{I}}$ $^{\text{f}}$ -(1 \rightarrow 3)- β -Gal $^{\text{p}}$ NAc-(1 \rightarrow 3)- α -Gal $^{\text{p}}$ -(1 \rightarrow 4)-Rib-ol-(5 \rightarrow P \rightarrow).

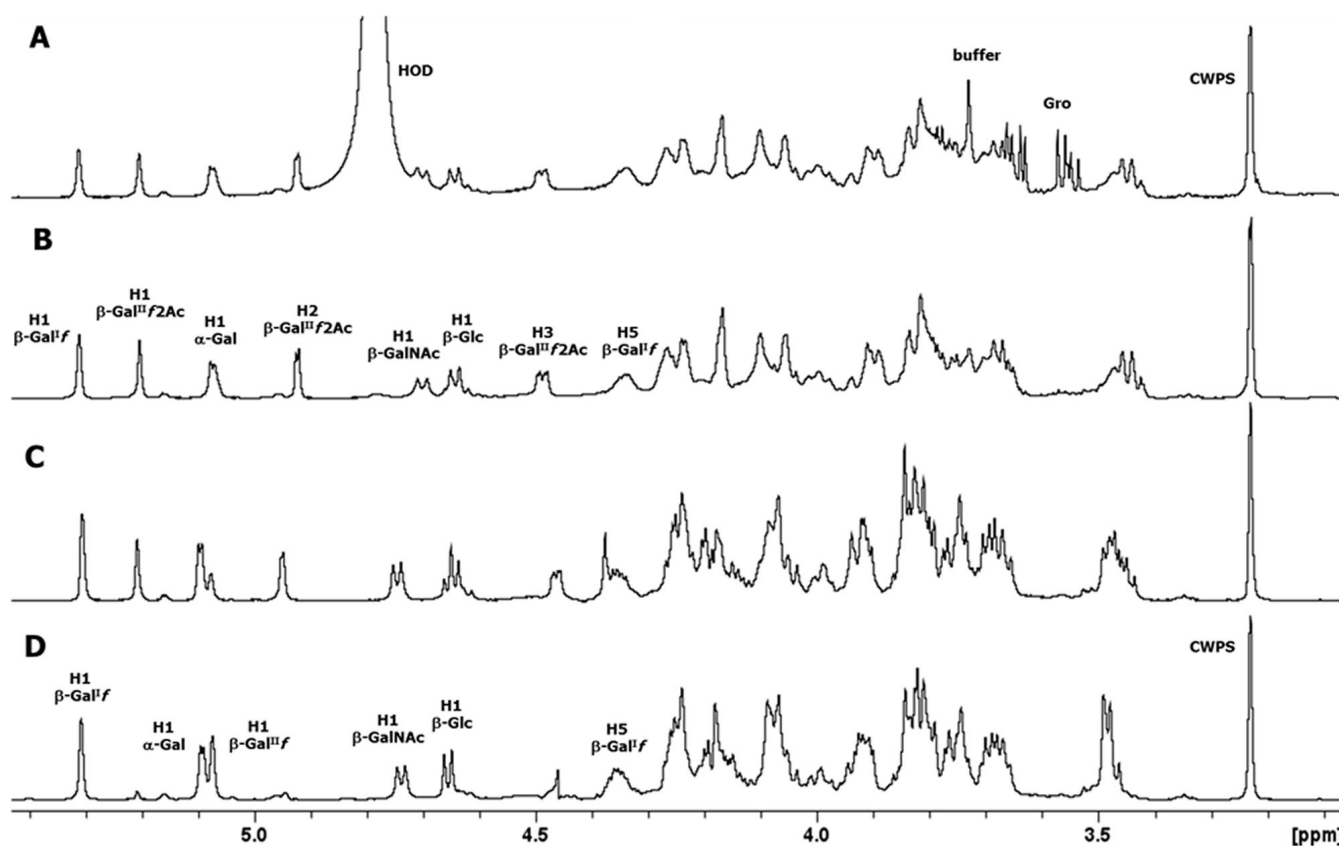


FIG 3 Expansion of the 1D proton NMR spectra of pneumococcal serotype 33X polysaccharide, some diagnostic anomeric, and ring protons are labeled. (A) 1D ^1H NMR at 500 MHz and 25°C, (B) 1D diffusion-ordered spectroscopy (DOSY) at 500 MHz and 25°C, (C) 1D DOSY at 600 MHz and 70°C (partially de-O-acetylated), and (D) 1D DOSY at 600 MHz and 60°C (de-O-acetylated polysaccharide).

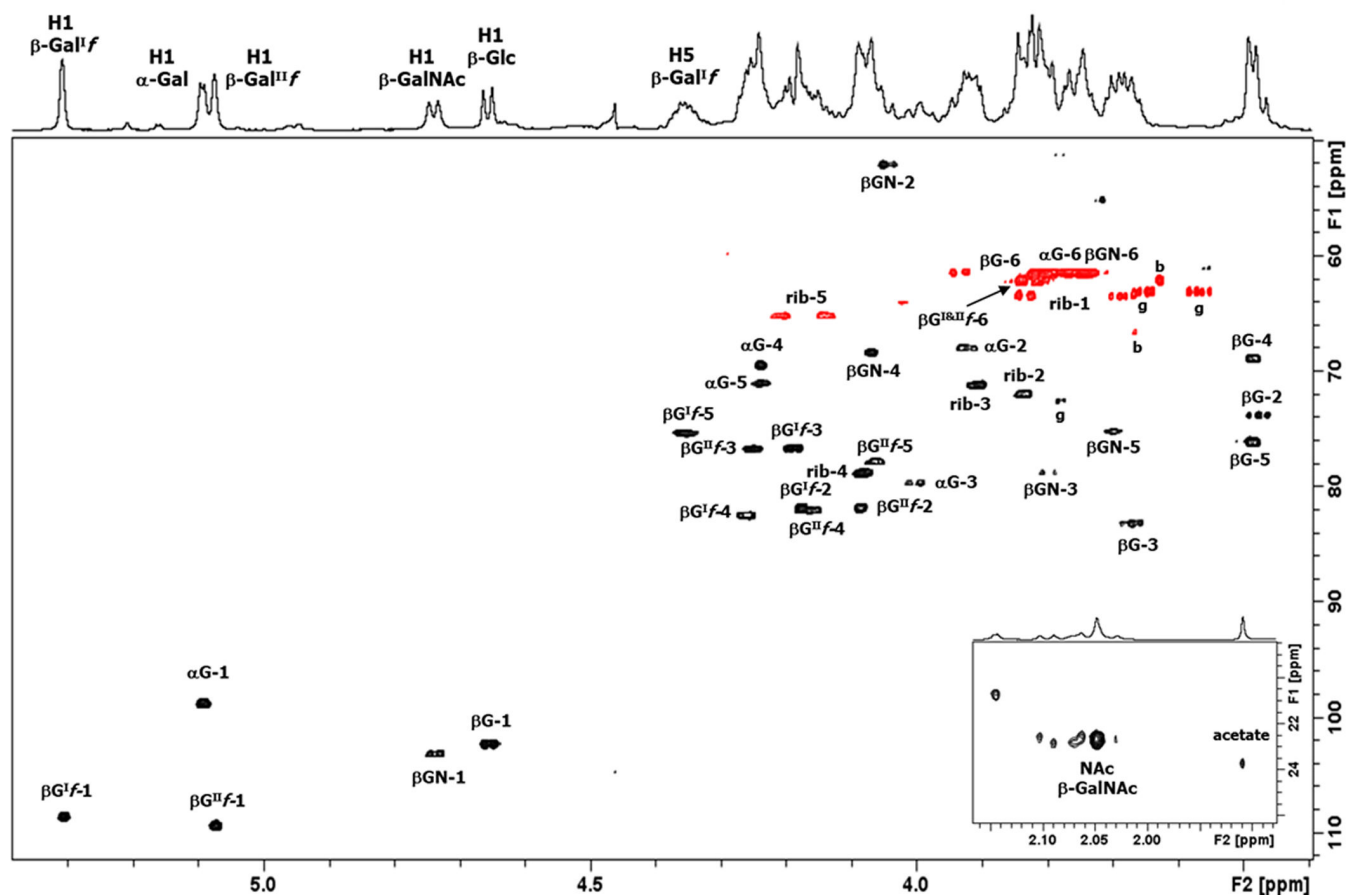


FIG 4 The expansion of the HSQC spectrum of polysaccharide 33X recorded at 600 MHz (60°C), the cross peaks from the methyl region of the spectrum are shown in the inset. All the hexasaccharide repeat unit (including ribitol-P) proton/carbon cross peaks have been labeled according to the carbon atom of the corresponding residue ($\beta G^I f = \beta\text{-Gal}^I f$, $\alpha G = \alpha\text{-Galp}$, $\beta G^I f = \beta\text{-Gal}^I f$, $\beta GN = \beta\text{-GalNAC}$, $\beta G = \beta\text{-Glc}$, and $\text{rib} = \text{ribitol}$). Additional peaks are due to buffer (B) and glycerol (G).

