Submitted as a Research Article for *PLoS Pathogens*

# Crosstalk and eavesdropping among quorum sensing peptide signals that regulate bacteriocin production in *Streptococcus pneumoniae*

Eric L. Miller[a,b,§], Morten Kjos[c,d,§], Monica I. Abrudan[e,g], Ian S. Roberts[a,1], Jan-Willem Veening[c,f,1] and Daniel E. Rozen[b,1]

[a] Faculty of Biology, Medicine, and Health, University of Manchester, Manchester, M13 9PL, UK.

[b] Institute of Biology Leiden, Leiden University, Leiden, 2333 BE, The Netherlands.

[c] Molecular Genetics Group, Groningen Biomolecular Sciences and Biotechnology Institute, Centre for Synthetic Biology, University of Groningen, Groningen, 9700 AE, The Netherlands.

[d] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, N-1432 Ås, Norway

[e] Wellcome Trust Sanger Institute, Genome Campus, Cambridge, CB10 1SA, UK

[f] Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland

[g] Faculty of Medicine, School of Public Health, Imperial College, London W2 1PG, United Kingdom

[§] Both authors contributed equally to this manuscript.

[1] Correspondence to: d.e.rozen@biology.leidenuniv.nl

Jan-Willem.Veening@unil.ch

i.s.roberts@manchester.ac.uk

Classification: Biological Sciences; Genetics / Microbiology

Short title: Intraspecific crosstalk in *blp* quorum sensing

Key words: Streptococcus, quorum sensing, intra-species competition, signaling, eavesdropping, crosstalk

**Abstract**

During colonization of the human nasopharynx, multiple strains of the Gram-positive pathogen *Streptococcus pneumoniae* coexist and compete with each other using secreted antimicrobial peptides called bacteriocins. The major class of pneumococcal bacteriocins is encoded by the *blp* operon, whose transcription is controlled by the secretion and detection of a polymorphic family of quorum sensing (QS) peptides. We examined the genomic association between *blp* QS signals (BlpC) and receptors (BlpH) across 4,096 pneumococcal genomes. Imperfect concordance between nine QS signal peptide types and five phylogenetically-related QS receptor groups suggested extensive crosstalk between signals (where cells produce signals that non-clonal cells can detect) and eavesdropping (where cells respond to signals that they do not produce). To test these possibilities, we quantified the response of reporter strains containing each of six different *blp* QS receptor variants to cognate and non-cognate synthetic peptide signals. As predicted, we found evidence for crosstalk in five of six tested signals and for eavesdropping in four of these receptors. These *in vitro* results were confirmed during interactions between pneumococcal colonies grown adjacent to one another, providing direct evidence that crosstalk and eavesdropping occur at endogenous, ecologically-relevant, levels of signal secretion. Finally, using a spatially explicit stochastic model, we show that eavesdropping genotypes gain evolutionary advantages during inter-strain competition, even when their affinity to non-cognate signals is as low as 10% of the affinity to their cognate signal. Our results highlight the importance of social interactions in mediating intraspecific competition among bacteria and clarify that diverse competitive interactions can be mediated by polymorphism in QS systems.

**Author Summary**

Quorum sensing (QS), where bacteria secrete and respond to chemical signals to coordinate population-wide behaviors, has revealed that bacteria are highly social. Here, we use a multifaceted approach to investigate how diversity in QS signals and receptors can modify social interactions controlled by the QS system that regulates antimicrobial peptide secretion in *Streptococcus pneumoniae*. We experimentally confirmed that single QS receptors can respond to multiple signals (eavesdropping) and that single QS signals activate multiple receptors (crosstalk).We also show that QS eavesdropping can differentially affect gene expression in neighboring colonies. Eavesdropping and crosstalk can potentially explain our finding from surveys of 4,096 pneumococcal genomes that 16.7% of strains express QS receptors that may be unable to detect the QS signal that they produce. Simulations of QS strains producing antimicrobial peptides revealed that eavesdropping can be evolutionarily beneficial even when their affinity for non-cognate signals is very weak. Our results demonstrate the importance of eavesdropping and crosstalk as drivers of the outcome of competitive interactions mediated by bacterial quorum sensing.

## Introduction

Quorum sensing (QS) is a mechanism of intercellular communication that allows bacterial populations to coordinately regulate gene expression in response to changes in population density. QS is controlled by the secretion and detection of diffusible signaling molecules that, at threshold concentrations, lead to increased signal secretion and the induction of coupled downstream pathways (1,2). By this process, QS ensures that metabolically costly products are only produced when this would benefit the bacterial population, i.e. when they are at high concentrations (2,3). QS systems are coordinated by the fact that cells are simultaneously capable of sending and detecting a specific signal (2–4), a characteristic that increases the likelihood that QS functions as a private message between clonemates that share evolutionary interests (5,6). Evolutionary biology research on QS has focused on the dynamics of QS signal-blind and/or QS detection-blind cells ('cheaters') that do not pay the cost of QS but that take advantage of any induced public goods created by QS-faithful cells ('cooperators') (7–12). However, outside the laboratory, bacteria often reside in multispecies and multi-strain communities, where secreted QS signals can be detected by any cell, not just by clonemates (3). Thus, although QS works as an effective means of gene regulation in the laboratory in single strain cultures, QS in nature may be less reliable because it is susceptible to signal eavesdropping (i.e. where a promiscuous QS receptor can detect a QS signal not produced by that genotype) and signal crosstalk (i.e. where a non-specific QS signal can activate QS receptors in genotypes that produce other QS signals) (Fig. 1A; 3, 7). This variation in QS signals and QS signal detection is both widespread in nature (14–17) and distinct from well-studied cheater/cooperator dynamics. For example, signal-blind bacteria that produce, but are incapable of responding to, QS signals can engage in signal crosstalk in order to manipulate the behavior of other cells, e.g. by inducing them to produce expensive public goods (8). Crosstalk and eavesdropping can occur even if all cells within a population are otherwise phenotypically wild-type if (i) QS signals and receptors are polymorphic and (ii) signals can bind and activate more than one receptor variant. Here we examine these issues using the polymorphic QS system regulating bacteriocin production in the Gram-positive opportunistic pathogen *Streptococcus pneumoniae*, where QS is integral for mediating intraspecific competition.

To initiate infection, *S. pneumoniae* must successfully colonize the nasopharynx and then persist during subsequent colonization attempts from other strains. Commensal carriage of *S. pneumoniae* is ubiquitous, affecting up to 88% of children worldwide (18,19), while between 5-52% of individuals are co-colonized with multiple strains (19–22). The interactions between different strains during colonization are widespread and dynamic, and the rate of clonal turnover — where one strain displaces another — occurs on a timescale of days to months (23,24). Among the key factors thought to mediate intraspecific competition among pneumococcal strains are small anti-microbial peptides with narrow target ranges called bacteriocins (25), many of which are regulated by QS. The genome of *S. pneumoniae*

encodes several bacteriocin families, the most diverse of which are the bacteriocins encoded by the *blp* (*b*acteriocin-*l*ike *p*eptides) operon (26,25). Our recent work revealed that the number of distinct combinations of bacteriocins and immunity genes at this operon can extend into the trillions, although phylogenetic and functional constraints reduce this number to several hundred realized combinations (27). As with other Gram-positive peptide signals, the QS signal peptide (BlpC) regulating the *blp* operon is constitutively produced at low levels, but is auto-induced at high levels once a threshold concentration has been reached (26). Secreted BlpC binds to the extracellular domain of the membrane-bound histidine kinase BlpH, and upon binding the kinase phosphorylates the response regulator BlpR (Fig. 1B; 14, 15), which initiates production of the *blp* bacteriocin and immunity genes and increases production of the BlpC signal (30). Additionally, *blpC* expression is enhanced by the induction of pneumococcal competence, which is regulated by the paralogous *com* QS signaling system (31). Both ABC transporters BlpAB (32) and ComAB (31,33) cleave the N-terminal, double-glycine leader sequence of BlpC before export of the mature peptide signal by the same transporters (Fig. 1B). Another QS signal, CSP (competence-stimulating-peptide), can induce the transcription of *blp* genes at a low rate through the QS receptor ComD and its associated response regulator, ComE (31,33). Using QS to regulate secretion presumably ensures that Blp bacteriocins are only produced when there is a sufficiently high cell number to allow these anticompetitor toxins to reach effective concentrations.

Importantly, both the BlpC signal and its receptor, BlpH, are highly polymorphic. Our survey of 4,096 genomes identified 29 amino acid variants of the BlpC gene and 156 amino acid variants of BlpH (27). What are the effects of this variation, and how does this diversity influence the competitive interactions between strains that are mediated by *blp* bacteriocins? One possibility is that each unique BlpC signal corresponds to a distinct set of BlpH receptors to which it specifically and exclusively binds. By this explanation, strains detect and respond only to their own signal to determine the threshold at which they induce the *blp* operon. Such exclusivity is found in the competence signaling system where the two dominant peptide signals, CSP1 and CSP2, only induce cells expressing the cognate receptor (34). Similarly, there is near perfect concordance between the signal and receptor carried by any single genome, suggesting that tight coupling of these loci is crucial for the activation of competence (35). An alternative possibility, considered in a recent experimental study (36), is that BlpC peptides cross-react via crosstalk or eavesdropping with different BlpH receptors, thereby leading to a scenario where competing strains interact socially to induce the production of either immunity or bacteriocins at densities that would be insufficient for activation by auto-induction. Bacterial strains may benefit from this cross-reactivity if they are forewarned of the threats from others, allowing them to induce their own bacteriocins or immunity. Alternatively, eavesdropping may be costly if strains with promiscuous receptors are induced to secrete bacteriocins at densities that are too low to provide sufficient benefits to offset the costs of their production. *S. pneumoniae* presents an ideal opportunity to study the evolution of QS systems beyond cheater/cooperator dynamics (10–12) in an easily manipulated, highly relevant study system in which much is already known about signal/receptor

dynamics (36).

To understand the incidence and consequences of crosstalk and eavesdropping in a QS system that contains enormous diversity, we first investigated the fidelity between BlpC signals and BlpH histidine kinases across thousands of *S. pneumoniae* genomes using a bioinformatics approach. These results then informed experiments that quantified the response of bacterial strains and interacting colonies to cognate and non-cognate peptide signals across the major signaling classes. Finally, these results were examined in light of a stochastic model that investigated the consequences of QS eavesdropping for bacteriocin regulation. Our results reveal the importance of QS signaling polymorphism on *blp* operon regulation and clarify its ecological effects on *S. pneumoniae* intraspecific interactions.

## Results

### Molecular diversity of *blpH* and *blpC*

We examined 4,418 *S. pneumoniae* genomes taken from six data sets of randomly collected strains (Maela, Massachusetts Asymptomatic, GenBank, Hermans, Georgia GenBank, and PMEN: 4,096 genomes in total) and two additional data sets that are intentionally biased to specific clonal sub-groups (Complex 3 and PMEN-1: 322 genomes in total). We identified *blpC* in 99.0%, *blpH* in 99.0%, and both *blpC* and *blpH* in 98.2% of the 4,418 genomes using a DNA reciprocal BLAST algorithm (27). We note that the few genomes apparently lacking a *blp* operon gene may still contain these genes, as the data sets contain incomplete draft genomes.

We found extensive allelic variation within *blpC*, which contains 37 alleles at the nucleotide level, 29 protein variants, and 20 different BlpC signal peptides, including signal peptides lacking a canonical double-glycine cleavage site. Nine of these peptide signal sequences were found in more than 0.5% of genomes (i.e., over 20 genomes; Table 1), and together these nine comprise ~98% of all signal variants. All signals under this 0.5% threshold were each confined to a single clade in the whole-genome phylogeny (Fig. S1). Each unique BlpC signal was designated with a letter from the NATO phonetic alphabet (Table 1). As expected for the genomes from intentionally biased samples, the PMEN-1 data set almost exclusively carried the Golf signal (93.8%; Table S1), while the Clonal Complex 3 data set almost exclusively carried the Kilo signal (97.6%; Table 1). The Bravo and Hotel signal peptides were exclusively found in strains collected as part of the Maela data set. Even though the Maela genomes composed the majority of genomes in our data, the relatively high proportion of Maela genomes with the Bravo and Hotel signals (3.8% and 8.2%, respectively) suggests that either natural selection or limited admixture prevented these signals from appearing in the other data sets. There was more variation in *blpH* (194 alleles at the nucleotide level across 156 amino acid variants) than in *blpC*, although this is a likely consequence of the fact that *blpH* is longer than *blpC*. Rarefaction curves of non-singleton protein variants of BlpC and BlpH show that the diversity of protein variants reached saturation after ~2,000 and ~3,000 sampled genomes, respectively (Fig. S2).

Indicators of molecular evolution showed that *blpH* and *blpC* are evolving rapidly (Table 2). *blpH* had the highest peak of nucleotide diversity across the seven *blp* regulatory genes (Fig. S3). Additionally, higher $d_N$ / $d_S$ ratios (i.e. a higher non-synonymous to synonymous mutation ratio) in the receptor versus the kinase region of this protein (Table 2) suggests stronger diversifying selection acting on peptide:receptor binding than on the intracellular kinase domain responsible for downstream signal transfer by phosphorylation. Inferred rates of recombination peaked within the second transmembrane domain of BlpH; however, specific peaks of either nucleotide diversity or recombination were not distinguished by predicted transmembrane regions (Fig. S3). Recombination also locally peaked within *blpC* (Fig. S3), which also had an increased $d_N$ / $d_S$ ratio when compared to estimates for housekeeping genes (Table 2).

