Pneumococcal Capsular Switching: A Historical Perspective

Kelly L. Wyres, University of Oxford
Lotte M. Lambertsen, Statens Serum Institut
Nicholas J. Croucher, Wellcome Trust Sanger Institute
Lesley McGee, Centers for Disease Control and Prevention
Ann von Gottberg, National Institute for Communicable Diseases
Josefina Linares, Bellvitge Hospital
Michael R. Jacobs, Case Western Reserve University
Karl G. Kristinsson, University of Iceland
Bernard W. Beall, Centers for Disease Control and Prevention
Keith P Klugman, Emory University

Only first 10 authors above; see publication for full author list.

Journal Title: Journal of Infectious Diseases
Volume: Volume 207, Number 3
Publisher: Oxford University Press (OUP): Policy B - Oxford Open Option C | 2013-02-01, Pages 439-449
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1093/infdis/jis703
Permanent URL: https://pid.emory.edu/ark:/25593/v1nxm

Final published version: http://dx.doi.org/10.1093/infdis/jis703

Copyright information:
© 2012 The Author.
This is an Open Access work distributed under the terms of the Creative Commons Attribution 3.0 Unported License (http://creativecommons.org/licenses/by/3.0/).

Accessed December 15, 2019 7:47 AM EST
Pneumococcal Capsular Switching: A Historical Perspective

Kelly L. Wyres,1 Lotte M. Lambertsen,3 Nicholas J. Croucher,2 Lesley McGee,4 Anne von Gottberg,7 Josefin Liñares,8 Michael R. Jacobs,6 Karl G. Kristinsson,3 Bernard W. Beall, Keith P. Klugman, Julian Parkhill,2 Regine Hakenbeck,10 Stephen D. Bentley,2 and Angela B. Brueggemann1

1Department of Zoology, University of Oxford, and 2Pathogen Genomics Team, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 3Department of Microbiology Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark; 4Streptococcus Laboratory, Centers for Disease Control and Prevention, and 5Hubert Department of Global Health, Emory University, Atlanta, Georgia; 6Department of Pathology, Case Western Reserve University, Cleveland, Ohio; 7Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases, Gauteng, South Africa; 8Department of Microbiology, Bellvitge Hospital-CIBERes-IDIBELL-UB, Barcelona, Spain; 9Clinical Microbiology Department, Landspitali University Hospital and University of Iceland, Reykjavik; and 10Department of Microbiology, University Kaiserslautern, Germany

Background. Changes in serotype prevalence among pneumococcal populations result from both serotype replacement and serotype (capsular) switching. Temporal changes in serotype distributions are well documented, but the contribution of capsular switching to such changes is unknown. Furthermore, it is unclear to what extent vaccine-induced selective pressures drive capsular switching.

Methods. Serotype and multilocus sequence typing data for 426 pneumococci dated from 1937 through 2007 were analyzed. Whole-genome sequence data for a subset of isolates were used to investigate capsular switching events.

Results. We identified 36 independent capsular switch events, 18 of which were explored in detail with whole-genome sequence data. Recombination fragment lengths were estimated for 11 events and ranged from approximately 19.0 kb to ≥58.2 kb. Two events took place no later than 1960, and the imported DNA included the capsular locus and the nearby penicillin-binding protein genes pbp2x and pbp1a.

Conclusions. Capsular switching has been a regular occurrence among pneumococcal populations throughout the past 7 decades. Recombination of large DNA fragments (>30 kb), sometimes including the capsular locus and penicillin-binding protein genes, predates both vaccine introduction and widespread antibiotic use. This type of recombination has likely been an intrinsic feature throughout the history of pneumococcal evolution.

Keywords. Capsule; serotype; switching; pneumococcus.

*Streptococcus pneumoniae* (the "pneumococcus") is an asymptomatic colonizer of the human nasopharynx and a major cause of otitis media, sinusitis, pneumonia, and meningitis, resulting in approximately 14.5 million annual serious disease episodes globally among young children [1]. The primary virulence determinant is its polysaccharide capsule, of which there are 94 known serotypes [2–6]. The capsule protects against phagocytosis during invasive pneumococcal disease and may also prevent clearance during nasopharyngeal colonization [7, 8]. Ninety-two capsule types are synthesized and exported through the "Wzy-dependent" pathway, whereby extracellular polymerization of component lipid-linked repeat units is preceded by "flippase-mediated" transfer across the cell membrane [2, 9]. The proteins involved are encoded by the capsule polysaccharide synthesis (cps) genes, located between *dexB* and *aliA* on the chromosome. The remaining 2 capsule types are synthesized through independent biochemical pathways.