Identification of O-acetylation at C2 of $\beta\text{-Gal}^I f$

NMR experiments performed on the fully and partially O-acetylated polysaccharide (Fig. 3A and C) confirmed the presence of 5-linked $\beta\text{-Gal}^I f$ 2Ac with diagnostic signals for H1 at 5.20, H2 at 4.93, H3 at 4.48 ppm, and the O-acetyl singlet at 2.14 ppm. TOCSY of the partially O-acetylated polysaccharide elucidated the proton spin systems of both 5-linked $\beta\text{-Gal}^I f$ 2Ac and the de-O-acetylated 5-linked $\beta\text{-Gal}^I f$ residues (Fig. S7). Finally,

TABLE 3 NMR chemical shift for serotype 33X polysaccharide (de-O-acetylated) and O-acetylated $\text{Gal}^I f$ (lower panel)

Residue	H1/C1	H2/C2	H3/C3	H4/C4	H5/C5	H6, 6'/C6
P→5)- $\beta\text{-D-Galp}^I f\text{-(1}\rightarrow^a$	5.31/108.5	4.18/81.8	4.19/76.6	4.26/82.4	4.35/75.2	~3.83/62.1
→3)- $\alpha\text{-D-Galp}\text{-(1}\rightarrow$	5.09/98.7	3.92/67.6	4.00/79.6 ^d	4.24/69.4	4.24/71.0	4.06, 3.79/61.5
→5)- $\beta\text{-D-Galp}^I f\text{-(1}\rightarrow$	5.07/109.3	4.09/81.8	4.25/76.6	4.16/82.0	4.06/77.8	~3.83/62.1
→3)- $\beta\text{-D-GalpNAC}\text{-(1}\rightarrow^b$	4.74/103.0	4.04/52.0	3.80/78.7	4.07/68.3	3.69/75.1	3.81/61.1
→3)- $\beta\text{-D-Glcp}\text{-(1}\rightarrow$	4.66/102.2	3.47/73.7	3.67/83.1	3.48/68.8	3.49/76.0	3.94, 3.76/61.3
→4)-Ribitol-(5→P→	3.83, 3.68/63.4	3.84/71.8	3.91/71.0	4.08/78.7	4.21, 4.14/65.1	
Chemical shifts for PS-33X O-acetylated polysaccharide (500 MHz)						
→5)- $\beta\text{-D-Galp}^I f$ 2Ac-(1→ ^c	5.20/107.3	4.94/84.2	4.48/75.2	4.24/82.5	4.10/77.7	~3.82/62.1

^a³¹P at 0.34 ppm.

^bN-acetyl of $\beta\text{-D-GalpNAC}$: C=O at 174.2 ppm and methyl ¹H at 2.05 ppm; ¹³C at 22.7 ppm.

^cO-acetyl of $\beta\text{-D-Galp}^I f$ 2Ac: C=O at 173.3 ppm and methyl ¹H at 2.14 ppm; ¹³C at 20.8 ppm.

^dLinkage carbons are underlined.

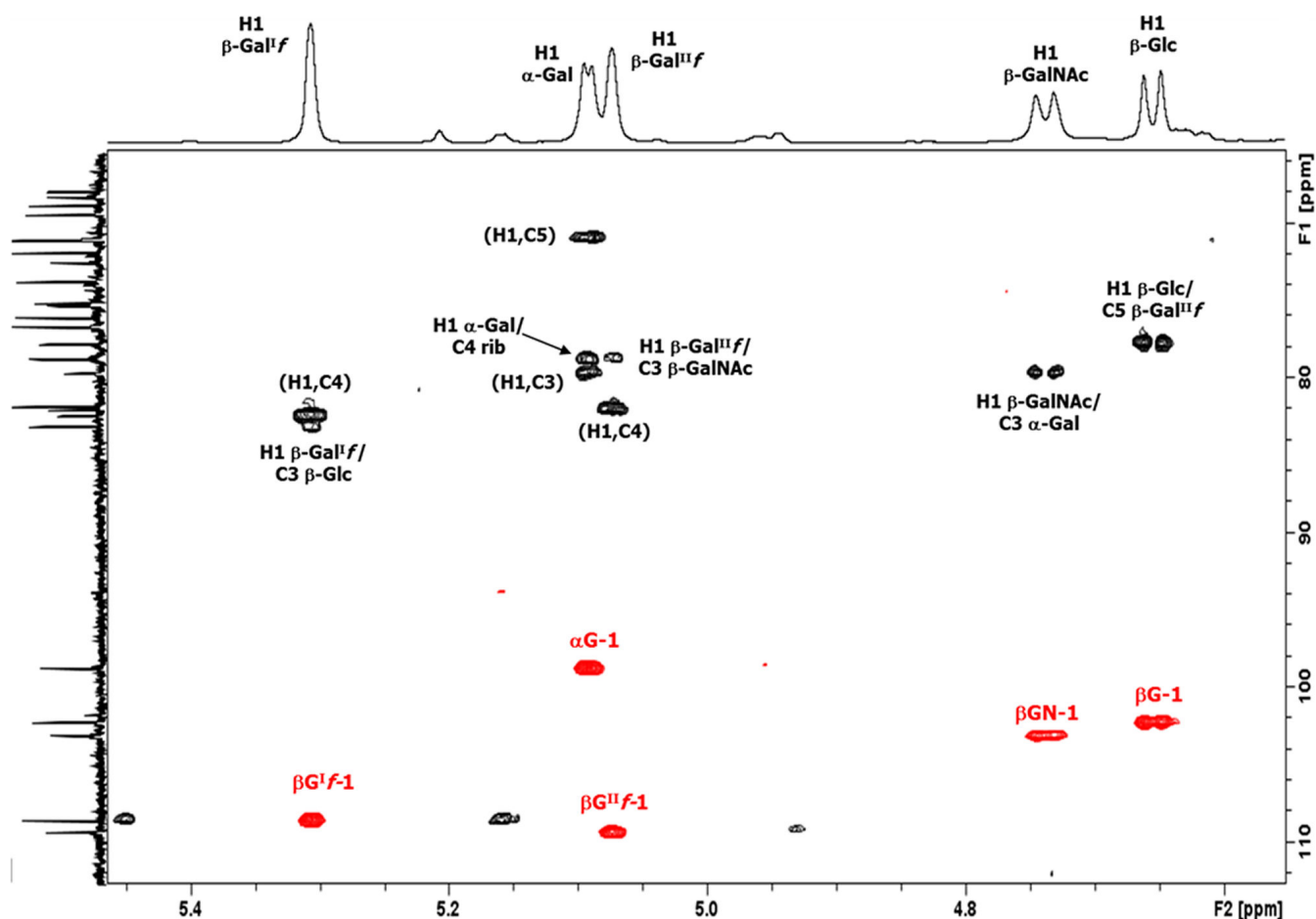


FIG 5 The 2D ^1H - ^{13}C overlay of the anomeric region for polysaccharide 33X: HSQC-DEPT (red)/HMBC (black) recorded at 600 MHz (333 K). The intra- and inter-residue HMBC cross peaks from H1 are labeled according to the carbon atom of the corresponding residue ($\beta\text{G}^{\text{I}f}$ = β -Gal $^{\text{I}f}$, αG = α -Galp, $\beta\text{G}^{\text{III}f}$ = β -Gal $^{\text{III}f}$, βGN = β -GalNAc, β -G = β -Glc, and rib = ribitol). The HMBC experiment was optimized for a coupling constant of 6 Hz and established the following linkages: H1 of β -Gal $^{\text{I}f}$ to C3 of β -Glc, H1 of α -Galp to C4 of ribitol, H1 of β -Gal $^{\text{III}f}$ to C3 of β -GalNAc, H1 of β -GalNAc to C3 of α -Galp, and H1 of β -Glc to C5 of β -Gal $^{\text{III}f}$.

HSQC-DEPT and HMBC of the fully O-acetylated polysaccharide, displayed as an overlay in Fig. 6, identified the carbon resonances for this residue. In addition, the inter-residue correlations confirmed the sequence and linkages of the serotype 33X O-acetylated hexasaccharide repeat unit. The NMR assignments for the 5-linked β -Gal $^{\text{III}f}$ 2Ac residue are provided in Table 3, lower panel.

Overall, the repeat unit of 33X is an O-acetylated hexasaccharide with components in common with both the 10B and 33B repeat units but not identical to either. Based on their known enzymatic activities (44), we were able to assign a function to all the enzymes encoded in the 33X *cps* locus to the linkages in the repeat unit, further supporting the proposed structure and its designation as a new pneumococcal serotype, hereby referred to as 33G (Fig. 7).

