



Gene Acquisition by a Distinct Phyletic Group within *Streptococcus pneumoniae* Promotes Adhesion to the Ocular Epithelium

Irena Antic,^a Kimberly M. Brothers,^b Maureen Stolzer,^a Han Lai,^a Evan Powell,^c Rory Eutsey,^a Rolando A. Cuevas,^a Xinyu Miao,^a Regis P. Kowalski,^b Robert M. Q. Shanks,^b Dannie Durand,^a N. Luisa Hiller^{a,c}

Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA^a; Department of Ophthalmology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA^b; Center of Excellence in Biofilm Research, Allegheny Health Network, Pittsburgh, Pennsylvania, USA^c

ABSTRACT *Streptococcus pneumoniae* (pneumococcus) displays broad tissue tropism and infects multiple body sites in the human host. However, infections of the conjunctiva are limited to strains within a distinct phyletic group with multilocus sequence types ST448, ST344, ST1186, ST1270, and ST2315. In this study, we sequenced the genomes of six pneumococcal strains isolated from eye infections. The conjunctivitis isolates are grouped in a distinct phyletic group together with a subset of nasopharyngeal isolates. The keratitis (infection of the cornea) and endophthalmitis (infection of the vitreous body) isolates are grouped with the remainder of pneumococcal strains. Phenotypic characterization is consistent with morphological differences associated with the distinct phyletic group. Specifically, isolates from the distinct phyletic group form aggregates in planktonic cultures and chain-like structures in biofilms grown on abiotic surfaces. To begin to investigate the association between genotype and epidemiology, we focused on a predicted surface-exposed adhesin (SspB) encoded exclusively by this distinct phyletic group. Phylogenetic analysis of the gene encoding SspB in the context of a streptococcal species tree suggests that *sspB* was acquired by lateral gene transfer from *Streptococcus suis*. Furthermore, an *sspB* deletion mutant displays decreased adherence to cultured cells from the ocular epithelium compared to the isogenic wild-type and complemented strains. Together these findings suggest that acquisition of genes from outside the species has contributed to pneumococcal tissue tropism by enhancing the ability of a subset of strains to infect the ocular epithelium causing conjunctivitis.

IMPORTANCE Changes in the gene content of pathogens can modify their ability to colonize and/or survive in different body sites in the human host. In this study, we investigate a gene acquisition event and its role in the pathogenesis of *Streptococcus pneumoniae* (pneumococcus). Our findings suggest that the gene encoding the predicted surface protein SspB has been transferred from *Streptococcus suis* (a distantly related streptococcal species) into a distinct set of pneumococcal strains. This group of strains distinguishes itself from the remainder of pneumococcal strains by extensive differences in genomic composition and by the ability to cause conjunctivitis. We find that the presence of *sspB* increases adherence of pneumococcus to the ocular epithelium. Thus, our data support the hypothesis that a subset of pneumococcal strains has gained genes from neighboring species that enhance their ability to colonize the epithelium of the eye, thus expanding into a new niche.

KEYWORDS *Streptococcus pneumoniae*, gene transfer, genomics, host-pathogen interactions, phylogenetic analysis


Received 11 May 2017 Accepted 14 September 2017 Published 25 October 2017

Citation Antic I, Brothers KM, Stolzer M, Lai H, Powell E, Eutsey R, Cuevas RA, Miao X, Kowalski RP, Shanks RMQ, Durand D, Hiller NL. 2017. Gene acquisition by a distinct phyletic group within *Streptococcus pneumoniae* promotes adhesion to the ocular epithelium. mSphere 2:e00213-17. <https://doi.org/10.1128/mSphere.00213-17>.

Editor Brandi M. Limbago, Centers for Disease Control and Prevention

Copyright © 2017 Antic 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 N. Luisa Hiller, lhiller@andrew.cmu.edu.

 Characterization of an adhesin in a distinct phyletic group on nonencapsulated pneumococci.

Streptococcal species are characterized by extensive intraspecies gene diversity that plays critical roles in tissue tropism and, consequently, disease outcomes (1–3). In group A streptococci (GAS), a significant association has been detected between gene content and the ability to colonize the skin versus the upper respiratory tract, leading to either impetigo or pharyngitis (1). In contrast, in the human pathogen *Streptococcus pneumoniae* (pneumococcus), an association between genomic background and site of infection had not been evident (4). The absence of this association is unexpected given the large number of diverse pneumococcal lineages and the variety of body sites that can be infected by this bacterium (2, 5, 6). Pneumococcus colonizes the nasopharynx, from where it can disseminate to tissues: frequently the middle ears and lungs and relatively rarely the eyes, heart, and brain (7). Pneumococcal conjunctivitis is a notable exception to the lack of association between phylogenetic background and site of infection: the genomic composition of strains causing pneumococcal conjunctivitis differs from that of strains causing disease in other tissues. This difference in gene content is so extensive that this set of strains localizes to a distinct phyletic group (8–10).

The best-characterized feature shared across conjunctivitis-associated pneumococcal strains is the absence of the genes encoding the polysaccharide capsule (8–11). The nonencapsulated strains in the distinct phyletic group are referred to as the classical nonencapsulated strains (8, 11). In the region syntenic to the capsular locus, these strains encode *aliC* and *aliD*, which are putative lipoproteins of unknown functions (9, 12–14). The capsule is the main pneumococcal virulence determinant: thus, the absence of the capsule has important clinical implications for the distinct phyletic group. Nonencapsulated strains are less virulent than encapsulated strains and are much less likely to disseminate in single-strain infections (15). Furthermore, the capsular structure is the target of the pneumococcal vaccine: thus nonencapsulated strains escape the vaccine (11).

Whereas the absence of capsule is a shared feature of all strains in the distinct phyletic group, this feature alone cannot be used as a marker. Many nonencapsulated strains are phylogenetically clustered with the majority of pneumococcal strains (16, 17). Nonencapsulated strains organized into the major pneumococcal phylogenetic branch are referred to as the sporadic nonencapsulated strains (8, 11). A subset of nonencapsulated strains has its origin in encapsulated strains that no longer encode the capsule due to mutations or deletions in the capsular locus; these are termed group I (11–13). The remainder of sporadic nonencapsulated strains, group II, carry noncapsular genes syntenic to the capsular ones (11). This locus carries either *pspK* (also known as *nspA*) or *aliD* (also known as *aliB* open reading frame 2 [ORF2]) (13). The strains carrying *aliD* may also carry *aliC* (also known as *aliB* ORF1) (11–14). Thus, the absence of capsule is a highly clinically relevant feature shared by conjunctivitis strains but not what drives the association between genomic background and ability to cause conjunctivitis.

Pneumococcal eye disease is not limited to conjunctivitis, where it infects the conjunctival epithelial layers. This pathogen can also infect the vitreous body inside the eye (endophthalmitis) and the cornea (keratitis) (18–20). It is unclear whether there are morphological and genomic features shared by all isolates that infect the human eye. In this study, we combine genomics, phylogenetics, and cell adhesion studies to gain insight into the genomes of isolates from multiple types of pneumococcal ocular infections and the tissue tropism associated with conjunctivitis isolates.

RESULTS

Sequencing of pneumococcal strains isolated from ocular infections. To compare the genomes of pneumococcal strains isolated from multiple types of eye infections, we sequenced six eye-associated strains. One isolate is from an endophthalmitis infection (strain E709), three are from keratitis infections (strains K2521, K2527, and K2557), and two are from conjunctivitis infections (strains B1598 and B1599). All isolates are deidentified clinical samples obtained at the Charles T. Campbell Ophthalmic

TABLE 1 Description of strain isolation, sequencing and assembly, and drug resistance profiles for eye-associated *S. pneumoniae* strains

Parameter	Result for strain:					
	E709	K2521	K2527	K2557	B1598	B1599
Isolation date (mo/day/yr)	5/13/13	1/28/13	2/21/13	6/19/13	10/19/12	11/3/12
Isolation site	UPMC	UPMC	UPMC	UPMC	UPMC	UPMC
Serotype ^a	23A	15B	19A	17F	NE	NE
MLST	338	3280	320	2355	2315	2315
Accession no.						
GenBank	JBOR00000000	JBOS00000000	JBOT00000000	JBOU00000000	JBOV00000000	JBOW00000000
BioProject	PRJNA235160	PRJNA235267	PRJNA235268	PRJNA235269	PRJNA235270	PRJNA235271
BioSample	SAMN02691894	SAMN02691895	SAMN02691896	SAMN02691897	SAMN02691898	SAMN02691899
CG (%)	39.6	39.6	39.8	39.8	39.6	39.6
No. of contigs	37	32	29	111	158	38
Read coverage (×)	55.22	73.71	63.13	30.39	52.07	49.71
Sequencing platform	SMRT	SMRT	SMRT	SMRT	SMRT	SMRT
Estimated genome size (bp)	2,155,121	2,184,812	2,180,290	2,051,681	2,902,565	2,422,872
Resistance	Gentamicin, amikacin	Tobramycin, polymyxin B	Gentamicin, tobramycin, polymyxin B	Gentamicin, polymyxin B, tobramycin, sulfonamide, ciprofloxacin, ^b ofloxacin ^b	Gentamicin, trimethoprim, polymyxin B, tobramycin	Gentamicin, trimethoprim, polymyxin B, tobramycin

^aSerotypes were determined based on genome sequence. NE, nonencapsulated.