Sequence analyses of 88 reference *cps* loci revealed variable lengths (approximately 10–30 kb) and a range of genes specific to capsule production [2]. The serotype-nonspecific genes were present among all
Wzy-dependent capsule types, whereas serotype-specific genes were only found among 1 or a subset of types. Unique serotypes, and further diversity within serogroups, evolved through a combination of mutation and interspecies/intrasppecies recombination [2, 10–13].

Serotype is a key determinant of invasive pneumococcal disease potential and prevalence; certain serotypes are more commonly associated with carriage and others more commonly with invasive pneumococcal disease [14, 15]. A 7-valent pneumococcal conjugate vaccine (PCV7) was introduced in the United States in 2000 [16] and, subsequently, into many other countries. Ten-valent and 13-valent vaccines (PHid-CV and PCV13, respectively) contain polysaccharides targeted against the original 7 serotypes (4, 6B, 9V, 14, 18C, 19F, and 23F) plus serotypes 1, 5, and 7F (both vaccines) and 3, 6A, and 19A (PCV13 only) and have now replaced PCV7 [17, 18].

Surveillance following PCV7 introduction showed a decline of vaccine type (VT) and an increase of nonvaccine type (NVT) pneumococci in disease and nasopharyngeal carriage [19]. This can be attributed to the phenomena of “serotype replacement,” the expansion of preexisting NVT pneumococci, and/or “serotype switching,” a change of serotype of a single clone by alteration or exchange of its cps locus [20]. (Serotype switching was first described by Griffith in 1928 [21] and was the focus of the transformation studies by Avery and colleagues in 1944 [22].) These effects are not completely independent: capsular switch variants can subsequently expand within a population. Both phenomena can be studied by comparison of the serotypes and genotypes present in populations before and after introduction of pneumococcal vaccines. Pneumococcal genotypes, as defined by multilocus sequence typing (MLST) [23], show serotype-specific associations [14, 24]; any isolate exemplifying a different genotype-serotype combination may represent a capsular switch variant. Such variants usually arise by recombination at the cps locus, and studies of a limited number of such strains indicated that recombination fragment sizes varied from approximately 21.9 kb to approximately 56.5 kb [25–28]. In some cases, the fragments also included part or all of the *php2x* and *php1a* genes (2 of the 3 primary penicillin-resistance determining genes, located approximately 8 kb upstream and approximately 7 kb downstream of the *cps* locus).

Vaccine-induced selective pressure is contributing to the postvaccination changes of serotype epidemiology, but natural fluctuations in serotype prevalence also play a role. Many studies have documented prevaccine temporal changes in relative serotype prevalence: all but 3 included pneumococci isolated no earlier than 1969; only 2 studies provided genotype data (Supplementary Materials). In this study, we used serotyping, MLST, and whole-genome data to study a large, genetically diverse collection of historical and modern pneumococci, with the aim to better understand the mechanisms and role of capsular switching in pneumococcal evolution.

**METHODS**

**Strains and Genome Sequencing**

A global collection of 426 pneumococci recovered during 1937–2007 (Supplementary Materials) were previously serotyped by the Quellung reaction and genotyped by MLST [23] to determine the sequence type (ST). Closely related isolates were assigned to clonal complexes (CCs) by a modified goeBURST method [29] (Supplementary Materials). CCs were named after the predicted founder ST. When no single founder ST could be determined, CCs were named NoneX, where X was the ST of lowest numeric value within the CC. Isolates of the same CC but with different serotypes were presumed to be capsular switch variants.

A total of 96 genetically diverse isolates from our collection were selected for whole-genome sequencing (Supplementary Materials and Table 1) on the Illumina platform as previously described [30]. Seven were excluded as technical failures. Raw sequence data were assembled using Velvet [31], and contigs were deposited in a BIGS database (BIGSdb) [32]. Sequence data were deposited in the European Nucleotide Archive (Supplementary Table 1).

