



# Global Distribution of Invasive Serotype 35D *Streptococcus pneumoniae* Isolates following Introduction of 13-Valent Pneumococcal Conjugate Vaccine

Stephanie W. Lo,<sup>a</sup> Rebecca A. Gladstone,<sup>a</sup> Andries J. van Tonder,<sup>a</sup> Paulina A. Hawkins,<sup>b,c</sup> Brenda Kwambana-Adams,<sup>d</sup> Jennifer E. Cornick,<sup>e,f</sup> Shabir A. Madhi,<sup>g,h</sup> Susan A. Nzenze,<sup>g,h</sup> Mignon du Plessis,<sup>i,j</sup> Rama Kandasamy,<sup>k</sup> Philip E. Carter,<sup>l</sup> Özgen Köseoglu Eser,<sup>m</sup> Pak Leung Ho,<sup>n</sup> Naima Elmdaghri,<sup>o,p</sup> Sadia Shakoore,<sup>q</sup> Stuart C. Clarke,<sup>r</sup> Martin Antonio,<sup>d,s,t</sup> Dean B. Everett,<sup>e,u</sup> Anne von Gottberg,<sup>i,j</sup> Keith P. Klugman,<sup>b</sup> Lesley McGee,<sup>c</sup> Robert F. Breiman,<sup>b,v</sup> Stephen D. Bentley,<sup>a</sup> The Global Pneumococcal Sequencing Consortium

<sup>a</sup>Infection Genomics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

<sup>b</sup>Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

<sup>c</sup>Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

<sup>d</sup>Vaccines and Immunity Theme, Medical Research Council Unit The Gambia, Banjul, Fajara, The Gambia

<sup>e</sup>Malawi Liverpool Wellcome Trust Clinical Research Programme, Blantyre, Malawi

<sup>f</sup>Institute of Infection & Global Health, University of Liverpool, Liverpool, United Kingdom

<sup>g</sup>Medical Research Council: Respiratory and Meningeal Pathogens Research Unit, University of the Witwatersrand, Johannesburg, South Africa

<sup>h</sup>Department of Science and Technology/National Research Foundation: Vaccine Preventable Diseases, University of the Witwatersrand, Johannesburg, South Africa

<sup>i</sup>Centre for Respiratory Disease and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa

<sup>j</sup>School of Pathology, University of the Witwatersrand, Johannesburg, South Africa

<sup>k</sup>Oxford Vaccine Group, Department of Paediatrics, University of Oxford, and the NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom

<sup>l</sup>Institute of Environmental Science and Research Limited, Kenepuru Science Centre, Porirua, New Zealand

<sup>m</sup>Hacettepe University Faculty of Medicine, Department of Medical Microbiology, Ankara, Turkey

<sup>n</sup>Department of Microbiology and Carol Yu Centre for Infection, The University of Hong Kong, Queen Mary Hospital, Hong Kong, China

<sup>o</sup>Department of Microbiology, Faculty of Medicine and Pharmacy, Hassan II University of Casablanca, Casablanca, Morocco

<sup>p</sup>Bacteriology-Virology and Hospital Hygiene Laboratory, University Hospital Centre Ibn Rochd, Casablanca, Morocco

<sup>q</sup>Department of Pathology and Laboratory Medicine and Department of Paediatrics and Child Health, The Aga Khan University, Karachi, Pakistan

<sup>r</sup>Faculty of Medicine and Institute of Life Sciences, University of Southampton, Southampton, United Kingdom

<sup>s</sup>Microbiology and Infection Unit, Warwick Medical School, Warwick, United Kingdom

<sup>t</sup>London School of Hygiene and Tropical Medicine, London, United Kingdom

<sup>u</sup>University of Edinburgh, The Queens Medical Research Institute, Edinburgh, United Kingdom

<sup>v</sup>Emory Global Health Institute, Emory University, Atlanta, Georgia, USA

**ABSTRACT** A newly recognized pneumococcal serotype, 35D, which differs from the 35B polysaccharide in structure and serology by not binding to factor serum 35a, was recently reported. The genetic basis for this distinctive serology is due to the presence of an inactivating mutation in *wciG*, which encodes an O-acetyltransferase responsible for O-acetylation of a galactofuranose. Here, we assessed the genomic data of a worldwide pneumococcal collection to identify serotype 35D isolates and understand their geographical distribution, genetic background, and invasiveness potential. Of 21,980 pneumococcal isolates, 444 were originally typed as se-

**Received** 9 February 2018 **Returned for modification** 27 February 2018 **Accepted** 24 April 2018

**Accepted manuscript posted online** 2 May 2018

**Citation** Lo SW, Gladstone RA, van Tonder AJ, Hawkins PA, Kwambana-Adams B, Cornick JE, Madhi SA, Nzenze SA, du Plessis M, Kandasamy R, Carter PE, Eser ÖK, Ho PL, Elmdaghri N, Shakoore S, Clarke SC, Antonio M, Everett DB, von Gottberg A, Klugman KP, McGee L, Breiman RF, Bentley SD, The Global Pneumococcal Sequencing Consortium. 2018. Global distribution of invasive serotype 35D *Streptococcus pneumoniae* isolates following introduction of 13-valent pneumococcal conjugate vaccine. *J Clin Microbiol* 56:e00228-18. <https://doi.org/10.1128/JCM.00228-18>.