**_blpC/blpH_ intragenomic pairing is highly biased but not exclusive**

Phylogenetic analysis of *blpC* revealed four well-supported clades (Fig. 2) containing the following signals: 1) Alpha, Bravo, and Kilo; 2) Golf and Hotel; 3) Charlie; and 4) Delta, Echo, and Foxtrot. With the exception of the Delta signal, within-group signals are differentiated by a single amino acid or stop codon. The relationships between signaling groups within these major clades are uncertain, although there is evidence (0.75 < posterior probability < 0.95) that the Hotel, Bravo, and Delta signals are each monophyletic within their respective larger clades.

After accounting for recombination, phylogenetic analysis of the receptor domain of *blpH* (residues 1-229) identified five paraphyletic clades that are broadly concordant with the divisions observed for BlpC signals (Fig. 3), although there are many exceptions to this correspondence. Across the five clades, the classification of *blpH* alleles correlated with the BlpC signal in at least 75% of cases: (Alpha / Bravo / Kilo Clade: 86.6%; Echo / Foxtrot Clade: 90.0%; Delta Clade: 100%: Charlie Clade: 86.2%; Golf / Hotel Clade: 75.0%). Evidence of extensive recombination affecting the *blpH* kinase, intergenic region, and *blpC* signal (Fig. S3) suggests that recombination has caused some of these mismatches; however, multiple recombination events in the same region obscure reconstructing these evolutionary events with confidence. Overall, from the 4,002 genomes with full-length *blpH* genes, 667 genomes (16.7%) show a lack of correspondence between signal and peptide, suggesting either that these strains are deficient in *blp* signaling or that these BlpH histidine kinase receptors can be cross-induced by non-cognate BlpC signals. Overall frequencies by signal and receptor class are summarized in Fig. 4A.

**Crosstalk and eavesdropping between BlpC signals and BlpH receptors**

To examine the incidence of crosstalk and eavesdropping between signals and receptors experimentally, we measured the responsiveness of each of the major BlpH clades to synthetic peptides from each BlpC class. We transformed a *S. pneumoniae* D39 strain lacking the native *blp* regulatory genes (*blpSRHC*) with constructs expressing one of six different BlpH histidine kinases alleles: $blpSRH^{D39}$ from the Alpha/Bravo/Kilo clade, $blpSRH^{PMEN\text{-}2}$ from the

Echo/Foxtrot clade, *blpSRH*$^{\text{Hermans-33}}$ from the Delta clade, *blpSRH*$^{\text{Hermans-1012}}$ and *blpSRH*$^{\text{PMEN-14}}$ from the Charlie clade, and *blpSRH*$^{\text{PMEN-18}}$ from the Golf/Hotel clade. These strains also contained a reporter cassette, in which the *blp*-promoter from either $P_{blpK}$ or $P_{blpT}$ controlled expression of firefly luciferase (*luc*), GFP ((*sf*)*gfp*), and β-galactosidase (*lacZ*) (31). Deletion of the *blpC* signal gene and the native *blpSRH* genes from the D39 ancestor ensured that the reporter strains would only be induced in response to exogenously added signal via the introduced *blpSRH* systems. By exposing cells to a concentration gradient of exogenous peptide, we could estimate the peptide concentration that induced the maximum response as well the minimum concentration required to elicit a response. While the maximum response indicates the overall influence of a given peptide on each receptor, the minimal concentration required to induce a response provides an indication of the sensitivity of each receptor to every potential peptide partner.

Figures 4A-B shows that five of six $P_{blpK}$ reporter strains were maximally induced by the BlpC signal carried by a significant majority of their wild type counterparts. However, we also see extensive evidence for crosstalk and eavesdropping between mismatched peptide:receptor pairs, demonstrating that some BlpH receptors are highly promiscuous while equally, several BlpC peptides can induce the *blp* operon in strains carrying non-complementary BlpH receptors. For example, *blpSRH*$^{\text{PMEN-2}}$ (Echo / Foxtrot BlpH clade) could be induced by 4 out of 6 synthetic peptides, and the strain with *blpSRH*$^{\text{Hermans-1012}}$ (Charlie BlpH clade) was strongly induced by the Echo and Foxtrot signals at 65% and 71% expression of its cognate signal. While there is clear evidence for cross-induction, these responses tended to be less sensitive to non-cognate peptides, with a minimum concentration required for induction of between 2-500-fold greater than with the cognate signal (Fig. 4C). By contrast, the strain with *blpSRH*$^{\text{Hermans-1012}}$ (Charlie BlpH clade) was more sensitive to the non-cognate Echo and Foxtrot signals (1 ng/ml and 3.9 ng/ml) than to its complementary Charlie signal (7.8 ng/ml; Fig. 4C). The reporter strain carrying *blpSRH*$^{\text{Hermans-33}}$ did not respond to any of the BlpC peptides, not even its cognate Delta BlpC (Fig. 4B-C). Interestingly, *blpSRH*$^{\text{Hermans-33}}$, as well as all other strains with *blpH* alleles in the Delta clade, contains a frameshift in the *blpR* gene, encoding the response regulator, thus preventing expression of the full-length *blpR*. This probably renders the QS systems non-functional and therefore not responsive to added peptide. All results with $P_{blpK}$ were mirrored with a different set of reporter strains that used the *blpT* promoter for the reporter cassette (Fig. S4).

We conclude from these results that crosstalk among quorum-dependent peptide BlpC signals is widespread and concentration dependent, with strains able to eavesdrop onto multiple signals using cross-responsive receptors. Furthermore, these results are highly concordant with the patterns of co-association observed in our bioinformatics survey of pneumococcal strains. When only considering genomes carrying *blpC* and *blpH* alleles potentially capable of *blp* activation (as determined in Fig. 4B and 4C), 88.0 % of the strains are predicted to autoinduce *blp* expression under appropriate conditions, i.e., their genomes contain functionally active *blpC*/*blpH* pairs. Notably, however, this also indicates that a substantial proportion of strains (12.0%, 364 of 3046 genomes) may not be able to autoinduce *blp*

expression since they carry *blpC*/*blpH* pairs that were inactive in our experimental assay; this is in addition to strains carrying Delta *blpC/blpH*, which was also unable to autoinduce *blp* expression in our assay.

**Cross-induction between colonies**

Pneumococci in the nasopharynx live in spatially structured colonies or biofilms. In order to determine if cross-induction between signaling cells could occur under these conditions, we examined interactions between neighboring colonies endogenously secreting either cognate or non-cognate signals (Fig. 5). In control assays, we first demonstrated that colonies were induced by exogenous addition of peptide to the plate surface; these results were concordant with those in Figure 4B in 14 of 15 combinations (Fig. S5). Next, we measured expression of reporter strains when grown adjacent to wild-type colonies that secreted BlpC peptides at endogenous levels (Fig. 5A). We observed a response in the reporter strains as estimated by increased LacZ activity in 3 out of 6 strains, with 2 examples of induction by non-cognate BlpC signals. Interestingly, when the reporter strain expressing the BlpH from Hermans-1012 was grown adjacent to its wild type counterpart, there was no induction; instead this strain was induced by PMEN-14, which also produced the Charlie signal. The same strain was also induced by PMEN-2, which produced the Foxtrot signal (which induces Hermans-1012 at a lower concentration than with its cognate signal; Fig. 4C), and strain PMEN-18 (Golf/Hotel BlpH clade) was induced by PMEN-14, which produced the Charlie signal (Fig. 5). This may suggest that in addition to differences in the binding affinities of BlpC and BlpH, strains may also vary in the concentration of the diffusible signals that they secrete, at least under these experimental conditions. Consistent with our *in vitro* assays with synthesized peptides, these results show that *blp* operon expression can be activated by crosstalk between neighboring competing colonies secreting peptides at wild-type concentrations.

**Evolutionary consequences of eavesdropping genotypes**

Because the *blp* operon is auto-induced via a quorum dependent process, cross-induction can potentially influence other strains by lowering the population density required for auto-induction. To examine the possible effects of cross-induction on bacteriocins, we developed a spatially explicit stochastic model to investigate conditions where genotypes with eavesdropping receptors may be favored over strains only able to respond to a single peptide signal. We further varied the signal affinity to eavesdropping receptors to determine how this altered the selective benefits of cross-induction. Simulations are initiated with cells of four genotypes randomly spaced upon a plane. The four genotypes each release their own QS signal at equal concentrations (Table S2). Cells bind these secreted signals in a concentration dependent manner, at which point they are induced to produce bacteriocins that kill susceptible neighbor cells at the cost of reduced growth for the producer (37). While two faithful-signaling genotypes are only able to respond to their own signals, the two other eavesdropping genotypes can respond to multiple signals. Our results shown in Fig. 6 lead to

two conclusions. First, we observe strong benefits to eavesdropping cells that depends on the degree of cross-sensitivity, or affinity, to non-cognate signals. Specifically, we found that higher affinity to non-cognate signal provides stronger ecological benefits. This results from earlier potential activation (Fig. S4) and secretion of bacteriocins in these cells, an effect that increases with greater affinity to non-cognate signals. Second, we find that the benefits to eavesdropping are strongly negative frequency-dependent, i.e. eavesdropping cells only gain benefits (in the form of earlier bacteriocin induction) when surrounded by faithful-signaling cells. When eavesdropping cells are rare, they benefit through maximum exposure to the alternative peptide, while after they increase in frequency they must rely solely on auto-induction. Because the benefits of eavesdropping are frequency-dependent, these simple simulations thus suggest that promiscuous receptor mutants with increased affinity to non-cognate signals will be able to rapidly invade populations of cells that can only respond to a single signal. Interestingly, the simulations also clarify that the affinity to non-cognate signals can be extremely low — even at 10% of the affinity to cognate signals — to provide benefits (Fig. 6).

## Discussion

Pneumococcal bacteriocins are believed to play a key role in mediating intraspecific competitive interactions (25). Here, we show that the QS system regulating *blp* bacteriocins is highly polymorphic, that there is widespread evidence that QS signals are cross-reactive (crosstalk), and that promiscuous receptors can detect and respond to non-cognate signals (eavesdropping). Assays between adjacent colonies revealed that both behaviors occur at endogenous concentrations of secreted peptides, while simulations revealed ecological benefits to strains that express promiscuous receptors. Together, these results suggest that social interactions influenced by QS signaling may have a strong influence on pneumococcal competition.

Previous surveys (28,38) of BlpC and BlpH identified four BlpC signals: the Alpha, Charlie, Foxtrot, and Golf signals in our nomenclature, which together represent ~75% of the strains in our sample (Table 1). By expanding our survey to thousands of strains, we identified several additional signal peptide families (Fig. 2): the Echo, Hotel, Delta, Bravo, and Kilo signals. Two of these (Bravo and Hotel) were found only in the Maela dataset, consistent with the idea of strong geographic structuring in this species; furthermore, because these signal variants appear in significant frequencies (3.8% and 8.1%, respectively, within the Maela data set), we infer that they are not actively selected against. The additional signals reported here suggest flexibility (and perhaps diversifying selection) in the mature signal peptide sequence, such as in the first two residues (which differ between the Alpha, Bravo, and Kilo signals, which share a co-occurring BlpH clade) and in signal residue 22, which differs between the Echo and Foxtrot signals despite the signals activating BlpH variants near identically in our experiments (Fig. 4B, Fig. 4C). Previous work suggested that differences in the electric charge of signal residue 14 is crucial for specificity (36). This residue is undoubtedly

important because it differentiates the Alpha/Bravo/Kilo, Echo/Foxtrot, and Golf/Hotel signal groups; however, other signal residues are also likely to be important for BlpH binding, as signals that are identical at this site 14 (e.g. Golf/Hotel, and Alpha/Bravo/Kilo/Charlie) differentially bind/activate BlpH (Fig. 4). Although rarefaction analysis indicates that we have essentially saturated the diversity of signal and receptor types (Fig. S2), more extensive sampling in other geographic locations, such as with the Global Pneumococcal Sequencing project (http://www.pneumogen.net/gps), will likely uncover further rare variants.

The concordance between the phylogenies of *blpC* and *blpH* and the extensive co-occurrence in individual genomes suggest that these genes are co-evolving (Fig. 2, Fig. 3). At the same time, both *blpC* and *blpH* are changing rapidly, as indicated by their relatively high levels of nucleotide diversity (Fig. S3) and their high non-synonymous / synonymous substitution rates (Table 2), which is consistent with the idea that the genes are evolving non-neutrally. The co-variation between BlpC and BlpH also allows inferences on the key residues within each gene that mediate their binding. We used mutual information based upon this co-variation to identify residues in BlpH that are correlated with the QS signal, and these results can serve as a guide for experimental approaches to unravel the specificity of the BlpC/BlpH interaction (Fig. S6, S7). Notably, our analysis of the BlpH receptor residues support previous findings that residues 17 and 119-124 are important for activation by BlpC signals (36), although additional residues co-vary with specific BlpC signals (Fig. S6, S7).

