bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The *Streptococcus pneumoniae* Competence-Induced BriC Peptide Promotes Nasopharyngeal Colonization and Impacts Biofilm Development

Running Title: Competence-induced peptide regulates biofilms and *in vivo* colonization

Surya D. Aggarwal[1]
Rory Eutsey[1]
Jacob West-Roberts[1]
Arnau Domenech[2]
Wenjie Xu[1]
Iman T. Abdullah[3,4]
Aaron P. Mitchell[1]
Jan-Willem Veening[2]
Hasan Yesilkaya[3]
N. Luisa Hiller [*1,5]

[1]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland
[3]Department of Infection, Immunity & Inflammation, University of Leicester, Leicester, UK
[4]Department of Biology, College of Science, University of Kirkuk, Iraq
[5]Center of Excellence in Biofilm Research, Allegheny Health Network, Pittsburgh, PA, USA

[*]**Correspondence:**
N. Luisa Hiller
lhiller@andrew.cmu.edu

Keywords: *Streptococcus pneumoniae*, biofilms, competence, colonization, transformation, comparative genomics

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Streptococcus pneumoniae (pneumococcus) is an opportunistic pathogen that causes otitis media, sinusitis, pneumonia, meningitis and sepsis. The progression to this pathogenic lifestyle is preceded by asymptomatic colonization of the nasopharynx. This colonization is associated with biofilm formation; the competence (Com) pathway influences the structure and stability of biofilms. However, how the Com pathway is linked to biofilm formation is unknown. Here, we identified a new competence-induced gene, called *briC*, and show that its product promotes biofilm development and stimulates colonization in a murine model. We show that expression of *briC* is induced by the master regulator of competence, ComE. Whereas *briC* does not substantially influence early biofilm development on abiotic surfaces, it significantly impacts later stages of biofilm development. Specifically, *briC* expression leads to increases in biofilm biomass and thickness at 72h. Consistent with the role of biofilms in colonization, *briC* promotes nasopharyngeal colonization in the murine model. The function of BriC appears to be conserved across pneumococci, as comparative genomics reveal that *briC* is widespread across isolates. Surprisingly, strains from clinically important PMEN1 and PMEN14 lineages, which are widely associated with colonization, encode a longer *briC* promoter. This long form captures an instance of genomic plasticity and functions as a competence-independent expression enhancer that may serve as a precocious point of entry into this otherwise competence-regulated pathway. Moreover, overexpression of *briC* by the longer promoter fully rescues the *comE*-deletion induced biofilm defect *in vitro*, and partially *in vivo*. These findings indicate that BriC may bypass the influence of competence in biofilm development and that such a pathway may be active in a subset of pneumococcal lineages. In conclusion, *briC* is a part of the complex molecular network that connects signaling of the competence pathway to biofilm development and colonization.

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Bacteria form sessile communities termed biofilms, where they interact with each other to engage in collaborative and/or competitive behaviors (Hall-Stoodley et al., 2004). In *Streptococcus pneumoniae* (pneumococcus), these cell-cell interactions are commonly mediated by secreted peptides that interact with both producing and neighboring cells of the same species, and induce changes in gene regulation that result in altered phenotypes (Shanker and Federle, 2017). These dynamic pneumococcal biofilms occur in chronic otitis media, chronic rhinosinusitis and nasopharyngeal colonization (Blanchette-Cain et al., 2013; Hall-Stoodley et al., 2006; Hoa et al., 2009; Marks et al., 2012a; Oggioni et al., 2006; Sanderson et al., 2006).

The ability to form biofilms is a critical component of pneumococcal disease (Chao et al., 2015). Biofilms serve as reservoirs for acute infections (Bogaert et al., 2004). In the middle ear, cells released from a biofilm are thought to be responsible for recurrent episodes of infection (Hall-Stoodley et al., 2006). Bacterial cells released from nasopharyngeal biofilms can seed pneumococcal transmission between individuals by being incorporated into nasal shedding. Alternatively, these cells can disseminate to tissues causing mild to severe diseases, such as otitis media, pneumonia, and sepsis (Bogaert et al., 2004). Pneumococcal cells released from biofilms display increased virulence relative to their planktonic or biofilm counterparts, suggesting that chronic biofilms set the stage for the stimulation of a virulence program activated upon the dispersal of cells (Marks et al., 2013). Moreover, pneumococci in a biofilm display decreased susceptibility to antibiotics, and are recalcitrant to treatment (Marks et al., 2012a). Thus, biofilms are an important component of pneumococcal epidemiology in transmission, maintenance of asymptomatic colonization, and development of disease.

The transcriptional program required for the initiation and the growth of pneumococcal biofilms has been the subject of numerous investigations. It is clear that at least two quorum

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sensing (QS) signal transduction pathways are critical for biofilm development: Com and Lux (Oggioni et al., 2006; Trappetti et al., 2011a, 2011b, Vidal et al., 2011, 2013). The competence (Com) pathway has been the subject of intense investigation for decades (Alloing et al., 1998; Guenzi et al., 1994; Havarstein et al., 1995a; Håvarstein et al., 1996; Pestova et al., 1996; Peterson et al., 2004; Tomasz, 1965). Competence is activated by a classic two-component system where the extracellular competence stimulating peptide (CSP, encoded by *comC*) binds to the surface exposed ComD histidine kinase receptor, inducing its autophosphorylation and the subsequent transfer of the phosphate group to its cognate regulator, ComE (Havarstein et al., 1995a; Pestova et al., 1996). Activation of the Com pathway leads to increased expression of 5-10% of the pneumococcal genome in two main waves of gene expression (Peterson et al., 2004). The first wave of induction is carried out directly by ComE; it upregulates a subset of competence genes (early genes) that include *comAB*, *comCDE*, as well as the alternative sigma factor, *comX*. The second wave of competence induction is regulated by ComX; it leads to an increase in the levels of at least 80 genes (late genes), that subsequently modulate important phenotypes such as transformation, metabolism, fratricide and biofilm formation (Claverys et al., 2006; Martin et al., 2010; Peterson et al., 2004). This competence program is upregulated during biofilm mode of growth *in vitro*, during interactions with human epithelial cells, and in lungs and brain after intranasal and intracranial challenges respectively in murine infection models (Aprianto et al., 2016; Oggioni et al., 2006; Trappetti et al., 2011a). Importantly, in cell culture models, *comC* is required for biofilm development (Trappetti et al., 2011a; Vidal et al., 2013). Thus, activation of the Com pathway is important for productive biofilm formation and critical for pneumococcal infection and adaptation.

The Lux QS system also plays a role in biofilm formation. In this system, Lux QS is controlled by the AI-2 autoinducer, which is secreted and sensed by both Gram-positive and Gram-negative species. LuxS is a node in the regulation of competence, fratricide, and biofilm

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

development (Armbruster et al., 2010; Vidal et al., 2011). Lux upregulates competence via ComE and ComX (Vidal et al., 2011). It contributes to bactericidal activity via upregulation of the choline binding murein hydrolase (CbpD*)*. Through lysis, this bacteriocidal activity increases the levels of extracellular DNA, which is a key ingredient in the extracellular polymeric substance (EPS) that makes up the biofilm. Thus, the Com and Lux systems provide the molecular framework to coordinate multi-cellular bacterial communities to form and develop robust biofilms during infection.

Whereas the role of Com signaling in biofilm development is well established, the molecules that connect competence to biofilms are poorly understood (Blanchette-Cain et al., 2013; Marks et al., 2012b; Oggioni et al., 2006; Vidal et al., 2011). In this study, we identify one such molecule that links competence and biofilms. We characterize the gene encoding BriC (**b**iofilm **r**egulating peptide **i**nduced by **C**ompetence), a novel colonization factor in the competence pathway. Levels of *briC* are regulated by ComE, and increased *briC* levels enhance transformation and biofilm development and promote nasopharyngeal colonization.

Results

Identification of a competence-regulated Gly-Gly peptide

We have identified the gene encoding a putative secreted peptide that is co-regulated with competence (*spd_0391* (D39); *spr_0388* (R6); *sp_0429* (TIGR4)). Based on the results presented in this manuscript, we have termed it **b**iofilm-**r**egulating peptide **i**nduced by **C**ompetence (BriC). BriC was identified in our previously described *in silico* screen designed to capture cell-cell communication peptides in the pneumococcal genome (Cuevas et al., 2017). The known double glycine (Gly-Gly) streptococcal peptides are exported and proteolytically processed by dedicated ABC transporters that recognize N-terminal sequences with the Gly-Gly leader peptide (LSXXELXXIXGG) (Havarstein et al., 1995b). To identify novel secreted

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pneumococcal peptides, we performed a computational analysis to search for proteins with N-termini that contain a Gly-Gly leader. To define this leader, we employed multiple iterations of Multiple Expectation Maximization for Motif Elicitation (MEME) on an input set that consisted of the alleles of two exported Gly-Gly peptides, the signaling molecule CSP and the bacteriocin BIP (Dawid et al., 2007; Havarstein et al., 1995a). This output consists of a position dependent probability matrix that captures the length and positional variability at each residue of the Gly-Gly motif. Next, we searched for this motif in a database of sixty streptococcal genomes, using the Motif Alignment and Search Tool (MAST). As described in our previous work, we defined a predicted secretome consisting of twenty-five sequence clusters, one of which corresponds to BriC.

To identify genes co-regulated with *briC*, we performed transcriptional studies using a NanoString probe set that reports on the abundance of the *briC* transcript as well as transcripts encoding a subset of pneumococcal regulators and cell wall proteins. We assessed the levels of *briC* transcript *in vitro* and *in vivo*. *In vitro* expression was measured by screening RNA extracted from mid-log planktonic cultures of a laboratory strain (R6-derivative (R6D)). *In vivo* expression was evaluated by analysis of middle-ear effusions recovered from chinchillas infected with a clinical PMEN1 strain. The mRNA levels of the *briC* were positively associated with *comC* and *comE in vitro* (strain R6D: R^2=0.61 and 0.79, respectively) *and in vivo* (strain PN4595-T23: R^2=0.92 and 0.88, respectively). It is noteworthy that when performing the first comprehensive cataloguing of CSP-regulated genes, Peterson and colleagues also observed changes in *briC* levels, however the association between *briC* and CSP was below the statistical threshold (Peterson et al., 2004). Thus, our gene expression analysis suggests that *briC* is induced by competence.

To directly test whether *briC* is a competence-regulated peptide, we employed fusion of the *briC* promoter to the *lacZ* reporter (R6 P*briC-lacZ*). Stimulation of the signal transduction

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

system that initiates competence by addition of CSP led to an induction of the β-galactosidase activity by over twenty-five-fold (**Fig. 1**). Induction of the *briC* promoter was specific to the CSP pherotype encoded by strain R6. The β-galactosidase activity was observed upon addition of CSP1, the CSP pherotype from strain R6, but not upon addition of the non-cognate CSP2 pherotype (**Fig. S1**). Thus, we conclude that *briC* is a competence-responsive gene.

Levels of *briC* transcripts are directly regulated by ComE

Our *in silico* analysis of the *briC* promoter in strains R6 and R6D revealed the presence of a ComE-binding site. ComE binds a well-defined sequence consisting of two imperfect direct repeats of nine nucleotides separated by a gap of twelve or thirteen base pairs (Ween et al., 1999). Our analysis of the putative *briC* promoter across thirty-five pneumococcal strains revealed an excellent match to the ComE-binding box (**Fig. 2A**). To further investigate the association between ComE and *briC*, we tested whether CSP-induction of *briC* requires ComE. We compared the CSP-induction of *briC* in a wild-type (R6D WT) strain to that of an isogenic *comE*-deletion mutant (R6DΔ*comE*), using qRT-PCR analysis. In WT cells, the addition of CSP triggered a significant increase in levels of *briC* at 10 minutes post-addition, with levels slowly decreasing by 15 minutes (**Fig. 2B**). This trend follows the temporal pattern observed for the levels of *comE* that has been associated with genes under direct controls of ComE (Peterson et al., 2004). In contrast, the transcript levels of *briC* were unaffected by CSP addition in the Δ*comE* strain, indicating that the expression of *briC* requires ComE (**Fig. 2B**). These results strongly suggest that *briC* is directly regulated by ComE.