**Editor** Daniel J. Diekema, University of Iowa College of Medicine

**Copyright** © 2018 Lo 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/).

Address correspondence to Stephanie W. Lo, [sl28@sanger.ac.uk](mailto:sl28@sanger.ac.uk), or Stephen D. Bentley, [sdb@sanger.ac.uk](mailto:sdb@sanger.ac.uk).

rotype 35B by PneumoCaT. Analysis of the *wciG* gene revealed 23 isolates from carriage ( $n = 4$ ) and disease ( $n = 19$ ) with partial or complete loss-of-function mutations, including mutations resulting in premature stop codons ( $n = 22$ ) and an in-frame mutation ( $n = 1$ ). These were selected for further analysis. The putative 35D isolates were geographically widespread, and 65.2% (15/23) of them was recovered after the introduction of pneumococcal conjugate vaccine 13 (PCV13). Compared with serotype 35B isolates, putative serotype 35D isolates have higher invasive disease potentials based on odds ratios (OR) (11.58; 95% confidence interval [CI], 1.42 to 94.19 versus 0.61; 95% CI, 0.40 to 0.92) and a higher prevalence of macrolide resistance mediated by *mefA* (26.1% versus 7.6%;  $P = 0.009$ ). Using the Quellung reaction, 50% (10/20) of viable isolates were identified as serotype 35D, 25% (5/20) as serotype 35B, and 25% (5/20) as a mixture of 35B/35D. The discrepancy between phenotype and genotype requires further investigation. These findings illustrated a global distribution of an invasive serotype, 35D, among young children post-PCV13 introduction and underlined the invasive potential conferred by the loss of O-acetylation in the pneumococcal capsule.

**KEYWORDS** 35D, PCV, novel serotype, whole-genome sequencing

*Streptococcus pneumoniae* (pneumococcus) is an important human pathogen that causes pneumonia, bacteremia, and meningitis. In 2015, >330,000 deaths globally in children of <5 years old were estimated to have been caused by pneumococci (1). The polysaccharide capsule of pneumococcus, which has almost 100 serological variants (serotypes), is a major virulence factor (2, 3). Pneumococcal conjugate vaccines (PCVs) targeting up to 13 serotypes have gradually been introduced into 139 countries since the early 2000s (<http://view-hub.org/viz/>). Simultaneously, a proportional increase in nonvaccine serotypes, such as serotype 35B, has been reported in various countries (4).

Recently, a serotype 35B variant, 35D, was identified in four pneumococcal isolates in Australia (5) and two in the United States (2, 6), all of which had an inactivating mutation in *wciG*, which encodes an O-acetyltransferase responsible for O-acetylation of a galactofuranose. Nuclear magnetic resonance (NMR) analysis on a single isolate representing this novel pneumococcal serotype verified that the serotype 35D capsule lacked O-acetylation but that it was otherwise identical to serotype 35B (2). Serologically, serotype 35D is distinct from serotype 35B by consistently not binding to factor serum 35a, but it displays variable reactivity to group 35 antiserum (2, 5, 6). *WciG* functionality has been shown to be the determinant of factor serum 35a recognition (2, 7).

Presence and absence of O-acetylation is one of the mechanisms for generating diversity in capsular structure, as shown by other serotype pairs such as 9V/9A (O-acetylation mediated by *WciE*) (8), 11A/11E (*WcjE*) (8), 15B/15C (*WciZ*) (8), 33A/33F (*WcjE*) (9), and 35C/42 (*WciG*) (7). It is noteworthy that the O-acetyl group in the capsular repeat unit is important for innate immune recognition (10) and is the target of vaccine-elicited antibodies (11). Loss of O-acetylation in serotype 11E is predicted to assist pneumococci in evading host immune and vaccine response and has been suggested to occur during invasive disease after initial colonization with the serotype 11A strain expressing an O-acetylated form of capsule (12). The role of loss of O-acetylation in pneumococcal survival during invasion among the other serotype pairs has remained unknown due to the rarity of serotypes 9A, 33A, and 42 for comparisons, and by the difficulty in differentiation between serotype 15B and 15C.

Although the serological profile and biochemical structure of serotype 35D have been described, there has not been an opportunity to comprehensively study this serotype across geographies and clinical considerations. Here, we assessed the genomic data on serotype 35D isolates from a worldwide pneumococcal collection to understand this serotype's geographical distribution, genetic background and potential invasiveness.

