We discovered 3 invasive, multidrug-resistant *Streptococcus pneumoniae* isolates of vaccine-refractory capsular serotype 3 that recently arose within the successful sequence type 271 complex through a serotype switch recombination event. Mapping genomic recombination sites within the serotype 3/sequence type 271 progeny revealed a 55.9-kb donated fragment that encompassed *cps3, pbp1a*, and additional virulence factors.

The polysaccharide capsule, of which there are >90 structurally and serologically unique types, is the primary pneumococcal virulence factor (5). Serotype 3 is historically associated with higher virulence than other serotypes and currently causes the largest proportion of IPD cases in the United States (>12% of all cases) (4). Although serotype 3 is included in PCV13, PCV13 provides poor protection against serotype 3 because of unique qualities of the serotype 3 capsule (6,7). In keeping with the ST271 heritage, the 3 serotype 3/ST271 isolates share the same antimicrobial-resistance mechanisms, including reduced affinities for β-lactams (mosaic PBPs), dual mechanisms for macrolide resistance (ErmB rRNA methylase and MefA/MsrD macrolide efflux system), clindamycin resistance (ErmB), cotrimoxazole resistance (altered FolA and FolP enzymes), and tetracycline resistance (TetM-mediated ribosome alteration) (Appendix 1, https://wwwnc.cdc.gov/EID/article/27/6/20-3629-App1.xlsx).

To reveal genomic regions within the 3 serotype 3/ST271 progeny resulting from recombination, we first identified likely recipient and donor strains involved in the serotype switch event (Appendix 2, https://wwwnc.cdc.gov/EID/article/27/6/20-3629-App2.pdf). Phylogenetic analysis using publicly available genome sequences revealed single contig ST271 genomes that shared highest relatedness with 3/ST271. Strain A026 (19F/ST271) recovered in China during 2006–2008 was the most highly related putative genetic recipient (Figure, panel A). Two 19F/271
invasive ABCs isolates recovered from infants (6299-05 in Tennessee in 2004 and 2012214924 in California in 2012) had the closest matching PBP type to serotype-switch 3/ST271 isolates, sharing 2 of 3 PBP sequences (PBP2B-16 and PBP2x-47) (Appendix 1). Both ABCs 19F/ST271 isolate genomes shared more relatedness with 3/ST271 than public available 19F/271 genomes (Figure, panel A).

By using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and whole-genome shotgun database (Appendix 2), we identified the likely cps3 donors (3/ST700 strains B20605 and 73D368810). B20605 and 73D368810 shared sequence identity to the 9522 bp region (dexB to pgm) encompassing the ≈5,000-bp cps3 operon of the 3/ST271 isolates. In contrast to serotype 3 strains in the United States that are typically basally

![Diagram A](image1)

![Diagram B](image2)

![Diagram C](image3)