BriC plays a key role in biofilm development

To investigate whether expression of *briC* plays a role in biofilm development, we compared biofilm formation across WT (R6D WT), *briC* deletion mutant (R6DΔ*briC*), and *briC*

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

185 complemented (R6DΔ*briC*::*briC*) strains grown on an abiotic surface at 24h and 72h post-
186 seeding. No difference was observed in biofilm biomass and thickness at 24h post-seeding,
187 suggesting that expression of *briC* does not contribute to early stages of biofilm formation (**Fig.**
188 **3A, B**). In contrast, at 72h post-seeding, Δ*briC* biofilms displayed significantly reduced
189 biomass and thickness when compared to WT (**Fig. 3C, D**). Further, biofilms with Δ*briC::briC*
190 cells restored the WT phenotype at this time-point (**Fig. 3C, D**). The indistinguishable biofilm
191 parameters of WT and Δ*briC* cells at 24h post-seeding suggests that there is no fitness-related
192 growth difference between the strains and indicates that the biofilm defect is biologically
193 relevant. These findings suggest that *briC* contributes to late biofilm development.
194
195 **BriC is widely distributed across pneumococcal strains**
196 To investigate the prevalence of *briC*, we investigated its distribution across the
197 pneumococcus and related streptococci. To place the distribution in the context of phylogeny,
198 we used a published species tree generated from a set of fifty-five genomes (Cuevas et al.,
199 2017; Kadam et al., 2017) (**Table S1**). The genomes encompass thirty-five pneumococcal
200 genomes that span twenty-nine multi-locus sequence types as well as eighteen serotypes and
201 nontypeable strains; eighteen genomes from related streptococcal species that also colonize
202 the human upper respiratory tract, namely *S. pseudopneumoniae*, *S. mitis*, and *S. oralis*; and
203 finally, two distantly related *S. infantis* strains as an outgroup. Using tblastn, we identified three
204 distinct polymorphic groups within this set of genomes, one is dominant across pneumococcus
205 and two are present in related streptococcal species and one pneumococcal strain (**Fig. 4, Fig.**
206 **S2 and File S1**).
207 In pneumococcus, *briC* is present in a majority of the strains (thirty-four out of thirty-five)
208 (**Fig. 4**). The extent of *briC* sequence conservation is high; nonetheless, the pneumococcal
209 strains display two widespread polymorphisms (**Fig. S2**). The first is at the C-terminus where

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

position –2 encodes either an alanine or a threonine. The second is in the putative N-terminal secretion sequence where the sequences encode either an asparagine or a glutamic acid. In other streptococci, *briC* homologs are encoded within a subset of *S. mitis* and *S. oralis* strains (**Fig. 4**). In addition, a strain from *S. oralis* subspecies tigurinus encodes two distinct copies of *briC*. The phylogenetic distribution of *briC* supports a conserved role across pneumococci and a subset of related streptococcal species.

Inter-strain differences in the putative *briC* promoter are associated with diverse regulation of *briC* in clinically important lineages

As described above, the *briC* promoter contains a ComE-binding box in all pneumococcal strains. Remarkably, a subset of strains encode for an extra 104 bp within the region upstream of *briC* (**Fig. 4**, **File S2**). The additional nucleotides are located after the ComE-binding site and before the transcriptional start site. In our curated sequences, the longer promoter is present in strains from the clinically important PMEN1 and PMEN14 lineages (**Fig. 4**, **Table S1**). To expand beyond our curated set, we investigated the distribution of these additional nucleotides in a set of 3,529 genome sequences obtained from two large pneumococcal studies (Chewapreecha et al., 2014; Croucher et al., 2013). We find that the longer promoter is present in 100% of the PMEN1 and PMEN14 strains in this expanded set.

To investigate how this genomic difference influences *briC* expression, we generated a LacZ reporter strain. The 263bp upstream of *briC* from the PMEN1 strain, PN4595-T23, were fused to *lacZ* to produce the P*briC*$_{long}$-*lacZ* reporter, and its reporter activity was compared to that of the P*briC*-*lacZ* generated with the fusion of 159bp upstream of *briC* obtained from strain R6. The function of these additional nucleotides could be strain-dependent (for example, via a regulator encoded only in a subset of strains). Thus, these reporter constructs were tested in both the R6 and the PMEN1 backgrounds, in the absence and presence of CSP treatment

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(**Fig. 5A, B)**. The presence of additional nucleotides dramatically increased the basal levels of *briC* in the absence of CSP, and this increase was observed in both R6 and PMEN1. Furthermore, both constructs were induced upon the addition of CSP. These findings suggest that the extra nucleotides serve as an expression enhancer; they increase levels of *briC* transcripts and this increase is CSP-independent. Thus, in some lineages, *briC* appears to be under the control of both CSP-dependent and CSP-independent regulation.

Expression of *briC* driven by the longer promoter bypasses the impact of competence induction in biofilm development

Next, we investigated the biological impact of the natural variations in the *briC* promoter on biofilm development. It has been well established that competence promotes biofilm development. Specifically, deletion of the *comC* (encodes CSP) and *comD* (encodes histidine kinase of competence TCS) genes lead to a reduction in *in vitro* biofilms in strains D39 and TIGR4 (Oggioni et al., 2006; Trappetti et al., 2011a). We have shown that the longer promoter of *briC* serves as an expression enhancer, wherein it drives *briC* expression in a competence-independent way (**Fig. 5**). Thus, we measured biofilm biomass and thickness for R6D WT, R6DΔ*comE*, and a R6DΔ*comE* strain where *briC* expression is driven by the longer promoter (R6DΔ*comE*::P*briC*$_{long}$-*briC*). At 72h post-seeding, a time-point where *briC* was found to show observable differences in biofilm parameters, we found that relative to the WT strain, the R6DΔ*comE* displayed approximately 15% and 23% reduction in biofilm biomass and maximum thickness respectively. These defects were fully rescued by increased expression of *briC* in the R6DΔ*comE*::P*briC*$_{long}$-*briC* strain (**Fig. 6A,B**). Thus, expression of *briC* is sufficient to rescue a competence-dependent biofilm defect. Further, these data suggest that the natural variations in the *briC* promoter are physiologically relevant.

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Since BriC is associated with the competence pathway and is able to rescue the biofilm defects associated with Com signaling, we investigated whether Com-associated transporters play a role in exporting BriC. The bacteriocin inducing peptide, which is a Gly-Gly peptide, is exported into the extracellular milieu via two ABC transporters, ComAB and BlpAB (Kjos et al., 2016; Wholey et al., 2016). In strains R6 and R6D, BlpAB is not functional due to a frameshift mutation that leads to an early stop codon (Son et al., 2011). Thus, we hypothesized that as a Gly-Gly peptide co-expressed with genes of the competence pathway, BriC may be exported via the ComAB transporter. We tested this hypothesis by investigating whether the role of BriC in contributing to biofilm development is impaired in a *comAB*-deletion mutant. We compared the biofilm biomass and thickness of a *comE/comAB*-double deletion mutant overexpressing *briC* (R6DΔ*comE*Δ*comAB::*P*briC*$_{long}$-*briC*) to that of a *comE*-deletion strain overexpressing *briC* (R6DΔ*comE::*P*briC*$_{long}$-*briC*) and a *comE*-deletion strain (R6DΔ*comE*) at 72h post-seeding. We found that the ability of the strain overexpressing *briC* to rescue the defect of a *comE*-deletion strain was hindered when the *comAB* transporter was also deleted (**Fig. S3**). However, the biofilm biomass and thickness of Δ*comE*Δ*comAB::*P*briC*$_{long}$-*briC* were still significantly higher than that of Δ*comE* cells. These findings suggested that ComAB may contribute to the secretion of BriC. However, our results did not exclude the possibility of other genes being involved in the export process.

BriC is important for *in vivo* colonization

During nasopharyngeal colonization, pneumococci form biofilms and upregulate the competence pathway. Thus, we investigated the role of *briC* in nasopharyngeal colonization using an experimental murine colonization model. Our *in vitro* investigations have been performed using strain R6D strain, which is defective in colonization due to the absence of a capsule. Thus, we performed colonization experiments with the serotype 2 D39 strain, which is

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the ancestor of strain R6 (Lanie et al., 2007). Mice were colonized with D39 WT, the *briC*-deletion mutant (D39Δ*briC*) or the *briC*-complemented (D39Δ*briC*::*briC*) strains. Comparison of the number of bacteria in nasal lavages immediately after inoculation revealed that mice in the three cohorts received the same number of bacteria. In contrast, nasal lavages at three and seven days post-inoculation revealed decreased levels of D39Δ*briC* relative to WT in the nasal wash (**Fig. 7A**). Furthermore, the WT levels were restored in the complemented strain (**Fig. 7A**). These findings indicate that *briC* encodes a novel colonization factor.

In *in vitro* biofilms, overexpression of *briC* driven by the long version of the promoter was found to restore the competence-dependent defect in biofilm development. Thus, we investigated whether expression driven by this *briC* promoter can restore the colonization defect associated with a *comE* deletion-mutant *in vivo*. We found that additions of this longer *briC* promoter in Δ*comE* cells (D39Δ*comE::*P*briC*$_{long}$-*briC)* partially rescues the colonization defect of the D39Δ*comE* strain. The numbers of bacterial cells of strain D39Δ*comE::*P*briC*$_{long}$-*briC* recovered from the nasal lavages at both three and seven days post-inoculation were significantly higher than the numbers of D39Δ*comE* cells recovered, but less than that of the D39 WT (**Fig. 7B)**. Thus, we conclude that BriC is a key contributor to the competence-induced stimulation of nasopharyngeal colonization observed in strain D39. Further, natural variations leading to a longer *briC* promoter appear to dampen the impact of competence in colonization.

Discussion

An important component of pneumococcal pathogenesis is its ability to form complex biofilm structures. Pneumococci in a biofilm mode of growth display decreased sensitivity to antibiotics and increased resistance to host immune responses (Marks et al., 2012a). These

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

properties make the bacteria recalcitrant to treatment and highlight the need to better understand the molecular mechanisms that drive biofilm development. Activation of the competence pathway is critical for biofilm development. Previous *in vitro* studies have demonstrated that while cell-adherence and early biofilm formation is competence-independent, an intact competence system is required in the later stages of biofilm development. It was shown that the competence pathway positively influences structure and stability of late stage biofilms (Trappetti et al., 2011a). However, the molecules downstream of competence activation by ComDE that regulate biofilm development remain poorly understood. In this study, we present BriC, a previously uncharacterized peptide, that we show is regulated by competence and plays a role in promoting biofilm development and nasopharyngeal colonization.

We have presented extensive evidence that *briC* is a competence regulated gene. We have shown that induction of *briC* is triggered by addition of CSP and requires ComE. Further, we have also shown that the *briC* promoter encodes the consensus ComE-binding box, and that *briC* expression follows the temporal pattern described for genes directly regulated by ComE. Previously, Peterson and colleagues have performed microarray studies to identify pneumococcal genes differentially regulated upon CSP stimulation (Peterson et al., 2004). They had categorized these genes into three categories - early genes regulated by ComE, late genes regulated by the ComX alternative sigma factor, or delayed genes that appeared to be stress-related. In their study, *briC* was found to be upregulated in a pattern consistent with early genes. However, the upregulation was not found to be statistically significant, and this study is the first validation of *briC* as a competence-regulated peptide.

We have provided evidence that *briC* stimulates biofilm development on abiotic surfaces and promotes nasopharyngeal colonization in a murine model. These findings are consistent with studies that show that pneumococcal biofilms contribute to colonization. Colonization of

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the upper respiratory tract is a requisite for pneumococcal dissemination to distant anatomical sites and subsequent disease (Bogaert et al., 2004). These sessile communities serve as a source of pneumococcal cells with an activated virulence-associated transcription program. That is, when compared to cells originating from a planktonic mode of growth, those originating from a biofilm mode of growth are more likely to cause disease upon infecting other tissues (Marks et al., 2013). In this manner, increased biofilm development likely heightens the risk for disease. Biofilms and competence are also associated with transformation efficiency. We have observed a mild but significant decrease in the transformation efficiency of *briC*-deletion mutants relative to WT R6D cells (**Fig. S4**). Finally, colonization of the upper respiratory tract is also a reservoir for pneumococcal transmission. Transmission occurs when cells migrate from the nasopharynx of one host to that of another. Thus, BriC's contribution to colonization may influence both disease severity and transmission.

While it has been established that CSP contributes to biofilm development, the competence-dependent genes that regulate biofilm development are not well understood (Oggioni et al., 2006; Trappetti et al., 2011a). Our finding that increased levels of *briC* can fully rescue biofilm defects from a *comE* deletion mutant *in vitro*, and partially rescue its colonization defects *in vivo* suggests that *briC* expression may bypass the requirement for competence in biofilm development. ComE is a key regulator of competence whose activity is required to regulate approximately 5-10% of the genome, and as such deletion of *comE* is expected to have substantial global consequences (Peterson et al., 2004). In this context, it is remarkable that overexpression of one gene (*briC*) in the *comE*-deletion mutant was able to significantly improve colonization in the murine model. These findings strongly suggested that BriC is a molecular link between competence, biofilm development, and colonization.

Our data suggests that many strains have multiple inputs to the regulation of *bric*. Shared across all strains is the regulation by ComE, the key regulator of the competence

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pathway. Competence is responsive to environmental cues, such as changes in cell density, pH, mutational burden in cells, and exposure to antibiotics (Claverys et al., 2006; Gagne et al., 2013; Hakenbeck and Chhatwal, 2007; Moreno-Gámez et al., 2017). Conversely, competence is inhibited by the degradation of CSP via the activity of the CiaHR TCS and the serine protease, HtrA (Mascher et al., 2003; Sebert et al., 2005). Factors altering competence will also alter *bric* levels due to its competence-dependent induction. Our comparative genomics suggest that a subset of pneumococcal lineages may encode an additional *briC*-regulatory element. Specifically, the *briC* promoter differs across strains, in that a subset of lineages encodes a promoter that is longer by 104bp (P*briC*$_{long}$) and has higher basal levels of *briC* expression. This longer promoter is constitutively active, even when competence is off.

