



Published in final edited form as:

J Infect Dis. 2025 April 15; 231(4): 871–882. doi:10.1093/infdis/jiae453.

Increased Proportions of Invasive Pneumococcal Disease Cases Among Adults Experiencing Homelessness Sets the Stage for New Serotype 4 Capsular-Switch Recombinant

Bernard Beall¹, Sopio Chochua², Ben Metcalf², Wuling Lin², Theresa Tran², Zhongya Li², Yuan Li², Meghan L. Bentz³, Mili Sheth³, Gunars Osis⁴, Lesley McGee²

¹Contractor to Division of Bacterial Diseases, Centers for Disease Control and Prevention, Eagle Global Scientific, LLC, Atlanta, Georgia, USA

²Respiratory Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

³Biotechnology Core Facility Branch, Division of Core Laboratory Services and Response, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

⁴Contractor to National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Abstract

Background.—The Centers for Disease Control and Prevention’s Active Bacterial Core surveillance (ABCs) identified increased serotype 4 invasive pneumococcal disease (IPD), particularly among adults experiencing homelessness (AEH).

Methods.—We quantified increased proportions of IPD cases in AEH during 2016–2022. Employing genomic-based characterization of IPD isolates, we identified serotype-switch variants. Recombinational analyses were used to identify the genetic donor and recipient strain that generated a serotype 4 progeny strain. We performed phylogenetic analyses of the serotype 4 progeny and of the serotype 12F genetic recipient to determine genetic distances.

Results.—We identified a cluster of 30 highly related (0–21 nucleotide differences) IPD isolates recovered during 2022–2023, corresponding to a serotype 4 capsular-switch variant. This strain arose through a multifragment recombination event between serotype 4/ST10172 and serotype

Correspondence: Bernard Beall, PhD, Division of Bacterial Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333 (bbeall@cdc.gov).

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.

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.

Potential conflicts of interest. All authors: No reported 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.

12F/ST220 parental strains. Twenty-five of the 30 cases occurred in Oregon. Of 29 cases with known residence status, 16 occurred in AEH. Variant emergence coincided with a 2.6-fold increase (57 to 148) of cases caused by the serotype 4/ST10172 donor lineage in 2022 compared to 2019 and its first appearance in Oregon. Most serotypes showed sequential increases of AEH IPD/all IPD ratios during 2016–2022 (eg, for all serotypes combined, 247/2198, 11.2% during 2022 compared to 405/5317, 7.6% for 2018–2019, $P < .001$). Serotypes 4 and 12F each caused more IPD than any other serotype in AEH during 2020–2022 (207 combined case isolates accounting for 38% of all IPD in AEH).

Conclusions.—Expansion and increased transmission of serotypes 4 and 12F among adults potentially led to recent genesis of an impactful hybrid serotype-switch variant.

Keywords

invasive pneumococcal disease; adults experiencing homelessness; serotype switch variant; multifragment recombination; western Active Bacterial Core Surveillance sites

Active Bacterial Core surveillance (ABCs) performs active and population-based invasive pneumococcal disease (IPD) surveillance in 10 states [1]. Serotypes 4 and 12F have been the 2 serotypes causing the highest numbers of IPD cases in adults experiencing homelessness (AEH) in ABCs in the USA during 2015–2019 [2–5]. These 2 serotypes frequently comprise IPD genomic clusters (defined as isolates differing by no more than 10 nucleotide polymorphisms) in adults [2, 3]. For 400 serotype 4 ABCs cases reported during 2010–2018 [4], all but 7 were from adults ≥ 18 years of age and 82% (328) occurred in 3 of the 10 ABCs sites (California, Colorado, and New Mexico). From 2010 to 2018, serotype 4 incidence increased from 0 to 301 cases per 100 000 AEH in Colorado and 0 to 274 per 100 000 AEH in New Mexico [4]. In 2018, serotype 4 incidence was 100–300 times higher among AEH versus adults not experiencing homelessness (ANH) in these 2 states.

We first discovered the serotype multilocus sequence type 4 (4/ST10172) lineage in a single IPD isolate from 2013 IPD surveillance within ABCs [6, 7]. During 2015–2018, 38 serotype 4/ST10172 cases were identified among AEH, all within Colorado and New Mexico, which accounted for 43.2% of all ABCs serotype 4 cases among AEH. A serotype 4/ST10172 strain recovered during 2022 was reported carrying a vancomycin-resistance determinant [8]. Serotype 12F, primarily the closely related 12F/ST220 and 12F/ST218 lineages, also disproportionately causes genomic clusters of IPD cases among AEH in western states [2, 3]. Here we report a recombination event involving 2 unrelated strains (serotype 4/ST10172 and 12F/ST220) previously associated with IPD genomic clusters among AEH [2–4], that culminated in a disease cluster caused by the hybrid serotype 4 progeny strain.

METHODS

ABCs Data

ABCs IPD cases are defined through isolation of *Streptococcus pneumoniae* from a normally sterile site in a surveillance area resident. The surveillance sites included entire states or selected counties in 10 states, representing approximately 32.2–35 million individuals during 2016–2022 [1]. Case medical charts were reviewed to attain demographic

information, including residence type and clinical information. AEH were defined as adults documented as homeless or residing in a shelter, mission, medical respite, or church community center at the time of positive culture. ABCs data were determined to be nonresearch, public health surveillance, such that review by an institutional review board was not required.

Isolate Characterization

Isolates were assigned serotypes and multilocus sequence types (STs) employing whole-genome sequencing [7, 9]. All ABCs isolates received since January 2015 are subjected to short-read sequencing with genome sequences deposited in the National Center for Biotechnology Information Sequence Read Archive under BioProject number PRJNA284954. A range of 76.6%–95.5% of case isolates were assigned serotypes in the 10 sites during 2016–2022, with the average being 87.2%. Of 19 457 isolates from ABCs IPD cases in 2016–2022, 17 017 (87.5%) were sent to the Centers for Disease Control and Prevention (CDC), of which 16 964 (99.7%) were assigned serotypes. For the incidence data, correction factors were used to estimate serotype 4 rates within each of the 4 western sites as follows: (reported serotype 4 cases) $\times$ (all IPD cases $\div$ all IPD cases assigned serotypes). Selected deduced serotype 4 isolates were subjected to Quellung testing and immunodiffusion tests against CDC type 4 typing antisera [10].

Long-Read Sequencing

The PacBio Microbial Multiplexing procedure was implemented (Pacific Biosciences). Genomic DNA extracted for short-read sequencing was also employed for 3 representative genomes (isolate identifiers 20234287 [4/ST10172], 20166336 [12F/ST220], and 20234295 [4/ST18704]). Libraries were generated with the SMRTbell Express Template Prep Kit 3.0. Libraries were size-selected using a 0.45 $\times$ Ampure bead cleanup (Pacific Biosciences) to remove small DNA (<5 kb). Final libraries were sequenced for 30 hours after 30 minutes preextension times on a Sequel II (Pacific Biosciences) using Sequel II Binding Kit 3.2. PacBio HiFi reads were utilized in Flye version 2.9 [11, 12] to conduct de novo assembly for the genome size of 2.13 Mb. Assembled circularized contigs were evaluated using BLAST+ version 2.9.0 [13]. Read coverages of 327–681 were obtained for the 3 genomes that were assigned accession numbers CP152418 (20234287), CP154876 (20155336), and CP154877 (20234295).

