Comparative genomic analysis of two ST320 *Streptococcus pneumoniae* isolates, representing serotypes 19A and 19F

Weronika Puzia[1,2], Jan Gawor[2], Robert Gromadka[2], Anna Skoczyńska[1,3] and Ewa Sadowy[4*]

Abstract

Background *Streptococcus pneumoniae* (pneumococcus) represents an important human pathogen, responsible for respiratory and invasive infections in the community. The efficacy of polysaccharide conjugate vaccines formulated against pneumococci is reduced by the phenomenon of serotype replacement in population of this pathogen. The aim of the current study was to obtain and compare complete genomic sequences of two pneumococcal isolates, both belonging to ST320 but differing by the serotype.

Results Here, we report genomic sequences of two isolates of important human pathogen, *S. pneumoniae*. Genomic sequencing resulted in complete sequences of chromosomes of both isolates, 2,069,241 bp and 2,103,144 bp in size, and confirmed the presence of *cps* loci specific for serotypes 19A and 19F. The comparative analysis of these genomes revealed several instances of recombination, which involved not only *S. pneumoniae* but also presumably other streptococci as donors.

Conclusions We report the complete genomic sequences of two *S. pneumoniae* isolates of ST320 and serotypes 19A and 19F. The detailed comparative analysis of these genomes revealed the history of several recombination events, clustered in the region including the *cps* locus.

Keywords *Streptococcus pneumoniae*, Serotype, Serotype switch, Complete genome, Recombination, Antimicrobial resistance, Tn*2010* transposon

Introduction

Streptococcus pneumoniae (pneumococcus) represents one of the leading human bacterial pathogen in community-acquired respiratory and invasive infections [1,

*Correspondence:
Ewa Sadowy
e.sadowy@nil.gov.pl
[1] Department of Epidemiology and Clinical Microbiology, National Medicines Institute, Warsaw, Poland
[2] DNA Sequencing and Synthesis Facility, Institute of Biochemistry and Biophysics PAS, Warsaw, Poland
[3] National Reference Centre for Bacterial Meningitis, National Medicines Institute, Warsaw, Poland
[4] Department of Molecular Microbiology, National Medicines Institute, Ul. Chelmska 30/34, 00-725, Warsaw, Poland

2]. Polysaccharide capsule, specifying the serotype constitutes a key virulence factor of pneumococcus [3]. The composition of capsule, determined by the *cps* locus shows a remarkable diversity in the population of this pathogen [4, 5] and represents a pivotal component of anti-pneumococcal non-conjugated and conjugated vaccines [6]. Introduction of the 7-valent polysaccharide conjugate vaccine (PCV7, against serotypes 4, 6B, 9 V, 14, 18C, 19F and 23F) into a mass vaccination of children resulted in not only decreased an incidence of invasive infections in this group but also contributed to a reduction of resistance levels to antimicrobials important for a therapy [7], due to the fact that limited number of pneumococcal serotypes was then in a significant part responsible for the appearance of multi-drug resistance (MDR)

Puzia *et al. BMC Genomic Data*   (2023) 24:19

Page 2 of 10

among pneumococci [8, 9]. Multilocus sequence typing (MLST) introduced into analyses of pneumococcal populations [10] played an important role in determining that circulation of certain pneumococcal epidemic clones greatly contributed to increasing the levels of MDR among pneumococci [11] before the PCV7 era. Such clones were identified by the Pneumococcal Molecular Epidemiology Network (PMEN) and named following their country of first isolation and main serotype associated with a given clone [12]. Introduction of the whole-genome sequencing (WGS) into microbiology opened entirely new possibilities for analyses of pneumococcal clones and their evolution [13].

Post-vaccine surveillance revealed a quick adaptation of *S. pneumoniae* to the selective pressure exerted by PCV7, resulting in serotype replacement [14], caused by an increased circulation of clones associated with non-vaccine serotypes (NVT) and changes of serotypes within established epidemic clones [15–17]. Such change, named a "serotype switch" is known to occur in pneumococcal populations thanks to a natural competence of these bacteria and involves an exchange of the *cps* locus, often with its adjacent sequences [18]. In the US, following the introduction of PCV7 an appearance of "vaccine escape recombinants" expressing non-vaccine serotype 19A instead of serotype 4 targeted by the PCV7 was observed in the clone of sequence type (ST) 695 [19]. These changes in pneumococcal population prompted introduction of two higher-valent vaccines, PCV10 (against PCV7 serotypes plus 1, 5 and 7F) and PCV13 (against PCV10 serotypes plus 3, 6A and 19A), into the market in 2009. These vaccines either replaced PCV7 or were introduced de novo into the national or local vaccination calendars. Serotype replacement by 19A post-PCV7 was observed in the US and several other countries due to both increases of prevalence of strains typically associated with 19A, such as e.g. the clonal complex CC199[19A] as well by spread of clones that had undergone serotype switch such as CC695[19A] mentioned above [16]. Such replacement by 19A, mostly multidrug resistant, has also been noticed in countries where PCV10 has been used [20, 21]. In turn, the use of PVC13 has greatly reduced the number of 19A infections and thus the level of antibiotic resistance [22, 23] although this beneficial effect may be reduced by emergence of resistance in other non-vaccine serotypes, e.g. in Spain [24].

In Poland, the National Reference Centre for Bacterial Meningitis (NRCBM, Warsaw) since 1997 constantly monitors and provides laboratory confirmation for invasive pneumococcal infections in all age groups. The PCV10 was introduced into the vaccination calendar in Poland in 2017 [25] but before all PCVs were available commercially and used on a voluntary basis. A gradual increase in the 19A serotype prevalence was observed in the 2010s [25, 26] and based on our preliminary data, isolates of ST320 were the most prevalent ones among 19A isolates in Poland before the vaccination era [27]. To improve our understanding of the phenomenon of exchange of *cps* and other genes among particular pneumococcal clones, we present the complete genomic sequences of two pneumococcal isolates, representing the same sequence type, ST320 as defined by the MLST scheme but demonstrating two different serotypes, 19A and 19F, obtained in Poland from invasive infections.