## MATERIALS AND METHODS

We retrospectively determined serotypes of 21,980 assembled pneumococcal genomes from the Global Pneumococcal Sequencing (GPS) project ( $n = 16,575$ ; May 2017; <http://www.pneumogenet/gps/>) and a compiled data set ( $n = 5,405$ ) by van Tonder et al. (13). DNA extraction was performed on a pure overnight culture derived from a single colony. Sequencing was performed on the Illumina HiSeq platform to produce paired-end reads of either 75 (in 2010 and 2011), 100 (in 2013 and 2014), or 125 (in 2015 and 2016) bp in length. *In silico* serotype was determined using the whole-genome sequence (WGS)-based serotyping method PneumoCaT (14). As the current version of PneumoCaT does not distinguish serotype 35D from serotype 35B, all samples that were initially typed as serotype 35B were included in this study. To differentiate these two serotypes, nucleotide sequences of *wciG* were extracted from the assembled genome sequences and aligned to a reference sequence of 35B *wciG* (GenBank accession number [KX021817](https://www.ncbi.nlm.nih.gov/nuclot/KX021817)) described by Geno et al. (2) using CLUSTALW (15). Nonsense and frameshift mutations that led to premature stop codons and in-frame insertions/deletions in *wciG* were predicted to result in complete loss of function and reduction of function of the WciG protein, respectively. Isolates with these mutations were *in silico* typed as serotype 35D, and their phenotypic serotype were determined by the Quellung reaction, tested on an overnight culture derived from a single colony (16). Phylogenetic analysis was performed on all serotype 35B and 35D isolates by constructing a maximum likelihood tree using RAxML v.8.2.X (17) based on single-nucleotide polymorphism sites extracted from a core gene alignment with Roary v.3.6.1 (18). An empirical odds ratio for invasive disease due to serotype 35B and 35D was calculated based on a pneumococcal collection of 3,333 randomly selected carriage ( $n = 1,260$ ) and disease ( $n = 2,073$ ) isolates from children aged <2 years old, collected during the pre-PCV ( $n = 1,691$ ), post-PCV7 ( $n = 678$ ), and post-PCV13 ( $n = 964$ ) eras using a previously described method (19). For each country, the random selection was carried out from a collection of disease isolates collected via laboratory-based surveillance and carriage isolates collected via cohort studies using the following criteria: 50% of the isolates represented the pre-PCV period ( $\leq 1$  year before) and 50% the post-PCV period ( $\geq 2$  years after primary and  $\geq 1$  after subsequent PCVs). The randomly selected collection in this study included 67 different serotypes plus nontypeable pneumococci. These isolates were collected in South Africa (carriage  $n = 721$ , disease  $n = 1,047$ ), Malawi (carriage  $n = 336$ , disease  $n = 60$ ), and the Gambia (carriage  $n = 1,016$ , disease  $n = 153$ ). Isolates from other locations in the GPS data set were either not randomly selected or consisted of only disease or only carriage isolates and thus could not be used to calculate odds ratios. Susceptibility to chloramphenicol, co-trimoxazole, erythromycin, penicillin, and tetracycline were predicted by the identification of resistant determinants in the assembled genomes using previously described pipelines (20–22). The epidemiological and phylogenetic data can be interactively visualized and analyzed online by using the Microreact tool ([https://microreact.org/project/GPS\\_serotype\\_35B\\_35D](https://microreact.org/project/GPS_serotype_35B_35D)).

## RESULTS AND DISCUSSION

Of 21,980 assembled pneumococcal genomes from the Global Pneumococcal Sequencing (GPS) project ( $n = 16,575$ ; May 2017) and a compiled data set ( $n = 5,405$ ) by van Tonder et al. (13), 444 isolates from disease ( $n = 173$ ), carriage ( $n = 270$ ), and an unknown source ( $n = 1$ ) were originally typed as serotype 35B by PneumoCaT (5). The *wciG* alignment revealed that 78.6% (349/444) of isolates were identical to the serotype 35B reference, 8.3% (37/444) had silent mutations, 7.9% (35/444) had missense mutations, 3.4% (15/444) had frameshift mutations, 1.6% (7/444) had nonsense mutations, and 0.2% (1/444) had an in-frame insertion. All frameshift mutations led to a premature stop codon that disrupted the coding region of *wciG*. Given that the latter three types of mutations lead to reduced function or a complete loss of function of WciG, the 23 isolates were designated serotype 35D (Table 1). The Quellung reaction of 20 viable isolates showed that 50% (10/20) were serologically typed as serotype 35D, 25% (5/20) as serotype 35B, and 25% (5/20) as a mixture of serotype 35B and 35D (Table 2). In all discrepant cases, we examined the *cps* locus sequences in an attempt to identify any gene loss and mixed *wciG* alleles. The *cps* locus region shared the same capsular genes with the serotype 35D reference (GenBank accession number [KY084476](https://www.ncbi.nlm.nih.gov/nuclot/KY084476)), and the mutations in *wciG* were supported by at least  $42\times$  depth of reads (median,  $80\times$ ; range,  $42\times$  to  $143\times$ ) with 100% consistency. The discrepancy between phenotype and genotype could be due to (i) our inability to capture the serotype diversity in a clinical sample, since the bacterial cultures subjected to DNA extraction and Quellung testing were derived from a single colony that could be different between experiments, and (ii) the possible interconvertibility between serotype 35B and 35D during bacterial culture *in vitro*. In all five isolates that were both positive and negative to antisera fs35a under one microscope (Table 2), the mutations in *wciG* were either a 1-bp insertion or deletion that occurred after a 6- to 7-bp homopolymer, highlighting the possibility of interconversion between serotype 35B and 35D during DNA replication. Metagenomic analysis

**TABLE 1** Genetic diversity of inactivating mutations in *wciG* of 29 serotype 35D *S. pneumoniae* isolates from the Global Pneumococcal Sequencing (GPS) project ( $n = 23$ ) and previous studies ( $n = 6$ )