^bIntermediate resistance.

Laboratory at the University of Pittsburgh Eye Center from October 2012 to May 2013. The strains were sequenced using single-molecule real-time (SMRT) technology, and the genome sequences have been deposited in GenBank (Table 1).

Pneumococcal genomes are conventionally classified using multilocus sequence typing (MLST) (which serves as a proxy for lineages), polysaccharide capsular types, and drug resistance profiles (2, 21). The six strains encompass five STs and four capsular types, as well as nonencapsulated types, and display various levels of resistance to a standard panel of antibiotics (Table 1). The conjunctivitis strains are nonencapsulated, while the keratitis and endophthalmitis strains are encapsulated. All strains displayed some degree of drug resistance: the broadest was observed for keratitis strain K2557, which is resistant to gentamicin, polymyxin B, tobramycin, and sulfonamide and has intermediate resistance to the fluoroquinolone antibiotics ciprofloxacin and ofloxacin (Table 1). In summary, strains isolated from the different ocular infections display diverse ST, serotype, and resistance profiles.

Phylogenetics and comparative genomics of pneumococcal strains isolated from ocular infections. To establish the phylogenetic relationship of these eye-associated strains, we compared the six genomes with a diverse set of 34 pneumococcal genomes from strains isolated from blood, lung, and nasopharynx (see Table S1 in Data Set S2 in the supplemental material). This highly curated set consists of genomes used for the first large-scale pneumococcal pangenome studies (2, 5), genomes from PCV-7-immunized children (22), and genomes from nonencapsulated isolates (23). Together this set reflects diversity in multilocus sequence types, serotypes, disease states, and geographical locations. We generated and aligned the core genomes of these sequences and produced a maximum likelihood phylogenetic tree (Fig. 1). In agreement with published work (8, 9, 24), the most prominent feature of the phylogenetic tree is the presence of a distinct and strongly supported branch that contains the conjunctivitis strains (Fig. 1).

A second noteworthy feature of the phylogenetic tree is the tight grouping between conjunctivitis and nasopharyngeal strains in the distinct phyletic group. Specifically, there is a short distance between the conjunctivitis strains B1599 and B1598 and the nonencapsulated nasopharyngeal carriage strain MNZ85; furthermore, there is strong

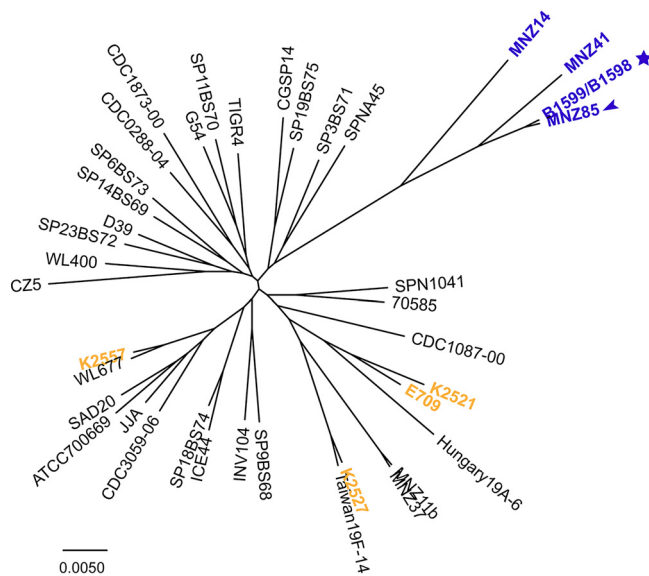


FIG 1 Phylogenetic analysis of pneumococcal strains. Maximum likelihood phylogeny of 40 pneumococcal strains generated from the core genome. Purple, distinct phyletic group; arrow, nasopharyngeal isolate; star, conjunctivitis isolates; yellow, endophthalmitis- and keratitis-associated strains. The scale bar indicates the number of substitutions per site. The numbers on the branch denote bootstrap values. The basal position of the distinct phyletic group presented in the context of other streptococcal species is illustrated in Fig. S3.

support for the branch that groups these strains together (Fig. 1). The mixture of nasopharyngeal and conjunctivitis isolates on this distinct phyletic group suggests that diversification of these strains is not a consequence of niche isolation and instead supports a model where strains migrate between the ocular epithelium and the nasopharynx. Finally, in stark contrast to the conjunctivitis strains, the endophthalmitis and keratitis isolates are not part of the distinct phyletic group (Fig. 1, yellow). The three keratitis isolates are distributed across the main group, suggesting the core genome is not associated with the ability of strains to cause keratitis. This arrangement demonstrates that belonging to the distinct phyletic group is not a feature shared by all strains that infect the eye—rather, it is a distinguishing feature of conjunctivitis. In summary, phylogenetic analysis of six pneumococcal strains isolated for three types of eye infection revealed that conjunctivitis isolates comprise a distinct phyletic group together with nonencapsulated nasopharyngeal strains, while keratitis and endophthalmitis isolates are clustered with the remainder of strains.

To identify the genes that distinguish the distinct phyletic group from other pneumococcal strains, we singled out the coding sequences (CDSs) that are present in the conjunctivitis isolates (B1598 and B1599) but absent in strains from the other major branch. We identified 77 coding sequences, many of which were grouped together in the genome (see Fig. S1 in the supplemental material). The most widespread functional feature of these genes is a predicted localization to the bacterial surface: that is, surface localization motifs (LPXTG or YSIRK), surface-related functions (such as ABC transporter), and/or motifs predicted to interact with the host (β -galactosidase, choline binding, sialidase, and two adhesins) (10, 25–29).

The association between this distinct phyletic group and conjunctivitis isolates (prominently those of ST448 and ST344) has been observed in three independent large-scale genomic studies (8–10). To compare our finding to previous work, we utilized a set of 616 genomes isolated in Massachusetts from 2007 to 2010 (6, 24). This set includes 10 strains from the distinct phyletic group identified by Croucher and colleagues (24), termed SC12, as well as 606 additional genomes. The data sets are highly consistent: genes captured exclusively in our distinct phyletic group are present in multiple SC12 genomes and are either absent or rare (<2%) in the 606 non-SC12

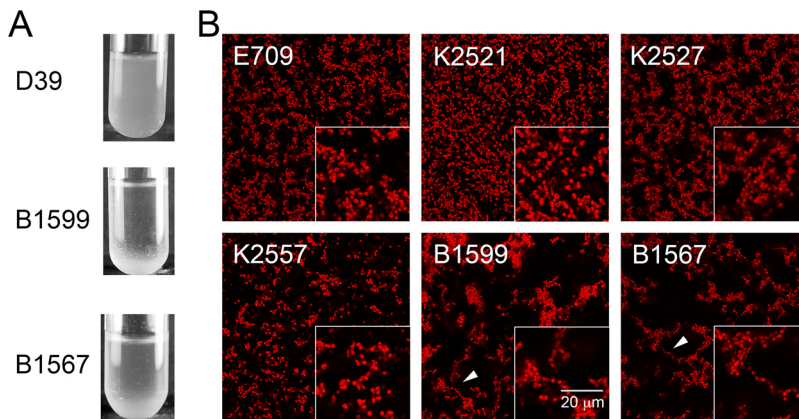


FIG 2 Conjunctivitis-associated strains B1599 and B1567 exhibit aggregates in planktonic culture and form abundant chain-like structures in biofilms. (A) Planktonic cultures of strains B1599, B1567, and D39. Conjunctivitis isolates precipitate at the bottom of the test tube, but no precipitate is observed for model strain D39. (B) Confocal images of 72-h biofilms fixed and stained with Syto59. Boxes on the bottom right display a magnified view. Conjunctivitis isolates B1599 and B1567 form chain-like structures (white arrows point to examples of chain-like structures); however, equivalent chains are not observed in the other eye-associated strains. The scale bar is the same for all micrographs.

genomes (Fig. S1). The set of uncharacterized CDSs, predicted to be surface exposed, are likely candidates for molecular components of tissue tropism to the ocular epithelium during conjunctivitis.

Morphological features that distinguish strains isolated from conjunctivitis relative to other eye infections. The core (Fig. 1) and distributed (Fig. S1) genes of conjunctivitis strains and of a subset of nasopharyngeal strains differ from those of the majority of pneumococcal strains. This raises the question as to whether and how these strains differ regarding morphology.