The longer promoter is encoded in the vast majority of strains from the PMEN1 lineage (Spanish-USA) and the PMEN14 (Taiwan-19F) lineages. These lineages are prominent in the clinical setting; they are multi-drug resistant and pandemic (Croucher et al., 2011, 2014; Wyres et al., 2012). This additional competence-independent regulation of the longer promoter may provide promoter-binding sites for additional regulators or reflects consequences of positional differences for the existing promoter binding sites. Our biofilm and colonization experiments suggest that encoding the longer *briC* promoter has functional consequences. We conclude that the response of *briC* to competence is ubiquitous, but that additional lineage-specific factors influence *briC* regulation and downstream phenotypic consequences.

We propose a model where *briC* encodes a signaling molecule with a role in biofilm development and colonization. First, the transcription of *briC* is induced by ComE through competence signal transduction pathway in all lineages, and possibly by additional regulator(s) in a subset of lineages. Once this Gly-Gly peptide is produced, we propose that it is exported through ABC transporters, a process in which ComAB plays a role. Based on a bioinformatic comparison with other Gly-Gly peptides we suggest that BriC is cleaved into its active form

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(BRIC) during export. It is tempting to speculate that BRIC is a new member of the expanding set of pneumococcal secreted peptides that signal to neighboring cells promoting population-level behaviors.

Materials & Methods

Bacterial strains & growth conditions

Three wild-type (WT) *Streptococcus pneumoniae* strains were used for this experimental work. The majority of studies were performed on a penicillin-resistant derivative of R6 (R6D); this strain was generated from a cross where parental strain R6 was recombined with Hungary19A and the recombinant was selected for penicillin resistance (Severin et al., 1996). The *briC* allele in R6D is identical to the allele present in the parental R6. This laboratory strain is non-encapsulated and does not colonize mice, thus mice colonization experiments were performed with the serotype 2 D39 strains (GenBank CP000410)(Paixão et al., 2015). The D39 strain contains the same *briC* allele as is present in the R6D strain, which has been used for most of the work in this study. Finally, for a representative of PMEN1, we used the carriage isolate, PN4595-T23 (GenBank ABXO01) graciously provided by Drs. Alexander Tomasz and Herminia deLancastre (Hiller et al., 2011).

Colonies were grown from frozen stocks by streaking on TSA-II agar plates supplemented with 5% sheep blood (BD BBL, New Jersey, USA). Colonies were then used to inoculate fresh Columbia broth (Remel Microbiology Products, Thermo Fisher Scientific, USA) and incubated at 37°C and 5% CO_2 without shaking. When noted, colonies were inoculated into acidic Columbia broth prepared by adjusting the pH of Columbia broth to 6.6 using 1M HCl. Acidic pH was used to inhibit the endogenous activation of competence.

Construction of mutants

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The mutant strains (R6DΔ*briC and* PN4595Δ*briC*) were constructed by using site-directed homologous recombination to replace the region of interest with erythromycin-resistance gene (*ermB*) or kanamycin-resistance gene (*kan*), respectively (**Table S2**). *Kan* and spectinomycin-resistance gene (*aad9*) were used to construct Δ*comE* strains in R6D and PN4595-T23 respectively. Briefly, the transformation construct was generated by assembling the amplified flanking regions and antibiotic resistance cassettes. ~2kb of flanking regions upstream and downstream of the gene of interest was amplified from parental strains by PCR using Q5 2x Master Mix (New England Biolabs, USA). The antibiotic resistance genes, *kan* and *aad9* were amplified from *kan-rpsL* Janus Cassette and pR412, respectively (provided by Dr. Donald A. Morrison), and *ermB* was amplified from *S. pneumoniae* SV35-T23. SV35-T23 is resistant to erythromycin because of the insertion of a mobile element containing *ermB* (Hiller et al., 2011). These PCR fragments were then assembled together by sticky-end ligation of restriction enzyme-cut PCR products. The deletion mutant in R6D is an overexpressor of the downstream peptide (*spr_0389*).

The *briC* complemented and overexpressor strains were generated using constructs containing the CDS of *briC* along with either its entire native promoter region or overexpressing promoter respectively, ligated at its 3' end with a kanamycin resistance cassette. The promoters used to overexpress *briC* included either the constitutive *amiA* promoter, or P*briC*$_{long}$. These were assembled with the amplified flanking regions by Gibson Assembly using NEBuilder HiFi DNA Assembly Cloning Kit. The construct was introduced in the genome of R6D downstream of the *bga* region (without modifying *bga*), a commonly employed site (Zähner and Hakenbeck, 2000). Primers used to generate the constructs are listed in **Table S3**. The R6DΔ*briC::briC* is also an overexpressor of the downstream peptide (*spr_0389*). The R6DΔ*comE::*P*briC*$_{long}$-*briC* strain was constructed by replacing *comE* with spectinomycin resistant cassette in the R6D P*briC*$_{long}$-*briC* strain. *comAB*-deletion mutant in a *briC*

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

433 overexpressor R6D genomic background strain (R6DΔ*comAB::*P*briC*$_{long}$-*briC*) was constructed
434 by transforming the R6D*briC*-OE with the genomic DNA of ADP226. ADP226 is a strain from
435 the D39 genomic background with *comAB* replaced by erythromycin resistance cassette. To
436 make the construct, the flanking regions and erythromycin resistance cassette were
437 amplified, and then assembled together by sticky-end ligation of restriction enzyme-cut PCR
438 products. The construct was then transformed into D39 ADP225 (unpublished) and selected
439 on Columbia blood agar supplemented with 0.25 μg mL^{-1} erythromycin.

440 The D39 *briC* deletion mutant (D39Δ*briC*), *briC* complemented (D39Δ*briC::briC*), *comE*
441 deletion mutant (D39Δ*comE*), and *briC* overexpressor in *comE* deletion background
442 (D39Δ*comE*::P*briC*$_{long}$-*briC*) strains were generated by transformation with the corresponding
443 constructs amplified from R6D.

444

445 **Construction of *lacZ* fusions**

446 Chromosomal transcriptional *lacZ*-fusions to the target promoters were generated to
447 assay promoter activity. These *lacZ*-fusions were generated via double crossover homologous
448 recombination event in the *bgaA* gene using modified integration plasmid pPP2. pPP2 was
449 modified by introducing *kan* in the multiple cloning site, in a direction opposite to *lacZ*. The
450 modified pPP2 was transformed into *E. coli* TOP10. The putative *briC* promoter regions were
451 amplified from R6 and PN4595-T23 strains, and modified to contain KpnI and XbaI restriction
452 sites, which were then assembled in the modified pPP2 plasmid by sticky-end ligation of the
453 enzyme digested products. These plasmids were transformed into *E. coli* TOP10 strain, and
454 selected on LB (Miller's modification, Alfa Aesar, USA) plates, supplemented with ampicillin
455 (100μg/ml). These plasmids were then purified by using E.Z.N.A. Plasmid DNA Mini Kit II
456 (OMEGA bio-tek, USA), and transformed into pneumococcal strains R6 and PN4595-T23 and
457 selected on Columbia agar plates supplemented with kanamycin (150μg/ml).

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

458

459 **Bacterial transformations**

460 For all bacterial transformations to generate mutants, target strains (R6D or D39) were
461 grown in acidic Columbia broth, and 1µg of transforming DNA along with 125µg/mL of CSP1
462 (sequence: EMRLSKFFRDFILQRKK; purchased from GenScript, NJ, USA) was added to them
463 when the cultures reached an OD_{600} of 0.05, followed by incubation at 37°C. After 2 hours, the
464 treated cultures were plated on Columbia agar plates containing the appropriate antibiotic;
465 erythromycin (2µg/ml), or kanamycin (150µg/ml). Resistant colonies were cultured in selective
466 media, and the colonies confirmed using PCR. Bacterial strains generated in this study are
467 listed in **Table S2**.

468 For transformation efficiency experiments, R6D strain was grown in acidic Columbia
469 broth until it reached an OD_{600} of 0.05. At this point, number of viable cells was counted by
470 plating serial dilutions on TSA-blood agar plates. Transformations were carried out by adding
471 either 100ng or 500ng of transforming DNA in the media supplemented with 125µg/mL of
472 CSP1 and incubated at 37°C for 30mins. For transforming DNA, we used either genomic DNA
473 or PCR products. The donor DNA contained spectinomycin-resistance gene (*aad9*) in the inert
474 genomic region between *spr_0515* and *spr_0516*. This construct was generated in PN4595-
475 T23, $spec^R$, followed by its amplification and transformation into R6D and Taiwan-19F strains
476 (Sp3063-00). The genomic DNA was extracted from Taiwan-19F, $spec^R$ strain. The purified
477 linear DNA was an amplimer of the region from R6D. After 30 minutes, the cultures were
478 plated on Columbia agar plates containing spectinomycin (100µg/ml), incubated overnight, and
479 colonies were counted the next day.

480

481 **RNA extraction**

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RNA extraction consists of sample collection, pneumococcal cell lysis, and purification of RNA. For qRT-PCR analysis, the strains (R6D and R6DΔ*comE*) were grown to an OD_{600} of 0.3 in acidic Columbia broth, followed by CSP1 treatment for 0, 10, or 15 minutes. For *in vitro* transcriptomic analysis using NanoString Technology, the R6D strain was grown to an OD_{600} of 0.1 in Columbia broth (in one experimental set, the samples were grown in sub-lethal concentration of penicillin (0.8µg/ml) for an hour). RNA was collected in RNALater (Thermo Fisher Scientific, USA) to preserve RNA quality and pelleted. For the *in vivo* experiments, the RNA was extracted from middle-ear chinchilla effusions infected with PN4595-T23 and PN4595-T23Δ*comE* strains, and preserved by flash freezing the effusion. In all the samples, the pneumococcal cell lysis was performed by re-suspending the cell pellet in an enzyme cocktail (2mg/ml proteinase K, 10mg/ml lysozyme, and 20µg/ml mutanolysin), followed by bead beating with glass beads (0.5mm Zirconia/Silica) in FastPrep-24 Instrument (MP Biomedicals, USA). Finally, RNA was isolated using the RNeasy kit (Genesee Scientific, USA) following manufacturer's instructions. For analysis with the NanoString, which does not require pure DNA, the output from the RNeasy kit was loaded on the machine without further processing. For analysis using qRT-PCR, contaminant DNA was removed by treating with DNase (2U/µL) at 37°C for at least 45 mins. The RNA concentration was measured by NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA) and its integrity was confirmed on gel electrophoresis. The purity of the RNA samples was confirmed by the absence of a DNA band on an agarose gel obtained upon running the PCR products for the samples amplified for *gapdh*.

NanoString Technology for transcriptional analysis

nCounter Analysis System from NanoString Technology provides a highly sensitive platform to measure gene expression both *in vitro* and *in vivo*, as previously described (Geiss

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al., 2008). Probes used in this study were custom-designed by NanoString Technology, and included housekeeping genes *gyrB* and *metG*, which were used as normalization controls. 5µL of extracted RNA samples were hybridized onto the nCounter chip following manufacturer's instructions. RNA concentration ranged from 80-200ng/µL for *in vivo* samples, and between 60-70ng total RNA for *in vitro* samples. A freely available software from manufacturers, nSolver, was used for quality assessment of the data, and normalization. The RNA counts were normalized against the geometric mean of *gyrB* and *metG* (Carvalho et al., 2011; Kim et al., 2013). Pearson's Correlation Coefficient was used to estimate correlation in the expression levels of different genes.

qRT-PCR for transcriptional analysis

High quality RNA was used as a template for first-strand cDNA synthesis SuperScript VILO synthesis kit (Invitrogen). The resulting product was then directly used for qRT-PCR using PerfeCTa SYBR Green SuperMix (Quantabio, USA) in an Applied Biosystems 7300 Instrument (Applied Biosystems, USA).16S rRNA counts were used for normalization. The raw data was then run through LinregPCR for expression data analysis, where the output expression data is displayed in arbitrary fluorescence units (N_0) that represent the starting RNA amount for the test gene in that sample (Ramakers et al., 2003; Ruijter et al., 2014). Fold-change relative to WT was then calculated for each individual experiment.

β-galactosidase assay

β-galactosidase assays were performed as previously described (Miller, JH, 1972) using cells that were grown in acidic Columbia broth to exponential phase. Cells were either left untreated, or independently treated with CSP1 (EMRLSKFFRDFILQRKK), CSP2 (EMRISRIILDFLFLRKK) or PhrA (LDVGKAD) (Kadam et al., 2017) (Genscript, USA) for 30 minutes and processed for analysis.

