

# The leaderless communication peptide (LCP) class of quorum-sensing peptides is broadly distributed among Firmicutes

Received: 30 March 2023

Accepted: 14 September 2023

Published online: 23 September 2023

 Check for updatesShifu Aggarwal<sup>1,2,7</sup>, Elaine Huang<sup>1,2,7</sup>, Hackwon Do<sup>1,2,3,7</sup>, Nishanth Makthal<sup>1,2</sup>, Yanyan Li<sup>4</sup>, Eric Baptiste<sup>5</sup>, Philippe Lopez<sup>5</sup>, Charles Bernard<sup>6</sup>✉ & Muthiah Kumaraswami<sup>1,2</sup>✉

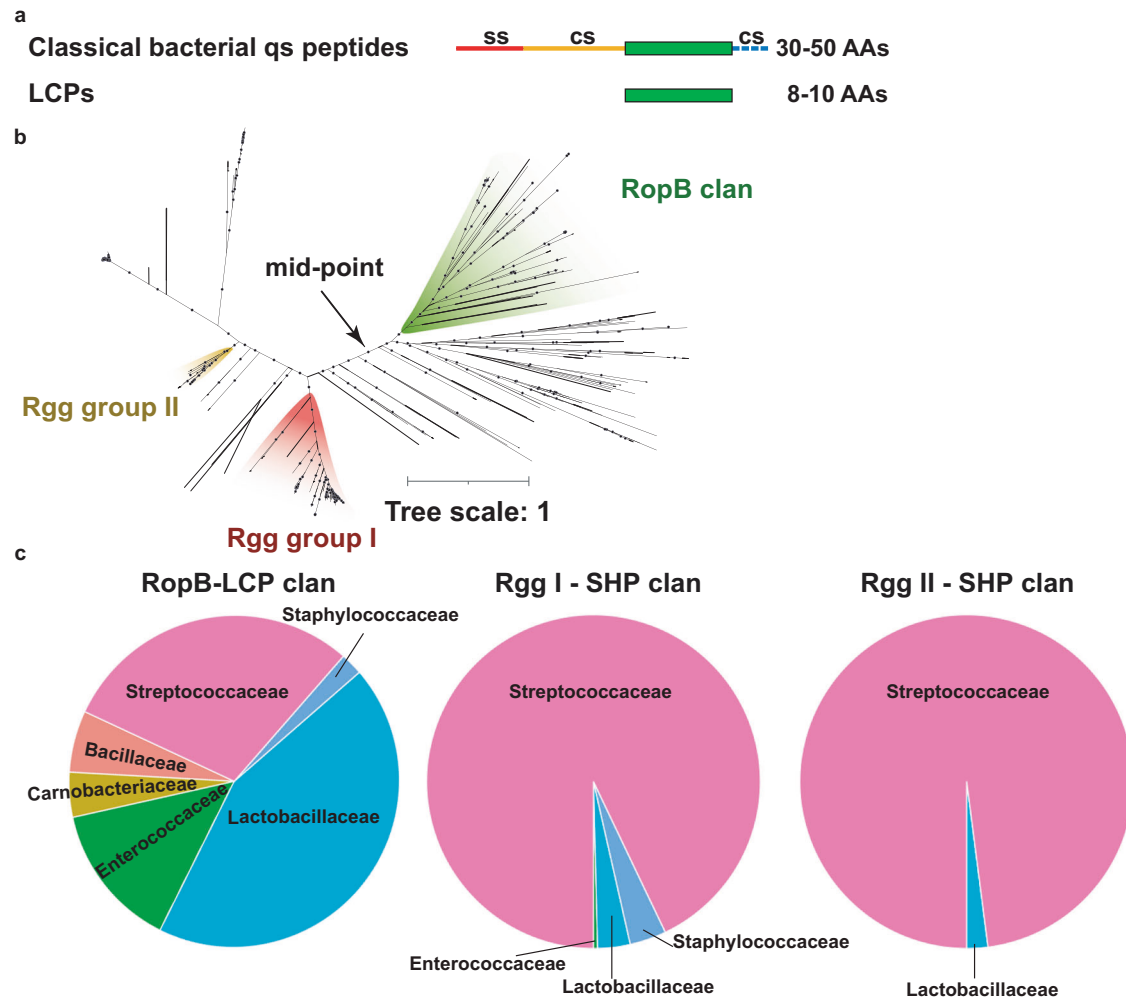
The human pathogen *Streptococcus pyogenes* secretes a short peptide (leaderless communication peptide, LCP) that mediates intercellular communication and controls bacterial virulence through interaction with its receptor, RopB. Here, we show that LCP and RopB homologues are present in other Firmicutes. We experimentally validate that LCPs with distinct peptide communication codes act as bacterial intercellular signals and regulate gene expression in *Streptococcus salivarius*, *Streptococcus porcinus*, *Enterococcus malodoratus* and *Limosilactobacillus reuteri*. Our results indicate that LCPs are more widespread than previously thought, and their characterization may uncover new signaling mechanisms and roles in coordinating diverse bacterial traits.

Bacteria use quorum sensing (qs) pathways to monitor the population density of their own and closely related competing species and coordinate population-wide phenotypic alterations that aid their survival<sup>1–6</sup>. The qs pathways control critical phenotypic traits such as bacterial virulence, biofilm formation, and antibiotic resistance<sup>1–6</sup>. The qs signaling involves signal production, signal export, signal recognition by extracellular or intracellular receptors, and the subsequent modulation of target gene expression<sup>1,3</sup>. The gram-positive bacteria typically use secreted cyclic or linear oligopeptides to carry out intercellular signaling<sup>6,7</sup>. The Rap-Rgg-NprR-PrgX-PlcR-AimR (RRNPPA)-family regulators in firmicutes and phages constitute the largest family of cytosolic qs receptors that bind to re-internalized linear oligopeptides to mediate gene regulation<sup>7</sup>. Invariably, the peptides are produced as long, inactive propeptides that contain the secretion signal sequence for their export and cleavage sites for membrane-bound or secreted proteases (Fig. 1a). The processed 5–8 amino acid peptides constitute the active intercellular signal that controls gene expression<sup>8,9</sup>.

Recently, we discovered a new class of leaderless communication peptide (LCP) in human pathogen *Streptococcus pyogenes* that controls bacterial virulence (Fig. 1a)<sup>10</sup>. Contrary to characterized bacterial peptide signals, the LCP, known as SpeB-inducing peptide (SIP), is synthesized from a 24 base pair ultrasmall open reading frame (usORF) encoding a mature 8 amino acid peptide without the amino acid sequences required for secretion and processing<sup>10</sup>. Consistent with the lack of hallmarks of bacterial peptide signals, the canonical machineries required for the export and import of qs peptides are not involved in SIP transport<sup>10</sup>. However, SIP is produced, secreted, and reimported into the bacterial cytosol. The cytosolic SIP binds to a RRNPPA-family transcription activator, Regulator of Protease B (RopB), which leads to the upregulation of an adjacently located virulence factor, a secreted cysteine protease SpeB, and increased bacterial virulence in animal models of infection<sup>10,11</sup>.

The lack of conformity of LCPs with the hallmarks of classical bacterial peptide signals and their presence as a 24-base pair usORF

<sup>1</sup>Center for Molecular and Translational Human Infectious Diseases Research, Houston Methodist Research Institute, Houston, TX 77030, USA. <sup>2</sup>Department of Pathology and Genomic Medicine, Houston Methodist Hospital, Houston, TX 77030, USA. <sup>3</sup>Research Unit of Cryogenic Novel Material, Korea Polar Research Institute, Incheon 21990, South Korea. <sup>4</sup>Communication Molecules and Adaptation of Microorganisms (MCAM), CNRS, Muséum National d'Histoire Naturelle, Paris, France. <sup>5</sup>Institut de Systématique, Evolution et Biodiversité (ISYEB), Sorbonne Université, CNRS, Muséum National d'Histoire Naturelle, Paris, France. <sup>6</sup>Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. <sup>7</sup>These authors contributed equally: Shifu Aggarwal, Elaine Huang, Hackwon Do. ✉ e-mail: [charles.bernard@cri-paris.org](mailto:charles.bernard@cri-paris.org); [mkumaraswami@houstonmethodist.org](mailto:mkumaraswami@houstonmethodist.org)



**Fig. 1 | Leaderless communication peptide (LCP) system is broadly distributed.** **a** Schematics showing the amino acid sequence characteristics of classical bacterial qs propeptides and LCPs. Ss—secretion signal sequence; cs—protease cleavage sites; cs with dashed lines—indicate the presence of additional cleavage sites in a subset of propeptides. **b** The unrooted phylogenetic tree comprises 953 RopB homologs and was inferred from a trimmed alignment of 204 sites. Blue dots indicate branches supported by >90% bootstraps. Branch lengths are proportional

to the expected number of substitutions per site, as indicated by the scale bar at the top left. Each highlighted clan corresponds to receptors for which the most likely translated adjacent uORF encodes a remarkable class of qs peptides. The two Rgg clans (red and yellow) are enriched in receptor-SHP propeptide pairs, while the RopB clan (green) is enriched in receptor-LCP pairs. **c** The distribution of RopB-LCP, Rgg I-SHPs, and Rgg II-SHPs systems in different bacterial taxa.

precluded their detection by gene annotation pipelines. Consequently, SIP remains as the only characterized LCP-based qs system, and the prevalence and distribution of LCPs in bacteria as well as their roles in bacterial pathophysiology and ecology remain unknown. To elucidate the full breadth of LCPs in bacteria, we designed a custom approach and scanned bacterial genomes for the presence of SIP-like LCPs. Here we report that LCPs are widespread among firmicutes encompassing a larger taxonomic diversity compared to other members of RRNPPA family regulators<sup>12</sup> and display a diverse set of communication codes with varying amino acid composition and lengths. We further show that a subset of the newly identified LCPs from distant bacterial species function as effective intercellular signals and control the expression of genes involved in diverse functions. Our findings point to the presence of a broadly distributed new class of qs signals and their potential roles in different aspects of bacterial pathophysiology.

## Results

### Leaderless communication peptide (LCP) system is broadly distributed

To assess the distribution of SIP-like LCP-based qs systems across bacteria, we employed a large-scale search strategy that includes: (i)

search for RopB homologs across a dataset of 129,001 bacterial genomes and 9,421 reference metagenomics-assembled genomes (MAGs), (ii) probing the genomic vicinity of RopB homologs for small ORFs with a preceding Shine-Dalgarno Ribosome-Binding-Site (RBS) motif that is indicative of their likely translation, (iii) identification of clans of RopB homologs for which the most likely translated adjacent sORF is ultrasmall and encodes a SIP-like LCP, and (iv) functional validation by assessing the regulatory activity of chosen subset of candidate LCPs.