Figure. Streptococcus pneumoniae serotype 3/ST271 lineage resulting from a recombination event between a 19F/ST271 recipient and 3/ST700 donor. A) Phylogenetic tree showing progeny serotype 3/ST271 isolates 20155315-S-ABC, 20170822-S-ABC, and 20182806-S-ABC (red arrow) most closely related to the putative recipient 19F/ST271 isolates 6299-05 and 2012214924 (blue arrows). The most closely related ST271 single contig reference is A026 (also indicated with blue arrow). Branch colors: yellow, additional 19F/ST271 Active Bacterial Core surveillance (ABCS) isolates and single contig references; red, single-locus variant single contig references or ABCs isolates; orange, double-locus variant single contig reference or ABCs isolates; blue, ST320 ABCs isolates. Zero-, single-, and double-locus variant, single contig references were identified by using the PubMLST database (https://pubmlst.org). Scale bar corresponds to 1,062 single nucleotide polymorphisms. B) Phylogenetic alignment of the 3 recombinant serotype 3/ST271 isolates and closest known genomic matches of the ST271 recipient lineage and a schematic of recombinant genome fragments, represented by rectangular blocks, that were predicted by Gubbins (10). Block locations and sizes are relative to the aligned genomes; red blocks represent sites in common between ≥2 isolates (1 site, <150 bp in length, was not counted among the 17 total shown), blue blocks sites unique to a given isolate, and gray blocks the serotype-switch fragment that replaced the corresponding cps19F region within the 19F/ST271 strain. The cps3 locus was not identified using Gubbins because of its complete divergence from cps19F and instead was identified using ProgressiveMauve (12) within the encompassed 55.9 kb fragment (panel C). Gray block contains cps3, pbp1a (PBP1A-17), trigger factor, and choline binding protein G genes. C) Schematic illustrating ancestral recombination event between the 3/ST700 donor (B20605) and 19F/ST271 recipient (A026) to yield the 3/ST271 progeny (20155315-S-ABC). Deduced crossover points, including coordinates in the progeny, are shown. luxS and sps8 genes are shown as blue/red hybrids. The minimum genes of the cps3 operon required for polysaccharide capsuole biosynthese is shown in gray (ugd, wchE, gtaB). Genes with arrows in black differ between B20605 and 73D368810 and are absent in the 3/ST271 isolates. ST, sequence type.
susceptible to antimicrobial drugs (4), the likely serotype 3/ST700 donor strain was predicted to have intermediate penicillin resistance attributable to the mosaic \( \text{pbp1a} \) gene (PPB1A-17), which was co-transferred with the donor \( \text{cps3} \) locus to the 19F/ST271 recipient (Appendix 1). Serotype 3/ST700 isolates are documented in several countries in Africa but not in the United States (https://pubmlst.org/spneumoniae).

To identify recombination sites, we mapped sequencing reads from the 3 3/ST271 isolates, 6299-05 (19F/ST271), and 2012214924 (19F/ST271) to A026 (19F/ST271) and then aligned the 6 genomes as described (8,9) (Appendix 2). The aligned 19F/ST271 and 3/ST271 genome sequences were input into Gubbins (10) (Appendix 2), which identified 17 recombinatorial fragments within 3/ST271 (median size 6,629 bp [range 363–59,159 bp]) (Figure, panel B). BEDTools (9) was applied to extract DNA sequences from recombinational fragments, and Prokka (11) was used to annotate sequences. Progressive Mauve (12) revealed the \( \text{cps3} \) operon in the 3 serotype 3/ST271 isolates on a 55.9-kb fragment that apparently originated from the serotype 3/ST700 donor (Figure, panel C). This region included the \( \text{cps3} \) operon, \( \text{pbp1a} \), and 2 additional virulence factor genes (trigger factor \( \text{tig} \) and choline binding protein G \( \text{cbpG} \)). The corresponding region in 3/ST700 isolate B20605 contains genes that differ from serotype 3/ST700 isolate 73D36881 and are absent in the 3/ST271 isolates (Figure, panel C). This difference indicates that additional gene insertions occurred within this potential recombination hotspot in B20605 and 73D36881 and that this region subsequently might have been lost in the serotype 3/ST271 progeny.

When we compared the genes encoded in the serotype-switch region of the recipient with the corresponding region from the progeny and donor, we found they encoded near-identical virulence factors as identified by using the Virulence Factors Database (http://www.mgc.ac.cn/VFs/main.htm). The \( \text{tig} \) and \( \text{cbpG} \) genes have roles in pathogenesis and are cell wall surface–localized, highly conserved among \( S. \ pneumoniae \), and immunogenic (13–15). Capsular polysaccharides provide the basis of approved \( S. \ pneumoniae \) vaccines. The limited efficacy of the serotype 3 component of PCV13 is probably attributable to high-level expression (thickness) of the serotype 3 capsule and its shedding by the bacteria (6,7). At the nucleotide level, \( \text{tig}^{\text{20155315-S-ABC}} \) and \( \text{tig}^{\text{20170822-S-ABC}} \) share 99.1% identity (11 SNPs apart); \( \text{cbpG}^{\text{20155315-S-ABC}} \) and \( \text{cbpG}^{\text{20170822-S-ABC}} \) share 51.0% identity, and \( \text{cps3} \) and \( \text{cps19F} \) operons are highly divergent (5). Moreover, the \( \text{pbp1a} \) gene within the 55.9-kb genomic fragment is identical between the 3 3/ST271 isolates and donor 3/ST700 strains and exhibits only 81.7% identity with the putative recipient 19F/ST271 strains. Recombination introduced a distinct virulence factor (replacement of the serotype 19F capsule with a structurally and serologically unique serotype 3 capsule) and introduced diversity in the surface protein virulence factors expressed in the serotype 19F/ST271 lineage.