Sequence Analyses

Prokka version 1.14.5 was used to obtain annotated open reading frames (ORFs) [14]. Recombinant regions of 3 representative strains were aligned to generate figures using Easy Fig version 2.2.3 [15]. Core genomic maximum parsimony genomic trees and pairwise SNP matrices were generated from short-read bacterial genome sequences employing kSNP3.0 with a kmer size of 19 [16]. Resultant core.tre files generated from kSNP3.0 were used to generate core genomic phylogenetic diagrams employing Mega7 [17]. Progressive Mauve [18] was employed to generate consensus 1 944 951 bp core genomes of two 4/ST18704 isolates plus the two most related 12F/ST220 isolates. The core alignment was subjected to Gubbins [19] to identify recombinant regions. Circular maps were generated employing

Proksee [20]. BLAST comparisons of ORFs within recombinant regions corresponding to the progeny, donor, and recipient strains were generated employing Proksee.

Statistical Analyses

Fisher exact tests were used for the different comparisons in this study evaluating underlying risk factors (experiencing homelessness, time periods, different serotypes). Differences were considered statistically significant at $P < .05$ (2-sided P values).

IPD Incidence

Denominators for IPD incidence calculations for adults ≥ 18 years of age within each of the 4 western ABCs sites were obtained from US Census Bureau estimates within ABCs counties [21]. For estimates of IPD among AEH, we used US Department of Housing and Urban Development (HUD) point in time data [22] for the 4 corresponding geographic areas only for years 2019 and 2022 because point in time data methodology was modified during 2020–2021.

RESULTS

Incidence of IPD Serotypes During 2016–2022 With Focus on Serotypes 4 and 12F

Overall IPD incidence calculated for ABCs areas during 2016–2019 was stable at between 9.2 and 9.6 cases per 100 000 individuals each year and, presumably due to nonpharmaceutical interventions, plummeted to 5.3–5.4 cases per 100 000 during 2020–2021, before increasing to 8.3 cases per 100 000 during 2022 [1]. There was a trend of disproportionate increases of IPD in AEH relative to ANH in prepandemic (2018–2019 compared to 2016–2017), pandemic (2020–2021 compared to 2018–2019), and postpandemic (2022 compared to 2018–2019) years (Table 1), as shown by the proportions of combined reported cases due to the 6 serotypes causing the most IPD in AEH (the first 6 depicted in Figure 1), combined reported cases due to the remaining serotypes, and combined reported cases due to all serotypes. Twenty-six of the 27 percentage comparisons (5 of the 6 individual serotypes, 6 serotypes combined, remaining serotypes, all serotypes) showed increases of the later time period compared to the earlier time period. All increased ratios for the combined cases due to the 6 serotypes causing the most IPD among AEH, as well as for all serotypes combined, were highly significant ($P < .001$).

Except for 2 isolates, during 2016–2022, serotype 4 IPD was only found in adults 19 years and older (634/636 isolates) and serotype 4 IPD affecting persons experiencing homelessness was completely restricted to adults. Of 1440 total cases recorded from persons experiencing homelessness during 2016–2022 (including 200 IPD cases for which isolates were not available), only 5 were from individuals younger than 18 years (1 lacking isolate, 2 serotype 22F, 1 serotype 12F, and 1 serotype 10A).

During 2016–2022, serotype 4 was the most frequent cause of IPD in AEH compared to any other serotype (Figure 1 and Table 1). During 2021–2022, the emergent serotype 4/ST10172 was the most common serotype 4 lineage (Figure 2). Serotype 12F ranked second as a cause of IPD among AEH (Figure 1), primarily due to the closely related ST220 and ST218

lineages, as seen during 2018–2019 [2, 3]. Also, the highest percentages of IPD cases among AEH relative to all adults of known residence status were caused by serotypes 12F and 4 during the entire period of 2016–2022 (237/631 [37.6%] and 155/614 [25.2%] for serotypes 4 and 12F, respectively, compared to 851/14771 [5.8%] for all remaining serotypes; $P < .001$ for the individual serotype 4 and serotype 12F proportions compared with the proportion for all remaining serotypes combined). Although IPD incidence in adults was lower during 2020–2022 compared to 2019 [1], reported serotype 4 IPD cases increased among all adults (both AEH and ANH) during 2020–2022, and more than doubled during 2022 relative to 2019 (Figure 1 and Table 1). Of all adult cases assigned serotypes in ABCs, only serotype 4 exhibited overall increases during each of the years 2020–2022 relative to 2019 (Table 1 and Figure 1).

As shown in Table 2, serotype 4 IPD was largely confined to western ABCs sites. Using ABCs, US census, and HUD data to calculate adult incidence for all IPD and serotype 4 IPD within the 4 western states, increases in IPD incidence due to all serotypes among AEH in 2022 relative to 2019 were observed in New Mexico and Oregon, with a statistically significant increase seen in New Mexico ($P = .014$) and New Mexico ($P = .002$) (Table 2). Within New Mexico and Oregon serotype 4 IPD significantly increased among both AEH and ANH. Remarkably, there was no reported serotype 4 IPD in Oregon during 2019, but there were 42 reported cases during 2022.

Serotype-Switch Strain Progeny Arising From 4/ST10172 and 12F Parental Strains

Among isolates recovered through ABCs between October 2022 and August 2023 and received by the CDC *Streptococcus* Laboratory for analysis between February 2023 and June 2024 there were 30 serotype 4 blood isolates recovered from adults aged 27–71 years, 16 of whom were AEH (22 men and 7 women; unknown residence status for 1 New Mexico case subsequently determined to reside outside of ABCs areas) that shared the new multilocus sequence type ST18704, a single-locus variant of ST220 independently recorded on 21 March 2023 from a serotype 4 isolate recovered from an adult with IPD in the Southwestern United States [6]. These isolates and the CDC serotype 4 reference strain were strongly serotype 4 and immunodiffusion tests indicated antigenically identical capsules (data not shown). The ABCs serotype 4/ST18704 cases occurred in the Oregon (25 isolates), Tennessee (2 isolates), Minnesota (1 isolate), and California (1 isolate) ABCs sites. Twenty-three of these bacteremic cases were diagnosed as pneumonia and 2 as meningitis (1 fatal).

Phylogenetic analysis of available serotype 4 isolates recovered during 2021–2022 (Figure 2) and ST18704 isolates received by the CDC *Streptococcus* Laboratory between 2022 and June 2023 (Figure 3) revealed that the serotype 4/ST18704 isolates were highly related, with pairwise SNP distance of 0–21 and an average of 6 SNPs (Supplementary Table 2). Twenty-three of the 29 patients (79.3%; 13 AEH and 10 ANH), including 20 of the 25 (80%) cases in Oregon, were persons with substance use disorder involving use of 1 or more of (in order of incidence) methamphetamine, opioids, alcohol, and cocaine (Figure 3). In contrast, 33 of 187 (17.6%) 2022–2023 IPD cases in Oregon caused by nonserotype 4 strains were persons with substance abuse disorder involving, in order of incidence,

methamphetamine, opioids or alcohol, and cocaine ($P < .001$). Only 2 of the 29 serotype 4/ST18704 ABCs case patients had reported history of pneumococcal vaccination and 21 of these individuals smoked tobacco. Although defined epidemiologic links were not identified, all 25 of the Oregon cases were listed from the same county (22 isolates) or the adjacent county (3 isolates). Both cases from Tennessee were from the same county and zip code. There was also extensive overlap between identified hospitals where patients were admitted (data not shown).

