

RESEARCH

Open Access



Streptococcus pneumoniae serotype 33H: a novel serotype with frameshift mutations in the acetyltransferase gene *wciG*

Sam Manna^{1,2,3*}, Belinda D. Ortika¹, Joel P. Werren⁴, Casey L. Pell¹, Ilche Gjuroski⁵, Stephanie W. Lo^{6,7}, Jason Hinds^{8,9}, Odgerel Tundev¹⁰, Eileen M. Dunne¹¹, Bradford D. Gessner¹¹, Fiona M. Russell^{1,2}, E. Kim Mulholland^{1,12}, Tuya Mungun¹⁰, Claire von Mollendorf^{1,2}, Stephen D. Bentley⁷, Markus Hilty^{4†}, Neil Ravenscroft^{13†} and Catherine Satzke^{1,2,3†}

Abstract

Background *Streptococcus pneumoniae* (the pneumococcus) is a leading cause of community-acquired pneumonia. Pneumococci are categorised into serotypes, based on the type of capsular polysaccharide produced, which has important implications for virulence, vaccine impact and global surveillance. Recently, we identified a novel serotype, which we named 33G, that is comprised of an O-acetylated hexasaccharide repeat unit. In this study, we report and describe variants of 33G, designated 33G-like, which we isolated from the nasopharynx of two adults hospitalised with pneumonia in Mongolia.

Methods Serological comparison of 33G and 33G-like pneumococci were conducted by Quellung serotyping. Genetic analysis of the capsular polysaccharide loci was performed using whole genome sequencing. Polysaccharide composition was determined using ¹H nuclear magnetic resonance.

Results By Quellung serotyping, 33G pneumococci type as both 10B and 33B whereas 33G-like pneumococci type as both 10B and 33F. Genomic analysis of the capsular polysaccharide locus revealed 33G-like loci are identical to 33G, except for frameshift mutations in the *wciG* gene which encodes an acetyltransferase responsible for the O-acetylation of beta-galactofuranose (β -Gal f) in the capsular polysaccharide repeat unit. We constructed an artificial 33G-like by deleting *wciG* in a 33G strain and confirmed this gene was responsible for the serological differences between 33G and 33G-like pneumococci. Lastly, ¹H nuclear magnetic resonance confirmed the O-acetylation present in the 33G polysaccharide is absent in the 33G-like polysaccharide.

Conclusions Here, we have provided serological, genetic and biochemical evidence that the 33G-like capsule differs to 33G and all other pneumococcal serotypes, meeting the requirements to be designated as a new serotype, which we have named 33H.

[†]Markus Hilty, Neil Ravenscroft and Catherine Satzke are Co-senior authors.

*Correspondence:
Sam Manna
sam.manna@mcri.edu.au

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Keywords *Streptococcus pneumoniae*, Capsule, Serotypes, vaccines

Background

Streptococcus pneumoniae (the pneumococcus) is a bacterial pathogen of global importance and a major causative agent of pneumonia as well as other diseases including sepsis and meningitis [1]. Surrounding the bacterial cell is a polysaccharide capsule, with over 100 biochemically distinct capsule types ('serotypes') reported to date [2, 3]. Pneumococcal serotype plays an important role in virulence and is the basis for current vaccine formulations.

Novel serotypes or molecular variants of existing serotypes have the potential to reduce the accuracy of serotyping data used to inform decisions around vaccine introduction or maintenance [4–6]. Furthermore, novel serotypes may be more virulent or transmissible. For example, discovered in the last 15 years, serotypes 11E and 35D are both reported to be more invasive than related members of serogroups 11 and 35 [7–9]. Therefore, it is essential we monitor and investigate these serotypes as they emerge so that prevention strategies can be adapted.

Recently, we reported a novel serotype of pneumococcus detected in Mongolia [2]. The 33G capsular polysaccharide (*cps*) locus is likely derived from a series of recombination events with several other serotypes and streptococcal species [2]. To date, over 25 isolates of this serotype have been identified across three countries [2, 10, 11]. The 33G capsular repeat unit consists of an O-acetylated hexasaccharide with no side chains. When typed by serological methods (Quellung reaction or latex agglutination), 33G pneumococci react with typing sera such that the interpretation yields two serotyping results (10B and 33B) [2].

During our investigation of serotype variants across the Asia-Pacific region, we isolated novel variants (33G-like). The 33G-like isolates exhibit similar but distinct properties to serotype 33G, and also differ to all other known serotypes. Here, we determined the serological, genetic and biochemical basis of these differences and provide evidence that 33G-like meets the criteria to be designated a new serotype, which we have named 33H.

Methods

Human ethics approval

See declarations section.

Pneumococcal identification and serotyping

Nasopharyngeal swabs from the adult pneumonia surveillance program [12] were collected, stored and tested in accordance with the World Health Organization recommendations [13]. Screening of swabs for the presence

of pneumococci, as well as culture and serotyping by DNA microarray (Senti-SP v 1.5, BUGS Bioscience) and Quellung was conducted as described previously [14, 15].