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

532

533 **Biofilm formation assay**

534 Pneumococcal cultures grown in Columbia broth were used to seed biofilms on abiotic
535 surfaces. When the cultures reached an OD_{600} of 0.05, each bacterial strain was seeded on
536 35MM glass bottom culture dishes (MatTek Corporation, USA). To promote biofilm growth, the
537 plates were incubated at 37°C and 5% CO_2. Every 24 hours, the supernatant was carefully
538 aspirated, followed by addition of the same volume of pre-warmed Columbia broth at one-fifth
539 concentration. The biofilm samples were fixed at two time-points: 24 and 72 hrs. For fixing, the
540 supernatants were carefully aspirated, and biofilms were washed thrice with PBS to remove
541 non-adherent and/or weakly adherent bacteria. Subsequently, biofilms were fixed with 4% PFA
542 (Electron Microscopy Sciences), washed three times with PBS, and prepared for confocal
543 microscopy.

544

545 **Confocal microscopy & quantification of biofilms**

546 Fixed biofilms were stained with SYTO59 Nucleic Acid Stain (Life Technologies, USA)
547 for 30 minutes, washed three times, and preserved in PBS buffer for imaging. Confocal
548 microscopy was performed on the stage of Carl Zeiss LSM-880 META FCS, using 561nm
549 laser line for SYTO59 dye. Stack were captured every 0.46 µm, imaged from the bottom to the
550 top of the stack until cells were visible, and reconstructed in Carl Zeiss black edition and
551 ImageJ. The different biofilm parameters (biomass, maximum thickness, and average
552 thickness over biomass) were quantified using COMSTAT2 plug-in available for ImageJ
553 (Heydorn et al., 2000). For depiction of representative reconstructed Z-stacks, the color levels
554 were adjusted using GNU Image Manipulation Program (GIMP). The colors were adjusted to
555 the same levels in an experiment across all the different conditions.

556

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

557 ***In vivo* transcriptomic analysis using chinchilla OM model**

558 All chinchilla experiments were conducted with the approval of Allegheny-Singer
559 Research Institute (ASRI) Institutional Animal Care and Use Committee (IACUC) A3693-
560 01/1000. Research grade young adult chinchillas (*Chinchilla lanigera*) weighing 400-600g were
561 acquired from R and R Chinchilla Inc., Ohio. Chinchillas were maintained in BSL2 facilities and
562 experiments were done under subcutaneously injected ketamine-xylazine anesthesia (1.7mg/kg
563 animal weight for each). Chinchillas were infected with 100 CFUs in 100µL of *S. pneumoniae*
564 PN4595-T23 by transbullar inoculation within each middle ear. For RNA extraction, chinchillas
565 were euthanized 48h post-inoculation of pneumococcus, and a small opening was generated
566 through the bulla to access the middle ear cavity. Effusions were siphoned out from the middle
567 ear and flash frozen in liquid nitrogen to preserve the bacterial RNA. Animals were euthanized
568 by administering an intra-cardiac injection of 1mL potassium chloride after regular sedation.
569
570 **Murine colonization model**
571 The role of *briC* in experimental pneumococcal colonization was assessed as previously
572 described (Al-Bayati et al., 2017; Kahya et al., 2017). For this, 10 weeks old female CD1 mice
573 (Charles River), weighing approximately 30-35 g were anesthetized with 2.5% isoflurane over
574 oxygen (1.5 to 2 liter/min), and administered intranasally with approximately $1X10^5$ CFU/mouse
575 in 20µl PBS. At predetermined time intervals, a group of 5 mice were euthanized by cervical
576 dislocation, and the nasopharyngeal lavage of each animal was obtained using 500µl PBS. The
577 pneumococci in nasopharyngeal wash were enumerated by plating the serial dilutions onto blood
578 agar plates.
579
580 **Statistical Tests**

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

581 The statistical differences among different groups were calculated by performing ANOVA
582 followed by Tukey's post-test, unless stated otherwise. p-values of less than 0.05 were
583 considered to be statistically significant.
584
585 **Distribution of *briC* across streptococcal strains**
586 To identify *briC* homologs we used tblastn with default parameters on the RAST
587 database to search the genome sequences of all fifty-five strains. Predicted protein sequences
588 were downloaded as well as nucleotide sequences for the *briC* homolog and 1500-bp flanking
589 regions surrounding the *briC* homolog. Predicted protein sequences for BriC were aligned
590 using NCBI Cobalt (Papadopoulos and Agarwala, 2007) and visualized using Jalview
591 (Waterhouse et al., 2009). Jalview was then used to generate a neighbor-joining tree from the
592 protein multiple sequence alignment, and Principal Component Analysis (PCA) was used to
593 further analyze variation among the samples. From the PCA results and phylogenetic analysis,
594 it was determined that three major groups existed within our dataset, with the largest group
595 having two pherotypes within it. The *briC* alleles were then organized in the context of the
596 species tree. For this we used a published phylogenetic tree (Cuevas et al., 2017; Kadam et
597 al., 2017). As previously described, the whole genome sequence (WGS) for these strains were
598 aligned using MAUVE (Darling et al., 2004, 2010), the core region was extracted and aligned
599 using MAFFT (FFT-NS-2) (Katoh et al., 2002). Model selection was performed using
600 MODELTEST (Posada and Crandall, 1998), and the phylogenetic tree was built with PhyML
601 3.0 (Guindon et al., 2010), model GTR+I(0.63) using maximum likelihood and 100 bootstrap
602 replicates. On the visualization, each allelic type is shape-coded, and the visualization was
603 generated using the Interactive Tree of Life (iTOL) (Letunic and Bork, 2016).
604
605 **Analysis of *briC* promoter region**

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

606 In order to examine the structure of the promoter region upstream of the *briC* gene, a
607 1500-bp flanking region on both sides of the *briC* gene was pulled from the RAST database
608 (Overbeek et al., 2014). Sequences were aligned using Kalign (Lassmann and Sonnhammer,
609 2005) and then visualized with Jalview (Waterhouse et al., 2009). The alignment revealed two
610 clear groups within the dataset: those with the promoter insertion and those without. These
611 promoter insertions were then used to mark the species tree with allelic variants as described
612 above. We observed this insertion in the representative isolates from two clinically important
613 lineages PMEN1 and PMEN14. To check the distribution of the longer promoter in a larger set
614 strains, we used PubMLST (Jolley and Maiden, 2010) to inspect 3,529 sequences with complete
615 MLST profiles from two large pneumococcal datasets (Chewapreecha et al., 2014; Croucher et
616 al., 2013). These include thirty-two ST81 (PMEN1), as well as seventy-nine ST236 (PMEN14)
617 and ten ST320 (PMEN14) strains.
618 For analysis of the ComE-binding box, the ComE consensus sequence was visualized
619 and extracted from the promoter for the pneumococcal strains. The logo was generated using
620 WebLogo (Crooks et al., 2004).
621
622 **Ethics statement**
623 Mouse experiments were performed at the University of Leicester under appropriate
624 project (permit no. P7B01C07A) and personal licenses according to the United Kingdom Home
625 Office guidelines under the Animals Scientific Procedures Act 1986, and the University of
626 Leicester ethics committee approval. The protocol was agreed by both the U.K. Home Office
627 and the University of Leicester ethics committee. Where specified, the procedures were carried
628 out under anesthetic with isoflurane. Animals were housed in individually ventilated cages in a
629 controlled environment, and were frequently monitored after infection to minimize suffering.
630 Chinchilla experiments were performed at the Allegheny-Singer Research Institute (ASRI) under

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

631 the Institutional Animal Care and Use Committee (IACUC) permit A3693-01/1000. Chinchillas
632 were maintained in BSL2 facilities, and all experiments with chinchillas were done under
633 subcutaneously injected ketamine-xylazine anesthesia (1.7mg/kg animal weight for each). All
634 chinchillas were maintained in accordance with the applicable portions of the Animal Welfare
635 Act, and the guidelines published in the DHHS publication, Guide for the Care and Use of
636 Laboratory Animals.
637
638 **Funding**
639 This work was supported by NIH grant R00-DC-011322 to L.H., Stupakoff Scientific
640 Achievement Award to S.A., as well as support from the Department of Biological Sciences at
641 Carnegie Mellon University.
642
643 **Acknowledgements**
644 We thank Drs. Alexander Tomasz and Herminia deLancastre for the PMEN1 strain PN4595-T23
645 and the R6D strain used in this study. We would also like to thank Dr. Donald A. Morrison for
646 *kan-rpsL* Janus cassette and the plasmid pR412. We thank Anagha Kadam for help on data
647 analyses, Rolando A. Cuevas for help analyzing biofilm images, and Emilio I. Rodriguez for his
648 support with experiments.
649
650 **Supplementary Tables**
651 **Table S1: Strains used in genomic comparisons and phylogenetic tree.**
652 **Table S2: Strains used in this experimental work.**
653 **Table S3: Primers used in this study.**
654

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Al-Bayati, F. A. Y., Kahya, H. F. H., Damianou, A., Shafeeq, S., Kuipers, O. P., Andrew, P. W., et al. (2017). Pneumococcal galactose catabolism is controlled by multiple regulators acting on pyruvate formate lyase. *Sci. Rep.* 7, 43587. doi:10.1038/srep43587.

Alloing, G., Martin, B., Granadel, C., and Claverys, J. P. (1998). Development of competence in Streptococcus pneumonaie: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. *Mol. Microbiol.* 29, 75–83.

Aprianto, R., Slager, J., Holsappel, S., and Veening, J.-W. (2016). Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biol.* 17, 198. doi:10.1186/s13059-016-1054-5.

Armbruster, C. E., Hong, W., Pang, B., Weimer, K. E. D., Juneau, R. A., Turner, J., et al. (2010). Indirect Pathogenicity of Haemophilus influenzae and Moraxella catarrhalis in Polymicrobial Otitis Media Occurs via Interspecies Quorum Signaling. *mBio* 1, e00102-10. doi:10.1128/mBio.00102-10.

Blanchette-Cain, K., Hinojosa, C. A., Akula Suresh Babu, R., Lizcano, A., Gonzalez-Juarbe, N., Munoz-Almagro, C., et al. (2013). Streptococcus pneumoniae biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. *mBio* 4, e00745-713. doi:10.1128/mBio.00745-13.

Bogaert, D., de Groot, R., and Hermans, P. (2004). Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect. Dis.* 4, 144–154. doi:10.1016/S1473-3099(04)00938-7.

Carvalho, S. M., Kloosterman, T. G., Kuipers, O. P., and Neves, A. R. (2011). CcpA Ensures Optimal Metabolic Fitness of Streptococcus pneumoniae. *PLoS ONE* 6, e26707. doi:10.1371/journal.pone.0026707.

Chao, Y., Marks, L. R., Pettigrew, M. M., and Hakansson, A. P. (2015). Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. *Front. Cell. Infect. Microbiol.* 4. doi:10.3389/fcimb.2014.00194.

Chewapreecha, C., Harris, S. R., Croucher, N. J., Turner, C., Marttinen, P., Cheng, L., et al. (2014). Dense genomic sampling identifies highways of pneumococcal recombination. *Nat. Genet.* 46, 305–309. doi:10.1038/ng.2895.

Claverys, J.-P., Prudhomme, M., and Martin, B. (2006). Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. *Annu. Rev. Microbiol.* 60, 451–475. doi:10.1146/annurev.micro.60.080805.142139.

Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi:10.1101/gr.849004.

Croucher, N. J., Chewapreecha, C., Hanage, W. P., Harris, S. R., McGee, L., van der Linden, M., et al. (2014). Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. *Genome Biol. Evol.* 6, 1589–1602. doi:10.1093/gbe/evu120.

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

694 Croucher, N. J., Finkelstein, J. A., Pelton, S. I., Mitchell, P. K., Lee, G. M., Parkhill, J., et al.
695 (2013). Population genomics of post-vaccine changes in pneumococcal epidemiology.
696 *Nat. Genet.* 45, 656–663. doi:10.1038/ng.2625.

697 Croucher, N. J., Harris, S. R., Fraser, C., Quail, M. A., Burton, J., van der Linden, M., et al.
698 (2011). Rapid pneumococcal evolution in response to clinical interventions. *Science*
699 331, 430–434. doi:10.1126/science.1198545.

700 Cuevas, R. A., Eutsey, R., Kadam, A., West-Roberts, J. A., Woolford, C. A., Mitchell, A. P., et
701 al. (2017). A novel streptococcal cell-cell communication peptide promotes
702 pneumococcal virulence and biofilm formation. *Mol. Microbiol.* 105, 554–571.
703 doi:10.1111/mmi.13721.

704 Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: Multiple Alignment
705 of Conserved Genomic Sequence With Rearrangements. *Genome Res.* 14, 1394–1403.
706 doi:10.1101/gr.2289704.

707 Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: Multiple Genome
708 Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* 5, e11147.
709 doi:10.1371/journal.pone.0011147.

710 Dawid, S., Roche, A. M., and Weiser, J. N. (2007). The blp bacteriocins of Streptococcus
711 pneumoniae mediate intraspecies competition both in vitro and in vivo. *Infect. Immun.*
712 75, 443–451. doi:10.1128/IAI.01775-05.

713 Gagne, A. L., Stevens, K. E., Cassone, M., Pujari, A., Abiola, O. E., Chang, D. J., et al. (2013).
714 Competence in Streptococcus pneumoniae Is a Response to an Increasing Mutational
715 Burden. *PLOS ONE* 8, e72613. doi:10.1371/journal.pone.0072613.

