



Genomic Characterization of the Emerging Pathogen *Streptococcus pseudopneumoniae*

Geneviève Garriss,^a Priyanka Nannapaneni,^a Alexandra S. Simões,^b Sarah Browall,^a Karthik Subramanian,^a Raquel Sá-Leão,^b Herman Goossens,^c Herminia de Lencastre,^{b,d} Birgitta Henriques-Normark^{a,e,f}

^aDepartment of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

^bLaboratory of Molecular Genetics, Instituto de Tecnologia Química e Biológica Antonio Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

^cLaboratory of Medical Microbiology, Vaccine & Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Antwerp, Belgium

^dLaboratory of Microbiology and Infectious Diseases, The Rockefeller University, New York, New York, USA

^eLee Kong Chian School of Medicine (LKC) and Singapore Centre on Environmental Life Sciences Engineering (SCELS), Nanyang Technological University, Singapore

^fClinical Microbiology, Karolinska University Hospital, Bioclinicum, Stockholm, Sweden

ABSTRACT *Streptococcus pseudopneumoniae* is a close relative of the major human pathogen *S. pneumoniae*. It is increasingly associated with lower-respiratory-tract infections (LRTI) and a high prevalence of antimicrobial resistance (AMR). *S. pseudopneumoniae* is difficult to identify using traditional typing methods due to similarities with *S. pneumoniae* and other members of the mitis group (SMG). Using whole-genome sequencing of LRTI isolates and a comparative genomic approach, we found that a large number of pneumococcal virulence and colonization genes are present in the core *S. pseudopneumoniae* genome. We also reveal an impressive number of novel surface-exposed proteins encoded by the genome of this species. In addition, we propose a new and entirely specific molecular marker useful for the identification of *S. pseudopneumoniae*. Phylogenetic analyses of *S. pseudopneumoniae* show that specific clades are associated with allelic variants of core proteins. Resistance to tetracycline and macrolides, the two most common types of resistance, were found to be encoded by Tn916-like integrating conjugative elements and Mega-2. Overall, we found a tight association of genotypic determinants of AMR and phenotypic AMR with a specific lineage of *S. pseudopneumoniae*. Taken together, our results shed light on the distribution in *S. pseudopneumoniae* of genes known to be important during invasive disease and colonization and provide insight into features that could contribute to virulence, colonization, and adaptation.

IMPORTANCE *S. pseudopneumoniae* is an overlooked pathogen emerging as the causative agent of lower-respiratory-tract infections and associated with chronic obstructive pulmonary disease (COPD) and exacerbation of COPD. However, much remains unknown on its clinical importance and epidemiology, mainly due to the lack of specific markers to distinguish it from *S. pneumoniae*. Here, we provide a new molecular marker entirely specific for *S. pseudopneumoniae* and offer a comprehensive view of the virulence and colonization genes found in this species. Finally, our results pave the way for further studies aiming at understanding the pathogenesis and epidemiology of *S. pseudopneumoniae*.

KEYWORDS infectious disease, *Streptococcus pseudopneumoniae*, *Streptococcus pneumoniae*, bacterial diagnostics, comparative genomics

Streptococcus pseudopneumoniae is a close relative of the human pathogen *S. pneumoniae*, and it was first described in 2004 (1). It belongs to the mitis group along with 13 other species, including some of the most common colonizers of the oral cavity, such as *S. mitis* (2). An increasing number of reports indicate that *S. pseudo-*

Citation Garriss G, Nannapaneni P, Simões AS, Browall S, Subramanian K, Sá-Leão R, Goossens H, de Lencastre H, Henriques-Normark B. 2019. Genomic characterization of the emerging pathogen *Streptococcus pseudopneumoniae*. mBio 10:e01286-19. <https://doi.org/10.1128/mBio.01286-19>.

Editor Joerg Vogel, University of Würzburg

Copyright © 2019 Garriss et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Birgitta Henriques-Normark, birgitta.henriques@ki.se.

G.G. and P.N. contributed equally to this work.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Keith Klugman, Emory University; Sven Hammerschmidt, University of Greifswald.

Received 17 May 2019

Accepted 23 May 2019

Published 25 June 2019

pneumoniae is a potential pathogen, usually associated with underlying medical conditions (3–5). It can be isolated from multiple invasive and noninvasive sites (6–9) and was reported as the probable causative agent in fatal septicemia cases (5). Experiments using multiple *S. pseudopneumoniae* strains in a mouse peritonitis/sepsis model have further underlined its pathogenic potential (10). *S. pseudopneumoniae* is also frequently associated with high rates of antimicrobial resistance (AMR), in particular to penicillin, macrolides, co-trimoxazole, and tetracycline (6–8).

Despite its emerging role as a pathogen, relatively little is known about the epidemiology, pathogenic potential, and genetic features of *S. pseudopneumoniae*. This problem is partially attributable to difficulties in distinguishing it from *S. pneumoniae* and *S. mitis*, highlighted by the incorrect identification of 50% of the publicly available genome sequences of *S. pseudopneumoniae* (11, 12). It is likely that infections due to *S. pseudopneumoniae* are overlooked or misdiagnosed due to lack of reliable measures to identify this species. *S. pseudopneumoniae* was originally described as optochin resistant if grown in the presence of 5% CO₂ but susceptible in ambient atmosphere, bile insoluble, and nonencapsulated (1), but exceptions to these phenotypes were later reported (4, 5, 7, 13). Molecular methods, such as PCR amplification of specific markers, mostly aim at identifying pneumococci and, thus, have limited value for the positive identification of *S. pseudopneumoniae*. The only molecular marker reported so far for the identification of *S. pseudopneumoniae*, SPS0002, is also found in a subset of *S. pneumoniae* strains (12). Understanding the clinical significance and epidemiology of *S. pseudopneumoniae* requires more discriminative identification methods and a more complete picture of its genetic diversity.

All *S. pseudopneumoniae* strains described to date lack a polysaccharide capsule, which is considered the major virulence factor of *S. pneumoniae* due to its inhibitory effect on complement-mediated opsonophagocytosis. In addition to the capsule, a plethora of other factors, and especially surface-exposed proteins, have been shown to significantly contribute to pneumococcal disease and colonization (reviewed in references 14 and 15), and some of these features have been identified in *S. pseudopneumoniae* (3, 9, 14, 16). Despite the lack of a capsule, naturally nonencapsulated pneumococci can cause disease, and the surface protein PspK, expressed by a subgroup of nonencapsulated pneumococci, promotes adherence to epithelial cells and mouse nasopharyngeal colonization to levels comparable with those of encapsulated pneumococci (17, 18). A comprehensive overview of the distribution of known and potentially new genes that could promote virulence and colonization in *S. pseudopneumoniae* is, however, still lacking.

In this study, we performed an extensive comparative genomic analysis with the aim of elucidating the molecular features that characterize *S. pseudopneumoniae* and distinguish it from its close relative, *S. pneumoniae*. We show that a substantial number of known pneumococcal virulence factors are conserved in *S. pseudopneumoniae*, and we identify a vast number of novel surface-exposed proteins. Finally, our results establish a tight association of AMR determinants with certain lineages and reveal the composite scenario of genetic elements that characterize this species. Importantly, we identified a genetic marker uniquely present in *S. pseudopneumoniae* that can allow the identification of this overlooked species.

RESULTS

Identification of *S. pseudopneumoniae* genomes. Whole-genome sequencing (WGS) followed by a phylogenetic analysis, including 147 genomes from various streptococci of the mitis group (SMG) species, was performed to classify 24 isolates collected from lower-respiratory-tract infections (LRTI) (19) that we suspected to be *S. pseudopneumoniae* ($n = 16$) or *S. mitis* ($n = 3$) or for which no definitive classification was possible to obtain using traditional typing methods and multilocus sequence analysis (MLSA) ($n = 5$). Twenty-one of 24 LRTI isolates clustered within the *S. pseudopneumoniae* clade (Fig. 1A). The 3 strains initially identified as *S. mitis* clustered within the *S. mitis* clade and are not discussed further in this study. As previously reported (11,