A Blastp search of RopB against the target dataset resulted in 19,280 hits (sequence identity  $\geq 25\%$ , mutual length coverage  $\geq 70\%$ , Supplementary Data 1) encoded in 15,776 genomes and 39 MAGs distributed across 468 taxa. These 19280 hits correspond to 975 unique protein sequences that are predicted to harbor C-terminal tetratricopeptide repeat domain, a hallmark structural element responsible for peptide recognition by the RRNPPA family of receptors<sup>7</sup> (Supplementary Data 1). Importantly, 478 out of 975 RopB homologs (~49%) are flanked by at least one candidate sORF with a high confidence upstream RBS motif (Supplementary Data 1)<sup>13,14</sup>.

To determine whether clans of RopB homologs are enriched for their association with LCPs, we inferred the phylogeny of RopB-like receptors and mapped the amino acid sequence of flanking sORF with

the strongest RBS motif on its corresponding receptor leaf of the tree (Fig. 1b, Supplementary Data 1). The two previously well-characterized clans of RopB homologs correspond to streptococcal Rgg regulators, which recognize adjacently encoded canonical propeptides that are post-translationally processed into mature short hydrophobic peptides (SHPs)<sup>15,16</sup>. Accordingly, the two Rgg clans appeared as two well-delineated clans in the RopB tree and were correctly mapped to SHP propeptides, which lends validation to our methodology to identify clans of qs systems (Fig. 1b and S1, Supplementary Data 1 and 2). Remarkably, RopB and its evolutionarily closest relatives represent a distinct clan of 183 receptors that were frequently associated with an ultrasmall peptide with amino acid composition similar to SIP (Fig. 1b, Supplementary Data 1 and 2). We term the new clan, RopB clan, and the receptors in RopB clan represent a new class of qs system that uses LCPs as signaling molecules.

The refined manual assessment of the candidate cognate LCP of each receptor within the RopB clan (Fig. 2 and Supplementary Data 2) revealed that the LCP system is broadly distributed in bacterial genomes of a large taxonomic diversity, spanning over *Streptococcaceae*, *Lactobacillaceae*, *Enterococcaceae*, *Carnobacteriaceae*, *Bacillaceae* and *Staphylococcaceae* families (Fig. 1c). Most of the candidate receptor-LCP pairs are encoded in the bacterial genomes, and a non-negligible number of LCP systems are present in either plasmids or integrative and conjugative elements (ICEs) (Fig. 3). However, receptor-LCP pairs were not detected in prophages or phages. The genetic elements encoding LCP-receptor pairs are found in various host-associated microbiomes, waste waters, and fermented food products. The LCP systems are prevalent in the genomes of several clinically relevant human pathogens such as *S. pyogenes* or *Enterococcus casseliflavus*, and in animal pathogens such as *S. porcinus* and *S. pseudoporcinus* (Fig. 3).

Most putative LCPs are 8 to 10 amino acids long with few exceptions such as LCPs from *Lysinibacillus sphaericus*, *Staphylococcus delphini*, *Enterococcus durans*, and *Granulicatella balaenopterae*. The predicted LCPs are highly hydrophobic and are mostly comprised of aliphatic and aromatic amino acids (Fig. 2 and S1). The peptides encode distinct communication codes with unique LCP amino acid sequences, suggesting that LCP-mediated communication is specific among closely related bacterial species or strains. Importantly, the characterized RopB-SIP system of *S. pyogenes* represents only a small subset of the identified LCP systems (Fig. 2), indicating the broader distribution of LCP systems and their potential unappreciated roles in bacterial pathophysiology.

The structure-guided multiple amino acid sequence alignment (MSA) analyses of RopB clan receptors indicate that the LCP-binding C-terminal domain (CTD) diversified faster than the N-terminal DNA-binding domain (DBD) (Figs. S2 and S3a)<sup>7,11,17</sup>. Among the structural elements of CTD, the  $\alpha 6$  helix is highly conserved relative to the rest of CTD (Figs. S2 and S3a). The  $\alpha 6$  helix constitutes a critical structural element for LCP binding as it forms the floor of the LCP binding pocket as well as engages in intramolecular interactions that provide the scaffold for the LCP-binding pocket of RopB (Supplementary Fig. 3b)<sup>7,11,17</sup>. In accordance with the suggested functional constraint on  $\alpha 6$  helix to bind LCPs, the site-wise dN/dS ratio analyses indicate that  $\alpha 6$  helix of candidate LCP receptors are under strong purifying selection against amino acid substitutions (Supplementary Fig. 3a). Similarly, the MSA analyses of 12 LCP-contacting amino acids of RopB with corresponding amino acids from candidate RopB-clan receptors suggest that these amino acids evolve slower and face stronger purifying selection compared to the remainder of CTD (Supplementary Fig. 4a, b). These observations are suggestive of faster diversification of TPR motifs likely due to their innate degeneracy<sup>18,19</sup> relative to LCP-receptor specificity diversification. In accordance with this, pairwise comparisons of evolutionary distances between LCP receptors, LCP-contacting residues in receptors, and candidate LCPs suggest that

the amino acid sequences of LCP receptors diverged faster than the LCP-contacting residues and amino acid sequences of putative LCPs, and both LCPs and LCP-contacting amino acids in RopB clan receptors diverge at similar evolutionary rate (Supplementary Fig. 5). Collectively, these observations are suggestive of the function of RopB clan receptors and candidate LCPs as qs receptor-signal pairs.

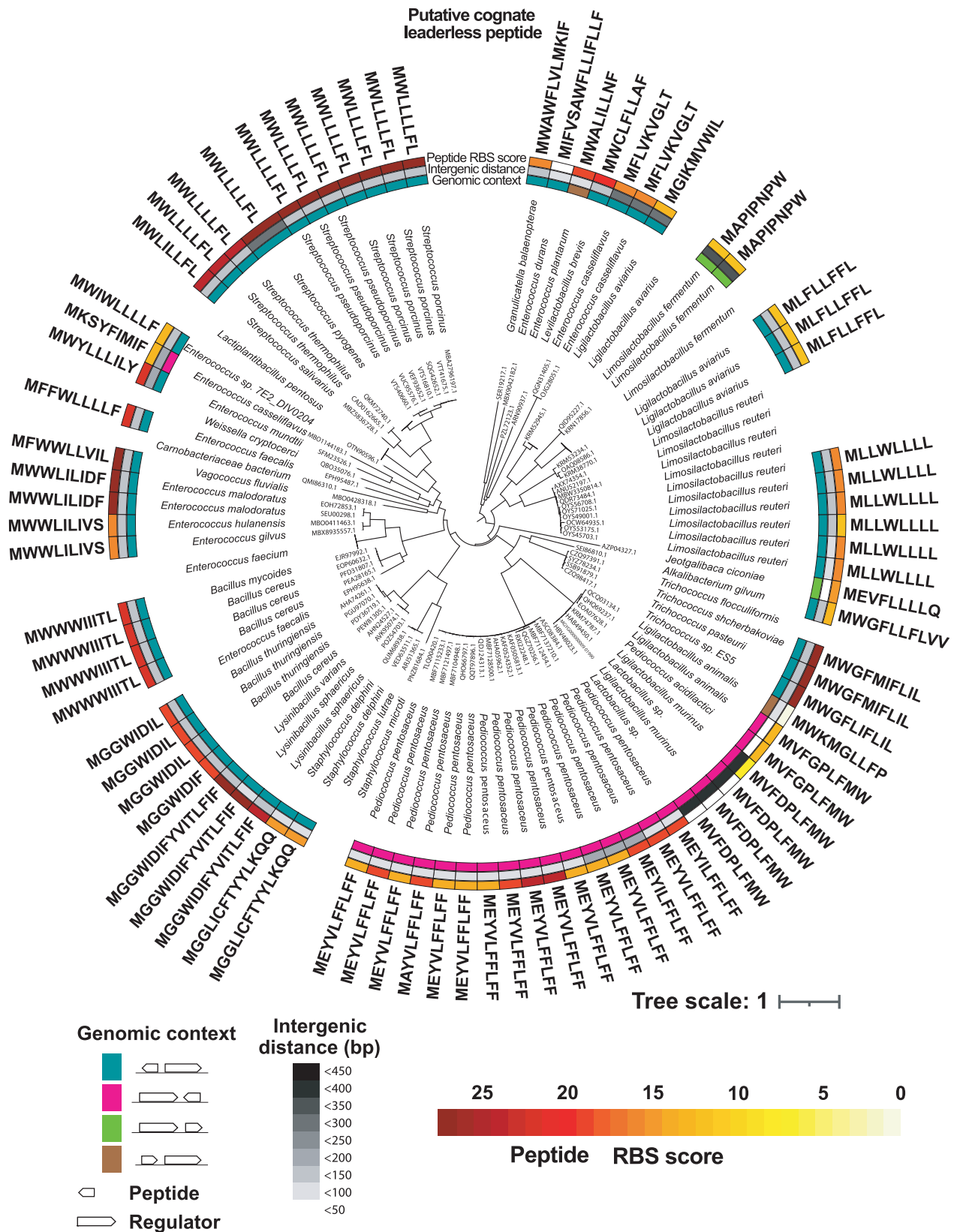
To understand the co-evolution of receptors and candidate LCPs, we aligned the 12 LCP-contacting amino acids of RopB with similarly located amino acids from RopB-clan receptors and compared them with the physicochemical characteristics of the corresponding LCPs (Supplementary Fig. 4a, b). Consistent with the preponderance of aliphatic and aromatic amino acids in majority of candidate LCPs (Supplementary Fig. 1d), the peptide-contacting amino acids are relatively well conserved among RopB clan receptors (Supplementary Fig. 4). However, compared to RopB, the peptide-contacting residues of RopB clan receptors from *Ligilactobacillus muralis*, *Pediococcus acidilactici*, *Ligilactobacillus animalis*, and *Enterococcus casseliflavus* are distinct (Supplementary Fig. 4). Accordingly, the physicochemical properties of corresponding LCPs also deviate from the typical signature of LCPs as they contain charged, polar, and proline residues (Supplementary Fig. 4). These observations suggest a tropism of RopB clan receptors for SIP-like LCPs, however, divergence exists to achieve alternative LCP specificities through receptor-LCP co-evolution. Finally, analyses of the RopB clan receptor-LCP pairs from *Bacillus cereus* and *Lysinibacillus sphaericus* revealed an evolutionary feature that is suggestive of peptidase-mediated processing of some LCPs similar to canonical RRNPP propeptides (Supplementary Fig. 4). The predicted LCP binding sites and the corresponding candidate LCPs of *B. cereus* and *L. sphaericus* have identical amino acid composition. However, the LCP from *L. sphaericus* have an additional eight amino acids in their C-terminus compared to *B. cereus* LCP (Supplementary Fig. 4). This observation suggests that LCPs may exist in longer precursor forms and the C-terminal appendages may be involved in peptidase-mediated cleavage of precursor form to release mature LCPs.