Methods

Bacterial isolates

The 3238/09 and 3641/15 isolates, obtained from patients' blood in hospitals in Wejherowo (54.6147 N, 18.2450 E) and Kraków (50.0120 N, 20.0012 E) were received by the NRCBM as a part of routine surveillance activity mandated by the Ministry of Health. The study was conducted as part of continuous surveillance in accordance with the World Health Medical Association 1966 Declaration of Helsinki and the EU rules of Good Clinical Practice, thus ethical approval and informed consent were not required. Upon arrival to the NRCBM isolates were streaked on the Columbia agar with 5% sheep blood (CBA) (Becton, Dickinson and Company, Franklin Lakes, NJ) and incubated for 18–24 hours at 37 °C with 5% CO_2. Re-identification of isolates was performed using classical microbiological methods, such as an evaluation of colony morphology, Gram staining and testing optochin susceptibility and deoxycholate solubility [28]. Isolates were stored in -80 °C in trypticase soy broth (TSB) (Becton, Dickinson and Company, Sparks, MD) with 40% horse serum and 15% glycerol until further analysis.

Serotype determination and antimicrobial susceptibility testing

The quellung reaction with serotype-specific antisera (SSI Diagnostica, Hillerod, Denmark) was used for serotype determination, as previously described [28, 29]. The minimum inhibitory concentrations (MIC) of 18 antibiotics were determined by the broth microdilution method (BMD) "in house" and interpreted as recommended by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) [30]. Susceptibility to erythromycin and clindamycin was additionally verified on the basis of the double disc test, and resistant isolates were assigned to specific phenotypes, including constitutive MLS_B ($cMLS_B$), inducible MLS_B ($iMLS_B$) and efflux-mediated resistance (M-phenotype). Multi drug-resistance (MDR) was defined as resistance to at least one agent from three

Puzia *et al. BMC Genomic Data*   (2023) 24:19

Page 3 of 10

or more antimicrobial groups. The quality control strain was *S. pneumoniae* ATCC 49619.

DNA isolation, WGS, read assembly, MLST and rMLST

For genomic sequencing bacteria were grown on CBA for 11–13 hours and isolation of DNA was performed using the SDS/phenol method [31]. In brief, bacteria were collected from CBA plates by washing and resuspending in 2 ml TE buffer. Six hundred microliters of such suspension were taken for the DNA isolation procedure. DNA concentration was determined using the Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA) and the quality of DNA preparations was evaluated by electrophoresis in 0.8% agarose (Prona Agarose, Burgos, Spain). Short-read sequencing was performed with the MiSeq instrument (Illumina Inc., San Diego, CA), using NEB Ultra II FS kit (New England Biolabs, Beverly, MA) for a library construction and MiSeq v3 600 cycle sequencing kit. Quality control and trimming was performed with the fastp software version 0.20.0 [32]. MLST and ribosomal MLST (rMLST) were performed following the established schemes [10, 33] and using the Internet-accessible databases https://pubmlst.org/organisms/streptococcus-pneumoniae and https://pubmlst.org/species-id, respectively (last accessed 1st June 2022) [34] to identify particular alleles and resulting STs and rSTs.

Long-read sequencing of both isolates was performed using the GridION instrument (Oxford Nanopore Technologies, Oxford, UK). Long read libraries were constructed using SQK-RBK004 kit and sequenced on R9.4.1 flowcell. Raw nanopore reads were basecalled and demultiplexed using guppy 4.2.2. Adaptor removal using porechop [35] and data quality filtering using NanoFilt [36] resulted in 35,962 and 46,547 reads, 265.15 and 560.43 Mbp of sequencing data with the N50 value of 11.8 kb and 28.6 kb, respectively for 3641/15 and 3238/09 isolates. Hybrid assembly of Illumina and Nanopore long reads was performed with the Unicycler version 0.4.6 software [37].

Genomic data analysis

Annotation of assembled genomic sequences was performed with the NCBI PGAP version 5.3 [38]. The Geneious Prime v.2022.1.1 software (Biomatters, Auckland, New Zealand) was used for visualization of genomes and additional analyses. The Average Nucleotide Identity (ANI) was calculated using an online ANI calculator (http://www.ezbiocloud.net/tools/ani; last accessed 27th April 2022) [39]. Ribosomal sequence types (rSTs) were established using the rMLST [33] database (https://pubmlst.org/species-id; last accessed 25th March 2022) [34]. Antimicrobial resistance determinants were detected and localized in the genomic

sequences using the ResFinder 3.0 [40] online service (https://cge.cbs.dtu.dk/services/ResFinder/; last accessed 25th March 2022). The presence of phages was determined by initial searches with PHASTER (https://phaster.ca/; last accessed 11th February 2022) [41] followed by manual analyses. Similarity of nucleotide sequences to these reported by others and potential functions of gene products were investigated by blastn and blastx searches, respectively, in GenBank (https://blast.ncbi.nlm.nih.gov/; last accessed 30th May 2022). Regions of recombination were identified with Gubbins [42] and visualized in Phandango [43]. Artemis Comparison Tool (ACT) [44] was used for sequence alignments and visualization. For all the software the default parameters were used. Gene localization is provided relative to the R6 genome of *S. pneumoniae* [45].