Pneumococcal serotype-switch progeny can have multiple genomically unlinked recombinant regions originating from the genetic donor of the *cps* operon [23]. Strain 4/ST18704 differs from 12F/ST220 strains from CDC bioinformatic pipeline data only in the *cps* serotype-specific markers (*wzy4* vs *wci12F*) and the multilocus sequence type *aroE* target (*aroE1* vs *aroE10*). In addition to serotype 4/ST10172 prevalence in western ABCs sites [2–4] (Figure 2), the *aroE1* marker made this strain the top serotype 4 candidate strain tracked in ABCs as the genetic donor for the putative 12F/ST220 to 4/ST18704 serotype-switch event. Seven recombinant regions were mapped in the representative 4/ST18704 genome that ranged from 3096 to 70 341 bp in size (Figure 4). Six widely separated recombinant regions in the progeny shared sequence identity or near identity with the strain 4/ST10172 reference genome, including the *cps4*-containing R2 and the *aroE1*-containing R4 fragments. One predicted 4652-bp recombinant fragment (fragment R5) differed extensively from both recipient and donor lineage strains. The number of genes containing 1 or more missense mutations relative to the recipient and donor strains is provided for each fragment, as well as providing representative gene identifiers, in Figure 4. A total of 174 ORFs, not counting the 7 ORFs within fragment R5 and 5 nonconserved *cps4* genes within fragment R2 (see GenBank *cps4* accession [CR931635](#)), were included within the 6 recombinant fragments from the putative serotype 4/ST10172 donor strain, 124 of which contained 1 or more missense mutations compared to the recipient 12F/ST220 reference single-contig genome. In contrast, only 2 of the 174 ORFs contained single missense substitutions compared to the reference 4/ST10172 donor sequence. The regions replaced primarily revealed the same genes/orientations shared between the donor/progeny and recipient sequences (Supplementary Figures 1 and 2). There was a precise replacement of the 17 *cps12F* structural genes with the 15 genes of the *cps4* operon (Figure 5), predicted to place the *cps4* operon under transcriptional control of the *cps12F* promoter and associated recipient lineage transcriptional regulators [24] in the progeny. None of these transcriptional regulators, with the exception of *malR*, mapped within recombinant fragments; the predicted MalR products of the recipient, donor, and progeny (fragment R6 in Figure 4 and Supplementary Figure 2) share amino acid sequence identity (data not shown).

Serotype 4/ST10172 expanded in number during 2021–2022 and first appeared in Oregon during 2022 (Figure 2). This coincided with the appearance of 25 serotype 4/ST18704 cases in Oregon that were recovered during 2022–2023 (Figure 2 and Figure 3). This coincidence conceivably supports a causal effect related to the density of the donor strain 4/ST10172 within Oregon during this period. Phylogenetic analyses of representative serotype 4/ST18704 core genomes together with 76 available core genomic sequences of 12F/ST220 recovered during 2015–2022 (Supplementary Figure 3) did not support this notion, in that it revealed a large genomic distance between 4/ST18704 and its closest 12F/ST220 neighbor

(difference of 761 SNPs). This SNP difference was reduced by subtracting SNPs shared between progeny and serotype 4/ST10172 donor strain within the 6 recombinant fragments that were not shared with the closest matching serotype 12F/ST220 recipient lineage strains (data not shown). The SNP distance of 308 is still indicative of a large temporal difference between recipient and progeny according to calculations of the generation of about 4 core SNPs per year [25]. Nonetheless, differences of >500 SNPs evident between individual 12F/ST220 strains (Supplementary Figure 3) suggest that the most related recipient strains have potentially not been recovered through ABCs.

DISCUSSION

Many pneumococcal strains appear to transmit in the population through a pediatric carriage reservoir, with their prevalence in IPD reflected in their prevalence in carriage [26, 27]. Because carriage serotypes 4 and 12F have been undetectable in colonization surveys carried out in both adults and children in the United States during the conjugate vaccine era [27–30], and are currently extremely rare causes of IPD in children [2], we propose that strains of these 2 serotypes are primarily rapidly transmitted directly between adults and cause IPD shortly after acquisition in carriage. This view is compatible with recent analyses indicating that invasive disease risk is highest upon first acquisition of a strain within the respiratory tract [31, 32]. Certain serotypes, such as 1, 4, and 5, have low carriage duration and quickly transition from carriage to cause respiratory disease within susceptible populations [32]. Such serotypes are most likely rapidly transmitted between adults when they are experiencing respiratory symptoms. Serotype 4 IPD occurs through apparent rapid spread of genetically unrelated or distantly related strains (ST10172, ST244, ST695, and now ST18704), particularly among AEH and other disadvantaged adults. Because these strains are genetically unrelated, it appears that the key factors facilitating this spread include the type 4 capsular polysaccharide and incompletely defined environmental factors within western ABCs sites as shown in this study and others [2–5, 33, 34].

Short carriage duration quickly transitioning to respiratory disease onset between transmission events is consistent with IPD clusters composed of identical or near-identical isolates. This view is also compatible with our frequent observations in previous studies of genomically identical or near-identical serotype 12F and 4 isolates recovered during short time frames from individuals within the same counties [2] that also share the same zip codes and hospital identifiers (data not shown) as we observed in the serotype 4/ST18704 genomic cluster described in this report. In addition, individual tight genomic clusters often occur among AEH [2–4] (as also described here for the serotype 4/ST18704 cluster), providing circumstantial support for possible epidemiologic links in shelters or encampments.

The convergence of IPD cases within Oregon during 2022 caused by both 4/ST10172 donor lineage and 4/ST18704 is suggestive of the 4/ST10172 and 12F/ST220 parental strain “substrates” concurrently circulating among disadvantaged adults within western ABCs sites. The large genomic distances between detected 12F/ST220 recipient strains and 4/ST18704 progeny seemingly refutes recent genesis of the progeny; however, at least some of this distance is artifactual because this distance is reduced by approximately 60% when SNPs caused by multifragment recombination are removed. In addition, while

Gubbins is effective in identifying recombinant regions of high SNP density, it will not detect recombinant regions of low SNP density, which are probably present within serotype 4/ST18704 progeny. Perhaps an even more impactful likely confounder of the apparent temporal distance between recipient and progeny is that the actual 12F/ST220 recipient strain could be much more related to the progeny than the closest recipient detected within ABCs.