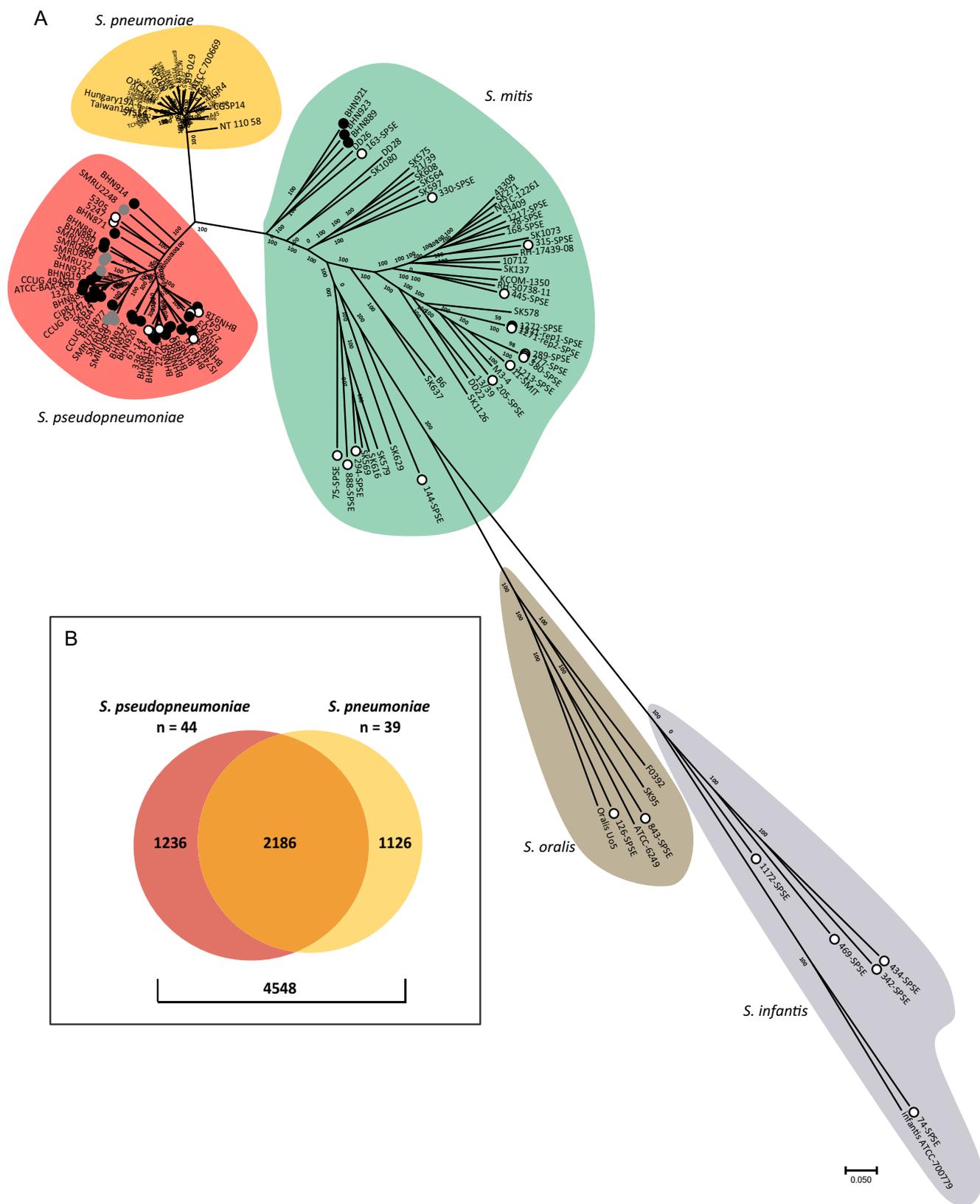


FIG 1 Phylogenetic and pangenome analysis of *S. pseudopneumoniae*. (A) Unrooted consensus parsimony phylogenetic tree based on all SNPs (1,230,968) of 147 genomes: LRTI isolates ($n = 24$) and publicly available *S. pseudopneumoniae* ($n = 38$), *S. pneumoniae* ($n = 39$), nontypeable *S. pneumoniae* ($n = 8$), *S. mitis* ($n = 36$), *S. oralis* ($n = 1$), and *S. infantis* ($n = 1$). Circles indicate LRTI isolates (black), NCBI genomes labeled as *S. pseudopneumoniae* (open), or nontypeable *S. pneumoniae* (gray). Background shading delineates clades of different species. The tree was built in kSNP and visualized in MEGA7 (51). (B) Pangenome of *S. pseudopneumoniae* and *S. pneumoniae* showing the distribution of shared and unique COGs.

TABLE 1 Phenotypic and genotypic characterization of *S. pseudopneumoniae* LRTI isolates

Parameter	No. (%) of strains (n = 29)
Phenotypic markers	
Optochin susceptibility	
5% CO ₂	3 (10.3)
Ambient atmosphere	19 ^a (65.5)
Bile solubility	1 ^b (3.4)
Genotypic markers	
PCR markers	
Pneumococcal <i>lytA</i>	2 (6.9)
<i>cpsA</i>	0 (0.00)
<i>spn9802</i>	28 (96.5)
Pneumococcus-specific 16S rRNA	25 (86.2)
RFLP signatures	
Pneumococcal/atypical <i>lytA</i>	0/29 (0.100)
<i>ply-mly</i>	5/24 (17.2/82.8)

^aEight strains did not grow in ambient atmosphere. The 18/19 strains susceptible in ambient atmosphere were resistant in CO₂.

^bTwo strains showed partial solubility.

12), 8 nontypeable *S. pneumoniae* NCBI genomes fell within the *S. pseudopneumoniae* clade, along with only 15/38 publicly available genomes currently classified as *S. pseudopneumoniae* (Fig. 1A). Based on our phylogenetic analysis, a total of 44 sequenced genomes were considered *S. pseudopneumoniae* and further analyzed (see Data Set S1 in the supplemental material). The pangenome of these 44 *S. pseudopneumoniae* genomes is composed of 3,447 clusters of orthologous genes (COGs), of which 44% are found in the core genome ($\geq 95\%$ isolates) and 56% in the accessory genome.

A single locus, SPPN_RS10375, can be used to identify *S. pseudopneumoniae*.

We then aimed to identify COGs present in *S. pseudopneumoniae* but absent from *S. pneumoniae*. We first defined the pangenome of these two species, using the 44 *S. pseudopneumoniae* genomes and 39 completed and fully annotated *S. pneumoniae* NCBI genomes, and found that 2,186/4,548 COGs (48%) were shared by both species (Fig. 1B). We identified 30 core COGs, present in each of the 44 *S. pseudopneumoniae* genomes, among the 1,236 COGs unique to *S. pseudopneumoniae* (Table S1). We then assessed the presence of each of these 30 COGs in other bacterial species by BLASTn analysis against all NCBI genomes, including the 8,358 *S. pneumoniae* genomes deposited at the time of the study. This revealed that only two COGs, represented by open reading frames (ORFs) SPPN_RS10375 and SPPN_RS06420, were found exclusively in *S. pseudopneumoniae* genomes. While SPPN_RS06420 has a G+C content challenging for the design of PCR primers (average of 27.1%), further analysis of SPPN_RS10375, which encodes a hypothetical protein, and its surrounding intergenic regions in the 44 genomes indicated that this 627-bp locus is a suitable candidate for a molecular marker. Eight clinical isolates not subjected to whole-genome sequencing and collected during the same LRTI study (19) that were either impossible to identify ($n = 4$) or suspected to be *S. pseudopneumoniae* ($n = 4$) were found to be positive by PCR for SPPN_RS10375, indicating that they belong to the *S. pseudopneumoniae* species (Fig. S1A). These strains were also positive for the recently published *S. pseudopneumoniae* marker SPS0002 (12) (Fig. S1B).

Among the 29 *S. pseudopneumoniae* LRTI isolates, 16 (55%) displayed the typical optochin susceptibility and bile solubility phenotypes originally attributed to this species (1) (Table 1). As previously described (7), the pneumococcus-specific markers 16S rRNA and *spn9802* were positive in the majority of the isolates. Discrepancies between the restriction fragment length polymorphism (RFLP) and PCR results used for detecting the pneumococcal variant of the autolysin gene *lytA* were found to be due to phage-encoded *lytA* genes similar enough to be identified by PCR as the pneumococcal *lytA* but lacking the BsaI restriction site used for RFLP analysis (20) (Fig. S1C).

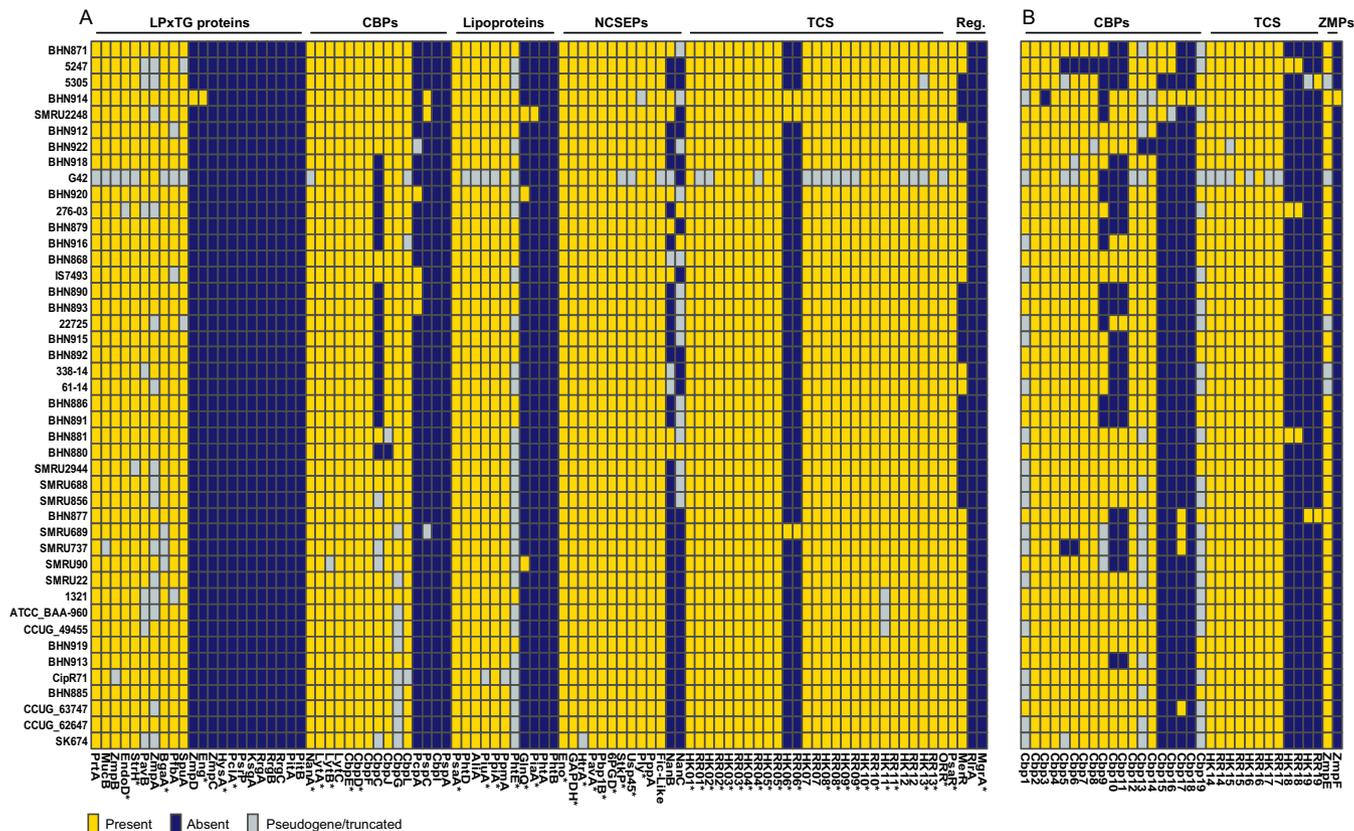


FIG 2 Distribution of known and putative new colonization and virulence factors in the 44 *S. pseudopneumoniae* genomes. (A) Known pneumococcal surface-exposed proteins, two-component systems (histidine kinase/response regulator pairs), and stand-alone regulators. Asterisks indicate core *S. pneumoniae* proteins according to Gamez et al. (21). (B) Distribution of novel choline-binding proteins, two-component systems, and zinc-metalloproteases. CBP, choline-binding protein; NCSEP, nonclassical surface-exposed protein; TCS, two-component system; HK, histidine kinase; RR, response regulator; Reg., stand-alone regulator; ZMPs, zinc metalloproteases.

The pneumococcal variant of the cytotoxin pneumolysin gene *ply* was detected by RFLP in three instances. However, this is due to the presence of the *Bsa*I restriction used for RFLP in some nonpneumococcal variants of *Ply* (see below and Fig. S2).