While the correlation between *blpH* clade and co-occurring BlpC signal is high, in some clades the correlation drops to 75.0%, and BlpH / BlpC mismatches (Fig. 3) are widespread across the pneumococcal phylogeny. This can be compared to the exceptionally tight, > 99% correlation between the ComD QS receptor and CSP signal also in *S. pneumoniae* (35). There are at least two explanations for this difference. First, we do not know if different BlpH variants are functionally distinct; all *blpH* alleles could, in principle, be most responsive to their co-occurring BlpC. This seems unlikely, given the high frequency (up to 45 signal:receptor pairs) of *blpH* clade / BlpC mismatches (Fig. 3). Second, weaker selection for a highly auto-inducing *blp* QS could explain the difference between the *blp* and *com* QS systems. After a recombination event that results in a sub-optimal BlpH/ BlpC pair for auto-induction, the BlpC signal may still be able to activate the co-occurring BlpH variant through crosstalk, albeit at a higher concentration of BlpC (Fig. 4C). While auto-induction may be decreased, such a genotype would gain an eavesdropping receptor that can potentially detect signals of surrounding genotypes. For comparison, there is no eavesdropping between CSP pherotypes in the *com* QS system, and very rare signal/receptor mismatches (34,35).

These signal/receptor mismatches can result in two outcomes for cell-to-cell communication. First, this can result in cells that are unable to detect the signal that they produce, rendering them unable to auto-induce. The lack of QS activation in strains producing the Delta signal (Hermans-33; Fig. 4) seemingly fits into this description; however, interestingly, this is not caused by signal / receptor mismatch because there is perfect concordance between the Delta

signal and the Delta *blpH* clade, and no tested signal activated strains with Delta *blpH*. Instead, all 143 strains carrying the Delta signal have a frameshift in *blpR*, which suggests functional deterioration of the QS system in these strains, which has not yet led to deterioration of *blpH* and *blpC*. These Delta BlpC strains are not simply 'cheater' cells, as they potentially continue to pay the cost of synthesizing BlpC if the *blpC* gene is actively transcribed. This suggests there may be weakened selection for functional *blp* QS.

The second outcome of signal/receptor mismatches for cell-to-cell communication is crosstalk and eavesdropping. We have ample evidence for crosstalk in the *blp* QS system, as all signal peptides except for the Alpha signal activated QS receptors in genotypes that produce other QS signals (Fig. 4B, Fig. 4C). Similarly, BlpH receptors (aside from the Alpha clade) were eavesdropping QS receptors able to detect more than one QS peptide signal (Fig. 4B, Fig. 4C). Each of the receptors we tested (except for the signal-blind BlpH Delta clade) was maximally induced with a single set of related signals and decreased to 0-71% with signals that the receptors were eavesdropping upon (Fig 4B). This suggests that there are no 'generalist' receptors that are able to listen to multiple signals with equal affinity. Crosstalk was observed in previous research on the *blp* system (37; see asterisks in Fig. 4C and Table 1 alternative signal names), and results from this study indicated that *blpH* alleles with more crosstalk were less sensitive to BlpC (36). However, the results reported here show that receptors from strains PMEN-2, Hermans-1012, and PMEN-14 were all highly sensitive to their complementary signal ($\leq$1.0 ng/ml) despite showing extensive crosstalk (Fig. 4C), thereby suggesting that the trade-off between crosstalk and sensitivity of *blpH* alleles is not a general phenomenon.

What are the potential consequences of crosstalk and eavesdropping? The result of crosstalk could be to manipulate other, non-clonal, strains into inducing their QS system at lower densities, thereby causing competing strains to secrete bacteriocins and induced immunity proteins earlier. At present, it is unclear how such crosstalk would be beneficial to cells producing cross-reactive signals, unless premature production of bacteriocins or immunity introduces energetic or other costs to cells responding at sub-quorum densities. Similar benefits are thought to exist for other bacterial public goods (8,39). By contrast, it is easier to envision the potential benefits of eavesdropping, which can both lead to earlier activation of bacteriocins (although this may also have attendant costs) and earlier induction of cross-reactive immunity. Importantly, our simulations suggest that this could be beneficial even if the affinity of promiscuous receptors is only 10% of the affinity for their cognate signal (Fig. 6). This value falls within the range of responses we measured experimentally (Fig. 4C). This level of responsiveness is also sufficient to induce the *blp* operon among adjacent colonies secreting peptides at endogenous levels (Fig. 5).

How does this amount of crosstalk specifically affect bacteriocin-mediated competition between *S. pneumoniae* strains? Three factors make it difficult to answer this question conclusively. First, the extensive variation in the kinase domain of BlpH, the response regulator BlpR, and the leader sequences of the *blp* bacteriocins (27) prevent a full understanding how signal concentrations translate into increased concentrations of exported bacteriocins. A systematic

approach to investigate each of these molecules and their variants in the laboratory will be required to address this question. This includes investigating the role of bacteriocin immunity, which could drive additional, immunity-based effects from crosstalk and eavesdropping. Second, a bioinformatics approach to examine evidence of selection in coordination with the BlpH receptor or BlpC signal is not possible due to the inability to align the entire *blp* operon and because recombination breaks up potential associations that are otherwise selected for. Third, the effects of crosstalk and eavesdropping will also depend on the activation of the *com* QS system, which promotes the expression and export of *blpC* at a low level (Fig. 1), even when the ABC-transporter genes *blpAB* are disrupted by early termination mutations (31,33). For example, we found that both wildtype strains D39 and PMEN-14 could activate *blp* expression in neighboring colonies (Fig. 5) despite having disrupted *blpA* (for PMEN-14) or disrupted *blpA* and *blpB* (for D39).

Signaling interactions *in vitro* can lead to complex ecological outcomes that may strongly influence competitive interactions between strains. As yet, however, it is unclear how these interactions will play out in the complex within-host environment of the human nasopharynx. In addition, it remains unclear how these interactions directly influence bacteriocin-mediated killing and immunity. Clearly, the heterogeneous conditions *in vivo* differ markedly between liquid cultures or even agar surfaces. Diffusion is more limited, while population densities may be strongly constrained overall and spatially. These factors, among others, may alter the level and dispersion of signal peptides as well as the sensitivity of individual strains to these signals. It will remain an important aim for future work to elucidate the influence of these real-world factors. More generally, our results reinforce the importance of social interactions among bacteria for mediating competitive dynamics. Many ecologically relevant bacterial traits are regulated by QS, and many of these systems, especially in Gram-positive peptide signaling systems, are polymorphic. While some of these systems (e.g. pneumococcal competence regulated by the *com* QS system) have only few signal types and show no cross-reactivity, many others signal are polymorphic with substantial cross-reactivity (e.g. Agr in *S. aureus* (17) and ComX in *B. subtilis* (40)). It remains to be investigated which of these polymorphic QS signals have ecological effects and which factors (such as co-colonization or extensive intraspecific competition) result in the evolution of crosstalk and eavesdropping.

## Materials and Methods

### Genomic data

We used *S. pneumoniae* genomes from eight publicly available sets, six of which contain strains that were randomly sampled from either cases of disease or asymptomatic carriage: 3,017 genomes from refugees in Maela, Thailand (41); 616 genomes from Massachusetts carriage strains (42); 295 genomes from GenBank, which include 121 genomes from Atlanta, Georgia, The United States (43); 142 genomes from Rotterdam, the Netherlands (Hermans data set) (27,44); and 26 PMEN (Pneumococcal Molecular Epidemiology Network) genomes (27,45). The PMEN-1 (46) and

Clonal Complex 3 (47) data sets, containing 240 and 82 genomes, respectively, were a result of targeted sampling for specific clonal complexes of *S. pneumoniae*; as such, these strains were excluded from analyses that assumed random sampling. In each of these genomes, we located the *blpC* and *blpH* alleles using a DNA reciprocal BLAST search as previous described (27). Genome assemblies were deposited in the European Nucleotide Archive with sample-IDs ERS852964 – ERS853134.

**Phylogenetic Analysis**

We examined if the rates of transition in changing BlpC signals were independent from the rates of transition between bacteriocin groups along the whole-genome phylogeny from (27). We did so by measuring the likelihood while estimating both independent and dependent transitions using BayesTraits (48) for each signal and bacteriocin group separately. We used a log ratio test to measure P-values from the likelihoods.

To infer the evolutionary history of *blpC*, we aligned the 37 alleles of *blpC* from all genomes, removed nucleotide sites caused by insertion mutations in single alleles, and reconstructed the phylogeny using Geneious 7.1.9 (49) and MrBayes 3.2.2 (49–51) with an HKY+γ nucleotide substitution model as determined by jModelTest 2.1.7 (52).

Because *blpH* contained evidence of recombination, we modified our approach to infer its evolutionary history. We focused on the receptor domain of *blpH* (residues 1-229) aligned the unique 163 *blpH* alleles using Geneious 7.1.9 (49). We then used Gubbins 1.4.2 (53) to detect and place recombination events onto a phylogenetic tree. To reconstruct the phylogeny while measuring the confidence for each clade, we used the filtered polymorphic sites from Gubbins, which replaced recombination fragments with N's, as input for MrBayes 3.2.2 (49–51) using a GTR+γ nucleotide substitution model as determined by jModelTest 2.1.7 (52).

**Sequence Analysis**

We measured the nucleotide diversity and $d_N$ / $d_S$ ratio of *blpH*, *blpC*, as well as the seven housekeeping genes used for MLST analysis (*aroE*, *ddlA*, *gdhA*, *glkA*, *lepB*, *recP*, and *xpt*) using DNAsp 5.10.01 (54). We separated *blpH* after nucleotide 687 (residue 229) to measure these values for the receptor and kinase domains separately (36).

We extracted the region 1500 bp before *blpT* to 100bp after *blpB* from the 4,096 randomly sampled genomes from a previous alignment with R6_uid57859 (35), which included sequences with interrupted BlpA and BlpB genes (11). We measured nucleotide entropy (measured as Shannon entropy by site) and recombination breakpoints (measured from previously calculated data from GeneConv 1.81a (35,55) for these sequences. Using only pairwise recombination breakpoints with unique start and stop positions prevented overestimating signs of recombination events

413 caused by the common descent of strains.

414 Mutual information was calculated using a custom Python script. For each signal, we grouped signals into
415 signal groups, such as Alpha:Bravo:Kilo or Golf:Hotel. We grouped amino acids as either acidic, aliphatic, aromatic,
416 basic, cyclic, or hydroxyl residues. We then normalized mutual information to correct for the entropy of signals and of
417 amino acids at each site.

418 Transmembrane domains in BlpH were predicted using TOPCONS 2.0 (56).

419

420 **Bacterial strains and growth conditions**

421 *S. pneumoniae* strains were grown as liquid cultures in C+Y medium (57) at 37°C and transformed as described
422 previously (31). For selection, *S. pneumoniae* was plated on Columbia agar plates supplemented with 2% defibrinated
423 sheep blood (Johnny Rottier, Kloosterzande, Netherlands) and 1 µg/ml tetracycline, 100 µg/ml spectinomycin or 0.25
424 µg/ml erythromycin, when appropriate. *E. coli* was grown in LB medium with shaking at 37°C or plated on LA
425 containing 100 µg/ml ampicillin.

426

427 Strain construction

428 Strains and plasmids used in this study are listed in Table S3.

429

430 Constructs for expression of *blpSRH* from different strains in *S. pneumoniae* D39.

431 The *blpSRH* genes, including the constitutive *blpS*-promoter (31), was amplified from the genome of *S.*
432 *pneumoniae* strains D39, PMEN-18 and Hermans-1012 using primers blpS-F-ClaI-SphI and blpH-R-Hermans-1012-
433 NotI-SpeI, from PMEN-2, using primers blpS-up-F-PMEN2-SphI and blpS-down-R-PMEN2-SpeI-NotI from PMEN14
434 with primers BlpS-PMEN14-F-SphI and BlpH-R-Hermans-1012-NotI-SpeI and from Hermans-33 with primers blpS-F-
435 ClaI-SphI and blpH-Hermans33/35-R-NotI-SpeI. The PCR products were digested with SphI and NotI and ligated into
436 the corresponding sites of plasmid pJWV25 (between the *bgaA* homology regions) and transformed into *E. coli* DH5α.
437 The resulting plasmids were verified by PCR and sequencing. The plasmids were then transformed into *S. pneumoniae*.
438 Correct integration of the $P_{blpS}$-*blpSRH* constructs into the non-essential *bgaA*-locus was verified by PCR. Primers used
439 for these constructs are listed in Table S4.

440

441 Deletion of *blpSRHC*.

442 The native *blp*-regulatory genes (*blpS*, *blpR*, *blpH*, *blpC*) of *S. pneumoniae* D39 were deleted by replacement
443 with an erythromycin-resistance cassette as described previously (31).