To understand the effects of recombination, a comparison of fitness and virulence of progeny and parental strains in mouse models might be valuable. Our study highlights the existence of the 3/ST271 strain, because currently no vaccine is available and few antimicrobial drugs are predicted to optimally target this strain should it gain traction. Asymptomatic nasopharyngeal carriage serves as the major reservoir of highly common noninvasive infections and precedes invasive infections. Three nearly isogenic invasive isolates of this clone appearing over an expanse of 3 years in 2 different states is a concerning sign of a successful foothold within the carriage reservoir. All 3 cases were caused by bacteremic pneumonia in middle-aged patients. Of particular interest is the isolation of this strain 16 months apart from blood specimens from the same person (isolates 20155315 and 20170822), which is a rare occurrence and suggests a failure to eradicate the organism and long-term carriage or potentially the reacquisition of the strain from persons in the community. The recovery of still another invasive isolate of this same clone in 2018 further suggests a level of fitness required for long-term survival in the carriage reservoir and potential to cause IPD. Further, after the review process of this manuscript, ABCs has reported recovery of 2 additional closely related 3/ST271 case isolates during 2019 from Minnesota (B. Beall, unpub. data, 2020). This new strain complex has unique features potentially guiding its success, including high antimicrobial resistance, a capsule ineffectively targeted by PCV13, and undefined parameters inherent to the predominant invasive lineage of the post-PCV7 era. These observations emphasize the clear need for a pneumococcal vaccine that effectively targets serotype 3.

Acknowledgments

We thank Zhongya Li, Hollis Walker, Theresa Tran, Sopio Chochua, and Lesley McGee for whole-genome sequencing. We also acknowledge the entire ABCs program for recovery of these novel invasive isolates.
About the Author

Dr. Scherer is an assistant professor in the Division of Infectious Diseases, Department of Medicine, School of Medicine, Emory University, Atlanta, Georgia, USA. Her research interests include understanding what immune responses generate durable immunity against infectious diseases, with a focus on antibody and B cell responses to viral and bacterial pathogens and their vaccines.

References


Address for correspondence: Bernard Beall, Centers for Disease Control and Prevention, 1600 Clifton Rd NE, Mailstop C02, Atlanta, GA 30329-4027, USA; email: bbeall@cdc.gov
Pneumococcal Serotype-Switch Variant of Multidrug-Resistant *Streptococcus pneumoniae* Sequence Type 271

Appendix 2

**Methods**

**Identification of Donor and Recipient Strains**

NCBI’s Nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to find SPN isolates with highest identity to 3/ST271 cps3 operon (dexB to pgm), with whole genome-shotgun (wgs) database and *Streptococcus pneumoniae* organism (taxid: 1313) as search parameters.

kSNP3 version 3.0.0 (https://sourceforge.net/projects/ksnp/files) was applied to find the closest single contig reference to 3/ST271 isolates, with a Kchooser determined kmer of 23 and “–core” command line argument to calculate the phylogenetic tree based on SNPs found in all genomes.

Interactive Tree of Life (iTOL version 5; https://itol.embl.de) was used to visualize phylogenetic trees resulting from kSNP3 by re-rooting the tree to the branch containing single and double locus variant single contig references.

Figure, panel A was generated using kSNP3 and iTOL as described above. Figure, panel B was generated using Prokka annotations of indication regions and Easyfig (https://mjsull.github.io/Easyfig). Figure, panel C (https://wwwnc.cdc.gov/EID/article/27/6/20-3629-F1.htm) was generated using gubbins_drawer.py.