Pneumococcal virulence and colonization genes are widely distributed in *S. pseudopneumoniae*. To gain insight into genetic features that could promote adhesion, virulence, and colonization, we investigated the presence of orthologues of 92 pneumococcal surface-exposed proteins, transcriptional regulators, and two-component signal transducing systems (TCSs) for which the distribution among pneumococcal genomes has been studied (21, 22). No orthologs were found in *S. pseudopneumoniae* for 16/92 proteins, including the subunits of both pneumococcal pili (RrgABC and PitAB), surface-exposed proteins PspP and PspA, and the stand-alone regulators MgrA and RlrA (Fig. 2A). Three of these sixteen proteins, HysA, PclA, and MgrA, are core *S. pneumoniae* features (21). Other core *S. pneumoniae* proteins were represented in only a very small subset of *S. pseudopneumoniae* strains, such as Eng ($n = 1$), PiaA ($n = 1$), GlnQ ($n = 3$), and the histidine kinase (HK) and response regulator (RR) that constitute TCS06 ($n = 3$). Of 61 surface-exposed proteins, 29 were found in the core *S. pseudopneumoniae* genome, including major virulence factors such as pneumolysin (Ply), NanA, and HtrA (Fig. 2A). The NanA variant found in *S. pseudopneumoniae* shares similar domains and has good similarity with pneumococcal NanA (61.2%). However, it differs strongly in its C-terminal region, where the LPxTG-anchoring domain normally found in *S. pneumoniae* NanA is replaced with a choline-binding domain (CBD). The *S. pseudopneumoniae* Ply proteins (sometimes referred to as Pply) are extremely well conserved (99.1% pairwise identity). Interestingly, while *S. pneumoniae* and *S. pseudopneumoniae* carry Ply in their core genome, it is found in only a small fraction (8%) of *S. mitis* strains.

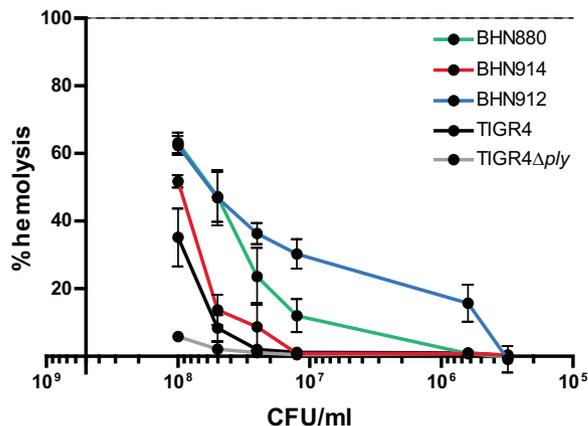


FIG 3 Hemolytic activity of *S. pseudopneumoniae*. The ability of *S. pseudopneumoniae* strains BHN880, BHN912, and BHN914 to lyse human blood was tested using a hemolysis assay. *S. pneumoniae* TIGR4 and TIGR4 Δ ply strains were used as controls. Results are expressed as the percentage of lysis compared to the positive control (100%; indicated by the dashed line) upon incubation with decreasing numbers of CFU/ml. Each data point represents the average from three independent experiments, and standard deviations are indicated.

Although they are closely related to pneumococcal Ply (97.4% pairwise identity), phylogenetic analysis shows that all *S. pseudopneumoniae* and *S. mitis* Ply variants fall in a phylogenetic clade distinct from that of their pneumococcal counterparts (7, 9) (Fig. S2). Hemolysis assays show that *S. pseudopneumoniae* strains encoding Ply proteins from each phylogenetic clade have a hemolytic activity comparable or superior to that of the reference *S. pneumoniae* TIGR4 strain (Fig. 3).

Pneumococcal LPxTG cell wall-anchored proteins were found to have the lowest levels of representation in *S. pseudopneumoniae*, with 12/23 being absent from all genomes. With the exception of TCS06 and HK11, all HK-RR pairs were core *S. pseudopneumoniae* proteins. Two of the three isolates encoding TCS06 also harbor a PspC-like protein in the same locus, such as that found in pneumococcal genomes. These two PspC-like proteins carry an LPxTG-anchoring domain and share limited similarity to each other (30.8%) and to their closest pneumococcal allele, PspC11.3 (32.9%) (23). The third genome encoding TCS06 carries a truncated gene encoding a PspC-like protein.

We further evaluated the presence in *S. pseudopneumoniae* of genes relevant for infection and colonization by investigating the presence of 356 *S. pneumoniae* genes differentially expressed in mouse models of invasive disease (IPD) and during epithelial cell contact (ECC) (24). We found that 94% are present in at least one *S. pseudopneumoniae* genome (Data Set S2). The use of draft *S. pseudopneumoniae* genomes (43/44), in contrast to fully assembled *S. pneumoniae* genomes, would likely result in an underestimation of the presence of these genes due to contig breaks. Hence, we considered genes present in 42/44 *S. pseudopneumoniae* genomes to belong to the core genome. A large fraction of IPD/ECC genes were found in the core genome of *S. pseudopneumoniae* (87%) and *S. pneumoniae* (74%) (Fig. 4A and Table S3). Of the 356 genes, 20 (5.6%) were absent from *S. pseudopneumoniae*, and among them was the gene encoding pneumococcal surface protein A (*pspA*), a known virulence factor present in the majority of *S. pneumoniae* genomes (34/39). The remaining genes belonged to various functional categories (Fig. 4B and Data Set S2).

Identification of an encapsulated *S. pseudopneumoniae* strain. Unexpectedly, our analysis revealed the presence of the capsular polysaccharide biosynthesis genes *cpsA* and *cpsC* in one instance (Data Set S2). Further analysis showed that one LRTI *S. pseudopneumoniae* isolate, BHN880, encodes a full capsular locus similar to the pneumococcal serotype 5 capsule and to the capsule locus of *S. mitis* strain 21/39 (Fig. 5). Gel diffusion assays typed BHN880 as pneumococcal serotype 5, which is supported by the high nucleotide identity (97.7%) between the regions encoding the sugar precursors of

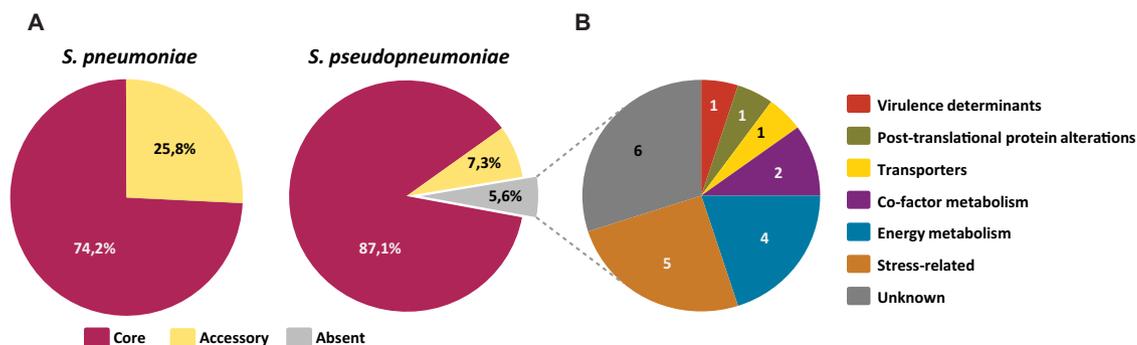


FIG 4 Distribution of IPD and ECC genes. (A) Pie charts representing the percentage of core, accessory, and absent genes from *S. pneumoniae* ($n = 39$) and *S. pseudopneumoniae* ($n = 44$) genomes. Due to the use of draft *S. pseudopneumoniae* genomes compared with complete *S. pneumoniae* genomes, genes were considered core in *S. pseudopneumoniae* when they were found in $\geq 95\%$ of the genomes. (B) Distribution of the 20 genes absent from *S. pseudopneumoniae* in the functional categories defined by Orihuela et al. (24).

BHN880 and serotype 5 capsular loci. The 43 remaining genomes do not encode a capsule and carry the NCC3-type locus, which encompasses genes *dexB*, *aliD*, and *glf* (also known as *cap* or *capN* [18, 25]).

***S. pseudopneumoniae* encodes a substantial number of new potential surface-exposed proteins and two-component systems.** We then investigated if *S. pseudopneumoniae* harbored additional features that could be relevant for colonization and adaptation by searching the proteome of the *S. pseudopneumoniae* species for novel choline-binding proteins (CBPs) and new TCSs. We found 19 previously undescribed proteins containing a choline-binding domain (CBD), which we named Cbp1 to Cbp19. In total, 4 of the 19 proteins belong to the core genome, while the others have various levels of presence among the 44 genomes (Fig. 2B). Each strain carried between 6 and 15 new CBPs, and some *S. pseudopneumoniae* genomes carried a total of 26 CBPs. Prediction of functional domains in these proteins indicates that the majority of these proteins have an SP1 signal peptide and a C-terminal CBD composed of 2 to 9 repeats (Fig. 6A). While no functional domain could be identified in the majority of cases, some proteins contained known domains, such as the trypsin-like serine protease domain and the G5 domain, the latter of which is frequently associated with zinc metalloproteases (ZMPs) such as the IgA protease, ZmpA. In addition, we found that *S. pseudopneumoniae* encodes two new putative ZMPs containing the HEMTH...E motif (26), which we named ZmpE and ZmpF. While ZmpE is present in most of the isolates, ZmpF is found in only one strain (BHN914) (Fig. 2B). ZmpE harbors the domains typically found in ZMPs, such as the pneumococcal ZmpA (Fig. 6B). ZmpF, however, lacks the

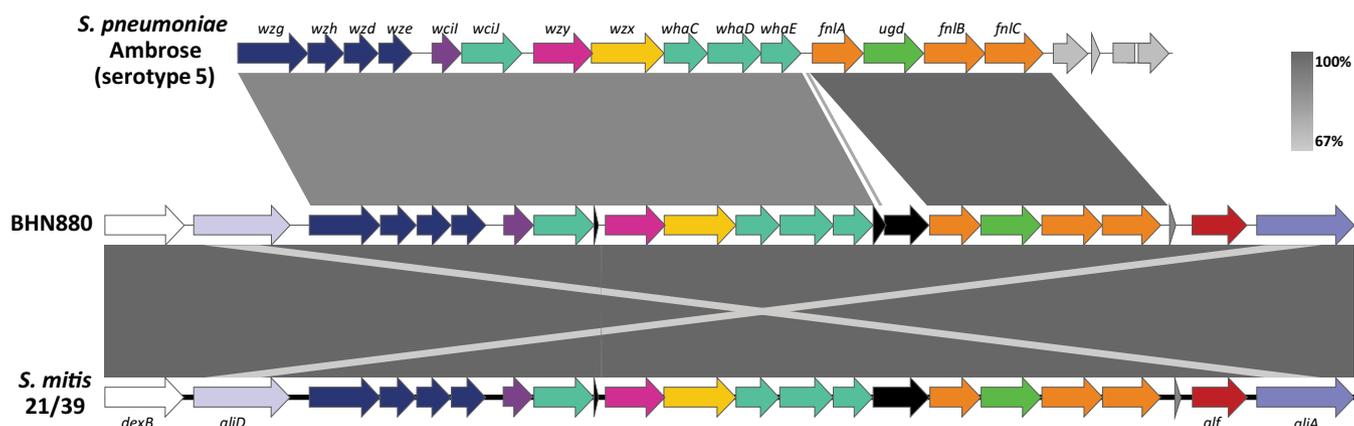


FIG 5 Pairwise alignment of the capsule locus of *S. pseudopneumoniae* strain BHN880 with *S. pneumoniae* Ambrose (serotype 5) and *S. mitis* 21/39. Colors and annotations are based on Bentley et al. (54). Gray shading indicates the degree of pairwise nucleotide identity. Locus tags of *dexB* and *aliA* homologs in BHN880 are E3V35_07635 and E3V35_07740, respectively.