**Nucleotide Sequence Analysis**

Eight CCs were represented by isolates subjected to whole-genome sequencing and were selected for further *cps* locus analysis; 7 CCs included capsular switch variants. Within each CC, the ancestral serotype was assumed to be the most common one or that of the oldest isolate. *cps* sequence alignments for study isolates and published loci were used to investigate serotype changes. Nucleotide sequences of the *cps* locus, and upstream/downstream flanking sequences where appropriate, were aligned using MUSCLE [33] and imported to MEGA5 [34] for visual inspection of variable sites. When recombination at the *cps* locus was predicted, potential donor representatives were sought from our whole-genome sequenced isolates, the *cps* locus reference sequences, and an additional 131 pneumococcal genomes retrieved from GenBank. Putative donors were identified as isolates for which the *cps* locus differed from that of the recombinant by a maximum of 3 nucleotides (excluding clusters of closely linked substitutions, which were considered to have arisen through recombination within the *cps* locus itself). MLST data were used to infer the CC most likely represented by the true donor isolate (the same as that of the donor representative). Recombination regions were identified as the minimum region over which the recombinant representative (hereafter, the “recombinant”) differed from the ancestral representative (hereafter, the “ancestor”) and was identical or highly similar (>99.7% sequence
identity) to the donor representative (hereafter, the “donor”). Where no suitably matched donor was identified, the maximum recombination region was estimated as the maximum length over which the recombinant differed from the ancestor (Supplementary Materials).

RESULTS

Indication of Serotype Switching

Among the entire collection of 426 pneumococci, 21 of 163 unique CCs were represented by isolates of 2–6 serotypes. At least 36 independent changes of serotype within CCs were represented in our collection, 34 of which predated the introduction of pneumococcal conjugate vaccines (Figure 1).

Within-CC cps Diversity

CC15

Seven of 10 CC15 (CC15, serotype 14) cps locus sequences were unique (Table 1). Isolates ICE13 and ICE50 shared wciY locus nucleotide substitutions with 4 CC124 representatives (isolates 14/9, USA6, Ala289, and ICE594), suggesting a possible recombination event between these CCs. The CC15 representatives had a 432-bp deletion in the wciY gene, resulting in a predicted 250-amino-acid truncation of the protein. Isolate CGSP14 [35] also had an additional 52-bp deletion within the wciY gene; if this represented a true deletion event rather than a sequencing/assembly error, the resultant protein would be further reduced to 38 amino acids.

CC66

Ten CC66 isolates, dated 1952–2005, represented 5 serotypes. The oldest isolate was serotype 7B; the next oldest was a serotype 9N from 1960. An approximately 53.4-kb region of sequence of the serotype 9N isolate, including pbp2x, the cps locus, and pbp1a, differed from the CC667B isolate but was highly similar to a CC3782N isolate dated 1952 (Table 2 and Figure 2). (Although the data strongly suggest that a CC3782N pneumococcus was the donor in this case, it is possible that another clone not represented in this sample was the true donor. Thus, we use the term “CCXX-like” when describing putative donors.) Within the cps locus, a second CC66N isolate differed from the oldest CC66N isolate by only 5 nucleotide substitutions (Table 1).

Sequences of the 2 CC6619F genomes suggested 2 independent capsular switches, but no potential 19F donors were identified. The putative imports were estimated as the regions over which these genomes clearly differed from those of CC66N; independent recombination breakpoints were indicated (Table 2 and Figure 2).

There were 3 independent capsular switches among 4 CC6614 representatives. On the basis of sequence similarity, the donors for 2 events were CC1514-like, while the third was CC12414-like. The breakpoints for 1 event could not be determined because of extensive diversity in the cps flanking regions, but breakpoints for the others were estimated (Table 2 and Figure 2). Interestingly, the 5’ breakpoints for 2 serotype 14 switches and one of the aforementioned serotype 19F switches were all 2.2–2.5 kb upstream of dexB (in the ATCC700669 reference genome [36], this corresponds to the clpL locus, which has been shown to be involved in modulating the expression of virulence-related genes [37] and in the development of penicillin nonsusceptibility [38]).

The single CC6623F representative in our collection differed from the CC66N isolates from approximately 2 kb upstream of dexB to approximately 3 kb downstream of aliA, suggesting a potential recombination import of approximately 30.4 kb, but no suitable donor was identified.