Analyses of the genomic context of representative putative LCP systems showed that the candidate genes regulated by LCPs are predominantly in a divergent context relative to the receptor (Fig. 2) and belong to 3 major categories: biosynthetic gene clusters (BGCs) predicted to produce ribosomally synthesized and post-translationally modified peptide (RiPPs) as well as non-ribosomally synthesized antimicrobials, ABC-type transporters, and type VII secretion systems that are typically involved in the translocation of virulence factors (Fig. 3). These observations suggest a broader and more diverse role for LCP systems in bacterial pathogenesis, physiology, and microbial ecology.

### LCP in *S. salivarius* mediates gene regulation

To investigate whether the LCPs other than SIP also act as intercellular signals, we characterized the putative cytosolic receptor-LCP pair from *S. salivarius* (RopB<sub>ss</sub>-LCP<sub>ss</sub>). The LCP<sub>ss</sub> is encoded in a megaplasmid and located downstream of *ropB<sub>ss</sub>* and transcribed divergently (Fig. 4a). The LCP<sub>ss</sub> encodes an eight amino acid hydrophobic peptide with a predicted amino acid sequence of MWLILLFL with no additional amino acids at either end (Fig. 4b). The genetic proximity of a 14-gene operon encoding a putative non-ribosomal peptide synthase biosynthesis gene cluster (*NRPS-BGC*) located immediately downstream of LCP<sub>ss</sub> and transcribed convergently (Fig. 4a) suggest that *NRPS-BGC* is the regulatory target of the RopB<sub>ss</sub>-LCP<sub>ss</sub> pathway.

In accordance with this, inactivation of *ropB<sub>ss</sub>* or LCP<sub>ss</sub> abrogated *NRPS-BGC* expression and *cis*-complementation of  $\Delta ropB_{ss}$  and LCP<sub>ss</sub>\* mutants with *ropB<sub>ss</sub>* and LCP<sub>ss</sub>, respectively, restored *NRPS-BGC* expression (Fig. 4c). Similarly, the addition of synthetic LCP<sub>ss</sub> containing the predicted amino acid sequence in native order (LCP<sub>ss</sub>), not in scrambled order (SCRA), restored WT-like *NRPS-BGC* expression in the LCP<sub>ss</sub>\* mutant (Fig. 4d). However, supplementation with synthetic



**Fig. 2 | Predicted LCPs encode diverse peptide communication codes.** The phylogenetic tree of the RopB clan, rooted at mid-point for visualization purpose, comprises 182 leaves and was inferred from a trimmed alignment of 229 sites. The label of a leaf corresponds to the NCBI protein id of the putative receptor, followed by the name of the encoding species. The most outer label in bold indicates the sequence of a putative LCP encoded in the genomic vicinity of a *ropB* homolog

(leaf) whenever detected. The three-color strips correspond to (i) the genomic orientation of the candidate receptor-LCP pair, (ii) the intergenic distance between the receptor and LCP ORFs and+ (iii) the confidence score, ranging from 1 to the 27 of the Shine-Dalgarno RBS motif identified upstream from the LCP's ORF (white meaning no RBS detected). Tips without labels correspond to collapsed receptors encoded by the same species and without LCP detected in the vicinity.





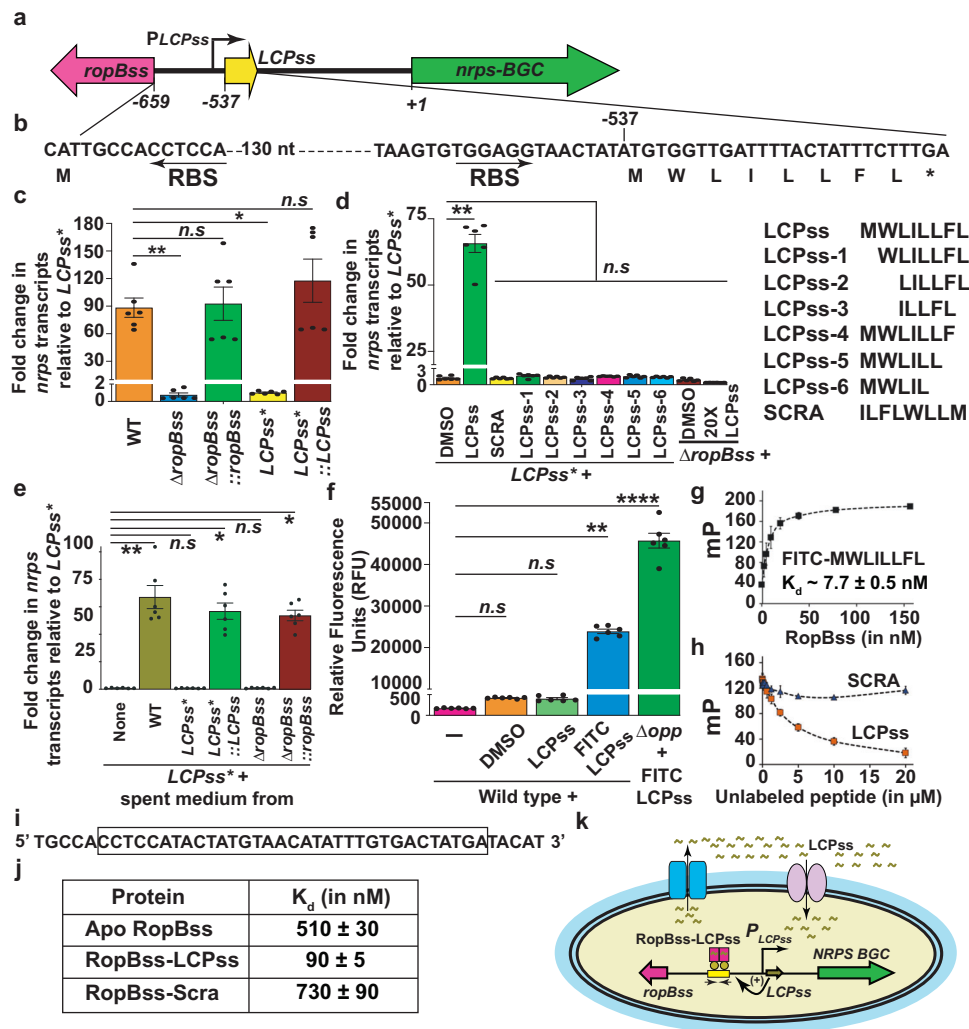
**Fig. 3 | Taxonomic, sequence, ecological and functional diversity of candidate LCP systems.** Each row depicts a representative of a group of receptor-LCP pairs with the same LCP sequence and the same species within the RopB clan. The stacked histogram highlights the proportion of receptor-LCP encoding elements in

chromosomes, plasmids, ICEs or unclassified encoding contigs within a group. The matrix shows the biomes in which the LCP systems of a group were detected. The last column indicates the adjacently located putative target regulon controlled by each representative LCP system.

LCP<sub>ss</sub> containing staggering truncations at either N-terminal or C-terminal ends (Fig. 4d) failed to activate *NRPS-BGC* expression in the LCP<sub>ss</sub>\* mutant, demonstrating that LCP<sub>ss</sub> encodes a mature LCP and lacks the hallmarks of canonical bacterial peptide signals (Fig. 4d). Furthermore, supplementation with even 20X molar excess of synthetic LCP<sub>ss</sub> failed to activate *NRPS-BGC* expression in  $\Delta$ ropB<sub>ss</sub> (Fig. 4d), indicating that LCP<sub>ss</sub> activity requires its cognate receptor RopB<sub>ss</sub>. However, despite the absence of the secretion signal sequence, the LCP<sub>ss</sub> is secreted and reinternalized into the cytosol and acts as an intercellular signal. This was demonstrated by the presence of LCP<sub>ss</sub> associated regulatory activity only in the secreted component of peptide-producing strains (WT, LCP<sub>ss</sub>\*::LCP<sub>ss</sub>, and  $\Delta$ ropB<sub>ss</sub>::ropB<sub>ss</sub>) and internalization of the exogenously added FITC-labeled synthetic LCP<sub>ss</sub> (Fig. 4e, f and Supplementary Fig. 6). Furthermore, inactivation of the canonical bacterial peptide import machinery oligopeptide permease ( $\Delta$ opp) did not affect LCP<sub>ss</sub> import (Fig. 4f)<sup>20</sup>, suggesting that unknown reimport mechanisms are involved in LCP<sub>ss</sub> import.

To elucidate the molecular mechanism of LCP<sub>ss</sub>-mediated signaling, we investigated the sequence-specific recognition of LCP<sub>ss</sub>

by cytosolic RopB<sub>ss</sub> by fluorescence polarization (FP) assay using FITC-labeled LCP<sub>ss</sub>. RopB<sub>ss</sub> binds LCP<sub>ss</sub> with high affinity ( $K_d$  ~8 nM) (Fig. 4g) and the pre-formed RopB<sub>ss</sub>-FITC-LCP<sub>ss</sub> complex was disrupted only by unlabeled LCP<sub>ss</sub>, not by non-specific SCRA (Fig. 4h), indicating that RopB<sub>ss</sub> recognition of LCP<sub>ss</sub> is sequence-specific. To explain the downstream consequences of RopB<sub>ss</sub>-LCP<sub>ss</sub> interactions, we hypothesized that LCP<sub>ss</sub> facilitates RopB<sub>ss</sub> interactions with target promoters and promotes RopB<sub>ss</sub>-dependent activation of *NRPS-BGC* expression. To map the operator sequences for RopB<sub>ss</sub> in LCP<sub>ss</sub> and *NRPS-BGC* promoters, we performed electrophoretic mobility shift assays (EMSA) using different DNA fragments that span LCP<sub>ss</sub> and LCP<sub>ss</sub>-*NRPS-BGC* intergenic region (Supplementary Fig. 7). RopB<sub>ss</sub> bound only to a 43-bp fragment located immediately upstream of the putative -35 hexamer of the LCP<sub>ss</sub> promoter and did not interact with the *NRPS-BGC* promoter (Fig. 4i and Supplementary Fig. 7). These results indicate that LCP<sub>ss</sub> and *NRPS-BGC* are likely expressed as a polycistronic transcript and RopB<sub>ss</sub> binding site is located in the LCP<sub>ss</sub> promoter (Supplementary Fig. 7). We further probed the 43-bp fragment for the presence of putative palindromes