Whole genome sequencing

Bacterial DNA was extracted from either pure isolates or plate sweeps cultured on horse blood agar or horse blood agar containing 5 µg/ml gentamicin (Thermo Fisher Scientific), respectively, using the QIAamp 96 DNA QIAcube HT kit [14]. For isolates, whole genome sequencing was conducted with the Illumina DNA Prep kit (Illumina) on the NovaSeq 6000 platform (2 × 150 bp paired end reads). Genome assembly and annotation was conducted as described previously [2]. For plate sweeps, deep sequencing was conducted as described above except that samples were sequenced to a greater depth (14 million reads). Verification of *wciG* mutations in the deep sequenced samples was conducted using snippy 4.6.0 (<https://github.com/tseemann/snippy>) against the 33G *cps* locus using default parameters. The 33H *cps* loci have been deposited in Genbank (accession numbers PQ281427 and PQ281428).

PCR and amplicon sequencing

PCR targeting the *wciG* gene was conducted using the KAPA HiFi HotStart ReadyMix (KAPA Biosystems). PCR reactions contained 1x KAPA HiFi HotStart ReadyMix, 1 µl of DNA template and 300 nM of primers flanking the *wciG* gene (for primer sequences, see Supplementary Table 1). PCRs were run under the following cycling conditions: 95 °C for 3 min, 35 cycles of 98 °C for 20 s, 60 °C for 15 s and 72 °C for 2 min, followed by a final extension at 72 °C for 3 min. Following cleanup using the Wizard SV gel and PCR cleanup system (Promega), the amplicons were sent to the Australian Genome Resource Facility for Sanger sequencing using the primers listed in Supplementary Table 1. Sequences were then assembled using cap3 10.2011 [16] and aligned to the reference *wciG* gene from 33G by Clustal Omega (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>).

Construction of the $\Delta wciG$ mutant

The $\Delta wciG$ mutant was constructed in PMP1612 (a known 33G strain characterised previously [2]) using an overlapping PCR product consisting of a Janus cassette (containing a kanamycin resistance marker) [17], flanked by ~3 kb of pneumococcal DNA sequences that are upstream and downstream of the *wciG* gene (5' and 3' homology arms, respectively). To construct the overlapping PCR product, each product was first amplified individually using the KAPA HiFi HotStart ReadyMix

(KAPA Biosystems). PCR reactions contained 1x KAPA HiFi HotStart ReadyMix, 1 µl of DNA template and 300 nM of primers (see Supplementary Table 1) and was run under the following cycling conditions; initial denaturation of 95 °C for 3 min, 35 cycles of 98 °C for 20 s, 65 °C for 15 s and 72 °C for 3 min followed by a final extension at 72 °C for 6 min. Overlapping PCR combining the amplicons was then conducted with the reaction containing 1x KAPA HiFi HotStart ReadyMix and 20ng of each of the three individual PCR products. This was run at initial denaturation of 95 °C for 3 min, 15 cycles of 98 °C for 20 s, 65 °C for 15 s and 72 °C for 8 min followed by a final extension at 72 °C for 16 min at which point 300 nM of the forward 5' homology arm primer and the reverse 3' homology arm primer were added. The reaction was then run under the following cycling conditions; initial denaturation of 95 °C for 3 min, 20 cycles of 98 °C for 20 s, 70 °C for 15 s and 72 °C for 8 min.

For transformation, PMP1612 was incubated in 500 µl CTM (1% [w/v] Casamino Acids, 0.5% [w/v] tryptone, 0.5% [w/v] NaCl, 1% [w/v] yeast extract, 16 µM K₂HPO₄, 0.2% [w/v] glucose, 150 µg/ml glutamine) and 55 ng CSP-2 competence stimulating peptide for 10 min at 37 °C with 5% CO₂. Following incubation, 10 µl of overlapping PCR product was added and the culture was incubated at 32 °C for 30 min. The culture was then moved to 37 °C with 5% CO₂ and incubated for 4 h before being cultured on horse blood agar plates supplemented with 500 µg/ml of kanamycin.

NMR analysis

Extraction, purification and ¹H NMR analysis of 33G and 33G-like polysaccharides was performed as described previously [2]. Briefly, 1D ¹H NMR of polysaccharides was recorded at 500 MHz at 25 °C using a Bruker Neo 500 MHz NMR spectrometer. Signals from low molecular weight species (e.g. solvent, amino acids and residuals from sample processing) were removed from the proton spectra using the DOSY Bruker programme ledbpg2s1d.

Results

Identification of 33G-like pneumococci in Mongolia

In Mongolia, we conducted pneumococcal carriage surveillance in adults hospitalised with pneumonia from 2019 to 2022. The participants in this program did not receive the 13-valent pneumococcal conjugate vaccine (PCV13), which was introduced into the infant immunisation schedule in 2016 [12]. Nasopharyngeal swabs were collected and screened for pneumococci by qPCR targeting the *lytA* gene. Culture positive samples were then serotyped by DNA microarray. Interestingly, DNA microarray identified two samples containing a presumptive serotype 33G, which DNA microarray designates as '35A/10B-like'. However, when these pneumococci were isolated and serotyped by Quellung reaction, they did not type as 33G as was expected. When typed by Quellung, 33G pneumococci yield two serotype results; 10B and 33B (including positive reactions with factor sera 10b, 10d and 33f) [2]. Instead, these pneumococcal isolates (PMP1615 and PMP1623, Table 1) typed as 10B and 33F (including positive reactions with factor sera 10b, 10d and 33b, Table 2). We designated these variants '33G-like'. To date, no other 33G-like were identified in our other studies including in over 9,500 nasopharyngeal swabs from healthy children [18, 19] as well as children hospitalised with pneumonia [20] in Mongolia, nor from the over 18,000 swabs from five other countries we have tested previously [2].