444

445 Reporter constructs

446 Two different *blp* promoters were used to monitor *blp*-expression; $P_{blpK}$, controlling expression of the bacteriocin
447 *blpK* and $P_{blpT}$, controlling expression of the functionally uncharacterized gene *blpT*. The $P_{blpK}$ and $P_{blpT}$ promoters have
448 previously been shown to be co-regulated (31) and act as reporters for BlpR activation across the tested pneumococcal
449 strains (as there is almost no variation in the DNA-binding motif of different BlpR). The reporter constructs $P_{blpK}$-*luc-*
450 *gfp-lacZ* and $P_{blpT}$-*luc-gfp-lacZ* (a tripartite reporter cassette) integrated into the non-essential *cep*-locus of *S.*
451 *pneumoniae* D39, has been described previously (31).

452

453 Luciferase assays

454 Luciferase assays were performed essentially as described before (31,58). Briefly, *S. pneumoniae* cultures pre-
455 grown to $OD_{600}$ 0.4 were diluted 100-fold in C+Y medium (pH 6.8) with 340 µg/ml luciferin. Luc-activity was
456 measured in 96-well plates at 37°C, and $OD_{600}$ and luminescence (as relative luminescence units, RLU) were
457 recorded every 10 minutes using Tecan Infinite 200 PRO. Synthetic peptides (BlpCs) were purchased from Genscript
458 (Piscataway, NJ). Different concentrations of BlpCs were added to the culture wells after 100 min or in the beginning of
459 the experiment, depending on the experiment. The data was plotted as RLU/OD over time to analyze induction of *blp*
460 expression.

461

462 **LacZ assays on agar plates**

463 LacZ assays for testing induction by neighbouring colonies on plates were performed on C+Y agar (pH 8.0)
464 covered with 40 µl of 40 mg/ml solution X-gal (spread on top of the plates). All strains were pre-grown to $OD_{600}$ 0.4,
465 before 2 µl of the wild-type strains (BlpC producers) were spotted and allowed to dry. Then 2 µl of the different
466 reporter strains were spotted next to the dried spot. The plates were incubated at 37°C over-night.

467 For induction with synthetic BlpC, C+Y agar plates (pH 7.2) were covered with 40 µl of 40 mg/ml solution X-
468 gal and 5 µl 1 mg/ml BlpC (spread on top of the plates), and different reporter strains were spotted on top. The plates
469 were incubated at 37°C over-night.

470

471 **Stochastic Model**

472 We built a spatial, stochastic model in which cells are modeled individually and interact in a grid. At each
473 discrete time point, each cell can divide with probability 80%, which produces an identical offspring in an empty,
474 randomly chosen adjacent position on the grid. Each cell also secretes one of four signaling molecules, which
475 accumulate and diffuse in the space around the cell (defined as “diffusion area”). When the amount of signaling

molecules within the sensitivity area (which is an area with half the radius of the diffusion area) of a cell reaches a defined threshold, this cell becomes induced and starts producing bacteriocins. A bacteriocin-producing cell can kill up to six neighboring cells depending on their genotype, as explained below. Induced cells, which produce bacteriocins, grow 20% slower than uninduced cells. Every cell has a 0.1% probability of death at each time point.

We modeled four genotypes, which differ in the signaling molecule and bacteriocins that they produce as well as in the number and identity of signals that they respond to (Table S2). Bacteriocins produced by genotypes 1 and 2 specifically could kill genotypes 3 and 4 and vice versa. Signals produced by genotype 1 could induce genotypes 1 and 2 and similarly, signals produced by genotype 3 could induce genotypes 3 and 4; we therefore classify genotypes 2 and 4 as “eavesdropping genotypes”. Genotypes 1 and 3 can only respond to their own signal, as “signal-faithful genotypes”. All four genotypes have equivalent growth rates, which are only variable depending on if a cell is induced or uninduced. Eavesdropping cells respond to signals that they do not produce with certain degrees of affinity. If we consider the affinity of a cell to its own signal as 100%, we ranged the affinity to the other signals in the case of eavesdropping genotypes as 0% - 90% for different simulations.

We used an initial grid of 400 by 400 positions, and we started with 10% of the grid randomly populated with eavesdropping and signal-faithful phenotypes at a range of proportions. As cells grow and interact, the center of the grid remains occupied, while the cell population can expand on the boundaries past the initial 400 by 400 grid. Simulations were run for until the number of cells reached 110% of initial grid size. We calculated fitness as the difference between relative frequencies of eavesdropping cells at the last time point and at the initial time point, i.e.:

$$(\text{Eavesdropping cells/Total cells})_{\text{Final Time Point}} - (\text{Eavesdropping cells/Total cells})_{\text{Initial Time Point}}$$

**Acknowledgements**

We would like to thank Frank Lake for technical assistance.

**Funding**

This work was supported by the Biotechnology and Biological Sciences Research Council (grant number BB/J006009/1) to DER and ISR and by the Wellcome Trust (105610/Z/14/Z) to the University of Manchester. MA is supported by the Biotechnology and Biological Sciences Research Council (grant number BB/M000281/1). Work in the Veening lab is supported by the EMBO Young Investigator Program, a VIDI fellowship (864.12.001) from the Netherlands Organisation for Scientific Research, Earth and Life Sciences (NWO-ALW) and ERC starting grant 337399-PneumoCell. MK is supported by a grant from The Research Council of Norway (250976/F20).

**References**

1. Miller MB, Bassler BL. Quorum sensing in bacteria. Annu Rev Microbiol. 2001;55:165–99.

2. Waters CM, Bassler BL. Quorum Sensing: Cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol. 2005;21(1):319–46.

3. Redfield RJ. Is quorum sensing a side effect of diffusion sensing? Trends Microbiol. 2002;10(8):365–70.

4. Bassler BL, Greenberg EP, Stevens AM. Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*. J Bacteriol. 1997;179(12):4043–5.

5. Strassmann JE, Gilbert OM, Queller DC. Kin discrimination and cooperation in microbes. Annu Rev Microbiol. 2011;65(1):349–67.

6. Schluter J, Schoech AP, Foster KR, Mitri S. The evolution of quorum sensing as a mechanism to infer kinship. PLOS Comput Biol. 2016;12(4):e1004848.

7. West SA, Griffin AS, Gardner A, Diggle SP. Social evolution theory for microorganisms. Nat Rev Microbiol. 2006;4(8):597–607.

8. Diggle SP, Griffin AS, Campbell GS, West SA. Cooperation and conflict in quorum-sensing bacterial populations. Nature. 2007;450(7168):411–4.

9. Heurlier K, Dénervaud V, Haas D. Impact of quorum sensing on fitness of *Pseudomonas aeruginosa*. Int J Med Microbiol. 2006;296(2–3):93–102.

10. Eldar A. Social conflict drives the evolutionary divergence of quorum sensing. Proc Natl Acad Sci. 2011;108(33):13635–40.

11. Son MR, Shchepetov M, Adrian P V, Madhi SA, de Gouveia L, von Gottberg A, et al. Conserved mutations in the pneumococcal bacteriocin transporter gene, *blpA*, result in a complex population consisting of producers and cheaters. MBio. 2011;2(5).

12. Pollak S, Omer-Bendori S, Even-Tov E, Lipsman V, Bareia T, Ben-Zion I, et al. Facultative cheating supports the coexistence of diverse quorum-sensing alleles. Proc Natl Acad Sci. 2016;113(8):2152–7.

13. Atkinson S, Williams P. Quorum sensing and social networking in the microbial world. J R Soc Interface. 2009;6(40):959–78.

14. Bouillaut L, Perchat S, Arold S, Zorrilla S, Slamti L, Henry C, et al. Molecular basis for group-specific activation of the virulence regulator PlcR by PapR heptapeptides. Nucleic Acids Res. 2008;36(11):3791–801.

15. Swem LR, Swem DL, Wingreen NS, Bassler BL. Deducing receptor signaling parameters from *in vivo* analysis: LuxN/AI-1 quorum sensing in *Vibrio harveyi*. Cell. 2008;134(3):461–73.

16. Ansaldi M, Dubnau D. Diversifying selection at the Bacillus quorum-sensing locus and determinants of modification specificity during synthesis of the ComX pheromone. J Bacteriol. 2004;186(1):15–21.

17. Ji G, Beavis R, Novick RP. Bacterial interference caused by autoinducing peptide variants. Science. 1997;276(5321):2027–30.

18. Regev-Yochay G, Raz M, Dagan R, Porat N, Shainberg B, Pinco E, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* by adults and children in community and family settings. Clin Infect Dis. 2004;38(5):632–9.

19. Wyllie AL, Chu MLJN, Schellens MHB, Gastelaars JVE, Jansen MD, Van Der Ende A, et al. *Streptococcus pneumoniae* in saliva of Dutch primary school children. PLoS One. 2014;9(7):1–8.

20. Sauver JS, Marrs CF, Foxman B, Somsel P, Madera R, Gilsdorf JR. Risk factors for otitis media and carriage of multiple strains of *Haemophilus influenzae* and *Streptococcus pneumoniae*. Emerg Infect Dis. 2000;6(6):622–30.

21. García-Rodríguez JA, Fresnadillo Martínez MJ. Dynamics of nasopharyngeal colonization by potential respiratory pathogens. J Antimicrob Chemother. 2002;50 Suppl S:59–73.

22. Brugger SD, Frey P, Aebi S, Hinds J, Muhlemann K. Multiple colonization with *S. pneumoniae* before and after introduction of the seven-valent conjugated pneumococcal polysaccharide vaccine. PLoS One. Public Library of Science; 2010;5(7):e11638.

23. Meats E, Brueggemann AB, Enright MC, Sleeman K, Griffiths DT, Crook DW, et al. Stability of serotypes during nasopharyngeal carriage of *Streptococcus pneumoniae*. J Clin Microbiol. 2003;41(1):386–92.

24. Turner P, Turner C, Jankhot A, Helen N, Lee SJ, Day NP, et al. A longitudinal study of *Streptococcus pneumoniae* carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. PLoS One. 2012;7(5).

25. Dawid S, Roche AM, Weiser JN. The *blp* bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both *in vitro* and *in vivo*. Infect Immun. 2007 Jan;75(1):443–51.

26. Lux T, Nuhn M, Hakenbeck R, Reichmann P. Diversity of bacteriocins and activity spectrum in *Streptococcus pneumoniae*. J Bacteriol. 2007 Nov;189(21):7741–51.

27. Miller EL, Abrudan MI, Roberts IS, Rozen DE. Diverse ecological strategies are encoded by *Streptococcus pneumoniae* bacteriocin-like peptides. Genome Biol Evol. 2016;8(4):1072–90.

28. De Saizieu A, Gardes C, Flint N, Wagner C, Kamber M, Mitchell TJ, et al. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. J Bacteriol. 2000;182(17):4696–703.

29. Reichmann P, Hakenbeck R. Allelic variation in a peptide-inducible two-component system of *Streptococcus pneumoniae*. FEMS Microbiol Lett. 2000;190(2):231–6.

30. De Saizieu a., Gardes C, Flint N, Wagner C, Kamber M, Mitchell TJ, et al. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. J Bacteriol. 2000;182(17):4696–703.

31. Kjos M, Miller E, Slager J, Lake FB, Gericke O, Roberts IS, et al. Expression of *Streptococcus pneumoniae* bacteriocins is induced by antibiotics via regulatory interplay with the competence system. PLOS Pathog. 2016;12(2):e1005422.

32. Håvarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol Microbiol. 1995;16(2):229–40.

33. Wei-Yun W, Kochan TJ, Storck DN, Dawid S. Coordinated bacteriocin expression and competence in *Streptococcus pneumoniae* contributes to genetic adaptation through neighbor predation. PLoS Pathog. 2016;

34. Iannelli F, Oggioni MR, Pozzi G. Sensor domain of histidine kinase ComD confers competence pherotype specificity in *Streptoccoccus pneumoniae*. FEMS Microbiol Lett. 2005;252(2):321–6.

35. Miller EL, Evans BA, Cornejo OE, Roberts IS, Rozen D. Pherotype polymorphism in *Streptococcus pneumoniae* and its effects on population structure and recombination [Internet]. bioRxiv. 2016. Available from: http://biorxiv.org/content/early/2016/08/17/070011.abstract

36. Pinchas MD, LaCross NC, Dawid S. An electrostatic interaction between BlpC and BlpH dictates pheromone specificity in the control of bacteriocin production and immunity in *Streptococcus pneumoniae*. J Bacteriol. 2015;197(7):1236–48.

37. Ruparell A, Dubern JF, Ortori CA, Harrison F, Halliday NM, Emtage A, et al. The fitness burden imposed by synthesizing quorum sensing signals. Sci Rep. Nature Publishing Group; 2016;6(August):33101.

38. Reichmann P, Hakenbeck R. Allelic variation in a peptide-inducible two-component system of *Streptococcus*

*pneumoniae*. FEMS Microbiol Lett. 2000 Sep;190(2):231–6.

39. West SA, Winzer K, Gardner A, Diggle SP. Quorum sensing and the confusion about diffusion. Trends Microbiol. Elsevier Ltd; 2012;20(12):586–94.

40. Stefanic P, Decorosi F, Viti C, Petito J, Cohan FM, Mandic-Mulec I. The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. Environ Microbiol. 2012;14(6):1378–89.

41. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet. 2014;46(3):305–9.

42. Croucher NJ, Finkelstein J a, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. Nature Publishing Group; 2013;45(6):656–63.