To explore this question, we investigated the morphology of conjunctivitis strains in planktonic and biofilm modes of growth. The conjunctivitis strain B1599 is of multilocus sequence type (MLST) 2315. However, ST488 (represented in Fig. 1 by isolate MNZ14) is the MLST most commonly associated with conjunctivitis; thus we also analyzed strain B1567 (ST448). Strain B1567 was isolated from a patient with conjunctivitis and was selected as a representative of ST488. First, we analyzed planktonic cultures of these two conjunctivitis strains (B1599 [ST2315] and B1567 [ST488]), comparing them to a model strain, D39. In contrast to canonical strains, such as the control, both conjunctivitis isolates formed aggregates in stationary-phase planktonic cultures, easily visualized at the bottom of the culture tubes (Fig. 2A). Second, we investigated biofilm-related phenotypes, given that pneumococci can colonize the epithelium by growing in a biofilm mode of growth (30–35). We employed confocal microscopy to compare biofilm growth of 3-day biofilms of conjunctivitis strains with those of other ocular strains. In contrast to previously characterized pneumococcus biofilms (36) and to the keratitis and endophthalmitis isolates, the majority of cells in strain B1599 were organized into long chain-like structures, often with over 20 connected cells (Fig. 2B). The long structures were also formed in strain B1567. To our knowledge, this phenotype has not been previously reported in wild-type pneumococcal strains—only pneumococcal mutants (37, 38). These observations raise the hypothesis that strains in the phyletic group that contains the conjunctivitis isolates display phenotypic differences apparent in planktonic and biofilm growth that distinguish them from most characterized pneumococcal isolates, as well isolates from other ocular infections.

Functional studies of pneumococcal SspB. To investigate the relationship between genotype and function in the distinct phyletic group, we focused on a predicted agglutinin receptor encoded exclusively by the distinct phyletic group (Fig. S1) (GenBank accession no. [KGI30072](#) and [OYL08640.1](#)). This predicted protein contains an N-terminal Sec-type signal sequence and a C-terminal LPXTG motif, strongly sug-

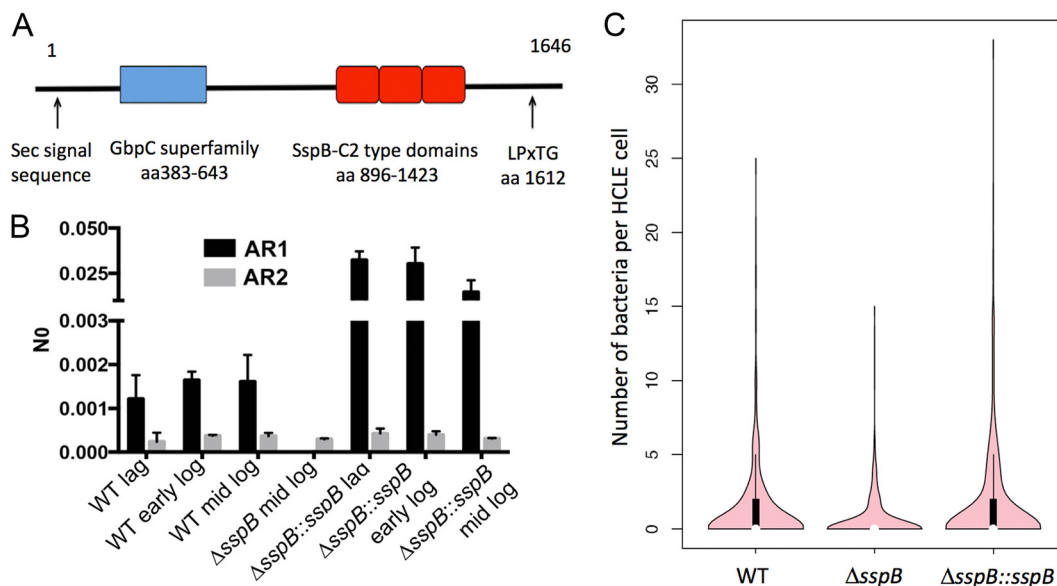


FIG 3 Functional analyses of the *sspB* gene encoding the predicted SspB adhesin. (A) Schematic of the predicted SspB protein (GenBank accession no. [KGI30072](#) and [OYL08640.1](#)) illustrating domains implicated in adhesion and biofilm formation. (B) Gene expression of *sspB* in *S. pneumoniae* B1599 (black bars); for comparison, we show the expression of a second related protein with SspC-C2 domains (WP_050568636) (gray bars). The y axis displays *N₀*, the number of fluorescence units representing the RNA amount in the respective samples. (C) Role of *sspB* in attachment to HCLE cells. HCLE cells were exposed to *S. pneumoniae* strain B1599, B1599 Δ *sspB*, or B1599 Δ *sspB::sspB* for 30 min, and HCLE cells with bacteria attached were enumerated. Experiments were performed in triplicate, and a total of 460 HCLE cells were analyzed for presence of bacteria and the number of bacteria attached. The results are plotted using a violin plot, generated in the R statistical package. The violin plot displays the distribution of the data: the pink areas display the density plot, the thick black bars represent the midspread of the data (interquartile range), the thin black lines display the 95% confidence interval, and the white circles correspond to the median.

gesting it is attached to the peptidoglycan cell wall by a sortase (25, 28). It contains a glucan-binding protein C (GbpC) domain (Conserved Domain Database [CDD] E value of $8.2e-80$), followed by three adhesin isopeptide-forming domains (SspB-C2 type) (CDD E values of $3.12e-69$, $4.61e-68$, and $3.57e-63$, respectively) (39) (Fig. 3A). Proteins containing the GbpC domain are found in several species of oral streptococci, where they participate in dextran binding and biofilm formation (40, 41). SspB-C2-type domains are present in oral streptococci as components of SspB (42). In *Streptococcus gordonii*, SspB is a three-domain adhesion unit that facilitates cross-species interactions in the oral cavity (43, 44). This predicted pneumococcal sequence is 37% identical to the streptococcal surface protein B (SspB) precursor from *S. gordonii* (GenBank accession no. [AAC44102.3](#)) (45), which led to the designation of SspBC1 for the pneumococcal homolog (X231_1085) (9). To avoid confusion between the protein name and domain names, we refer to the pneumococcal protein KGI30072 as SspB. We found that the gene encoding SspB is expressed in planktonic cultures of strain B1599, as determined by quantitative reverse transcription-PCR (qRT-PCR) (Fig. 3B, black bars). The predicted surface localization and the presence of adhesive domains are consistent with a role for SspB in host interactions, and its restricted genomic distribution is consistent with a tropism to the ocular epithelium.

To test whether *sspB* has a role in adhesion to the ocular epithelium, we generated a deletion mutant (B1599 Δ *sspB*) and an overexpressor strain (B1599 Δ *sspB::sspB*) (see Table S2 in Data Set S2). The levels of *sspB* in the overexpression strain vary from 9- to 26-fold higher than the levels of the wild-type strains, depending on the stage of planktonic growth (Fig. 3B). The mutant and wild-type strains displayed the same morphology regarding aggregation in planktonic culture and chain-like structures in a biofilm (see Fig. S2 in the supplemental material). Thus, *sspB* alone is not responsible for the differences characterized in Fig. 2. Next, we investigated whether the gene encoding SspB influences adhesion to human corneal limbal epithelial (HCLE) cells. HCLE cells

were selected because they are found at the interface between the conjunctiva and the cornea, and they are representative of the ocular surface in that they produce many mucins and compounds associated with cornea and conjunctiva (46, 47). Independently, each strain was allowed to adhere to HCLE cells, and the number of bacteria attached per cell was enumerated after gentle washing to remove nonadherent bacteria. Tukey's test was used to establish the statistical significance of the differences in adherence to 460 HCLE cells among the three strains; these cells were gathered over three independent experiments (48). The attachment for the deletion mutant strain was significantly different from that of the wild-type ($P = 0.0019$) and complement ($P = 0.000003$) strains. Specifically, the deletion of *sspB* led to a decrease in the number of HCLE cells with any bacteria attached, as well as a decrease in the number of bacteria attached per HCLE cell (Fig. 3C). Overexpression of the gene encoding SspB in the deletion strain restored the wild-type phenotype. These data suggest that SspB plays a role in adhesion to the ocular epithelium.