Genetic analysis of the 33G-like *cps* locus

We next conducted whole genome sequencing on the two 33G-like isolates. Each isolate belonged to a different multi-locus sequence type and global pneumococcal sequencing cluster as inferred from pubMLST [22, 23] and popPUNK [24], respectively (ST6318 and GPSC687 for PMP1615, ST2754 and GPSC230 for PMP1623). The 33G pneumococci identified in Mongolia previously belong to these same lineages ($n=3$ for ST2754 and GPSC230, $n=17$ for ST6318 and GPSC687) [2].

We next compared the *cps* locus from each 33G-like isolate to the 33G *cps* locus. The 33G-like and 33G *cps* loci were identical except that both 33G-like isolates had

Table 1 Pneumococcal strains used in this study

Strain	Serotype	GPSC ^a	Source/Description	Year	Reference
PMP1615	33G-like	687	Nasopharyngeal isolate from a participant (aged 30–39 years) from the Sukhbaatar district in Ulaanbaatar, Mongolia, hospitalised with pneumonia	2020	This study
PMP1623	33G-like	230	Nasopharyngeal isolate from a participant (aged 40–49 years) from the Sukhbaatar district in Ulaanbaatar, Mongolia, hospitalised with pneumonia	2020	This study
PMP1612	33G	687	Nasopharyngeal isolate from a participant (aged 30–39 years) from the Chingeltei district in Ulaanbaatar, Mongolia, hospitalised with pneumonia. Used as a control 33G strain for this study	2020	[2]
Δ <i>wciG</i>	Artificial 33G-like	687	<i>wciG</i> mutant constructed in PMP1612 background	N/A	This study

^a Global Pneumococcal Sequencing Cluster [21]

Table 2 Quellung reaction profile of representative 33G and 33G-like Pneumococci with '+' and '-' denoting a positive and negative reaction with typing Sera from SSI diagnostica, respectively

	Sera	33G	33G-like
First pool	A	-	-
	B	-	-
	C	-	-
	D	-	-
	E	+	+
	F	-	-
	G	-	-
	H	-	-
	I	-	-
Second pool	P	-	-
	Q	-	-
	R	-	-
	S	+	+
Group	T	+	+
	10	+	+
Group 10 factors	33	+	+
	10b	+	+
	10d	+(weak) ^a	+(weak) ^a
Group 33 factors	10f	-	-
	33b	-	+
	33e	-	-
	33f	+	-
	6a	-	-
	20b	-	-
Serotype result		10B and 33B	10B and 33F

^a Reaction is weak with one 1 μ l loopful of 10d factor sera. A stronger reaction was obtained using two 1 μ l loopfuls of sera

frameshift mutations in the *wciG* gene (Supplementary Figs. 1 and 2, Fig. 1). This gene encodes an acetyltransferase responsible for the O-acetylation of the galactofuranose (Gal_f) residue in the polysaccharide repeat unit [25–27]. PMP1615 has a single nucleotide deletion in *wciG*, whereas PMP1623 has two frameshift mutations

(a single nucleotide insertion and an 11 bp deletion in *wciG*), (Table 3). These mutations are predicted to result in the generation of premature stop codons and therefore *wciG* is unlikely to encode a functional acetyltransferase in these 33G-like isolates (Supplementary Figs. 3 and 4). All mutations were verified by traditional Polymerase Chain Reaction (PCR) amplification and Sanger sequencing of *wciG* from these isolates (Supplementary Fig. 5).

The same mutations were also detected when we re-cultured the nasopharyngeal samples and conducted traditional PCR targeting *wciG* using DNA extracted from a sweep of bacterial growth from the culture plate (Supplementary Fig. 6). We also conducted deep sequencing on the DNA extracted from the plate sweep. For the sample from which PMP1615 is derived, 99.3% (278/280) of reads covering position 313 of *wciG* matched the mutant allele observed in the isolate. Similarly, for the sample from which PMP1623 was isolated, 100% of reads covering positions 163 (298/298) and 766 (141/141) of *wciG* matched the mutant allele in the isolate.

To verify that the mutations in *wciG* were responsible for the serological differences observed in the Quellung reaction, we deleted *wciG* in PMP1612 (a 33G isolate) to create an artificial 33G-like strain (Supplementary Figs. 7 and 8). Consistent with the 33G-like pneumococci, $\Delta wciG$ typed as both 10B and 33F by Quellung whereas its parent 33G strain (PMP1612) typed as both 10B and 33B, which is consistent with the reaction profiles shown in Table 2. These data confirm the serological differences between 33G and 33G-like pneumococci are attributable to *wciG*.

Elucidation of 33G-like polysaccharide repeat unit structure

To verify the lack of O-acetylated Gal_f in the 33G-like capsule, the polysaccharide from both 33G-like pneumococcal isolates (PMP1615 and PMP1623), a 33G isolate (PMP1612) and the artificial 33G-like ($\Delta wciG$)