**Fig. 4 | Intercellular signaling and gene regulation by *S. salivarius* LCP.**

**a** Schematic representation of genetic elements in *S. salivarius* (ss) encoding *ropBss*, *LCPss*, and a biosynthetic gene cluster (BGC) encoding non-ribosomal peptide synthase (*nrps*). The divergently transcribed *ropBss* and *LCPss* along with predicted transcription start site (*PLCPss*, bent arrow) of *LCPss* are shown. The numbers below denote the nucleotide positions relative to the first nucleotide of the start codon of *nrps-BGC*. **b** The nucleotide sequence of the *ropBss-LCPss* intergenic region, coding sequence of *LCPss*, and corresponding predicted amino acid sequence of *LCPss* are shown. The ribosomal-binding sites (RBS) of *LCPss* and *RopBss* are denoted by arrows. **c** Analysis of *nrps* transcript levels in the indicated strains by qRT-PCR. **d** Addition of full-length synthetic *LCPss* activates *nrps* expression in *LCPss* inactivated mutant (*LCPss*<sup>\*</sup>). The amino acid sequences of the synthetic *LCPss* variants used are shown (right). **e** qRT-PCR based *nrps* transcript level analyses in *LCPss*<sup>\*</sup> mutant grown in spent medium from the indicated strains. **f** Cytosolic fluorescence corresponding to FITC-labeled *LCPss* indicative of the import of exogenous *LCPss* as assessed by fluorescence measurements. Unmodified or FITC-labeled *LCPss* was added at a final concentration of 1  $\mu$ M to either WT or oligopeptide permease-inactivated mutant ( $\Delta$ *opp*). After 30 min incubation, fluorescence in the clarified

cell lysates was measured using excitation and emission wavelengths of 480 nm and 520 nm, respectively. **g** Analysis of the binding between purified *RopBss* and fluoresceinated *LCPss* by fluorescence polarization (FP) assay. **h** Ability of *LCPss* or SCRA peptide to compete with the FITC-labeled *LCPss-RopBss* complex for binding. A preformed *RopB* (250 nM)-labeled *LCPss* (10 nM) complex was challenged with the indicated unlabeled peptides. **i** Nucleotide sequence of the identified *RopBss* binding site in *LCPss* promoter used in the DNA-binding studies is shown. **j** Summary of the affinity of different forms of *RopBss* to *LCPss* promoter as assessed by FP assays. **k** Proposed model for *LCPss* signaling. *LCPss* is produced, exported, and internalized into the cytosol. The recognition of *LCPss* by *RopBss* promotes high affinity interactions between *RopBss* and binding site in *LCPss* promoter, which leads to the upregulation of *LCPss* and *NRPS-BGC*. In (c, d, e, f), data are derived from three biological replicates and analyzed in duplicates. In (g, h), data are derived from three independent experiments. In (c–h), data graphed represent mean values  $\pm$  s.e.m. *P* values in (c, d, e, f) were calculated by Kruskal-Wallis test. In (c), \* - *P* = 0.0259, \*\* - *P* = 0.073. In (d), \* - *P* = 0.014, \*\* - *P* = 0.0027. In (e), \* - *P* = 0.0114, \*\* - *P* = 0.0016, \*\*\*\* - *P* < 0.0001. n.s not significant. Source data are provided as a Source Data file.

and found an inverted repeat with a 12 bp half site –4 bp spacer –12 bp half site motif that likely constitutes *RopBss* binding site (Supplementary Fig. 7e–f). However, the *RopBss* binding site differs from *RopB-GAS* binding site in several aspects including the motif arrangement, length, and nucleotide composition. The *RopB-GAS* binds to a palindrome with a 9 bp half site –7 bp spacer –9 bp half site motif (25 bp long)<sup>10</sup> compared to the 12 bp half site –4 bp spacer –12 bp half site motif (26 bp long) of *RopBss*. The half site of the palindrome in *RopB-GAS* binding site has a nucleotide composition of GTTACGTNT<sup>10</sup>, which varies from *RopBss* binding site that has

nucleotide composition of ATGTAACATATT (Supplementary Fig. 7f). These findings indicate that the receptors recognize operator sequences of different length and nucleotide composition in the target promoters. However, consistent with the role of *RopBss* and *RopB-GAS* as transcription activators<sup>10</sup> and their likely role in the recruitment of RNA polymerase to defective promoters<sup>21</sup>, the binding sites for both receptors are located upstream of and around the –35 region of *LCP* promoters.

To delineate the influence of *LCPss* on *RopBss*-promoter interactions, we assessed *RopBss*-DNA interactions in the presence and

absence of LCP<sub>ss</sub> by FP assay using FITC-labeled oligoduplex containing the identified RopB<sub>ss</sub> binding site (Fig. 4i). The addition of LCP<sub>ss</sub> resulted in high affinity interactions between RopB<sub>ss</sub> and the cognate DNA sequences ( $K_d \sim 90$  nM) compared to that of apo- or SCRA-bound RopB<sub>ss</sub> ( $K_d > 500$  nM) (Fig. 4j and Supplementary Fig. 7g, h), indicating that LCP<sub>ss</sub> binding promotes high affinity interactions between RopB<sub>ss</sub> and LCP<sub>ss</sub> promoter. Based on these observations, we proposed a model for LCP<sub>ss</sub> signaling and LCP<sub>ss</sub>-dependent transcription activation of *NRPS-BGC* by RopB<sub>ss</sub> (Fig. 4k).

### LCP system regulates streptococcal virulence factor production

To assess the functionality of LCPs in other bacteria, we first assessed the regulatory activity of LCP from the swine pathogen *S. porcinus*. The amino acid sequence of *S. porcinus* LCP (LCP<sub>sp</sub>) is identical to that of SIP (Fig. 5a, b). The coding region of LCP<sub>sp</sub> is flanked upstream by *ropB* in the divergent direction and downstream by a gene encoding cysteine protease (*speB<sub>sp</sub>*) that is transcribed convergently (Fig. 5a). Consistent with the role of LCP<sub>sp</sub> as an intercellular signal that controls *speB<sub>sp</sub>* expression, supplementation of *S. porcinus* with synthetic LCP<sub>sp</sub> triggered early induction of *speB<sub>sp</sub>* expression, while the non-cognate SCRA<sub>sp</sub> had no effect on gene regulation (Fig. 5c). Since the secreted cysteine protease SpeB is critical for the virulence of *S. pyogenes*<sup>22,23</sup>, we reason that LCP<sub>sp</sub>-mediated activation of *speB<sub>sp</sub>* expression may impact the pathogenic traits of *S. porcinus*.

### LCPs from different clades of the RopB clan receptors phylogeny mediate intercellular communication and gene regulation

To test whether LCP from non-streptococcal genus is functional, we characterized the regulatory activity of LCP from *Enterococcus malodoratus* (LCP<sub>em</sub>) (Fig. 2). The LCP<sub>em</sub> is 9 amino acid long and its amino acid sequence is distinct from SIP (Fig. 5d, e). The *ropB<sub>em</sub>* is divergently transcribed from LCP<sub>em</sub>. However, unlike the other characterized LCPs (above), there are no convergently transcribed genes downstream of LCP<sub>em</sub> (Fig. 5d). Instead, there are two genes encoding T7 secretion system-effector pair (T7SS) located downstream of *ropB<sub>em</sub>* and transcribed convergently from *ropB<sub>em</sub>*. Additionally, there are two hypothetical genes located downstream of LCP<sub>em</sub> but transcribed divergently from LCP<sub>em</sub> (Fig. 5d). Since the gene arrangement is distinct from other characterized LCP systems and regulatory influence of LCP<sub>em</sub> on these genes is unknown, we investigated the effect of synthetic LCP<sub>em</sub> on the expression of genes in both directions. Supplementation of *E. malodoratus* with LCP<sub>em</sub> induced only the expression of genes encoding T7SS and its effector and the induction was specific for LCP<sub>em</sub> (Fig. 5f). Contrarily, the LCP<sub>em</sub> had no influence on the expression profile of the two hypothetical genes located downstream of LCP<sub>em</sub> (data not shown). These findings demonstrate that the LCP<sub>em</sub> that is dissimilar to SIP acts as an intercellular signal and controls the production of an *E. malodoratus* T7 secretion system/effector system.

To investigate the functionality of a LCP system from a more distant LCP system, we assessed the regulatory activity of LCP from *Limosilactobacillus reuteri* (LCP<sub>lr</sub>) (Figs. 2, 6). Unlike other *L. reuteri* strains, the *L. reuteri* DSM32035 strain has a naturally occurring stop codon at amino acid position 4 of the putative LCP<sub>lr</sub> (Fig. 6b). The predicted untruncated full length LCP<sub>lr</sub> is 8 amino acid long (Fig. 6b, c) with the characteristic aliphatic and aromatic amino acid composition (Fig. 6a–c). The *ropB<sub>lr</sub>* is divergently transcribed from LCP<sub>lr</sub> (Fig. 6a). Two hypothetical genes encoding a putative ABC-type transporter are located downstream of LCP<sub>lr</sub> and transcribed convergently from LCP<sub>lr</sub> (Fig. 6a). Supplementation of synthetic LCP<sub>lr</sub> to the exponential growth of *L. reuteri* activated the naturally silent LCP pathway and induced the expression of genes encoding the ABC-type transporter. The induction was specific for LCP<sub>lr</sub> as the SCRA<sub>lr</sub> failed to activate the expression of ABC transporter (Fig. 6c). These results indicate that LCP from a distant species, LCP<sub>lr</sub>, functions effectively as a qs signal and mediates gene regulation.

## Discussion

Here, we report the widespread prevalence of a new class of bacterial qs signals that remain largely uncharacterized for their roles in bacterial communication<sup>10</sup>. LCPs are unique bacterial qs peptides due to their synthesis from unannotated <30 base pairs usORFs and the absence of hallmarks of characterized bacterial peptide signals (Fig. 2). Notwithstanding the enormous advances in the understanding of bacterial quorum sensing over several decades<sup>1,3,8</sup>, the unique features of LCPs prevented their identification and obscured the critical knowledge of the roles of an entire class of qs signals in bacterial group behaviors. To overcome this, we devised a tailor-made approach to detect LCPs and show that LCPs are prevalent among diverse bacterial species and encode an array of peptide communication codes (Figs. 1, 2, 3). Importantly, the discovery of LCPs indicates that usORFs as small as 24 base pairs can encode a functional bacterial qs signal thus rewriting the rules for the composition of bacterial qs peptides (Fig. 2)<sup>10</sup>. Many RRNPPA regulators are categorized as orphan receptors due to the absence of classical peptide ligands in their genetic vicinity<sup>8,12,24,25</sup>. In light of our findings, we reason that careful analyses of usORFs and their translated products in the genetic vicinity of orphan RRNPPA receptors may identify their cognate LCP ligands. In this regard, the methodology used here provides an effective tool to perform homology-based search to detect additional LCP systems. We anticipate that such analyses may uncover additional LCPs, expand LCP codes, elucidate new roles for LCPs in bacterial pathophysiology, and reveal an even broader distribution of LCPs than indicated by our study.