Results and discussion

Bacterial serotypes and antimicrobial susceptibility

The 3238/09 and 3641/15 isolates represented serotypes 19F and 19A, respectively. Of these two serotypes, 19F was included in the past in PCV7 and is present in two conjugate vaccines used currently (PCV10, PCV13) in mass vaccination, which significantly reduced its incidence. In contrast, 19A is targeted only by PCV13, which contributed to curbing 19A infections after their increase associated with the serotype replacement following PCV7/PCV10 use [20–23]. The 3238/09 and 3641/15 isolates were both multidrug resistant (MDR) demonstrating resistance to penicillin (MICs 8 and 4 mg/L, respectively), ampicillin (8 and 16 mg/L, respectively), amoxicillin (8 and 4 mg/L, respectively), cefuroxime (both 64 mg/L), cefotaxime (4 and 8 mg/L, respectively), cefepime (16 and 8 mg/L, respectively), erythromycin (both > 32 mg/L), clindamycin (both > 32 mg/L), tetracycline (8 and 16 mg/L, respectively) and trimethoprim-sulfamethoxazole (16 and 8 mg/L, respectively). They were susceptible to increased exposure to levofloxacin (both 1 mg/L) and moxifloxacin (both 0.12 mg/L), and susceptible to meropenem (1 and 0.5 mg/L, respectively), doxycycline (0.5 and 1 mg/L, respectively), linezolid (both 0.5 mg/L), chloramphenicol (4 and 2 mg/L, respectively), rifampicin (0.0075 and 0.03 mg/L, respectively) and vancomycin (0.5 and 0.25 mg/L, respectively). Except for the rifampicin (two dilutions), the isolates differed by no more than a single dilution in MIC values. The MDR phenotype is frequently observed worldwide for pneumococci representing serotypes 19A and 19F [46, 47]. In Poland, between 2011 and 2013 86.7 and 70.5% of invasive 19A and 19F pneumococci, respectively, were MDR while this phenotype characterized 21.6% invasive *S. pneumoniae* in general [26].

Table 1 Summary of genome data for two ST320 isolates of *S. pneumoniae*

Isolate	Serotype	ST	rST	Genome size (bp)	%GC	Number of CDSs	Number of pseudo-genes	SRA accession number for Illumina reads	SRA accession number for Oxford Nanopore reads	GenBank accession number	Sequencing depth (x) ONT, Illumina
3238/09	19F	320	573	2,103,144	39.8	2089	124	SRR17689170	SRR17689169	CP091451	237x, 74x
3641/15	19A	320	572	2,069,241	39.9	2025	125	SRR17689168	SRR17689167	CP091450	123x, 88x

ST, Sequence type, *rST* Ribosomal sequence type, *CDSs* Coding DNA sequences, *SRA* Sequence Read Archive

Puzia *et al. BMC Genomic Data*   (2023) 24:19

Page 5 of 10

Illumina sequencing, MLST and rMLST

Sequencing performed with Miseq resulted in 276,713 and 305,632 paired-end reads, respectively, corresponding to 142.39 Mbp and 164.73 Mbp of sequence for the 3238/09 and 3641/15 isolates, respectively. Both these isolates belonged to ST320 (Table 1). Among 805 isolates of ST320 reported to the PubMLST database for *S. pneumoniae* (date accessed 11th April 2022), the complete serotype was available for 789 isolates, among which serotype 19A and 19F were characteristic for 685 and 101 isolates, respectively. Both serotypes showed a global distribution and were observed in a similar time span (1998–2019 for 19A and 2000–2020 for 19F). ST320 is a double locus variant (DLV) of ST236, originally described as characteristic for the Taiwan[19F]-14 PMEN clone, associated with serotype 19F and nonsusceptible to penicillin, tetracycline and erythromycin but in contrast to our isolates sensitive to clindamycin [12, 48]. Both ST320 and ST236 are included in a large clonal complex named CC320/271 with either ST236 or ST271 (a single locus variant, SLV, of ST320) considered its likely ancestor [16, 49, 50]. In the recently introduced core genome MLST (cgMLST) scheme CC320/271 corresponds to GPSC1 [50]. The 3238/09 and 3641/15 isolates were associated with rST573 and rST572, respectively, differing by three loci (*rpsG, rpsI* and *rplM*) out of 53 rMLST loci [33].

Complete genomes of 19A and 19F isolates and their features

The WGS of two analysed isolates yielded complete closed chromosomes (Table 1) with the ANI equal 99.69%. Both genomes showed the presence of four complete rRNA loci and 58 tRNA genes. The *cps* locus of the 3238/09 isolate of serotype 19F was in 99.98% identical to the *cps* of the Taiwan[19F]-14 isolate (CP000921.1) and belonged to the subtype 19F-I [51]. The *cps* locus of the 3641/15 isolate of serotype 19A was in 100% identical to its counterparts in some other members of

CC320/271, such as the 19A-ST320_99–176 isolate from Korea (CP063829.1) and the SP61 isolate from Germany (CP018137.1) of ST2432, an SLV of ST320. The structure of this *cps* locus represented the subtype 19A-III [51]. Both isolates carried the identical intact loci determining biosynthesis of pilus-1 (P1) and pilus-2 (P2) types of pili. These structures are considered important for pneumococcal colonization and disease [52–54], yet only a minor part of *S. pneumoniae* population i.e. below 30% carries pili genes [54, 55]. The presence of both pili types is a characteristic feature of some CC320/271 isolates [53, 55].

Reduced susceptibility to penicillin and other β-lactams in pneumococci is associated with changes in some of so-called penicillin-binding proteins (PBPs), in particular Pbp1a, Pbp2b and Pbp2x [56]. While the *pbp1a* gene was the same in both isolates, their *pbp2b* and *pbp2x* genes demonstrated 95.0 and 97.5% identity, respectively, and this difference most likely resulted from recombination events (see below). The *pbp2b* in the 19F isolate was shared with a number of isolates belonging to CC320/271, such as NUBL-1080, RMV7, SP64, SP61, 19A-ST320_99–176 and TCH8431/19A (LC198130.1, OV904788.1, CP018138.1, CP018137.1, CP063829.1 and CP001993.1, respectively) and in the 19A isolate this gene was novel, with the closest hit (98.4% identity) to *pbp2b* of the URAspn6056 isolate of unknown serotype from Portugal (AM779405.1) [57]. The *pbp2x* in the 19F isolate was identical solely to *pbp2x* of the NUBL-1080 and RMV7 isolates mentioned above, and the 19A isolate harboured a novel gene, identical in 98.7% to *pbp2x* of Tw03–308 of 6B serotype (KC522447.1) [58].