The incidence data among AEH within western ABCs sites that were reported in this article are unlikely to be quantitatively accurate due to the subjectivity of both numerators (patient medical charts) and denominators (point in time snapshots of the number of individuals in shelters, temporary housing, and in unsheltered settings) [22]. Nonetheless, we have demonstrated that serotype 4 IPD continues to expand in western ABCs sites during the post-Covid pandemic, and especially among AEH and individuals with substance abuse problems [2, 3]. Invasive serotype 4/ST18704 isolates have now been observed in Oregon, New Mexico, California, Tennessee, and Minnesota. The work presented here highlights the potentially recent genesis of a new strain, which might continue to expand and disproportionally impact vulnerable individuals. It is striking that the single ST18704 ABCs isolates recovered in California and Minnesota were both from AEH. This study indicates that continued transmission of serotype 4 among vulnerable populations might contribute to the generation of additional invasive hybrid strains. Such recombinant strains are unknown entities in their capabilities to cause significant disease burdens over the long term. For example, *cps* promoter regions that bind to different transcriptional activators vary markedly between different serotypes resulting in differences in capsule expression, and this variation is likely to contribute to different levels of virulence between strains [24]. In strain 4/ST18704, the placement of the *cps4* operon under the transcriptional control of the *cps12F* promoter (and corresponding ST220 clone transcriptional regulators) is an unknown factor that might contribute to the virulence and fitness of this strain. Another unknown factor (among many) is the replacement of the *pbp1b* gene with an allele encoding 4 amino acid changes relative to the 12F/ST220 recipient (fragment R7). A single substitution within the pneumococcal Pbp1b transglycosylase domain was associated with meningitis and experimentally found to confer tolerance to antimicrobials [35]. Additional genes associated with virulence that map within the 4/ST18704 recombinant regions include the *ccs4* competence-induced gene (fragment R1), the branched chain amino acid transporter gene *livJ* (fragment R3), *psr* family and associated glycosyl transferase genes involved in anchoring wall polymers to wall peptidoglycan (fragment R4), and the choline-binding protein gene *pcpA* (fragment R7) [36–39].

The remarkably high proportion of serotype 4 IPD cases associated with AEH is suggestive of shared immunological susceptibility within a subset of adults. A conjugate vaccine scheduled for use in adults in June of 2024 (V116) is designed to prevent IPD and targets pneumococcal serotypes of past and current impact [40]. V116 does not contain the serotype 4 polysaccharide, and overall serotype 4 IPD incidence is relatively low. However, within western ABCs sites, serotype 4 IPD has had considerable and increasing impact in adults over the past decade, especially in AEH. More encompassing pneumococcal disease surveillance within US areas could provide more insights into serotype 4 incidence in adult populations, such as the recent resurgences of serotype 4 IPD in Canada [41] and in 2022–

2023 in the Navajo Nation who live in parts of Arizona, New Mexico, and Utah [42]. Continued and expanded monitoring of disease-caused by serotype 4 pneumococcal strains is needed to better understand serotype 4 disease impact and specific demographic factors that contribute to IPD disparities among different populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

We are very grateful to everyone involved in Active Bacterial Core surveillance: <https://www.cdc.gov/abcs/index.html>. We thank Dr Adam L. Cohen for insightful questions, suggestions, and criticisms, and Robert E. Gertz, Jr for serologic testing of the isolates.

Financial support.

This work was funded by the Centers for Disease Control and Prevention as part of the authors' normal responsibilities.

References

1. ABCs Bactfacts Interactive Data Dashboard. Active bacterial core surveillance reports for 1997–2021. <https://www.cdc.gov/abcs/bact-facts/data-dashboard.html>. Accessed 13 July 2024.
2. Metcalf BJ, Chochua S, Walker H, et al. Invasive pneumococcal strain distributions and isolate clusters associated with persons experiencing homelessness during 2018. *Clin Infect Dis* 2021; 72:e948–56. [PubMed: 33150366]
3. Beall B, Chochua S, Li Z, et al. Invasive pneumococcal disease clusters disproportionately impact persons experiencing homelessness, injecting drug users, and the western United States. *J Infect Dis* 2022; 226:332–41. [PubMed: 35172327]
4. Beall B, Walker H, Tran T, et al. Upsurge of conjugate vaccine serotype 4 invasive pneumococcal disease clusters among adults experiencing homelessness in California, Colorado, and New Mexico. *J Infect Dis* 2021; 223:1241–9. [PubMed: 32798216]
5. Steinberg J, Bressler SS, Orell L, et al. Invasive pneumococcal disease and potential impact of pneumococcal conjugate vaccines among adults, including persons experiencing homelessness—Alaska, 2011–2020. *Clin Infect Dis* 2024; 78: 172–8. [PubMed: 37787072]
6. Public databases for molecular typing and microbial genome diversity. <https://pubmlst.org/organisms/streptococcus-pneumoniae>. Accessed 1 February 2023.
7. Metcalf BJ, Gertz RE Jr, Gladstone RA, et al. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. *Clin Microbiol Infect* 2016; 22:60.e9–29.
8. Chochua S, Beall B, Lin W, et al. The emergent invasive serotype 4 ST10172 strain acquires vanG type vancomycin-resistance element: a case of a 66-year-old with bacteremic pneumococcal pneumonia [published online ahead of print 8 August 2024]. *J Infect Dis* doi: 10.1093/infdis/jiae393.
9. Metcalf BJ. CDC streptococcal bioinformatics pipelines. <https://github.com/BenJamesMetcalf>. Accessed 1 December 2023.
10. Gertz RE Jr, Pimenta FC, Chochua S, et al. Nonpneumococcal strains recently recovered from carriage specimens and expressing capsular serotypes highly related or identical to pneumococcal serotypes 2, 4, 9A, 13, and 23A. *mBio* 2021; 12:e01037–21.
11. Kolmogorov M, Yuan J, Lin Y, Pevzner P. Assembly of long error-prone reads using repeat graphs. *Nat Biotech* 2019; 32:1009.
12. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. Assembly of long error-prone reads using de Bruijn graphs. *Proc Nat Acad Sci* 2016; 113: E8396–8405. [PubMed: 27956617]

13. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2008; 10:421.
14. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; 30:2068–9. [PubMed: 24642063]
15. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011; 27:1009–10. [PubMed: 21278367]
16. Gardner SN, Slezak T, Hall BG. kSNP3.0: sNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 2015; 31: 2877–8. [PubMed: 25913206]
17. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; 33:1870–4. [PubMed: 27004904]
18. Darling AE, Mau B, Perna NT. Progressive mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010; 5:e11147.
19. Croucher NJ, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using gubbins. *Nucleic Acids Res* 2015; 43:e15. [PubMed: 25414349]
20. Grant JR, Enns E, Marinier E, et al. Proksee: in-depth characterization and visualization of bacterial genomes nucleic acids research. *Nucleic Acids Res* 2023; 51:W484–92. [PubMed: 37140037]
21. United States Census Bureau. Population and housing unit estimates. <https://www.census.gov/programs-surveys/popest.html>. Accessed 15 July 2023.
22. US Department of Housing and Urban Development. <https://www.hudexchange.info/programs/coc/coc-homeless-populations-and-subpopulations-report>. Accessed 15 July 2023.
23. Golubchik T, Brueggemann AB, Street T, et al. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat Genet* 2012; 44:352–55. [PubMed: 22286217]
24. Glanville DG, Gazioglu O, Marra M, et al. Pneumococcal capsule expression is controlled through a conserved, distal cis-regulatory element during infection. *PLoS Pathog* 2023; 19:e1011035.
25. Croucher NJ, Chewapreecha C, Hanage WP, et al. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. *Genome Biol Evol* 2014; 6:1589–602. [PubMed: 24916661]
26. Kellner JD, McGeer A, Cetron MS, et al. The use of *Streptococcus pneumoniae* nasopharyngeal isolates from healthy children to predict features of invasive disease. *Pediatr Infect Dis J* 1998; 17:279–86. [PubMed: 9576381]
27. Huang SS, Hinrichsen VL, Stevenson AE, et al. Continued impact of pneumococcal conjugate vaccine on carriage in young children. *Pediatrics* 2009; 124:e1–11. [PubMed: 19564254]
28. Sharma D, Baughman W, Holst A, et al. Pneumococcal carriage and invasive disease in children before introduction of the 13-valent conjugate vaccine: comparison with the era before 7-valent conjugate vaccine. *Pediatr Infect Dis J* 2013; 32: e45–53. [PubMed: 23080290]
29. Desai AP, Sharma D, Crispell EK, et al. Decline in pneumococcal nasopharyngeal carriage of vaccine serotypes after the introduction of the 13-valent pneumococcal conjugate vaccine in children in Atlanta, Georgia. *Pediatr Infect Dis J* 2015; 34:1168–74. [PubMed: 26226445]
30. Milucky J, Carvalho MG, Roupheal N, et al. *Streptococcus pneumoniae* colonization after introduction of 13-valent pneumococcal conjugate vaccine for US adults 65 years of age and older, 2015–2016. *Vaccine* 2019; 37:1094–100. [PubMed: 30685247]
31. Dominguez-Huttinger E, Boon NJ, Clarke TB, Tanaka RJ. Mathematical modeling of *Streptococcus pneumoniae* colonization, invasive infection and treatment. *Front Physiol* 2017; 8:115. [PubMed: 28303104]
32. Metcalf BJ, Waldetoft KW, Beall BW, Brown SP. Variation in pneumococcal invasiveness metrics is driven by serotype carriage duration and initial risk of disease. *Epidemics* 2023; 45:100731.
33. Beall B, Chochua S, Gertz RE Jr, et al. A population-based descriptive atlas of invasive pneumococcal strains recovered within the U.S. During 2015–2016. *Front Microbiol* 2018; 9:2670. [PubMed: 30524382]

34. Varghese J, Chochua S, Tran T, et al. Multistate population and whole genome sequence-based strain surveillance of invasive pneumococci recovered in the USA during 2017. *Clin Microbiol Infect* 2020; 26:512.e1–10.
35. Li Y, Metcalf BJ, Chochua S, et al. Genome-wide association analyses of invasive pneumococcal isolates identify a missense bacterial mutation associated with meningitis. *Nat Commun* 2019; 10:178. [PubMed: 30643125]
36. Hirose Y, Yamaguchi M, Goto K, Sumitomo T, Nakata M, Kawabata S. Competence-induced protein ccs4 facilitates pneumococcal invasion into brain tissue and virulence in meningitis. *Virulence* 2018; 9:1576–87. [PubMed: 30251911]
37. Molzen TE, Burghout P, Bootsma HJ, et al. Genome-wide identification of *Streptococcus pneumoniae* genes essential for bacterial replication during experimental meningitis. *Infect Immun* 2011; 79:288–97. [PubMed: 21041497]
38. Ye W, Zhang J, Shu Z, Yin Y, Zhang X, Wu K. Pneumococcal LytR protein is required for the surface attachment of both capsular polysaccharide and teichoic acids: essential for pneumococcal virulence. *Front Microbiol* 2018; 9:1199. [PubMed: 29951042]
39. Walker MM, Novak L, Widener R, et al. PcpA promotes higher levels of infection and modulates recruitment of myeloid-derived suppressor cells during pneumococcal pneumonia. *J Immunol* 2016; 196:2239–48. [PubMed: 26829988]
40. Platt HL, Bruno C, Buntinx E, et al. Safety, tolerability, and immunogenicity of an adult pneumococcal conjugate vaccine, V116 (STRIDE-3): a randomised, double-blind, active comparator controlled, international phase 3 trial [published online ahead of print 1 July 2024]. *Lancet Infect Dis* 2024 doi: S1473–3099(24)00344-X.
41. Kellner JD, Ricketson LJ, Demczuk WHB, et al. Whole-genome analysis of *Streptococcus pneumoniae* serotype 4 causing outbreak of invasive pneumococcal disease, Alberta, Canada. *Emerg Infect Dis* 2021; 27: 1867–75. [PubMed: 34152965]
42. Navajo Epidemiology Center. Serotype 4 invasive pneumococcal disease (IPD) information for providers. <https://nec.navajo-nsn.gov/>. Accessed 22 July 2024.

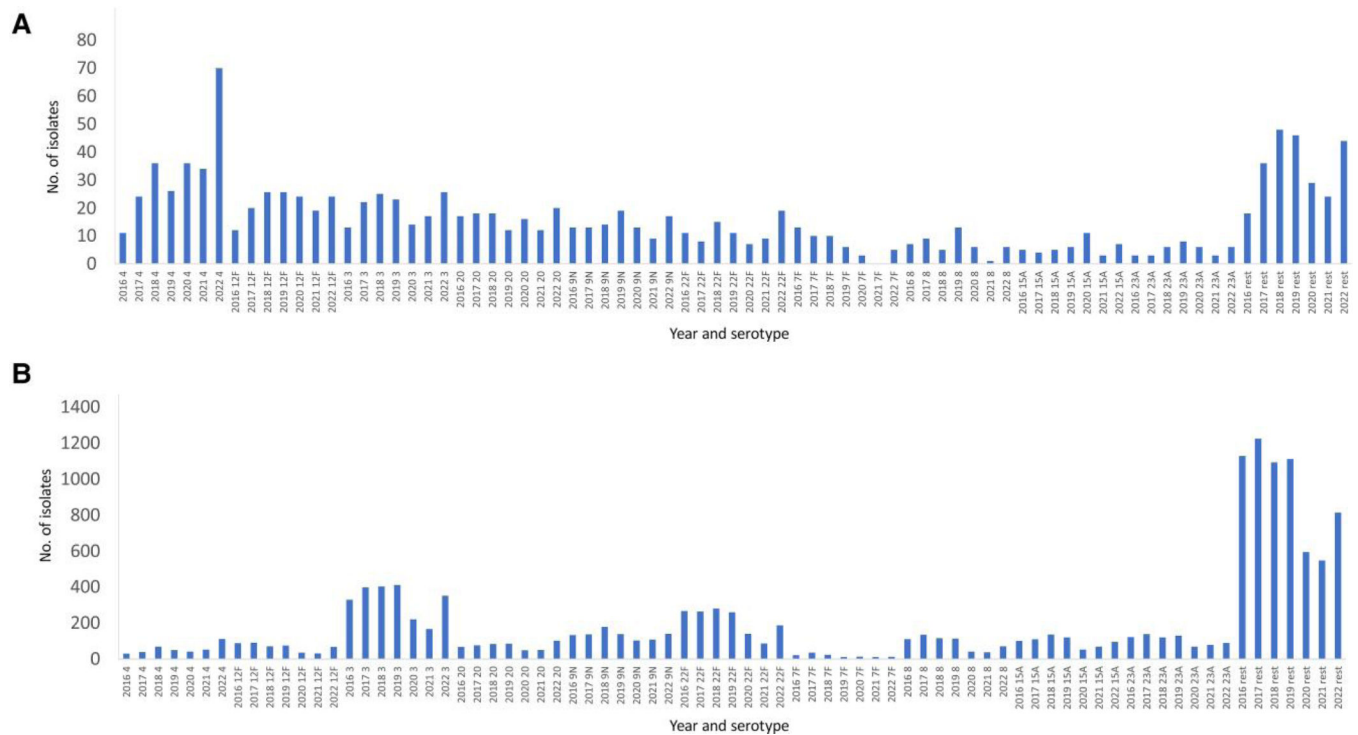


Figure 1.