A

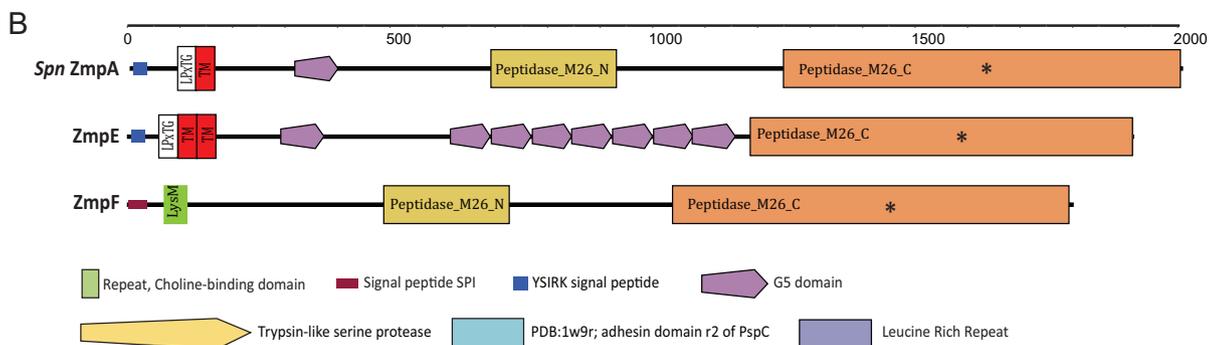
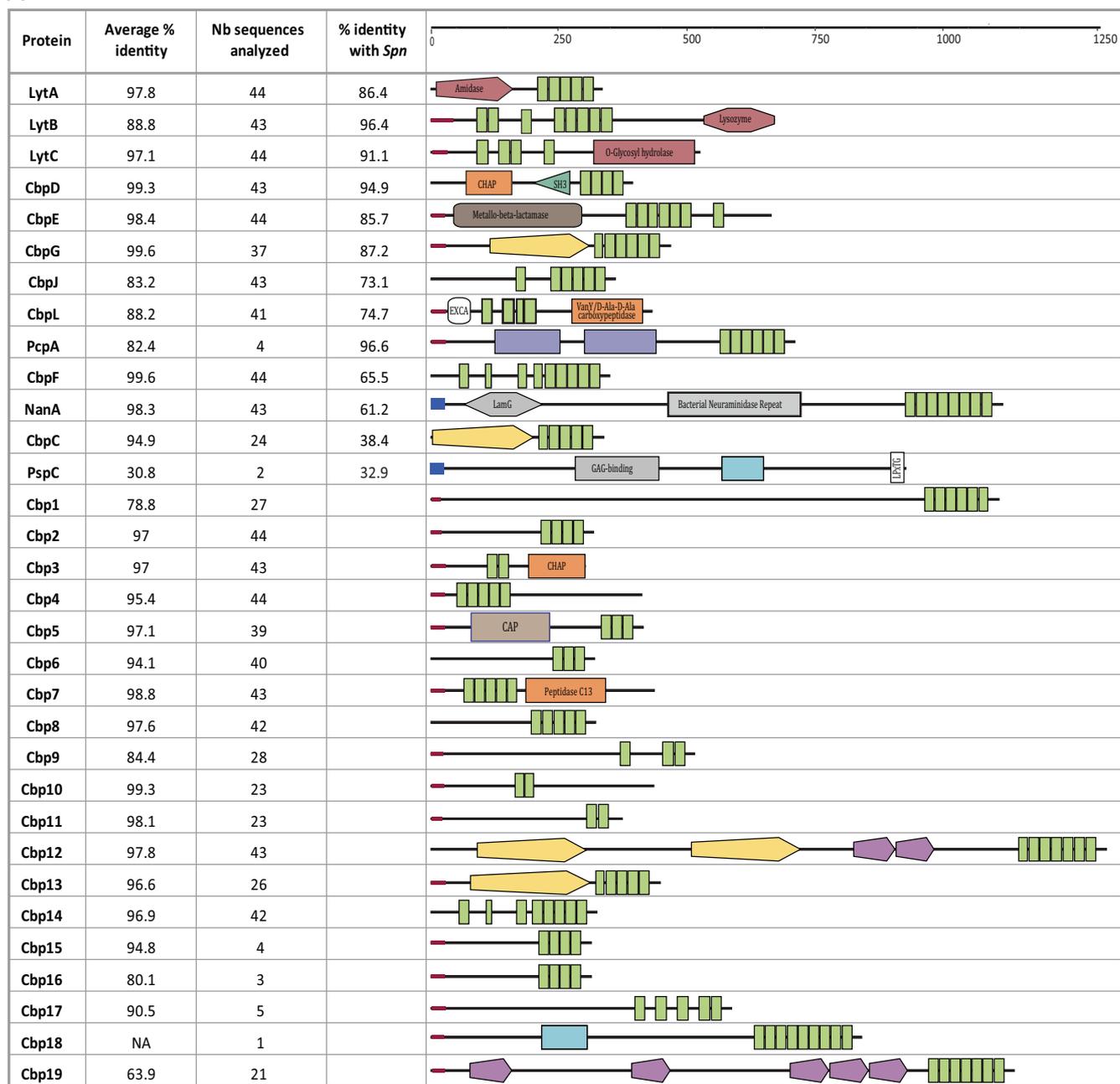


FIG 6 Domain prediction of choline-binding proteins and new zinc metalloproteases of *S. pseudopneumoniae* based on SMART (57). (A) Choline-binding proteins. The average percent identity of each CBP in *S. pseudopneumoniae* and the number of proteins analyzed are indicated. Percent identity with (Continued on next page)

TABLE 2 Novel two-component signaling systems of *S. pseudopneumoniae*

TCS	RR ^a	HK ^b	Species of closest homologue	Family of regulators	Associated gene category
14	SPPN_RS00570	SPPN_RS00565	<i>S. mitis</i>	LytTR	Bacteriocins
15	SPPN_RS11635	SPPN_RS01890	<i>S. mitis</i>	YesN	Ferric iron transport
16	SPPN_RS03570	SPPN_RS03565	<i>S. pseudoporcinus</i> , <i>S. canis</i>	OmpR	Potassium transport ^c
17	SPPN_RS07705	SPPN_RS07700	<i>S. parasanguinis</i>	LytTR/YesN	Thiamine biosynthesis ^d
18	E3V59_10390	E3V59_10385	<i>S. mitis</i>	YesN/AraC	Sugar transport
19	E3V34_05540	E3V34_05535	<i>S. suis</i>	CitB	Bacitracin export

^aLocus tag of the response regulator (RR). Locus tag in IS7493 was used when present.

^bLocus tag of the sensor histidine kinase (HK). Locus tag in IS7493 was used when present.

^cSimilar to KdpD/KdpE from *E. coli* (61).

^dSimilar to TCS02 of *S. thermophilus* (62).

typical LPxTG motif and transmembrane domain and instead carries a LysM domain, which is thought to bind peptidoglycan.

We found six additional HK-RR pairs in the *S. pseudopneumoniae* pangenome, 4 of which are core features (Table 2 and Fig. 2B). We named these TCS14 to TCS19. A more detailed analysis of their genetic loci revealed that TCS14 is found next to genes encoding a ComC/Blp family peptide and bacteriocins. These genes that encode ComC/Blp peptides are distinct from those encoding ComC and BlpC associated with TCS12 and TCS13, respectively, present elsewhere in the *S. pseudopneumoniae* genome. The remaining five TCSs are genetically linked to genes predicted to encode ABC transporters involved in iron, potassium, and sugar transport, thiamine biosynthesis, and bacitracin export.

Bacteriophages are tightly associated with *S. pseudopneumoniae*. Among the 44 *S. pseudopneumoniae* strains, 27 carried at least one putatively full-length prophage, and the remaining 17 strains all carried phage genes, although the presence of full-length prophages could not be confirmed. Twenty-one of the full-length prophages shared a highly related novel integrase ($\geq 90.5\%$ nucleotide identity), which we termed *int_{Sppn1}*, and in 19 cases these prophages were found integrated between SPPN_RS05275 (encoding a putative CYTH domain protein) and SPPN_RS05395 (encoding a putative GTP pyrophosphokinase) (Table S2). The integration site of the remaining 2 full-length prophages encoding *Int_{Sppn1}* could not be confirmed, as they were found alone in a contig without chromosomal flanking sequences. *Int_{Sppn1}* was found in the other 23 strains. However, a full-length prophage could not be confirmed in these strains. In all strains, except for strains G42 and ATCC BAA_960, *int_{Sppn1}* was associated with some phage genes. Six strains carried a second putatively full-length prophage encoding an integrase closely related to that of pneumococcal group 2a prophages, *int2a* (27). These prophages were found between SPPN_RS07570 and SPPN_RS07555, which are the homologs of the genes flanking the phage group 2a integration site in pneumococci (28). Twenty-three other strains harbored *int2a*; however, the presence of more than one phage per strain severely impaired our ability to confirm the completeness of the phages they were associated with, as phage sequences were split between various contigs.

***S. pseudopneumoniae* clades are characterized by distinct alleles of a peptide pheromone and different patterns of antibiotic resistance.** A single-nucleotide polymorphism (SNP)-based phylogenetic tree, which was constructed using the 793 *S. pseudopneumoniae* core COGs, shows that the species is divided into three clades (Fig. 7A). Clades II and III encompass most of the isolates, while clade I is composed of five isolates which fall closer to *S. pneumoniae* genomes (Fig. 7A and Fig. S3A). All three

FIG 6 Legend (Continued)

S. pneumoniae (*Spn*) was calculated using the proteins from IS7493 and *S. pneumoniae* TIGR4, except in the following cases: NanA (R6) and PspC (allele PspC11.3; AF276622.1). Representations of domains found in each CBP are based on the variant found in IS7493. In the absence of the protein in IS7493, the analysis was based on BHN914 (PspC, Cbp15, Cbp16, Cbp17, Cbp18, and Cbp19), BHN879 (Cbp1), and BHN886 (Cbp19). Nb, number. (B) Zinc metalloproteases ZmpE and ZmpF from *S. pseudopneumoniae*. ZmpA from *S. pneumoniae* (*Spn* ZmpA) is included for comparison. Asterisks indicate the ZMP motif HEMTH...E (26). Domain prediction is based on ZmpE from IS7493 and ZmpF from BHN914. Locus tags can be found in Table S3.