Both isolates harboured the *tet*(M) tetracycline resistance gene, and *erm*(B) and *mef*(A) macrolide resistance genes, in concordance with the observed phenotypes. All three genes were located in the same genomic region and its analysis demonstrated the presence of

(See figure on next page.)

Fig. 1 Recombination events affecting genomes of Polish invasive *S. pneumoniae* isolates of ST320 and serotypes 19A and 19F elucidated using Gubbins and visualized in PHANDANGO. The annotated 3238/09 genome used as reference, with short vertical blue lines of various thickness representing CDSs; phylogenetic trees on the left; pink bars, recombinations shared by clusters of isolates; violet bars, recombinations unique for isolates at terminal branches. **A** Analysis of the 3238/09 and 3641/15 genomes from this study and the complete SP64 and TCH8431-19A genomes of ST320 and serotype 19A from GenBank (CP018138.1 and CP001993.1, respectively). Recombination blocks, distinguishing the 3641/15 genome from the 3238/09 genome indicated above the reference genome; the approximately 161-kb *cps* region shadowed. **B** Details of the *cps* region from (**A**) depicted for the 3238/09 and 3641/15 genomes. The putative mobile genetic element (MGE) [49] and the *cps* operon marked by thick black bars; localization of *pbp2x, dexB, aliA, pbp1a* indicated with thin arrows; recombination blocks, distinguishing the 3641/15 genome from the 3238/09 genome indicated below. **C** Analysis of the approximately 0.3 Mb contig harbouring the *cps* operon from the SN39039 isolate [49] and the corresponding parts of the 3238/09, 3641/15, SP64 and TCH8431-19A genomes. MGE, *cps, pbp2x, dexB, aliA, pbp1a* indicated as in (**B**); the 76.5 kb recombining fragment, acquired by the SN39039 isolate [49] indicated by a thick double-headed arrow; recombination blocks upstream and downstream *cps* specific for the 3641/15 isolate described in the lower part of the figure, with the recombining segment R3 shadowed in grey
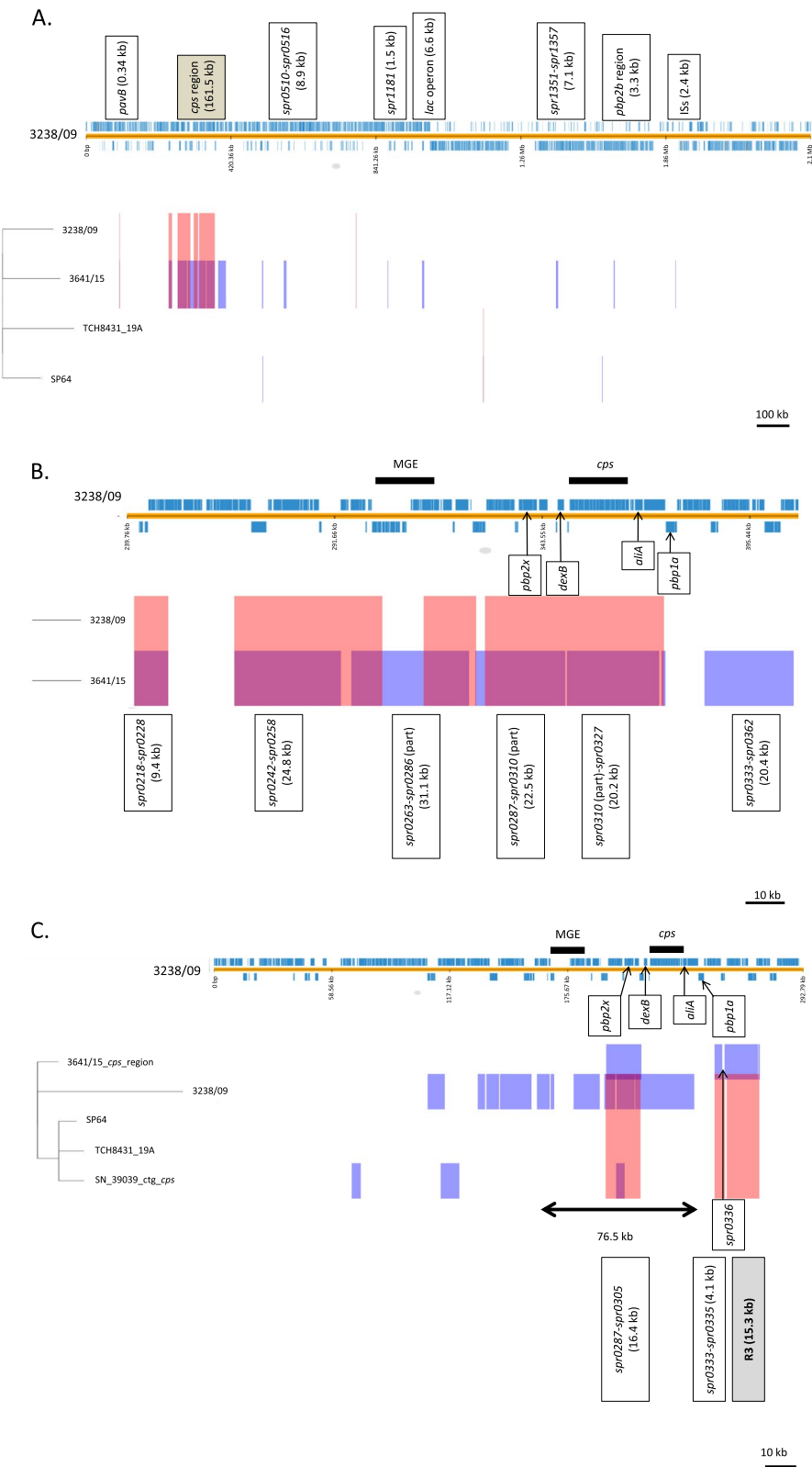
Puzia *et al. BMC Genomic Data*   (2023) 24:19

Page 6 of 10

Fig. 1 (See legend on previous page.)

