

- 1 **Global distribution of invasive serotype 35D *Streptococcus pneumoniae* post-**  
2 **PCV13 introduction**
- 3 **Stephanie W. Lo,<sup>a,#</sup> Rebecca A. Gladstone,<sup>a</sup> Andries J. van Tonder,<sup>a</sup> Paulina A.**  
4 **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>**  
5 **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**  
6 **Köseoglu Eser,<sup>m</sup> Pak Leung Ho,<sup>n</sup> Naima Elmdaghri,<sup>o,p</sup> Sadia Shakoor,<sup>q</sup> Stuart C.**  
7 **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.**  
8 **Klugman,<sup>b</sup> Lesley McGee,<sup>c</sup> Robert F. Breiman,<sup>b,v</sup> Stephen D. Bentley,<sup>a,#</sup> The Global**  
9 **Pneumococcal Sequencing Consortium**
- 10 <sup>a</sup>Infection Genomics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome  
11 Campus, Hinxton, Cambridge, CB10 1SA, UK
- 12 <sup>b</sup>Hubert Department of Global Health, Rollins School of Public Health, Emory University,  
13 Atlanta, GA 30322, USA
- 14 <sup>c</sup>Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA  
15 30333, USA
- 16 <sup>d</sup>Vaccines and Immunity Theme, Medical Research Council Unit The Gambia, P.O Box 273,  
17 Banjul, Fajara, The Gambia
- 18 <sup>e</sup>Malawi Liverpool Wellcome Trust Clinical Research Programme, P.O. Box 30096, Blantyre,  
19 Malawi
- 20 <sup>f</sup>Institute of Infection & Global Health, University of Liverpool, Liverpool, L69 7BE,  
21 UK <sup>g</sup>Medical Research Council: Respiratory and Meningeal Pathogens Research Unit,  
22 University of the Witwatersrand, Johannesburg, South Africa
- 23 <sup>h</sup>Department of Science and Technology/National Research Foundation: Vaccine Preventable  
24 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
- <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, OX3 9DU, UK
- <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, 06100, 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, B.P. 9154, Hassan II University of Casablanca, Morocco
- <sup>p</sup>Bacteriology-Virology and Hospital Hygiene Laboratory, University Hospital Centre Ibn Rochd, 1, Rue des, Casablanca, Morocco
- <sup>q</sup>Department of Pathology and Laboratory Medicine and Department of Paediatrics and Child Health, The Aga Khan University, Karachi 74800, Pakistan;
- <sup>r</sup>Faculty of Medicine and Institute of Life Sciences, University of Southampton, SO17 1BJ, UK;
- <sup>s</sup>Microbiology and Infection Unit, Warwick Medical School, Warwick, CV4 7AL, UK; <sup>t</sup>London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK;
- <sup>u</sup>University of Edinburgh, The Queens Medical Research Institute, Edinburgh, EH16 4TJ, UK;
- <sup>v</sup>Emory Global Health Institute, Emory University, Atlanta, GA 30322, USA
- Running Head: Epidemiology of serotype 35D Streptococcus pneumoniae

51 #Address correspondence to Stephanie W. Lo, [stephanie.lo@sanger.ac.uk](mailto:stephanie.lo@sanger.ac.uk) and Stephen D.  
52 Bentley, [sdb@sanger.ac.uk](mailto:sdb@sanger.ac.uk).

