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

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

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-
Protype 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.pneumogen.net/gps/index.html) and a cohort 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) 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 by the identification of resistant determinants in the assembled genomes using previously described pipelines (20–22). The epidemiological and 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-PCV (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 (≤1 year before) and 50% the post-PCV period (>2 years after primary and ≥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).

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), and the mutations in wciG were supported by at least 42× depth of reads (median, 80×; range, 42× to 143×) 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...
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 \textit{wciG} in this study strongly indicated the presence of serotype 35D pneumococci in the sample. Thus, the 23 \textit{in silico} serotype 35D isolates were selected for further analysis.

The mutation patterns of \textit{wciG} among the \textit{in silico} serotype 35D isolates were diverse. The \textit{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 \textit{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) (\textit{n} = 9), CC198 (\textit{n} = 6), and CC156 (\textit{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

\begin{table}[h]
\centering
\caption{Genetic diversity of inactivating mutations in \textit{wciG} of 29 serotype 35D \textit{S. pneumoniae} isolates from the Global Pneumococcal Sequencing (GPS) project (\textit{n} = 23) and previous studies (\textit{n} = 6)\label{tab:1}}
\begin{tabular}{|l|l|l|l|l|l|l|}
\hline
Type of mutation (\textit{n}) & Clonal complex and sequence type (\textit{n}) & Isolation & Geographical location(s) (\textit{n}) & Yr(s) (\textit{n}) & Site(s)\textsuperscript{a} (\textit{n}) & Reference or source \\
\hline
Frameshift mutation (18)\textsuperscript{a} & 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\textsuperscript{d} \\
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) & 1 & CC156 & USA & 2015 & Blood (2) & 6 \\
792_968del_177bp\textsuperscript{b} & 1 & CC558 & Australia & 2016 & Unknown & 5 \\
523_524ins_15bp & 1 & CC558 & USA & 2009 & Blood & GPS \\
Nonsense mutation (7) & 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) & 1 & Unknown & USA & Unknown & Unknown & (2) \\
\textit{G533A, G679A}\textsuperscript{c} & 1 & Unknown & USA & Unknown & Unknown & (2) \\
\hline
\textsuperscript{a}All frameshift mutations resulted in a premature stop codon. \\
\textsuperscript{b}The in-frame deletion rendered \textit{WciG}, an acetyltransferase, nonfunctional. It was evidenced by the serological profiles reported by Chochua et al. (6) and Staples et al. (5). \\
\textsuperscript{c}The resulting amino acid changes were R178K and A227T. The substitution led to a nonfunctional \textit{WciG}, confirmed by serological test and NMR spectroscopic analysis. \\
\textsuperscript{d}These two isolates were reported in a previous study by Croucher et al. (23) and \textit{in silico} serotype was updated as serotype 35D in this study. \\
\textsuperscript{e}CSF, cerebrospinal fluid.
\end{tabular}
\end{table}
### 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

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Country</th>
<th>CC</th>
<th>Yr</th>
<th>wcG mutation(s)*</th>
<th>Pool G</th>
<th>Type</th>
<th>Group 35</th>
<th>Antiserum</th>
<th>Phenotypic serotype</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>3431-06</td>
<td>USA</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>35D</td>
</tr>
<tr>
<td>165471</td>
<td>Australia</td>
<td>CC198</td>
<td>2016</td>
<td>306,307insA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>SAMDU-00005305</td>
<td>Australia</td>
<td>C558</td>
<td>2015</td>
<td>36delA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>16549</td>
<td>Australia</td>
<td>CC452</td>
<td>2016</td>
<td>663,696del_34bp</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>16535</td>
<td>Australia</td>
<td>C558</td>
<td>2016</td>
<td>755,808del_54bp</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>20152877</td>
<td>USA</td>
<td>CC156</td>
<td>2015</td>
<td>792,968del_177bp</td>
<td>ND</td>
<td>ND</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>CH2075</td>
<td>USA</td>
<td>C558</td>
<td>2007</td>
<td>162,163insT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B</td>
<td></td>
</tr>
<tr>
<td>3025</td>
<td>USA</td>
<td>C558</td>
<td>2004</td>
<td>162,163insT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B</td>
<td></td>
</tr>
<tr>
<td>GPS_US_2010209945_R1</td>
<td>USA</td>
<td>C558</td>
<td>2009</td>
<td>523,524ins_15bp</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B</td>
<td></td>
</tr>
<tr>
<td>GPS_GM_1130</td>
<td>The Gambia</td>
<td>CC198</td>
<td>2014</td>
<td>T731G (L244*)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B</td>
<td></td>
</tr>
<tr>
<td>GPS_GM_1148</td>
<td>The Gambia</td>
<td>CC198</td>
<td>2014</td>
<td>T731G (L244*)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B</td>
<td></td>
</tr>
<tr>
<td>GPS_ZA_2370</td>
<td>South Africa</td>
<td>C9813</td>
<td>2012</td>
<td>382,385delATAT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>GPS_ZA_2636</td>
<td>South Africa</td>
<td>C558</td>
<td>2013</td>
<td>914,929del_16bp</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>2012215593</td>
<td>USA</td>
<td>C558</td>
<td>2012</td>
<td>914,929del_16bp</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>201215608</td>
<td>USA</td>
<td>C558</td>
<td>2012</td>
<td>C104A (S35*)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>GPS_ZA_2559</td>
<td>South Africa</td>
<td>C156</td>
<td>2013</td>
<td>C220T (Q74*)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>GPS_NP_7242</td>
<td>Nepal</td>
<td>Singleton*</td>
<td>2014</td>
<td>C220T (Q74*)</td>
<td>+</td>
<td>ND</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>201220613</td>
<td>USA</td>
<td>C558</td>
<td>2012</td>
<td>C23A (S108*)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>2013208723</td>
<td>USA</td>
<td>C558</td>
<td>2012</td>
<td>86,87insG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>GPS_MW_D38253_R1</td>
<td>Malawi</td>
<td>C156</td>
<td>2006</td>
<td>86,87insG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>GPS_MW_BR609</td>
<td>Malawi</td>
<td>C156</td>
<td>2015</td>
<td>86,87insG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35D</td>
<td></td>
</tr>
<tr>
<td>PI0167</td>
<td>Senegal</td>
<td>Singleton*</td>
<td>2011</td>
<td>86,87insG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B/D</td>
<td></td>
</tr>
<tr>
<td>GPS_NZ_1550720</td>
<td>New Zealand</td>
<td>C558</td>
<td>2013</td>
<td>86,87insG</td>
<td>+</td>
<td>+</td>
<td>ND</td>
<td>+</td>
<td>35B/D</td>
<td></td>
</tr>
<tr>
<td>GPS_ZA_2487</td>
<td>South Africa</td>
<td>C9813</td>
<td>2012</td>
<td>86,87insG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B/D</td>
<td></td>
</tr>
<tr>
<td>GPS_MW_BR526C</td>
<td>Malawi</td>
<td>C156</td>
<td>2015</td>
<td>705,707insT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B/D</td>
<td></td>
</tr>
<tr>
<td>PI0258</td>
<td>Cameroon</td>
<td>C156</td>
<td>2012</td>
<td>86delG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>35B/D</td>
<td></td>
</tr>
<tr>
<td>GPS_GM_0028</td>
<td>The Gambia</td>
<td>C198</td>
<td>2013</td>
<td>92,93insC</td>
<td>ND*</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>35B/D</td>
<td></td>
</tr>
<tr>
<td>GPS_GM_0600</td>
<td>The Gambia</td>
<td>C198</td>
<td>2009</td>
<td>312delA</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>35B/D</td>
<td></td>
</tr>
<tr>
<td>GPS_GM_0320</td>
<td>The Gambia</td>
<td>C198</td>
<td>2009</td>
<td>T434G (L145*)</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>35B/D</td>
<td></td>
</tr>
</tbody>
</table>

