Rich Repertoire of Quorum Sensing Protein Coding Sequences in CPR and DPANN Associated with Interspecies and Interkingdom Communication

Charles Bernard,[a,b] Romain Lannes,[a*] Yanyan Li,[b] Éric Bapteste,[a] Philippe Lopez[a]

aInstitut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d'Histoire Naturelle, Paris, France
bUnité Molécules de Communication et Adaptation des Micro-organismes (MCAM), CNRS, Museum National d'Histoire Naturelle, Paris, France

ABSTRACT The bacterial candidate phyla radiation (CPR) and the archaeal DPANN superphylum are two novel lineages that have substantially expanded the tree of life due to their large phylogenetic diversity. Because of their ultrasmall size, reduced genome, and lack of core biosynthetic capabilities, most CPR and DPANN members are predicted to be sustained through their interactions with other species. How the few characterized CPR and DPANN symbionts achieve these critical interactions is, however, poorly understood. Here, we conducted an *in silico* analysis on 2,597 CPR/DPANN genomes to test whether these ultrasmall microorganisms might encode homologs of reference proteins involved in the synthesis and/or the detection of 26 different types of communication molecules (quorum sensing [QS] signals), since QS signals are well-known mediators of intra- and interorganismic relationships. We report the discovery of 5,693 variants of QS proteins distributed across 63 CPR and 6 DPANN phyla and associated with 14 distinct types of communication molecules, most of which were characterized as interspecies QS signals.

IMPORTANCE The selection of predicted genes for interspecies communication within the CPR and DPANN genomes sheds some light onto the underlying mechanisms supporting their inferred symbiotic lifestyle. Also, considering the lack of core pathways such as the *de novo* synthesis of nucleotides or amino acids in the CPR and DPANN lineages, the persistence of these genes highlights how determinant social traits can be for the survival of some microorganisms. Finally, the considerable number of variants of QS proteins identified among the 69 CPR and DPANN phyla substantially expands our knowledge of prokaryotic communication across the tree of life and suggests that the multiplicity of "dialects" in the microbial world is probably larger than previously appreciated.

KEYWORDS CPR, DPANN, interkingdom signaling, microbial communication, microbiology, quorum sensing

The recent efforts at sequencing the DNA extracted from diverse environments enabled access to genomes of microorganisms with no isolated representatives, which together contributed to expand our vision of life's diversity (1). Most of this expansion is attributable to the discovery of two novel microbial lineages, the candidate phyla radiation (CPR), estimated to account for more than 26% of the currently known bacterial diversity (2), and the archaeal DPANN superphylum (for Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota) (3). Although little is known about these lineages, they already challenge our perspectives on the biology of prokaryotes; CPR and DPANN microorganisms have small to ultrasmall cell sizes (some can pass through 0.22-μm-pore-size filters [4]) and reduced genome sizes, and most of them lack core genes in pathways considered essential in other