53  
54 **ABSTRACT (249/250)**

55 A newly recognized pneumococcal serotype 35D, which differs from the 35B polysaccharide  
56 in structure and serology by not binding to factor serum 35a, was recently reported. The  
57 genetic basis for this distinctive serology is due to the presence of an inactivating mutation in  
58 *wciG*, which encodes an O-acetyltransferase responsible for O-acetylation of a  
59 galactofuranose. Here, we assessed the genomic data of a worldwide pneumococcal  
60 collection to identify serotype 35D isolates and understand their geographical distribution,  
61 genetic background and invasiveness potential. Of 21,980 pneumococcal isolates, 444 were  
62 originally typed as serotype 35B by PneumoCaT. Analysis of *wciG* revealed 23 isolates from  
63 carriage (n=4) and disease (n=19) with partial or complete loss-of-function mutations,  
64 including mutations resulting in pre-mature stop codons (n=22) and an in-frame mutation  
65 (n=1). These were selected for further analysis. The putative 35D isolates were  
66 geographically widespread and 65.2% (15/23) of them was recovered after PCV13  
67 introduction. Compared with serotype 35B, putative serotype 35D isolates have higher  
68 invasive disease potentials based on odds ratio (OR) (11.58; 95% CI, 1.42-94.19 vs 0.61;  
69 95% CI, 0.40-0.92) and a higher prevalence of macrolide resistance mediated by *mefA*  
70 (26.1% vs 7.6%, p=0.009). Using Quellung, 50% (10/20) of viable isolates were serotype  
71 35D, 25% (5/20) serotype 35B, and 25% (5/20) a mixture of 35B/35D. The discrepancy  
72 between phenotype and genotype requires further investigation. These findings illustrated a  
73 global distribution of an invasive serotype 35D among young children post-PCV13  
74 introduction and underlined the invasive potential conferred by the loss of O-acetylation in  
75 the pneumococcal capsule.

76

## INTRODUCTION

*Streptococcus pneumoniae* (pneumococcus) is an important human pathogen that causes pneumonia, bacteremia, and meningitis. In 2015, >330,000 deaths globally in children <5 years old were estimated to have been caused by pneumococci (1). Its polysaccharide capsule, 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 non-vaccine 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 USA (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 was otherwise identical to serotype 35B (2). Serologically, it is distinct from serotype 35B strain 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

103 (11). Loss of O-acetylation in serotype 11E is predicted to assist pneumococci in evading  
104 host immune and vaccine response, and has been suggested to occur during invasive disease  
105 after initial colonization with the serotype 11A strain expressing an O-acetylated form of  
106 capsule (12). The role of loss of O-acetylation in pneumococcal survival during invasion  
107 among the other serotype pairs has remained unknown due to the rarity of serotype 9A, 33A,  
108 and 42 for comparisons, and the difficulty in differentiation between serotype 15B and 15C.

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

114

## 115 RESULTS AND DISCUSSION

116 Of 21,980 assembled pneumococcal genomes from the Global Pneumococcal  
117 Sequencing (GPS) project (n=16,575, May 2017) and a compiled dataset (n=5,405) by Van  
118 Tonder *et al.* (13), 444 isolates from disease (n=173), carriage (n=270), and unknown source  
119 (n=1) were originally typed as serotype 35B by PneumoCaT (5). The *wciG* alignment  
120 revealed that 78.6% (349/444) of isolates were identical to the serotype 35B reference, 8.3%  
121 (37/444) had silent mutations, 7.9% (35/444) had missense mutations, 3.4% (15/444) had  
122 frameshift mutations, 1.6% (7/444) had non-sense mutations, and 0.2% (1/444) had an in-  
123 frame insertion. All frameshift mutations led to a pre-mature stop codon which disrupted the  
124 coding region of *wciG*. Given that the latter three types of mutations lead to reduced function  
125 or a complete loss of function of WciG, the 23 isolates were designated serotype 35D (Table  
126 1). The Quellung reaction of 20 viable isolates showed that 50% (10/20) were serologically  
127 typed as serotype 35D, 25% (5/20) serotype 35B and 25% (5/20) a mixture of serotype 35B

128 and 35D (Table 2). In all discrepant cases, we examined the *cps* locus sequences in an  
129 attempt to identify any gene loss and mixed *wciG* alleles. The *cps* locus region shared the  
130 same capsular genes with the serotype 35D reference (accession number KY084476), and the  
131 mutations in *wciG* were supported by at least 42X depth of reads (median: 80X; range: 42X  
132 to 143X) with 100% consistency. The discrepancy between phenotype and genotype could be  
133 due to 1) our inability to capture the serotype diversity in a clinical sample, since the bacterial  
134 culture subject to DNA extraction and Quellung testing were derived from a single colony  
135 that could be different between experiments; and 2) the possible inter-convertibility between  
136 serotype 35B and 35D during bacterial culture *in vitro*. In all five isolates which were both  
137 positive and negative to antisera fs35a under one microscope (Table 2), the mutations in  
138 *wciG* were either a 1-bp insertion or deletion that occurred after a 6- to 7-bp homopolymer,  
139 highlighting the possibility of inter-conversion between serotype 35B and 35D during DNA  
140 replication. Metagenomic analysis of clinical samples to snapshot the serotype diversity and  
141 investigation into the inter-convertibility of serotype 35B and 35D will potentially explain the  
142 discrepancy between the phenotypes and genotypes observed in this study. Considering the  
143 limitation of this study and our recent understanding of the genetic basis that differentiates  
144 serotype 35B and 35D (2, 6, 7), the non-silent mutations detected in *wciG* in this study  
145 strongly indicated the presence of serotype 35D pneumococci in the sample. Thus, the 23 *in*  
146 *silico* serotype 35D isolates were selected for further analysis.