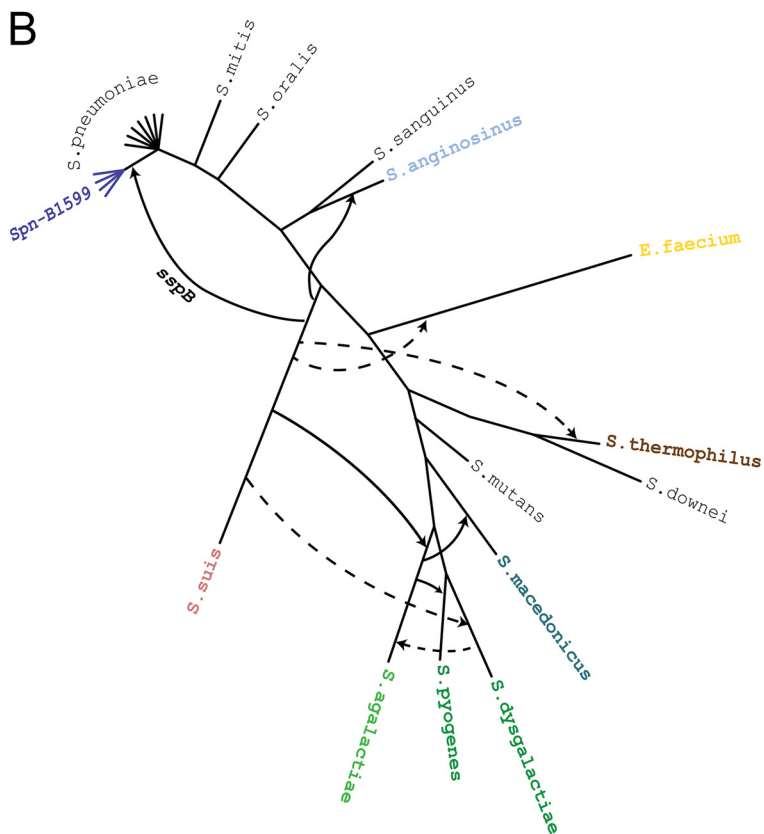
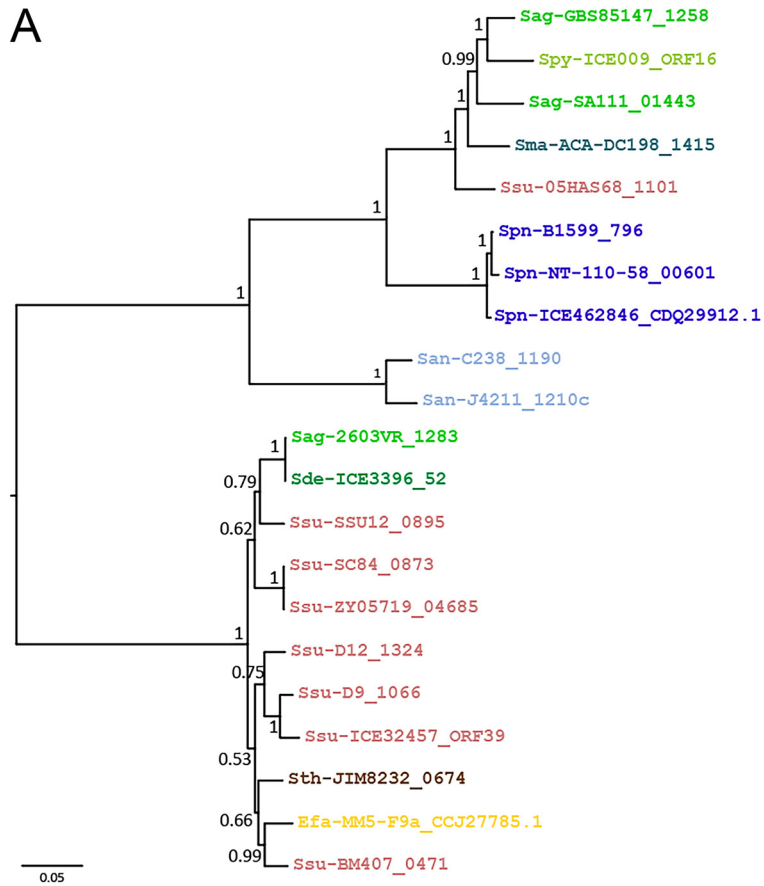
Origin of the pneumococcal *sspB*. A scan of the genomic region surrounding *sspB* using CONJscan-T4SSscan software (49) identified genes encoding components of type IV secretion systems and relaxases associated with integrative conjugative elements (ICEs), consistent with a foreign origin for *sspB*. The closest relatives of the region carrying *sspB* and neighboring ICE components were identified via BLASTn search, resulting in 26 genomic regions (see Table S3 in Data Set S2). Of these, five were derived from ICE sequences, further supporting the inference that *sspB* is part of a mobile genetic element. One sequence is from outside the streptococcal genus: an ICE from *Enterococcus faecium*. Two sequences are from within *S. pneumoniae*: an ICE from isolate 403790 and a nontypeable pneumococcal strain (NT_110_58) previously localized to a distinct phyletic group (8). All other matches are streptococcal sequences from species outside *S. pneumoniae*: 13 *Streptococcus suis* strains, 2 *Streptococcus anginosus* strains, 6 strains from various pyogenic species, and one each from *Streptococcus thermophilus* and *Streptococcus gallolyticus* subsp. *macedonicus* (*Streptococcus macedonicus*).

While this patchy phyletic distribution, covering multiple dispersed groups within the genus, is suggestive of lateral gene transfer, phylogenetic analysis provides stronger evidence for lateral transfer. Explicit comparison of a gene tree with the associated species tree can distinguish between parallel loss and horizontal transfer and infer the specific transfer events that occurred. A Bayesian gene tree (Fig. 4A) was constructed from a codon-aware multiple alignment of *sspB* sequences extracted from the 21 closely related genomic regions remaining after redundant sequences, and sequences lacking ORF predictions were removed (see Table S3 in Data Set S2). Interspecies transfers were inferred by reconciling this gene tree with a previously published streptococcal species tree (see Fig. S3 in the supplemental material) (50) using Notung 2.9 (51).

Reconciliation infers the minimal set of events that explains the topological incongruence between gene and species trees. Reconciliation with Notung 2.9 yielded four possible evolutionary scenarios with minimal event histories (Fig. 4A to D). The four scenarios are largely in agreement, differing only in the events involving pyogenic strains, which are unrelated to our central question. In particular, all four scenarios predict a single horizontal transfer from *Streptococcus suis* to *Streptococcus pneumoniae*. This suggests that the pneumococcal *sspB* was acquired from *Streptococcus suis* at the base of the distinct phyletic lineage.

DISCUSSION

The clustering of conjunctivitis isolates into a distinct phyletic group is the only clear instance of niche specialization in pneumococcus known to date. In this study, we compare the genomes and evolutionary histories of pneumococcal strains isolated from various body sites, focusing on multiple types of eye infections. We observe that conjunctivitis isolates cluster with a subset of nasopharyngeal isolates and away from all other strains, including those isolated from patients with keratitis and endophthal-



mitis. Phenotypically, our conjunctivitis isolates display aggregation in planktonic culture and chain-like structures in biofilms grown on an abiotic surface. Furthermore, we demonstrate that the *sspB* gene was acquired from *S. suis* by gene transfer and that its product plays a role in adhesion to the ocular epithelium. Our work combines comparative genomics, phylogenetics, and cell biology to explore the evolution and molecular mechanisms that underlie this unique instance of pneumococcus niche specialization.

The long branch that splits off the distinct phyletic group from other pneumococcal strains highlights the extensive differences in their core genomes and thus provides evidence of strain differentiation. This differentiation does not appear to be the consequence of niche separation, as some members of the distinct phyletic group are nasopharyngeal isolates and likely cocolonize with pneumococcal isolates from the main branch. Another mechanism promoting strain differentiation in pneumococcus and other bacteria is restriction-modification systems (52, 53). Whereas genomes in the distinct phyletic group do encode a type I restriction-modification system absent in the major branch, it is only encoded in a subset of these distinct strains (Fig. S1). This limited distribution suggests that this restriction-modification system is not driving strain differentiation. Thus, the molecular mechanisms driving this instance of strain differentiation remain a fascinating topic of study.

The strains in the distinct phyletic group share a set of genes not observed in strains outside this group. Do these differences in gene content between strains translate into differences in transmission routes and/or tissue tropism? Pneumococcal strains from the main branch are transmitted by nasal shedding, and the nasopharynx is the main reservoir of circulating strains (54). The phylogenetic grouping of conjunctivitis and nasopharyngeal strains is consistent with this model, in that strains from the distinct phyletic group may be transmitted to and from the nasopharynx and then disseminated to the ocular epithelium. Alternatively, strains acquired directly into the eye may have the ability to colonize the nasopharynx. Once in contact with the conjunctiva, a strain from the distinct phyletic branch may be able to overcome a specific host barrier and gain access to the ocular epithelium. Phylogenetic positioning of the keratitis and endophthalmitis isolates outside the distinct phyletic group suggests that this ability is not associated with other eye tissues or the ability to survive immunity in the eye. Instead, the ability to colonize the ocular epithelium could be guided by features specific to the conjunctiva, such as adhesion and/or colonization of its epithelium, or by an intermediate niche, such as eyelids, eyelashes, or even fingers that people use to rub their eyes. Moreover, these strains may be more resistant to desiccation or more competent at survival in the tear film, ultimately allowing productive infection of the conjunctiva. We postulate that the distinct phyletic group encodes proteins that allow these strains to overcome a host barrier associated with transmission and/or colonization of the ocular epithelium.