Figure 1. Temporal distribution of capsular switch variants among the historical isolate collection. Gray triangles indicate isolation dates for the oldest recombinant representatives of 36 independent capsular switch events identified among the historical pneumococcal collection. Abbreviation: PCV7, 7-valent pneumococcal conjugate vaccine.
Table 1. Genetic Evolution at the *cps* Locus Among Isolates Representing the Same Pneumococcal Serotypes and Clonal Complexes (CCs)

<table>
<thead>
<tr>
<th>CC&lt;sup&gt;Serotype&lt;/sup&gt;</th>
<th>Isolate</th>
<th>Year</th>
<th>Nucleotide Substitution(s) as Compared to Ancestor (<em>gene</em>)</th>
<th>Recombination Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>15&lt;sup&gt;14&lt;/sup&gt;</td>
<td>14/5</td>
<td>1967</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PMEN10</td>
<td>1987</td>
<td>A8695G (<em>wchM</em>); missing bp 11 509–11 525 and bp 12 423–end&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPhINV200</td>
<td>1995</td>
<td>C12439T (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICE13</td>
<td>1998</td>
<td>C12466A and C12467A (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala243</td>
<td>1998</td>
<td>C12439T (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA13338</td>
<td>1999</td>
<td>C12439T (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala317</td>
<td>2001</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICE50</td>
<td>2003</td>
<td>C12439T, C12466A, C12467A, and G12502A (<em>wciY</em>); 22-bp insertion at position 11 509</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SP14-BS292</td>
<td>Unknown</td>
<td>T1976A (<em>wzh</em>) and C12439T (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td>66&lt;sup&gt;9N&lt;/sup&gt;</td>
<td>9N/6</td>
<td>1960</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USA8</td>
<td>2001</td>
<td>A4399G (<em>wchA</em>), A14000G, T14014G, C14037T, and T14042C (between <em>ugd</em> and IS1381 putative transposase)</td>
<td></td>
</tr>
<tr>
<td>113&lt;sup&gt;18C&lt;/sup&gt;</td>
<td>18C/2</td>
<td>1939</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18C/1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1940</td>
<td>A2203G (<em>wzd</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18C/3</td>
<td>1968</td>
<td>G3478A (<em>wze</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Netherlands18C-36</td>
<td>1980</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USA2</td>
<td>1999</td>
<td>C4622T (<em>wchA</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICE501</td>
<td>2002</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>124&lt;sup&gt;14&lt;/sup&gt;</td>
<td>14/2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1952</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14/4</td>
<td>1961</td>
<td>A3109G and G3478A (<em>wze</em>), A5616G and C5617T (<em>wchK</em>), and T6662A (<em>wzy</em>)</td>
<td><em>cps</em> bp 12 330–12 467</td>
</tr>
<tr>
<td></td>
<td>PMEN35</td>
<td>1980</td>
<td>None; missing bp 12 437–end&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14/7</td>
<td>1992</td>
<td>A4141G (<em>wchA</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala292</td>
<td>1998</td>
<td>C3816A and G4087A (<em>wchA</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USA6</td>
<td>1999</td>
<td>C43A (*wzg), A12403G, C12466A, C12467A, and G12502A (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala263</td>
<td>2002</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala289</td>
<td>2002</td>
<td>C12466A, C12467A, and G12502A (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICE46</td>
<td>2003</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICE594</td>
<td>2005</td>
<td>C12466A, C12467A, and G12502A (<em>wciY</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCR11974</td>
<td></td>
<td>G2455T (<em>wzd</em>)</td>
<td></td>
</tr>
<tr>
<td>156&lt;sup&gt;162&lt;/sup&gt;</td>
<td>9V/5</td>
<td>1991</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PMEN3</td>
<td>1993</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GA08780</td>
<td>1997</td>
<td>T5230C (<em>wchO</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SP9-B568</td>
<td>Unknown</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>191&lt;sup&gt;7F&lt;/sup&gt;</td>
<td>7F/3</td>
<td>1962</td>
<td>NA; missing bp 20 430–end&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7F/5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1962</td>
<td>A16G (<em>wzd</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Netherlands7F-39</td>
<td>1984</td>
<td>None; missing bp 20 434–end&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7F/4</td>
<td>1986</td>
<td>None; missing bp 20 434–end&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICE22</td>
<td>1993</td>
<td>C20435T, A20436G, T20441C, and A20450G (<em>glf</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CDC1087–00</td>
<td>1999</td>
<td>T10697C (<em>wchF</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USA16</td>
<td>2003</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>218&lt;sup&gt;12F&lt;/sup&gt;</td>
<td>12F/5</td>
<td>1988</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Denmark12F-34</td>
<td>1995</td>
<td>G16593T (between <em>fnlA</em> and <em>fnlB</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12F/6</td>
<td>1996</td>
<td>G16593T (between <em>fnlA</em> and <em>fnlB</em>) and C18121T (<em>fnlC</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USA18</td>
<td>1999</td>
<td>G16593T (between <em>fnlA</em> and <em>fnlB</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CDC0288-04</td>
<td>2003–2004</td>
<td>C7997T (<em>wzy</em>), G16593T (between <em>fnlA</em> and <em>fnlB</em>), and C18121T (<em>fnlC</em>)</td>
<td></td>
</tr>
</tbody>
</table>
**CC113**
Eleven CC113 isolates represented serotypes 18C (n = 6), 35C (n = 2), and 9V, 17F, and 18B (n = 1 each). Serotype 18C was assumed to be the ancestral serotype and was also represented by 1 of the 2 oldest isolates, both dated 1939. The change from serotype 18C to 18B was due to a single nucleotide substitution within the \(\text{wciX} \) gene [2], while each of the other serotype changes was associated with a recombination event. Both CC113\(^{35C}\) isolates likely arose from a single capsular switch event (Table 2 and Figure 2). The serotype 17F change appears to have involved DNA acquisition from a CC57417F-like pneumococcus by a CC113\(^{35C}\) pneumococcus. The 5' region of the CC113\(^{17F}\) \(\text{cps}\) locus was distinct from that of 3 other serotype 17F isolates in our collection but differed by only 1 nucleotide substitution from the CC113\(^{35C}\) representatives. The CC113\(^{17F}\) genome was highly similar to that of a CC574\(^{17F}\) isolate between position 549 of the \(\text{wze}\) locus and approximately 3 kb downstream of \(\text{pbp1a}\), after which the CC113\(^{17F}\) genome resembled that of the CC113\(^{18C}\) isolates.