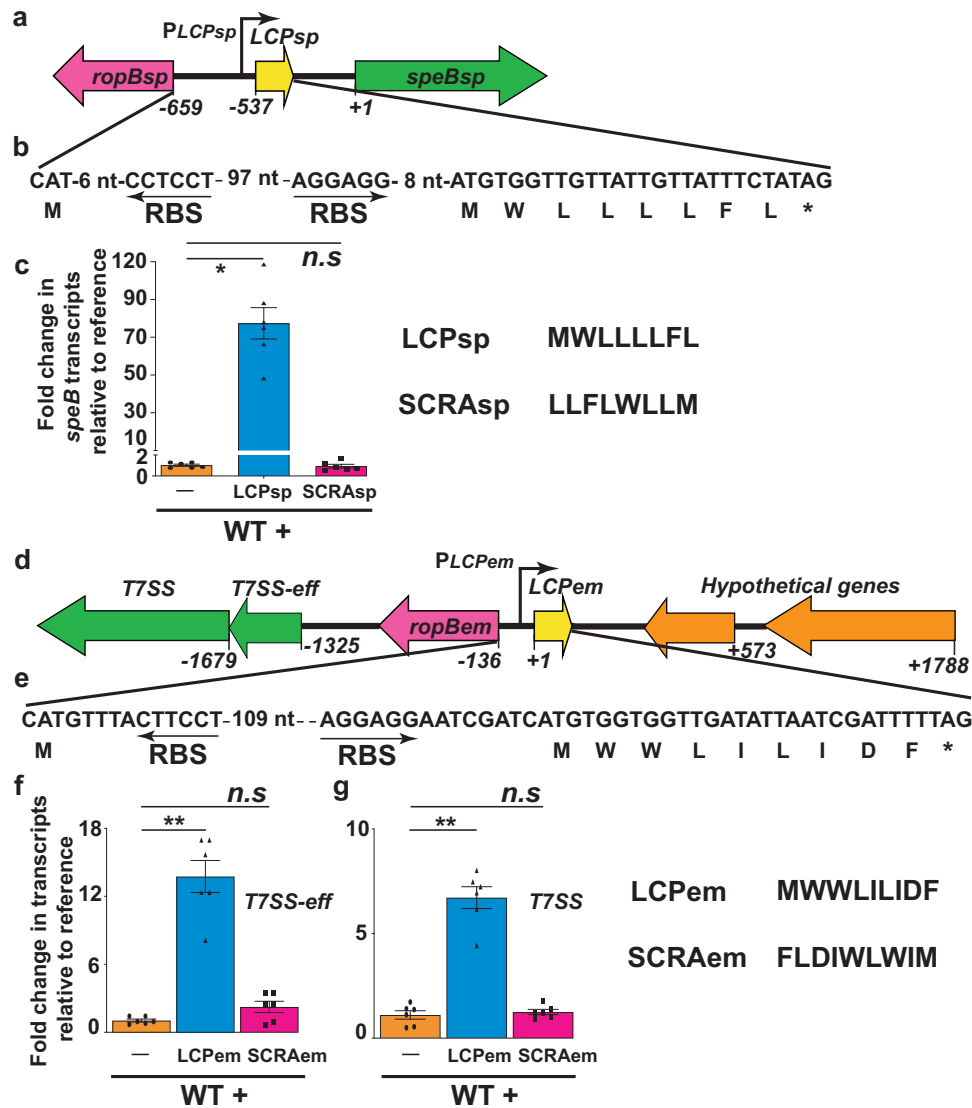
We demonstrate that, despite their presence as mature peptides without accessory sequences, several predicted LCPs act as intercellular signals and upregulate genes critical for various bacterial phenotypes (Figs. 4, 5, 6)<sup>10</sup>. Similarly, the genetic proximity of other uncharacterized LCPs to RopB homologs and the presence of a strong RBS upstream of their predicted start codons suggest that these usORFs also encode functional LCPs and participate in the coordination of bacterial group behaviors (Figs. 2, 3). Given the preponderance of LCPs in firmicutes, a prevalent phylum in human gut and oral microbiota<sup>26,27</sup>, it will be interesting to explore the role of LCPs in communication among human-associated microbial communities, the functions they regulate, and their impact on human health.

On a broader note, the fundamental aspects of LCPs including their evolutionary origin and implications of usORFs encoding mature peptide signals, export and import mechanisms, regulatory mechanisms, genes and phenotypes regulated, and their contribution to microbial behavior remain unknown. Thus, our discovery of the broadly distributed LCPs may open new avenues of research and uncover exciting new knowledge regarding molecular mechanisms of qs signaling, gene regulation, and roles for LCPs in coordinating population-wide bacterial traits.

## Methods

### Distribution of putative signaling peptide-receptor pairs across the RopB-Rgg phylogeny

The complete genomes, chromosomes, and scaffolds of bacteria available on the NCBI Assembly database as of October 2021 were accessed using the following query string: ("Bacteria"[Organism] AND "latest genbank"[filter] AND ("complete genome"[filter] OR "scaffold level"[filter] OR "chromosome level"[filter])). The 448,667,348 protein sequences corresponding to these assemblies were downloaded from the NCBI using Genbank as a source database<sup>28</sup>. The metadata, protein sequences, genomes and annotations of the 9,246 high-quality species-representative MAGs of the four MAGs catalogs available on Mgnify (cow-rumen, human-gut, human-oral and marine) were downloaded from the following location on the Mgnify ftp server: [http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify\\_genomes](http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes)<sup>29</sup>. The search for



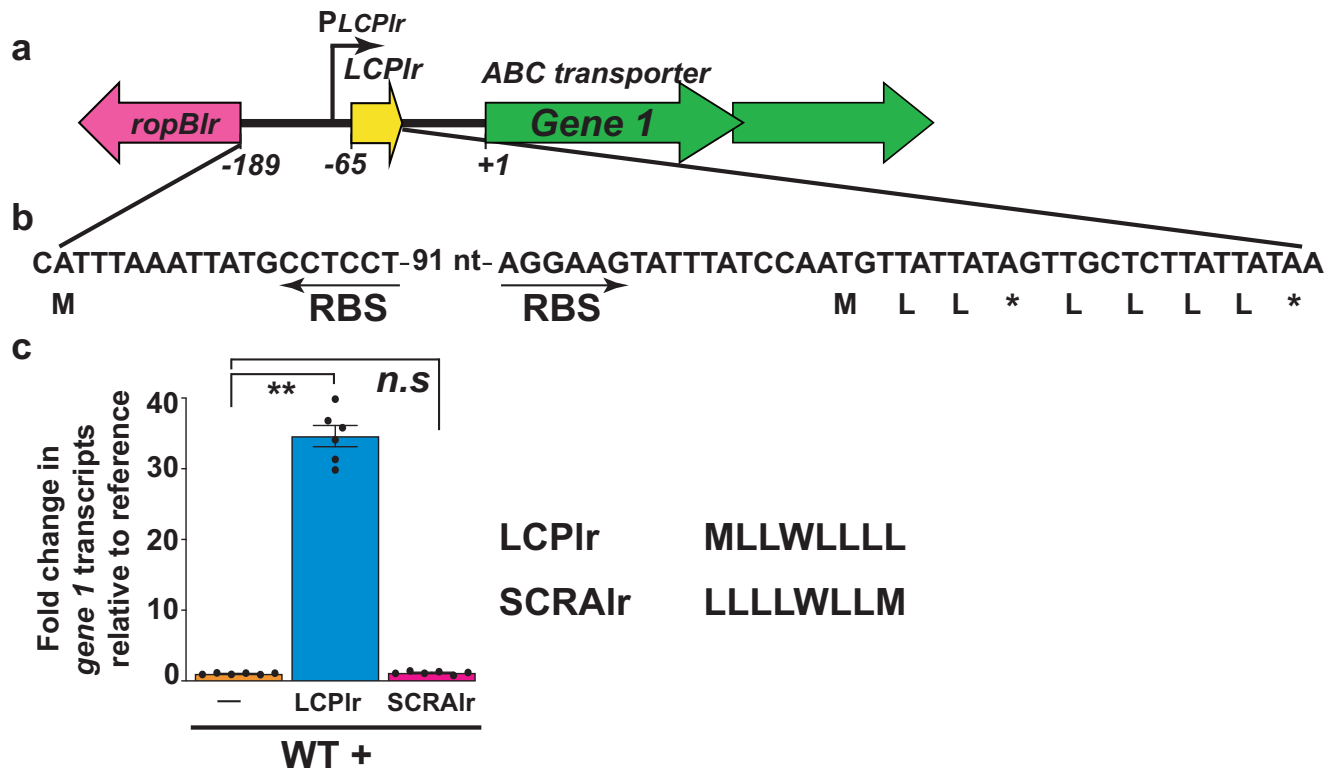
**Fig. 5 | Diverse LCPs mediate intercellular signaling in streptococcus and enterococcus.** **a** Schematic representation of genetic elements in *S. porcinus* (sp) encoding *ropB<sub>sp</sub>*, *LCP<sub>sp</sub>*, and secreted cysteine protease *speB<sub>sp</sub>*. The *ropB<sub>sp</sub>* and *LCP<sub>sp</sub>* are divergently transcribed. The bent arrow above indicates the transcription start site of *LCP<sub>sp</sub>*. The numbers below denote the nucleotide positions relative to the first nucleotide of the start codon of *speB<sub>sp</sub>*. **b** The nucleotide sequence of the *ropB<sub>sp</sub>*-*LCP<sub>sp</sub>* intergenic region, coding sequence of *LCP<sub>sp</sub>*, and corresponding predicted amino acid sequence of *LCP<sub>sp</sub>* are shown. The ribosomal-binding sites (RBS) of *LCP<sub>sp</sub>* and *ropB<sub>sp</sub>* are marked by arrows. **c** Addition of synthetic *LCP<sub>sp</sub>* causes early induction of *speB* expression in WT *S. porcinus*. *speB* transcript levels were assessed by qRT-PCR and the fold change in *speB* expression relative to unsupplemented growth (reference) is shown. **d** Schematics of genetic elements in *E. malodoratus* (*em*) encoding *ropB<sub>em</sub>*, *LCP<sub>em</sub>*, and genes encoding putative T7 secretion system (*T7SS*) and cognate effector (*T7SS-eff*). The bent arrow above

indicates the predicted transcription start site (*PLCP<sub>em</sub>*) of *LCP<sub>em</sub>*. The numbers below denote the nucleotide positions relative to the first nucleotide of the start codon of *LCP<sub>em</sub>*. **e** The nucleotide sequence of the *ropB<sub>em</sub>*-*LCP<sub>em</sub>* intergenic region, coding sequence of *LCP<sub>em</sub>*, and corresponding predicted amino acid sequence of *LCP<sub>em</sub>* are shown. The ribosomal-binding sites of *LCP<sub>em</sub>* and *ropB<sub>em</sub>* are underlined. **f** Addition of synthetic *LCP<sub>em</sub>* causes early induction of *T7SS* and *T7SS-eff* expression in WT *E. malodoratus*. Transcript levels were assessed by qRT-PCR and the fold change in *gene* expression relative to unsupplemented growth (reference) is shown. In **c**, **f**, data are derived from three biological replicates analyzed in duplicate and data graphed represent mean values  $\pm$  s.e.m. *P* values in (**c**, **f**) were calculated by Kruskal-Wallis test. In (**c**), \* - *P* = 0.0189, n.s not significant. In panel **f**, \*\* - *P* = 0.0015, n.s - not significant, whereas in (**g**), \*\* - *P* = 0.0049, n.s - not significant. Source data are provided as a Source Data file.