In this study, we captured a notable association between genotype and phenotype. First, we observed long chain-like structures only within biofilms of wild-type strains from this distinct phyletic group. To our knowledge, this phenotype has not been

FIG 4 Phylogenetic reconciliation reveals a history of transfers in the origin of *sspB*. (A) Bayesian phylogeny of *sspB* gene sequences constructed with MrBayes v.3.2.6 from codon-aware multiple alignment of nucleic acid sequences. The tree is midpoint rooted. Branches are labeled with posterior probabilities representing statistical support. The scale bar represents the number of nucleotide substitutions per site. Colors are by species according to the labeling in panel B. (B) Interspecies *sspB* gene transfers in the context of the *Streptococcus* phylogeny. Transfers were inferred by reconciling the gene tree (A) with a species tree (see Fig. S4 in the supplemental material) based on the streptococcal species tree from Richards et al. (50), generated from a core set of 136 genes sampled from 44 streptococcal species. The evolutionary history contains all species in panel A (in color), plus additional species (in black) to provide a representative sample of well-studied taxa with at least one species for each major taxonomic group. The distinct phyletic lineage in pneumococcus is shown in dark blue. Four evolutionary scenarios were inferred with Notung 2.9 (see Fig. S5 in the supplemental material), each with 9 transfers, represented as arrows from the donor to the recipient species. (Solid and dashed lines represent transfers inferred in the upper and lower subtrees in panel A, respectively.) All evolutionary scenarios support a horizontal transfer from *S. suis* to the base of the distinct lineage (dark blue).

reported for biofilms of wild-type strains grown on abiotic surfaces. It has only been observed in cell culture (55) and with genetic mutants of *S. pneumoniae* (38, 56). Furthermore, there is precedence for a positive association between long chains and increased adherence (55), such that a similar trend may hold for strains in the distinct phyletic branch. Second, we observed aggregates in planktonic culture for the conjunctivitis strains. In a related experiment, Valentino and colleagues reported clumping of conjunctivitis isolates after addition of gp340, a glycoprotein found in tears (9). It remains to be tested whether these three phenotypes have the same molecular basis. Our studies with the *sspB* deletion demonstrate that SspB is not required for these phenotypes. These phenotypes could be due to other proteins, or perhaps the involvement of SspB is obscured due to redundancy in protein function.

Finally, our findings suggest that the gene encoding SspB contributes to adhesion to the ocular epithelium. It is likely that multiple redundant adhesins influence epithelial binding. For instance, a second putative adhesin with SspB-C2 domains (WP_050568636, previously referred to as SspB-C2 [9]) is also expressed in the distinct phyletic group (Fig. S1). The molecule contains multiple SspB-C2 domains (CDD E values of $1.89e-70$, $1.8e-69$, $2.17e-67$, and $6.9e-66$, respectively) and a C-terminal LPXTG motif, consistent with a cell wall attachment adhesin (Fig. S1). It seems probable that the propensity to form long chain-like structures in early biofilms and/or to aggregate in culture is widespread across isolates from the distinct phyletic group and, together with unique surface adhesins, plays an important role in interactions with the host.

The *sspB* gene is located within an ICE, and reconciliation between the streptococcal species tree and the *sspB* gene tree suggests *sspB* was acquired by gene transfer from *S. suis*. Furthermore, multiple *sspB* xenologs are contained within ICEs, such as MB56Spyo009, ICESde3396, ICESu32457, and ICESsu32457 (57–60). Similarly, the reconciliation suggests these ICEs were also acquired by gene transfer from *S. suis*. Thus, it is likely that *S. suis* serves as an ICE reservoir dispersing these elements to many species within the streptococcal genus (60). It is tempting to speculate that ICE-associated adhesins may modify bacterial interactions with the host, either promoting or suppressing dissemination to specific hosts, body sites, and/or tissues.

The distinct phyletic group in the pneumococcal species tree provides a striking example of strain differentiation and tissue tropism. It generates many open questions on how these strains differ from other pneumococcal strains regarding evolution, transmission routes, morphology, gene expression, and host-pathogen interactions. Studies of these strains will provide exciting insight into the evolution and the biology of pneumococcus.

MATERIALS AND METHODS

Bacterial strains. The *Streptococcus pneumoniae* isolates B1567, B1598, B1599, E709, K2521, K2527, and K2557 were obtained from patient ocular infections and stored by the Charles T. Campbell Eye Microbiology Laboratory at the University of Pittsburgh Medical Center (UPMC) Eye Center (Table 1).

Bacterial growth conditions. Frozen bacterial stocks were streaked onto Trypticase soy agar plates containing 5% sheep blood (BD BBL). All *S. pneumoniae* strains were grown in Columbia broth (Thermo Scientific) at 37°C with 5% CO₂ without shaking. Medium was supplemented with antibiotics at 1 μg/ml for tetracycline and 100 μg/ml for spectinomycin.

Pacific Biosciences SMRT sequencing. Genomic DNA (10 μg) was extracted from strains E709, K2527, K2521, K2557, B1598, and B1599. Following digestion into ~10-kb fragments, the DNA was end repaired, purified with AMPure PB beads, and ligated to SMRTbell hairpin adapters. The SMRTbell libraries were further purified and quantified with a NanoDrop spectrophotometer and an Agilent 2100 Bioanalyzer. Polymerase-bound libraries were loaded onto a PacBio RS for sequencing with two SMRT cells per strain after completing primer and polymerase binding.

Multilocus sequence typing of strains. Sequences for the seven MLST alleles with available whole-genome sequences were extracted from sequence data and are listed in Table 1. For strain B1567, the ST was acquired by Sanger sequencing of PCR amplicons (21). For typing, the allele sequences were submitted to the *S. pneumoniae* MLST website (<https://pubmlst.org/spneumoniae>) (61). The serotypes were predicted by comparing the capsular locus to the sequences deposited in NCBI: specifically 15B (GenBank accession no. [CR931664.1](#)), 17F (GenBank accession no. [CR931670](#)), 19A (GenBank accession no. [CR931675.1](#)), and 23A (GenBank accession no. [CR931683.1](#)). Note that serotype 15C differs from 15B

only by an additional two nucleotides within a stretch of TA repeats, wherein the additional bases appear to lead to a premature stop in a predicted *O*-acetyltransferase (62).

Gene annotation. Genomes were submitted to RAST for CDS prediction and annotation (63).

Generation of *S. pneumoniae* species tree. Forty streptococcal strains were selected for phylogenetic analysis (see Table S1 in Data Set S2). These genomes correspond to those used for the first large-scale pneumococcal pangenome studies (2, 5), as well as additional genomes from PCV-7-immunized children (22), and nonencapsulated strains (23). Combined these strains reflect a large variety of multilocus sequence types (MLSTs) and serotypes, as well as strains isolated from different disease states and geographic locations. The whole-genome sequences (WGSs) for all 40 strains were aligned using MAUVE (64), and the core region corresponding to 1,345,780 total sites and 92,728 informative sites was extracted from the MAUVE output files. Alignment of the core region was performed using MAFFT (MODEL) (65), and model selection was performed using ModelTest (66). The phylogenetic tree was built with PhyML 3.0, model GTR+I(0.63), using maximum likelihood analysis and 100 bootstrap replicates (67).

Gene clustering and selection of genes unique to distinct phylogenetic clade. The CDSs were organized into gene clusters as previously described (68). Briefly, similar genes were identified by tfasty36 (FASTA v.3.6 package) for six-frame translation homology searches of all predicted proteins against all possible translations (69). The output was parsed such that genes with at least 70% identity over 70% of their sequence were grouped into gene clusters allowing the strains to be analyzed for presence or absence of clusters. We selected genes present in strains B1598, B1599, MNZ14, MNZ85, and MNZ41 but absent in the remaining pneumococcal strains.

We utilized a set of 616 genomes isolated in Massachusetts from 2007 to 2010 (6) to compare our findings to previous work. This set, analyzed by Croucher and colleagues, contains 10 strains from a distinct phylogenetic branch termed SC12, as well as 606 additional genomes. To compare the sets, we used BLASTp and an E value cutoff of $1e-10$ to query the 77 genes unique to our distinct phyletic group against a database of 1,231,516 sequences from the 616 genomes in the set isolated in Massachusetts from 2007 to 2010. The BLAST results were parsed to exclude hits with less than 70% identity and/or less than 70% coverage. A Python script was used to parse the BLAST output into a matrix correlating the query with the output. The data were represented as a heat map using the ggplot2 package in the R statistical package (70). All positive hits are plotted in Fig. S1. (Genomes without any hits are not represented in the heat map.)

Phyletic distribution. Sequences were retrieved using NCBI BLASTn to search the nonredundant (nr) database, restricted to *Bacillus*, *Lactobacillus*, and *Streptococcus* (taxid: 91061). The query corresponded to a 26,160-bp region of B1599 (within B1599_contig 202) that comprises *sspB* and adjacent genes. The resulting matches were curated to include only sequences with an E value of zero and a maximum bit score above 8,100; all of these matches entailed high-scoring pairs (HSPs) with at least 74% identity distributed along the full length of the query sequence, suggesting that these are genomic regions that are homologous to the query sequence in its entirety. The final set includes 26 genomic sequences, of which 5 are ICE sequences (see Table S3 in Data Set S2).