147 The mutation patterns of *wciG* among the *in silico* serotype 35D isolates were diverse.  
148 The *wciG* mutation patterns in the 23 serotype 35D isolates were different from the 6  
149 serotype 35D isolates reported previously (2, 5, 6). In total, there were twenty mutation  
150 patterns observed in 29 serotype 35D isolates from ten countries across four continents  
151 (Table 1). The most common naturally deficient WciG was due to 86\_87insG, which  
152 occurred within a 6-bp homopolymeric stretch of guanine. It was first observed in an isolate

153 from Malawi in 2006, prior to the introduction of PCV7, and was also found in isolates from  
154 Senegal in 2011, South Africa and the USA in 2012, and New Zealand in 2015. Isolates with  
155 this mutation were sporadically distributed on the phylogenetic tree (Figure 1), suggesting  
156 that the mutations had arisen independently on multiple occasions. The convergence of  
157 mutations may imply that this site is a mutational hotspot.

158 The majority of serotype 35D isolates belonged to the clonal complexes (CC)558  
159 (n=9), CC198 (n=6), and CC156 (n=5) that were primarily associated with serotype 35B (6,  
160 14, 15). The CC558 and CC156 lineages accounted for most of the increase in serotype 35B  
161 after the introduction of PCV13 in the USA (6), while CC198 is the major serotype 35B  
162 lineage in The Gambia (unpublished data). Based on a high-resolution SNP tree (Figure 1),  
163 serotype 35D pneumococci emerged among closely related serotype 35B isolates within  
164 different clusters. Together with the unrelated mutations observed in *wciG*, this strongly  
165 indicated that serotype 35B is the progenitor of serotype 35D.

166 When compared with serotype 35B isolates, serotype 35D isolates were more likely to  
167 be recovered from sterile anatomical sites including cerebrospinal fluid (CSF; n=9), blood  
168 (n=8), lung aspirate (n=1), and joint aspirate (n=1) than among carriage isolates (n=4) [82.6%  
169 (19/23) vs 36.7% (154/420);  $p < 0.001$  by Fisher's exact test]. Based on a larger  
170 pneumococcal collection (n=3,333) randomly selected from the GPS project database, the  
171 empirical odds ratio (OR) for invasive disease due to serotype 35D is 11.58 (95% confidence  
172 interval 95% CI: 1.42-94.19), whereas the OR for serotype 35B is 0.61 (95% CI: 0.40-0.92).  
173 The increased invasive capacity in serotype 35D strain could be a result of evasion of the  
174 immune response targeting the capsule O-acetyl group. The observation in serotype 35B/35D  
175 coincides with a previous study on serotype 11A/11E, in which serotype 11E strains with a  
176 loss or reduced amount of acetylation in the capsule were found to be significantly associated  
177 with invasive pneumococcal disease (12, 16). The emergence of serotype 35D is likely



178 explained by Calix et al.'s hypothesis (12) that pneumococcal capsule structure undergoes  
179 microevolution during progression from carriage to infection in response to divergent  
180 selection pressure in early mucosal colonization compared to later in a sterile site. This model  
181 of microevolution needs to be further investigated by characterizing the serotype dynamic  
182 over the development of invasive disease *in vivo*.