DISCUSSION

Traditionally, pneumococci are classified into serotypes based on differences in serological reactions using the Quellung reaction (46–48). Different serotypes produce a distinct capsular polysaccharide (48–51) encoded by a unique *cps* locus (30). Previously, seven isolates from Thailand (17, 18) and South Africa (16) with the serotype designation 33X (also known as 10X) were identified. No investigation had been conducted into the structural properties of the capsule, and it was therefore not known whether 33X is a

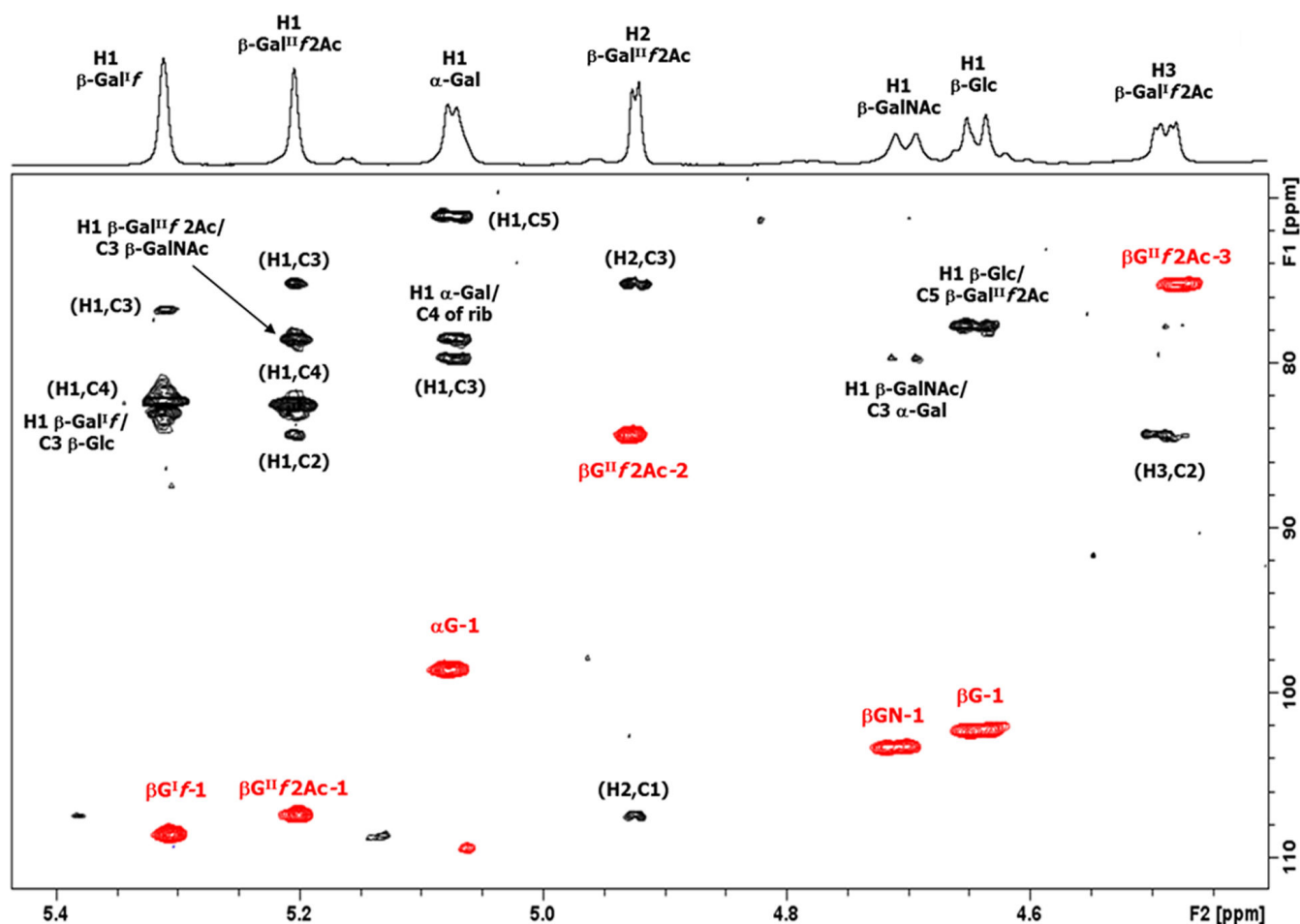


FIG 6 The 2D ^1H - ^{13}C overlay of the anomeric region for O-acetylated polysaccharide 33X: HSQC-DEPT (red)/HMBC (black) recorded at 500 MHz (298 K). The intra- and inter-residue HMBC cross peaks from diagnostic anomeric proton signals are labeled according to the carbon atom of the corresponding residue (βG^f , = β -Gal f , αG = α -Gal p , $\beta\text{G}^{11}f2\text{Ac}$ = β -Gal $^{11}f2\text{Ac}$, βGN = β -GalNAc, $\beta\text{-G}$ = β -Glc, and rib = ribitol). The HMBC experiment was optimized for a coupling constant of 6 Hz.

novel pneumococcal serotype. In this study, we provide a comprehensive investigation of this putative new serotype using 20 isolates that we identified in Mongolia.

Consistent with the previous reports (17, 18), the 33X isolates from Mongolia possess a unique *cps* locus compared with known serotypes. Using the gold standard Quellung reaction, we show that 33X pneumococci exhibit a unique serological profile, typing as both 10B and 33B. Importantly, pre-incubation of serum from mice colonized with 33X pneumococci resulted in greater ELISA inhibition than pre-incubation with either 10B or 33B pneumococci. Serotype 35A was not explored further given the negative Quellung reaction with group 35 sera. Together, these data indicate that serological responses to 33X are distinct from the related serotypes 10B and 33B. Using 1D and 2D NMR and sugar composition analyses, we interrogated the 33X capsule, demonstrating a structure that possessed features of both 10B and 33B, yet was not identical to either nor any other known pneumococcal serotype. Taken together, these data provide the requisite genetic, serological, and biochemical evidence that 33X is a new pneumococcal serotype, hereby referred to as 33G.

We have proposed placing this new serotype in serogroup 33 rather than 10 since five of the residues in the 33G hexasaccharide repeat unit are conserved in 33B compared with four for 10B. Like 33B, 33G and other serotypes in serogroup 33 (33C and 33D) also possess a hexasaccharide backbone containing ribitol-5-P. In addition, 33G has an O-acetylated β -Gal f residue, found in all members of serogroup 33 and lacking in

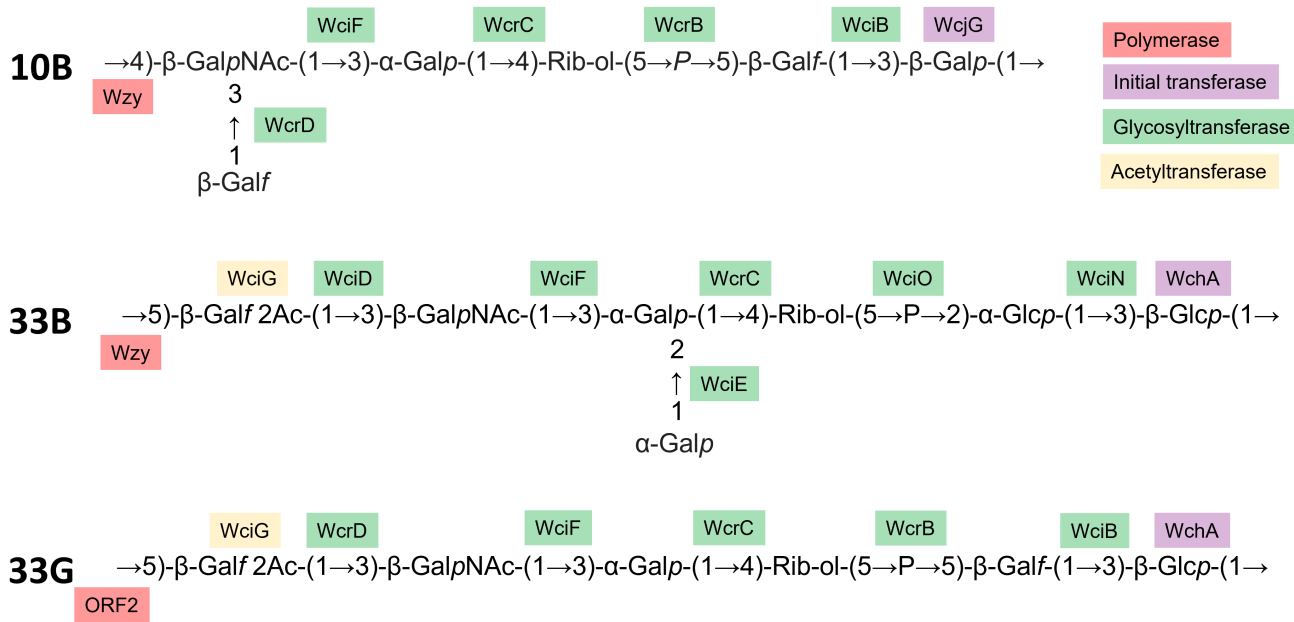


FIG 7 Polysaccharide repeat unit structure of 10B (45), 33B (43), and new serotype 33G (this study), as well as proposed enzymatic assignments based on the literature (15, 44).

serogroup 10 capsules (42, 43). The presence of O-acetyl groups in the 33G capsule is likely to have important biological relevance given that O-acetylation impacts the structure of the polysaccharide and is an epitope in pneumococcal capsules readily recognized by the immune system (48, 52). Such immune pressure has led to the emergence of new serotypes that lack the O-acetyl modification (15, 53–55). Lastly, a characteristic feature found in all serotypes in serogroup 10 is a terminal side chain β-Galf (11, 42), which is lacking in 33G.