Gene tree reconstruction. To analyze the evolutionary origin of the *sspB* gene, we extracted the sequences that contain *sspB* and inferred a gene tree. Genomes lacking ORF predictions were discarded. For each of the remaining genomes, the *sspB* nucleotide and amino acid sequences were extracted from the GenBank file using GenBank_to_fasta.py, downloaded from the Rocap lab website (http://rocaplab.ocean.washington.edu/tools/genbank_to_fasta/). This resulted in 21 sequences after removing redundant sequences and adding B1599 *sspB* to the set.

Multiple alignment of the SspB protein sequences was performed using MAFFT (65), with the "E-INS-I" option via the Jalview dashboard (71). The alignment was converted into a codon-aware nucleic acid alignment using PAL2NAL (Fig. S1) (72), which was trimmed manually in Jalview to remove columns with more than 25% gaps. The best phylogenetic model for each of the three codon positions GTR + G, HKY + G, and HKY + G, was selected by the Bayesian information criterion using MODELGENERATOR (73) in Topali (74). A gene tree (Fig. 4A) was then constructed from the trimmed alignment using MrBayes v.3.2.6 (75); model parameters were fit by MrBayes for each codon site. The Markov chain Monte Carlo (MCMC) process was run for 500,000 generations with a sampling frequency of 15 generations and default settings for all other MCMC parameters. The gene tree was then midpoint rooted in FigTree, which was also used in figure generation (<http://tree.bio.ed.ac.uk/software/figtree/>).

Gene tree-species tree reconciliation. To infer the history of evolutionary events during *sspB* evolution, the resulting gene tree was reconciled with a species tree (Fig. S3), adapted from the tree from Richards et al. (50), for a reduced set of taxa. The Richards species tree was constructed from a concatenation of a core set of 136 genes across 44 streptococcus species (46 strains), representing 8 major groups. All species that encode a putative homolog of *sspB* were retained in the species tree for this study, as were additional species (that do not harbor a putative *sspB* homolog) to provide a broad representation of the streptococcal genus, including at least one species for each of the major taxonomic groups identified by Richards et al. (*mitis*, *sanguinis*, *anginosus*, *salivarius*, *downei*, *mutans*, *pyogenes*, and *bovis*). When an *sspB* homolog was predicted in more than one strain per species, all strains were added to the species tree. Strain relationships are unresolved in this tree; if more than two strains were included for a single species, their relationships are represented as a nonbinary node (i.e., polytomies).

We utilized Notung 2.9 to reconcile the *sspB* gene tree with this nonbinary species tree under a duplication, transfer, and loss model (see Fig. S5A in the supplemental material). Notung infers the history of events that minimizes the weighted sum of events when fitting a gene tree to a species tree; in this analysis, we used weights of 3.0, 1.5, and 1.6 for transfers, duplications, and losses, respectively.

Notung does not infer events between taxa within an unresolved clade. We chose to represent strain relationships as nonbinary nodes (i.e., polytomies) in order to focus the analysis to interspecies transfers and not intraspecies transfers. A schematic (Fig. 4B; Fig. S5B to D) displaying the predicted history of gene transfers on the species tree was generated in FigTree and Adobe Illustrator.

Biofilm growth and imaging. All strains were grown in Columbia broth to an optical density at 600 nm (OD_{600}) of 0.05 before seeding the culture onto MatTek dishes. At 24 and 48 h, medium was exchanged using diluted (1/5 with water) Columbia broth. At 72 h, biofilms were washed 2 times with phosphate-buffered saline (PBS) and fixed with 4% paraformaldehyde (PFA) for 30 min. Fixed biofilms were stained with Syto59 fluorescent dye according to the manufacturer's instructions (LifeTech). Biofilms were imaged using a Zeiss 510 Meta Confocor3 laser scanning microscope, and images were processed using ImageJ.

Bacterial cell aggregation assays. Strains D39, B1599, and B1567 were inoculated into Columbia broth and grown until an OD_{600} of 0.05. Each culture was diluted 10-fold in full-strength Columbia broth and incubated overnight. Photographs were taken the following day, after 18 to 20 h of growth. Test tubes were photographed with Olympus Pen E-P1 for documentation.

Construction of the deletion mutant and complemented strain. The *sspB* deletion mutant in strain B1599 was generated by replacement of this gene with a spectinomycin resistance cassette. Specifically, we amplified the 1-kb upstream and downstream regions of *sspB* and ligated the flanking sequences to the resistance cassette by sticky end ligation with T4 DNA ligase. We amplified the ligation mixture by PCR to generate the transforming DNA (primers in Table S4 in Data Set S2). The *sspB* complement strain (B1599 Δ *sspB*::*sspB*) was generated in the Δ *sspB* background by reintroducing the *sspB* gene into a conserved intragenic region (contig 208, position 161851) previously used for complementation (53).

Bacterial transformations. For all bacterial transformations, about 1 μ g of transforming DNA was added to the growing culture of a target strain at an OD_{600} of 0.05, supplemented with 125 μ g/ml of CSP2 (sequence EMRISRILDFLFLRKK [purchased from GenScript, Piscataway, NJ]), and incubated at 37°C. After 4 h, the treated cultures were plated on Columbia agar containing 100 μ g/ml spectinomycin. Resistant colonies were cultured in media, the region of interest was amplified by PCR, and the amplicon was submitted for Sanger sequencing (Genewiz, Inc.) to verify the sequence of the mutants.

mRNA isolation and qRT-PCR analysis. Strains B1599, B1599 Δ *sspB*, and B1599 Δ *sspB*::*sspB* were grown in Columbia broth until reaching OD_{600} s of 0.05, 0.2, and 0.5. At each time point, 5 ml of culture was collected and mixed with RNAlater. Pelleted cultures were frozen until RNA extraction. The RNeasy Plus minikit from Qiagen was used to extract and purify RNA from each sample. Each sample was DNase treated to remove DNA contamination. Expression of *sspB* was assayed by qRT-PCR and normalized to GAPDH (glyceraldehyde-3-phosphate dehydrogenase). Primers for each locus were designed using Roche Universal Probe Library software. The experiment was performed in triplicate, and data were analyzed using LinReg PCR software. Statistical analysis was performed with Wilcoxon's paired rank test using GraphPad.

Mammalian tissue culture conditions. Human corneal limbal epithelial (HCLE) cells were cultured in keratinocyte serum-free medium (KSFM [Gibco catalog no. 10724-011]) containing 25 μ g/ml bovine pituitary extract (Gibco catalog no. 13028-014), 10.2 ng/ml embryonic growth factor (Gibco catalog no. 10450-013), 100 μ g/ml penicillin, and 100 μ g/ml streptomycin (Corning catalog no. 30-002-CL).

Bacterial attachment to HCLE cells. HCLEs were seeded into 12-well MatTek glass bottom dishes (MatTek P12G-1.5-14-F) in antibiotic-free KSFM at a density of 1.50×10^5 cells per well and allowed to adhere overnight at 37°C with 5% CO_2 . *Streptococcus pneumoniae* strains B1599 (wild type), B1599 Δ *sspB*, and B1599 Δ *sspB*::*sspB* were streaked onto blood agar and grown overnight at 37°C with 5% CO_2 . Bacteria were scraped off of the blood agar with an inoculating loop, added to 2.5 ml Columbia broth, and grown for 5 h to an OD_{600} of 0.3. Cultures were pelleted by centrifugation at 14,000 rpm for 2 min. The pellets were washed two times in PBS and resuspended in 1 ml Columbia broth. Two hundred microliters of each strain was added to each well of HCLE cells containing 1 ml of antibiotic-free KSFM. The plate was incubated at 37°C with 5% CO_2 for 30 min. After incubation, HCLEs were washed gently two times with PBS and supplemented with fresh KSFM. The cells were imaged on an Olympus Fluoview FV-1000 laser scanner confocal microscope with a 60 \times objective. Ten fields per treatment group were imaged, and the number of bacteria on each HCLE was manually counted using Fluoview image viewing software version 3.1. The experiment was repeated on three separate days with similar results.

Accession number(s). We have submitted the complete genome sequences of the following six genomes to GenBank under the accession numbers given in parentheses: E709 (JBOR000000000), K2521 (JBOS000000000), K2527 (JBOT000000000), K2557 (JBOU000000000), B1598 (JBOV000000000), and B1599 (JBOW000000000).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSphere.00213-17>.

FIG S1, TIF file, 2.6 MB.

FIG S2, TIF file, 6.5 MB.

FIG S3, TIFF file, 0.8 MB.

FIG S4, TIF file, 2.8 MB.

FIG S5, TIFF file, 13.9 MB.

DATA SET S1, TXT file, 0.1 MB.