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

191 Among the 23 serotype 35D isolates, 87.0% (20/23) had at least one resistance  
192 determinant conferring resistance to commonly used antibiotics including penicillin (65.2%,  
193 15/23), erythromycin (30.4%, 7/23), cotrimoxazole (21.7%, 5/23), and tetracycline (4.3%,  
194 1/23). Similar to the previous studies on serotype 35B (6, 14), the penicillin-resistant isolates  
195 in this study were predominantly CC558 (60.0%, 9/15), followed by CC156 (35.7%, 5/15)  
196 and a singleton of ST373 (6.7%, 1/15). Macrolide resistance mediated by *mefA* was  
197 significantly higher in serotype 35D isolates than in serotype 35B isolates (Table 4). Five of  
198 six serotype 35D isolates harboring *mefA* were from the USA, where macrolides are  
199 recommended for use as an empirical therapy for pneumonia in children (17-19); they all  
200 belonged to CC558, a major contributor to penicillin resistance in USA after introduction of  
201 PCV13 (14). Unlike the highly invasive but usually antibiotic susceptible serotype 1,  
202 pneumococci expressing serotype 35B (lower-invasive capsule) are more likely to be



203 commensal in the nasopharynx which could allow them to acquire antibiotic resistance  
204 determinants via horizontal gene transfer from other nasopharyngeal bacteria; a subsequent  
205 switch to serotype 35D (high-invasive capsule) would then transform the antibiotic resistant  
206 strain into a more virulent form.

207       The limitation of this study is that the carriage and disease isolates included for  
208 calculating the invasiveness index were sampled in different cities in each country; all  
209 isolates included were collected between 2007 and 2015 from children aged < 2 years-old.  
210 Ideally, the carriage and disease isolates should be geography, time, and age-matched. In this  
211 instance, we calculated OR for invasiveness separately for each country: the ORs for invasive  
212 disease due to serotype 35B and 35D in The Gambia were 0.37 (95% CI: 0.09-1.56) and 20.3  
213 (95% CI: 2.10-196.42), respectively. The ORs could not be calculated for invasive disease as  
214 all serotype 35D isolates in South Africa and Malawi were from disease. The ORs for disease  
215 due to 35B in South Africa and Malawi were 0.68 (95% CI: 0.40-1.16) and 0.72 (95% CI:  
216 0.11 – 2.15), respectively. The ORs by country were consistent with the ORs calculated from  
217 the combined datasets of all three countries. Another limitation was that the effects of an in-  
218 frame insertion of 15bp and the missense mutations in *wciG* on the protein function have not  
219 been evaluated. Removing these samples from all comparisons of serotype 35B and 35D did  
220 not alter the conclusions drawn from the statistical analyses.

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

224

## 225 MATERIALS AND METHODS

226       We retrospectively determined serotypes of 21,980 assembled pneumococcal  
227 genomes from the GPS project (n=16,575, May 2017, <http://www.pneumogen.net/gps/>) and a

228 compiled dataset (n=5,405) by Van Tonder *et al.* (13). DNA extraction was performed on a  
229 pure overnight culture derived from a single colony. Sequencing was performed on the  
230 Illumina HiSeq platform to produce paired-end reads of either 75 (2010-2011), 100 (2013-  
231 2014) or 125 base pairs (2015-2016) in length. *In silico* serotype was determined using the  
232 whole genome sequence (WGS) based serotyping method PneumoCaT (20). As the current  
233 version of PneumoCaT does not distinguish serotype 35D from serotype 35B, all samples  
234 that were initially typed as serotype 35B were included in this study. To differentiate these  
235 two serotypes, nucleotide sequences of *wciG* were extracted from the assembled genomes  
236 and aligned to a reference sequence of 35B *wciG* (accession number KX021817) described  
237 by Geno *et al.* using CLUSTALW (2, 21). Mutations such as nonsense and frameshift  
238 mutations that led to pre-mature stop codon and in-frame insertion/deletions in *wciG* were  
239 predicted to reduce the function or a complete loss of function of the WciG protein. Isolates  
240 with these mutations were *in silico* typed as serotype 35D and their phenotypic serotype were  
241 determined by the Quellung reaction tested on an overnight culture derived from a single  
242 colony (22). Phylogenetic analysis was performed on all serotype 35B and 35D isolates by  
243 constructing a maximum likelihood tree using RAxML v.8.2.X (23) based on single  
244 nucleotide polymorphic sites (SNPs) extracted from a core gene alignment with Roary v.3.6.1  
245 (24). An empirical odds ratio for invasive disease due to serotype 35B and 35D was  
246 calculated based on a pneumococcal collection of 3,333 randomly selected carriage (n=1,260)  
247 and disease (n=2,073) isolates from children aged < 2 years-old, collected during the pre-  
248 PCV (n=1,691), post-PCV7 (n=678), and post-PCV13 (n=964) eras using a previously  
249 described method (25). For each country, the random selection was carried out from a  
250 collection of disease isolates collected via laboratory-based surveillance and carriage isolates  
251 via cohort-studies using the following criteria: 50% isolates represented pre-PCV period ( $\leq 1$   
252 year before) and 50% post-PCV period ( $\geq 2$  years after primary,  $\geq 1$  after subsequent PCVs).