There was little to no variation at the \(\text{cps}\) locus among 5 CC113\(^{18C}\) representatives (Table 1). A sixth isolate differed by 1 nucleotide in the \(\text{wchA}\) gene and by a putative import of approximately 1.1 kb (marked by 44 nucleotide substitutions) spanning the \(\text{rmlA}\) and \(\text{rmlC}\) genes (Figure 3A).

**CC124**
Twelve of 14 CC124 representatives were serotype 14 and represented 8 unique but highly similar \(\text{cps}\) sequences (Table 1). Two isolates were serotypes 9L and 11C, but no suitable donors were identified, and the \(\text{cps}\) flanking regions of both isolates were highly divergent from those of the CC124\(^{14}\) representatives for \(\geq 30\) kb in either direction. Thus, the putative recombination breakpoints could not be estimated.

**CC156/162**
Seven isolates (dated 1952–2006) representing serotypes 9V, 9A, and 19A were assigned to CC156/162. Four isolates were serotype 9V; the oldest was serotype 9A. The CC156/162\(^{9A}\) \(\text{cps}\) locus differed from that of the oldest CC156/162\(^{9V}\) locus by 2 nucleotides [2] (Table 2 and Figure 2). Three of the 4 CC156/162\(^{9V}\) \(\text{cps}\) loci were identical, and the other differed by 1 nucleotide.

Analysis of the \(\text{cps}\) loci and flanking regions of the 2 CC156/162\(^{19A}\) representatives (GenBank accession numbers AGOR01000001–AGOR01000024 and AGQA01000001–AGQA01000007) suggested a single capsular switch event. Approximately 50.4-kb regions of the CC156/162\(^{19A}\) genomes were highly similar to those of a CC199\(^{19A}\) isolate (Table 2 and Figure 2) and included \(\text{pbp2x}\), the \(\text{cps}\) locus, and \(\text{pbp1a}\). The regions directly flanking the putative import did not resemble those of the CC156/162\(^{9V}\) representatives, and thus the true recombination breakpoints could not be estimated. However, we deduced that the original import was \(\geq 50.4\) kb.

**CC191**
Seven of the CC191 representatives were serotype 7F, and the eighth, and oldest, was serotype 7A. Serotypes 7A and 7F differed by 3 nucleotides (Table 2 and Figure 4). There was a repeated motif (5'–CTA AGA TGA ATA-3') within the \(\text{wzc}\) gene, and the number of repeats differed between isolates (n = 3, 4, or 6). Repeat motifs are difficult to sequence accurately, so further speculations about such changes cannot be made. Apart from the repeat region, the CC191\(^{7F}\) \(\text{cps}\) loci each differed by a maximum of 4 nucleotides (Table 1 and Figure 3B).