DATA SET S2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Nicholas Stella for extensive help with HCLE studies and Annette McLeod and Jacob West-Roberts for assistance with data presentation and figure preparation. We also thank Matthew Salyers for help with photographing bacterial cultures. We are grateful to Haibing Teng from the Molecular Biosensor and Imaging Facility (MBIC) at Carnegie Mellon University for assistance and support with confocal imaging and Garth Ehrlich and the Allegheny Health Network for support with genome sequencing. We are grateful to the Charles T. Campbell Laboratory of Ophthalmic Microbiology for sharing clinical strains. This work was supported by NIH grants EY024785 to the Department of Ophthalmology, University of Pittsburgh, EY024785 to K.M.B., EY027331 to R.M.Q.S., and R00-CD-011322 to N.L.H., by support from Carnegie Mellon University and Allegheny Health Network, and by NSF grant DBI1262593 to D.D. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors do not have any conflicts of interest to report.

REFERENCES

- Bessen DE, Kumar N, Hall GS, Riley DR, Luo F, Lizano S, Ford CN, McShan WM, Nguyen SV, Hotopp JCD, Tettelin H. 2011. Whole-genome association study on tissue tropism phenotypes in group A *Streptococcus*. *J Bacteriol* 193:6651–6663. <https://doi.org/10.1128/JB.05263-11>.
- 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>.
- Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome.” *Proc Natl Acad Sci U S A* 102:13950–13955. <https://doi.org/10.1073/pnas.0506758102>.
- Kulohoma BW, Cornick JE, Chaguza C, Yalcin F, Harris SR, Gray KJ, Kiran AM, Molyneux E, French N, Parkhill J, Faragher BE, Everett DB, Bentley SD, Heyderman RS. 2015. Comparative genomic analysis of meningitis- and bacteremia-causing pneumococci identifies a common core genome. *Infect Immun* 83:4165–4173. <https://doi.org/10.1128/IAI.00814-15>.
- Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, Keefe R, Ehrlich NE, Shen K, Hayes J, Barbadora K, Klimke W, Dernovoy D, Tatusova T, Parkhill J, Bentley SD, Post JC, Ehrlich GD, Hu FZ. 2007. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. *J Bacteriol* 189:8186–8195. <https://doi.org/10.1128/JB.00690-07>.
- Croucher NJ, Finkelstein JA, Pelton SI, Parkhill J, Bentley SD, Lipsitch M, Hanage WP. 2015. Population genomic datasets describing the post-vaccine evolutionary epidemiology of *Streptococcus pneumoniae*. *Sci Data* 2:150058. <https://doi.org/10.1038/sdata.2015.58>.
- Kadioglu A, Weiser JN, Paton JC, Andrew PW. 2008. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* 6:288–301. <https://doi.org/10.1038/nrmicro1871>.
- Hilty M, Wüthrich D, Salter SJ, Engel H, Campbell S, Sá-Leão R, de Lencastre H, Hermans P, Sadowy E, Turner P, Chewapreecha C, Diggle M, Pluschke G, McGee L, Köseoglu Eser Ö, Low DE, Smith-Vaughan H, Endimiani A, Küffer M, Dupasquier M, Beaudouin E, Weber J, Brüggemann R, Hanage WP, Parkhill J, Hathaway LJ, Mühlemann K, Bentley SD. 2014. Global phylogenomic analysis of nonencapsulated *Streptococcus pneumoniae* reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. *Genome Biol Evol* 6:3281–3294. <https://doi.org/10.1093/gbe/evu263>.
- Valentino MD, McGuire AM, Rosch JW, Bispo PJM, Burnham C, Sanfilippo CM, Carter RA, Zegans ME, Beall B, Earl AM, Tuomanen EI, Morris TW, Haas W, Gilmore MS. 2014. Unencapsulated *Streptococcus pneumoniae* from conjunctivitis encode variant traits and belong to a distinct phylogenetic cluster. *Nat Commun* 5:5411. <https://doi.org/10.1038/ncomms6411>.
- Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat Commun* 5:5471. <https://doi.org/10.1038/ncomms6471>.
- Keller LE, Robinson DA, McDaniel LS. 2016. Nonencapsulated *Streptococcus pneumoniae*: emergence and pathogenesis. *mBio* 7:e01792-15. <https://doi.org/10.1128/mBio.01792-15>.
- 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>.
- 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>.
- Salter SJ, Hinds J, Gould KA, Lambertsen L, Hanage WP, Antonio M, Turner P, Hermans PWM, Bootsma HJ, O'Brien KL, Bentley SD. 2012. Variation at the capsule locus, cps, of mistyped and non-typable *Streptococcus pneumoniae* isolates. *Microbiology* 158:1560–1569. <https://doi.org/10.1099/mic.0.056580-0>.
- Murrah KA, Pang B, Richardson S, Perez A, Reimche J, King L, Wren J, Swords WE. 2015. Nonencapsulated *Streptococcus pneumoniae* causes otitis media during single-species infection and during polymicrobial infection with nontypeable *Haemophilus influenzae*. *Pathog Dis* 73. <https://doi.org/10.1093/femspd/ftu011>.
- 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>.
- Hiller NL, Ahmed A, Powell E, Martin DP, Eutsey R, Earl J, Janto B, Boissy RJ, Hogg J, Barbadora K, Sampath R, Lonergan S, Post JC, Hu FZ, Ehrlich GD. 2010. Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog* 6:e1001108. <https://doi.org/10.1371/journal.ppat.1001108>.
- Marquart ME, O'Callaghan RJ. 2013. Infectious keratitis: secreted bacterial proteins that mediate corneal damage. *J Ophthalmol* 2013:e369094. <https://doi.org/10.1155/2013/369094>.