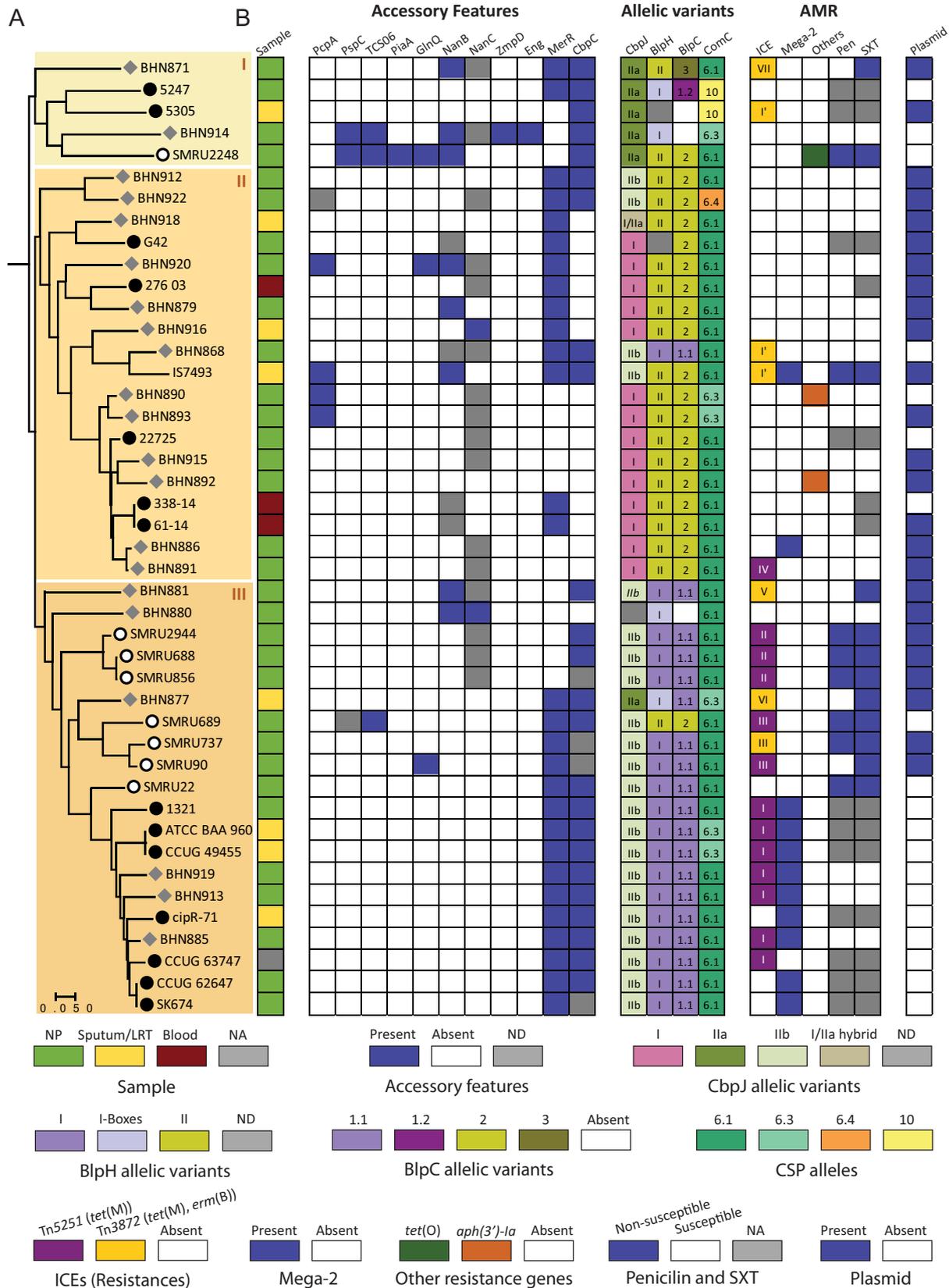


FIG 7 Phylogenetic distribution of accessory features and allelic variants. (A) Core genome species tree-based SNPs from the 793 single-copy core genes of 44 *S. pseudopneumoniae* genomes. Circles indicate isolates from LRTI (black), NCBI genomes labeled as *S. pseudopneumoniae* (open), or nontypeable *S. pneumoniae* (gray). The tree was built in panX (52) and visualized in MEGA7 (51). Clades are delineated by the background shading. (B) Distribution of accessory features and allelic variants of surface-exposed proteins, regulatory genes, peptide (Continued on next page)

clades are composed of strains isolated from the nasopharynx and from sputum or lower-respiratory-tract samples. The three blood isolates belonged to clade II. No specific association between accessory virulence/colonization features and specific phylogenetic clades could be seen, except for PcpA, which was found exclusively in clade II, and CbpC, which was found in most strains of clade I and in all strains of clade III (Fig. 7B). Isolates carrying PiaA, ZmpD, and Eng, three features found only once, belonged to clade I. The presence of CbpC correlated with specific alleles of the protein encoded by the neighboring gene, CbpJ (Fig. 7B and Fig. S3B). Strains which carried variant I of CbpJ were exclusively found in clade II and in all cases were devoid of CbpC. Interestingly, the major clades II and III were characterized by distinct alleles of the histidine kinase HK13 (BlpH) and of BlpC, the peptide pheromone which controls the expression of bacteriocins in *S. pneumoniae* (Fig. 7B and Fig. S3C). Four variants of BlpH, which did not specifically cluster with a specific clade, were found to be similar to BlpH-I in boxes 1 and 2, which are important for interaction with BlpC (29). As expected, BlpH variants were almost strictly associated with specific variants of BlpC, BlpC*Spp*1.1 and BlpC*Spp*2. The latter is identical to BlpC 6A (29), while the former differs from BlpC R6 by one amino acid in the leader peptide sequence (Fig. S3D). Two strains carried other BlpC alleles, BlpC*Spp*1.2, which is identical to BlpC R6, and BlpC*Spp*3, which is unique. Unlike the case for BlpC, most strains had the same CSP phenotype. Besides CSP6.1 and CSP6.3, which have previously been described in *S. pseudopneumoniae* (30), 2 new alleles of ComC were found, CSP6.4 and CSP10 (Fig. 7B and Fig. S3E).

Additionally, we investigated the presence of antibiotic resistance. Resistance to erythromycin and tetracycline were the most common among our LRTI isolates (Table 3). More than half of the *S. pseudopneumoniae* genomes ($n = 24$) harbored genes encoding resistance to tetracycline [*tet(M)*], 14- and 15-membered macrolides [*mef(E)* and *msr(D)*], and/or macrolides, lincosamides, and streptogramin B (MLS_B antibiotics) [*erm(B)*]. *mef(E)* and *msr(D)* were encoded by Mega-2 elements (macrolide efflux genetic assembly) integrated within the coding sequence of a DNA-3-methyladenine glycosylase homolog to SP_RS00900 of *S. pneumoniae* TIGR4 (Fig. S4A). Integration of Mega-2 in this site has been previously reported in *S. pneumoniae* (31, 32). Nine of the 11 strains carrying Mega-2 belonged to a subset of clade III, and the presence of this element was almost strictly associated with the absence of a plasmid (Fig. 7B). *tet(M)* and *erm(B)* genes were found within the Tn916-like integrating conjugative elements (ICEs) Tn5251 (33) and Tn3872 (34) (Fig. S4B). Tn5251 and Tn3872 ICEs were integrated in 7 different integration sites (Fig. 7B and Fig. S4C). Four of the integration sites were unique, while the other 3 were shared by two or more strains. ICE integration sites were mostly shared by closely related strains. One strain carried the *tet(O)* gene, which also encodes tetracycline resistance, and 2 other strains carried an aminoglycoside-3'-phosphotransferase [*aph(3')*-Ia] gene.

Phenotypic resistance to penicillin and co-trimoxazole (SXT) had high prevalence in other reports (4, 6–8) and were available for many of the NCBI genomes. We found that these types of resistance were also strongly associated with clade III. Taken together, 19/20 strains (95.2%) of clade III carried at least one genetic element encoding an AMR determinant or were shown to be resistant to at least one antibiotic (Fig. 7B). In contrast, a relatively small percentage of strains belonging to clade II (31.6%) were associated with AMR.

DISCUSSION

Correct identification of SMG isolates remains a challenge to this day and impairs our understanding of their epidemiology and contribution to human disease. The high

FIG 7 Legend (Continued)

pheromones, genotypic and phenotypic antibiotic resistance, and plasmids. Description of the colors is indicated in the key. Supporting information on allelic variants can be found in Fig. S3B to E. Roman numerals in the ICE column refer to integration sites (Fig. S4). ICE, Mega-2, and other types of resistance refer to genotypic resistance; penicillin (Pen) and co-trimoxazole (SXT) refer to phenotypic resistance (Table 3 and references 5, 16, and 58–60). NP, nasopharynx; ND, pseudogenes/truncated; NA, data not available.

TABLE 3 Antibiotic susceptibility profile of LRTI *S. pseudopneumoniae* isolates^a

Strain	MIC ($\mu\text{g/ml}$) for:					
	SXT	Penicillin G	Erythromycin	Clindamycin	Tetracycline	Levofloxacin
BHN868	0.064	0.0004	2	0.5	32	0.5
BHN871	1.5	0.047	1.5	0.5	64	1
BHN877	4	0.064	4	0.5	32	1
BHN879	0.047	0.0004	0.19	0.125	0.5	0.5
BHN880	0.25	0.012	0.19	0.125	0.25	1
BHN881	2	0.125	6	0.5	48	0.75
BHN885	0.032	0.0008	3	0.094	32	1
BHN886	0.032	0.0008	4	0.094	0.094	0.5
BHN890	0.064	0.0008	0.25	0.125	0.25	1
BHN891	0.064	0.0008	0.125	0.094	32	0.75
BHN892	0.064	0.0008	0.125	0.064	0.25	0.5
BHN893	0.047	0.0008	0.125	0.125	0.5	0.75
BHN912	0.047	0.0008	0.125	0.094	0.5	1
BHN913	0.064	0.004	2	0.125	32	1.5
BHN914	0.064	0.016	0.125	0.125	0.5	0.75
BHN915	0.064	0.016	0.19	0.125	0.5	0.75
BHN916	0.094	0.0008	0.125	0.125	0.38	0.75
BHN918	0.064	0.016	0.25	0.125	0.5	0.75
BHN919	0.125	0.008	4	0.125	32	0.75
BHN920	0.064	0.006	0.125	0.125	0.38	1
BHN922	0.032	0.008	0.19	0.125	0.38	1
BHN1333	0.064	0.008	0.125	0.125	0.25	0.75
BHN1334	0.094	0.008	0.125	0.125	0.19	1.5
BHN1335	0.19	0.008	0.19	0.19	0.38	0.75
BHN1336	0.094	0.008	0.125	0.094	0.19	0.5
BHN1337	0.047	0.008	0.094	0.094	0.38	1.5
BHN1338	0.023	0.008	0.125	0.125	0.38	2
BHN1339	0.064	0.008	0.19	0.125	0.5	1
BHN1340	0.38	0.016	8	0.5	0.38	1.5
Nonsusceptible ^b	10.34 (3)	0 (0)	31.03 (9)	17.24 (5)	27.59 (8)	0 (0)

^aMICs were interpreted using the CLSI guidelines for viridans streptococci (44), except for SXT, which was interpreted according to EUCAST guidelines (45) for non-meningitis *S. pneumoniae* isolates.