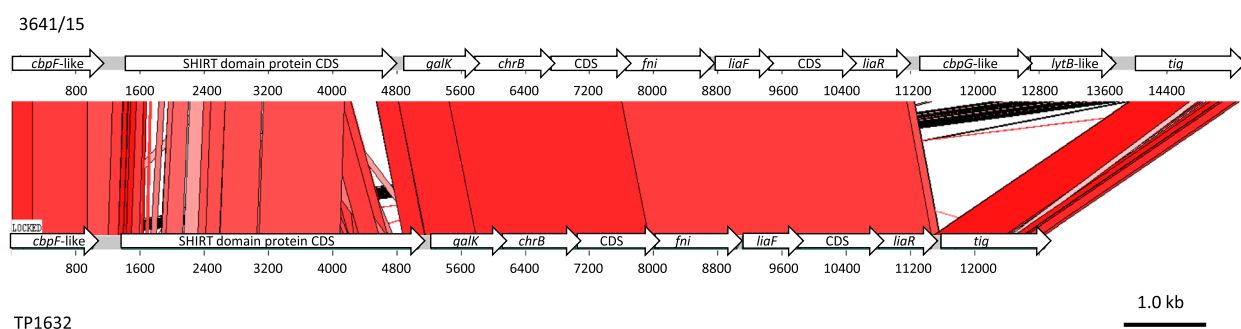Puzia *et al. BMC Genomic Data*   (2023) 24:19

Page 7 of 10

Fig. 2 The 15.3 kb recombining segment R3 from 3641/15 isolate, compared to the corresponding part of the genome from the TP1632 isolate of *S. toyakuensis* sp. nov. using ACT. CDSs depicted as blue rectangles, homology blocks indicated in red

the Tn*2010*-type transposon with the right terminus located in the counterpart of *spr1764* and the left terminus positioned upstream an ORF corresponding to *spr1775* of the R6 genome [45]. Tn*2010* is a 26.4 kb composite derivative of Tn*916* with insertions of *erm*(B) and *mega* elements [59, 60]. The observed localization of Tn*2010* is characteristic for several genomes of isolates associated with CC320/271 belonging to 19A and 19F serotypes [59, 60]. Tn*2010* showed 99.8% identity in the two isolates due to the presence of a unique 42-bp deletion in the putative replication initiation protein gene, corresponding to ORF20 in the original Tn*916*.

A single phage, 18,529 bp in size was located in the 3238/09 genome between the CDSs corresponding to *spr0003* and *spr0004* in the R6 genome. This phage occurred also in the same genetic localization in other genomes of pneumococci belonging to CC320/271 such as the ST556, RMV7, Taiwan19F-14, 19A-19,087, 19A-19,339, 19A-19,343, 19A-ST320_99−176 and TCH8431/19A isolates (GenBank accession numbers: CP003357.2, OV904788.1, CP035237.1, CP071916.1, CP071917.1, CP071918.1, CP063829.1 and CP001993.1, respectively) and was in 99.99% identical to the *Streptococcus* spp. satellite phage Javan759 [61].

No plasmids were found in any of the two isolates, which is in agreement with a very rare occurrence of plasmids in *S. pneumoniae* in general [62].

Putative recombination events affecting genomes of 19A and 19F isolates

Analysis of two genomes revealed several instances of possible recombination events, especially in the region surrounding the *cps* locus (Fig. 1A). Two other complete genomes of *S. pneumoniae* of ST320 and serotype 19A, SP64 and TCH8431/19A available in GenBank

(CP018138.1 and CP001993.1, respectively; date accessed 28th August 2020) were also included in these analyses. An internal recombination in the *pavB* gene (*spr0075*), whose product is a virulence factor involved in binding of host fibronectin and plasminogen [63], resulted in two different variants of *pavB* in 3238/09 and 3641/15. The difference between *pbp2b* of our two isolates, mentioned above was due to a local recombination event involving approximately the 3.3 kb segment carrying *pbp2b* (*spr1517*) and parts of the adjacent genes *recR* and *ybbH*. Other recombination events affected structural genes, e.g. the *metEF* genes and CDSs of unknown function in the region corresponding to *spr0510-spr0516*, *gdhA* (*spr1181),* most of the *lac* operon (*spr1070-spr1077*) and the *glnHQP* genes, residing in the region corresponding to *spr1351-spr1357*. Recombination occurred also in a locus harbouring presumably functional and degenerate insertion sequences (*spr1701-spr1702*). The approximately 161-kb region including the *cps* operon was affected by a number of recombination events (Fig. 1B). Frequent recombinations in this area in CC320/271 were observed also by others [49] but the reason of this phenomenon is unclear. Apart from the exchange of the *cps* operon itself, five other recombination blocks, differentiating genomes of two Polish isolates were observed in this area. While one of these blocks contained the *pbp2x* gene, resulting in its different variants in the two isolates as described above, the *pbp1a* gene located downstream *cps* remained unaffected. In general, recombination blocks detected by analysis of the two genomes varied in size from 0.34 kb (*pavB*) to 31.1 kb (*spr0263-spr0286* in the *cps* region), in a good agreement with sizes of exchanged fragments observed for other vaccine escape recombinants [64]. It was proposed that at least four independent recombinations resulting in the change of original serotype 19F to 19A occurred in CC320/271 and