19. Miller JJ, Scott IU, Flynn HW, Smiddy WE, Corey RP, Miller D. 2004. Endophthalmitis caused by *Streptococcus pneumoniae*. *Am J Ophthalmol* 138:231–236. <https://doi.org/10.1016/j.ajo.2004.03.008>.
20. Parmar P, Salman A, Kalavathy CM, Jesudasan CAN, Thomas PA. 2003. Pneumococcal keratitis: a clinical profile. *Clin Exp Ophthalmol* 31:44–47. <https://doi.org/10.1046/j.1442-9071.2003.00598.x>.
21. Enright MC, Spratt BG. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* 144:3049–3060. <https://doi.org/10.1099/00221287-144-11-3049>.
22. Frazão N, Hiller NL, Powell E, Earl J, Ahmed A, Sá-Leão R, de Lencastre H, Ehrlich GD, Tomasz A. 2013. Virulence potential and genome-wide characterization of drug resistant *Streptococcus pneumoniae* clones selected in vivo by the 7-valent pneumococcal conjugate vaccine. *PLoS One* 8:e74867. <https://doi.org/10.1371/journal.pone.0074867>.
23. Keller LE, Thomas JC, Luo X, Nahm MH, McDaniel LS, Robinson DA. 2013. Draft genome sequences of five multilocus sequence types of nonencapsulated *Streptococcus pneumoniae*. *Genome Announc* 1:e00520-13. <https://doi.org/10.1128/genomeA.00520-13>.
24. Croucher NJ, Finkelstein JA, Pelton SJ, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* 45:656–663. <https://doi.org/10.1038/ng.2625>.
25. Bae T, Schneewind O. 2003. The YSIIRK-G/S motif of staphylococcal protein A and its role in efficiency of signal peptide processing. *J Bacteriol* 185:2910–2919. <https://doi.org/10.1128/JB.185.9.2910-2919.2003>.
26. Bergmann S, Hammerschmidt S. 2006. Versatility of pneumococcal surface proteins. *Microbiology* 152:295–303. <https://doi.org/10.1099/mic.0.28610-0>.
27. Jeong JK, Kwon O, Lee YM, Oh DB, Lee JM, Kim S, Kim EH, Le TN, Rhee DK, Kang HA. 2009. Characterization of the *Streptococcus pneumoniae* BgaC protein as a novel surface β -galactosidase with specific hydrolysis activity for the Gal β 1-3GlcNAc moiety of oligosaccharides. *J Bacteriol* 191:3011–3023. <https://doi.org/10.1128/JB.01601-08>.
28. Schneewind O, Missiakas DM. 2012. Protein secretion and surface display in Gram-positive bacteria. *Philos Trans R Soc Lond B Biol Sci* 367:1123–1139. <https://doi.org/10.1098/rstb.2011.0210>.
29. Zähler D, Hakenbeck R. 2000. The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *J Bacteriol* 182:5919–5921. <https://doi.org/10.1128/JB.182.20.5919-5921.2000>.
30. Blanchette-Cain K, Hinojosa CA, Akula Suresh Babu R, Lizcano A, Gonzalez-Juarbe N, Munoz-Almagro C, Sanchez CJ, Bergman MA, Orihuela CJ. 2013. *Streptococcus pneumoniae* biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. *mBio* 4:e00745-13. <https://doi.org/10.1128/mBio.00745-13>.
31. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, Forbes M, Greenberg DP, Dice B, Burrows A, Wackym PA, Stoodley P, Post JC, Ehrlich GD, Kerschner JE. 2006. Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *JAMA* 296:202–211. <https://doi.org/10.1001/jama.296.2.202>.
32. Hoa M, Syamal M, Sachdeva L, Berk R, Cotichia J. 2009. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. *Ann Otol Rhinol Laryngol* 118:292–298. <https://doi.org/10.1177/000348940911800410>.
33. Marks LR, Parameswaran GI, Hakansson AP. 2012. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. *Infect Immun* 80:2744–2760. <https://doi.org/10.1128/IAI.00488-12>.
34. Post JC, Hiller NL, Nistico L, Stoodley P, Ehrlich GD. 2007. The role of biofilms in otolaryngologic infections: update 2007. *Curr Opin Otolaryngol Head Neck Surg* 15:347–351. <https://doi.org/10.1097/MOO.0b013e3282b97327>.
35. Sanderson AR, Leid JG, Hunsaker D. 2006. Bacterial biofilms on the sinus mucosa of human subjects with chronic rhinosinusitis. *Laryngoscope* 116:1121–1126. <https://doi.org/10.1097/01.mlg.0000221954.05467.54>.
36. Hall-Stoodley L, Nistico L, Sambanthamoorthy K, Dice B, Nguyen D, Mershon WJ, Johnson C, Hu FZ, Stoodley P, Ehrlich GD, Post JC. 2008. Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in *Streptococcus pneumoniae* clinical isolates. *BMC Microbiol* 8:173. <https://doi.org/10.1186/1471-2180-8-173>.
37. Berg KH, Stamsås GA, Straume D, Håvarstein LS. 2013. Effects of low PBP2b levels on cell morphology and peptidoglycan composition in *Streptococcus pneumoniae* R6. *J Bacteriol* 195:4342–4354. <https://doi.org/10.1128/JB.00184-13>.
38. Damjanovic M, Kharat AS, Eberhardt A, Tomasz A, Vollmer W. 2007. The essential tacF gene is responsible for the choline-dependent growth phenotype of *Streptococcus pneumoniae*. *J Bacteriol* 189:7105–7111. <https://doi.org/10.1128/JB.00681-07>.
39. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DL, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43:D222–D226. <https://doi.org/10.1093/nar/gku1221>.
40. Biswas I, Drake L, Biswas S. 2007. Regulation of gbpC expression in *Streptococcus mutans*. *J Bacteriol* 189:6521–6531. <https://doi.org/10.1128/JB.00825-07>.
41. Kagami A, Okamoto-Shibayama K, Yamamoto Y, Sato Y, Kizaki H. 2007. One of two gbpC gene homologues is involved in dextran-dependent aggregation of *Streptococcus sobrinus*. *Oral Microbiol Immunol* 22:240–247. <https://doi.org/10.1111/j.1399-302X.2006.00347.x>.
42. Forsgren N, Lamont RJ, Persson K. 2010. Two intramolecular isopeptide bonds are identified in the crystal structure of the *Streptococcus gordonii* SspB C-terminal domain. *J Mol Biol* 397:740–751. <https://doi.org/10.1016/j.jmb.2010.01.065>.
43. Brooks W, Demuth DR, Gil S, Lamont RJ. 1997. Identification of a *Streptococcus gordonii* SspB domain that mediates adhesion to *Porphyromonas gingivalis*. *Infect Immun* 65:3753–3758.
44. Jakubovics NS, Strömberg N, van Dolleweerd CJ, Kelly CG, Jenkinson HF. 2005. Differential binding specificities of oral streptococcal antigen I/II family adhesins for human or bacterial ligands. *Mol Microbiol* 55:1591–1605. <https://doi.org/10.1111/j.1365-2958.2005.04495.x>.
45. Holmes AR, Gilbert C, Wells JM, Jenkinson HF. 1998. Binding properties of *Streptococcus gordonii* SspA and SspB (antigen I/II family) polypeptides expressed on the cell surface of *Lactococcus lactis* MG1363. *Infect Immun* 66:4633–4639.
46. Gipson IK, Spurr-Michaud S, Argüeso P, Tisdale A, Ng TF, Russo CL. 2003. Mucin gene expression in immortalized human corneal-limbal and conjunctival epithelial cell lines. *Invest Ophthalmol Vis Sci* 44:2496–2506. <https://doi.org/10.1167/iovs.02-0851>.
47. Williamson YM, Gowrisankar R, Longo DL, Facklam R, Gipson IK, Ades EP, Carlone GM, Sampson JS. 2008. Adherence of nontypeable *Streptococcus pneumoniae* to human conjunctival epithelial cells. *Microb Pathog* 44:175–185. <https://doi.org/10.1016/j.micpath.2007.08.016>.
48. Tukey JW. 1949. Comparing individual means in the analysis of variance. *Biometrics* 5:99–114. <https://doi.org/10.2307/3001913>.
49. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, de la Cruz F, Rocha EP. 2014. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res* 42:5715–5727. <https://doi.org/10.1093/nar/gku194>.
50. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ. 2014. Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol Evol* 6:741–753. <https://doi.org/10.1093/gbe/evu048>.
51. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28:i409–i415. <https://doi.org/10.1093/bioinformatics/bts386>.
52. Budroni S, Siena E, Dunning Hotopp JC, Seib KL, Serruto D, Nofroni C, Comanducci M, Riley DR, Daugherty SC, Angiuoli SV, Covacci A, Pizza M, Rappuoli R, Moxon ER, Tettelin H, Medini D. 2011. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc Natl Acad Sci U S A* 108:4494–4499. <https://doi.org/10.1073/pnas.1019751108>.
53. Eutsey RA, Powell E, Dordel J, Salter SJ, Clark TA, Korlach J, Ehrlich GD, Hiller NL. 2015. Genetic stabilization of the drug-resistant PMEN1 *Streptococcus pneumoniae* lineage by its distinctive DpnIII restriction-modification system. *mBio* 6:e00173-15. <https://doi.org/10.1128/mBio.00173-15>.
54. Zangari T, Wang Y, Weiser JN. 2017. *Streptococcus pneumoniae* transmission is blocked by type-specific immunity in an infant mouse model. *mBio* 8:e00188-17. <https://doi.org/10.1128/mBio.00188-17>.
55. Rodriguez JL, Dalia AB, Weiser JN. 2012. Increased chain length promotes pneumococcal adherence and colonization. *Infect Immun* 80:3454–3459. <https://doi.org/10.1128/IAI.00587-12>.
56. De Las Rivas B, García JL, López R, García P. 2002. Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-

- acetylglucosaminidase: the chain-dispersing murein hydrolase. *J Bacteriol* 184:4988–5000. <https://doi.org/10.1128/JB.184.18.4988-5000.2002>.
57. Bjørkeng EK, Hjerde E, Pedersen T, Sundsfjord A, Hegstad K. 2013. ICESluvan, a 94-kilobase mosaic integrative conjugative element conferring interspecies transfer of VanB-type glycopeptide resistance, a novel bacitracin resistance locus, and a toxin-antitoxin stabilization system. *J Bacteriol* 195:5381–5390. <https://doi.org/10.1128/JB.02165-12>.
 58. Davies D. 2003. Understanding biofilm resistance to antibacterial agents. *Nat Rev Drug Discov* 2:114–122. <https://doi.org/10.1038/nrd1008>.
 59. Huang J, Liang Y, Guo D, Shang K, Ge L, Kashif J, Wang L. 2016. Comparative genomic analysis of the ICESa2603 family ICEs and spread of erm(B)- and tet(O)-carrying transferable 89K-subtype ICEs in swine and bovine isolates in China. *Front Microbiol* 7:55. <https://doi.org/10.3389/fmicb.2016.00055>.
 60. Palmieri C, Magi G, Mingoia M, Bagnarelli P, Ripa S, Varaldo PE, Facinelli B. 2012. Characterization of a *Streptococcus suis* tet(O/W/32/O)-carrying element transferable to major streptococcal pathogens. *Antimicrob Agents Chemother* 56:4697–4702. <https://doi.org/10.1128/AAC.00629-12>.
 61. Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. <https://doi.org/10.1186/1471-2105-11-595>.
 62. van Selm S, van Cann LM, Kolkman MAB, van der Zeijst BAM, van Putten JPM. 2003. Genetic basis for the structural difference between *Streptococcus pneumoniae* serotype 15B and 15C capsular polysaccharides. *Infect Immun* 71:6192–6198. <https://doi.org/10.1128/IAI.71.11.6192-6198.2003>.
 63. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
 64. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
 65. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>.
 66. Posada D, Crandall KA. 1998. ModelTest: testing the model of DNA substitution. *Bioinformatics* 14:817–818. <https://doi.org/10.1093/bioinformatics/14.9.817>.
 67. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>.
 68. Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD. 2007. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol* 8:R103. <https://doi.org/10.1186/gb-2007-8-6-r103>.
 69. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 70. R Foundation for Statistical Computing. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
 71. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
 72. Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612. <https://doi.org/10.1093/nar/gkl315>.
 73. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29. <https://doi.org/10.1186/1471-2148-6-29>.
 74. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, Wright F. 2009. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25:126–127. <https://doi.org/10.1093/bioinformatics/btn575>.
 75. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>.