**CC218**
Eight CC218 isolates represented serotypes 12F (n = 5) and 7F (n = 3). Between approximately 9 kb upstream of \(\text{pbp2x}\) and approximately 4.8 kb downstream of \(\text{pbp1a}\), CC218\(^{12F}\) and CC191\(^{7F}\) sequences were very similar, although differentiating nucleotides were identified (Table 2). The clustering of 44 nucleotide substitutions within the \(\text{cps}\) locus was indicative of recombination events (Figure 3B), but whether these events occurred before or after the \(\text{cps}\) switch is unknown. Note that approximately 585 bp were missing between \(\text{pbp2x}\) and \(\text{dexB}\) in both CC218\(^{12F}\) and CC191\(^{7F}\) representatives. Table 1 provides additional details about nucleotide substitutions among the \(\text{cps}\) loci of CC218\(^{12F}\) and CC218\(^{12F}\) isolates.
Two CC574 isolates were examined: a serotype 17F isolate from 1952 and a serotype 2 isolate from 1956. The sequences differed from position 326 of *dexB* to approximately 5 kb downstream of *pbp1a*. A region of the serotype 2 representative (totaling approximately 35 kb, from approximately 0.6 kb downstream of *dexB* to approximately 5 kb downstream of *pbp1a*) was highly similar to that of a CCNone1282 representative dated 1916 (Table 2 and Figure 4). A short, 314-bp region of unknown sequence in the *dexB* locus was also present.
Figure 2. Genetic changes leading to capsular switching within clonal complexes (CCs) 66, 113, and 156/162. Bars represent the cps locus and flanking regions. Filled arrows indicate the relative positions of the dexB (left) and alIA (right) loci. Open arrows indicate the relative positions of the pbp2x (left) and pbp1a (right) loci. Vertical bisecting lines indicate single nucleotide substitutions. Recombination fragments are indicated by coloring. Identical sequences are indicated by identical coloring. Regions bounded by solid lines are depicted at their maximum length. Regions not bounded by solid lines are depicted at their minimum length (Table 2). The asterisk indicates a single-base-pair deletion. The schematic is drawn approximately to scale.

Figure 3. Variable site maps for the cps loci of serogroups 18 and 7. Rows represent the nucleotide sequences of independent isolates and are labeled as isolate name (year). Nucleotide differences relative to row 1 are shown. Periods indicate an identical base, and hyphens indicate a missing base. Numbers above each column indicate the nucleotide position within the alignment. Nucleotide substitutions marking putative recombination regions are marked in bold and underlined. A, CC113 isolates include 1 of serotype 18B (18B/2) and 6 of serotype 18C. B, CC191 isolates include 1 of serotype 7A (7A/2) and 7 of serotype 7F. CC218 isolates represent serotype 7F (B*). The arrow indicates the position of a single-base-pair insertion in the 7A/2 sequence, compared with the other serogroup 7 sequences.
DISCUSSION

Within our collection of historical and modern pneumococci, we identified 36 independent capsular switching events. Approximately 94% of the variants were isolated prior to the introduction of PCV7 and were roughly evenly distributed through time (Figure 1). The collection was not designed for inferring a capsular switching rate; nevertheless, these data imply that this phenomenon has been a regular occurrence (ie, there is evidence of capsular switching within a diverse range of CCs) every decade throughout the past 7 decades).

Analysis of the *cps* loci of 10 representatives of CC66 indicated multiple independent changes to the same serotype, supporting the notion that capsular switching may occur regularly among pneumococci. Within this CC, an initial capsular switch from serotype 7B to 9N no later than 1952 was followed by at least 2 independent changes to serotype 19F no later than 1972 and 2005 and by at least 3 independent changes to serotype 14, each no later than the mid-1990s. This is consistent with previous studies that demonstrated multiple changes of serotype within the same CC: CC156/162 9V → 14 [39–41], CC8123F 19F/A [30], and CC695 4–19A [25, 28].

We studied 18 capsular switching events in detail. Three of these were presumably the result of nucleotide substitution and/or deletion, a finding consistent with previous work [2]. The remaining 15 events appeared to be due to recombination, and breakpoints for 11 events could be estimated by comparison of the *cps* loci and flanking sequences of the putative ancestors, donors, and recombinants. The capsular switch recombination fragments identified here (ie, imports of various lengths, inserted at different points around the *cps* locus, with or without the adjacent *pbp* sequences) are consistent with fragments detected in previously published studies [25, 42, 43]. Figure 5 depicts the exchange of *cps* loci between CCs, as inferred from nucleotide sequence analyses for 15 capsular switching events. Donor CCs could not be predicted for a total of 7 events because donor isolates were not present among this collection. Arrow labels indicate serotypes associated with the transferred *cps* loci and isolation year of the oldest recombinant pneumococcus in this collection.