253 The randomly selected collection in this study included 67 different serotypes plus non-  
254 typeable pneumococci. They were collected in South Africa (carriage n=721, disease  
255 n=1,047), Malawi (carriage n=336, disease n=60), and The Gambia (carriage n=1,016,  
256 disease n=153). Isolates from other locations in the GPS dataset were either not randomly  
257 selected, or consisted of only disease or carriage isolates and thus could not be used to  
258 calculate odds ratios. Susceptibility to chloramphenicol, cotrimoxazole, erythromycin,  
259 penicillin, and tetracycline were predicted by the identification of resistant determinants in  
260 the assembled genomes using previously described pipelines (26-28). The epidemiological  
261 and phylogenetic data can be interactively visualized and analyzed online using the  
262 Microreact tool at [https://microreact.org/project/GPS\\_serotype\\_35B\\_35D](https://microreact.org/project/GPS_serotype_35B_35D)

263  
264 Figure 1 Maximum likelihood phylogenetic tree was constructed using 56,848 SNPs  
265 extracted from a 1.02-Mb codon alignment of 1,141 core genes from 444 serotype 35B and  
266 35D *S. pneumoniae* isolates. The tree is colored according to the geographic location of each  
267 samples' isolation. This analysis used an unrelated non-typeable isolate as the outgroup on  
268 which to root the tree. Clonal complex (CC) and mutations in *wciG* are shown to the right of  
269 the tree. Singleton sequence types and minor CCs with <5 isolates in this study are indicated  
270 in pink and grey, respectively.

## 271 272 **ACKNOWLEDGEMENTS**

273 This work is funded by the Bill and Melinda Gates Foundation (grant code OPP1034556), the  
274 Wellcome Trust Sanger Institute, and the Centers for Disease Control and Prevention, Atlanta,  
275 USA. We thank The Centre for Genomic Pathogen Surveillance (CGPS), Sanger Institute for  
276 their excellent visualization tool, Microreact. The authors gratefully acknowledge Dr.  
277 Bernard W. Beall and Dr. 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 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 Institute of Environmental Science and Research Limited, New Zealand and Olga Hattingh from National Institute for Communicable Disease, South Africa for performing Quellung test. The authors declare no additional conflicts of interest.

#### DISCLAIMER

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.

#### REFERENCE

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, abstr International Symposium on Pneumococci and Pneumococcal Diseases, Glasgow.
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.
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* doi: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.
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* doi: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.