716 Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., et al. (2008).
717 Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat.*
718 *Biotechnol.* 26, 317–325. doi:10.1038/nbt1385.

719 Guenzi, E., Gasc, A.-M., Sicard, M. A., and Hakenbeck, R. (1994). A two-component signal-
720 transducing system is involved in competence and penicillin susceptibility in laboratory
721 mutants of Streptococcus pneumoniae. *Mol. Microbiol.* 12, 505–515.
722 doi:10.1111/j.1365-2958.1994.tb01038.x.

723 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010).
724 New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing
725 the Performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi:10.1093/sysbio/syq010.

726 Hakenbeck, R., and Chhatwal, S. (2007). *Molecular Biology of Streptococci*. Horizon Scientific
727 Press.

728 Hall-Stoodley, L., Costerton, J. W., and Stoodley, P. (2004). Bacterial biofilms: from the Natural
729 environment to infectious diseases. *Nat. Rev. Microbiol.* 2, 95–108.
730 doi:10.1038/nrmicro821.

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

731 Hall-Stoodley, L., Hu, F. Z., Gieseke, A., Nistico, L., Nguyen, D., Hayes, J., et al. (2006). Direct
732 Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis
733 Media. *JAMA J. Am. Med. Assoc.* 296, 202–211. doi:10.1001/jama.296.2.202.

734 Havarstein, L. S., Coomaraswamy, G., and Morrison, D. A. (1995a). An unmodified
735 heptadecapeptide pheromone induces competence for genetic transformation in
736 Streptococcus pneumoniae. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11140–11144.

737 Havarstein, L. S., Coomaraswamy, G., and Morrison, D. A. (1995b). An unmodified
738 heptadecapeptide pheromone induces competence for genetic transformation in
739 Streptococcus pneumoniae. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11140–11144.

740 Håvarstein, L. S., Gaustad, P., Nes, I. F., and Morrison, D. A. (1996). Identification of the
741 streptococcal competence-pheromone receptor. *Mol. Microbiol.* 21, 863–869.
742 doi:10.1046/j.1365-2958.1996.521416.x.

743 Heydorn, A., Nielsen, A. T., Hentzer, M., Sternberg, C., Givskov, M., Ersbøll, B. K., et al.
744 (2000). Quantification of biofilm structures by the novel computer program COMSTAT.
745 *Microbiol. Read. Engl.* 146 (Pt 10), 2395–2407.

746 Hiller, N. L., Eutsey, R. A., Powell, E., Earl, J. P., Janto, B., Martin, D. P., et al. (2011).
747 Differences in Genotype and Virulence among Four Multidrug-Resistant Streptococcus
748 pneumoniae Isolates Belonging to the PMEN1 Clone. *PLoS ONE* 6, e28850.
749 doi:10.1371/journal.pone.0028850.

750 Hoa, M., Syamal, M., Sachdeva, L., Berk, R., and Coticchia, J. (2009). Demonstration of
751 nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis
752 media. *Ann. Otol. Rhinol. Laryngol.* 118, 292–298.

753 Jolley, K. A., and Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome
754 variation at the population level. *BMC Bioinformatics* 11, 595. doi:10.1186/1471-2105-
755 11-595.

756 Kadam, A., Eutsey, R. A., Rosch, J., Miao, X., Longwell, M., Xu, W., et al. (2017). Promiscuous
757 signaling by a regulatory system unique to the pandemic PMEN1 pneumococcal
758 lineage. *PLOS Pathog.* 13, e1006339. doi:10.1371/journal.ppat.1006339.

759 Kahya, H. F., Andrew, P. W., and Yesilkaya, H. (2017). Deacetylation of sialic acid by
760 esterases potentiates pneumococcal neuraminidase activity for mucin utilization,
761 colonization and virulence. *PLOS Pathog.* 13, e1006263.
762 doi:10.1371/journal.ppat.1006263.

763 Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid
764 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30,
765 3059–3066.

766 Kim, W., Park, H. K., Hwang, W.-J., and Shin, H.-S. (2013). Simultaneous Detection of
767 Streptococcus pneumoniae, S. mitis, and S. oralis by a Novel Multiplex PCR Assay
768 Targeting the gyrB Gene. *J. Clin. Microbiol.* 51, 835–840. doi:10.1128/JCM.02920-12.

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

769 Kjos, M., Miller, E., Slager, J., Lake, F. B., Gericke, O., Roberts, I. S., et al. (2016). Expression
770 of Streptococcus pneumoniae Bacteriocins Is Induced by Antibiotics via Regulatory
771 Interplay with the Competence System. *PLOS Pathog.* 12, e1005422.
772 doi:10.1371/journal.ppat.1005422.

773 Lanie, J. A., Ng, W.-L., Kazmierczak, K. M., Andrzejewski, T. M., Davidsen, T. M., Wayne, K.
774 J., et al. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of
775 Streptococcus pneumoniae and comparison with that of unencapsulated laboratory
776 strain R6. *J. Bacteriol.* 189, 38–51. doi:10.1128/JB.01148-06.

777 Lassmann, T., and Sonnhammer, E. L. (2005). Kalign – an accurate and fast multiple
778 sequence alignment algorithm. *BMC Bioinformatics* 6, 298. doi:10.1186/1471-2105-6-
779 298.

780 Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display
781 and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242-245.
782 doi:10.1093/nar/gkw290.

783 Marks, L. R., Davidson, B. A., Knight, P. R., and Hakansson, A. P. (2013). Interkingdom
784 signaling induces Streptococcus pneumoniae biofilm dispersion and transition from
785 asymptomatic colonization to disease. *mBio* 4. doi:10.1128/mBio.00438-13.

786 Marks, L. R., Parameswaran, G. I., and Hakansson, A. P. (2012a). Pneumococcal interactions
787 with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and
788 in vivo. *Infect. Immun.* 80, 2744–2760. doi:10.1128/IAI.00488-12.

789 Marks, L. R., Reddinger, R. M., and Hakansson, A. P. (2012b). High Levels of Genetic
790 Recombination during Nasopharyngeal Carriage and Biofilm Formation in
791 Streptococcus pneumoniae. *mBio* 3, e00200-12. doi:10.1128/mBio.00200-12.

792 Martin, B., Granadel, C., Campo, N., Hénard, V., Prudhomme, M., and Claverys, J.-P. (2010).
793 Expression and maintenance of ComD-ComE, the two-component signal-transduction
794 system that controls competence of Streptococcus pneumoniae. *Mol. Microbiol.* 75,
795 1513–1528. doi:10.1111/j.1365-2958.2010.07071.x.

796 Mascher, T., Zähner, D., Merai, M., Balmelle, N., de Saizieu, A. B., and Hakenbeck, R. (2003).
797 The Streptococcus pneumoniae cia regulon: CiaR target sites and transcription profile
798 analysis. *J. Bacteriol.* 185, 60–70.

799 Miller, JH (1972). "Assay of b-galactosidase.," in *Experiments in Molecular Genetics*. Cold
800 Spring Harbor, New York: Laboratory Press, 352–355.

801 Moreno-Gámez, S., Sorg, R. A., Domenech, A., Kjos, M., Weissing, F. J., Doorn, G. S., et al.
802 (2017). Quorum sensing integrates environmental cues, cell density and cell history to
803 control bacterial competence. *Nat. Commun.* 8, 854. doi:10.1038/s41467-017-00903-y.

804 Oggioni, M. R., Trappetti, C., Kadioglu, A., Cassone, M., Iannelli, F., Ricci, S., et al. (2006).
805 Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis.
806 *Mol. Microbiol.* 61, 1196–1210. doi:10.1111/j.1365-2958.2006.05310.x.

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi:10.1093/nar/gkt1226.

Paixão, L., Oliveira, J., Veríssimo, A., Vinga, S., Lourenço, E. C., Ventura, M. R., et al. (2015). Host glycan sugar-specific pathways in Streptococcus pneumonia: galactose as a key sugar in colonisation and infection. *PloS One* 10, e0121042. doi:10.1371/journal.pone.0121042.

Papadopoulos, J. S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23, 1073–1079. doi:10.1093/bioinformatics/btm076.

Pestova, E. V., Håvarstein, L. S., and Morrison, D. A. (1996). Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. *Mol. Microbiol.* 21, 853–862.

Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V., Snesrud, E. C., Luo, P., et al. (2004). Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Mol. Microbiol.* 51, 1051–1070.

Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinforma. Oxf. Engl.* 14, 817–818.

Ramakers, C., Ruijter, J. M., Deprez, R. H. L., and Moorman, A. F. M. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* 339, 62–66.

Ruijter, J. M., Lorenz, P., Tuomi, J. M., Hecker, M., and van den Hoff, M. J. B. (2014). Fluorescent-increase kinetics of different fluorescent reporters used for qPCR depend on monitoring chemistry, targeted sequence, type of DNA input and PCR efficiency. *Mikrochim. Acta* 181, 1689–1696. doi:10.1007/s00604-013-1155-8.

Sanderson, A. R., Leid, J. G., and Hunsaker, D. (2006). Bacterial biofilms on the sinus mucosa of human subjects with chronic rhinosinusitis. *The Laryngoscope* 116, 1121–1126. doi:10.1097/01.mlg.0000221954.05467.54.

Sebert, M. E., Patel, K. P., Plotnick, M., and Weiser, J. N. (2005). Pneumococcal HtrA protease mediates inhibition of competence by the CiaRH two-component signaling system. *J. Bacteriol.* 187, 3969–3979. doi:10.1128/JB.187.12.3969-3979.2005.

Severin, A., Figueiredo, A. M., and Tomasz, A. (1996). Separation of abnormal cell wall composition from penicillin resistance through genetic transformation of Streptococcus pneumoniae. *J. Bacteriol.* 178, 1788–1792.

Shanker, E., and Federle, M. J. (2017). Quorum Sensing Regulation of Competence and Bacteriocins in Streptococcus pneumoniae and mutans. *Genes* 8. doi:10.3390/genes8010015.

Son, M. R., Shchepetov, M., Adrian, P. V., Madhi, S. A., de Gouveia, L., von Gottberg, A., et al. (2011). Conserved mutations in the pneumococcal bacteriocin transporter gene,

bioRxiv preprint doi: https://doi.org/10.1101/245902; this version posted January 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

blpA, result in a complex population consisting of producers and cheaters. *mBio* 2. doi:10.1128/mBio.00179-11.

Tomasz, A. (1965). Control of the Competent State in Pneumococcus by a Hormone-Like Cell Product: An Example for a New Type of Regulatory Mechanism in Bacteria. *Nature* 208, 155–159. doi:10.1038/208155a0.

Trappetti, C., Gualdi, L., Meola, L. D., Jain, P., Korir, C. C., Edmonds, P., et al. (2011a). The impact of the competence quorum sensing system on Streptococcus pneumoniae biofilms varies depending on the experimental model. *BMC Microbiol.* 11, 75. doi:10.1186/1471-2180-11-75.

Trappetti, C., Potter, A. J., Paton, A. W., Oggioni, M. R., and Paton, J. C. (2011b). LuxS Mediates Iron-Dependent Biofilm Formation, Competence, and Fratricide in Streptococcus pneumoniae ▿. *Infect. Immun.* 79, 4550–4558. doi:10.1128/IAI.05644-11.

Vidal, J. E., Howery, K. E., Ludewick, H. P., Nava, P., and Klugman, K. P. (2013). Quorum-sensing systems LuxS/autoinducer 2 and Com regulate Streptococcus pneumoniae biofilms in a bioreactor with living cultures of human respiratory cells. *Infect. Immun.* 81, 1341–1353. doi:10.1128/IAI.01096-12.

Vidal, J. E., Ludewick, H. P., Kunkel, R. M., Zähner, D., and Klugman, K. P. (2011). The LuxS-Dependent Quorum-Sensing System Regulates Early Biofilm Formation by Streptococcus pneumoniae Strain D39 ▿. *Infect. Immun.* 79, 4050–4060. doi:10.1128/IAI.05186-11.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi:10.1093/bioinformatics/btp033.

Ween, O., Gaustad, P., and Håvarstein, L. S. (1999). Identification of DNA binding sites for ComE, a key regulator of natural competence in Streptococcus pneumoniae. *Mol. Microbiol.* 33, 817–827.

Wholey, W.-Y., Kochan, T. J., Storck, D. N., and Dawid, S. (2016). Coordinated Bacteriocin Expression and Competence in Streptococcus pneumoniae Contributes to Genetic Adaptation through Neighbor Predation. *PLOS Pathog.* 12, e1005413. doi:10.1371/journal.ppat.1005413.

Wyres, K. L., Lambertsen, L. M., Croucher, N. J., McGee, L., von Gottberg, A., Liñares, J., et al. (2012). The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success. *Genome Biol.* 13, R103. doi:10.1186/gb-2012-13-11-r103.

Zähner, D., and Hakenbeck, R. (2000). The Streptococcus pneumoniae Beta-Galactosidase Is a Surface Protein. *J. Bacteriol.* 182, 5919–5921.