Type of mutation ( $n$ ) or <i>wciG</i> nucleotide mutation	Clonal complex and/or sequence type ( $n$ )	Isolation:			Reference or source	
		Geographical location(s) ( $n$ )	Yr(s) ( $n$ )	Site(s) <sup>e</sup> ( $n$ )		
Frameshift mutation (18) <sup>a</sup>						
86_87insG	6	CC156 (2), CC558 (2), CC198 (1), CC9813 (1)	Malawi (2), New Zealand (1), Senegal (1), South Africa (1), United States (1)	2006 (1), 2011 (1), 2012 (2), 2015 (2)	CSF (3), blood (2), joint pus (1)	GPS
914_929del_16bp	2	CC558	South Africa, United States	2012 (1), 2013 (1)	CSF (1), blood (1)	GPS
162_163insT	2	CC558	United States	2004 (1), 2007 (1)	Nasopharynx (2)	GPS <sup>d</sup>
92_93insC	1	CC198	The Gambia	2013	Blood	GPS
705_706insT	1	CC156	Malawi	2015	CSF	GPS
86delG	1	CC156	Cameroon	2012	CSF	GPS
312delA	1	CC198	The Gambia	2009	Nasopharynx	GPS
382_385_del_4bp	1	CC9813	South Africa	2012	CSF	GPS
306_307insA	1	CC198	Australia	2016	Unknown	5
36delA	1	CC558	Australia	2015	Unknown	5
663_696del_34bp	1	CC452	Australia	2016	Unknown	5
In-frame deletion/insertion (3)						
792_968del_177bp <sup>b</sup>	1	CC156	USA	2015	Blood (2)	6
755_808del_54bp <sup>b</sup>	1	CC558	Australia	2016	Unknown	5
523_524ins_15bp	1	CC558	USA	2009	Blood	GPS
Nonsense mutation (7)						
C220T	2	CC156, ST373	Nepal, South Africa	2013 (1), 2014 (1)	CSF (1), nasopharynx (1)	GPS
T732G	2	CC198	The Gambia	2014 (2)	CSF (1), blood (1)	GPS
C104A	1	CC558	USA	2012	Blood	GPS
C323A	1	CC558	USA	2012	Blood	GPS
T434G	1	CC198	The Gambia	2009	Lung aspirate	GPS
Missense mutations (1)						
G533A, G679A <sup>c</sup>	1	Unknown	USA	Unknown	Unknown	(2)

<sup>a</sup>All frameshift mutations resulted in a premature stop codon.

<sup>b</sup>The in-frame deletion rendered *WciG*, an acetyltransferase, nonfunctional. It was evidenced by the serological profiles reported by Chochua et al. (6) and Staples et al. (5).

<sup>c</sup>The resulting amino acid changes were R178K and A227T. The substitution led to a nonfunctional *WciG*, confirmed by serological test and NMR spectroscopic analysis.

<sup>d</sup>These two isolates were reported in a previous study by Croucher et al. (23) and *in silico* serotype was updated as serotype 35D in this study.

<sup>e</sup>CSF, cerebrospinal fluid.

of clinical samples to snapshot the serotype diversity and investigation into the interconvertibility of serotype 35B and 35D will potentially explain the discrepancy between the phenotypes and genotypes observed in this study. Considering the limitation of this study and our recent understanding of the genetic basis that differentiates serotype 35B and 35D (2, 6, 7), the nonsilent mutations detected in *wciG* in this study strongly indicated the presence of serotype 35D pneumococci in the sample. Thus, the 23 *in silico* serotype 35D isolates were selected for further analysis.

The mutation patterns of *wciG* among the *in silico* serotype 35D isolates were diverse. The *wciG* mutation patterns in the 23 serotype 35D isolates were different from those of the 6 serotype 35D isolates reported previously (2, 5, 6). In total, there were 20 mutation patterns observed in 29 serotype 35D isolates from 10 countries across four continents (Table 1). The most common naturally deficient *WciG* was due to 86\_87insG, which occurred within a 6-bp homopolymeric stretch of guanine. It was first observed in an isolate from Malawi in 2006, prior to the introduction of PCV7, and was also found in isolates from Senegal in 2011, South Africa and the United States in 2012, and New Zealand in 2015. Isolates with this mutation were sporadically distributed on the phylogenetic tree (Fig. 1), suggesting that the mutations had arisen independently on multiple occasions. The convergence of mutations may imply that this site is a mutational hot spot.

The majority of serotype 35D isolates belonged to clonal complex 558 (CC558) ( $n = 9$ ), CC198 ( $n = 6$ ), and CC156 ( $n = 5$ ), which were primarily associated with serotype 35B (6, 24, 25). The CC558 and CC156 lineages accounted for most of the increase in serotype 35B isolates after the introduction of PCV13 in the United States (6), while

**TABLE 2** Serological profiles of 29 serotype 35D *S. pneumoniae* isolates from the Global Pneumococcal Sequencing (GPS) project (n = 23) and previous studies (n = 6) tested by Quellung reaction