RopB homologs against these two target datasets was performed by BLASTP version 2.9.0+ with the option `-max_target_seqs` set to 1,000,000,000 to ensure that no RopB hits will be missing from the output file<sup>30</sup>. The output of BLASTP was filtered with the 25% identity and 70% mutual coverage cutoffs. These thresholds were chosen to capture a significant diversity of sequences while offering good guarantees on their expected function of receptor of small hydrophobic signaling peptides (irrespective of whether these peptides underwent post-translational cleavage (like SHPs) or not (like SIPs)), as these values correspond to the similarity observed between RopB and Rgg receptors in *S. pyogenes*. The presence of the

C-terminal peptide-interacting tetratricopeptide repeats of the RopB-Rgg superfamily in RopB homologs was determined by hmsearch (HMMER version 3.3), using the TIGR01716 HMM profile from the TIGRFAM database as a query (E-value < 1E-5)<sup>31</sup>. The upstream region (-450 to +20 bp from start codon) and the downstream region (-20 to +450 bp from stop codon) of each of the *ropB* homologs were either extracted from the nucleotide fasta file of MAGs (MGnify) or downloaded from the NCBI (Genbank dataset) via [https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nuccore&id=<genomic\\_accession>&seq\\_start=<region\\_start>&seq\\_stop=<region\\_end>&rettype=fasta](https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nuccore&id=<genomic_accession>&seq_start=<region_start>&seq_stop=<region_end>&rettype=fasta).





**Fig. 6 | LCP from a diverse species mediate intercellular signaling in *Limosilactobacillus reuteri*.** **a** Schematic representation of genetic elements in *L. reuteri* (*lr*) encoding *ropB<sub>lr</sub>*, *LCP<sub>lr</sub>*, ABC transporter (*Gene 1*). The *ropB<sub>lr</sub>* and *LCP<sub>lr</sub>* are divergently transcribed. The bent arrow above indicates the transcription start site of *LCP<sub>lr</sub>*. The numbers below denote the nucleotide positions relative to the first nucleotide of the start codon of *gene 1*. **b** The nucleotide sequence of the *ropB<sub>lr</sub>*-*LCP<sub>lr</sub>* intergenic region, coding sequence of *LCP<sub>lr</sub>*, and corresponding predicted amino acid sequence of *LCP<sub>lr</sub>* are shown. The ribosomal-binding sites (RBS) of *LCP<sub>lr</sub>*

and *RopB<sub>lr</sub>* are marked by arrows. **c** Addition of synthetic *LCP<sub>lr</sub>* causes early induction of *gene 1* expression in WT *L. reuteri*. *Gene 1* transcript levels were assessed by qRT-PCR and the fold change in expression relative to unsupplemented growth (reference) is shown. In (c), data are derived from three biological replicates analyzed in duplicate and data graphed represent mean values  $\pm$  s.e.m. *P* values in (c) were calculated by Kruskal-Wallis test. \*\* - *P* = 0.0024, n.s not significant. Source data are provided as a Source Data file.

The reverse complements of the two flanking regions of each *ropB* homolog were then generated for the minus strand to facilitate subsequent sORF calling and RBS motif detection. The Orfpy python software version 0.0.4 was launched against all these nucleotide flanking regions with the options `--strand f --min 18 --max 300 --start ATG` to select for any putative sORF encoding a micropeptide between 6 and 100 aa<sup>32</sup>. Nested micropeptides were then extracted based on the occurrence of an in frame ATG codon between the start and the stop codons of a longer ORF, if the resulting length was  $\geq 6$ aa. The hierarchy of 27 regular expressions introduced by the Prodigal ORF-calling tool to detect Shine-Dalgarno RBS motifs was then applied against the upstream region of each detected ORF (-21 to -1bp from start codon)<sup>13</sup>, consistent with the fact that  $\geq 90\%$  of the protein coding genes encoded by Firmicutes are preceded by a Shine-Dalgarno RBS<sup>14</sup>. In a preliminary effort to assign a unique peptide to each RopB homolog in a stringent way, the neighboring micropeptide preceded by the RBS motif with the highest bin/score according to Prodigal was selected, providing that its score was above 13 and its usage across prokaryotes is greater than 3% according to Omotajo et al. (i.e bins 27, 24, 22, 19, 16 and 15)<sup>14</sup>. In case of an RBS score equality between peptides, a divergent context, and a small intergenic distance to *ropB* were prioritized. The phylogeny of RopB was then reconstructed as follows. First, RopB-like proteins were reduced to clusters of identical proteins using CD-HIT version 4.8.1 with the option `-c 1`<sup>33</sup>. Second, a multiple sequence alignment of the dereplicated RopB homologs was produced with mafft v7.453 with the options `-maxiterate 1000 -localpair` for high accuracy<sup>34,35</sup>, and

trimmed with trimAl version 1.2.rev59 with the option `-automated 1` optimized for maximum likelihood phylogenetic reconstruction<sup>36</sup>. Third, the trimmed alignment of 174 sites was given as input to IQ-TREE version multicore version 1.6.12 to infer a maximum likelihood phylogenetic under the LG+G model with 1000 ultrafast bootstraps<sup>37</sup>. Finally, the topology of the tree and the peptide mapping onto the leaves of the tree were visualized using ITOL<sup>38</sup>. Remarkable clans of receptors were delimited based on the topology of the tree, and the amino-acid profile of mapped peptides. Receptor-peptide pairs within RopB and Rgg clans were further manually refined by carefully examining the list of all peptides encoded by a sORF in the vicinity of each representative *ropB* or *rgg* gene. SHPs were defined according to an amino-acid profile similar to that of the reference sequences listed in<sup>15,16,39</sup>. LCPs were identified based on a high RBS score, the consistency of mapped peptides across adjacent leaves of the tree, a genomic context preferably divergent relative to the receptor, a short length and an enrichment in aromatic and aliphatic residues (such as SIP).

#### Computational characterization of candidate receptor-LCP pairs within the RopB clan

The phylogeny of members of the RopB clan was inferred using mafft<sup>34,35</sup> and IQ-tree with the same parameters as cited above. The comparative analysis of amino-acid type enrichment and physico-chemical properties between LCPs, SHPs and PrgX-like peptides were conducted as follows. First, mature SHP peptides and enterococcal PrgX-like peptides secreted via the same PptAB translocon as SHPs were retrieved from the Quorumpeps database<sup>39</sup>. The aaComp function

of the Peptides R package was used to create a matrix storing the abundance of each amino-acid type within each input quorum sensing peptide<sup>40</sup>. This matrix was used to create a heatmap with the ComplexHeatmap R package, with the quorum sensing peptides being clustered according their amino-acid type composition, using a Pearson correlation as the distance metric and the ward.D2 algorithm as the clustering method<sup>41</sup>. The 66 physico-chemical descriptors of each sequence were obtained by combining the output of the `blosu-mIndices`, `crucianiProperties`, `fsgaiVectors`, `kider-aFactors`, `mshimScores`, `ProtFP`, `stScales`, `tScales`, `vhseScales` and `zScales` functions of the Peptides R package. A PCA was then performed on this space of 66 dimensions with the `prcomp` function. The encoding element analysis was conducted as follows. First, the nucleotide fasta and the genbank files of each genomic accession / contig encoding a candidate receptor-LCP pair were fetched from the NCBI. Plasflow version 1.1 was run with default parameters against all these contigs in order to classify each contig according to one of these three categories: chromosome, plasmid or unclassified<sup>42</sup>. The genbank files were given as input to ICEscreen version 1.0.4 to identify regions corresponding to ICEs or IMEs within contigs<sup>43</sup>. Virsorter2 version 2.2.3 was launched with default parameters against the concatenated nucleotide fasta file and returned possible prophage regions within contigs<sup>44</sup>. The coordinates of ICEs, IMEs and prophages were then intersected with the coordinates of candidate LCP system(s) on corresponding contigs to define putative systems included within these mobile genetic elements. Because the provirus boundary detection algorithm in VirSorter2 purposely tends to overextend to host regions, the Phaster webtool was further used to confirm or infirm the prediction by VirSorter2 that a candidate LCP-based system is included within a prophage region (the few cases were all infirmed)<sup>45</sup>. The biome distribution analysis of LCP systems was performed as follows. First, the `isolation_source` tag was extracted from the genbank files of contigs encoding a candidate LCP system to obtain their environment of isolation. Next, a Blastp search of each receptor associated with a putative LCP was launched (sequence identity  $\geq 90\%$ , mutual coverage  $\geq 90\%$ ) against the MGnify database of 623,796,864 clusters of  $\geq 90\%$  identical proteins: [https://ftp.ebi.ac.uk/pub/databases/metagenomics/peptide\\_database/current\\_release/mgy\\_clusters.fa](https://ftp.ebi.ac.uk/pub/databases/metagenomics/peptide_database/current_release/mgy_clusters.fa) to obtain the biome of highly similar environmental sequences. The functional prediction of the target regulon of candidate LCP systems was conducted by running the Antismash webtool against the reference contig of each representative receptor-LCP pair<sup>46</sup>. Each BGC found right next to a receptor-LCP pair was assessed as the putative target regulon. When no adjacent BGC was identified, we carefully looked at the genomic orientation and the functional annotation of each neighboring genes. Of the two flanking regions, if one happened to be in a co-directional context relative to the LCP, this region was chosen as the putative target regulon.