The 33G *cps* locus is a mosaic of genes from multiple sources including three pneumococcal serogroups (35, 10, and 33) as well as other streptococci. The emergence of such a complex *cps* is the result of several recombination events. The pneumococcal *cps* locus is a known recombination hotspot, and examples of recombination with the *cps* locus of other pneumococci have been reported previously (6, 17). Similar recombination events with non-pneumococcal streptococci also occur, as exemplified by the recently described serotype 10D, which contains capsular genes from serotypes 6C, 39, and oral streptococci (11). The earliest detection of 33G was in an invasive disease isolate from South Africa in 2007 (16). However, with limited epidemiological data and a small number of isolates ($n = 27$ in total across three countries including the present study), it is difficult to determine when and where 33G emerged.

The chemical and NMR studies performed on the serotype 33G polysaccharide elucidated the O-acetylated repeat unit as →5)-β-Galf(1→3)-β-Glcp-(1→5)-β-Galf 2Ac-(1→3)-β-GalpNAc-(1→3)-α-Galp-(1→4)-Rib-ol-(5→P→). This structure explains the cross-reactivity observed with serotypes 10B and 33B by Quellung due to the common structural features seen between these serotypes. Serotypes 10B and 10F contain the tetrasaccharide β-GalpNAc-(1→3)-α-Galp-(1→4)-Rib-ol-(5→P→5)-β-Galf-(1→, but the β-GalpNAc is a branch point 3,4-linked and 4,6-linked, respectively, not 3-linked as in 33G. Serotypes 33B and 33D contain the pentasaccharide →3)-β-Glcp-(1→5)-β-Galf 2Ac-(1→3)-β-GalpNAc-(1→3)-α-Galp-(1→4)-Rib-ol-(5→P→, but the α-Galp is a branch point with a terminal α-Galp group on C-2 (43), not only 3-linked as in 33G. Additionally, the large deletions in the *wzy* gene in the 33G *cps* locus suggest that it is unlikely to encode a functional polymerase. However, *orf2* is predicted to encode a *wzy* homolog,

suggesting the acquisition of ORF2 from another streptococcal species has replaced the ancestral Wzy as the polymerase responsible for linking the 33G repeat units together.

Our findings have implications for surveillance using both phenotypic and genotypic serotyping methods. For researchers relying on serological typing methods, like the Quellung reaction, 33G will yield two serotype results (10B and 33B). Therefore, there is potential for 33G to be mistyped as 10B or 33B (or even 10F in some cases if the operator misses the weak reaction with 10d factor sera). Interestingly, 33G pneumococci did not react with any of the sera recognizing serogroup 35 despite a large portion of the 33G *cps* locus with high homology to serogroup 35. This likely occurred because the serogroup 35-derived genes include the transcriptional regulatory region (*wzg-wze*), the initial transferase (*wchA*), and one glycosyltransferase (*wciB*), which either have no impact on the capsule structure or are not responsible for antigenic sites recognized by group 35 typing sera. Not surprisingly, most molecular tools using whole genome sequence data are not yet able to accurately serotype 33G (Table S2). We have made the 33G *cps* sequences from this study publicly available (Genbank accession numbers [OR509570–OR509589](#)) so that *cps* databases can be updated. However, the users of seroBA will not need to update the database as the 33G *cps* locus is already detected (called 10X). The US Centers for Disease Control and Prevention (CDC) method for qPCR serotyping (56) does not include primers for 10B or 33B and would therefore be unlikely to detect serotype 33G. DNA microarray designates 33G as 35A/10B-like due sequences in the 33G *cps* locus matching probes for some but not all genes for the 35A and 10B *cps* loci. The TaqMan Array Card for pneumococcal serotyping (38) designates 33G as 10B due to the detection of *wcrD* in 10B, whereas the targets for other relevant serotypes such as 33B (*wciN*) and 35A (*wcrK*) are absent from the 33G *cps* locus. Serotypes 10B and 33B are not included in any licensed vaccines, so it is unlikely 33G would be misidentified as a vaccine serotype in surveillance studies.

Serotype 33G has been found in healthy carriers [reference (57) and this study], in the nasopharynx of pneumonia patients (this study) as well as in invasive disease (16). These data indicate that 33G can not only colonize but also cause disease. A total of 27 serotype 33G isolates have now been detected, including in Thailand ($n = 5$) (18), South Africa ($n = 2$) (16), and Mongolia (this study, $n = 20$). We have tested over 18,000 swabs from across Asia and have not detected 33G in any of our previous pneumococcal carriage studies in healthy children or children with acute respiratory infection including Lao PDR (58), Fiji (59), Vietnam (60), Indonesia (61), or Papua New Guinea (62). It will be interesting to see whether 33G will be identified in other countries in the future, especially as sequence-based analysis becomes more accessible.

The discovery of serotype 33G has raised multiple questions that should form the basis of future research. Firstly, similar to the recently described serotype 33E (15), 33G lacks a branching monosaccharide, which are typically highly immunogenic (63). Thus, future research should focus on identifying whether the production of a 33G capsule provides enhanced immune evasion capabilities or other biological advantages. Similarly, the function of *orf2* requires experimental investigation. Based on the bioinformatic analysis, we propose *orf2* encodes a Wzy homolog replacing the ancestral *wzy* gene, which contains deletions and is likely non-functional. However, the role of ORF2 as the polymerase needs to be experimentally validated. Lastly, serotype 33G was only detected in Mongolia from 2018 onward, following the phased introduction of PCV13 from 2016. However, there are insufficient data to determine whether the emergence of 33G in Mongolia is related to vaccine introduction or not. Future studies should focus on gaining a better understanding of the geographic distribution of 33G, the prevalence of this serotype following vaccine introduction, as well as the potential of this serotype to cause pneumococcal disease, particularly in the post-PCV era.

MATERIALS AND METHODS

Pneumococcal identification and DNA microarray serotyping from nasopharyngeal swabs

Nasopharyngeal swabs were collected, stored, transported, and tested following World Health Organization guidelines (39). Screening of swabs for pneumococci was conducted by *lytA* qPCR (24, 58). Culture of pneumococci from swabs was conducted on selective horse blood agar plates supplemented with 5 µg/mL gentamicin (Thermo Fisher Scientific). DNA was then extracted from these cultures (QIAamp 96 DNA QIAcube HT kit) and tested by DNA microarray (Senti-SPv1.5 slides) as described previously (25) to determine serotype(s). Pneumococcal isolates of interest were purified and confirmed as pneumococci using standard identification tests (39), including optochin sensitivity and genotyping by MLST and GPSC.

Whole-genome sequencing

DNA was extracted from pneumococcal isolates using the QIAamp 96 DNA QIAcube HT kit as described previously (25). Libraries were constructed using the Illumina DNA Prep kit (Illumina). Sequencing was performed using NovaSeq as 2 × 150 bp paired end. Read quality was assessed with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) with low-quality reads removed using Trimmomatic (64). Genomes were assembled *de novo* using Spades 3.15.4 (65). Genome annotation was conducted with RAST (29).

Molecular serotyping of 33X isolates

Quellung serotyping was performed as described previously (40). Genome assemblies or Fastq reads of 33X isolates were run through whole-genome sequencing-based serotyping tools PneumoCaT (33), seroBA (34), SeroCall (35), PneumoKITy (36), and PfaSTer (37) using default parameters. DNA microarray was performed as described above. Pneumococcal serotyping using the TaqMan Array Card (SPN_CDC_V1 TAC) provided by Streptococcus Laboratory, Centers for Disease Control and Prevention and was performed as previously described (38).