prokaryotic lineages, such as the *de novo* synthesis of nucleotides, amino acids or lipids (3, 5). These unusual traits suggest that the majority of these ultrasmall species might depend on other organisms to survive (6) or might even be obligate symbionts (5), a suggestion supported by the few endobiotic (7) and epibiotic (8, 9) relationships uncovered between a CPR or a DPANN and another microorganism(s). Nevertheless, the biological mechanisms underlying these interactions are currently poorly understood.
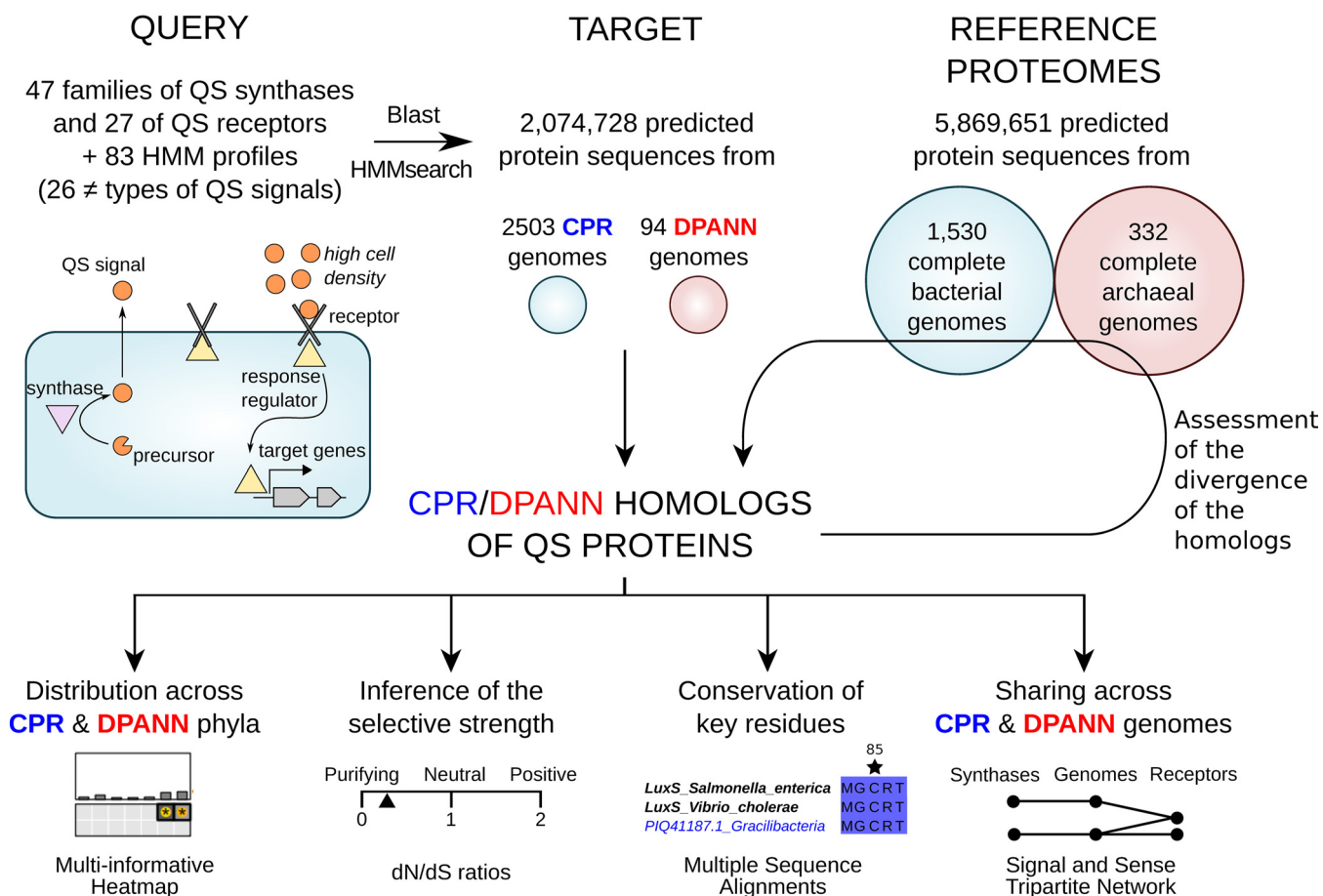
However, two recent studies suggest that interspecies communication via quorum sensing (QS) signals could be a new avenue of investigation to study how CPR and DPANN members may promote and/or maintain their interactions with other species (10, 11). The first study shows that the CPR member "*Candidatus* Saccharibacteria" TM7x HMT-952 interacts with and modulates the QS circuit of its obligate epibiont, *Actinomyces odontolyticus* XH001, which upregulates the biofilm formation pathway in this host and results into an increased stability of the TM7x-XH001 collective within dual-species biofilms (10). The second study hints at the presence of putative QS receptors in two DPANN genomes, "*Candidatus* Woesarchaeota" and "*Candidatus* Micrarchaeota" (11).

A canonical quorum sensing mechanism involves the collective emission of a diffusible signal molecule whose concentration correlates with the cellular density of the emitting microbial population. Upon reaching a threshold concentration (reflecting the quorum of the emitting population), the signal robustly binds to its cognate receptors, within or at the surface of individual cells, and is then transduced to coordinate the emergence of a collective behavior (1–14). Many symbionts, parasites, and even viruses rely on complete QS systems (one or several QS synthase[s] plus a cognate QS receptor) to orchestrate collective behaviors upon reaching a significant population density that either benefits to the host (15–17) or facilitates host invasion (18–20). Hence, QS often dynamically regulates the nature of the interaction between coevolving species. Importantly, the specificity, or secrecy, of a QS signal may range from the intraspecies to the interspecies and even interkingdom levels of recognition and may thus entail relationships of coinfluence within a community of organisms (21–23). It is also worthy of mention that microbiological entities can rely on QS receptors, either uncoupled or coupled with QS synthases, to eavesdrop on exogenous signals, which provides a means for collecting cues about the density or the physiological status of a host and to trigger biological processes accordingly (24–27). Conversely, orphan QS synthases may as well be selected for the influence that the produced signal might exert on the biology of other organisms, to the benefit of the emitting population. Indeed, some QS signals have been reported to inhibit or activate various receptors in nonemitting species and, therefore, to promote different types of host manipulation (28, 29).