Strain name	Country	CC	Yr	w <i>al</i> G mutation(s) <sup>f</sup>	Pool G	Type		Group	Antiserum				Phenotypic serotype	Reference or source	
						29	42		fs35a	fs35b	fs35c	fs29b			fs42a
3431-06	USA	N/A	N/A	G533A, G679A	+	ND	ND	35	-	-	+	-	-	35D	2
165471	Australia	CC198	2016	306_307insA	+	+	+	35	-	-	+	-	-	35D	5
SAMDU-00005305	Australia	CC558	2015	36delA	+	+	+	35	-	-	+	-	-	35D	5
16549	Australia	CC452	2016	663_696del_34bp	+	+	+	35	-	-	+	-	-	35D	5
16535	Australia	CC558	2016	755_808del_54bp	+	+	+	35	-	-	+	-	-	35D	5
CH20275	USA	CC156	2015	792_968del_177bp	+	ND	ND	35	-	-	+	-	-	35D	6
3025	USA	CC558	2007	162_163insT	+	+	+	35	+	+	+	+	+	35B	GPS <sup>e</sup>
GPS_US_2010209945_R1	USA	CC558	2004	162_163insT	+	+	+	35	+	+	+	+	+	35B	GPS <sup>e</sup>
GPS_GM_1130	The Gambia	CC198	2009	523_524ins_15bp	+	+	+	35	+	+	+	+	+	35B	GPS
GPS_GM_1148	The Gambia	CC198	2014	T731G (L244*)	+	+	+	35	+	+	+	+	+	35B	GPS
GPS_ZA_2370	South Africa	CC9813	2012	382_385delATAT	+	+	+	35	-	-	+	-	-	35D	GPS
GPS_ZA_2636	South Africa	CC558	2013	914_929del_16bp	+	+	+	35	-	-	+	-	-	35D	GPS
2012215593	USA	CC558	2012	914_929del_16bp	+	+	+	35	-	-	+	-	-	35D	GPS
2012215608	USA	CC558	2012	C104A (S35*)	+	+	+	35	-	-	+	-	-	35D	GPS
GPS_ZA_2559	South Africa	CC156	2013	C220T (Q74*)	+	+	+	35	-	-	+	-	-	35D	GPS
GPS_NP_7242	Nepal	Singleton <sup>d</sup>	2014	C220T (Q74*)	+	ND	ND	35	-	-	+	-	-	35D	GPS
2012220613	USA	CC558	2012	C323A (S108*)	+	+	+	35	-	-	+	-	-	35D	GPS
2013208723	USA	CC558	2012	86_87insG	+	+	+	35	-	-	+	-	-	35D	GPS
GPS_MW_D38253_R1	Malawi	CC156	2006	86_87insG	+	+	+	35	-	-	+	-	-	35D	GPS
GPS_MW_BKR609	Malawi	CC156	2015	86_87insG	+	+	+	35	-	-	+	-	-	35D	GPS
PI0167	Senegal	CC198	2011	86_87insG	+	+	+	35	-	-	+	-	-	35D	GPS
GPS_NZ_155P0720	New Zealand	CC558	2013	86_87insG	+	+	ND	35	+	+	+	+	+	35B/D	GPS
GPS_ZA_2487	South Africa	CC9813	2012	86_87insG	+	+	+	35	+	+	+	+	+	35B/D	GPS
GPS_MW_BKR5WC	Malawi	CC156	2015	705_706insT	+	+	+	35	-	-	+	-	-	35B/D	GPS
PI0258	Cameroon	CC156	2012	86delG	+	+	+	35	+	+	+	+	+	35B/D	GPS
GPS_GM_0282	The Gambia	CC198	2013	92_93insC	ND <sup>a</sup>	ND	ND	35	ND	ND	ND	ND	ND	ND	GPS
GPS_GM_0600	The Gambia	CC198	2009	312delA	ND	ND	ND	35	ND	ND	ND	ND	ND	ND	GPS
GPS_GM_0320	The Gambia	CC198	2009	T434G (L145*)	ND	ND	ND	35	ND	ND	ND	ND	ND	ND	GPS

<sup>a</sup>ND, data not available.

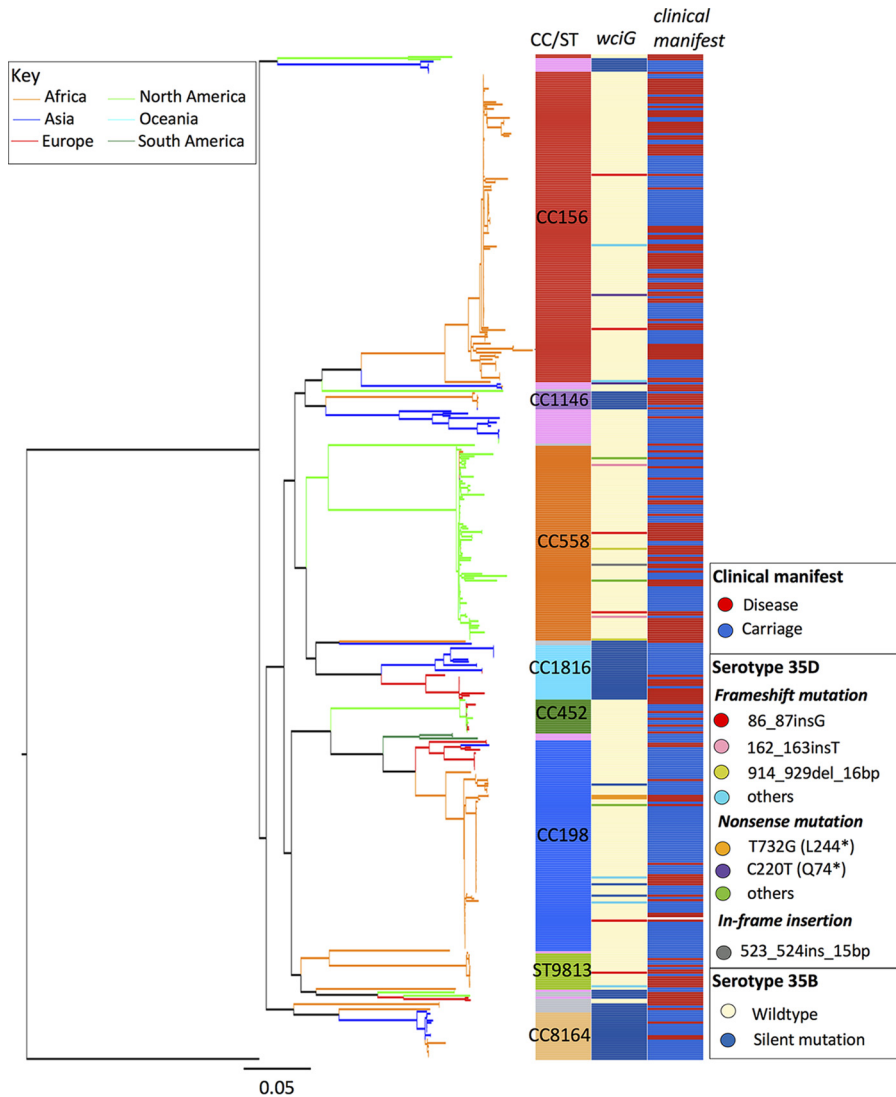
<sup>b</sup>Under the microscope, cells that were derived from a single-colony overnight culture showed both positive and negative to the antisera tested.