Competitive ELISA

Serum was generated by administering 2×10^3 CFU of 33X pneumococci (strain PMP1486) to 5-day-old C57BL/6 mice intranasally without anesthesia as described previously (66). At 7 weeks post-infection, blood was collected by cardiac puncture, allowed to clot overnight at 4°C, and subjected to centrifugation to collect serum. Animal work was approved by the Murdoch Children's Research Institute Animal Ethics Committee (protocol A945) in accordance with the Australian code for the care and use of animals for scientific purposes (67). To prepare bacterial stocks for the ELISA, pneumococcal strains of serotypes 33X (PMP1486), as well as SSI strains of 10B, 33B, and 14 (hereby named PMP738, PMP790, and PMP829, respectively) were grown to OD 0.85–0.95 in THY broth [3% (wt/vol) Todd-Hewitt broth, 0.5% (wt/vol) yeast extract] at 37°C and 5% CO₂ and stored in 80% (vol/vol) glycerol until required. Wells of high-binding Nunc Maxisorp plates were coated with PBS-washed 33X pneumococci (100 µL per well) and incubated overnight at 4°C. Competition ELISAs were conducted using a method adapted from previously established approaches (41, 68). Briefly, serum was incubated with PBS-washed preparations of pneumococci of different serotypes (33X, 10B, 33B, or 14) in 10% (vol/vol) fetal calf serum (FCS) overnight at 4°C. Serum was diluted to 1:1,000 in these reactions, which was determined to be the optimal dilution for high inhibition. The 96-well plates were blocked by adding 10% FCS for 1 h at 37°C. After washing with PBS containing 0.05% (vol/vol) tween, 50 µL of sera was added to the well and incubated for 1 h at 37°C and washed again. Biotinylated goat anti-mouse IgG (diluted 1:500 in 10% FCS; Sigma-Aldrich) was added to the wells and incubated for 1 h at 37°C. Wells

were washed, streptavidin-Horseradish Peroxidase conjugate (diluted 1:500 in 10% FCS) added and incubated for 1 h at 37°C. A 1:1 mixture of KPL TMB Peroxidase Substrate and KPL Peroxidase Substrate Solution B (Sera Care) was added to each well (50 μ L) and incubated in the dark for ~12 min. The reaction was then stopped with the addition of 1 M H_3PO_4 and optical density read at 450 nm (and 630 nm as a reference) to calculate the percentage of inhibition.

Extraction of 33X capsular polysaccharide

Pneumococcal strains (PMP1486, PMP1488, PMP1489, and PMP1521) were streaked from frozen stocks onto Columbia sheep blood agar plates and incubated overnight at 37°C and 5% CO_2 . Resultant colonies were then inoculated into modified Lacks medium (69) supplemented with 20 mM glucose and grown to an optical density at 600 nm of 0.5. Cultures were subjected to centrifugation at $3,200 \times g$ in a Heraeus centrifuge (Thermo Fisher Scientific, USA) and washed with ice-cold ultrapure water (in-house product). After another centrifugation, the supernatants were removed, and the pellets were resuspended again in ultrapure water. Buffer-saturated phenol (Thermo Fisher Scientific, USA) with a final concentration of 1% (vol/vol) was added to the suspension and incubated at room temperature overnight. Incubated samples were subjected to centrifugation for 30 min with $3,200 \times g$ at 4°C, and supernatants were collected. The collected supernatants were incubated with 3 μ L benzonase nuclease (Sigma-Aldrich, USA) overnight at 37°C. Then 20 mg/mL Proteinase K solution (Roche, Switzerland) was added and again incubated overnight at 37°C. Next, the solution was transferred to a Millipore Amicon Ultra 30 kDa cut off membrane centrifugal filter unit (Merck Millipore, USA; prewashed twice to remove glycerol) and subjected to centrifugation for 25 min at $3,200 \times g$. The remaining solution above the filter was collected and evaporated under reduced pressure (25 mbar). Dried polysaccharide samples were submitted for chemical and NMR analysis.

Sugar composition and linkage analysis

Composition analysis was carried out by chemical derivatization of the sample both as alditol acetates and trimethylsilyl methyl glycosides. Alditol acetates were obtained after hydrolysis with 2 M trifluoroacetic acid (TFA) for 1 h at 125°C, followed by reduction with sodium borohydride ($NaBD_4$) and peracetylation with acetic anhydride (70). Trimethylsilyl methyl glycosides were obtained by derivatization with the reagent Sylon HTP (Merck) after methanolysis of the polysaccharide with 3 M HCl in methanol at 85°C for 16 h (71). To determine the position of the glycosidic linkages, the polysaccharide 33X was permethylated following the protocol developed by Harris (72), hydrolyzed with 2 M TFA for 1 h at 125°C, reduced with sodium borodeuteride ($NaBD_4$), and peracetylated with acetic anhydride to give a mixture of partially methylated alditol acetates.

The derivatized samples were analyzed by GLC using an Agilent Technologies 6,850 gas chromatograph equipped with a flame ionization detector, using He as the carrier gas and a Zebron ZB-5 MSi capillary column (Phenomenex, 30 m \times 250 μ m \times 0.25 μ m). The following temperature programs were used: for alditol acetates, 3 min at 150°C, 150°C–270°C at 3°C/min, and 2 min at 270°C; for trimethylsilyl methyl glycosides, 1 min at 150°C, 150°C–280°C at 5°C/min, and 2 min at 280°C; for partially methylated alditol acetates, 1 min at 90°C, 90°C–140°C at 25°C/min, 140°C–200°C at 5°C/min, 200°C–280°C at 10°C/min, and 10 min at 280°C. GLC-MS analyses were carried out on an Agilent Technologies 7,890 A gas chromatograph coupled to an Agilent Technologies 5975C VL MSD using the same column and the temperature programs of the GLC analyses. Values of the integrated area of the partially methylated alditol acetates were corrected by the effective carbon response factors (73).

NMR analysis

A sample of polysaccharide 33X (1–2 mg from PMP1486, PMP1488, PMP1489, and PMP1521) was dissolved in deuterium oxide (D_2O ; Sigma-Aldrich, USA), subjected to

centrifugation for 5 min at $16,800 \times g$, transferred to 1.7 mm NMR tubes (Bruker, USA) and submitted for NMR analysis. Initial studies were conducted on a Bruker Neo 500 MHz NMR spectrometer equipped with a 1.7-mm triple resonance probe (TXI) probehead; the probe temperature was set at 25°C. Further NMR experiments were performed after recovery of the sample, cycles of D₂O exchange, and transfer to a 5-mm tube, on a Bruker Avance III 600 MHz NMR spectrometer equipped with a BBO Prodigy cryoprobe and processed using standard Bruker software (Topspin 3.2). The probe temperature was set at 60°C or 70°C. 1D (¹H, ³¹P, and ¹³C) and 2D, COSY, TOCSY, NOESY, HSQC, and HMBC NMR experiments were performed. 2D COSY and NOESY experiments were recorded with pre-saturation of HOD, whereas the TOCSY experiments were performed using DOSY to remove signals from low molecular weight components (ledbpgpml2s2d). 2D TOCSY experiments were recorded using a mixing time of 180 ms and the 1D variants using 180 or 200 ms. 2D NOESY experiments were recorded using a mixing time of 300 ms and the 1D variants using 300 or 500 ms. The HSQC experiment was optimized for $J = 145$ Hz (for directly attached ¹H-¹³C correlations), and the HMBC experiments were optimized for a coupling constant of 6 Hz (for long-range ¹H-¹³C correlations) and 10 Hz (for ¹H-³¹P correlations). To improve sensitivity by performing many scans, the 2D experiments were recorded using non-uniform sampling: 40% for homonuclear and 20%–30% for heteronuclear experiments. Spectra were referenced to residual CWPS: ¹H signal at 3.23 ppm, ¹³C signal at 54.5 ppm, and the shielded ³¹P signal at 1.30 ppm (74).

ACKNOWLEDGMENTS

We thank all the staff involved in the recruitment and nasopharyngeal swab collection in the carriage surveys and pneumonia surveillance programs in Mongolia as well as the Translational Microbiology Group at the Murdoch Children's Research Institute. We also thank all the participants and their families involved in these studies. Thank you to Beth Temple for epidemiological advice regarding nasopharyngeal swab data. We thank Lesley McGee and Srinivasan Velusamy for providing the pneumococcal TaqMan Array Cards.

This study was funded by a Robert Austrian Research Award in Pneumococcal Vaccinology awarded to SM by ISPPD (funded by Pfizer) and the National Health and Medical Research Council (NHMRC) Center of Research Excellence for Pneumococcal Disease Control in the Asia-Pacific (GNT1196415). The Mongolia carriage surveys and pneumonia surveillance programs were funded by the Bill and Melinda Gates foundation (Grant number OPP1115490), The Gavi Alliance (contract number PP61690717A2), and a Pfizer clinical research collaboration agreement with MCRI (contract number WI236621) for which MCRI was the study sponsor. Capsule structure work was funded by the Swiss National Science Foundation (Grant number 197083). The work at the Murdoch Children's Research Institute was also supported by the Victorian Government's Operational Infrastructure Support Program. Funders were neither involved in the collection and analysis of data nor in the decision to publish.

Conceptualization: S.M. and C.S. Methodology and study design: S.M., J.H., F.M.R., E.K.M., O.T., T.M., C.V.M., E.M.D., B.D.G., P.V.L., P.C., N.R., M.H., and C.S. Data analysis: S.M., C.L.P., P.V.L., P.C., N.R., M.H., and C.S. Experimental investigation: S.M., B.D.O., B.B., I.G., J.P.W., C.L.P., E.N., S.W.L., J.H., S.D.B., and N.R. Funding acquisition: S.M., E.K.M., F.M.R., M.H., and C.S. Writing of the original manuscript draft: S.M., M.H., N.R., and C.S. Review and editing of the manuscript for submission: S.M., B.D.O., J.P.W., B.B., C.L.P., E.N., I.G., S.W.L., J.H., S.D.B., F.M.R., E.K.M., O.T., E.M.D., B.D.G., T.M., C.V.M., P.V.L., P.C., N.R., M.H., and C.S.