The 10 most frequently occurring serotypes and remaining serotypes among invasive pneumococcal disease isolates recovered from adults experiencing homelessness (*A*) and the recovery of these same serotypes from adults not experiencing homelessness (*B*) during 2016–2022. Year and serotype are given on the x-axis.

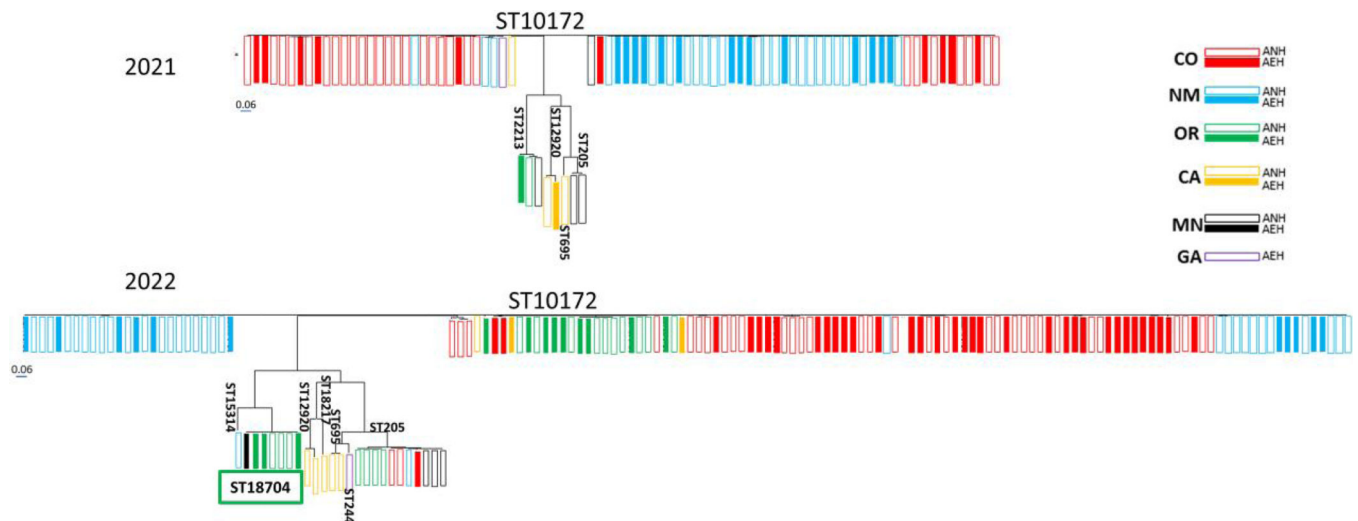
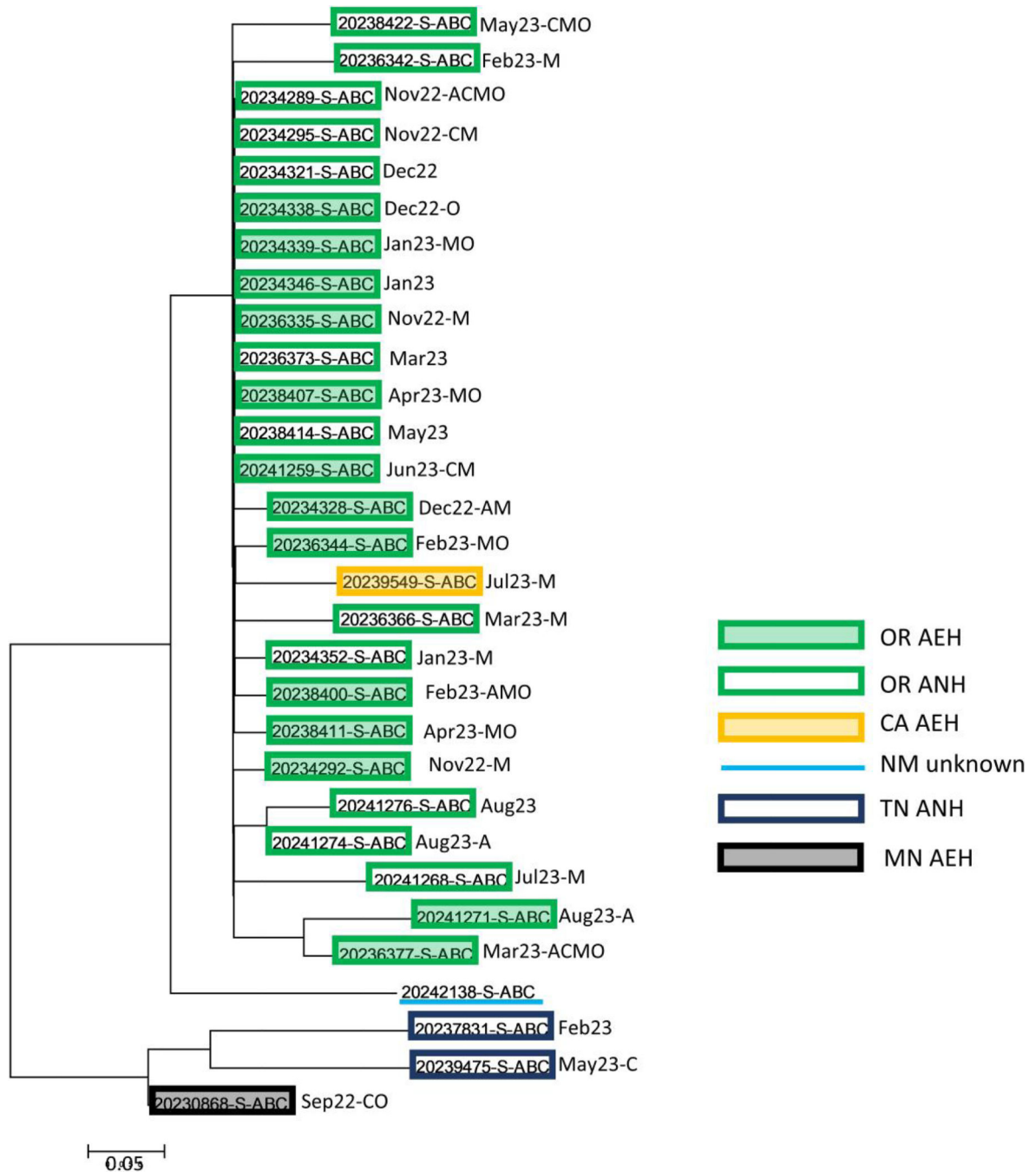


Figure 2.