^bValues are shown as % (n).

genetic relatedness of *S. pneumoniae* and *S. pseudopneumoniae*, exemplified by our result that they share 50% of their pangenomes, is likely due to their ability to acquire genetic material through natural transformation. Even though *S. pseudopneumoniae* causes milder infections than *S. pneumoniae* and is associated with underlying diseases (5, 8), it has been isolated from normally sterile body sites (5, 7). A causative agent has not been identified in a significant percentage of LRTI ($\approx 40\%$) and community-acquired pneumonia cases, both in the community and hospital settings (35), and it is possible that a fraction of these cases are due to potential pathogens such as *S. pseudopneumoniae*, which might be discarded as commensals and for which reliable identification methods are lacking. Hence, some LRTI isolates included in this study could only be classified using WGS and phylogenetic analyses. By performing a thorough comparative genomic analysis, we identified for the first time a genetic marker that is entirely specific for *S. pseudopneumoniae*.

Our results show that pneumococcal genes known to be differentially regulated under infection- and colonization-relevant conditions are widespread in *S. pseudopneumoniae* and that only a surprisingly small percentage (5.6%) of them are absent from *S. pneumoniae*. We further report the first *S. pseudopneumoniae* isolate encoding and expressing a capsule. The lack of transposase genes on either side of the capsule locus and its higher similarity to the capsular locus of an *S. mitis* strain suggest it was acquired from *S. mitis* rather than from a pneumococcal strain. Interestingly, a high prevalence of pneumococcal serotype 5 antigens in urine samples in the absence of culture confirmation has been reported in one study of community-acquired pneumonia cases in the United States (36). The possibility for *S. pseudopneumoniae* and other SMG species (37) to express the pneumococcal serotype 5 capsule should be taken into

consideration when interpreting results based solely on serotype-specific assays. Further studies are needed to understand the role of the capsule in *S. pseudopneumoniae* and to evaluate the prevalence of encapsulated isolates in larger clinical collections.

The multiple pneumococcal virulence and colonization factors found in the core genome of *S. pseudopneumoniae* confirm earlier observations (3, 16). The presence of some crucial virulence factors, such as pneumolysin (Ply), could mark an important difference between *S. pseudopneumoniae* and the more commensal *S. mitis*, especially in light of our recent study in which Ply was shown to drive internalization of pneumococci within nonlysosomal compartments of immune cells commonly found in the lungs (38). In addition, the presence of large numbers of surface-exposed proteins could provide an advantage for adhesion and colonization, as was described for nonencapsulated pneumococci (17, 18). Surface-exposed proteins are important players for successful pneumococcal colonization, which constitutes the first step of pneumococcal disease, and display a wide variety of functions, from virulence to fitness and antibiotic tolerance (39, 40). In this scenario, the lack of a capsule might avoid restricting the ability of surface-exposed proteins to interact with their ligands on host cells (24). The large number of two-component signaling systems in *S. pseudopneumoniae* suggests that it is equipped to fine-tune its response to different environmental cues.

Our observations reveal a composite scenario of genetic elements in *S. pseudopneumoniae*, where prophages are abundant and plasmids and AMR-encoding ICEs are found in a large number of isolates. The fact that the core genome phylogeny delineates clades that harbor different genetic elements suggests that small differences in their core genome play a role in the maintenance or exclusion of these elements. Our findings suggest multiple acquisition events and subsequent clonal expansion of Tn916-like ICEs in *S. pseudopneumoniae* or intrachromosomal mobilization. Most of the strains carrying a Mega-2 element were found in a subset of the same clade, suggesting that its presence is mainly driven through clonal expansion, as was suggested for *S. pneumoniae* (31, 32). Besides genetic determinants of AMR, phenotypic resistance also showed a tight association with a specific lineage. In *S. pneumoniae*, longer durations of carriage are associated with increased prevalence of resistance (41). No specific known virulence factor except for PcpA could be specifically associated with a given *S. pseudopneumoniae* clade. However, the 4 PcpA⁺ strains as well as the 3 septicemia isolates belong to the same phylogenetic clade, which is also characterized by fewer AMR determinants. It will be interesting in the future to evaluate the relative virulence of strains belonging to different clades.

In conclusion, our single specific molecular marker for identifying *S. pseudopneumoniae* from other SMG species will be a useful resource for better understanding the clinical importance of this species. Moreover, our results reveal the impressive amount of surface-exposed proteins encoded by some strains and shed light on the overall distribution in *S. pseudopneumoniae* of genes known to be important during pneumococcal invasive disease and colonization.

MATERIALS AND METHODS

Bacterial isolates and molecular typing. Thirty-two alpha-hemolytic strains isolated from sputum or nasopharyngeal swabs of patients with lower-respiratory-tract infections collected during the GRACE study (19) and presenting atypical results in traditional biochemical tests to identify *S. pneumoniae* were included in this study. Isolates were tested for optochin susceptibility and bile solubility (7, 42) by PCR for pneumococcal markers (*lytA*, *cpsA*, *spn_9802*, 16SrRNA) and by RFLP for pneumococcus-specific signatures (*lytA*, *ply-*mly**) (7). BHN880 was serotyped by gel diffusion (43). MICs to all antibiotics were determined using Etests (bioMérieux) and interpreted using the Clinical and Laboratory Standards Institute (CLSI) guidelines for viridans streptococci (44), except for SXT, which was interpreted using the European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoints for non-meningitis *S. pneumoniae* (45).

Whole-genome sequencing, assembly, and phylogenetic analysis. Chromosomal DNA was prepared from overnight cultures on blood agar plates using the genomic DNA buffer set and Genomic-tip 100/G (Qiagen) by following the manufacturer's instructions. Long DNA insert sizes were used, and libraries were prepared with the Illumina TruSeq HT DNA sample preparation kit. Two-hundred-fifty-bp-long paired-end reads were generated. Adapters were removed from the demultiplexed reads, and reads

were quality trimmed using Trimmomatic (46). The 24 genomes were assembled *de novo* with SPAdes (v3.1.1) (47), annotated with PROKKA (v1.11) (48, 49), and deposited in NCBI ([SOQB00000000](https://doi.org/10.1093/mbio/mbaa000) to [SOQV00000000](https://doi.org/10.1093/mbio/mbaa000000) [see Data Set S1 in the supplemental material]). Assembly metrics were calculated with QUAST 4.5.4 (49). kSNP 3.1 (50) was used to generate the SNP-based phylogenetic tree of SMG genomes. The optimum k-mer value of 19, estimated from Kchooser, and a consensus parsimony tree based on all the SNPs generated by kSNP were used (50). The phylogenetic tree was visualized in MEGA7 (51).

Pangenome analysis, construction of SPPN species tree, and identification of virulence factors.

The pangenome analysis of orthologous gene clusters, species trees, and their respective gene trees were analyzed using panX (52) for the 39 completed *S. pneumoniae* genomes (pan:SPN), 44 *S. pseudopneumoniae* genomes (pan:SPPN), and both species (pan:SPPN-SPN) with default cutoff values. pan:SPPN analysis resulted in 885 core genes (strict core; 100% present in all strains), and the core genome tree/species tree for the SPPN species was constructed based on the core genome SNPs, including only single-copy core genes ($n = 793$). Using pan:SPPN-SPN, all COGs were queried for *S. pneumoniae* locus tags corresponding to 356 IPD/ECC genes (24) and 92 well-studied pneumococcal genes (21) listed in Data Set S2 and Table S3. Proteins listed in Table S3 were analyzed using a 70% length cutoff to score proteins as present; conservation of synteny was confirmed for all proteins. Genetic loci of proteins scored as absent were manually checked for contig breaks and pseudogenes.

Molecular markers and PCR assay. Thirty COGs unique to the 44 *S. pseudopneumoniae* genomes and absent from the 39 *S. pneumoniae* genomes (Table S1) were filtered from the pangenome analysis (pan:SPPN_SPN) and subjected to BLAST searches against all NCBI genomes, which included 8,358 *S. pneumoniae* genomes. The 44 nucleotide sequences of the two unique ORFs (SPPN_RS10375 and SPPN_RS06420) were aligned using the ClustalW algorithm in Geneious, version 10.1.3 (<https://www.geneious.com>), with default parameters (gap open cost, 15; gap extend cost, 6.66). The upstream (70 bp) and downstream (329 bp) intergenic regions of SPPN_RS10375 were included. Primers SPPN_RS10375F (5'-CTAATTGCTACTGCTATTTCCGGTG-3') and SPPN_RS10375R (5'-CTGATACCTGCAACAAAATCGAAG-3') were designed in regions of 100% identity. PCR was performed using Phusion flash high-fidelity PCR master mix (ThermoFisher) by following the manufacturer's instructions with an annealing temperature of 50°C. One μ l of lysate, prepared by resuspending 2 to 3 isolated colonies in 100 μ l Tris-EDTA containing 0.1% Triton and incubating at 98°C for 5 min, was used as the template. PCR products were run on a 1.2% agarose gel stained with GelRed (Biotium).

In silico identification of new putative virulence features. A database was built using the concatenated sequence of all proteins from the 44 *S. pseudopneumoniae* genomes and was queried for the conserved choline-binding domain COG5263 and the peptidase_M26 domain pfam07580/cl06563 to identify novel choline-binding proteins and ZMPs, respectively, using the NCBI Batch CD-Search tool (53). Novel two-component signal transduction systems (TCSs) were identified by finding proteins containing the HATPase domain of histidine kinase (cd00075/smart00387/pfam02518) that were immediately preceded or followed by a DNA-binding regulator possessing the signal-receiver domain cd00156.

Analysis of capsular loci. Homologues of *cpsA* and *wzg* were searched for in pan:SPPN_SPN using gene family SP_RS01690. The locus was subsequently checked manually for the presence of the complete locus [BHN880_01411 to BHN880_01431]. The retrieved *cps* locus was subjected to a BLASTn search to identify the closest homologs. Pairwise alignment with the *S. pneumoniae* Ambrose serotype 5 locus ([CR931637.1](https://doi.org/10.1093/mbio/mbaa000000)) (54) and *S. mitis* 21/39 ([AYRR01000010.1](https://doi.org/10.1093/mbio/mbaa000000)) *cps* locus was performed using Easyfig (55).