AUTHOR AFFILIATIONS

¹Infection, Immunity, and Global Health, Murdoch Children's Research Institute, Melbourne, Australia

²Department of Pediatrics, The University of Melbourne, Melbourne, Australia

³Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

⁴Institute for Infectious Diseases, University of Bern, Bern, Switzerland

⁵Department of Life Sciences, University of Trieste, Trieste, Italy

⁶Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland

⁷Parasites and Microbes, Wellcome Sanger Institute, Hinxton, United Kingdom

⁸Institute for Infection and Immunity, St. George's, University of London, London, United Kingdom

⁹BUGS Bioscience, London Bioscience Innovation Center, London, United Kingdom

¹⁰National Center for Communicable Diseases, Ministry of Health, Ulaanbaatar, Mongolia

¹¹Pfizer Vaccines, Collegeville, Pennsylvania, USA

¹²Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

¹³Department of Chemistry, University of Cape Town, Rondebosch, South Africa

PRESENT ADDRESS

Barbara Bellich, Department of Advanced Translational Diagnostics, Institute for Maternal and Child Health, IRCCS "Burlo Garofolo", Trieste, Italy

AUTHOR ORCIDs

Sam Manna  <http://orcid.org/0000-0002-7694-9605>

Stephanie Lo  <http://orcid.org/0000-0002-2182-0222>

Catherine Satzke  <http://orcid.org/0000-0003-3164-8849>

DATA AVAILABILITY

The 33G capsule locus sequences have been deposited in GenBank (accession no. [OR509570-OR509589](https://doi.org/10.1093/nar/51/12/6789)).

ETHICS APPROVAL

Participants were recruited into either the carriage survey or pneumonia surveillance studies in accordance with the inclusion criteria described previously (19, 21, 22). Ethics approval was obtained from the Medical Ethics Review Committee at the Mongolian Ministry of Health and the Royal Children's Hospital Human Research Ethics Committee (HREC reference numbers 33203 and 38045). Written consent was obtained by participants or their parents/caregivers before being included in the study.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (Spectrum03579-23-s0001.docx). Fig. S1 to S7; Tables S1 and S2.

REFERENCES

1. WahI B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, Lukšić I, Nair H, McAllister DA, Campbell H, Rudan I, Black R, Knoll MD. 2018. 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* 6:e744–e757. [https://doi.org/10.1016/S2214-109X\(18\)30247-X](https://doi.org/10.1016/S2214-109X(18)30247-X)
2. Weiser JN, Ferreira DM, Paton JC. 2018. *Streptococcus pneumoniae*: transmission, colonization and invasion. *Nat Rev Microbiol* 16:355–367. <https://doi.org/10.1038/s41579-018-0001-8>
3. Ganaie FA, Saad JS, Lo SW, McGee L, Bentley SD, van Tonder AJ, Hawkins P, Keenan JD, Calix JJ, Nahm MH. 2023. Discovery and characterization of pneumococcal serogroup 36 capsule subtypes, serotypes 36A and 36B. *J Clin Microbiol* 61:e0002423. <https://doi.org/10.1128/jcm.00024-23>
4. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, Page AJ, Marttinen P, Bentley LJ, Ochoa TJ, et al. 2019. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine* 43:338–346. <https://doi.org/10.1016/j.ebiom.2019.04.021>

5. Blackwell GA, Hunt M, Malone KM, Lima L, Horesh G, Alako BTF, Thomson NR, Iqbal Z, Hanage WP. 2021. Exploring bacterial diversity via a curated and searchable snapshot of archived DNA sequences. *PLoS Biol* 19:e3001421. <https://doi.org/10.1371/journal.pbio.3001421>
6. Manna S, Dunne EM, Ortika BD, Pell CL, Kama M, Russell FM, Mungun T, Mulholland EK, Hinds J, Satzke C. 2018. Discovery of a *Streptococcus pneumoniae* serotype 33F capsular polysaccharide locus that lacks *wcjE* and contains a *wcyO* pseudogene. *PLoS One* 13:e0206622. <https://doi.org/10.1371/journal.pone.0206622>
7. Manna S, Ortika BD, Dunne EM, Holt KE, Kama M, Russell FM, Hinds J, Satzke C. 2018. A novel genetic variant of *Streptococcus pneumoniae* serotype 11A discovered in Fiji. *Clin Microbiol Infect* 24:428. <https://doi.org/10.1016/j.cmi.2017.06.031>
8. Kapatai G, Sheppard CL, Troxler LJ, Litt DJ, Furrer J, Hilty M, Fry NK. 2017. Pneumococcal 23B molecular subtype identified using whole genome sequencing. *Genome Biol Evol* 9:2122–2135. <https://doi.org/10.1093/gbe/evx092>
9. Manna S, Spry L, Wee-Hee A, Ortika BD, Boelsen LK, Batinovic S, Mazarakis N, Ford RL, Lo SW, Bentley SD, Russell FM, Blyth CC, Pomat WS, Petrovski S, Hinds J, Licciardi PV, Satzke C. 2022. Variants of *Streptococcus pneumoniae* serotype 14 from Papua New Guinea with the potential to be mistyped and escape vaccine-induced protection. *Microbiol Spectr* 10:e0152422. <https://doi.org/10.1128/spectrum.01524-22>
10. Dunne EM, Tikkanen L, Balloch A, Gould K, Yoannes M, Phuanukoonnon S, Licciardi PV, Russell FM, Mulholland EK, Satzke C, Hinds J. 2015. Characterization of 19A-like 19F pneumococcal isolates from Papua New Guinea and Fiji. *New Microbes New Infect* 7:86–88. <https://doi.org/10.1016/j.nmni.2015.06.008>
11. Ganaie F, Saad JS, McGee L, van Tonder AJ, Bentley SD, Lo SW, Gladstone RA, Turner P, Keenan JD, Breiman RF, Nahm MH. 2020. A new pneumococcal capsule type, 10D, is the 100th serotype and has a large cps fragment from an oral streptococcus. *mBio* 11:e00937–20. <https://doi.org/10.1128/mBio.00937-20>
12. Ganaie F, Maruhn K, Li C, Porambo RJ, Elverdal PL, Abeygunwardana C, van der Linden M, Duus JØ, Sheppard CL, Nahm MH. 2021. Structural, genetic, and serological elucidation of *Streptococcus pneumoniae* serogroup 24 serotypes: discovery of a new serotype, 24C, with a variable capsule structure. *J Clin Microbiol* 59:e0054021. <https://doi.org/10.1128/JCM.00540-21>
13. Geno KA, Saad JS, Nahm MH. 2017. Discovery of novel pneumococcal serotype 35D, a natural WciG-deficient variant of serotype 35B. *J Clin Microbiol* 55:1416–1425. <https://doi.org/10.1128/JCM.00054-17>
14. Kjeldsen C, Slott S, Elverdal PL, Sheppard CL, Kapatai G, Fry NK, Skovsted IC, Duus JØ. 2018. Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D. *Carbohydr Res* 463:24–31. <https://doi.org/10.1016/j.carres.2018.04.011>
15. Ganaie FA, Saad JS, Lo SW, McGee L, van Tonder AJ, Hawkins PA, Calix JJ, Bentley SD, Nahm MH. 2023. Novel pneumococcal capsule type 33E results from the inactivation of glycosyltransferase WciE in vaccine type 33F. *J Biol Chem* 299:105085. <https://doi.org/10.1016/j.jbc.2023.105085>
16. van Tonder AJ, Gladstone RA, Lo SW, Nahm MH, du Plessis M, Cornick J, Kwambana-Adams B, Madhi SA, Hawkins PA, Benisty R, Dagan R, Everett D, Antonio M, Klugman KP, von Gottberg A, Breiman RF, McGee L, Bentley SD, The Global Pneumococcal Sequencing Consortium. 2019. Putative novel cps loci in a large global collection of pneumococci. *Microbial Genomics* 5. <https://doi.org/10.1099/mgen.0.000274>
17. Mostowy RJ, Croucher NJ, De Maio N, Chewapreecha C, Salter SJ, Turner P, Aanensen DM, Bentley SD, Didelot X, Fraser C. 2017. Pneumococcal capsule synthesis locus *cps* as evolutionary hotspot with potential to generate novel serotypes by recombination. *Mol Biol Evol* 34:2537–2554. <https://doi.org/10.1093/molbev/msx173>
18. van Tonder AJ, Bray JE, Quirk SJ, Haraldsson G, Jolley KA, Maiden MCJ, Hoffmann S, Bentley SD, Haraldsson Á, Erlendsdóttir H, Kristinsson KG, Brueggemann AB. 2016. Putatively novel serotypes and the potential for reduced vaccine effectiveness: capsular locus diversity revealed among 5405 pneumococcal genomes. *Microb Genom* 2:000090. <https://doi.org/10.1099/mgen.0.000090>
19. La Vincente SF, von Mollendorf C, Ulziibayar M, Satzke C, Dashtseren L, Fox KK, Dunne EM, Nguyen CD, de Campo J, de Campo M, Thomson H, Surenkhand G, Dembereluren S, Bujinlkham S, Do LAH, Narangerel D, Cherian T, Mungun T, Mulholland EK. 2019. Evaluation of a phased pneumococcal conjugate vaccine introduction in Mongolia using enhanced pneumonia surveillance and community carriage surveys: a study protocol for a prospective observational study and lessons learned. *BMC Public Health* 19:333. <https://doi.org/10.1186/s12889-019-6639-y>
20. Chan J, Mungun T, Batsaivan P, Ulziibayar M, Suuri B, Otgonbayar D, Luvsantseren D, Nguyen CD, Narangerel D, Dunne EM, Fox K, Hinds J, Nation ML, Pell CL, Mulholland EK, Satzke C, von Mollendorf C, Russell FM, PneuCAPTIVE Mongolia Research Group. 2021. Direct and indirect effects of 13-valent pneumococcal conjugate vaccine on pneumococcal carriage in children hospitalised with pneumonia from formal and informal settlements in Mongolia: an observational study. *Lancet Reg Health West Pac* 15:100231. <https://doi.org/10.1016/j.lanwpc.2021.100231>
21. von Mollendorf C, Dunne EM, La Vincente S, Ulziibayar M, Suuri B, Luvsantseren D, Narangerel D, Ortika BD, Pell CL, Nation ML, Alamrousi A, Hinds J, Dembereluren S, Nguyen C, Mungun T, Mulholland EK, Satzke C. 2019. Pneumococcal carriage in children in Ulaanbaatar, Mongolia before and one year after the introduction of the 13-valent pneumococcal conjugate vaccine. *Vaccine* 37:4068–4075. <https://doi.org/10.1016/j.vaccine.2019.05.078>
22. von Mollendorf C, Ulziibayar M, Gessner BD, Do LAH, Nguyen CD, Beavon R, Suuri B, Luvsantseren D, Narangerel D, Jenney A, Dunne EM, Satzke C, Darmaa B, Mungun T, Mulholland EK. 2021. Evaluation of the impact of childhood 13-valent pneumococcal conjugate vaccine introduction on adult pneumonia in Ulaanbaatar, Mongolia: study protocol for an observational study. *BMC Public Health* 21:1731. <https://doi.org/10.1186/s12889-021-11776-8>
23. von Mollendorf C, Ulziibayar M, Nguyen CD, Batsaikhan P, Suuri B, Luvsantseren D, Narangerel D, de Campo J, de Campo M, Tsolmon B, Dembereluren S, Dunne EM, Satzke C, Mungun T, Mulholland K. 2023. Impact of pneumococcal conjugate vaccine introduction on pneumonia incidence rates in children 2–59 months of age in Mongolia, 2015–2021: population-based surveillance. *SSRN*. <https://doi.org/10.2139/ssrn.4485625>
24. Carvalho M da G, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, Steigerwalt A, Whaley M, Facklam RR, Fields B, Carlone G, Ades EW, Dagan R, Sampson JS. 2007. Evaluation and improvement of real-time PCR assays targeting *lytA*, *ply*, and *psaA* genes for detection of pneumococcal DNA. *J Clin Microbiol* 45:2460–2466. <https://doi.org/10.1128/JCM.02498-06>
25. Satzke C, Dunne EM, Porter BD, Klugman KP, Mulholland EK, PneuCarriage project group. 2015. The PneuCarriage project: a multi-centre comparative study to identify the best serotyping methods for examining pneumococcal carriage in vaccine evaluation studies. *PLoS Med* 12:e1001903. <https://doi.org/10.1371/journal.pmed.1001903>
26. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3:124. <https://doi.org/10.12688/wellcomeopenres.14826.1>
27. Enright MC, Spratt BG. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology (Reading)* 144 (Pt 11):3049–3060. <https://doi.org/10.1099/00221287-144-11-3049>
28. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. 2019. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 29:304–316. <https://doi.org/10.1101/gr.241455.118>
29. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>
30. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitz E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kalltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG. 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2:e31. <https://doi.org/10.1371/journal.pgen.0020031>
31. Rasmussen LH, Dargis R, Højholt K, Christensen JJ, Skovgaard O, Justesen US, Rosenvinge FS, Moser C, Lukjancenko O, Rasmussen S, Nielsen XC. 2016. Whole genome sequencing as a tool for phylogenetic analysis of