Core genome phylogeny of available serotype 4 invasive pneumococcal disease isolates occurring during 2021 and 2022, with colors indicators of ABCs site and residence status. For the year 2021 tree, the analysis involved all 86 genomes and 14 154 nucleotide positions. For the year 2022 tree, the analysis involved 156 of 168 serotype 4 isolate genomes and 15 355 nucleotide positions. Distance scales and multilocus sequence type references for the different branches for the tree lengths are provided for both trees. Abbreviations: ABCs, Active Bacterial Core surveillance; AEH, adult experiencing homelessness; ANH, adult not experiencing homelessness; CA, California; CO, Colorado; GA, Georgia; MN, Minnesota; NM, New Mexico; OR, Oregon.

**Figure 3.**

Core genome phylogeny of 30 recombinant serotype 4/ST18704 isolates. ABCs site and residence status are indicated by color and the distance scale is provided. There were 60 variable positions in the final dataset. Abbreviations: ABCs, Active Bacterial Core surveillance; CA, California; MN, Minnesota; NM, New Mexico; OR, Oregon; TN, Tennessee; Substance abuse disorder abbreviations shown in varying combinations: methamphetamine, M; opioids, O; alcohol (A); cocaine (C).

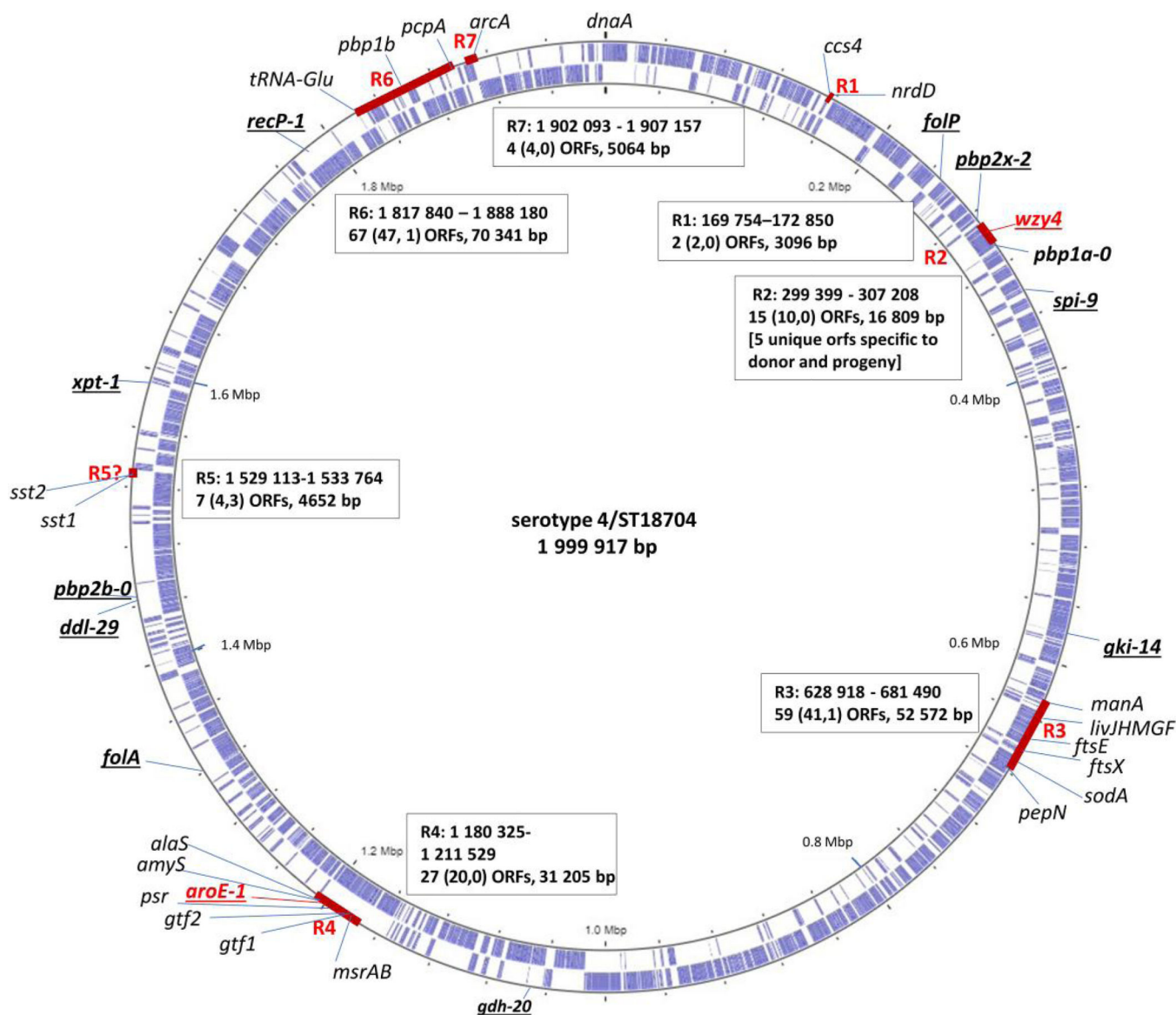


Figure 4.

Serotype 4/ST18704 genome coordinates. Base 1 lies at the origin of replication region, closely upstream of the *dnaA* gene, with forward and reverse open reading frames (ORFs) indicated by bands on outside and inside, respectively. Recombinant regions R1 through R7 are shown that share identity or near identity (>99.9%) with the serotype 4/ST10172 genetic donor. In boxes are included the genome coordinates of the recombinant fragment and number of ORFs. In parentheses are the number of these ORFs with substitutions relative to the recipient 12F/ST220 reference strain separated by a comma from the number of these ORFs with substitutions relative to the donor serotype 4/ST10172 reference strain. The size of each recombinant fragment is provided. The genes flanking the end of each recombinant fragment are listed, as well as 1 or more internal genes. Underlined genes in bold (*aroE-1* and *wzy4* included within recombinant fragments R2 and R4, respectively) represent bioinformatic pipeline output for serotype (*wzy4*), PBP gene types (*pbp1a-0*,

pbp2b-0, and *pbp2x-2*), cotrimoxazole-phenotype (*folA* and *folP*), and multilocus sequence type (*aroE-1*, *gdh-20*, *gki-14*, *recP-1*, *spi-9*, *xpt-1*, and *ddl-29*).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Author Manuscript

Table 1.