43. Chancey ST, Agrawal S, Schroeder MR, Farley MM, Tettelin HH, Stephens DS. Composite mobile genetic elements disseminating macrolide resistance in *Streptococcus pneumoniae*. Front Microbiol. 2015;6(February):1–14.

44. Bogaert D, Engelen MN, Timmers-Reker AJM, Elzenaar KP, Peerbooms PGH, Coutinho RA, et al. Pneumococcal carriage in children in the Netherlands: A molecular epidemiological study. J Clin Microbiol. 2001;39(9):3316–20.

45. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, et al. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. J Clin Microbiol. 2001;39(7):2565–71.

46. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011;331(6016):430–4.

47. Croucher NJ, Mitchell AM, Gould KA, Inverarity D, Barquist L, Feltwell T, et al. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. PLoS Genet. 2013;9(10).

48. Pagel M, Meade A. BayesTraits v.2. 2013.

49. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.

50. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics.

2001;17(8):754–5.

51. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–4.

52. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. Nature Publishing Group; 2012;9(8):772–772.

53. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43(3):e15–e15.

54. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.

55. Sawyer SA. GENECONV: A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis; 1999.

56. Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. 2015;43(May):1–7.

57. Martin B, Garcia P, Castanié M-P, Claverys J-P. The *recA* gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction. Mol Microbiol. 1995;15(2):367–379.

58. Slager J, Kjos M, Attaiech L, Veening J-W. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. Cell. 2014 Oct;157(2):395–406.

## Figure Legends

**Figure 1. QS eavesdropping, crosstalk, and regulation.** A) Eavesdropping occurs when a QS receptor of a cell is activated by a QS signal that the cell does not produce, such as activation of the blue QS receptor by both the cognate blue square signal and non-cognate green triangle signal. Crosstalk occurs when a QS signal activates more than one receptor, such as the green triangle signal activating both the cognate green QS receptor and the non-cognate blue QS receptor. B) *blp* QS regulation. External BlpC signal binds to histidine kinase receptor BlpH. This activates response regulator BlpR through phosphorylation, which increases transcription of *blpABC*, *blpT*, the *blp* bacteriocins (including *blpK*), and immunity genes. Pre-BlpC is processed and transported out of the cell by ABC transporters ComAB and BlpAB. Similarly, QS signal CSP binds to histidine kinase receptor ComD, thereby phosphorylating response regulator ComE, which increases transcription of the *blp* operon (although to a lower amount than BlpR) as well as *com*-specific genes.

**Figure 2.** Bayesian unrooted phylogenetic tree of *blpC*. Taxa are colored by mature BlpC signal with the signal designation followed by the number of genomes containing the allele. Internal nodes show the posterior probabilities of clades; we collapsed clades with less than 0.75 posterior probability.

**Figure 3.** Bayesian unrooted phylogenetic tree of *blpH* alleles. The outer ring shows the number of 4,096 genomes with each *blpH* allele, color-coded by their co-occurring BlpC signal and on a log scale. The inner ring denotes the *blpH* clade type, and recombination events within *blpH* are shown as solid green lines. Mismatches between *blpH* clade and BlpC signal are indicated by dashed lines. Internal nodes show the posterior probabilities of clades; we collapsed clades with less than 90.0% posterior probability.

**Figure 4.** A) Proportion of each BlpC signal within genomes containing each *blpH* clade. The phylograms are simplified versions of Fig. 1 and Fig. 2. B) The relative maximal expression levels of *luc* following addition of 1 µg/ml of synthesized BlpC signal peptide. The maximum expression level for each reporter strain was set to 1. Raw data is found is Supplementary Fig. 8 C) The minimum concentration of synthesized BlpC signal peptide required for *luc* induction in reporter strains with different BlpH. Asterisks indicate receptor/signal pair *blp* activation reported in (36). Example of raw data is provided in Supplementary Fig. 9 The Bravo, Kilo, and Hotel signal peptides were not synthesized and are denoted with slashes.

**Figure 5.** LacZ induction by neighboring colonies on agar plates. A) The wild-type strains were spotted next to the reporter strains (see box), and induction of *blp* expression by the wild-type produced BlpC is shown as faint blue

671 colonies. The experiment was repeated three times with the same result, and a representative photo of the plates is
672 shown. B) Summary of the results from B. Squares in white indicate no induction of the reporter strain for colony pairs,
673 while black and blue indicate induction by complementary and on-complementary BlpCs, respectively.
674
675 **Figure 6.** Average fitness of eavesdropping genotypes that produce bacteriocins in response to multiple signals in a
676 spatially explicit, stochastic model. Simulations were started with five proportions of eavesdropping genotypes mixed
677 with signal-faithful genotypes, as indicated on the x-axis. Absolute fitness values on the y-axis above 1.0 indicate that
678 the genotype can increase in frequency in the population. Affinity to other genotypes' signals are a percentage of
679 affinity to a genotype's own signal for eavesdropping genotypes. Error bars link the 25% and 75% quantiles for the
680 final eavesdropping genotypes' fitness across 100 simulations.

**Table 1.** Predicted *blpC* mature peptide signals in 4,096 genomes from six unbiased genome sets

| Signal | Signal Group | Mature AA Sequence[a] | Frequency in Unbiased Sets | Genomes with Mature Signal | Reference Genomes with Mature Signal | Alternative Name for Signal |
|---|---|---|---|---|---|---|
| Alpha | Alpha/Bravo/Kilo | **G**W**WE**EL**L**HETILSKFKI**TK**ALEL**PIQL** | 0.125 | D39 | D39, R6 | R6 |
| Bravo | Alpha/Bravo/Kilo | **G**L**WE**EL**L**HETILSKFKI**TK**ALEL**PIQL** | 0.028 | | | |
| Kilo | Alpha/Bravo/Kilo | E**WE**EL**L**HETILSKFKI**TK**ALEL**PIQL** | 0.007 | OXC141 | | |
| Charlie | Charlie | **G**L**WE**DI**L**YSLNIIKHNN**TK**GLHH**PIQL** | 0.291 | Hermans-1012, PMEN-14 | SPNA45, INV200 | 6A |
| Delta | Delta | **G**W**W**KDL**L**HRFNVIEQNN**TK**GFNQ**PIQL** | 0.035 | Hermans-33 | ECC_3510 | |
| Echo | Echo/Foxtrot | **G**W**WE**DF**L**YRFNIIEQKN**TK**GFHQ**PIQL** | 0.087 | | | |
| Foxtrot | Echo/Foxtrot | **G**W**WE**DF**L**YRFNIIEQKN**TK**GFYQ**PIQL** | 0.125 | PMEN-2 | P164 | P164 |
| Golf | Golf/Hotel | **G**LWEDL**L**YNINRYAHYI**T** | 0.223 | PMEN-18 | TIGR4, INV104 | T4 |
| Hotel | Golf/Hotel | **G**L**WE**DL**L**YNINRYAHYI**T**QELHH**PIQL** | 0.060 | | | |
| Other, or no 'GG' cleavage site | | | 0.009 | | | |
| No *blpC* | | | 0.011 | | | |

[a] Conserved residues are in bold. Residues that differentiate signals within a signal group are underlined.

684 **Table 2.** Nucleotide diversity and $d_N$ / $d_S$ for selected genes

685

686

| Gene | Function | Nucleotide Diversity ($\pi$) | $d_N / d_S$ | Number of Sites | Number of Unique Sequences |
|---|---|---|---|---|---|
| *blpC* | QS Signal | 0.124 | 0.965 | 126 | 34 |
| *blpH* (receptor) | Histidine Kinase | 0.110 | 0.377 | 687 | 98 |
| *blpH* (entire) | Histidine Kinase | 0.090 | 0.266 | 1353 | 161 |
| *blpH* (kinase) | Histidine Kinase | 0.069 | 0.146 | 666 | 103 |
| *ddlA* | Housekeeping Gene | 0.044 | 0.064 | 1041 | 170 |
| *lepB* | Housekeeping Gene | 0.027 | 0.026 | 612 | 90 |
| *aroE* | Housekeeping Gene | 0.022 | 0.146 | 852 | 123 |
| *glkA* | Housekeeping Gene | 0.022 | 0.061 | 975 | 143 |
| *xpt* | Housekeeping Gene | 0.020 | 0.083 | 579 | 111 |
| *gdhA* | Housekeeping Gene | 0.015 | 0.085 | 1485 | 185 |
| *recP* | Housekeeping Gene | 0.011 | 0.091 | 1974 | 231 |

687

688 **Table S1.** Predicted *blpC* mature peptide signals and their frequencies in each data set

| | Unbiased Sets | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Signal | All Unbiased Sets | Maela | Mass. Asymptomatic | GenBank | Hermans | Georgia GenBank | PMEN | Complex 3 | PMEN-1 |
| Charlie | 0.291 | 0.302 | 0.235 | 0.282 | 0.289 | 0.314 | 0.308 | 0.024 | 0.004 |
| Golf | 0.223 | 0.185 | 0.420 | 0.190 | 0.197 | 0.256 | 0.154 | | 0.938 |
| Alpha | 0.125 | 0.100 | 0.175 | 0.184 | 0.275 | 0.215 | 0.308 | | |
| Foxtrot | 0.125 | 0.130 | 0.091 | 0.195 | 0.092 | 0.099 | 0.192 | | |
| Echo | 0.087 | 0.099 | 0.039 | 0.069 | 0.092 | 0.041 | 0.038 | | |
| Hotel | 0.060 | 0.081 | | | | | | | |
| Delta | 0.035 | 0.042 | 0.016 | 0.011 | 0.028 | | | | |
| Bravo | 0.028 | 0.038 | | | | | | | |
| Kilo | 0.007 | 0.002 | 0.018 | 0.052 | 0.007 | 0.017 | | 0.976 | |
| Other, or no 'GG' cleavage site | 0.012 | 0.002 | | 0.021 | 0.050 | | | 0.058 | |
| No *blpC* | 0.011 | 0.010 | 0.003 | 0.017 | | 0.008 | | | |
| Total Strains | 3017 | 616 | 174 | 142 | 121 | 26 | 82 | 240 | |

689
690

691
692 **Table S2.** Properties of genotypes in the stochastic model.

| Genotype | Classification | Bacteriocin Produced | Susceptible to Bacteriocin | Signal produced | Receptor Sensitive to Signals |
|---|---|---|---|---|---|
| 1 | Faithful-signaling | 1 | 2 | 1 | 1 |
| 2 | Eavesdropping | 1 | 2 | 2 | 1,2 |
| 3 | Faithful-signaling | 2 | 1 | 3 | 3 |
| 4 | Eavesdropping | 2 | 1 | 4 | 3,4 |

693
694

**Table S3.** Strains and plasmids used.

| Strain / Plasmid | Characteristics | References |
|---|---|---|
| *S. pneumoniae* strain | | |
| D39 | Serotype 2 strain, disrupted *blpA* and *blpB*. | [a] |
| PMEN-2 | Serotype 6E strain, intact *blpA* and *blpB* | [b,c] |
| PMEN-14 | Serotype 19F strain, disrupted *blpA*, intact *blpB* | [b,c] |
| PMEN-18 | Serotype 14 strain, disrupted *blpA*, intact *blpB* | [b,c] |
| Hermans-33 | Intact *blpA* and *blpB* | [c,d] |
| Hermans-1012 | Disrupted *blpA* and *blpB* | [c,d] |
| FB141 | D39, *Δcep::*($P_{blpT}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$ | [e] |
| FB149 | D39, *Δcep::*($P_{blpK}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$ | [e] |
| K-D39 | D39, *Δcep::*($P_{blpK}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{D39}$ | This study |
| K-PMEN-14 | D39, *Δcep::*($P_{blpK}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{PMEN-14}$ | This study |
| K-PMEN18 | D39, *Δcep::*($P_{blpK}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{PMEN-18}$ | This study |
| K-Hermans-33 | D39, *Δcep::*($P_{blpK}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{Hermans-33}$ | This study |
| K-Hermans-1012 | D39, *Δcep::*($P_{blpK}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{Hermans-1012}$ | This study |
| K-PMEN-2 | D39, *Δcep::*($P_{blpK}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{PMEN-2}$ | This study |
| T-D39 | D39, *Δcep::*($P_{blpT}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{D39}$ | This study |
| T-PMEN-2 | D39, *Δcep::*($P_{blpT}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{PMEN-2}$ | This study |
| T-Hermans-1012 | D39, *Δcep::*($P_{blpT}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{Hermans-1012}$ | This study |
| T-PMEN-18 | D39, *Δcep::*($P_{blpT}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{PMEN-18}$ | This study |
| T-PMEN-14 | D39, *Δcep::*($P_{blpT}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{PMEN-14}$ | This study |
| T-Hermans-33 | D39, *Δcep::*($P_{blpT}$-*LGZ*, $spc^R$), *ΔblpSRHC::*$ery^R$, *ΔbgaA::*$P_{blpS}$-$blpSRH^{Hermans-33}$ | This study |
| Plasmid | | |
| pPEP1-$P_{blpT}$-LGZ | $cam^R$, *cep'*, $spc^R$-$P_{blpT}$-*luc-gfp-lacZ,'cep* | [e] |
| pPEP1-$P_{blpK}$-LGZ | $cam^R$, *cep'*, $spc^R$-$P_{blpK}$-*luc-gfp-lacZ,'cep* | [e] |
| pJWV25 | $amp^R$, *bgaA'*, $tet^R$-*'gfp,'bgaA* | [f] |
| pJWV25-$blp^{D39}$ | $amp^R$, *bgaA'*, $tet^R$-$P_{blpS}$-$blpSRH^{D39}$,*'bgaA* | This study |
| pJWV25-$blp^{PMEN-2}$ | $amp^R$, *bgaA'*, $tet^R$-$P_{blpS}$-$blpSRH^{PMEN-2}$,*'bgaA* | This study |
| pJWV25-$blp^{Hermans-1012}$ | $amp^R$, *bgaA'*, $tet^R$-$P_{blpS}$-$blpSRH^{Hermans-1012}$,*'bgaA* | This study |
| pJWV25-$blp^{Hermans-33}$ | $amp^R$, *bgaA'*, $tet^R$-$P_{blpS}$-$blpSRH^{Hermans-33}$,*'bgaA* | This study |
| pJWV25-$blp^{PMEN-14}$ | $amp^R$, *bgaA'*, $tet^R$-$P_{blpS}$-$blpSRH^{PMEN-14}$,*'bgaA* | This study |
| pJWV25-$blp^{PMEN-18}$ | $amp^R$, *bgaA'*, $tet^R$-$P_{blpS}$-$blpSRH^{PMEN-18}$,*'bgaA* | This study |


[a] Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med*. 1944;79: 137–158.