### Co-evolution analysis of RopB-LCP pairs

The MSA of receptors from RopB clan paired with a candidate LCP was performed with `mafft` with the parameters `-maxiterate 1000 -localpair` for high accuracy. Coordinates of structural elements of RopB were extracted from AlphaFold prediction for the N-terminal DNA binding domain (AlphaFoldDB identifier O85731) and from the structurally resolved C-terminal LCP binding domain (PDB identifier: 6DQL). Structural elements and LCP-contacting residues were mapped onto the MSA, using RopB amino-acids as anchors. Sequences of LCP-contacting residues in RopB homologs were extracted from the previous MSA. Site-wise dN/dS ratios were derived as follows. First, coding sequences (CDS) of aligned receptors were fetched from the NCBI Genbank database. Second, the MSA was trimmed with `trimAl` to retain only sites with less than 20% gaps<sup>36</sup>. Third, a phylogenetic tree was derived from the trimmed alignment (with the same pipeline). Fourth, the protein MSA was converted into corresponding CDS MSA using

amino acid to codon mapping. Site-wise dN/dS were finally derived from the CDS MSA and the phylogenetic tree using HYPHY v2.5.51 with FEL-contrast method<sup>47,48</sup>. Comparisons of pairwise evolutionary distances between receptors, contacting residues and LCPs were derived as following. For each category (receptors, contact residues, LCPs), all vs all pairwise alignments were performed. Since sequences of LCPs and contacting residues are ultrasmall, exact Smith-Waterman alignment algorithm implemented in the `fasta36` software v36.3.8i was used to ensure each alignment is the optimal one<sup>49</sup>. To derive an evolutionary distance metrics that is comparable between elements of different length (e.g., long receptors vs short LCPs / short contacting residues sequence), we used the following formula that normalize mutual bit scores between two sequences by the bit scores of their self-alignment:  $1 - (\text{bitscore A vs B} + \text{bitscore B vs A}) / (\text{bitscore A vs A} + \text{bitscore B vs B})$ .

**Bacterial strains, plasmids, and growth conditions.** Bacterial strains and plasmids used in this study are listed in Supplementary Table 1. *S. salivarius* DS85\_40B is a previously described strain isolated from the dietary supplements and food products whose genome has been fully sequenced<sup>50</sup>. *Escherichia coli* DH5 $\alpha$  strain was used as the host for plasmid constructions and BL21(DE3) strain was used for recombinant protein overexpression. *S. salivarius*, *S. porcinus*, and *E. malodoratus* were grown routinely in Todd-Hewitt broth containing 0.2% (w/v) yeast extract (THY; DIFCO) at 37 °C. *Limosilactobacillus reuteri* was grown in DeMan, Rogosa and Sharpe (MRS) broth at 37 °C under microaerophilic growth conditions in a sealed chamber provided with anaerobe sachets (BD GasPak EZ Container Systems). When required, kanamycin was added to a final concentration of 100  $\mu\text{g/ml}$ . Chloramphenicol was used at a final concentration of 15  $\mu\text{g/ml}$ . All bacterial growth experiments were done in triplicate on three separate occasions for a total of nine replicates. Overnight cultures were inoculated into fresh media to achieve an initial absorption at 600 nm ( $A_{600}$ ) of 0.03. Bacterial growth was monitored by measuring the absorption at 600 nm ( $A_{600}$ ). The *E. coli* strain used for protein overexpression was grown in Luria Broth Miller (LB broth; Thermo Scientific).

**Construction of isogenic mutant and cis-complemented strains.** Isoallelic strains containing either single codon changes or inactivation of entire coding region were generated as previously described<sup>51</sup>. A DNA fragment with approximately 600 bp on either side of the coding region of interest was amplified using the primers listed in Supplementary Table 2 and cloned into the multi-cloning site of the temperature-sensitive plasmid pJL1055<sup>52</sup>. The resultant plasmids were introduced into *S. salivarius* by competence-based DNA uptake. Briefly, overnight *S. salivarius* growth was diluted in 0.3 ml of chemically defined medium<sup>10</sup> and incubated at 37 °C for 75 min. Subsequently, synthetic competence-stimulating (ComS) peptide with the amino acid sequence of LPYFAGCL and plasmid was added to the cells and incubated for 3 h at 37 °C. Cells were plated on agar plates containing appropriate antibiotics. The plasmid pJL1055 containing the intact genes were used to generate *S. salivarius* genetic revertant strains that has the gene of interest re-introduced into its original genetic locus. Colonies with plasmid incorporated into the *S. salivarius* megaplasmid were selected for subsequent plasmid curing. DNA sequencing was then performed to ensure that no spurious mutations were introduced.

**Construction of *opp<sub>ss</sub>*-inactivated mutant strain.** Insertional inactivation of the *opp<sub>ss</sub>* gene in WT *S. salivarius* was performed by methods described below. Briefly, a PCR fragment containing a spectinomycin resistance (*spc*) cassette with the fragment of gene to be deleted on either side was generated in a three-step PCR process. Subsequently, the plasmid with the *spc* gene disruption cassette was introduced into the parent strain by electroporation, and the gene was disrupted

through homologous recombination. The isogenic mutant strains were selected by growth on spectinomycin-containing medium. Inactivation of the gene was confirmed by DNA sequencing. Primers used for the construction of the isogenic  $\Delta opp_{ss}$  mutant strain are listed in Supplementary Table 2.

**Transcript level analyses by qRT-PCR.** *S. salivarius*, *S. porcinus*, and *E. malodoratus* strains were grown in Todd-Hewitt Broth to late-exponential phase ( $A_{600} \sim 2.0$ ), early exponential phase ( $A_{600} \sim 0.6$ ), and mid-exponential phase ( $A_{600} \sim 1.0$ ) respectively. *L. reuteri* was grown in MRS broth to late-exponential phase (6 h post inoculation of the seed culture at 1:100 dilution). Cells were incubated with the corresponding LCP and cultures were immediately mixed with two volumes of ice-cold acetone and ethanol mixture and stored at  $-80^{\circ}\text{C}$  till use. Bacteria were harvested by centrifugation and RNA isolation and purification were performed using a RNeasy kit (Qiagen). Purified RNA was analyzed for quality and concentration with an Agilent 2100 Bioanalyzer. cDNA was synthesized from the purified RNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and qRT-PCR was performed with a Real-Time PCR Detection System (Bio-Rad). Comparison of transcript levels was done using  $\Delta C_T$  method of analysis using the housekeeping genes mentioned in Supplementary Table 2.

**Culture supernatant swap assay.** To assess the presence or absence of LCP<sub>ss</sub> associated regulatory activity in the stationary growth phase culture supernatant of *S. salivarius* that induces *NRPS-BGC* expression, the indicated strains were grown to late-exponential phase ( $A_{600} \sim 2.0$ ). Cell-free culture supernatants were prepared by centrifugation and filtering through  $0.22\ \mu\text{m}$  membrane filter. The cell pellets of the wild type *S. salivarius* grown to mid-exponential growth phase ( $A_{600} \sim 2.0$ ) were resuspended in the secretome prepared from the mid-exponential growth phase of the indicated strains and incubated at  $37^{\circ}\text{C}$  for 1 h. Transcript level analyses was performed by qRT-PCR as described above.

**Synthetic peptide addition assay.** Synthetic peptides of high purity (>90% purity) obtained from Peptide 2.0 (Chantilly, VA) were suspended in 100% DMSO to prepare a 10 mM stock solution. Stock solutions were aliquoted and stored at  $-20^{\circ}\text{C}$  until use. Working stocks were prepared by diluting in DMSO.

**Peptide reimport studies by microscopy.** To demonstrate the internalization of FITC-LCP<sub>ss</sub>, *S. salivarius* cells were grown to late-exponential phase of growth ( $A_{600} \sim 2.0$ ), incubated with the indicated concentrations of unmodified or FITC-labeled LCP<sub>ss</sub> for 30 min at  $37^{\circ}\text{C}$ . Cells were harvested by centrifugation, and washed three times with sterile PBS. Cells were fixed on the coverslip using 1% glutaraldehyde and 3% formaldehyde. The images were taken using a Nikon Eclipse TiN-STORM super resolution microscope equipped with iXon3 897 EM-CCD camera.

**Peptide reimport studies by fluorescence measurements.** To demonstrate the cytosolic internalization of exogenously added FITC-labeled LCP<sub>ss</sub>, indicated *S. salivarius* strains were grown to late-exponential phase of growth ( $A_{600} \sim 2.0$ ) and incubated with either the indicated synthetic peptide or the carrier for the synthetic peptides (DMSO) for 30 min at  $37^{\circ}\text{C}$ . Cells were harvested by centrifugation, washed three times with sterile PBS, and resuspended in equal volume of PBS. Cells were lysed by FastPrep-24 (MP Biomedicals) and lysates were clarified by centrifugation at  $18,000 \times g$  at  $4^{\circ}\text{C}$  for 30 min. Samples were analyzed in 100  $\mu\text{l}$  volume using an excitation and emission wavelengths of 490 and 520 nm, respectively. Readings were taken using a Biotek microplate reader (Biotek) and fluorescence measurements in relative fluorescence units were reported.

**Recombinant RopBss overexpression and purification.** The *ropBss* gene of *S. salivarius* was cloned into plasmid pET-28a and protein was overexpressed in *E. coli* strain BL21 (DE3) after induction with 1 mM IPTG at  $25^{\circ}\text{C}$  for 16 h. Cell pellets were suspended in 50 ml of buffer A (20 mM Tris HCl pH 8.0, 100 mM NaCl and 1 mM Tris 2-carboxyethyl phosphine hydrochloride (TCEP)) supplemented with DNase I to a final concentration of  $5\ \mu\text{g}/\text{ml}$ . Cells were lysed by a Microfluidizer LM10 (Microfluidics) and cell debris was removed by centrifugation at  $18,000 \times g$  for 30 min. RopBss was purified by affinity chromatography using a Ni-NTA agarose column. The concentrated RopBss was further purified by size exclusion chromatography with a Superdex 200 G column and stored in buffer B (20 mM Tris HCl pH 8.0, 200 mM NaCl, 1 mM TCEP and 1 mM EDTA). The protein was purified to >95% homogeneity and concentrated to a final concentration of  $\sim 5\ \text{mg}/\text{ml}$ .

**Electrophoretic mobility shift assays.** Probes containing different fragments of the *LCP<sub>ss</sub>* and *LCP<sub>ss</sub>-NRPS-BGC* intergenic region were annealed by heating equimolar mixture of top and bottom strand oligonucleotides at  $95^{\circ}\text{C}$  for 5 min followed by slow cooling to room temperature. Binding reactions were carried out in 20  $\mu\text{l}$  volume of binding buffer (20 mM Tris pH 7.5, 0.8 M NaCl, 12% glycerol and 100  $\mu\text{M}$  Zinc sulfate) containing 0.5  $\mu\text{M}$  of oligoduplex and increasing concentrations of RopBss. After 15-min incubation at  $37^{\circ}\text{C}$ , the reaction mixtures were resolved on a 10% native polyacrylamide gel supplemented with 5% glycerol (native polyacrylamide gel electrophoresis (PAGE)) for 90 min at 100 V at  $4^{\circ}\text{C}$  in Tris Borate buffer with 5% glycerol. The gels were stained with ethidium bromide and analyzed on a BioRad Gel Electrophoresis Systems.