<sup>c</sup>This isolate was tested in two different laboratories and exhibited as both positive to antiserum fs35a in one laboratory and negative in another.

<sup>d</sup>Isolate GPS\_NP\_7242 belong to ST373, a singleton that does not belong to any clonal complex.

<sup>e</sup>These two isolates were reported in a previous study by Croucher et al. (23) and *in silico* serotype was updated as serotype 35D in this study.

<sup>f</sup>\*, stop codon.



**FIG 1** Maximum likelihood phylogenetic tree was constructed using 56,848 single-nucleotide polymorphisms (SNPs) extracted from a 1.02-Mb codon alignment of 1,141 core genes from 444 serotype 35B and 35D *S. pneumoniae* isolates. The tree is colored according to the geographic location of each sample's isolation. This analysis used an unrelated nontypeable isolate as the outgroup on which to root the tree. Clonal complex (CC) and mutations in *wciG* are shown to the right of the tree. Singleton sequence types and minor CCs with <5 isolates in this study are indicated in pink and gray, respectively.

CC198 is the major serotype 35B lineage in the Gambia (unpublished data). Based on a high-resolution single-nucleotide polymorphic tree (Fig. 1), serotype 35D pneumococci emerged among closely related serotype 35B isolates within different clusters. Together with the unrelated mutations observed in *wciG*, this strongly indicated that serotype 35B is the progenitor of serotype 35D.

Compared with serotype 35B isolates, serotype 35D isolates were more likely to be recovered from sterile anatomical sites, including cerebrospinal fluid (CSF;  $n = 9$ ), blood ( $n = 8$ ), lung aspirate ( $n = 1$ ), and joint aspirate ( $n = 1$ ), than among carriage isolates ( $n = 4$ ) (82.6% [19/23] versus 36.7% [154/420];  $P < 0.001$  by Fisher's exact test). Based on a larger pneumococcal collection ( $n = 3,333$ ) randomly selected from the GPS project database, the empirical odds ratio (OR) for invasive disease due to serotype 35D is 11.58 (95% confidence interval, 1.42 to 94.19), whereas the OR for serotype 35B is 0.61 (95% CI, 0.40 to 0.92). The increased invasive capacity in serotype 35D strains could be a result of evasion of the immune response targeting the capsule O-acetyl group. The observation in serotype 35B/35D coincides with a previous study on serotype 11A/11E,

**TABLE 3** The prevalence of serotype 35B and 35D *S. pneumoniae* from South Africa ( $n = 1,768$ ), the Gambia ( $n = 1,169$ ) and Malawi ( $n = 396$ ) in each vaccine period

Vaccine period <sup>a</sup>	No. of isolates (%) for serotype:		Odds ratio (95% confidence interval) for serotype:	
	serotype 35B	serotype 35D	35B	35D
Pre-PCV ( $n = 1691$ )	36 (2.12)	1 (0.06)	Baseline	Baseline
Post-PCV7 ( $n = 678$ )	12 (1.77)	0	0.83 (0.4–1.6)	
Post-PCV13 ( $n = 964$ )	69 (7.16)	7 (0.73)	3.54 (2.4–5.4) <sup>b</sup>	12.36 (1.5–100.6) <sup>b</sup>

<sup>a</sup>Based on the year of PCV introduction, we grouped each year of collection into three categories, as follows: pre-PCV period (years when no conjugated vaccine was used and the year of PCV7 introduction); post-PCV7 (the second year of PCV7 introduction until the year when a higher-valency PCV was introduced); and post-PCV13 (the second year of PCV13 introduction until the end of the study year). PCV7 was introduced in South Africa and the Gambia in 2009; PCV13 was introduced in South Africa, the Gambia, and Malawi in 2011.

<sup>b</sup> $P$  value < 0.05.

in which serotype 11E strains with a loss or reduced amount of acetylation in the capsule were found to be significantly associated with invasive pneumococcal disease (12, 26). The emergence of serotype 35D is likely explained by Calix et al.'s hypothesis (12) that pneumococcal capsule structure undergoes microevolution during progression from carriage to infection in response to divergent selection pressure in early mucosal colonization compared to later in a sterile site. This model of microevolution needs to be further investigated by characterizing the serotype dynamic over the development of invasive disease *in vivo*.

Compared with the pre-PCV era, the prevalence of serotype 35D has not increased more than serotype 35B after the introduction of PCV13. (OR, 12.36; 95% CI, 1.5 to 100.6 versus OR, 3.54; 95% CI, 2.4 to 5.4; Table 3) in the randomly selected pneumococcal collection. A large proportion of 35D isolates (65.2%, 15/23) were collected after the rollout of PCV13. The post-PCV introduction isolates were all invasive isolates and were recovered in six countries (Cameroon, Malawi, New Zealand, South Africa, the Gambia, and the United States), highlighting that this invasive serotype is present in the residual pneumococcal population worldwide and could potentially be an example of serotype replacement.