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 1

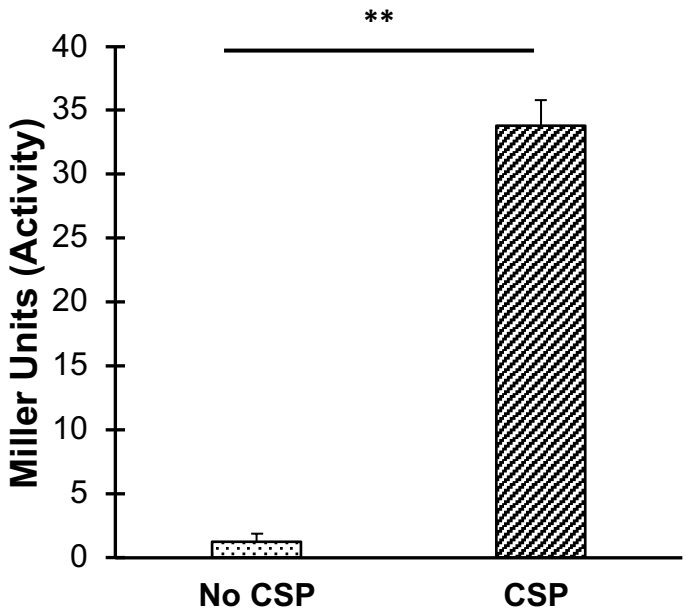
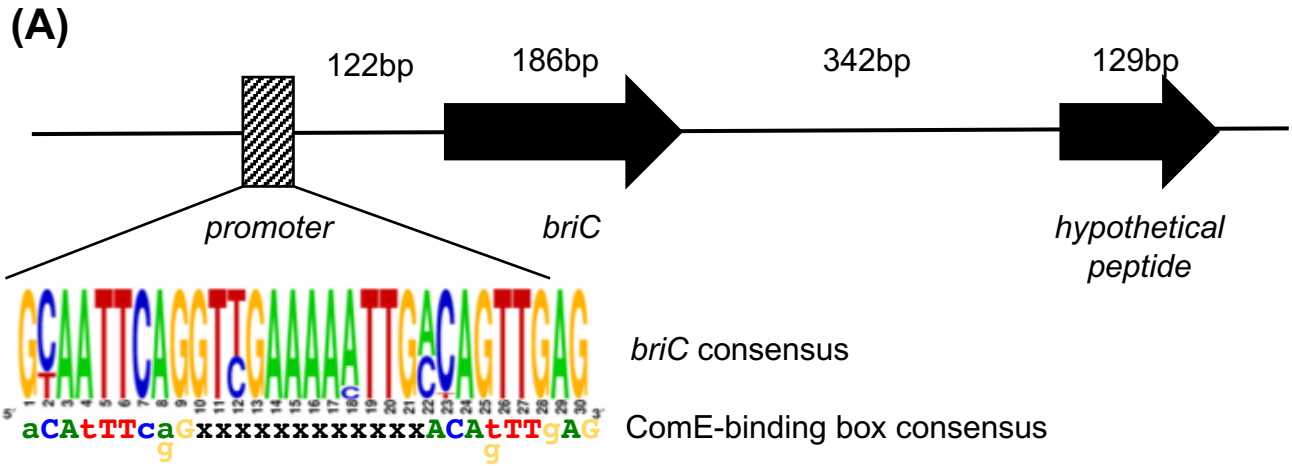


Fig. 1. Expression of *briC* is induced by CSP. β-galactosidase assay measuring P*briC-lacZ* activity in pneumococcal R6 cells grown to exponential phase in Columbia Broth at pH 6.6 followed by treatment with CSP1 for 30 minutes or untreated. Y-axis denotes P*briC-lacZ* expression levels in Miller Units. Activity is expressed in nmol p-nitrophenol/min/ml. Error bars represent standard error of the mean for biological replicates *(n=3)*; ** $p<0.01$ using Student's two-tailed paired *t*-test.

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 2



certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 2

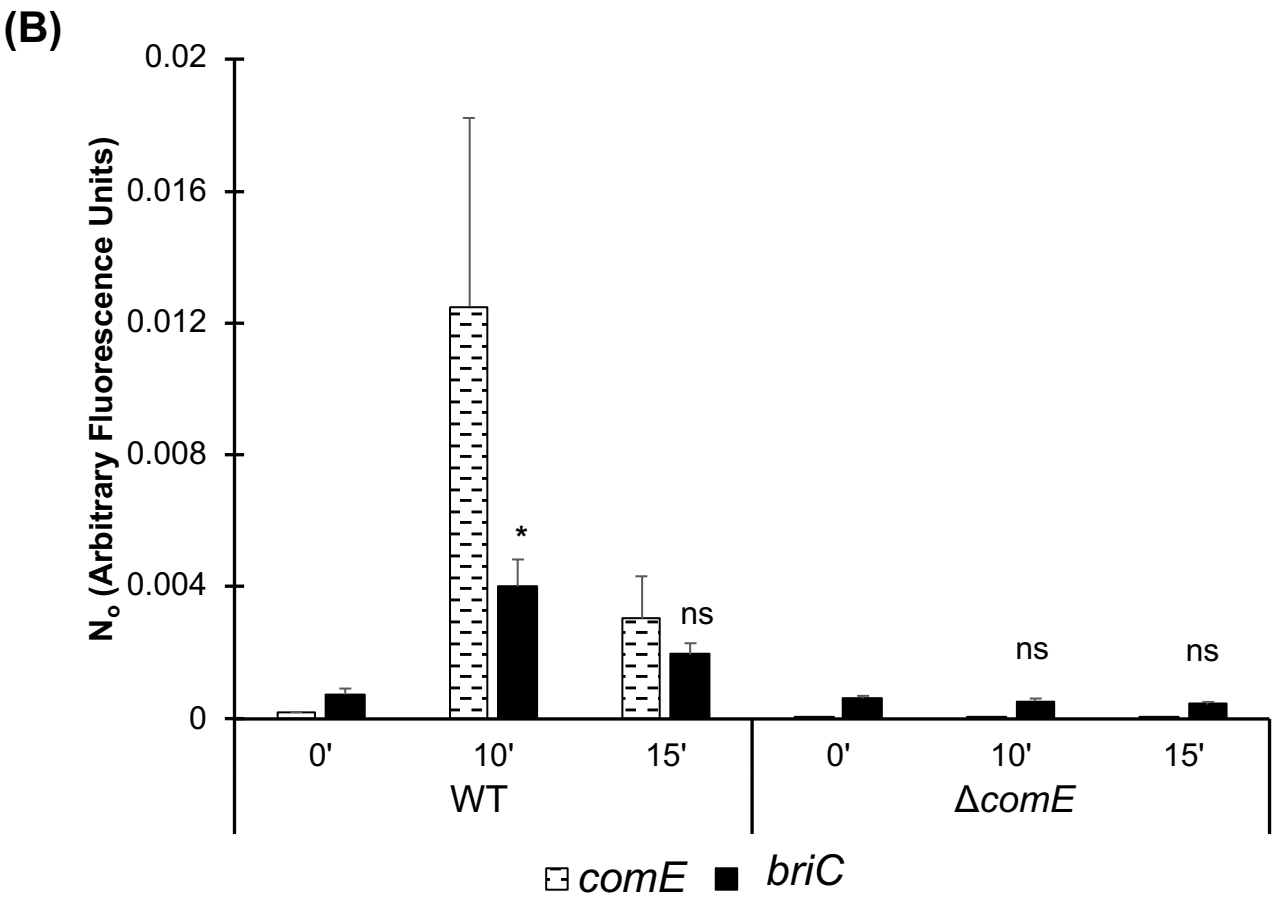
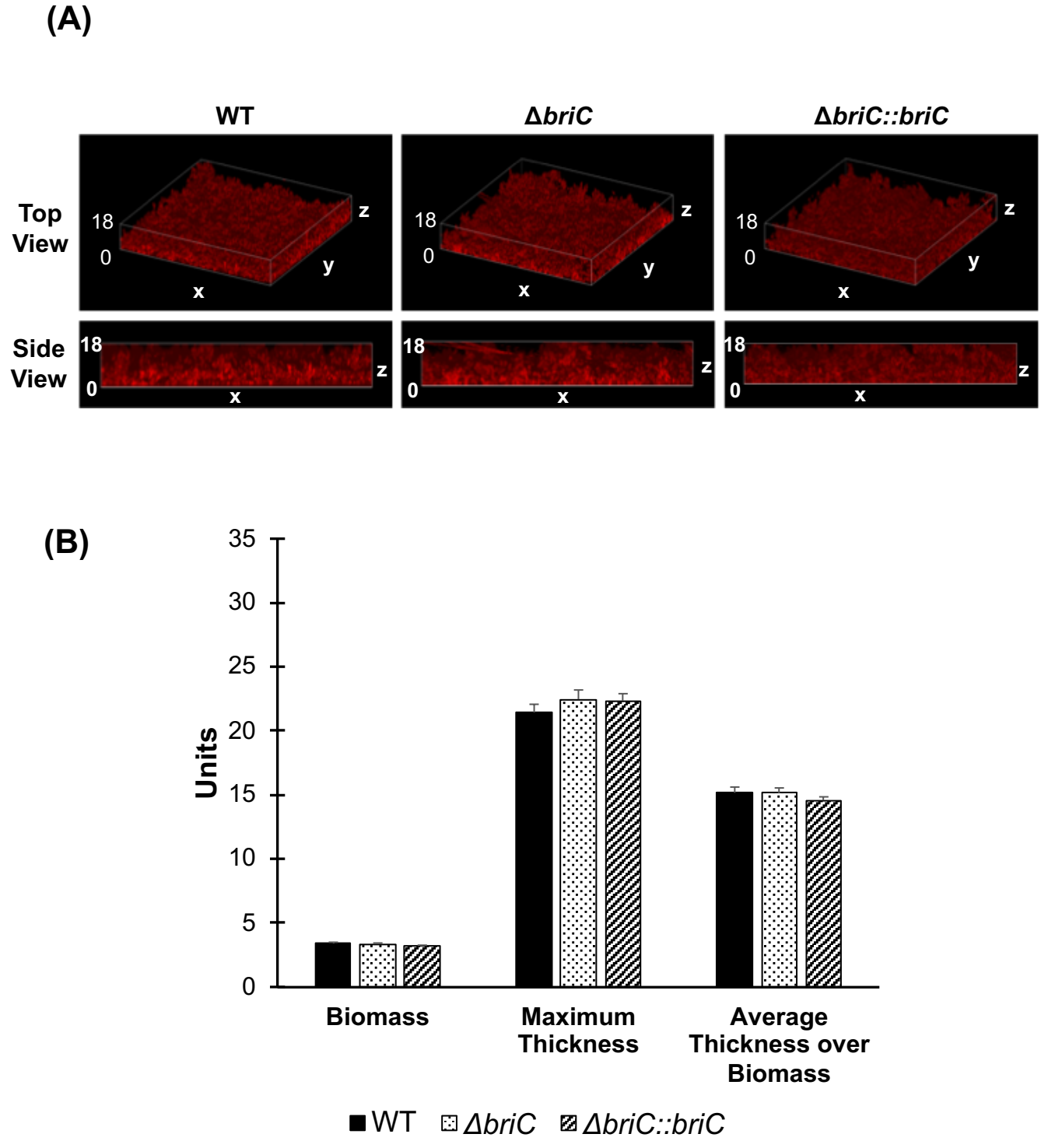


Fig. 2. CSP-induction of *briC* is ComE-dependent. (A) Genomic organization of the *briC* locus in strain R6D. Black arrows: coding sequences; box with diagonal lines: ComE binding-box within the putative *briC* promoter region; expanded region: logo generated from the predicted *briC* ComE-binding box in thirty-four pneumococcal genomes aligned with the published ComE-binding box consensus sequence. **(B)** mRNA transcript levels of *briC* (solid black) and *comE* (dashed black lines) as measured by qRT-PCR in R6D WT & R6DΔ*comE* cells. Cells were grown in Columbia broth at pH 6.6 to an OD_{600} of 0.3, and then treated with CSP1 for either 0', 10' or 15'. Data was normalized to 16S rRNA levels. Y-axis denotes normalized concentrations of mRNA levels in arbitrary fluorescence units as calculated from LinRegPCR. Error bars represent standard error of the mean calculated for biological replicates *(n=3)*; 'ns' denotes non-significant, * $p<0.05$ using ANOVA followed by Tukey's post-test relative to the respective 0' CSP treatment. Further, *briC* levels are also significantly higher in WT relative to Δ*comE* cells for the same time points post-CSP treatment ($p<0.05$).