Hemolysis assay. Bacteria were grown overnight on blood agar plates at 37°C in 5% CO₂. *S. pneumoniae* strains were grown into C+Y medium until exponential phase (optical density at 620 nm [OD₆₂₀] of 0.4). *S. pseudopneumoniae* strains were grown in C+Y medium to an OD₆₂₀ of 0.3 and then inoculated into a secondary culture, which was grown to an OD₆₂₀ of 0.25. Dilutions were made to obtain the desired concentration of bacterial cells, and viable counts were performed to retrospectively confirm bacterial numbers. Blood from healthy human donors (obtained from Karolinska University Hospital) was diluted 1:100 in phosphate-buffered saline–0.5 mM dithiothreitol, mixed 1:1 with 2-fold serial dilutions of *S. pneumoniae* or *S. pseudopneumoniae* cultures in 96-well plates, and incubated at 37°C for 1 h. After 50 min of incubation, 0.1% Triton X-100 was added to the positive-control wells. Cells were spun down at 400 \times g for 15 min, and the absorbance of the supernatants was measured at 540 nm in a microplate reader. Percentage of lysis compared to the positive control was calculated. All strains were tested in triplicate.

In silico identification of AMR determinants, plasmids, and phages. The 44 genomes were screened in Resfinder 3.0 (56) for the acquired AMR genes (90% identity threshold, minimum length of 60%). Chromosomal genes flanking Tn916-like ICEs were defined by using BLASTn to retrieve the loci in strain IS7493 ([NC_015875.1](https://doi.org/10.1093/mbio/mbaa000000)) of the genes located immediately upstream of the integrase and immediately downstream of *orf24* of Tn5251 ([FJ711160.1](https://doi.org/10.1093/mbio/mbaa000000)). Genome assemblies were queried for genes associated with known *S. pneumoniae* and *S. mitis* phages (SPH_0026, IPP61_00001, SPH_0070, SP670_2134, SP670_0091, SM1p01, SPPN_RS05280, and HMPREF1112_1362) and the *S. pseudopneumoniae* plasmid pDRPIS7493 ([NC_015876.1](https://doi.org/10.1093/mbio/mbaa000000)). Phage sequences were manually analyzed and deemed full length if they started with an integrase gene, ended with a lytic amidase, and were ≥ 30 kb in length.

Data availability. All 21 *S. pseudopneumoniae* sequenced genomes have been deposited in GenBank under BioProject code [PRJNA528011](https://doi.org/10.1093/mbio/mbaa000000), BioSample numbers [SAMN11166137](https://doi.org/10.1093/mbio/mbaa000000) to [SAMN11166157](https://doi.org/10.1093/mbio/mbaa000000), and accession numbers [SOQV00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQU00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQT00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQS00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQR00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQQ00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQP00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQO00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQN00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQM00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQL00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQK00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQJ00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQI00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQH00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQG00000000](https://doi.org/10.1093/mbio/mbaa000000), [SOQF00000000](https://doi.org/10.1093/mbio/mbaa000000),

SOQE00000000, SOQD00000000, SOQC00000000, and SOQB00000000. The accession numbers of the other genomes used in this study are listed in Data Set S1.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.01286-19>.

FIG S1, PDF file, 0.9 MB.

FIG S2, EPS file, 1.4 MB.

FIG S3, EPS file, 2.3 MB.

FIG S4, EPS file, 1 MB.

DATA SET S1, XLSX file, 0.02 MB.

DATA SET S2, XLSX file, 0.03 MB.

TABLE S1, PDF file, 0.04 MB.

TABLE S2, PDF file, 0.03 MB.

TABLE S3, PDF file, 0.04 MB.

ACKNOWLEDGMENTS

LRTI samples were collected as part of the GRACE (genomics to combat resistance against antibiotics in CA-LRTI in Europe) project. We thank the general practitioners, the GRACE study team, and the patients for taking part in this study. We thank the European Commission for the financial support of the GRACE project. We also thank Ingrid Andersson, Christina Johansson, Gunnel Möllerberg, Eva Morfeldt, and Jessica Darenberg at the Swedish Institute for Infectious Disease Control for excellent technical assistance. We acknowledge support from the National Genomics Infrastructure in Stockholm, funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, Stockholm City Council, the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. This work was partially supported by ONEIDA project (LISBOA-01-0145-FEDER-016417), cofunded by FEEI (Fundos Europeus Estruturais e de Investimento), from Programa Operacional Regional Lisboa 2020, and by national funds from Fundação para a Ciência e a Tecnologia, Portugal.

REFERENCES

- Arbique JC, Poyart C, Trieu-Cuot P, Quesne G, Carvalho MDGS, Steigerwalt AG, Morey RE, Jackson D, Davidson RJ, Facklam RR. 2004. Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. *J Clin Microbiol* 42:4686–4696. <https://doi.org/10.1128/JCM.42.10.4686-4696.2004>.
- Zheng W, Tan TK, Paterson IC, Mutha NV, Siow CC, Tan SY, Old LA, Jakubovics NS, Choo SW. 2016. StreptoBase: an oral *Streptococcus mitis* group genomic resource and analysis platform. *PLoS One* 11:e0151908. <https://doi.org/10.1371/journal.pone.0151908>.
- Leegaard TM, Bootsma HJ, Caugant DA, Eleveld MJ, Mannsaker T, Froholm LO, Gaustad P, Hoiby EA, Hermans PW. 2010. Phenotypic and genomic characterization of pneumococcus-like streptococci isolated from HIV-seropositive patients. *Microbiology* 156:838–848. <https://doi.org/10.1099/mic.0.035345-0>.
- Keith ER, Podmore RG, Anderson TP, Murdoch DR. 2006. Characteristics of *Streptococcus pseudopneumoniae* isolated from purulent sputum samples. *J Clin Microbiol* 44:923–927. <https://doi.org/10.1128/JCM.44.3.923-927.2006>.
- Fuursted K, Littauer PJ, Greve T, Scholz CF. 2016. Septicemia with *Streptococcus pseudopneumoniae*: report of three cases with an apparent hepatic or bile duct association. *Infect Dis* 48:636–639. <https://doi.org/10.3109/23744235.2016.1157896>.
- Keith ER, Murdoch DR. 2008. Antimicrobial susceptibility profile of *Streptococcus pseudopneumoniae* isolated from sputum. *Antimicrob Agents Chemother* 52:2998. <https://doi.org/10.1128/AAC.01526-07>.
- Rolo D, S Simões A, Domenech A, Fenoll A, Liñares J, de Lencastre H, Ardanuy C, Sá-Leão R. 2013. Disease isolates of *Streptococcus pseudopneumoniae* and non-typeable *S. pneumoniae* presumptively identified as atypical *S. pneumoniae* in Spain. *PLoS One* 8:e57047. <https://doi.org/10.1371/journal.pone.0057047>.
- Laurens C, Michon AL, Marchandin H, Bayette J, Didelot MN, Jean-Pierre H. 2012. Clinical and antimicrobial susceptibility data of 140 *Streptococcus pseudopneumoniae* isolates in France. *Antimicrob Agents Chemother* 56:4504–4507. <https://doi.org/10.1128/AAC.06374-11>.
- Simoes AS, Sa-Leao R, Eleveld MJ, Tavares DA, Carrico JA, Bootsma HJ, Hermans PW. 2010. Highly penicillin-resistant multidrug-resistant pneumococcus-like strains colonizing children in Oeiras, Portugal: genomic characteristics and implications for surveillance. *J Clin Microbiol* 48:238–246. <https://doi.org/10.1128/JCM.01313-09>.
- Harf-Monteil C, Granello C, Le Brun C, Monteil H, Riegel P. 2006. Incidence and pathogenic effect of *Streptococcus pseudopneumoniae*. *J Clin Microbiol* 44:2240–2241. <https://doi.org/10.1128/JCM.02643-05>.
- Jensen A, Scholz CF, Kilian M. 2016. Re-evaluation of the taxonomy of the Mitis group of the genus *Streptococcus* based on whole genome phylogenetic analyses, and proposed reclassification of *Streptococcus dentisani* as *Streptococcus oralis* subsp. *dentisani* comb. nov., *Streptococcus tigurinus* as *Streptococcus oralis* subsp. *tigurinus* comb. nov., and *Streptococcus oligofermentans* as a later synonym of *Streptococcus cristatus*. *Int J Syst Evol Microbiol* 66:4803–4820. <https://doi.org/10.1099/ijsem.0.001433>.
- Croxen MA, Lee TD, Azana R, Hoang LM. 2018. Use of genomics to design a diagnostic assay to discriminate between *Streptococcus pneumoniae* and *Streptococcus pseudopneumoniae*. *Microb Genom* 4:e000175. <https://doi.org/10.1099/mgen.0.000175>.
- Wessels E, Schelfaut JJ, Bernards AT, Claas EC. 2012. Evaluation of several biochemical and molecular techniques for identification of *Streptococcus pneumoniae* and *Streptococcus pseudopneumoniae* and their detection in