- clinical strains of Mitis group streptococci. *Eur J Clin Microbiol Infect Dis* 35:1615–1625. <https://doi.org/10.1007/s10096-016-2700-2>
32. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27:1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>
 33. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry NK. 2016. Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ* 4:e2477. <https://doi.org/10.7717/peerj.2477>
 34. Epping L, van Tonder AJ, Gladstone RA, Bentley SD, Page AJ, Keane JA, The Global Pneumococcal Sequencing Consortium. 2018. SeroBA: rapid high-throughput serotyping of *Streptococcus pneumoniae* from whole genome sequence data. *Microb Genom* 4:e000204. <https://doi.org/10.1099/mgen.0.000204>
 35. Knight JR, Dunne EM, Mulholland EK, Saha S, Satzke C, Tothpal A, Weinberger DM. 2021. Determining the serotype composition of mixed samples of pneumococcus using whole-genome sequencing. *Microb Genom* 7:1–10. <https://doi.org/10.1099/mgen.0.000494>
 36. Sheppard CL, Manna S, Groves N, Litt DJ, Amin-Chowdhury Z, Bertran M, Ladhani S, Satzke C, Fry NK. 2022. PneumoKITy: a fast, flexible, specific, and sensitive tool for *Streptococcus pneumoniae* serotype screening and mixed serotype detection from genome sequence data. *Microb Genom* 8:000904. <https://doi.org/10.1099/mgen.0.000904>
 37. Lee JT, Li X, Hyde C, Liberator PA, Hao L. 2023. PfaSTer: a machine learning-powered serotype caller for *Streptococcus pneumoniae* genomes. *Microb Genom* 9:001033. <https://doi.org/10.1099/mgen.0.001033>
 38. Pholwat S, Sakai F, Turner P, Vidal JE, Houpt ER. 2016. Development of a taqman array card for pneumococcal serotyping on isolates and nasopharyngeal samples. *J Clin Microbiol* 54:1842–1850. <https://doi.org/10.1128/JCM.00613-16>
 39. Satzke C, Turner P, Virolainen-Julkunen A, Adrian PV, Antonio M, Hare KM, Henao-Restrepo AM, Leach AJ, Klugman KP, Porter BD, Sá-Leão R, Scott JA, Nohynek H, O'Brien KL, WHO Pneumococcal Carriage Working Group. 2013. Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the world health organization pneumococcal carriage working group. *Vaccine* 32:165–179. <https://doi.org/10.1016/j.vaccine.2013.08.062>
 40. Habib M, Porter BD, Satzke C. 2014. Capsular serotyping of *Streptococcus pneumoniae* using the quellung reaction. *J Vis Exp* 84:e51208. <https://doi.org/10.3791/51208>
 41. Licciardi PV, Balloch A, Russell FM, Mulholland EK, Tang MLK. 2010. Antibodies to serotype 9V exhibit novel serogroup cross-reactivity following infant pneumococcal immunization. *Vaccine* 28:3793–3800. <https://doi.org/10.1016/j.vaccine.2010.03.033>
 42. Richardson NI, Kuttel MM, Ravenscroft N. 2022. Modeling of pneumococcal serogroup 10 capsular polysaccharide molecular conformations provides insight into epitopes and observed cross-reactivity. *Front Mol Biosci* 9:961532. <https://doi.org/10.3389/fmolb.2022.961532>
 43. Lin FL, Vinogradov E, Deng C, Zeller S, Phelan L, Green BA, Jansen KU, Pavliak V. 2014. Structure elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 33C, 33D, and revised structure of serotype 33B. *Carbohydr Res* 383:97–104. <https://doi.org/10.1016/j.carres.2013.11.006>
 44. Aanensen DM, Mavroidi A, Bentley SD, Reeves PR, Spratt BG. 2007. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol* 189:7856–7876. <https://doi.org/10.1128/JB.00837-07>
 45. Yang J, Nahm MH, Bush CA, Cisar JO. 2011. Comparative structural and molecular characterization of *Streptococcus pneumoniae* capsular polysaccharide serogroup 10. *J Biol Chem* 286:35813–35822. <https://doi.org/10.1074/jbc.M111.255422>
 46. Dochez AR, Gillespie LJ. 1913. A biologic classification of pneumococci by means of immunity reactions. *JAMA* 61:727. <https://doi.org/10.1001/jama.1913.04350100005003>
 47. Cooper G, Rosenstein C, Walter A, Peizer L. 1932. The further separation of types among the pneumococci hitherto included in group IV and the development of therapeutic antisera for these types. *J Exp Med* 55:531–554. <https://doi.org/10.1084/jem.55.4.531>
 48. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH. 2015. Pneumococcal capsules and their types: past, present, and future. *Clin Microbiol Rev* 28:871–899. <https://doi.org/10.1128/CMR.00024-15>
 49. Griffith F. 1928. The significance of pneumococcal types. *J Hyg (Lond)* 27:113–159. <https://doi.org/10.1017/s0022172400031879>
 50. Heidelberger M, Avery OT. 1923. The soluble specific substance of pneumococcus. *J Exp Med* 38:73–79. <https://doi.org/10.1084/jem.38.1.73>
 51. Heidelberger M, Avery OT. 1924. The soluble specific substance of pneumococcus: second paper. *J Exp Med* 40:301–317. <https://doi.org/10.1084/jem.40.3.301>
 52. Rajam G, Carlone GM, Romero-Steiner S. 2007. Functional antibodies to the O-acetylated pneumococcal serotype 15B capsular polysaccharide have low cross-reactivities with serotype 15C. *Clin Vaccine Immunol* 14:1223–1227. <https://doi.org/10.1128/CI.00184-07>
 53. Calix JJ, Nahm MH. 2010. A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. *J Infect Dis* 202:29–38. <https://doi.org/10.1086/653123>
 54. Geno KA, Spencer BL, Bae S, Nahm MH. 2018. Ficolin-2 binds to serotype 35B pneumococcus as it does to serotypes 11A and 31, and these serotypes cause more infections in older adults than in children. *PLoS One* 13:e0209657. <https://doi.org/10.1371/journal.pone.0209657>
 55. Brady AM, Calix JJ, Yu J, Geno KA, Cutter GR, Nahm MH. 2014. Low invasiveness of pneumococcal serotype 11A is linked to ficolin-2 recognition of O-acetylated capsule epitopes and lectin complement pathway activation. *J Infect Dis* 210:1155–1165. <https://doi.org/10.1093/infdis/jiu195>
 56. Velusamy S, Tran T, Mongkolrattanothai T, Walker H, McGee L, Beall B. 2020. Expanded sequential quadriplex real-time polymerase chain reaction (PCR) for identifying pneumococcal serotypes, penicillin susceptibility, and resistance markers. *Diagn Microbiol Infect Dis* 97:115037. <https://doi.org/10.1016/j.diagmicrobio.2020.115037>
 57. Turner P, Turner C, Jankhot A, Helen N, Lee SJ, Day NP, White NJ, Nosten F, Goldblatt D, Beall B. 2012. A longitudinal study of *Streptococcus pneumoniae* carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. *PLoS One* 7:e38271. <https://doi.org/10.1371/journal.pone.0038271>
 58. Satzke C, Dunne EM, Choummanivong M, Ortika BD, Neal EFG, Pell CL, Nation ML, Fox KK, Nguyen CD, Gould KA, Hinds J, Chanthongthip A, Xeuatvongsa A, Mulholland EK, Sychareun V, Russell FM. 2019. Pneumococcal carriage in vaccine-eligible children and unvaccinated infants in Lao PDR two years following the introduction of the 13-valent pneumococcal conjugate vaccine. *Vaccine* 37:296–305. <https://doi.org/10.1016/j.vaccine.2018.10.077>
 59. Dunne EM, Satzke C, Ratu FT, Neal EFG, Boelsen LK, Matanitobua S, Pell CL, Nation ML, Ortika BD, Reyburn R, Jenkins K, Nguyen C, Gould K, Hinds J, Tikoduadua L, Kado J, Rafai E, Kama M, Mulholland EK, Russell FM. 2018. Effect of ten-valent pneumococcal conjugate vaccine introduction on pneumococcal carriage in Fiji: results from four annual cross-sectional carriage surveys. *Lancet Glob Health* 6:e1375–e1385. [https://doi.org/10.1016/S2214-109X\(18\)30383-8](https://doi.org/10.1016/S2214-109X(18)30383-8)
 60. Temple B, Nation ML, Dai VTT, Beissbarth J, Bright K, Dunne EM, Hinds J, Hoan PT, Lai J, Nguyen CD, Ortika BD, Phan TV, Thuy HNL, Toan NT, Uyen DY, Satzke C, Smith-Vaughan H, Huu TN, Mulholland K. 2021. Effect of a 2+1 schedule of ten-valent versus 13-valent pneumococcal conjugate vaccine on pneumococcal carriage: results from a randomised controlled trial in Vietnam. *Vaccine* 39:2303–2310. <https://doi.org/10.1016/j.vaccine.2021.02.043>
 61. Murad G, Dunne EM, Sudigdoadi S, Fadlyana E, Tarigan R, Pell CL, Watts E, Nguyen CD, Satzke C, Hinds J, Dewi MM, Dhamayanti M, Sekarwana N, Rusmil K, Mulholland EK, Kartasasmita C. 2019. Pneumococcal carriage, density, and co-colonization dynamics: a longitudinal study in Indonesian infants. *Int J Infect Dis* 86:73–81. <https://doi.org/10.1016/j.ijid.2019.06.024>
 62. Blyth CC, Britton KJ, Nguyen CD, Sapura J, Kave J, Nivio B, Chan J, Satzke C, Ford R, Kirarock W, Lehmann D, Pomat W, Russell FM. 2022. Effectiveness of 13-valent pneumococcal conjugate vaccine against hypoxic pneumonia and hospitalisation in Eastern Highlands Province, Papua New Guinea: an observational cohort study. *Lancet Reg Health West Pac* 22:100432. <https://doi.org/10.1016/j.lanwpc.2022.100432>