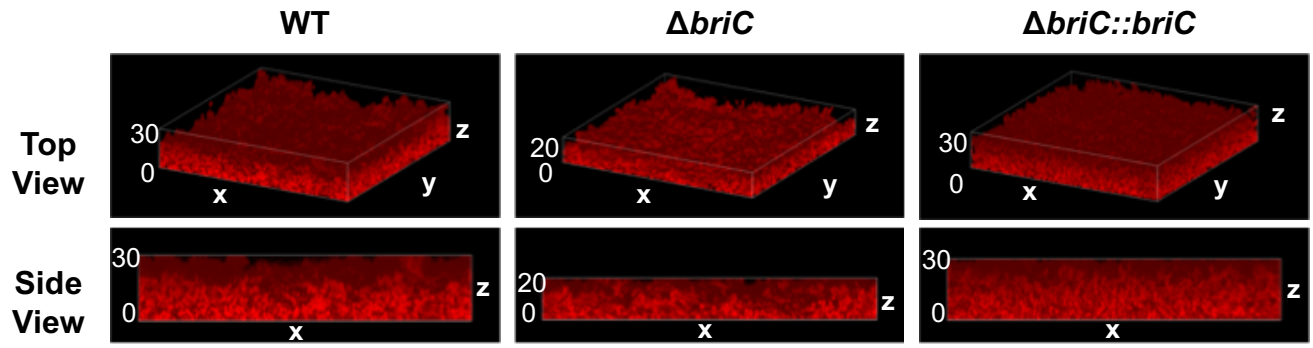
certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

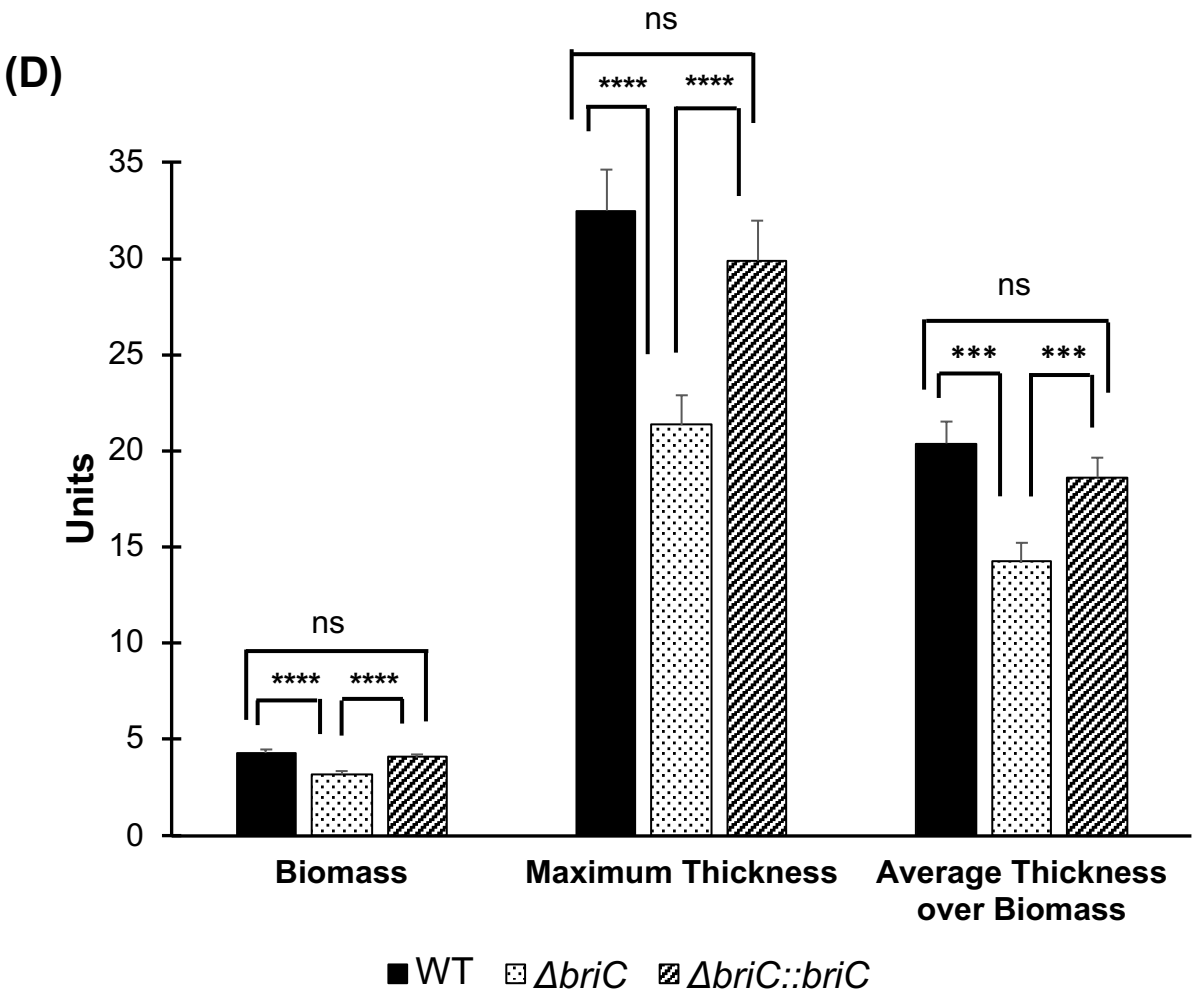
Figure 3



certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 3



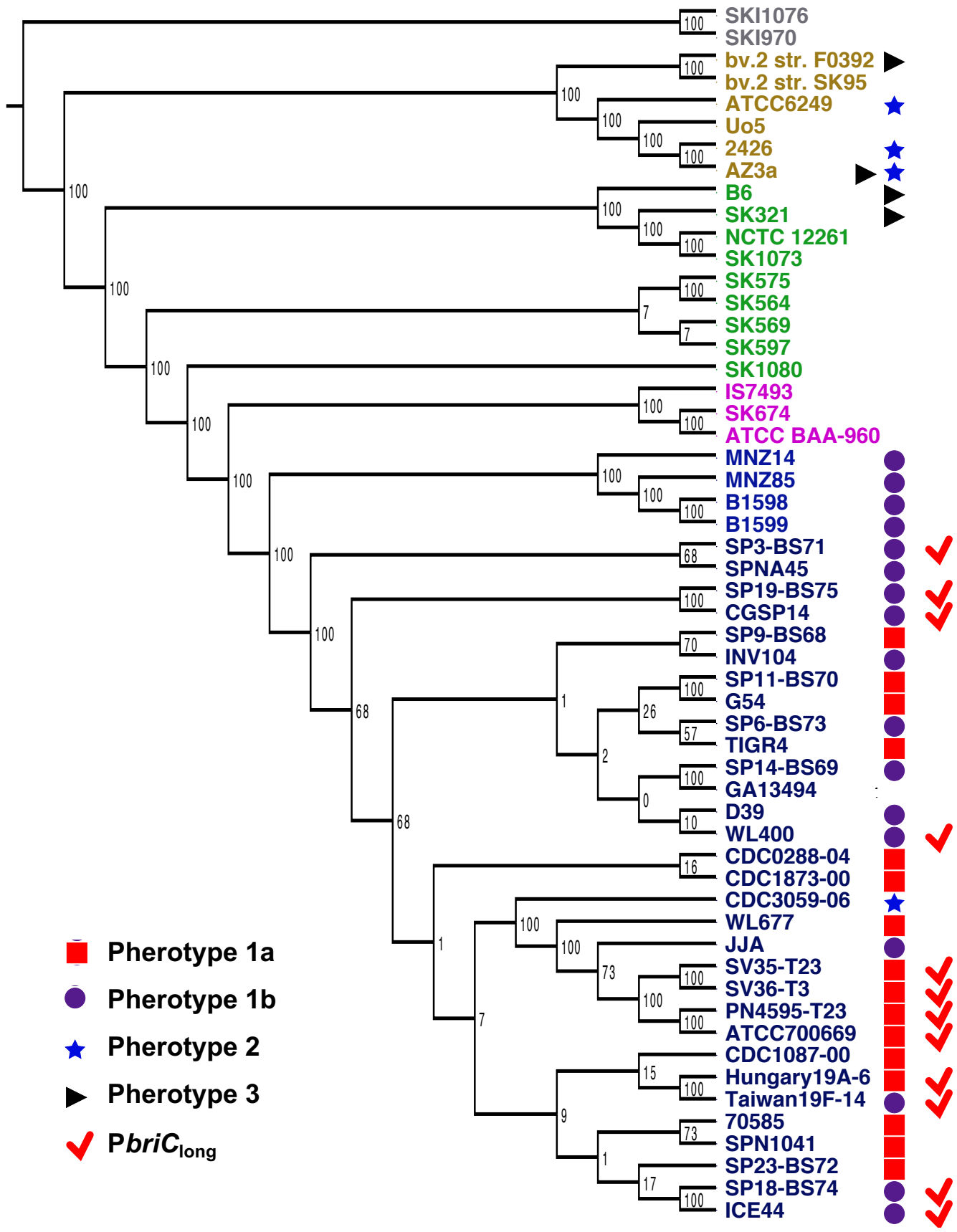


certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Fig. 3. BriC stimulates late biofilm development. Representative confocal microscopy images showing top & side views of the reconstructed biofilm stacks of WT, Δ*briC* and Δ*briC::briC* cells of strain R6D stained with SYTO59 dye at **(A)** 24-hr, and **(C**) 72-hr. 'x', 'y' and 'z' represent different axes of the reconstructed Z-stack with the numbers representing thickness in μm. COMSTAT2 quantification of **(B)** 24-hr, and **(D)** 72-hr biofilm images. Y-axis denotes units of measurement: $\mu m^3/\mu m^2$ for biomass, and μm for maximum thickness and average thickness over biomass. Error bars represent standard error of the mean calculated for biological replicates *(n=3)*; "ns" denotes non-significant comparisons, *** $p<0.001$, and **** $p<0.0001$ using ANOVA followed by Tukey's post-test.

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 4



certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Fig. 4. Distribution of the genomic region encoding BriC across streptococcal strains. Distribution of *briC* alleles in fifty-five streptococcal genomes. The *briC* alleles are visualized against a maximum likelihood tree of streptococcal genomes generated from the core genome, where the numbers on the branches represent bootstrap values. Species are color-coded as follows: *S. pneumoniae* (blue), *S. pseudopneumoniae* (pink), *S. mitis* (green), *S. oralis* (beige), and *S. infantis* (grey). The shapes at the tip of the branches illustrate *briC* alleles: 1a (red square), 1b (purple circle), 2 (blue star), and 3 (black triangle). Types “1a” and “1b” represent variants of the alleles widespread across pneumococcal strains; type “3” denotes a group with high variability. The BriC coding sequences are aligned in Figure S2. The red tick marks genomes which have a longer promoter, which in PMEN1 strains leads to increase in basal levels of *briC* in a CSP-independent manner.