*ND, data not available.
*Under the microscope, cells that were derived from a single-colony overnight culture showed both positive and negative to the antiserum tested.
*This isolate was tested in two different laboratories and exhibited as both positive to antiserum fs35a in one laboratory and negative in another.
*Isolate GPS_NP_7242 belong to ST37, a singleton that does not belong to any clonal complex.
*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.
*Stop codon.

This isolate was tested in two different laboratories and exhibited as both positive to antiserum fs35a in one laboratory and negative in another.
*Isolate GPS_NP_7242 belong to ST37, a singleton that does not belong to any clonal complex.
*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.
*Stop codon.
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.
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 predominately 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 3

<table>
<thead>
<tr>
<th>Vaccine period</th>
<th>No. of isolates (%) for serotype:</th>
<th>Odds ratio (95% confidence interval) for serotype:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>serotype 35B</td>
<td>serotype 35D</td>
</tr>
<tr>
<td>Pre-PCV (n = 1691)</td>
<td>36 (2.12)</td>
<td>1 (0.06)</td>
</tr>
<tr>
<td>Post-PCV7 (n = 678)</td>
<td>12 (1.77)</td>
<td>0</td>
</tr>
<tr>
<td>Post-PCV13 (n = 964)</td>
<td>69 (7.16)</td>
<td>7 (0.73)</td>
</tr>
</tbody>
</table>

*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.

### TABLE 4

<table>
<thead>
<tr>
<th>Antibiotic resistance determinant(s)</th>
<th>No. of isolates (%) for serotype:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35B (n = 421)</td>
</tr>
<tr>
<td>ermB</td>
<td>3 (0.7)</td>
</tr>
<tr>
<td>mefA</td>
<td>32 (7.6)</td>
</tr>
<tr>
<td>tetM</td>
<td>36 (8.6)</td>
</tr>
<tr>
<td>folA I100L and folP insertion</td>
<td>140 (33.3)</td>
</tr>
</tbody>
</table>
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 \(\text{wciG}\) 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 Preven-
tion (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 Ben-
bachir, 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 Zea-
land, 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

Lukišč I, Nair H, McAllister D, Campbell H, Rudan I, Knoll M. 2016. Global
burden of \(\text{Streptococcus pneumoniae}\) in children younger than 5 years in
the pneumococcal conjugate vaccines (PCV) era: 2000–2015. Interna-
tional Symposium on Pneumococci and Pneumococcal Diseases,
Glasgow, Scotland.

2. Geno KA, Saad JS, Nahm MH. 2017. Discovery of novel pneumococcal
serotype 35D, a natural \(\text{WciG}\)-deficient variant of serotype 35B. J Clin

capsule synthesis locus \(cps\) as evolutionary hotspot with potential to
https://doi.org/10.1093/molbev/msx173.

4. Balsells E, Guillot L, Nair H, Kyaw MH. 2017. Serotype distribution of
\(\text{Streptococcus pneumoniae}\) 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 \(\text{Streptococcus pneumoniae}\)
serogroup 35 variants in Australian patients. Clin Microbiol Infect 23:

Invasive serotype 35B pneumococci including an expanding serotype
https://doi.org/10.3201/eid2306.170071.

O-acetytransferase functionality differentiates pneumococcal serotypes
00822-17.

8. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins
M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kaltoft