63. Ravenscroft N, Omar A, Hlozek J, Edmonds-Smith C, Follador R, Serventi F, Lipowsky G, Kuttel MM, Cescutti P, Faridmoayer A. 2017. Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F. *Carbohydr Res* 450:19–29. <https://doi.org/10.1016/j.carres.2017.08.006>
64. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
65. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
66. Manna S, McAuley J, Jacobson J, Nguyen CD, Ullah M, Sebina I, Williamson V, Mulholland EK, Wijburg O, Phipps S, Satzke C, Rosenberg HF. 2022. Synergism and antagonism of bacterial-viral coinfection in the upper respiratory tract. *mSphere* 7:e0098421. <https://doi.org/10.1128/msphere.00984-21>
67. NHMRC. Australian code for the care and use of animals for scientific purposes | NHMRC. Available from: <https://www.nhmrc.gov.au/about-us/publications/australian-code-care-and-use-animals-scientific-purposes#block-views-block-file-attachments-content-block-1>. Retrieved 28 Jul 2023.
68. Cohen JM, Khandavilli S, Camberlein E, Hyams C, Baxendale HE, Brown JS. 2011. Protective contributions against invasive *Streptococcus pneumoniae* pneumonia of antibody and Th17-cell responses to nasopharyngeal colonisation. *PLoS One* 6:e25558. <https://doi.org/10.1371/journal.pone.0025558>
69. Adams MH, Roe AS. 1945. A partially defined medium for cultivation of pneumococcus. *J Bacteriol* 49:401–409. <https://doi.org/10.1128/jb.49.4.401-409.1945>
70. Albersheim P, Nevins DJ, English PD, Karr A. 1967. A method for the analysis of sugars in plant cell-wall polysaccharides by gas-liquid chromatography. *Carbohydr Res* 5:340–345. [https://doi.org/10.1016/S0008-6215\(00\)80510-8](https://doi.org/10.1016/S0008-6215(00)80510-8)
71. Kakehi K, Honda S. 1988. Silyl ethers of carbohydrates, p 43–85. In *Analysis of carbohydrates by GLC and MS*. CRC Press.
72. Harris PJ, Henry RJ, Blakeney AB, Stone BA. 1984. An improved procedure for the methylation analysis of oligosaccharides and polysaccharides. *Carbohydr Res* 127:59–73. [https://doi.org/10.1016/0008-6215\(84\)85106-x](https://doi.org/10.1016/0008-6215(84)85106-x)
73. Sweet DP, Shapiro RH, Albersheim P. 1975. Quantitative analysis by various g.l.c. response-factor theories for partially methylated and partially ethylated alditol acetates. *Carbohydr Res* 40:217–225. [https://doi.org/10.1016/S0008-6215\(00\)82604-X](https://doi.org/10.1016/S0008-6215(00)82604-X)
74. Vialle S, Sepulcri P, Dubayle J, Talaga P. 2005. The teichoic acid (c-polysaccharide) synthesized by *Streptococcus pneumoniae* serotype 5 has a specific structure. *Carbohydr Res* 340:91–96. <https://doi.org/10.1016/j.carres.2004.10.021>