Puzia *et al. BMC Genomic Data*   (2023) 24:19

Page 8 of 10

one of these events, involving an approximately 76.5 kb region yielded a strain with ST320 and 19A serotype, represented by the SN39039 isolate [49]. In the case of other recombinant analysed in the same study, the 8312–05 isolate representing ST236[19A], a potential donor of the *cps19A* genes belonged to ST199, however for SN39039 such donor could not be identified [49]. To investigate the relationship of SN39039 with Polish isolates, the approximately 0.3 Mb contig harbouring the *cps19A* operon from SN39039 was included in analysis with the corresponding parts of four previously analysed genomes (Fig. 1C). While the *cps19A* operon sequences in 3641/15 and SN39039 were identical, our analysis revealed three recombination blocks upstream and downstream *cps19A* distinguishing the 3641/15 isolate from SN39039. The 16.4-kb recombination block located upstream from 3641/15 showed 95.7% identity to the non-capsulated NT 100_58 strain [65] and of the two blocks identified downstream *cps19A,* the 4.1-kb region showed 96.4% identity to the KK1157 isolate (AP018044.1). Thus, the direct donor(s) of both these sequence blocks could not be determined. The 15.3-kb region marked R3 was most the most peculiar one, since it was most similar to a 12.9-kb fragment of the TP1632 genome of a newly proposed species, *Streptococcus toyakuensis* [66] (Fig. 2). These two fragments shared 82.0% identity. Two additional genes, presumably encoding CbpG-like and LytB-like proteins were present in R3 of the 3641/15 isolate. Both isolates in this part of the genome harboured a gene encoding a putative adhesin (pfam18655) containing Streptococcal High Identity Repeats in Tandem (SHIRT) domains reported in some species of viridians streptococci [67]. The deduced sequence of this protein was 1100 and 1224 amino acids long and harboured four and six SHIRT domains in 3641/15 and TP1632, respectively. To our knowledge, this is the first observation of the gene specifying an SHIRT domain protein in *S. pneumoniae.*

Conclusions

In our study we report the complete genomic sequences of two isolates of *S. pneumoniae,* demonstrating serotypes 19A and 19F and belonging to ST320. The detailed comparative analysis of these genomes suggested several recombination events, particularly affecting the region including the *cps* locus, determining the biosynthesis of capsular polysaccharide, the major virulence factor of pneumococcus and the target of antipneumococcal vaccines. The complexity of recombination and gene acquisition events as well as lack of sequences of direct potential donors in GenBank precluded a complete reconstruction of evolutionary history of the *cps* region in the 3641/15 isolate, a presumable vaccine escape recombinant.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. AlonsoDeVelasco E, Verheul A, Verhoef J, Snippe H. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. Microbiol Rev. 1995;59:591–603. https://doi.org/10.1128/mr.59.4.591-603.1995.
2. WHO. Pneumococcal conjugate vaccines in infants and children under 5 years of age: WHO position paper – February 2019. WER. 2019;94:85–104.
3. Austrian R. Some observations on the pneumococcus and on the current status of pneumococcal disease and its prevention. Rev Infect Dis. 1981;3(Suppl):S1–S17. https://doi.org/10.1093/clinids/3.supplement_1.s1.
4. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. 2006;2:e31. https://doi.org/10.1371/journal.pgen.0020031.
5. Aanensen DM, Mavroidi A, Bentley SD, Reeves PR, Spratt BG. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol. 2007;189:7856–76. https://doi.org/10.1128/JB.00837-07.
6. LaFon DC, Nahm MH. Measuring immune responses to pneumococcal vaccines. J Immunol Methods. 2018;461:37–43. https://doi.org/10.1016/j.jim.2018.08.002.
7. Whitney CG, Farley MM, Hadler J, Harrison LH, Bennett NM, Lynfield R, et al. Active bacterial Core surveillance of the emerging infections program network. Decline in invasive pneumococcal disease after the

Puzia *et al. BMC Genomic Data*   (2023) 24:19

Page 9 of 10

introduction of protein-polysaccharide conjugate vaccine. N Engl J Med. 2003;348:1737–46. https://doi.org/10.1056/NEJMoa022823.

8. Kyaw MH, Lynfield R, Schaffner W, Craig AS, Hadler J, Reingold A, et al. Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae. N Engl J Med. 2006;354(14):1455–63. https://doi.org/10.1056/NEJMoa051642.

9. Hampton LM, Farley MM, Schaffner W, Thomas A, Reingold A, Harrison LH, et al. Prevention of antibiotic-nonsusceptible *Streptococcus pneumoniae* with conjugate vaccines. J Infect Dis. 2012;205:401–11. https://doi.org/10.1093/infdis/jir755.

10. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae:* identification of clones associated with serious invasive disease. Microbiology. 1998;144:3049–60. https://doi.org/10.1099/00221287-144-11-3049.

11. Klugman KP. The successful clone: the vector of dissemination of resistance in *Streptococcus pneumoniae*. J Antimicrob Chemother. 2002;50 Suppl S2:1–5. https://doi.org/10.1093/jac/dkf500.

12. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, et al. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. J Clin Microbiol. 2001;39:2565–71. https://doi.org/10.1128/JCM.39.7.2565-2571.2001.

13. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011;331:430–4. https://doi.org/10.1126/science.1198545.

14. Spratt BG, Greenwood BM. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? Lancet. 2000;356:1210–1. https://doi.org/10.1016/S0140-6736(00)02779-3.

15. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet. 2011;378:1962–73. https://doi.org/10.1016/S0140-6736(10)62225-8.

16. Beall BW, Gertz RE, Hulkower RL, Whitney CG, Moore MR, Brueggemann AB. Shifting genetic structure of invasive serotype 19A pneumococci in the United States. J Infect Dis. 2011;203:1360–8. https://doi.org/10.1093/infdis/jir052.

17. Sheppard CL, Groves N, Andrews N, Litt DJ, Fry NK, Southern J, et al. The genomics of *Streptococcus pneumoniae* carriage isolates from UK children and their household contacts, pre-PCV7 to post-PCV13. Genes (Basel). 2019;10:687. https://doi.org/10.3390/genes10090687.