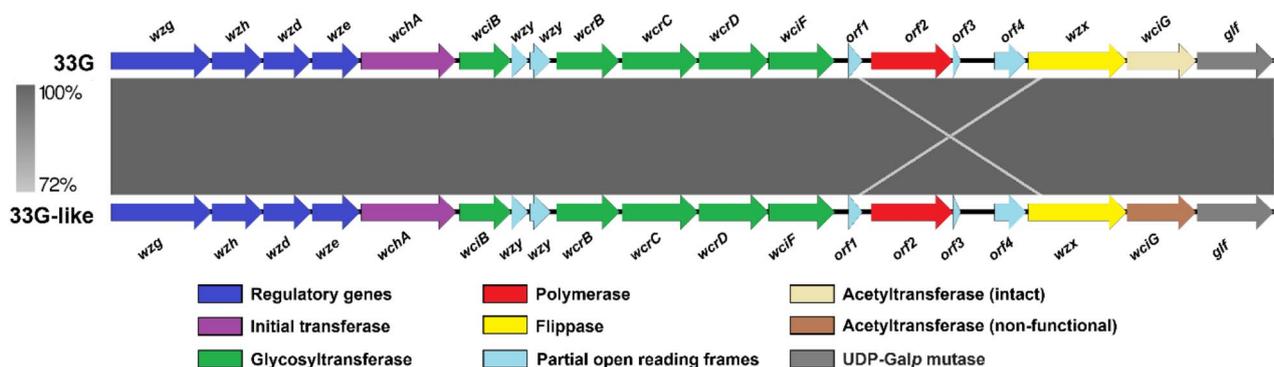


Fig. 1 Schematic comparison of the 33G and 33G-like *cps* loci. Shading indicates DNA sequence identity (%). The light grey cross indicates a related sequence between *orf1* and *wxz*, which may be a remnant of the recombination event that gave rise to the 33G *cps* locus. Image was created using Easyfig version 2.2.5 [28]. For full DNA sequence alignments, see Supplementary Figs. 1 and 2

Table 3 Frameshift mutations in *wciG* of 33G-like Pneumococci relative to *wciG* in 33G

Isolate	Gene size (bp) ^a	Mutation	Nucleotide position ^b	Amino acid change	Amino acid length ^c	Predicted consequence
PMP1615	998	Single nucleotide deletion (A)	313	Amino acid 105 change from Ile to Leu	105	Frameshift generating premature stop codon at codon 106
PMP1623	989	Single nucleotide insertion (T) Deletion of 11 bp (ATTTCTCACT) ^d	163 766 ^d	Amino acid 55 change from Met to Tyr	95	Frameshift generating premature stop codon at codon 96

^a Size of in frame gene in 33G=999 bp. ^b Nucleotide positions are inferred from the snippy 4.6.0 tool (<https://github.com/tseemann/snippy>) and are approximate because all mutations are located in homopolymeric/ambiguous regions and therefore the exact position cannot be ascertained. This also applies to all DNA alignments in supplementary data. ^c full length of *WciG* in 33G=332 amino acids. ^d As the 11 bp deletion occurs within a 14 bp region flanked by 'ACT' on either end, it is plausible the deletion is ACTATTTCTC starting at position 763