**Figure 4.** Genetic changes leading to capsular switching within clonal complexes (CCs) 191, 218, and 574. Bars represent the *cps* locus and flanking regions. Filled arrows indicate the relative positions of the *dexB* (left) and *aliA* (right) loci. Open arrows indicate the relative positions of the *pbp2x* (left) and *pbp1a* (right) loci. Vertical bisecting lines indicate single nucleotide substitutions. Recombination fragments are indicated by coloring. Regions bounded by solid lines are depicted at their maximum length. Regions not bounded by solid lines are represented at their minimum length (Table 2). The asterisk indicates a single-base-pair insertion. The schematic is drawn approximately to scale.

**Figure 5.** Inter-clonal complex (CC) *cps* locus transfers inferred in this study. Circles represent pneumococcal CCs as named. Arrows indicate transfer of *cps* loci between CCs, as inferred from nucleotide sequence analyses for 15 capsular switching events. Donor CCs could not be predicted for a total of 7 events because donor isolates were not present among this collection. Arrow labels indicate serotypes associated with the transferred *cps* loci and isolation year of the oldest recombinant pneumococcus in this collection.
conclusions about the number of independent capsular switches and the range of recombination fragment sizes would remain unchanged.

Our analyses did not indicate any recombination breakpoint hot spots around the \textit{cps} locus, which had also not been indicated by any previous studies [25, 27, 28, 30]. It is impossible to know whether the putative imports were acquired through single or multiple recombination events, but the former is most parsimonious. The import lengths were estimated to range from approximately 19.0 to $\geq$58.2 kb (and apart from 1 example, always included the entire \textit{cps} locus) and did not show any trends toward increasing/decreasing lengths through time.

The CC218$^{12F\rightarrow7F}$ and CC66$^{7B\rightarrow9N}$ events were characterized by large (>50 kb) recombination imports and must have taken place no later than 1952 and 1960, respectively. Large-scale recombination between pneumococci has therefore clearly been occurring for decades, although the technology capable of detecting and detailing such events has only recently become available [25, 28, 30, 42]. Additionally, both of these putative imports included the \textit{pbp2x} + \textit{cps} + \textit{pbp1a} loci, as reported for other events [27, 28] and for another newly characterized event in this study (CC156/162$^{9V\rightarrow19A}$).

Recently reported in vivo \textit{pbp2x} + \textit{cps} + \textit{pbp1a} recombination events were detected soon after widespread vaccination began in the United States. No such recombinants had been reported before in nature, and the imported \textit{pbp2x} and \textit{pbp1a} sequences conferred penicillin nonsusceptibility. Thus, it was theorized that vaccine-induced and/or antibiotic-induced selective pressures may play a role in driving these genetic changes [28, 44]. Penicillin was introduced in the 1940s; oral penicillins were not available until the mid-1950s, and their use was initially limited. Consequently, our analyses question the above theory because the CC218$^{12F\rightarrow7F}$ and CC66$^{7B\rightarrow9N}$ events were both associated with penicillin-susceptible pneumococci and occurred before the widespread use of penicillins and before the introduction of PCV7. Our new data suggest that recombination of the \textit{cps} locus and flanking regions might be “normal” biological processes, the evolution of which has undoubtedly been influenced by naturally occurring immunity and other selective pressures. Presumably, vaccine-induced immune pressures and/or the pressure of antibiotic use subsequently influence the spread and maintenance of advantageous genes (and/or alleles) by selecting recombinants that are best able to survive. However, the potential negative effects should not be underestimated: the CC695$^{4\rightarrow19A}$ “vaccine escape” capsular switches in the United States were first recovered from patients with pediatric invasive pneumococcal disease only 3 years after PCV7 introduction and, 2 years later, were the third-most common serotype 19A CC causing invasive pneumococcal disease among all age groups [20]. These strains were penicillin nonsusceptible, owing to the simultaneous acquisition of altered \textit{pbp2x} and \textit{pbp1a} genes. Consequently, the increase in prevalence of these strains likely contributed to the increase in pneumococcal penicillin nonsusceptibility in the United States after PCV7 introduction [20, 45].