[b] McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R, Hryniewicz W, Lefévre JC, Tomasz A, et al. 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. *J. Clin. Microbiol*. 39:2565–2571.

[c] Miller EL, Abrudan MI, Roberts IS, Rozen DE. Diverse ecological strategies are encoded by *Streptococcus pneumoniae* bacteriocin-like peptides. *Genome Biol Evol*. 2016;8(4):1072–90.

[d] Bogaert D, Engelen MN, Timmers-Reker AJM, Elzenaar KP, Peerbooms PGH, Coutinho RA, et al. Pneumococcal carriage in children in the Netherlands: A molecular epidemiological study. *J Clin Microbiol*. 2001;39(9):3316–20.

[e] Kjos M, Miller E, Slager J, Lake FB, Gericke O, Roberts IS, et al. Expression of *Streptococcus pneumoniae* bacteriocins is induced by antibiotics via regulatory interplay with the competence system. *PLOS Pathog*. 2016;12(2):e1005422.

[f] Eberhardt, Wu, Errington, Vollmer, Veening. Cellular localization of choline-utilization proteins in *Streptococcus pneumoniae* using novel fluorescent reporter systems. *Mol Microbiol*. 2009. 74(2):395-408

716 **Table S4.** Primer sequences used in this study.

717

| Primer name | Sequence (5’ → 3’) |
|---|---|
| blpS-F-ClaI-SphI | TCTGGTACCGCATGCGTCTTACTTCTGGCAACTGTG |
| blpH-R-NotI-SpeI | TATGCGGCCGCTCCACTAGTTATCATTCTGCATGTATCACAGT |
| blpS-up-F-PMEN2-SphI | ATGCGCATGCTGTTTTGATACTGTCAGTCTATC |
| blpS-down-R-PMEN2-SpeI-NotI | ACGTGCGGCCGCACTAGTAAGAGAACGCACTCTCGGTC |
| blpS-PMEN14-F-SphI | ATGCGCATGCGCTAAGGCAAGGATTCTGGATGG |
| blpH-Hermans33/35-R-NotI-SpeI | TATGCGGCCGCTCCACTAGTCATCATTCTGTATGTATCATAGT |

718

719

720 **Figure S1.** Phylogenetic relationship of 4,096 randomly sampled *S. pneumoniae* strains alongside 322 biased sampled
721 genomes, from (35). The data set for each strain is in the inner ring, while the BlpC signal is in the outer ring, which is
722 colored as in Fig. 2 and Fig. 3.
723
724 **Figure S2.** Rarefaction of non-singleton amino acid variants from randomly sampled genomes. Error bars show one
725 standard deviation across 10,000 bootstraps. Variants occurring in only one genome out of the 4,096 randomly sampled
726 genomes were excluded.
727
728 **Figure S3.** Diversity and recombination in the *blp* regulatory region. Nucleotide diversity, as measured by Shannon
729 entropy, and the number of unique recombination breakpoints are shown for A) the *blp* regulatory region and B) for the
730 *blpH-blpC* region, both using a sliding window of 30 bp. Red lines indicate the start and end of genes; blue regions
731 indicate predicted transmembrane domains for BlpH. BlpH and BlpC in the bottom panel are shaded grey with the
732 proportion of genomes containing each gene segment.
733
734 **Figure S4.** Relative maximal expression level from $P_{blpT}$ (as measured by RLU/OD) following addition of 1 µg/ml
735 BlpC. The relative maximal expression levels of *luc* in a $P_{blpT}$-reporter strain following addition of 1 µg/ml of
736 synthesized BlpC signal peptide. The maximum expression level for each reporter strain was set to 1.
737
738 **Figure S5.** LacZ induction on agar plates by exogenous BlpC. A) LacZ induction from exogenous BlpC of LacZ
739 reporter strains derived from *S. pneumoniae* D39 and containing the $P_{blpK}$-*lacZ* reporter along with a construct
740 constitutively expressing the *blpSRH* genes from different strains ($P_{blpS}$-*blpSRH*). B) Summary of the results presented
741 in A).
742
743 **Figure S6.** Mutual information between the BlpH receptor and the BlpC signal. For each of the five signal groups, we
744 calculated the mutual information of the signal (focal signal vs. non-focal signal) and each BlpH receptor residue (using
745 amino acid chemical class) that co-occurred in each of the 4,096 genomes. Shaded blue regions indicate predicted
746 transmembrane domains, and shaded red regions indicate predicted residues outside of the cell. The bottom panel shows
747 the total entropy of each residue, again using amino acid class.
748 We found several signal/residue combinations with mutual information at least 0.3 and at least twice as high as
749 other signals for the same residue within the BlpH receptor: Alpha/Bravo/Kilo and residues 10, 13, 42, 49, 51, 57, 58,

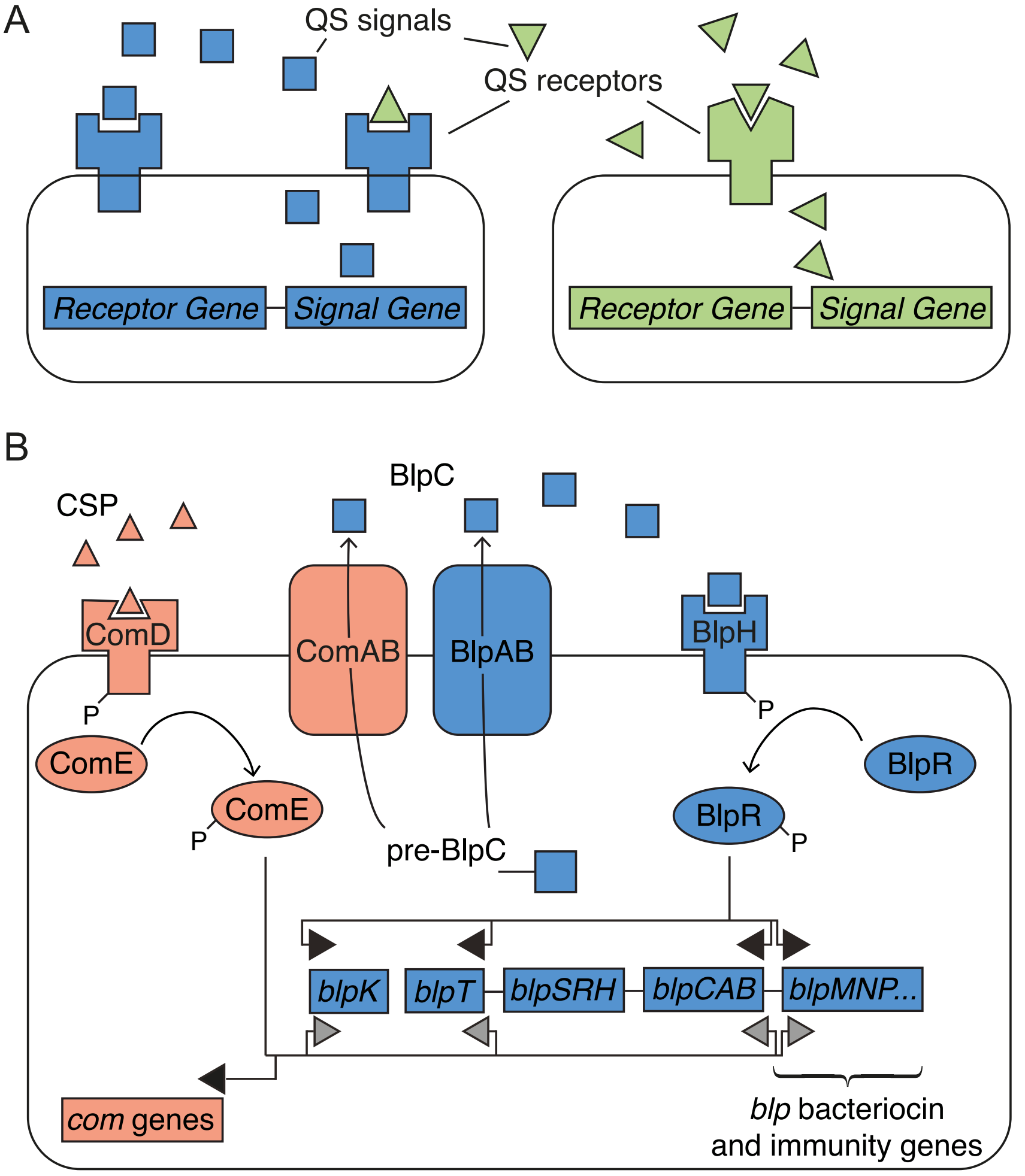
64, 115, 117, 118, 120-122, 126, 128, 131, 132, 135, 148, 153, 182, 223; Delta and residue 9; Echo/Foxtrot and residues 14, 17, 79, 152, 157, 178; Golf/Hotel and residues 187 and 191. However, no single amino acid site within the BlpH receptor can predict which signal co-occurs with 100% accuracy; the highest mutual information between receptor residues and signals are residues 152 and 178 with the Echo/Foxtrot signal (mutual information of 0.673 and 0.675, respectively).

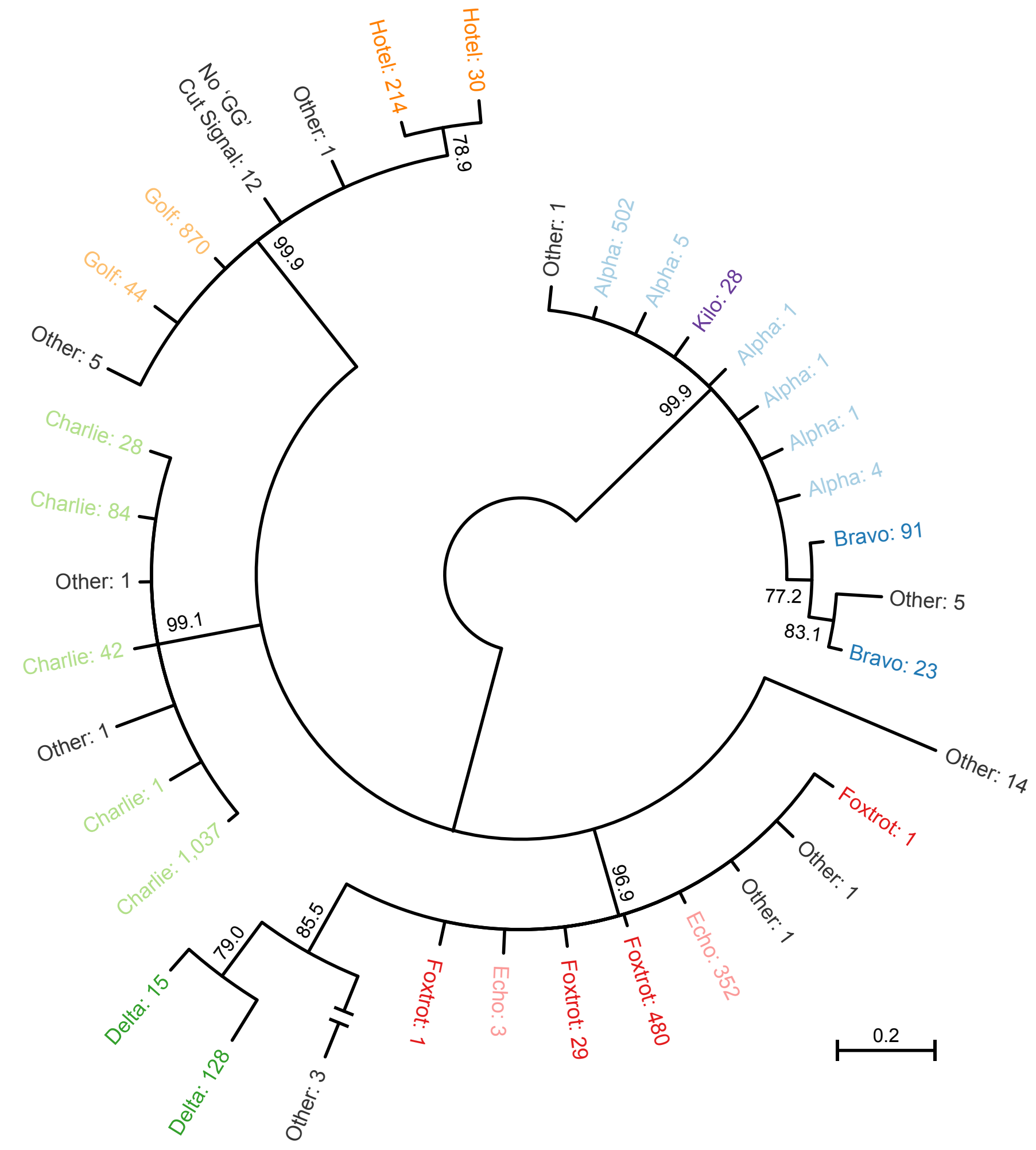
**Figure S7.** Mutual information between the BlpH kinase and the BlpC signal. For each of the five signal groups, we calculated the mutual information of the signal (focal signal vs. non-focal signal) and each BlpH kinase residue (using amino acid chemical class) across 4,096 genomes. The bottom panel shows the total entropy of each residue, again using amino acid class.