- respiratory samples. *J Clin Microbiol* 50:1171–1177. <https://doi.org/10.1128/JCM.06609-11>.
14. Johnston C, Hinds J, Smith A, van der Linden M, Van Eldere J, Mitchell TJ. 2010. Detection of large numbers of pneumococcal virulence genes in streptococci of the mitis group. *J Clin Microbiol* 48:2762–2769. <https://doi.org/10.1128/JCM.01746-09>.
 15. Mitchell AM, Mitchell TJ. 2010. *Streptococcus pneumoniae*: virulence factors and variation. *Clin Microbiol Infect* 16:411–418. <https://doi.org/10.1111/j.1469-0691.2010.03183.x>.
 16. Shahinas D, Thornton CS, Tamber GS, Arya G, Wong A, Jamieson FB, Ma JH, Alexander DC, Low DE, Pillai DR. 2013. Comparative genomic analyses of *Streptococcus pseudopneumoniae* provide insight into virulence and commensalism dynamics. *PLoS One* 8:e65670. <https://doi.org/10.1371/journal.pone.0065670>.
 17. Keller LE, Jones CV, Thornton JA, Sanders ME, Swiatlo E, Nahm MH, Park IH, McDaniel LS. 2013. PspK of *Streptococcus pneumoniae* increases adherence to epithelial cells and enhances nasopharyngeal colonization. *Infect Immun* 81:173–181. <https://doi.org/10.1128/IAI.00755-12>.
 18. Park IH, Kim KH, Andrade AL, Briles DE, McDaniel LS, Nahm MH. 2012. Nontypeable pneumococci can be divided into multiple *cps* types, including one type expressing the novel gene *pspK*. *mBio* 3:e00035-12. <https://doi.org/10.1128/mBio.00035-12>.
 19. Ieven M, Coenen S, Loens K, Lammens C, Coenjaerts F, Vanderstraeten A, Henriques-Normark B, Crook D, Huygen K, Butler CC, Verheij TJM, Little P, Zlateva K, van Loon A, Claas ECJ, Goossens H, GRACE Consortium. 2018. Aetiology of lower respiratory tract infection in adults in primary care: a prospective study in 11 European countries. *Clin Microbiol Infect* 24:1158–1163. <https://doi.org/10.1016/j.cmi.2018.02.004>.
 20. Simões AS, Tavares DA, Rolo D, Ardanuy C, Goossens H, Henriques-Normark B, Linares J, de Lencastre H, Sá-Leão R. 2016. IyA-based identification methods can misidentify *Streptococcus pneumoniae*. *Diagn Microbiol Infect Dis* 85:141–148. <https://doi.org/10.1016/j.diagmicrobio.2016.03.018>.
 21. Gamez G, Castro A, Gomez-Mejia A, Gallego M, Bedoya A, Camargo M, Hammerschmidt S. 2018. The variome of pneumococcal virulence factors and regulators. *BMC Genomics* 19:10. <https://doi.org/10.1186/s12864-017-4376-0>.
 22. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Massignani V. 2010. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol* 11:R107. <https://doi.org/10.1186/gb-2010-11-10-r107>.
 23. Iannelli F, Oggioni MR, Pozzi G. 2002. Allelic variation in the highly polymorphic locus *pspC* of *Streptococcus pneumoniae*. *Gene* 284:63–71. [https://doi.org/10.1016/S0378-1119\(01\)00896-4](https://doi.org/10.1016/S0378-1119(01)00896-4).
 24. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI. 2004. Microarray analysis of pneumococcal gene expression during invasive disease. *Infect Immun* 72:5582–5596. <https://doi.org/10.1128/IAI.72.10.5582-5596.2004>.
 25. Hathaway LJ, Stutzmann Meier P, Bättig P, Aebi S, Mühlemann K. 2004. A homologue of *aliB* is found in the capsule region of nonencapsulated *Streptococcus pneumoniae*. *J Bacteriol* 186:3721–3729. <https://doi.org/10.1128/JB.186.12.3721-3729.2004>.
 26. Gilbert JV, Plaut AG, Wright A. 1991. Analysis of the immunoglobulin A protease gene of *Streptococcus sanguis*. *Infect Immun* 59:7–17.
 27. Zhou M, Yang Q, Kudinha T, Zhang L, Xiao M, Kong F, Zhao Y, Xu YC. 2016. Using matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) complemented with selected 16S rRNA and *gyrB* genes sequencing to practically identify clinical important viridans group streptococci (VGS). *Front Microbiol* 7:1328. <https://doi.org/10.3389/fmicb.2016.01328>.
 28. Romero P, Croucher NJ, Hiller NL, Hu FZ, Ehrlich GD, Bentley SD, Garcia E, Mitchell TJ. 2009. Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. *J Bacteriol* 191:4854–4862. <https://doi.org/10.1128/JB.01272-08>.
 29. Pinchas MD, LaCross NC, Dawid S. 2015. An electrostatic interaction between BlpC and BlpH dictates pheromone specificity in the control of bacteriocin production and immunity in *Streptococcus pneumoniae*. *J Bacteriol* 197:1236–1248. <https://doi.org/10.1128/JB.02432-14>.
 30. Leung MH, Ling CL, Ciesielczuk H, Lockwood J, Thurston S, Charalambous BM, Gillespie SH. 2012. *Streptococcus pseudopneumoniae* identification by phenotype: a method to assist understanding of a potentially emerging or overlooked pathogen. *J Clin Microbiol* 50:1684–1690. <https://doi.org/10.1128/JCM.00131-12>.
 31. Chancey ST, Agrawal S, Schroeder MR, Farley MM, Tettelin H, Stephens DS. 2015. Composite mobile genetic elements disseminating macrolide resistance in *Streptococcus pneumoniae*. *Front Microbiol* 6:26. <https://doi.org/10.3389/fmicb.2015.00026>.
 32. Gay K, Stephens DS. 2001. Structure and dissemination of a chromosomal insertion element encoding macrolide efflux in *Streptococcus pneumoniae*. *J Infect Dis* 184:56–65. <https://doi.org/10.1086/321001>.
 33. Santoro F, Oggioni MR, Pozzi G, Iannelli F. 2010. Nucleotide sequence and functional analysis of the *tet(M)*-carrying conjugative transposon Tn5251 of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* 308:150–158. <https://doi.org/10.1111/j.1574-6968.2010.02002.x>.
 34. McDougal LK, Tenover FC, Lee LN, Rasheed JK, Patterson JE, Jorgensen JH, LeBlanc DJ. 1998. Detection of Tn917-like sequences within a Tn916-like conjugative transposon (Tn3872) in erythromycin-resistant isolates of *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 42:2312–2318. <https://doi.org/10.1128/AAC.42.9.2312>.
 35. Woodhead M, Blasi F, Ewig S, Garau J, Huchon G, Ieven M, Ortvist A, Schaberg T, Torres A, van der Heijden G, Read R, Verheij TJ, Joint Taskforce of the European Respiratory Society and European Society for Clinical Microbiology and Infectious Diseases. 2011. Guidelines for the management of adult lower respiratory tract infections—full version. *Clin Microbiol Infect* 17(Suppl 6):E1–E59. <https://doi.org/10.1111/j.1469-0691.2011.03672.x>.
 36. Sherwin RL, Gray S, Alexander R, McGovern PC, Graepel J, Pride MW, Purdy J, Paradiso P, File TM, Jr. 2013. Distribution of 13-valent pneumococcal conjugate vaccine *Streptococcus pneumoniae* serotypes in US adults aged ≥50 years with community-acquired pneumonia. *J Infect Dis* 208:1813–1820. <https://doi.org/10.1093/infdis/jit506>.
 37. Pimenta F, Gertz RE, Jr, Park SH, Kim E, Moura I, Milucky J, Roupael N, Farley MM, Harrison LH, Bennett NM, Bigogo G, Feikin DR, Breiman R, Lessa FC, Whitney CG, Rajam G, Schiffer J, da Gloria Carvalho M, Beall B. 2018. *Streptococcus infantis*, *Streptococcus mitis*, and *Streptococcus oralis* strains with highly similar *cps5* loci and antigenic relatedness to serotype 5 pneumococci. *Front Microbiol* 9:3199.
 38. Subramanian K, Neill DR, Malak HA, Spelmink L, Khandaker S, Dalla Libera Marchiori G, Dearing E, Kirby A, Yang M, Achour A, Nilvebrant J, Nygren PA, Plant L, Kadioglu A, Henriques-Normark B. 2019. Pneumolysin binds to the mannose receptor C type 1 (MRC-1) leading to anti-inflammatory responses and enhanced pneumococcal survival. *Nat Microbiol* 4:62–70. <https://doi.org/10.1038/s41564-018-0280-x>.
 39. Bergmann S, Hammerschmidt S. 2006. Versatility of pneumococcal surface proteins. *Microbiology* 152:295–303. <https://doi.org/10.1099/mic.0.28610-0>.
 40. Novak R, Braun JS, Charpentier E, Tuomanen E. 1998. Penicillin tolerance genes of *Streptococcus pneumoniae*: the ABC-type manganese permease complex Psa. *Mol Microbiol* 29:1285–1296. <https://doi.org/10.1046/j.1365-2958.1998.01016.x>.
 41. Lehtinen S, Blanquart F, Croucher NJ, Turner P, Lipsitch M, Fraser C. 2017. Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. *Proc Natl Acad Sci U S A* 114:1075–1080. <https://doi.org/10.1073/pnas.1617849114>.
 42. Ruoff KL, Whitley RA, Beighton D. 1999. *Streptococcus*, p 283–296. In Murray PR, Baron EJ, Pfaller MA, Tenover FC, Tenover RH (ed), *Manual of clinical microbiology*, 7th ed. ASM Press, Washington, D.C.
 43. Henriques Normark B, Christensson B, Sandgren A, Noren B, Sylvan S, Burman LG, Olsson-Liljequist B. 2003. Clonal analysis of *Streptococcus pneumoniae* nonsusceptible to penicillin at day-care centers with index cases, in a region with low incidence of resistance: emergence of an invasive type 35B clone among carriers. *Microb Drug Resist* 9:337–344. <https://doi.org/10.1089/107662903322762761>.
 44. CLSI. 2018. Performance standards for antimicrobial susceptibility testing, 28th ed. CLSI supplement M100. Clinical and Laboratory Standards Institute, Wayne, PA.
 45. EUCAST. 2018. Breakpoint tables for interpretation of MICs and zone diameters, version 8.0. http://www.eucast.org/clinical_breakpoints/.
 46. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 47. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new

- genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
48. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 49. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
 50. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31:2877–2878. <https://doi.org/10.1093/bioinformatics/btv271>.
 51. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
 52. Ding W, Baumdicker F, Neher RA. 2018. panX: pan-genome analysis and exploration. *Nucleic Acids Res* 46:e5. <https://doi.org/10.1093/nar/gkx977>.
 53. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200–D203. <https://doi.org/10.1093/nar/gkw1129>.
 54. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kalltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG. 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2:e31. <https://doi.org/10.1371/journal.pgen.0020031>.
 55. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27:1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>.
 56. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>.
 57. Letunic I, Bork P. 2018. Twenty years of the SMART protein domain annotation resource. *Nucleic Acids Res* 46:D493–D496. <https://doi.org/10.1093/nar/gkx922>.
 58. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen DM, Mather AE, Page AJ, Salter SJ, Harris D, Nosten F, Goldblatt D, Corander J, Parkhill J, Turner P, Bentley SD. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* 46:305–309. <https://doi.org/10.1038/ng.2895>.
 59. Ikryannikova LN, Ischenko DS, Lominadze GG, Kanygina AV, Karpova IY, Kostryukova ES, Mayansky NA, Skvortsov VS, Ilina EN, Govorun VM. 2016. The mystery of the fourth clone: comparative genomic analysis of four non-typeable *Streptococcus pneumoniae* strains with different susceptibilities to optochin. *Eur J Clin Microbiol Infect Dis* 35:119–130. <https://doi.org/10.1007/s10096-015-2516-5>.
 60. Alvarado M, Martin-Galiano AJ, Ferrandiz MJ, Zaballos A, de la Campa AG. 2017. Upregulation of the PatAB transporter confers fluoroquinolone resistance to *Streptococcus pseudopneumoniae*. *Front Microbiol* 8:2074. <https://doi.org/10.3389/fmicb.2017.02074>.
 61. Freeman ZN, Dorus S, Waterfield NR. 2013. The KdpD/KdpE two-component system: integrating K(+) homeostasis and virulence. *PLoS Pathog* 9:e1003201. <https://doi.org/10.1371/journal.ppat.1003201>.
 62. Thevenard B, Rasoava N, Fourcassie P, Monnet V, Boyaval P, Rul F. 2011. Characterization of *Streptococcus thermophilus* two-component systems: in silico analysis, functional analysis and expression of response regulator genes in pure or mixed culture with its yogurt partner, *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Int J Food Microbiol* 151:171–181. <https://doi.org/10.1016/j.ijfoodmicro.2011.08.019>.