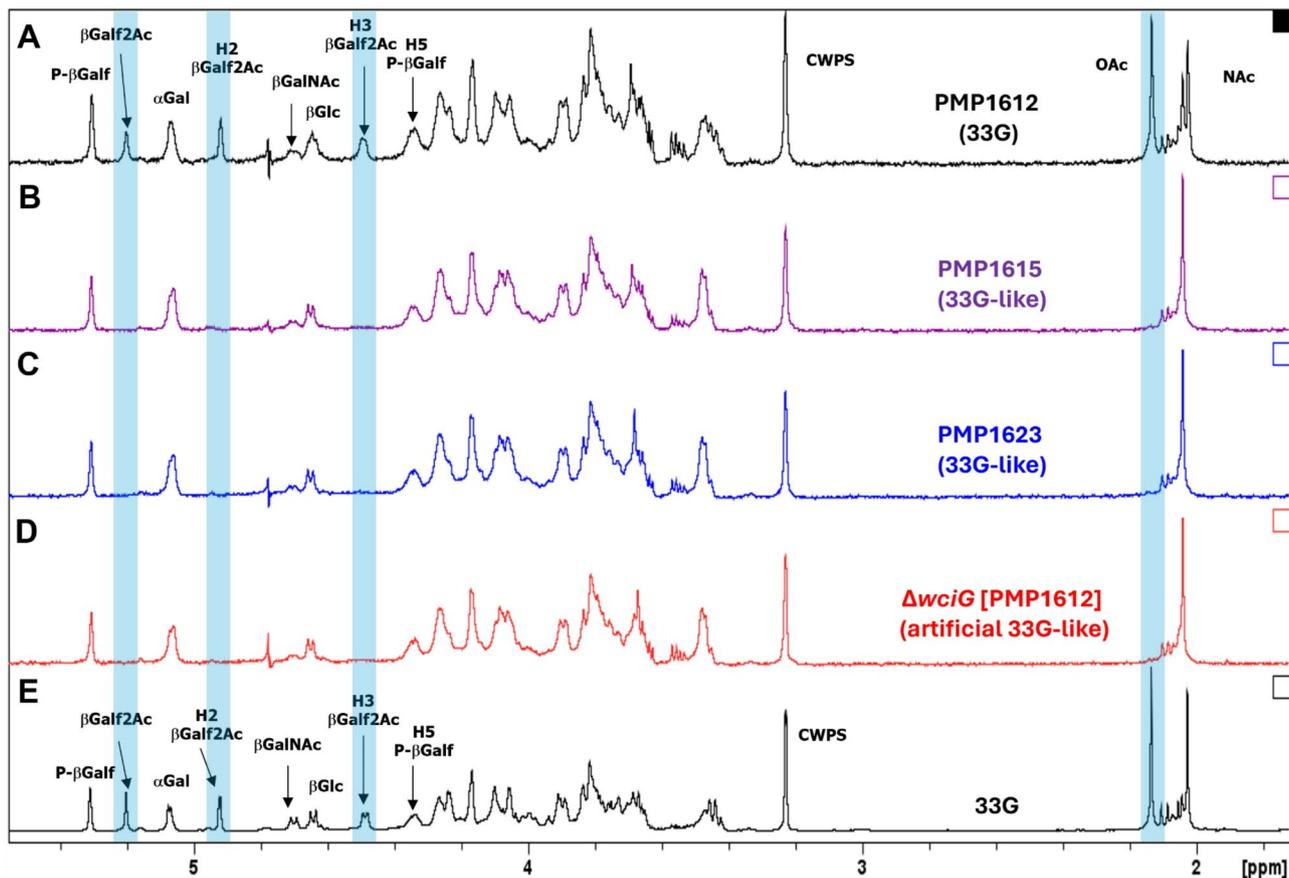


Fig. 2 ¹H diffusion-ordered spectroscopy (DOSY) NMR spectra of 33G (A), 33G-like (B-C) and artificial 33G-like (D) pneumococcal isolates identified in this study compared with the 33G polysaccharide spectrum reported previously (E) [2] at 500 MHz and 25 °C. CWPS=cell wall polysaccharide. Highlighted areas denote the location of acetylation peaks

[PMP1612]) strain were purified and analysed by ¹H nuclear magnetic resonance (NMR). PMP1612 polysaccharide (serotype 33G) exhibited the same NMR spectra as reported for this serotype previously [2], including an O-acetyl signal at 2.14 ppm and the diagnostic signal for the presence of O-acetylated *Gal*f (β-*Gal*f 2Ac) at 5.20, 4.93 and 4.48 ppm for H1, H2 and H3, respectively (Fig. 2). In contrast, all four of these acetylation signals were absent in the 33G-like isolates and in the artificial *wciG* mutant strain (Fig. 2). This biochemical analysis,

taken together with the genetic and serological differences to 33G and other previously described serotypes, confirm that 33G-like is a novel pneumococcal serotype, which we hereby name '33H' (Fig. 3).

Discussion

Previously, we characterised isolates from Mongolia that expressed a novel capsule and were therefore designated a new serotype, which we named 33G [2]. Now, we describe two new isolates in Mongolia that we designated

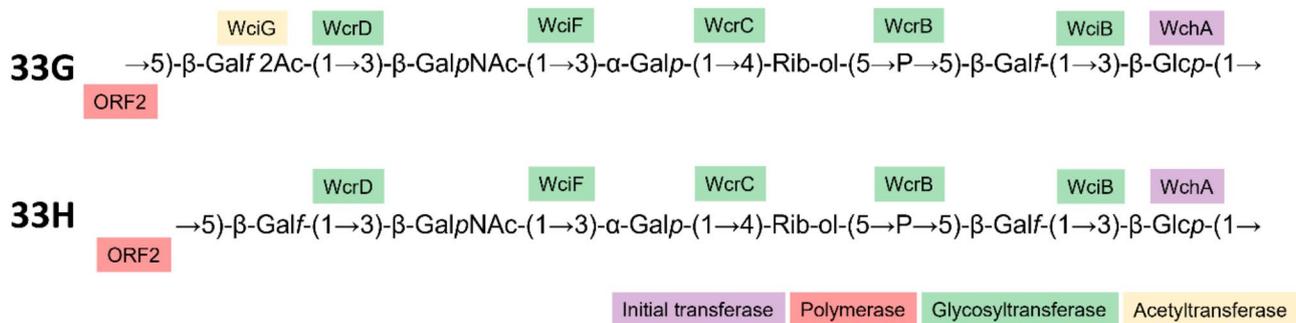


Fig. 3 Polysaccharide repeat unit structure of 33G [2] and the new serotype 33H (33G-like) identified in this study. Enzymatic assignments have been designated based on the literature [29]

33G-like. The serological properties of the 33G-like pneumococci differed to the traditional 33G pneumococci as indicated by different reactions to the group 33 factor sera. Genetic analysis confirmed that in contrast with 33G, 33G-like pneumococci have frameshift mutations in *wciG*. The *wciG* gene in the *cps* locus encodes an acetyltransferase responsible for the O-acetylation of Galf [26, 27]. Biochemical analysis confirmed the absence of acetylated Galf in the 33G-like capsule repeat unit. Taken together, these data satisfy the criteria for 33G-like to be designated as a novel serotype, which we have named 33H.

Novel pneumococcal serotypes resulting from small mutations in genes within the *cps* locus have been described previously. The most recent example is serotype 33E, which has a *cps* locus and capsule structure identical to 33G, except for a frameshift mutation in the *wciE* glycosyltransferase gene resulting in the lack of the α-Galp branch [26]. Serotypes 11E, 33F and 9A have mutations in the *wciE* gene, which encodes an acetyltransferase, but otherwise have identical *cps* loci and capsule structures to 11A, 33A and 9V, respectively [30–32]. Serotype 15C is identical to 15B, except for mutations in the *wciZ* acetyltransferase gene [33]. The closest analogy to 33G and 33H may be serotypes 35B and 35D, where 35D capsules lack O-acetylation due to mutations in *wciG* compared with 35B [34]. Interestingly, the *wciG* genes from serotypes 35B and 33G are over 85% identical (Supplementary Fig. 9) and the same two single nucleotide mutations observed in *wciG* in our 33H isolates have also been reported in the *wciG* gene of 35D isolates [7]. Both mutations occur in homopolymeric regions, which are susceptible to slipped-strand mispairing leading to the insertion or deletion of a single nucleotide [4, 33, 35].