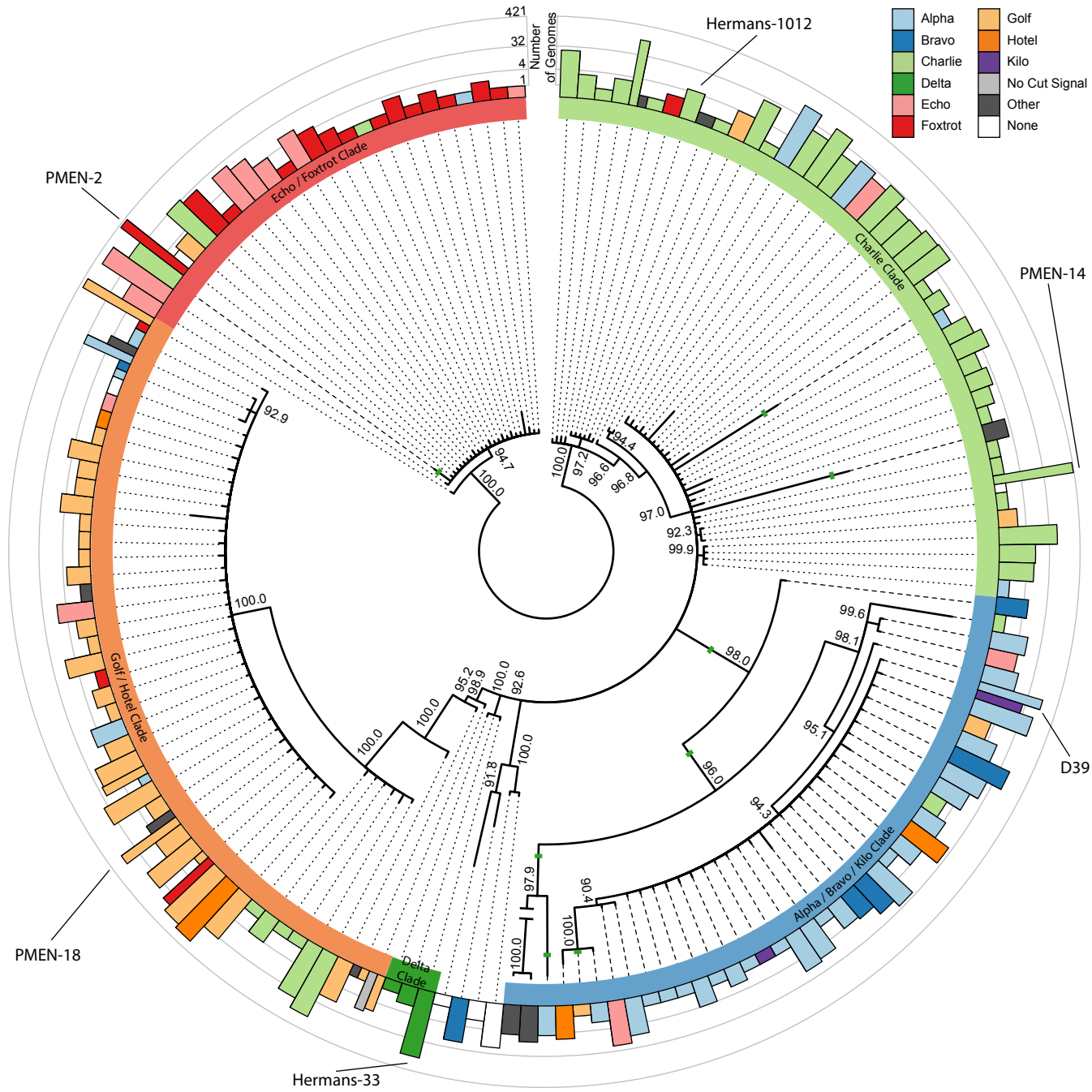
Higher mutual information was present between BlpH kinase residues and signals compared to the BlpH receptor and signal (Supplementary Fig. 6), which was unexpected, as the signal is not predicted to interact with the kinase. However, this region is more physically linked to the BlpC signal than the BlpH receptor, which would allow recombination to co-transfer both the signal and kinase in a single event.

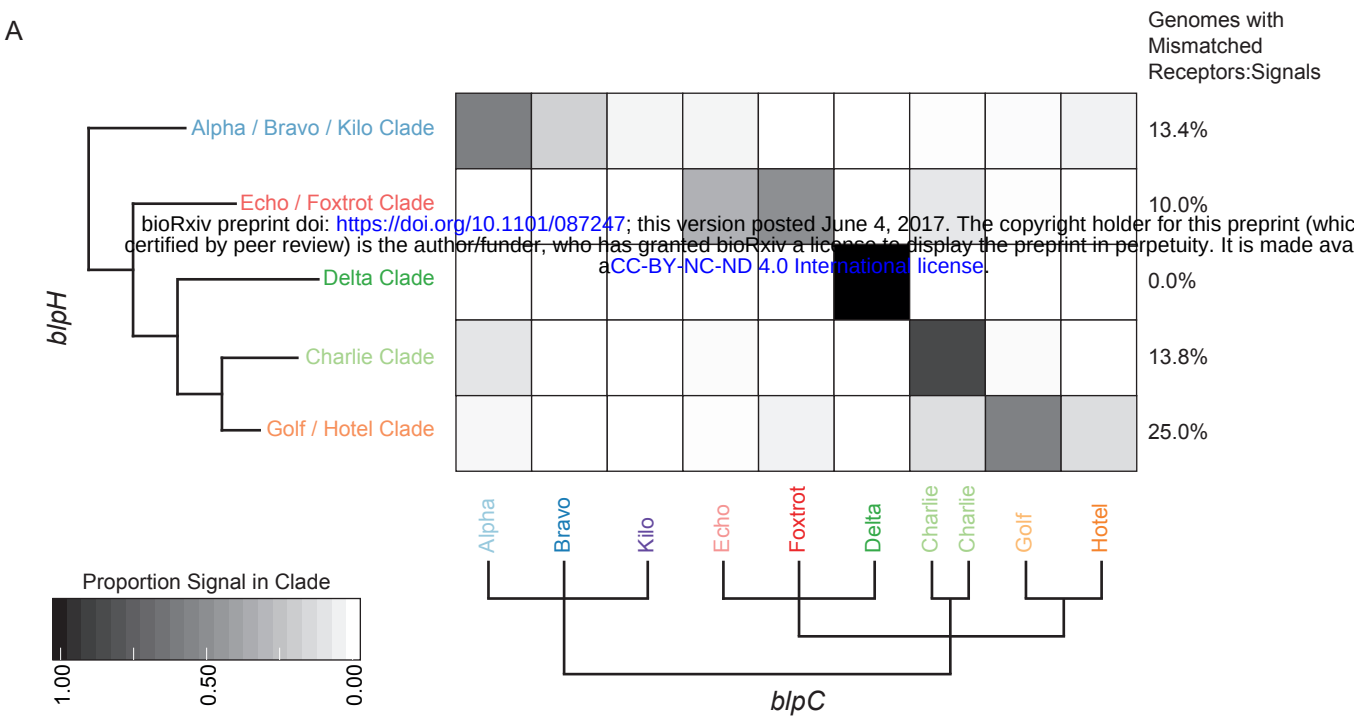
**Figure S8.** Raw data for Figure 1B. Reporter strains $P_{blpK}$-*luc* containing BlpSRH from different strains (as indicated above the plots) were grown in C+Y in a microtiter plate reader. Synthetic BlpC (final concentration 1µg/ml) was added after 100 minutes and expression of luciferase was followed, and is given as relative luminescence units per OD (RLU/OD).

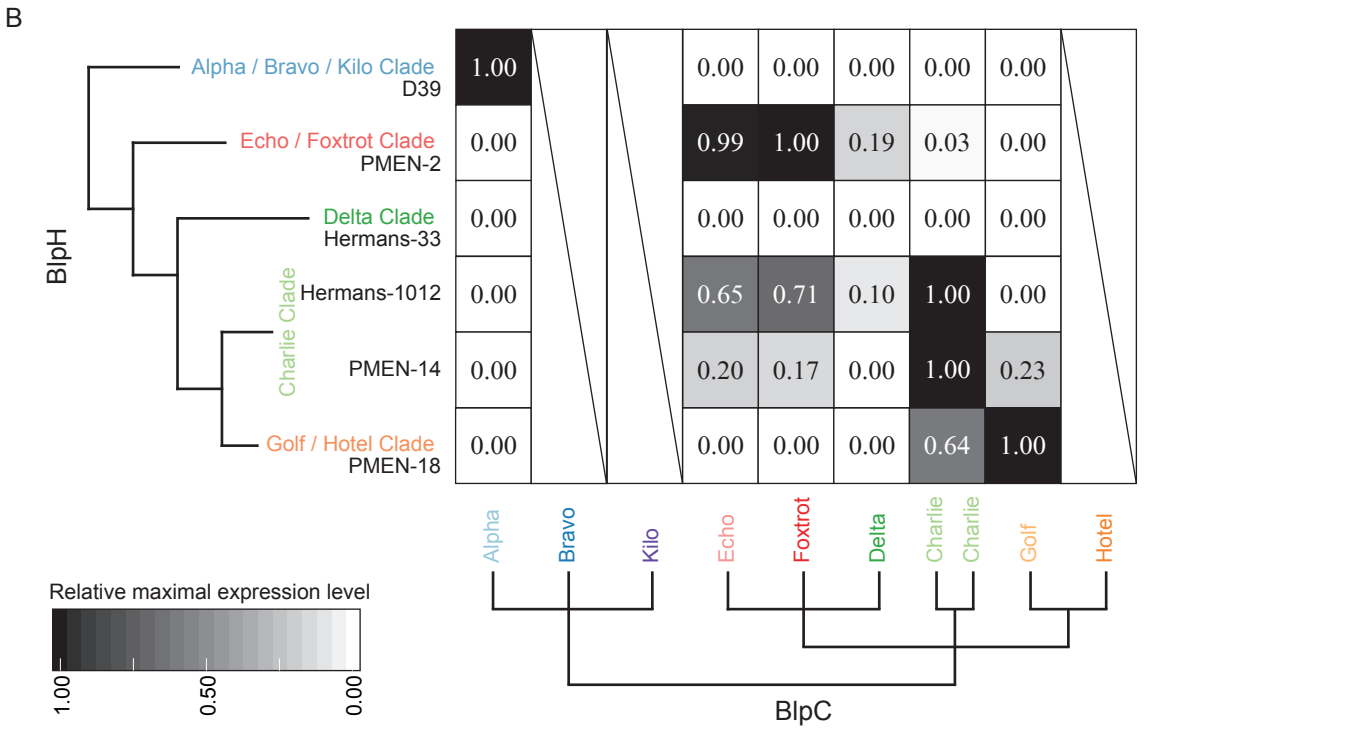
**Figure S9.** Example of raw data for Figure 1C. Reporter strain $P_{blpK}$-*luc* containing BlpSRH from D39 was grow in C+Y medium containing different concentrations of $BlpC^{Alpha}$. Reporter expression, given as RLU/OD, was followed. The minimal concentration of $BlpC^{Delta}$ to induce *luc* expression in this case was 3.9 ng/ml.