Comparisons of Invasive Pneumococcal Disease Proportions in AEH During Different Time Periods With Previous Time Periods: Prepandemic 2018–2019 vs 2016–2017, Pandemic 2020–2021 vs 2018–2019, and Postpandemic 2022 vs 2018–2019

Serotype	Year	No. Reported Cases	AEH/(AEH + ANH), % ^a	Comparison of Time Periods, AEH/(AEH + ANH), % ^b			
				2016–2017 vs 2018–2019	2018–2019 vs 2020–2021	2018–2019 vs 2022	P Value
4	2016	11/41, 26.8%					
	2017	24/63, 38.0%					
	2018	36/105, 34.2%					
	2019	26/76, 34.2%					
	2020	36/77, 46.7%					
	2021	34/87, 40.5%					
	2022	70/182, 38.4%					
	2016	12/100, 12.0%		35/104, 33.7%	62/181, 34.2%	62/181, 34.2%	
	2017	20/105, 19.0%					
	2018	28/96, 29.2%					
12F	2019	28/101, 27.7%					
	2020	24/60, 40.0%					
	2021	19/50, 38.0%					
	2022	24/67, 35.8%		32/205, 15.6%	56/197, 28.4%	56/197, 28.4%	
	2016	13/342, 3.8%		56/197, 28.4%	43/110, 39.1%	24/67, 35.8%	<i>P</i> = .325
	2017	22/400, 5.5%					
	2018	25/402, 6.2%					
	2019	23/412, 5.5%					
	2020	14/228 (6.1)					
	2021	17/178 (9.6)					
20	2022	31/353 (8.8)		35/742, 4.7%	48/814, 5.9%	48/814, 5.9%	
	2016	17/85 (20.0)		48/814, 5.9%	31/406, 7.6%	31/353, 8.8%	<i>P</i> = .098
	2017	18/91 (19.8)					
	2018	18/101 (17.8)					
	2019	12/96 (12.5)					

Serotype	Year	No. Reported Cases	Comparison of Time Periods, AEH/(AEH + ANH), % ^b					
			AEH/(AEH + ANH), % ^a		2016–2017 vs 2018–2019		2018–2019 vs 2020–2021	
					P Value		P Value	
9N	2020	16/65 (24.6)						
	2021	12/64 (18.8)					30/197, 15.2%	30/197, 15.2%
	2022	20/120 (16.7)					28/129, 21.7%	20/120, 16.7%
	2016	13/146 (8.9)						<i>P</i> = .190
	2017	13/146 (8.9)						<i>P</i> = .849
22F	2018	14/191 (7.3)						
	2019	19/156 (12.2)						
	2020	13/114 (11.4)						
	2021	9/114 (7.9)					33/347, 9.5%	33/347, 9.5%
	2022	17/155 (11.0)					22/228, 9.6%	17/155, 11.0%
4 + 12F + 3+20 + 9N + 22F	2016	11/277 (4.1)						<i>P</i> > .999
	2017	8/251 (3.2)						
	2018	15/273 (5.5)						
	2019	11/253 (4.3)						
	2020	7/141 (5.0)						
Remaining serotypes minus 4 + 12F + 3 + 20 + 9N + 22F	2021	9/90 (10.0)					26/526, 4.9%	26/526, 4.9%
	2022	19/184 (10.3)					16/231, 6.9%	19/184, 10.3%
	2016	77/1097 (7.0)						<i>P</i> = .353
	2017	97/1042 (9.3)						
	2018	136/1167 (11.7)						
	2019	119/1094 (10.9)						
	2020	110/685 (16.0)						
	2021	100/683 (14.6)					255/2261, 11.3%	255/2261, 11.3%
	2022	181/1057 (17.1)					210/1368 17.9%	181/1057, 17.1%
	2016	46/1403 (2.6)						<i>P</i> ≤ .001
	2017	61/1664 (3.7)						<i>P</i> ≤ .001
	2018	74/1525 (4.9)						
	2019	79/1534 (5.1)						
	2020	55/801 (6.9)						

Serotype	Year	No. Reported Cases	Comparison of Time Periods, AEH/(AEH + ANH), % ^b					
			2016–2017 vs 2018–2019		2018–2019 vs 2020–2021		2018–2019 vs 2022	
			AEH/(AEH + ANH), % ^a	P Value	AEH/(AEH + ANH), % ^a	P Value	AEH/(AEH + ANH), % ^a	P Value
All serotypes combined	2021	31/645	(4.8)		153/3059, 5.0%		153/3059, 5.0%	
	2022	67/1141	(5.9)		86/1446, 5.9%		67/1141, 5.9%	
	2016	123/2500	(4.9)					
	2017	165/2706	(6.1)					
	2018	207/2692	(7.7)					
	2019	198/2628	(7.5)					
	2020	164/1486	(11.0)					
	2021	132/1328	(11.0)		405/5317, 7.6%		405/5317, 7.6%	
	2022	247/2198	(11.2)		296/2814, 10.5%		247/2198, 11.2%	
				P ≤ .001		P ≤ .001		P ≤ .001

P values provided from Fischer exact test, 2-tailed.

Abbreviations: AEH, adults experiencing homelessness; ANH, adults not experiencing homelessness.

^a Percentages for 2020–2021 and 2022 that were greater than the percentages for each of the years 2016–2019 indicated in bold.

^b Greater of the 2 percentages and statistically significant P values ($P < .05$) indicated in bold.

Table 2.

Estimated Invasive Pneumococcal Disease Incidence in Adults Experiencing and Not Experiencing Homelessness in 4 Western ABCs Sites^a

Cases	California			Colorado			New Mexico			Oregon		
	AEH	ANH		AEH	ANH		AEH	ANH		AEH	ANH	
ABCs sites 2019												
Population	17327	2 872	196	4897 ^b	1 539	287	2726	1 666	493	4653	1 483	050
All cases (cases with serotype)	47 (33)	229	(196)	76 (65)	212	(176)	37 (33)	291	(255)	16 (16)	124	(119)
All cases/100 000	271	8.0		1552	13.8		1357	17.5		349	8.4	
Estimated serotype 4 cases (cases with serotype)	2.6 (2)	15.3	(13)	22.2 (19)	20.5	(17)	5.6 (5)	18.0	(16)	0(0)	0(0)	
Estimated serotype 4 cases/100 000	15	0.53		453	1.3		206	1.1		0	0	
ABCs sites 2022												
Population	19482	2 909	213	7243	1705	439	2076	1 664	010	5905	1 462	010
All cases (cases with serotypes)	29 (22)	127	(103)	110(101)	245	(201)	55 (39)	330	(272)	32 (28)	126	(104)
All cases /100 000	149 (<i>P</i> = .01373)	4.4 (<i>P</i> < .001)		1518	14.4		2649 (<i>P</i> = .002)	19.8		542	8.6	
Estimated serotype 4 cases (cases with serotype)	5.3 (4)	8.6	(7)	41.1 (34)	37.8	(31)	24.0 (17)	63.1	(52)	19.6 (14)	19.4	(16)
Estimated serotype 4 cases/100 000	27.2	0.3		567	2.2		1156 (<i>P</i> = .001)	3.8 (<i>P</i> = .001)		331.9 (<i>P</i> = .001)	1.3 (<i>P</i> = .001)	

Abbreviations: ABCs, Active Bacterial Core surveillance; AEH, adults experiencing homelessness; ANH, adults not experiencing homelessness; HUD, Department of Housing and Urban Development; PIT, point in time.

^aAEH population estimates used 2019 and 2022 PIT surveys for adults ≥ 18 years of age from HUD. ANH used adult (≥ 18 years of age) estimates for given ABCs regions for 2019 (1 April 2020) and 2022 (1 July 2023) from the US Census Bureau. Significant changes in incidence during 2022 relative to 2019 are shown in bold.

^bOverestimated population due to HUD PIT survey area consisted of 6 county metropolitan Denver area while ABCs only consists of 4 of these counties plus the low population Elbert county.