Acetyl groups on capsular polysaccharides are immunodominant epitopes and therefore are readily recognised by the immune system [36, 37]. It has therefore been proposed that there may be selective pressure on O-acetylated serotypes to lose this modification. The loss of acetylation is associated with increased immune evasion [7, 38]. For 11A/11E and 35B/35D, there is a

disproportionate distribution of these serotypes isolated in carriage versus invasive pneumococcal disease, with the acetyltransferase mutant serotypes more commonly associated with invasive pneumococcal disease and acetyltransferase proficient serotypes more commonly identified in carriage [7–9]. These observations have led to a within-host microevolutionary model where it is proposed that the acetyltransferase proficient serotypes 11A and 35B show greater fitness in carriage and transmission but reduced fitness during invasive disease. In contrast, acetyltransferase mutant serotypes 11E and 35D would be more fit in a disease state but are likely an evolutionary dead-end because they are less likely to be transmitted from sites such as the blood [8]. Interestingly, this may not be the case for 33G/33H since the 33H isolates described in this study (PMP1615 and PMP1623) were isolated from the nasopharynx of pneumonia patients. Additionally, deep sequencing of plate sweeps of bacteria cultured directly from the nasopharyngeal swab only detected 33H with over 99% of sequence reads matching the mutated *wciG* allele (33H). Although we cannot exclude within-host evolution, the deep sequencing results suggest it is more likely that these pneumococci were already 33H at the time of acquisition. Future studies should focus on the association between acetylation status and invasiveness for 33G/33H.

The accuracy of methods used for pneumococcal serotyping is essential for generating reliable data for disease surveillance and/or measuring vaccine impact. Unfortunately, most molecular approaches for serotyping are currently unable to distinguish 33H from 33G. This is because most of the *cps* loci are identical, except for the mutations in the *wciG* gene, of which the site/type can vary by strain. However, distinguishing 33G and 33H with a molecular approach could be undertaken using PCR amplification and sequencing of the *wciG* gene, or by whole genome sequencing and manual interrogation of the *cps* locus for *wciG* mutations. Additionally, it may be possible to create a custom database in whole genome sequencing serotyping tools such as PneumoCaT or SeroBA, where a screen for *wciG* could be incorporated

to determine if the gene is in frame, similar to screens already conducted for other serotypes that contain frameshifted capsule genes [39, 40].

Our study has identified 33H in nasopharyngeal swabs in patients hospitalised with pneumonia in Mongolia. Interestingly, one of the putative 33G invasive isolates from South Africa (GPS_ZA_887, ENA sample accession SAMEA2553998, isolated from the blood of a 5 month old child) identified in the Global Pneumococcal Sequencing project [21] contains a single T insertion in *wciG* at nucleotide position 558, resulting in a frameshift and formation of a premature stop codon, suggesting this isolate is serotype 33H. Therefore, evidence indicates that 33H is present in at least two countries and can cause invasive disease. Future studies should focus on understanding the epidemiology of 33H in other countries. Lastly, given 33H reacts with the same pool, group and factor sera as 33F, it is evident that there is cross-reactivity between these serotypes. Importantly, serotype 33F is a component of recent vaccine formulations including PCV15 and PCV20 [41]. Thus, future studies should explore the potential of these vaccines to provide cross-protection against 33H.

Conclusion

Here we provide genetic, serological and biochemical evidence that the 33G-like capsular polysaccharide differs to the related serotype 33G, as well as all other known pneumococcal serotypes. Taken together, these data satisfy the criteria to designate 33G-like as a new capsular serotype, hereby named 33H. Future work investigating 33H including its geographical distribution and potential for enhanced immune evasion over 33G, will shed light on the evolution of this serotype as well as improve our understanding of drivers mediating the emergence of new serotypes.

Abbreviations

cps	Capsular polysaccharide locus
GPSC	Global pneumococcal sequencing cluster
MLST	Multi-locus sequence type
GalF	Galactofuranose
PCR	Polymerase chain reaction
NMR	Nuclear magnetic resonance
GalF 2Ac	O-acetylated galactofuranose
DOSY	Diffusion-ordered spectroscopy
CWPS	Cell wall polysaccharide

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41479-025-00162-2>.

Supplementary Material 1

Acknowledgements

We thank all the participants and staff involved in the adult pneumonia surveillance study in Mongolia. We also thank Laura Boelsen for assistance

with deep sequencing as well as the Translational Genomics Unit (MCRI) for next generation sequencing services. Thanks to Ashleigh Wee-Hee and other MCRI staff for technical assistance.

Author contributions

Conceptualisation: SM, CS. Methodology and study design: SM, JH, OT, EMD, BDG, FMR, EKM, TM, CVM, MH, NR, CS. Data analysis: SM, BDO, CLP, MH, NR, CS. Experimental investigation: SM, BDO, JPW, CLP, IG, SL, JH, SDB, MH, NR. Funding acquisition: SM, FMR, EKM, CVM, CS. Writing of the original manuscript draft: SM, CS. Review and editing of the manuscript for submission: SM, BDO, JPW, CLP, IG, SL, JH, OT, EMD, BDG, FMR, EKM, TM, CVM, SDB, MH, NR, CS.