Among the 23 serotype 35D isolates, 87.0% (20/23) had at least one resistance determinant conferring resistance to commonly used antibiotics, including penicillin (65.2%, 15/23), erythromycin (30.4%, 7/23), co-trimoxazole (21.7%, 5/23), and tetracycline (4.3%, 1/23). Similar to the previous studies on serotype 35B (6, 24), the penicillin-resistant isolates in this study were predominantly CC558 (60.0%, 9/15), followed by CC156 (35.7%, 5/15) and a singleton of sequence type 73 (ST373) (6.7%, 1/15). Macrolide resistance mediated by *mefA* was significantly higher in serotype 35D isolates than in serotype 35B isolates (Table 4). Five of six serotype 35D isolates harboring *mefA* were from the United States, where macrolides are recommended for use as an empirical therapy for pneumonia in children (27–29); they all belonged to CC558, a major contributor to penicillin resistance in the United States after introduction of PCV13 (24). Unlike the highly invasive but usually antibiotic-susceptible serotype 1, pneumococci expressing serotype 35B (lower-invasive capsule) are more likely to be commensal in the nasopharynx, which could allow them to acquire antibiotic resistance determinants

**TABLE 4** Antimicrobial resistant determinants in serotype 35B and 35D *S. pneumoniae* isolates from the Global Pneumococcal Sequencing (GPS) project

Antibiotic resistance determinant(s)	No. of isolates (%) for serotype:		$P$ value
	35B ( $n = 421$ )	35D ( $n = 23$ )	
<i>ermB</i>	3 (0.7)	1 (4.3)	0.192
<i>mefA</i>	32 (7.6)	6 (26.1)	0.009
<i>tetM</i>	36 (8.6)	1 (4.3)	0.710
<i>folA</i> I100L and <i>folP</i> insertion	140 (33.3)	5 (21.7)	0.361

via horizontal gene transfer from other nasopharyngeal bacteria; a subsequent switch to serotype 35D (high-invasive capsule) would then transform the antibiotic-resistant strain into a more virulent form.

The limitation of this study is that the carriage and disease isolates included for calculating the invasiveness index were sampled in different cities in each country; all isolates included were collected between 2007 and 2015 from children aged <2 years old. Ideally, the carriage and disease isolates should be geography-, time-, and age-matched. In this instance, we calculated ORs for invasiveness separately for each country. The ORs for invasive disease due to serotype 35B and 35D in the Gambia were 0.37 (95% CI, 0.09 to 1.56) and 20.3 (95% CI, 2.10 to 196.42), respectively. The ORs could not be calculated for invasive disease, as all serotype 35D isolates in South Africa and Malawi were from disease. The ORs for disease due to 35B in South Africa and Malawi were 0.68 (95% CI, 0.40 to 1.16) and 0.72 (95% CI, 0.11 to 2.15), respectively. The ORs by country were consistent with the ORs calculated from the combined data sets of all three countries. Another limitation was that the effects of an in-frame insertion of 15 bp and the missense mutations in *wciG* on the protein function have not been evaluated. Removing these samples from all comparisons of serotype 35B and 35D did not alter the conclusions drawn from the statistical analyses.

This study highlighted the global distribution of an invasive serotype, 35D, among young children in the post-PCV13 era and underlined the invasive potential conferred by the loss of O-acetylation in the pneumococcal capsule.

#### ACKNOWLEDGMENTS

This work is funded by the Bill and Melinda Gates Foundation (grant OPP1034556), the Wellcome Trust Sanger Institute, and the Centers for Disease Control and Prevention (Atlanta, GA).

We thank the Centre for Genomic Pathogen Surveillance (CGPS), Sanger Institute for their excellent visualization tool, Microreact. We gratefully acknowledge Bernard W. Beall and Cynthia Whitney for their careful and critical reading of our manuscript and insightful comments. We appreciate Stephen I. Pelton from Boston University School of Medicine and Public Health, who kindly provided us with two putative serotype 35D isolates. We thank Belabbes Houria, Idrissa Diawara, Khalid Zerouali, Mohamed Benbachir, and Houria Belabbes from Hassan II University of Casablanca and University Hospital Centre Ibn Rochd, Morocco, and Furqan Kabir and Shahida Qureshi from Aga Khan University, Pakistan, for sample and metadata collection. We also thank Julie Morgan from the Institute of Environmental Science and Research Limited, New Zealand, and Olga Hattingh from the National Institute for Communicable Disease, South Africa, for performing the Quellung test.

We declare that we have no additional conflicts of interest. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