18. Coffey TJ, Enright MC, Daniels M, Morona JK, Morona R, Hryniewicz W, et al. Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. Mol Microbiol. 1998;27:73–83. https://doi.org/10.1046/j.1365-2958.1998.00658.x.

19. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog. 2007;3:e168. https://doi.org/10.1371/journal.ppat.0030168.

20. Isturiz R, Sings HL, Hilton B, Arguedas A, Reinert RR, Jodar L. *Streptococcus pneumoniae* serotype 19A: worldwide epidemiology. Expert Rev Vaccines. 2017;16:1007–27. https://doi.org/10.1080/14760584.2017.1362339.

21. Cassiolato AP, Almeida SCG, Andrade AL, Minamisava R, Brandileone MCC. Expansion of the multidrug-resistant clonal complex 320 among invasive *Streptococcus pneumoniae* serotype 19A after the introduction of a ten-valent pneumococcal conjugate vaccine in Brazil. PLoS One. 2018;13:e0208211. https://doi.org/10.1371/journal.pone.0208211.

22. Moore MR, Link-Gelles R, Schaffner W, Lynfield R, Lexau C, Bennett NM, et al. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: analysis of multisite, population-based surveillance. Lancet Infect Dis. 2015;15:301–9. https://doi.org/10.1016/S1473-3099(14)71081-3.

23. Metcalf BJ, Gertz RE Jr, Gladstone RA, Walker H, Sherwood LK, Jackson D, et al. Active bacterial Core surveillance team. 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.e9–60.e29. https://doi.org/10.1016/j.cmi.2015.08.027.

24. Sempere J, Llamosí M, López Ruiz B, Del Río I, Pérez-García C, Lago D, et al. Effect of pneumococcal conjugate vaccines and SARS-CoV-2 on antimicrobial resistance and the emergence of *Streptococcus pneumoniae* serotypes with reduced susceptibility in Spain, 2004-20: a national

surveillance study. Lancet Microbe. 2022;3:e744–52. https://doi.org/10.1016/S2666-5247(22)00127-6.

25. Polkowska A, Skoczyńska A, Paradowska-Stankiewicz I, Stefanoff P, Hryniewicz W, Kuch A, et al. Pneumococcal meningitis before the introduction of 10-valent pneumococcal conjugate vaccine into the National Childhood Immunization Program in Poland. Vaccine. 2019;37:1365–73. https://doi.org/10.1016/j.vaccine.2018.12.028.

26. Skoczyńska A, Kuch A, Sadowy E, Waśko I, Markowska M, Ronkiewicz P, et al. Participants of a laboratory-based surveillance of community acquired invasive bacterial infections (BINet). Recent trends in epidemiology of invasive pneumococcal disease in Poland. Eur J Clin Microbiol Infect Dis. 2015;34(4):779–87. https://doi.org/10.1007/s10096-014-2283-8.

27. Puzia W, Ronkiewicz P, Kuch A, Hryniewicz W, Sadowy E, Skoczyńska A. *Streptococcus pneumoniae* of serotype 19A before introduction of population based vaccination in Poland. Poster P0359. 29th ECCMID Amsterdam, 13–16.04; 2019.

28. World Health Organization & Centers for Disease Control and Prevention ((U.S.)((2011)(. Laboratory methods for the diagnosis of meningitis caused by *Neisseria meningitidis, Streptococcus pneumoniae*, and *Haemophilus influenzae*: WHO manual, 2nd ed. World Health Organization. https://apps.who.int/iris/handle/10665/70765.

29. Neufeld F. Über die agglutination der pneumokokken und über die theorie der agglutination. Z Hyg Infekt. 1902;40:54–72. https://doi.org/10.1007/BF02140530.

30. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. 2022;Version 12.0. http://www.eucast.org.

31. Wilson K. Preparation of genomic DNA from bacteria. Curr Protoc Mol Biol. 2001;56:2.4.1–5.

32. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. 2018; https://www.biorxiv.org/content/early/2018/04/09/274100. Accessed 20 Sept 2022.

33. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology (Reading). 2012;158:1005–15. https://doi.org/10.1099/mic.0.055459-0.

34. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124. https://doi.org/10.12688/wellcomeopenres.14826.1.

35. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017;3:e000132. https://doi.org/10.1099/mgen.0.000132.

36. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666–9. https://doi.org/10.1093/bioinformatics/bty149.

37. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

38. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44:6614–24. https://doi.org/10.1093/nar/gkw569.

39. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110:1281–6. https://doi.org/10.1007/s10482-017-0844-4.

40. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4. https://doi.org/10.1093/jac/dks261.

41. Arndt D, Grant J, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44:W16–21. https://doi.org/10.1093/nar/gkw387.

42. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15. https://doi.org/10.1093/nar/gku1196.

43. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics. 2018;34:292–3. https://doi.org/10.1093/bioinformatics/btx610.

44. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis comparison tool. Bioinformatics. 2005;21:3422–3. https://doi.org/10.1093/bioinformatics/bti553.
45. Hoskins J, Alborn WE Jr, Arnold J, Blaszczak LC, Burgett S, DeHoff BS, et al. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J Bacteriol. 2001;183:5709–17. https://doi.org/10.1128/JB.183.19.5709-5717.2001.
46. Golden AR, Rosenthal M, Fultz B, Nichol KA, Adam HJ, Gilmour MW, et al. Characterization of MDR and XDR *Streptococcus pneumoniae* in Canada, 2007-13. J Antimicrob Chemother. 2015;70:2199–202. https://doi.org/10.1093/jac/dkv107.
47. Kim SH, Song JH, Chung DR, Thamlikitkul V, Yang Y, Wang H, et al. Changing trends in antimicrobial resistance and serotypes of *Streptococcus pneumoniae* isolates in Asian countries: an Asian network for surveillance of resistant pathogens (ANSORP) study. Antimicrob Agents Chemother. 2012;56(3):1418–26. https://doi.org/10.1128/AAC.05658-11.
48. Shi ZY, Enright MC, Wilkinson P, Griffiths D, Spratt BG. Identification of three major clones of multiply antibiotic-resistant *Streptococcus pneumoniae* in Taiwanese hospitals by multilocus sequence typing. J Clin Microbiol. 1998;36:3514–9. https://doi.org/10.1128/JCM.36.12.3514-3519.1998.
49. Croucher NJ, Chewapreecha C, Hanage WP, Harris SR, McGee L, van der Linden M, et al. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. Genome Biol Evol. 2014;6:1589–602. https://doi.org/10.1093/gbe/evu120.
50. Gladstone RA, Lo SW, Goater R, Yeats C, Taylor B, Hadfield J, et al. The global pneumococcal sequencing consortium. Visualizing variation within global pneumococcal sequence clusters (GPSCs) and country population snapshots to contextualize pneumococcal isolates. Microb. Genom. 2020;6:e000357. https://doi.org/10.1099/mgen.0.000357.
51. Elberse K, Witteveen S, van der Heide H, van de Pol I, Schot C, van der Ende A, et al. Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. PLoS One. 2011;6:e25018. https://doi.org/10.1371/journal.pone.0025018.
52. Ness S, Hilleringmann M. *Streptococcus pneumoniae* type 1 pilus - a multifunctional tool for optimized host interaction. Front Microbiol. 2021;12:615924. https://doi.org/10.3389/fmicb.2021.615924.
53. Bagnoli F, Moschioni M, Donati C, Dimitrovska V, Ferlenghi I, Facciotti C, et al. A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. J Bacteriol. 2008;190:5480–92. https://doi.org/10.1128/JB.00384-08.
54. Iovino F, Nannapaneni P, Henriques-Normark B, Normark S. The impact of the ancillary pilus-1 protein RrgA of *Streptococcus pneumoniae* on colonization and disease. Mol Microbiol. 2020;113:650–8. https://doi.org/10.1111/mmi.14451.
55. Dzaraly ND, Muthanna A, Mohd Desa MN, Taib NM, Masri SN, Rahman NIA, et al. Pilus islets and the clonal spread of piliated *Streptococcus pneumoniae*: a review. Int J Med Microbiol. 2020;310:151449. https://doi.org/10.1016/j.ijmm.2020.151449.
56. Zapun A, Contreras-Martel C, Vernet T. Penicillin-binding proteins and beta-lactam resistance. FEMS Microbiol Rev. 2008;32:361–85. https://doi.org/10.1111/j.1574-6976.2007.00095.x.
57. Dias R, Félix D, Caniça M. Diversity of penicillin binding proteins among clinical *Streptococcus pneumoniae* strains from Portugal. Antimicrob Agents Chemother. 2008;52:2693–5. https://doi.org/10.1128/AAC.01655-07.
58. Ko KS, Baek JY, Song JH. Capsular gene sequences and genotypes of "serotype 6E" *Streptococcus pneumoniae* isolates. J Clin Microbiol. 2013;51:3395–9. https://doi.org/10.1128/JCM.01645-13.
59. Del Grosso M, Northwood JG, Farrell DJ, Pantosti A. The macrolide resistance genes *erm*(B) and *mef*(E) are carried by Tn*2010* in dual-gene *Streptococcus pneumoniae* isolates belonging to clonal complex CC271. Antimicrob Agents Chemother. 2007;51:4184–6. https://doi.org/10.1128/AAC.00598-07.
60. Li Y, Tomita H, Lv Y, Liu J, Xue F, Zheng B, et al. Molecular characterization of *erm*(B)- and *mef*(E)-mediated erythromycin-resistant *Streptococcus pneumoniae* in China and complete DNA sequence of Tn*2010*. J Appl Microbiol. 2011;110:254–65. https://doi.org/10.1111/j.1365-2672.2010.04875.x.
61. Rezaei Javan R, Ramos-Sevillano E, Akter A, Brown J, Brueggemann AB. Prophages and satellite prophages are widespread in *Streptococcus*

and may play a role in pneumococcal pathogenesis. Nat Commun. 2019;10:4852. https://doi.org/10.1038/s41467-019-12825-y.
62. Santoro F, Iannelli F, Pozzi G. Genomics and genetics of *Streptococcus pneumoniae*. Microbiol Spectr. 2019;7. https://doi.org/10.1128/microbiolspec.GPP3-0025-2018.
63. Jensch I, Gámez G, Rothe M, Ebert S, Fulde M, Somplatzki D, et al. PavB is a surface-exposed adhesin of *Streptococcus pneumoniae* contributing to nasopharyngeal colonization and airways infections. Mol Microbiol. 2010;77:22–43. https://doi.org/10.1111/j.1365-2958.2010.07189.x.
64. Golubchik T, Brueggemann AB, Street T, Gertz RE Jr, Spencer CC, Ho T, et al. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. Nat Genet. 2012;44:352–5. https://doi.org/10.1038/ng.1072.
65. Hilty M, Wüthrich D, Salter SJ, Engel H, Campbell S, Sá-Leão R, et al. Global phylogenomic analysis of nonencapsulated *Streptococcus pneumoniae* reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. Genome Biol Evol. 2014;6:3281–94. https://doi.org/10.1093/gbe/evu263.
66. Wajima T, Hagimoto A, Tanaka E, Kawamura Y, Nakaminami H. Identification and characterisation of a novel multidrug-resistant streptococcus, *Streptococcus toyakuensis* sp. nov., from a blood sample. J Glob Antimicrob Resist. 2022;29:316–22. https://doi.org/10.1016/j.jgar.2022.04.018.
67. Whelan F, Lafita A, Gilburt J, Dégut C, Griffiths SC, Jenkins HT, et al. Periscope proteins are variable-length regulators of bacterial cell surface interactions. Proc Natl Acad Sci U S A. 2021;118:e2101349118. https://doi.org/10.1073/pnas.2101349118.

Publisher's Note