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 5

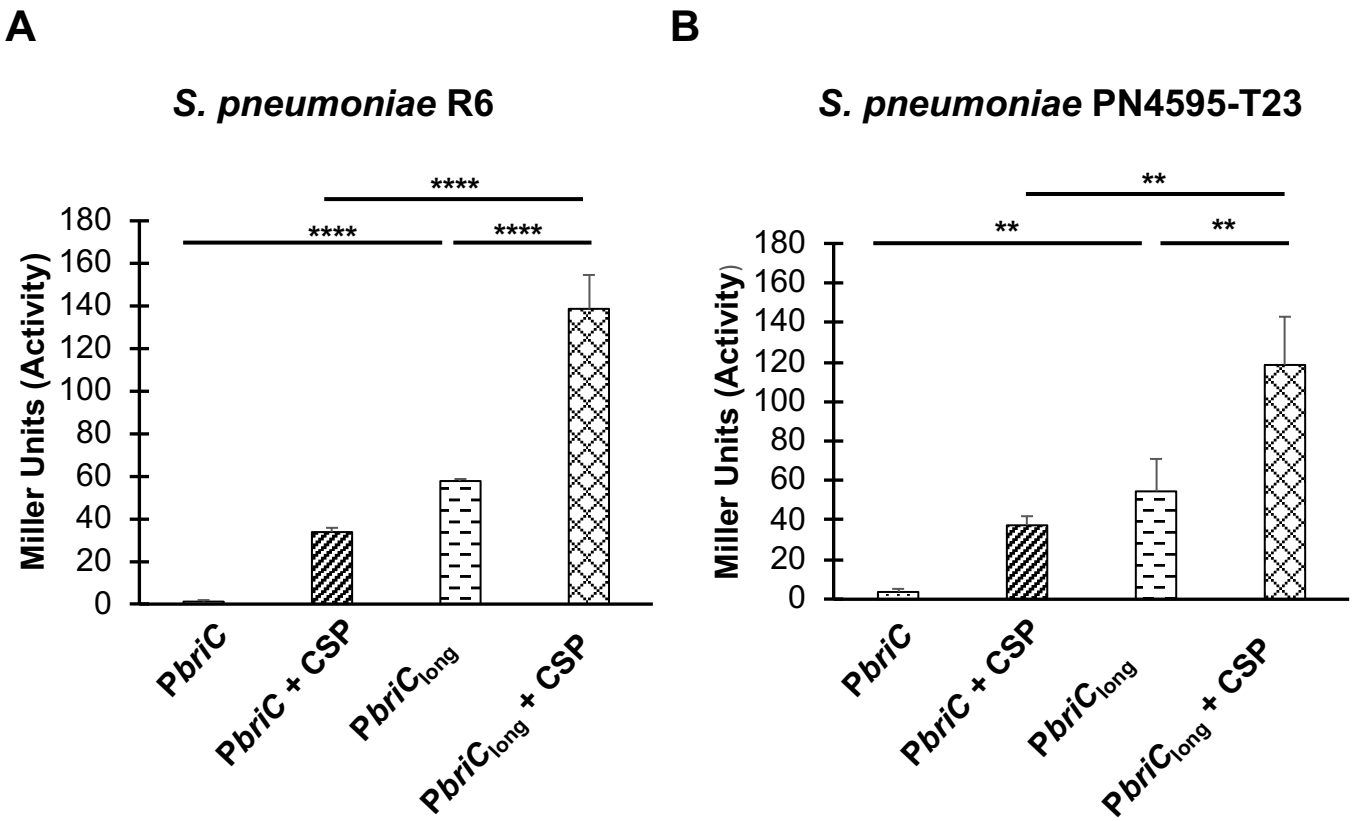
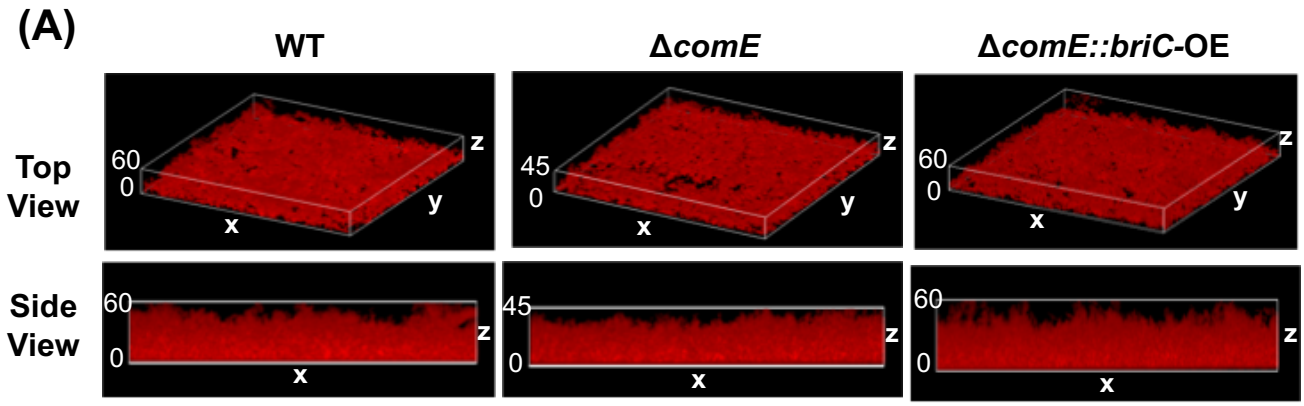
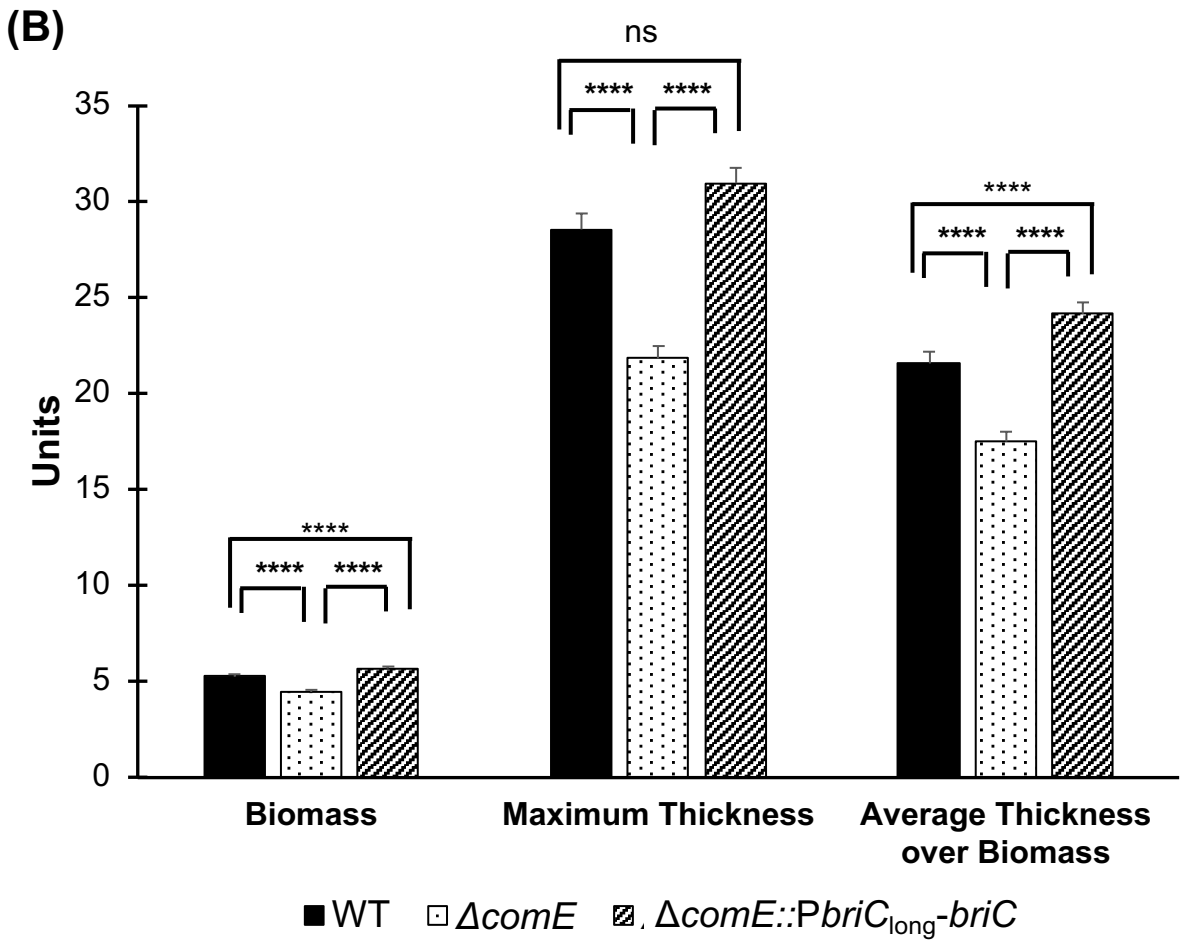


Fig. 5. Longer *briC* promoter is associated with an increase in the basal levels of *briC*. β-galactosidase assay comparing the LacZ activity of the R6 (short promoter, P*briC-lacZ)* and PN4595-T23 (longer promoter, P$briC_{long}$*-lacZ)* promoters. Both promoter activities were tested in **(A)** strain R6 and **(B)** strain PN4595-T23. Cells were grown in Columbia broth at pH 6.6 until mid-log phase, followed by either no treatment or treatment with CSP for 30 minutes. Y-axis denotes promoter activity in Miller Units expressed in nmol p-nitrophenol/min/ml. Error bars represent standard error of the mean for biological replicates *(n=3)*; ** $p<0.01$, & **** $p<0.0001$ using ANOVA followed by Tukey's post-test.

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 6





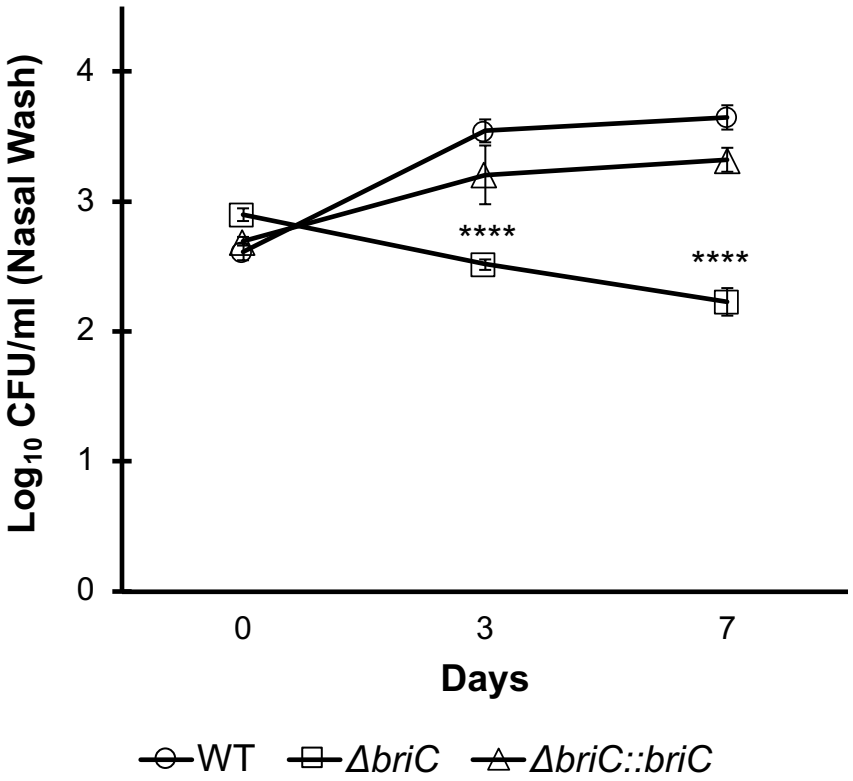
certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Fig. 6. BriC plays a pivotal role in regulating biofilm development. (A) Representative confocal microscopy images showing top & side views of the reconstructed biofilm stacks of WT, Δ*comE* and Δ*comE::*P*briC*$_{long}$-*briC* cells of strain R6D stained with SYTO59 dye at 72-hr. 'x', 'y' and 'z' represent different axes of the reconstructed Z-stack with the numbers representing thickness in μm. **(B)** COMSTAT2 quantification of 72-hr biofilm images. Y-axis denotes units of measurement: $\mu m^3/\mu m^2$ for biomass, and μm for maximum thickness and average thickness over biomass. Error bars represent standard error of the mean calculated for biological replicates *(n=6)*; "ns" denotes non-significant comparisons, and **** $p<0.0001$ using ANOVA followed by Tukey's post-test.

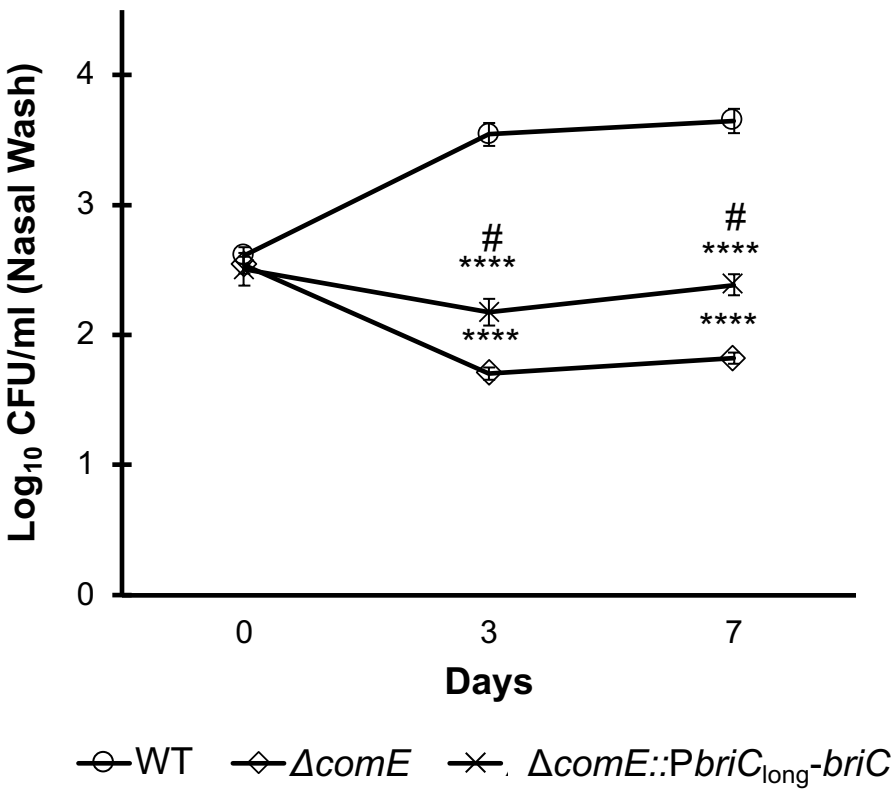
certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Figure 7

(A)



(B)



certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC-ND 4.0 International license.

Fig. 7. BriC contributes to pneumococcal colonization of the mouse nasopharynx. CD1 mice were infected intranasally with 20µl PBS containing approximately 1 X 10^5 CFU of (A) WT, Δ*briC*, and Δ*briC::bri*C (B) WT, Δ*comE*, and Δ*comE::*P*briC*$_{long}$-*briC* cells of the pneumococcal strain D39. At predetermined time points (0, 3 & 7 days post-infection), at least five mice were culled, and the pneumococcal counts in the nasopharyngeal washes were enumerated by plating on blood agar. Y-axis represents Log_{10} counts of CFU recovered from nasal washes. X-axis represents days post-inoculation. Each data point represents the mean of data from at least five mice. Error bars show the standard error of the mean. **** p<0.0001 relative to the WT strain, and # p<0.0001 relative to the Δ*comE* strain, calculated using ANOVA and Tukey post-test.