**Read Mapping**

Bowtie2 version 2.2.4 (https://github.com/BenLangmead/bowtie2) was applied to map whole genome sequencing reads from ABCs isolates 20155315, 20170822, 20182806, 6299–05 (ERR600161), and 2012214924 (ERR600302) to A026 with the “–very-sensitive-local” option.
Samtools suite version 0.1.18 (http://www.htslib.org) was applied to order the mapped reads with reference to A026 using the “view,” “sort” and “index” commands.

FreeBayes version 0.9.21 (https://github.com/ekg/freebayes) was used to align the indexed, sorted genome sequences of 20155315-S-ABC, 20170822-S-ABC, 20182806-S-ABC, ERR600161, ERR600302, and A026, with the “-j” (omit indels) and “- = ” options and specifying a base and mapping quality of 20 (“-q 20” and “-m 20”).

Vcflib suite version 2015.10.19 (https://github.com/vcflib/vcflib#vcflib) was applied to remove weakly called variants in the read mapping alignment file generated by Bowtie2, Samtools, and Freebayes with the vcffilter command, and the following “-f” and “-g” options: -f “QUAL >200 & DP >15 & TYPE = mnp | QUAL >200 & DP >15 & TYPE = snp” -g “GQ >30 & AO >10 & GT = 1”

Bedtools suite version 2.17.0 (https://github.com/arq5x/bedtools2) genomecov utility was applied to generate a file describing regions of low coverage (<10x) with the “-bga” option. The maskfasta utility was used to mask out all low coverage regions with N’s.

**Gubbins Recombination Analysis**

The input for Gubbins version 2.3.1 (https://github.com/sanger-pathogens/gubbins) was aligned sequences (20155315-S-ABC, 20170822-S-ABC, 20182806-S-ABC, ERR600161, ERR600302, and A026) where low coverage regions had been masked with N’s. Default settings were used with the “-i” option set for 20 iterations and with A026 specified as the outgroup. The gubbins_drawer.py command was used to generate a PDF of the putative recombinant regions using the output .tre and .embl files.

A .bed file was generated from the recombination sites’ coordinates using the Gubbins output .gff file.

Bedtools suite version 2.17.0 getfasta utility was used with this .bed file to generate an output file with all of the recombination site sequences extracted from the original Gubbins input file. NCBI’s Nucleotide BLAST was then used to identify SPN isolates with highest identity to these recombination sites, with nucleotide collection (nr/nt) database and *Streptococcus pneumoniae* organism (taxid: 1313) as search parameters. Recombinant sites <300 bp in length
(n = 3; 121 bp, 36 bp, and 110 bp) were omitted from the analysis. No regions with identity to serotype 3 were found.

Extracted regions were also annotated using Prokka version 1.14.3 (https://github.com/tseemann/prokka) with “--addgenes” and “--compliant” options; however, a cps3 operon was not identified.

**Progressive Mauve Recombination Analysis and Prokka Annotation**

Abacas version 1.0 (http://abacas.sourceforge.net) was used to sort and index whole genome assemblies from 20155315-S-ABC, 20170822-S-ABC, 20182806-S-ABC, ERR600161, and ERR600302 with reference to A026 with the “-m,” “-N,” and “-a” options.

Aligned genomes were annotated with Prokka version 1.14.3 using “--addgenes” and “--compliant” options with and without “--genus Streptococcus” and “--species pneumoniae” options.

Prokka annotations were input into progressiveMauve version 2.4.0 (http://darlinglab.org/mauve/download.html). Initially the cps3 operon and ultimately the ~55.9kb recombination site were identified by manually searching for regions of dissimilarity between the three 3/ST271 isolates 20155315-S-ABC, 20170822-S-ABC, 20182806-S-ABC and the two 19F/ST271 isolates ERR600161 and ERR600302 within the appended sequences (“--a” option in Abacas).

Pairwise alignments were conducted using EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle) or Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo).

**GenBank Accession Numbers**

Isolates were submitted to GenBank (accession nos. 73D36881, B20605 (NZ_FIXO01000000, NZ_FIXN01000000), A026 (CP006844.1).