- 314 7. Geno KA, Bush CA, Wang M, Jin C, Nahm MH, Yang J. 2017. WciG O-  
315 Acetyltransferase Functionality Differentiates Pneumococcal Serotypes 35C and 42. *J*  
316 *Clin Microbiol* 55:2775-2784.
- 317 8. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitsch E, Collins M,  
318 Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kalløft MS,  
319 Barrell B, Reeves PR, Parkhill J, Spratt BG. 2006. Genetic analysis of the capsular  
320 biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2:e31.
- 321 9. Spencer BL, Saad JS, Shenoy AT, Orihuela CJ, Nahm MH. 2017. Position of O-  
322 Acetylation within the Capsular Repeat Unit Impacts the Biological Properties of  
323 Pneumococcal Serotypes 33A and 33F. *Infect Immun* 85.
- 324 10. Brady AM, Calix JJ, Yu J, Geno KA, Cutter GR, Nahm MH. 2014. Low invasiveness  
325 of pneumococcal serotype 11A is linked to ficolin-2 recognition of O-acetylated  
326 capsule epitopes and lectin complement pathway activation. *J Infect Dis* 210:1155-65.
- 327 11. Rajam G, Carlone GM, Romero-Steiner S. 2007. Functional antibodies to the O-  
328 acetylated pneumococcal serotype 15B capsular polysaccharide have low cross-  
329 reactivities with serotype 15C. *Clin Vaccine Immunol* 14:1223-7.
- 330 12. Calix JJ, Dagan R, Pelton SI, Porat N, Nahm MH. 2012. Differential occurrence of  
331 *Streptococcus pneumoniae* serotype 11E between asymptomatic carriage and invasive  
332 pneumococcal disease isolates reflects a unique model of pathogen microevolution.  
333 *Clin Infect Dis* 54:794-9.
- 334 13. van Tonder AJ, Bray JE, Quirk SJ, Haraldsson G, Jolley KA, Maiden MC, Hoffmann  
335 S, Bentley SD, Haraldsson A, Erlendsdottir H, Kristinsson KG, Brueggemann AB.  
336 2016. Putatively novel serotypes and the potential for reduced vaccine effectiveness:  
337 capsular locus diversity revealed among 5405 pneumococcal genomes. *Microb*  
338 *Genom* 2:000090.
- 339 14. Metcalf BJ, Gertz RE, Jr., Gladstone RA, Walker H, Sherwood LK, Jackson D, Li Z,  
340 Law C, Hawkins PA, Chochua S, Sheth M, Rayamajhi N, Bentley SD, Kim L,  
341 Whitney CG, McGee L, Beall B, Active Bacterial Core surveillance t. 2016. Strain  
342 features and distributions in pneumococci from children with invasive disease before  
343 and after 13-valent conjugate vaccine implementation in the USA. *Clin Microbiol*  
344 *Infect* 22:60 e9-60 e29.
- 345 15. Beall B, McEllistrem MC, Gertz RE, Jr., Boxrud DJ, Besser JM, Harrison LH,  
346 Jorgensen JH, Whitney CG, Active Bacterial Core Surveillance/Emerging Infections  
347 Program N. 2002. Emergence of a novel penicillin-nonsusceptible, invasive serotype  
348 35B clone of *Streptococcus pneumoniae* within the United States. *J Infect Dis*  
349 186:118-22.
- 350 16. Calix JJ, Brady AM, Du VY, Saad JS, Nahm MH. 2014. Spectrum of pneumococcal  
351 serotype 11A variants results from incomplete loss of capsule O-acetylation. *J Clin*  
352 *Microbiol* 52:758-65.
- 353 17. Brogan TV, Hall M, Williams DJ, Neuman MI, Grijalva CG, Farris RW, Shah SS.  
354 2012. Variability in processes of care and outcomes among children hospitalized with  
355 community-acquired pneumonia. *Pediatr Infect Dis J* 31:1036-41.
- 356 18. Hersh AL, Shapiro DJ, Pavia AT, Shah SS. 2011. Antibiotic prescribing in  
357 ambulatory pediatrics in the United States. *Pediatrics* 128:1053-61.
- 358 19. Kronman MP, Hersh AL, Feng R, Huang YS, Lee GE, Shah SS. 2011. Ambulatory  
359 visit rates and antibiotic prescribing for children with pneumonia, 1994-2007.  
360 *Pediatrics* 127:411-8.
- 361 20. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry  
362 NK. 2016. Whole genome sequencing of *Streptococcus pneumoniae*: development,