#### REFERENCES

1. Wahl B, O'Brien KL, Greenbaum A, Liu L, Chu Y, Black R, Majumder A, Lukšić I, Nair H, McAllister D, Campbell H, Rudan I, Knoll M. 2016. Global burden of *Streptococcus pneumoniae* in children younger than 5 years in the pneumococcal conjugate vaccines (PCV) era: 2000–2015. International Symposium on Pneumococci and Pneumococcal Diseases, Glasgow, Scotland.
2. 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>.
3. 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>.
4. Balsells E, Guillot L, Nair H, Kyaw MH. 2017. Serotype distribution of *Streptococcus pneumoniae* causing invasive disease in children in the post-PCV era: a systematic review and meta-analysis. *PLoS One* 12:e0177113. <https://doi.org/10.1371/journal.pone.0177113>.
5. Staples M, Graham RM, Hicks V, Strachan J, Goncalves da Silva A, Peverall J, Wicks V, Jennison AV. 2017. Discovery of *Streptococcus pneumoniae* serogroup 35 variants in Australian patients. *Clin Microbiol Infect* 23:476–479. <https://doi.org/10.1016/j.cmi.2016.12.029>.
6. Chochua S, Metcalf BJ, Li Z, Walker H, Tran T, McGee L, Beall B. 2017. Invasive serotype 35B pneumococci including an expanding serotype switch lineage, United States, 2015–2016. *Emerg Infect Dis* 23:922–930. <https://doi.org/10.3201/eid2306.170071>.
7. Geno KA, Bush CA, Wang M, Jin C, Nahm MH, Yang J. 2017. *wciG* O-acetyltransferase functionality differentiates pneumococcal serotypes 35C and 42. *J Clin Microbiol* 55:2775–2784. <https://doi.org/10.1128/JCM.00822-17>.
8. 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>.
9. Spencer BL, Saad JS, Shenoy AT, Orihuela CJ, Nahm MH. 2017. Position of O-acetylation within the capsular repeat unit impacts the biological properties of pneumococcal serotypes 33A and 33F. *Infect Immun* 85: e00132-17. <https://doi.org/10.1128/IAI.00132-17>.
10. 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/jju195>.
11. 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/CVI.00184-07>.
12. Calix JJ, Dagan R, Pelton SI, Porat N, Nahm MH. 2012. 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* 54:794–799. <https://doi.org/10.1093/cid/cir953>.
13. van Tonder AJ, Bray JE, Quirk SJ, Haraldsson G, Jolley KA, Maiden MC, Hoffmann S, Bentley SD, Haraldsson A, Erlendsdottir 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>.
14. 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>.
15. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
16. Statens Serum Institut. 2015. *Streptococcus pneumoniae*: textbook in diagnosis, serotyping, virulence factors and enzyme-linked immunosorbent assay (ELISA) for measuring pneumococcal antibodies. Statens Serum Institut, Copenhagen, Denmark.
17. Liu K, Linder CR, Warnow T. 2011. RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. *PLoS One* 6:e27731. <https://doi.org/10.1371/journal.pone.0027731>.
18. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>.
19. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. 2003. Clonal relationships between invasive and carriage streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis* 187:1424–1432. <https://doi.org/10.1086/374624>.
20. Metcalf BJ, Chochua S, Gertz RE, Li Z, Walker H, Tran T, Hawkins PA, Glennen A, Lynfield R, Li Y, McGee L, Beall B. 2016. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. *Clin Microbiol Infect* 22: 1002.e1–1002.e8. <https://doi.org/10.1016/j.cmi.2016.08.001>.
21. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr, Walker H, Hawkins PA, Tran T, Whitney CG, McGee L, Beall BW. 2016. Penicillin-binding protein transpeptidase signatures for tracking and predicting  $\beta$ -lactam resistance levels in *Streptococcus pneumoniae*. *mBio* 7:e00756-16. <https://doi.org/10.1128/mBio.00756-16>.
22. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr, Walker H, Hawkins PA, Tran T, McGee L, Beall BW, Active Bacterial Core surveillance t. 2017. Validation of beta-lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. *BMC Genomics* 18:621. <https://doi.org/10.1186/s12864-017-4017-7>.
23. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* 45: 656–663. <https://doi.org/10.1038/ng.2625>.
24. Metcalf BJ, Gertz RE, Jr, Gladstone RA, Walker H, Sherwood LK, Jackson D, Li Z, Law C, Hawkins PA, Chochua S, Sheth M, Rayamajhi N, Bentley SD, Kim L, Whitney CG, McGee L, Beall B; Active Bacterial Core surveillance team. 2016. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. *Clin Microbiol Infect* 22: 60.e9–60.e29. <https://doi.org/10.1016/j.cmi.2015.08.027>.
25. Beall B, McEllistrem MC, Gertz RE, Jr, Boxrud DJ, Besser JM, Harrison LH, Jorgensen JH, Whitney CG; Active Bacterial Core Surveillance/Emerging Infections Program Network. 2002. Emergence of a novel penicillin-nonsusceptible, invasive serotype 35B clone of *Streptococcus pneumoniae* within the United States. *J Infect Dis* 186:118–122. <https://doi.org/10.1086/341072>.
26. Calix JJ, Brady AM, Du VY, Saad JS, Nahm MH. 2014. Spectrum of pneumococcal serotype 11A variants results from incomplete loss of capsule O-acetylation. *J Clin Microbiol* 52:758–765. <https://doi.org/10.1128/JCM.02695-13>.
27. Brogan TV, Hall M, Williams DJ, Neuman MI, Grijalva CG, Farris RW, Shah SS. 2012. Variability in processes of care and outcomes among children hospitalized with community-acquired pneumonia. *Pediatr Infect Dis J* 31:1036–1041.
28. Hersh AL, Shapiro DJ, Pavia AT, Shah SS. 2011. Antibiotic prescribing in ambulatory pediatrics in the United States. *Pediatrics* 128:1053–1061. <https://doi.org/10.1542/peds.2011-1337>.
29. Kronman MP, Hersh AL, Feng R, Huang YS, Lee GE, Shah SS. 2011. Ambulatory visit rates and antibiotic prescribing for children with pneumonia, 1994–2007. *Pediatrics* 127:411–418. <https://doi.org/10.1542/peds.2010-2008>.