Funding

This study was funded by a Robert Austrian Research Award in Pneumococcal Vaccinology awarded to Dr Manna from ISPPD (funded by Pfizer) as well as the National Health and Medical Research Council (NHMRC) Centre of Research Excellence for Pneumococcal Disease Control in the Asia-Pacific (GNT1 196415). The adult pneumonia surveillance study in Mongolia was funded by a Pfizer clinical research collaboration agreement with the Murdoch Children's Research Institute (MCRI) for which MCRI was the study sponsor. The work at MCRI was supported by the Victorian Government's Operational Infrastructure Support Program. Funders were not involved in the collection and analysis of data nor the decision to publish.

Data availability

The 33H cps loci have been deposited in Genbank (accession numbers PQ281427 and PQ281428). All other data generated during this study are included in the published article.

Declarations

Ethics approval and consent to participate

The 33G-like isolates described in this study were derived from a hospital-based pneumonia surveillance program of adults in Mongolia, which involved collection of nasopharyngeal swabs for pneumococcal detection and serotyping. Further details of the study protocol have been described previously [12]. Study approval was obtained from the National Ethics Committee for Health Research in Mongolia as well as the Royal Children's Hospital Human Research Ethics Committee in Australia (HREC number 38045B). Written informed consent was acquired from all participants prior to study enrolment.

Consent for publication

Not applicable.

Competing interests

CVM, EKM, TM, BDG, and CS 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. CS and EKM are investigators on a Merck Investigator Studies Program grant funded by MSD for a study unrelated to this work. SM, CVM, and CS have received honoraria (Pfizer (CS, SM, and CVM) and MSD (CS)) for presentations at symposia or attendance at expert advisory meetings unrelated to this study. MH is an investigator on a research grant funded by Pfizer for a study unrelated to this work. EMD and BDG are employed by Pfizer and may hold Pfizer stock or stock options. JH is co-founder and shareholder of BUGS Bioscience Ltd., a not-for-profit spin-out company of St George's, University of London.

Author details

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

²Department of Paediatrics, 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 Chemistry and Biochemistry, University of Bern, Bern, Switzerland

⁶Milner Center for Evolution, Life Sciences Department, University of Bath, Bath, United Kingdom