- 363 evaluation and verification of targets for serogroup and serotype prediction using an  
364 automated pipeline. *PeerJ* 4:e2477.
- 365 21. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H,  
366 Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.  
367 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-8.
- 368 22. Institut SS. 2015. *Streptococcus Pneumoniae: Textbook in Diagnosis, Serotyping,*  
369 *Virulence Factors and Enzyme-linked Immunosorbent Assay (ELISA) for Measuring*  
370 *Pneumococcal Antibodies.* Statens Serum Institut.
- 371 23. Liu K, Linder CR, Warnow T. 2011. RAxML and FastTree: comparing two methods  
372 for large-scale maximum likelihood phylogeny estimation. *PLoS One* 6:e27731.
- 373 24. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush  
374 D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome  
375 analysis. *Bioinformatics* 31:3691-3.
- 376 25. Brueggemann AB. 2003. Clonal Relationships between Invasive and Carriage  
377 *Streptococcus pneumoniae* and Serotype- and Clone-Specific Differences in Invasive  
378 Disease Potential.
- 379 26. Metcalf BJ, Chochua S, Gertz RE, Li Z, Walker H, Tran T, Hawkins PA, Glennen A,  
380 Lynfield R, Li Y, McGee L, Beall B. 2016. Using whole genome sequencing to  
381 identify resistance determinants and predict antimicrobial resistance phenotypes for  
382 year 2015 invasive pneumococcal disease isolates recovered in the United States.  
383 *Clinical Microbiology and Infection* 22:1002.e1-1002.e8.
- 384 27. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr., Walker H, Hawkins PA, Tran T,  
385 Whitney CG, McGee L, Beall BW. 2016. Penicillin-Binding Protein Transpeptidase  
386 Signatures for Tracking and Predicting beta-Lactam Resistance Levels in  
387 *Streptococcus pneumoniae*. *MBio* 7.
- 388 28. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr., Walker H, Hawkins PA, Tran T,  
389 McGee L, Beall BW, Active Bacterial Core surveillance t. 2017. Validation of beta-  
390 lactam minimum inhibitory concentration predictions for pneumococcal isolates with  
391 newly encountered penicillin binding protein (PBP) sequences. *BMC Genomics*  
392 18:621.
- 393 29. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD,  
394 Hanage WP, Lipsitch M. 2013. Population genomics of post-vaccine changes in  
395 pneumococcal epidemiology. *Nat Genet* 45:656-63.
- 396

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)

<i>wciG</i> nucleotide mutation	n	Clonal Complex/ Sequence Type	Geographical location of isolation	Year of isolation	Site of isolation	Ref.
<b><u>Frameshift mutation (n=18)<sup>a</sup></u></b>						
86_87insG	6	CC156 (n=2), CC558 (n=2), CC198 (n=1), CC9813 (n=1)	Malawi (n=2), New Zealand (n=1), Senegal (n=1), South Africa (n=1), USA (n=1)	2006 (n=1), 2011 (n=1), 2012 (n=2), 2015 (n=2)	CSF (n=3), blood (n=2), joint pus (n=1)	GPS
914_929del_16bp	2	CC558	South Africa, USA	2012 (n=1), 2013 (n=1)	CSF (n=1), blood (n=1)	GPS
162_163insT	2	CC558	USA	2004 (n=1), 2007 (n=1)	Nasopharynx (n=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)
<b><u>In-frame deletion/insertion (n=3)</u></b>						
792_968del_177bp <sup>b</sup>	1	CC156	USA	2015	Blood (n=2)	(6)
755_808del_54bp <sup>b</sup>	1	CC558	Australia	2016	unknown	(5)
523_524ins_15bp	1	CC558	USA	2009	Blood	GPS
<b><u>Nonsense mutation (n=7)</u></b>						
C220T	2	CC156, ST373	Nepal, South Africa	2013 (n=1), 2014 (n=1)	CSF (n=1),	GPS



					nasopharynx (n=1)	
T732G	2	CC198	The Gambia	2014 (n=2)	CSF (n=1), blood (n=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
<b><u>Missense mutation (n=1)</u></b>						
G533A, G679A <sup>c</sup>	1	unknown	USA	unknown	unknown	(2)

<sup>a</sup>All frameshift mutations resulted in pre-mature stop codon.

<sup>b</sup>The in-frame deletion rendered the WciG, an acetyltransferase, non-functional. 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 non-functional 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. (29) and *in silico* serotype was updated as serotype 35D in this study. CSF, cerebrospinal fluid