Another interesting finding was that a CC199$^{19A}$-like representative was the most probable donor of the 19A \textit{cps} locus and flanking \textit{pbps} to the CC156/162$^{19A}$ isolates described in this study. CC199$^{19A}$ representatives were also the donors of the \textit{cps} locus + \textit{pbps} to the vaccine escape progeny [25, 28] and an ST320$^{19A\rightarrow19A}$ \textit{cps} locus switch [27]. Future studies will attempt to uncover an explanation of why CC199$^{19A}$ representatives appear to be “good” \textit{cps} locus donors.

Our analyses also revealed recombination at the \textit{cps} locus that did not result in capsular switching, among isolates belonging to the same ancestral lineages (CC113$^{18C}$ and CC191$^{7F}$) and different ancestral lineages (CC191$^{7F}$ vs CC218$^{7F}$ and CC15$^{14}$ vs CC124$^{44}$). Similar events have been noted within the CC81$^{231}$ lineage [26] and among serogroup 6 and 19 isolates [10–12]. We speculate that this is more likely to be the result of recombination whereby some genes are coincidentally exchanged, rather than the result of exchange that occurred directly in response to selection pressure. It could also be the result of DNA repair mediated by recombination at or near the \textit{cps} locus. A comparison of the \textit{cps} loci of CC15$^{14}$ and CC124$^{14}$ representatives also indicated a deletion in the CC15$^{14}$ \textit{wciY} gene, resulting in a predicted truncation of the putative glycerol phosphotransferase encoded by this gene. Notwithstanding the deletion, the CC15$^{14}$ isolates were successfully serotyped by the Quellung reaction, indicating successful capsule production. Indeed, an in silico analysis of the predicted \textit{cps} protein coding regions failed to identify a specific reaction catalyzed by the serotype 14 version of this protein [46].

Given the overall capacity for \textit{cps} locus recombination, the associations between genotypes and serotypes are puzzling, as is our evidence of a high level of \textit{cps} sequence conservation within some CCs (eg, CC191$^{7F}$ and CC218$^{12F}$). Perhaps there is some synergism between serotype and the genetic background of a strain that conveys an advantage to certain combinations over others. This sort of synergistic effect has been invoked to explain changes in the pneumococcal population in South Korea, where there was a prevaccine reduction of multidrug-resistant CC271/320$^{19F}$ pneumococci and replacement by similarly multidrug-resistant CC271/320$^{19A}$ pneumococci over several years [47]. It is also possible that genes, such as those encoding sugar-biosynthesis enzymes, outside the \textit{cps} locus contribute to capsular expression and facilitate a genotype/serotype association [2].

Our collection of pneumococci provided a unique opportunity to study evolution at the \textit{cps} locus over approximately 70 years. Capsular switching with/without simultaneous \textit{pbp}}
transfer has occurred regularly and prior to both PCV7 introduction and widespread antibiotic use. It is highly likely that the proliferation of newly generated NVT capsular switch variants will continue to be favored by PHID-CV and PCV13 vaccination programs, as was the case after PCV7 implementation. Penicillin-nonsusceptible variants will have an even greater advantage. Although the magnitude of these selective forces relative to those favoring established genotype/VT associations remains unclear, the implementation of vaccine programs across the globe will most likely favor the intercontinental spread of NVT and penicillin-nonsusceptible pneumococci.

Supplementary Data

Supplementary materials are available at The Journal of Infectious Diseases online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank Dalia Denapaite, Irma Ochigava, and Shwan Rachid, for DNA sequencing and MLST analysis; Andries van Tonder, for retrieval of genomic sequence data from GenBank; and the clinicians, microbiologists, and investigators of the Active Bacterial Core surveillance program of the Emerging Infections Program Network, United States.

Financial support. This work was supported by the Wellcome Trust (grant 083511/Z/07/Z to A. B. B. and grant 098051 to J. P. and S. D. B.), Stiftung Rheinland Pfalz für Innovation (to R. H.), and Ciber de Enfermedades Respiratorias (Instituto de Salud Carlos III, Madrid Spain [to J. L.]). A. B. B. is a Wellcome Trust Career Development Fellow.

Potential conflicts of interest. K. P. K. received research funding from and is a consultant for Pfizer Vaccines and is a consultant for GlaxoSmithKline Biologicals and Merck. A. B. B. received research funding from GlaxoSmithKline Biologicals. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References