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

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

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

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

¹¹Pfizer Vaccines, Collegeville, Pennsylvania, United States of America

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

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

Received: 2 October 2024 / Accepted: 20 February 2025

Published online: 25 March 2025

References

- Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, et al. Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type B disease in children in the era of conjugate vaccines: global, regional, and National estimates for 2000–15. *Lancet Glob Heal*. 2018;6:e744–57.
- Manna S, Werren JP, Ortika BD, Bellich B, Pell CL, Nikolaou E et al. *Streptococcus pneumoniae* serotype 33G: genetic, serological, and structural analysis of a new capsule type. *Microbiol Spectr*. 2024;12.
- Yu J, Ravenscroft N, Davey P, Liyanage R, Lorenz O, Kuttel MM, et al. New Pneumococcal serotype 20 C is a WciG O-acetyltransferase deficient variant of canonical serotype 20B. *Microbiol Spectr*. 2024. <https://doi.org/10.1128/SPECTRUM.02443-24>.
- Manna S, Ortika BD, Dunne EM, Holt KE, Kama M, Russell FM et al. A novel genetic variant of *Streptococcus pneumoniae* serotype 11A discovered in Fiji. *Clin Microbiol Infect*. 2018;24:428.e1–428.e7.
- Manna S, Dunne EM, Ortika BD, Pell CL, Kama M, Russell FM, et al. Discovery of a *Streptococcus pneumoniae* serotype 33F capsular polysaccharide locus that lacks *WcjE* and contains a *WcyO* pseudogene. *PLoS ONE*. 2018;13:e0206622.
- Manna S, Spry L, Wee-Hee A, Ortika BD, Boelsen LK, Batinovic S, et al. Variants of *Streptococcus pneumoniae* serotype 14 from Papua new Guinea with the potential to be mistyped and escape vaccine-induced protection. *Microbiol Spectr*. 2022;10:e0152422.
- Lo SW, Gladstone RA, Van Tonder AJ, Hawkins PA, Kwambana-Adams B, Cornick JE et al. Global distribution of invasive serotype 35D *Streptococcus pneumoniae* isolates following introduction of 13-valent Pneumococcal conjugate vaccine. *J Clin Microbiol*. 2018;56.
- Nahm MH, Yu J, Calix JJ, Ganaie F. Ficolin-2 lectin complement pathway mediates capsule-specific innate immunity against invasive Pneumococcal disease. *Front Immunol*. 2022;13.
- Calix JJ, Dagan R, Pelton SI, Porat N, Nahm MH. Differential occurrence of *Streptococcus pneumoniae* serotype 11E between asymptomatic carriage and invasive Pneumococcal disease isolates reflects a unique model of pathogen microevolution. *Clin Infect Dis*. 2012;54:794–9.
- van Tonder AJ, Gladstone RA, Lo SW, Nahm MH, du Plessis M, Cornick J, et al. Putative novel *Cps* loci in a large global collection of Pneumococci. *Microb Genomics*. 2019;5:e000274.
- Mostowy RJ, Croucher NJ, De Maio N, Chewapreecha C, Salter SJ, Turner P, et al. Pneumococcal capsule synthesis locus *Cps* as evolutionary hotspot with potential to generate novel serotypes by recombination. *Mol Biol Evol*. 2017;34:2537–54.
- von Mollendorf C, Ulziibayar M, Gessner BD, Do LAH, Nguyen CD, Beavon R et al. 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*. 2021;21.
- Satzke C, Turner P, Virolainen-Julkunen A, Adrian PV, Antonio M, Hare KM, et al. Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the world health organization Pneumococcal carriage working group. *Vaccine*. 2013;32:165–79.
- Satzke C, Dunne EM, Porter BD, Klugman KP, Mulholland EK, Vidal JE, et al. The PneuCarriage project: A multi-centre comparative study to identify the best serotyping methods for examining Pneumococcal carriage in vaccine evaluation studies. *PLoS Med*. 2015;12:e1001903.
- Habib M, Porter BD, Satzke C. Capsular serotyping of *Streptococcus pneumoniae* using the quellung reaction. *J Vis Exp*. 2014;84:e51208.
- Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res*. 1999;9:868.
- Sung CK, Li H, Claverys JP, Morrison DA. An *RpsL* cassette, Janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl Environ Microbiol*. 2001;67:5190–6.
- von Mollendorf C, Dunne EM, La Vincente S, Ulziibayar M, Suuri B, Luvsantseren D, et al. Pneumococcal carriage in children in Ulaanbaatar, Mongolia before and one year after the introduction of the 13-valent Pneumococcal conjugate vaccine. *Vaccine*. 2019;37:4068–75.
- von Mollendorf C, Mungun T, Ulziibayar M, Skoko P, Boelsen L, Nguyen C, et al. Effect of Pneumococcal conjugate vaccine six years post-introduction on Pneumococcal carriage in Ulaanbaatar, Mongolia. *Nat Commun*. 2024;15:6577.
- Chan J, Mungun T, Batsaivan P, Ulziibayar M, Suuri B, Otgonbayar D, et al. 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 Heal - West Pac*. 2021;15:100231.
- Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, et al. International genomic definition of Pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine*. 2019;43:338–46.
- Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*. 2018;3:124.
- Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology*. 1998;144:3049–60.
- Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and flexible bacterial genomic epidemiology with poppunk. *Genome Res*. 2019;29:304–16.
- Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitz E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 Pneumococcal serotypes. *PLoS Genet*. 2006;2:0262–9.
- Ganaie FA, Saad JS, Lo SW, McGee L, van Tonder AJ, Hawkins PA, et al. Novel Pneumococcal capsule type 33E results from the inactivation of glycosyltransferase *WciE* in vaccine type 33F. *J Biol Chem*. 2023;299:105085.
- Lin FL, Vinogradov E, Deng C, Zeller S, Phelan L, Green BA, et al. Structure Elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 33 C, 33D, and revised structure of serotype 33B. *Carbohydr Res*. 2014;383:97–104.
- Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics*. 2011;27:1009.
- Aanensen DM, Mavroidi A, Bentley SD, Reeves PR, Spratt BG. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol*. 2007;189:7856.
- Calix JJ, Nahm MH. A new Pneumococcal serotype, 11E, has a variably inactivated *WcjE* gene. *J Infect Dis*. 2010;202:29–38.
- Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kaltoft MS, Reeves PR, et al. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol*. 2007;189:7841–55.
- Calix JJ, Saad JS, Brady AM, Nahm MH. Structural characterization of *Streptococcus pneumoniae* serotype 9A capsule polysaccharide reveals role of Glycosyl 6-O-acetyltransferase *WcjE* in serotype 9V capsule biosynthesis and immunogenicity. *J Biol Chem*. 2012;287:13996–4003.
- Van Selm S, Van Cann LM, Kolkman MAB, Van der Zeijst BAM, Van Putten JPM. Genetic basis for the structural difference between *Streptococcus pneumoniae* serotype 15B and 15 C capsular polysaccharides. *Infect Immun*. 2003;71:6192–8.
- Geno KA, Saad JS, Nahm MH. Discovery of novel Pneumococcal serotype 35D, a natural *WciG*-deficient variant of serotype 35B. *J Clin Microbiol*. 2017;55:1416–25.
- Pericone CD, Bae D, Shchepetov M, McCool T, Weiser JN. Short-sequence tandem and nontandem DNA repeats and endogenous hydrogen peroxide production contribute to genetic instability of *Streptococcus pneumoniae*. *J Bacteriol*. 2002;184:4392.

36. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, et al. Pneumococcal capsules and their types: past, present, and future. *Clin Microbiol Rev.* 2015;28:871–99.
37. Rajam G, Carlone GM, Romero-Steiner S. Functional antibodies to the O-acetylated Pneumococcal serotype 15B capsular polysaccharide have low cross-reactivities with serotype 15 C. *Clin Vaccine Immunol.* 2007;14:1223–7.
38. Brady AM, Calix JJ, Yu J, Geno KA, Cutter GR, Nahm MH. 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.* 2014;210:1155–65.
39. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, et al. Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ.* 2016;4:e2477.
40. Epping L, van Tonder AJ, Gladstone RA, Bentley SD, Page AJ, Keane JA. SeroBA: rapid high-throughput serotyping of *Streptococcus pneumoniae* from whole genome sequence data. *Microb Genomics.* 2018;4:e000186.
41. Kobayashi M, Farrar JL, Gierke R, Britton A, Childs L, Leidner AJ, et al. Use of 15-Valent Pneumococcal conjugate vaccine and 20-Valent Pneumococcal conjugate vaccine among U.S. Adults: updated recommendations of the advisory committee on immunization Practices - United States, 2022. *MMWR Morb Mortal Wkly Rep.* 2022;71:109–17.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.