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 the Quellung reactions

Strain name	Country	CC	Year	<i>wciG</i> mutation	Pool G	Type 29	Type 42	Group 35	fs35a	fs35b	fs35c	fs29b	fs42a	Phenotypic serotype	Ref.
3431-06	USA	N/A	N/A	G533A, G679A	+	N	N	-	-	-	+	+	-	35D	(2)
16S471	Australia	CC198	2016	306_307insA	+	+	+	+	-	-	+	+	-	35D	(5)
SAMDU-00005305	Australia	CC558	2015	36delA	+	+	+	+	-	-	+	+	-	35D	(5)
16S49	Australia	CC452	2016	663_696del_34bp	+	+	+	+	-	-	+	+	-	35D	(5)
16S35	Australia	CC558	2016	755_808del_54bp	+	+	+	+	-	-	+	+	-	35D	(5)
20152877	USA	CC156	2015	792_968del_177bp	+	N	N	+	-	-	+	+	-	35D	(6)
CH2075	USA	CC558	2007	162_163insT	+	+	-	+	+	-	+	+	-	35B	GPS <sup>c</sup>
3025	USA	CC558	2004	162_163insT	+	+	-	+	+	-	+	+	-	35B	GPS <sup>c</sup>
GPS_US_2010209945_R1	USA	CC558	2009	523_524ins_15bp	+	+	-	+	+	-	+	+	-	35B	GPS
GPS_GM_1130	The Gambia	CC198	2014	T731G (L244*)	+	+	-	+	+	-	+	+	-	35B	GPS
GPS_GM_1148	The Gambia	CC198	2014	T731G (L244*)	+	+	-	+	+	-	+	+	-	35B	GPS
GPS_ZA_2370	South Africa	CC9813	2012	382_385delATAT	+	+	+	+	-	-	+	+	-	35D	GPS
GPS_ZA_2636	South Africa	CC558	2013	914_929del_16bp	+	+	+	+ <sup>b</sup>	-	-	+	+	-	35D	GPS
2012215593	USA	CC558	2012	914_929del_16bp	+	+	-	-	-	-	+	+	-	35D	GPS
2012215608	USA	CC558	2012	C104A (S35*)	+	+	-	-	-	-	+	+	-	35D	GPS
GPS_ZA_2559	South Africa	CC156	2013	C220T (Q74*)	+	+	+	+	-	-	+	+	-	35D	GPS
GPS_NP_7242	Nepal	Singleton <sup>d</sup>	2014	C220T (Q74*)	+	+	N	+	-	-	+	+	-	35D	GPS
2012220613	USA	CC558	2012	C323A (S108*)	+	+	-	-	-	-	+	+	-	35D	GPS
2013208723	USA	CC558	2012	86_87insG	+	+	-	-	-	-	+	+	-	35D	GPS

GPS_MW_D38253_R1	Malawi	CC156	2006	86_87insG	+	+	-	-	-	-	+	+	-	35D	GPS
GPS_MW_BKR609	Malawi	CC156	2015	86_87insG	+	+	-	-	-	-	+	+	-	35D	GPS
PI0167	Senegal	CC198	2011	86_87insG	+	+	-	+	+ <sup>b</sup>	-	+	+	-	35B/D	GPS
GPS_NZ_15SP0720	New Zealand	CC558	2013	86_87insG	+	+	N	+	+ <sup>c</sup>	-	+	+	-	35B/D	GPS
GPS_ZA_2487	South Africa	CC9813	2012	86_87insG	+	+	+	+	+ <sup>b</sup>	-	+	+	-	35B/D	GPS
GPS_MW_BKR5WC	Malawi	CC156	2015	705_706insT	+	+	-	+ <sup>b</sup>	+ <sup>b</sup>	-	+	+	-	35B/D	GPS
PI0258	Cameroon	CC156	2012	86delG	+	+	-	+	+ <sup>b</sup>	-	+	+	-	35B/D	GPS
GPS_GM_0282	The Gambia	CC198	2013	92_93insC	N	N	N	N	N	N	N	N	N	N	GPS
GPS_GM_0600	The Gambia	CC198	2009	312delA	N	N	N	N <sup>a</sup>	N	N	N	N	N	N	GPS
GPS_GM_0320	The Gambia	CC198	2009	T434G (L145*)	N	N	N	N	N	N	N	N	N	N	GPS

<sup>a</sup>N, 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 exhibit both positive to antisera 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. (29) and *in silico* serotype was updated as serotype 35D in this study.

Table 3. The prevalence of serotype 35B and 35D *S. pneumoniae* from South Africa (n=1768), The Gambia (n=1169) and Malawi (n=396) in each vaccine period

Vaccine period <sup>a</sup>	No. of isolates (%)		Odds ratio (95% confidence interval)	
	serotype 35B	serotype 35D	serotype 35B	serotype 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 to 1.6)	-
Post-PCV13 (n=964)	69 (7.16)	7 (0.73)	3.54 (2.4 to 5.4)*	12.36 (1.5 to 100.6)*

<sup>a</sup>Based on the year of PCV introduction, we grouped each year of collection into three categories: 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-valent PCV was introduced); 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.

\*p value < 0.05

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

Antibiotic resistance determinants	No. of isolates (%)		P value
	serotype 35B (n=421)	serotype 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