**Fluorescence polarization assay to assess protein-peptide and protein-DNA binding interactions.** Fluorescence polarization-based RopB<sub>ss</sub>-ligand binding experiments were performed with a Biotek microplate reader (Biotek) using the intrinsic fluorescence of fluorescein labeled DNA or synthetic peptides. The polarization (P) of the labeled DNA or synthetic peptides increases as a function of protein binding, and equilibrium dissociation constants were determined from plots of millipolarization ( $P \times 10^{-3}$ ) against protein concentration.

For RopB<sub>ss</sub>-DNA-binding studies, 1 nM 5'-fluoresceinated oligoduplex in binding buffer (20 mM Tris-HCl pH 8.5, 200 mM NaCl, 1 mM TCEP and 25% DMSO) was titrated against increasing concentrations of purified RopB<sub>ss</sub> and the resulting change in polarization measured. Samples were excited at 490 nm and emission measured at 530 nm. The RopB<sub>ss</sub>-peptide-binding studies were performed in a peptide-binding buffer composed of 20 mM potassium phosphate pH 6.0, 75 mM NaCl, 2% DMSO, 1 mM EDTA and 0.0005% Tween 20. All data were plotted using KaleidaGraph and the resulting plots were fitted with the equation  $P = \{(P_{\text{bound}} - P_{\text{free}})[\text{protein}]/(K_D + [\text{protein}])\} + P_{\text{free}}$ , where P is the polarization measured at a given protein concentration,  $P_{\text{free}}$  is the initial polarization of the free ligand,  $P_{\text{bound}}$  is the maximum polarization of specifically bound ligand and  $[\text{protein}]$  is the protein concentration. Nonlinear least squares analysis was used to determine  $P_{\text{bound}}$ , and  $K_D$ . The binding constant reported is the average value from at least three independent experimental measurements.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Data supporting the findings of this study are available in this article and its Supplementary Information files, or from the corresponding author upon request. Source data are provided with this paper.



## References

1. Azimi, S., Klementiev, A. D., Whiteley, M. & Diggle, S. P. Bacterial quorum sensing during infection. *Annu. Rev. Microbiol.* **74**, 201–219 (2020).
2. Hughes, D. T. & Sperandio, V. Inter-kingdom signalling: communication between bacteria and their hosts. *Nat. Rev. Microbiol.* **6**, 111–120 (2008).
3. Mukherjee, S. & Bassler, B. L. Bacterial quorum sensing in complex and dynamically changing environments. *Nat. Rev. Microbiol.* **17**, 371–382 (2019).
4. Parker, C. T. & Sperandio, V. Cell-to-cell signalling during pathogenesis. *Cell. Microbiol.* **11**, 363–369 (2009).
5. Somerville, G. A. & Proctor, R. A. At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci. *Microbiol. Mol. Biol. Rev.* **73**, 233–248 (2009).
6. Thoendel, M., Kavanaugh, J. S., Flack, C. E. & Horswill, A. R. Peptide signaling in the staphylococci. *Chem. Rev.* **111**, 117–151 (2011).
7. Do, H. & Kumaraswami, M. Structural mechanisms of peptide recognition and allosteric modulation of gene regulation by the RRNPP family of quorum-sensing regulators. *J. Mol. Biol.* **428**, 2793–2804 (2016).
8. Neiditch, M. B., Capodagli, G. C., Prehna, G. & Federle, M. J. Genetic and structural analyses of RRNPP intercellular peptide signaling of Gram-positive bacteria. *Annu. Rev. Gen.* **51**, 311–333 (2017).
9. Perez-Pascual, D., Monnet, V. & Gardan, R. Bacterial cell–cell communication in the host via RRNPP peptide-binding regulators. *Front. Microbiol.* **7**, 706 (2016).
10. Do, H. et al. Leaderless secreted peptide signaling molecule alters global gene expression and increases virulence of a human bacterial pathogen. *Proc. Natl. Acad. Sci. USA* **114**, E8498–E8507, <https://doi.org/10.1073/pnas.1705972114> (2017).
11. Do, H. et al. Environmental pH and peptide signaling control virulence of *Streptococcus pyogenes* via a quorum-sensing pathway. *Nat. Commun.* **10**, 1–14 (2019).
12. Felipe-Ruiz, A., Marina, A. & Rocha, E. P. Structural and genomic evolution of RRNPPA systems and their pheromone signaling. *MBio* **13**, e02514–e02522 (2022).
13. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* **11**, 1–11 (2010).
14. Omotajo, D., Tate, T., Cho, H. & Choudhary, M. Distribution and diversity of ribosome binding sites in prokaryotic genomes. *BMC Genomics* **16**, 1–8 (2015).
15. Ibrahim, M. et al. A genome-wide survey of short coding sequences in streptococci. *Microbiology* **153**, 3631–3644 (2007).
16. Fleuchot, B. et al. Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. *Mol. Microbiol.* **80**, 1102–1119 (2011).
17. Makthal, N. et al. Structural and functional analysis of RopB: a major virulence regulator in *Streptococcus pyogenes*. *Mol. Microbiol.* **99**, 1119–1133 (2016).
18. Zeytuni, N. & Zarivach, R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* **20**, 397–405 (2012).
19. Cerveny, L. et al. Tetratricopeptide Repeat Motifs in the World of Bacterial Pathogens: Role in Virulence Mechanisms. *Infect. Immun.* **81**, 629–635 (2013).
20. Slamti, L. & Lereclus, D. The oligopeptide ABC-importers are essential communication channels in Gram-positive bacteria. *Res. Microbiol.* **170**, 338–344 (2019).
21. Busby, S. J. Transcription activation in bacteria: ancient and modern. *Microbiology* **165**, 386–395 (2019).
22. Carroll, R. K. & Musser, J. M. From transcription to activation: how group A streptococcus, the flesh-eating pathogen, regulates SpeB cysteine protease production. *Mol. Microbiol.* **81**, 588–601 (2011).
23. Shelburne, S. A. III et al. An amino-terminal signal peptide of Vfr protein negatively influences RopB-dependent SpeB expression and attenuates virulence in *Streptococcus pyogenes*. *Mol. Microbiol.* **82**, 1481–1495 (2011).
24. Even-Tov, E., Omer Bendori, S., Pollak, S. & Eldar, A. Transient duplication-dependent divergence and horizontal transfer underlie the evolutionary dynamics of bacterial cell–cell signaling. *PLoS Biol.* **14**, e2000330 (2016).
25. Parashar, V., Konkol, M. A., Kearns, D. B. & Neiditch, M. B. A plasmid-encoded phosphatase regulates *Bacillus subtilis* biofilm architecture, sporulation, and genetic competence. *J. Bacteriol.* **195**, 2437–2448 (2013).
26. Zarco, M., Vess, T. & Ginsburg, G. The oral microbiome in health and disease and the potential impact on personalized dental medicine. *Oral. Dis.* **18**, 109–120 (2012).
27. Hou, K. et al. Microbiota in health and diseases. *Signal Transduct. Target. Ther.* **7**, 135 (2022).
28. Benson, D., Cavanaugh, M., Clark, K., Karsch, I. M. DJ, L. J. Ostell, & Sayers EW. GenBank. *Nucl. Acids Res.* **41**, D36–D42 (2013).
29. Mitchell, A. L. et al. MGnify: the microbiome analysis resource in 2020. *Nucl. Acids Res.* **48**, D570–D578 (2020).
30. Altschul, S., Warren, G., Webb, M., Eugene, W. M. & David, J. L. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
31. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comp. Biol.* **7**, e1002195 (2011).
32. Singh, U. & Wurtele, E. S. orfipy: a fast and flexible tool for extracting ORFs. *Bioinformatics* **37**, 3019–3020 (2021).
33. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).
34. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl. Acids Res.* **30**, 3059–3066 (2002).
35. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
36. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
37. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
38. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucl. Acids Res.* **47**, W256–W259 (2019).
39. Wynendaele, E. et al. Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. *Nucl. Acids Res.* **41**, D655–D659 (2013).
40. Osorio, D., Rondón-Villarreal, P. & Torres, R. Peptides: A package for data mining of antimicrobial peptides. *Small* **12**, 44–444 (2015).
41. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
42. Krawczyk, P. S., Lipinski, L. & Dziembowski, A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucl. Acids Res.* **46**, e35–e35 (2018).
43. Chiapello, H. et al. ICEScreen, Nancy-Université; Inrae, (2022).
44. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 1–13 (2021).
45. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucl. Acids Res.* **44**, W16–W21 (2016).
46. Blin, K. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucl. Acids Res.* **49**, W29–W35 (2021).



47. Pond, S. L. K., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
48. Kosakovsky Pond, S. L., Wisotsky, S. R., Escalante, A., Magalis, B. R. & Weaver, S. Contrast-FEL—a test for differences in selective pressures at individual sites among clades and sets of branches. *Mol. Biol. Evol.* **38**, 1184–1198 (2021).
49. Pearson, W. R. Finding protein and nucleotide similarities with FASTA. *Curr. Protoc. Bioinforma.* **53**, 1–3.9. 25 (2016).
50. Barnaba, T. J., Gangiredla, J., Mammel, M. K., Lacher, D. W. & Tartera, C. Draft genome sequences of 18 streptococcus strains isolated from live dietary supplements and cultured food products. *Microbiol. Resour. Announc.* **11**, e00266–00222, <https://doi.org/10.1128/mra.00266-22> (2022).
51. Carroll, R. K. et al. Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. *J. Clin. Invest.* **121**, 1956–1968 (2011).
52. Li, J., Kasper, D. L., Ausubel, F. M., Rosner, B. & Michel, J. L. Inactivation of the  $\alpha$  C protein antigen gene, bca, by a novel shuttle/suicide vector results in attenuation of virulence and immunity in group B Streptococcus. *Proc. Natl. Acad. Sci. USA* **94**, 13251–13256 (1997).

## Acknowledgements

This work was supported in part by the National Institutes of Health grants R01 AI146048 and R01AI162748 to M.K. We thank Carmen Tartera for providing *S. salivarius* strain.

## Author contributions

S.A., E.H., H.D., N.M., Y.L., E.B., P.L., C.B., and M.K. designed and performed research; E.H., H.D., N.M., S.A., C.B., and M.K. analyzed data; S.A., E.H., H.D., C.B., and M.K. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41719-3>.

**Correspondence** and requests for materials should be addressed to Charles Bernard or Muthiah Kumaraswami.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023