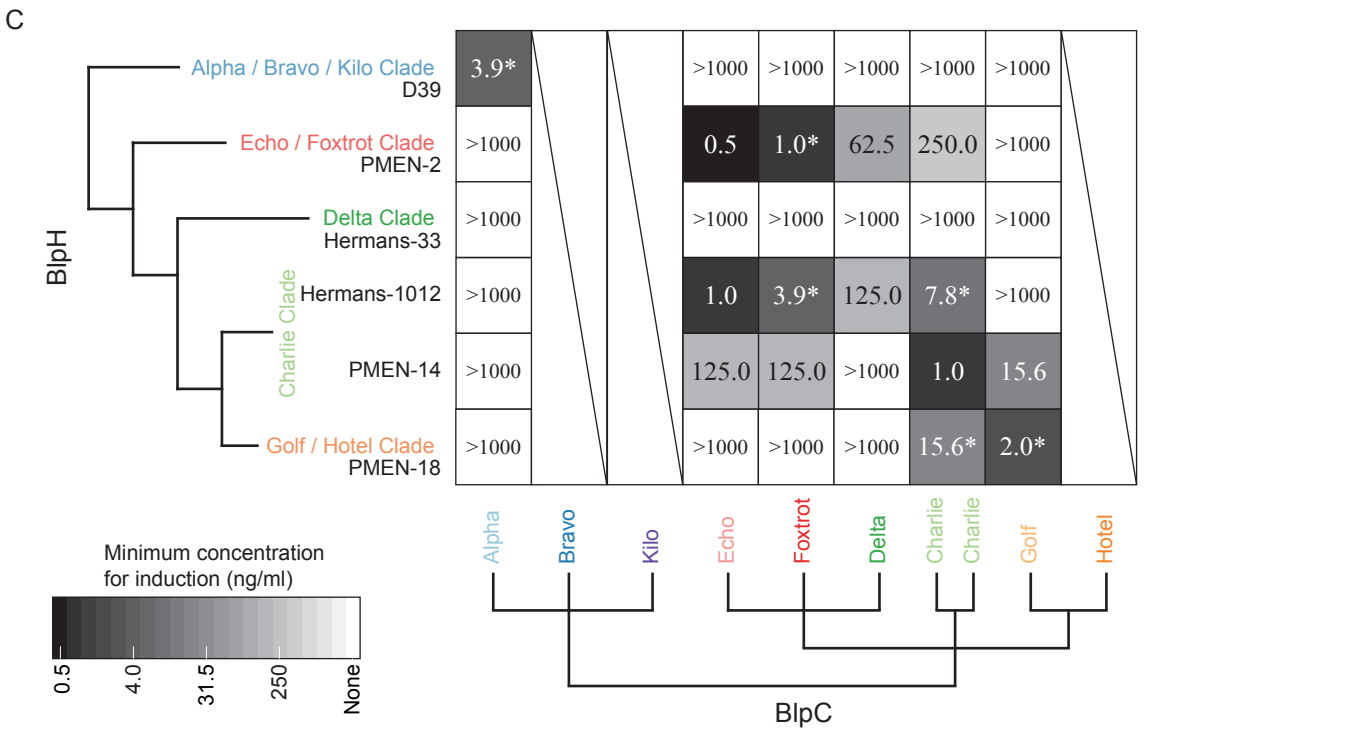

A
QS signals
QS receptors
Receptor Gene
Signal Gene
Receptor Gene
Signal Gene
B
BlpC
CSP
ComD
ComAB
BlpAB
BlpH
P
P
ComE
ComE
P
BlpR
BlpR
P
pre-BlpC
blpK
blpT
blpSRH
blpCAB
blpMNP...
com genes
blp bacteriocin
and immunity genes

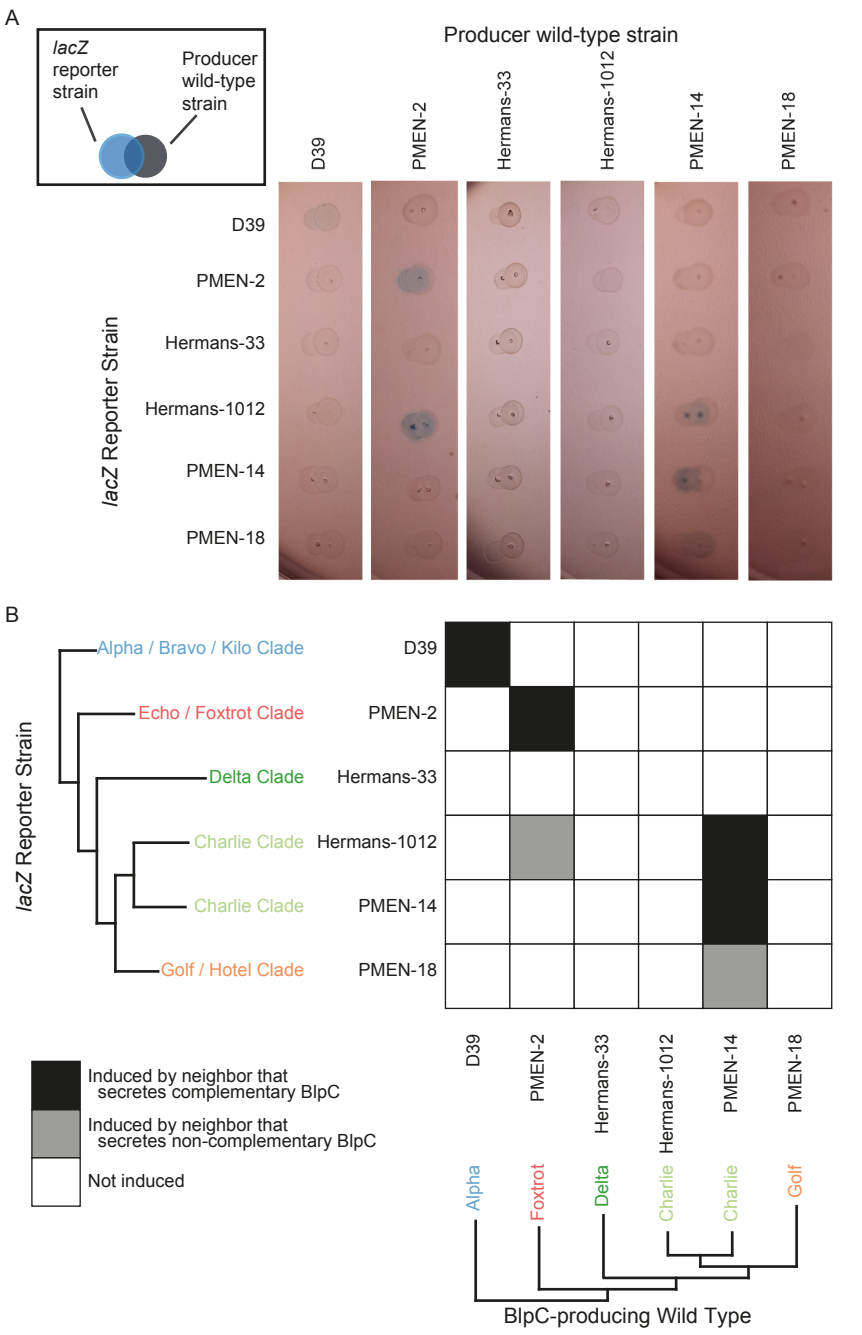

Hotel: 30
Hotel: 214
Other: 1
No 'GG' Cut Signal: 12
Golf: 870
Golf: 44
Other: 5
Charlie: 28
Charlie: 84
Other: 1
Charlie: 42
Other: 1
Charlie: 1
Charlie: 1,037
Delta: 15
Delta: 128
Other: 3
Foxtrot: 1
Echo: 3
Foxtrot: 29
Foxtrot: 480
Echo: 352
Other: 1
Other: 1
Foxtrot: 1
Other: 14
Bravo: 23
Bravo: 91
Other: 5
Alpha: 4
Alpha: 1
Alpha: 1
Alpha: 1
Kilo: 28
Alpha: 5
Alpha: 502
Other: 1
78.9
99.9
99.9
99.1
77.2
83.1
96.9
85.5
79.0
0.2

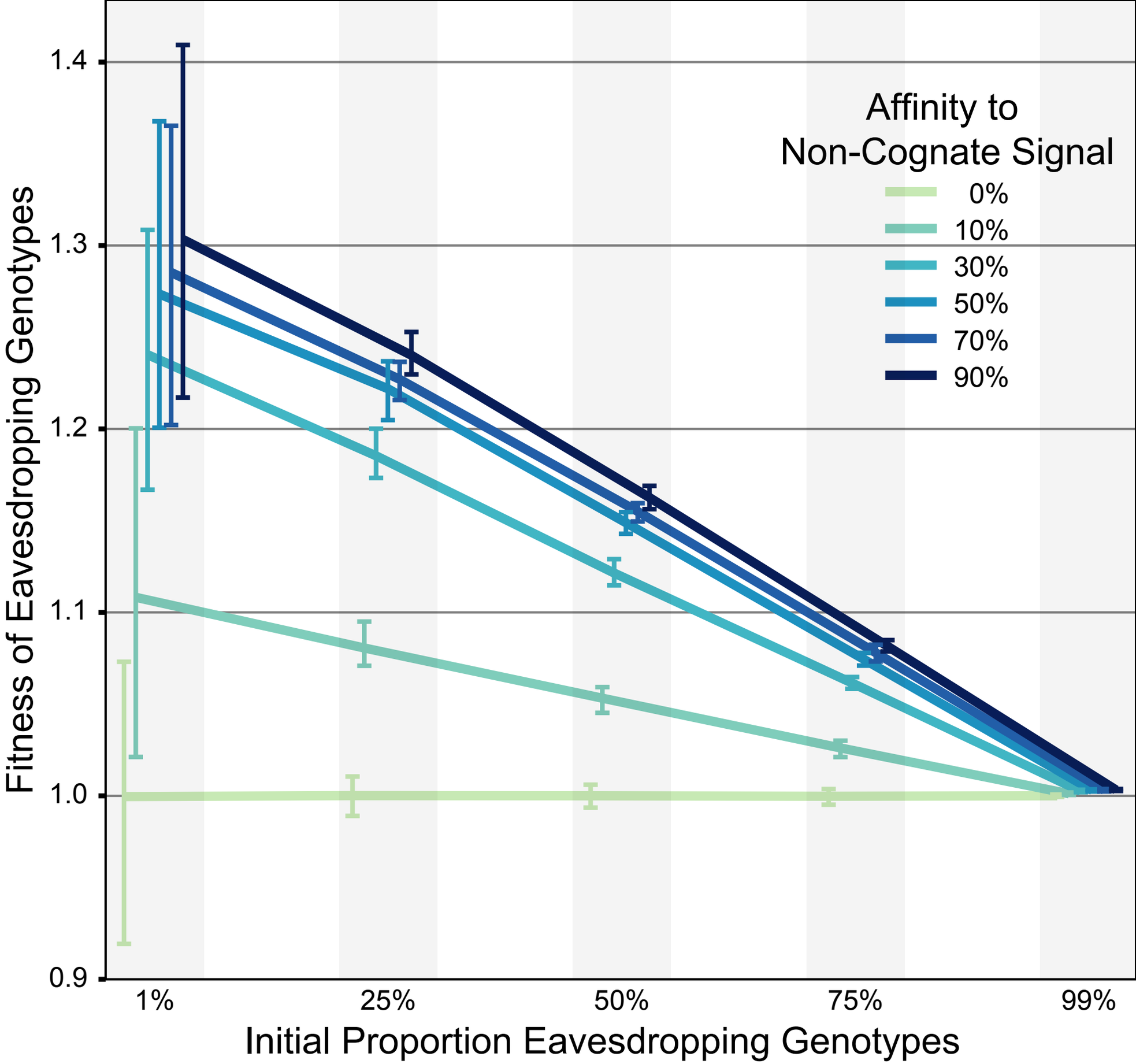

Alpha
Bravo
Charlie
Delta
Echo
Foxtrot
Golf
Hotel
Kilo
No Cut Signal
Other
None
421
32
4
1
Number of Genomes
Hermans-1012
PMEN-2
PMEN-14
D39
PMEN-18
Hermans-33
Echo / Foxtrot Clade
Charlie Clade
Golf / Hotel Clade
Delta Clade
Alpha / Bravo / Kilo Clade
92.9
94.7
100.0
100.0
97.2
96.6
96.8
94.4
97.0
92.3
99.9
100.0
100.0
100.0
95.2
98.9
100.0
92.6
100.0
91.8
100.0
98.0
99.6
98.1
95.1
96.0
94.3
97.9
100.0
90.4
100.0




A
lacZ reporter strain
Producer wild-type strain
Producer wild-type strain
D39
PMEN-2
Hermans-33
Hermans-1012
PMEN-14
PMEN-18
lacZ Reporter Strain
D39
PMEN-2
Hermans-33
Hermans-1012
PMEN-14
PMEN-18
B
lacZ Reporter Strain
Alpha / Bravo / Kilo Clade
Echo / Foxtrot Clade
Delta Clade
Charlie Clade
Charlie Clade
Golf / Hotel Clade
D39
PMEN-2
Hermans-33
Hermans-1012
PMEN-14
PMEN-18
Induced by neighbor that secretes complementary BlpC
Induced by neighbor that secretes non-complementary BlpC
Not induced
D39
PMEN-2
Hermans-33
Hermans-1012
PMEN-14
PMEN-18
Alpha
Foxtrot
Delta
Charlie
Charlie
Golf
BlpC-producing Wild Type


Affinity to
Non-Cognate Signal
0%
10%
30%
50%
70%
90%
Fitness of Eavesdropping Genotypes
1.4
1.3
1.2
1.1
1.0
0.9
1%
25%
50%
75%
99%
Initial Proportion Eavesdropping Genotypes