Hence, considering that QS signals are key mediators of intra- to interorganismic communication modalities and that the survival of most CPR and DPANN members seems conditioned by their ability to promote and maintain interactions with other species, we designed a computational study to test whether CPR and DPANN might rely on QS genes to achieve their crucial interspecies interactions, as hinted by preliminary studies. In all CPR and DPANN genomes available on the NCBI website, we tested for the presence of homologs of experimentally characterized QS synthases and QS receptors, associated with 26 different types of QS signal molecules, to assess whether and via which "language(s)" CPR and DPANN do communicate.

RESULTS

CPRs and DPANNs encode homologs of reference proteins involved in diverse bacterial communication systems. A total of 423 reference protein sequences, distributed across 74 protein families (47 QS families of synthases and 27 families of receptors) and associated with 26 different types of QS signals (22 bacterial, 1 archaeal, 2 eukaryotic, and 1 viral), were used as a BLAST query data set (see Table S1 in the supplemental material). The 2,074,728 protein sequences predicted from the detected

QUERY

47 families of QS synthases and 27 of QS receptors + 83 HMM profiles (26 ≠ types of QS signals)

Blast
HMMsearch

QS signal

high cell density
receptor

synthase
response regulator
precursor
target genes

TARGET

2,074,728 predicted protein sequences from

2503 **CPR** genomes
94 **DPANN** genomes

REFERENCE PROTEOMES

5,869,651 predicted protein sequences from

1,530 complete bacterial genomes
332 complete archaeal genomes

Assessment of the divergence of the homologs

CPR/DPANN HOMOLOGS OF QS PROTEINS

Distribution across **CPR** & **DPANN** phyla

Multi-informative Heatmap

Inference of the selective strength

Purifying   Neutral   Positive

0   1   2

dN/dS ratios

Conservation of key residues

85
★
LuxS_Salmonella_enterica   MG C R T
LuxS_Vibrio_cholerae   MG C R T
PIQ41187.1_Gracilibacteria MG C R T

Multiple Sequence Alignments

Sharing across **CPR** & **DPANN** genomes

Synthases   Genomes   Receptors

Signal and Sense Tripartite Network

FIG 1 Study design. The first step consists in the identification of variants of QS synthases and receptors in the CPR and DPANN genomes available on the NCBI, using both BLAST queries and hidden Markov models (HMM) corresponding to reference, experimentally characterized QS proteins. Then, using BLAST, these CPR/DPANN homologs are queried against a library of protein sequences corresponding to the complete genomes of *Archaea* and representative *Bacteria* available on the NCBI, in order to assess how divergent their sequences are from their best match among well-studied prokaryotes. Further analyses are undertaken to (i) display the distribution of these homologs of QS proteins across CPR and DPANN phyla, (ii) estimate the selective strength acting on their genes, (iii) visualize the conservation level of key residues in their sequences whenever structural analyses have been conducted on the query QS proteins, and (iv) depict which CPR and DPANN genomes encode highly similar QS synthases (cosignalers) and receptors (cosensors).

coding DNA sequences (CDS) of 2,503 CPR genomes and 94 DPANN genomes constituted the target data set. A homolog was defined as a protein whose sequence identity to a reference QS protein was no less than 25%, with over 75% mutual coverage and an E value below 1e−5 in a BLAST search (30, 31; see also Materials and Methods). These thresholds offer a good trade-off between functional reliability and permissive stringency, according to their application on a set of 76 experimentally validated LuxI synthases retrieved from the Sigmol Database (32) (see Fig. S1 in the supplemental material). A stringent "significance" label was assigned to homologs with an E value of <1e−20. Additionally, the homologs found by BLAST at a first iteration were allowed to serve as queries in a second BLAST search, as long as the remote homologs exhibited 75% mutual coverage with the initial query reference QS protein(s). When available, hidden Markov models (HMM) built from our own multiple-sequence alignments (MSAs) or retrieved from reference databases (Pfam and CATH_Gene3d) were used to check whether the CPR and DPANN sequences harbored the key, conserved residues of the query protein families (Fig. 1).

We detected a total of 5,693 homologs, matching 24 query reference QS protein families and associated with 14 distinct QS signals; 2,003 proteins were found at the first BLAST iteration (24 families/14 signals), including 1,508 hits characterized by an E value of <1e−20 (18 families/12 signals), whereas 1,568 proteins identified by BLAST were
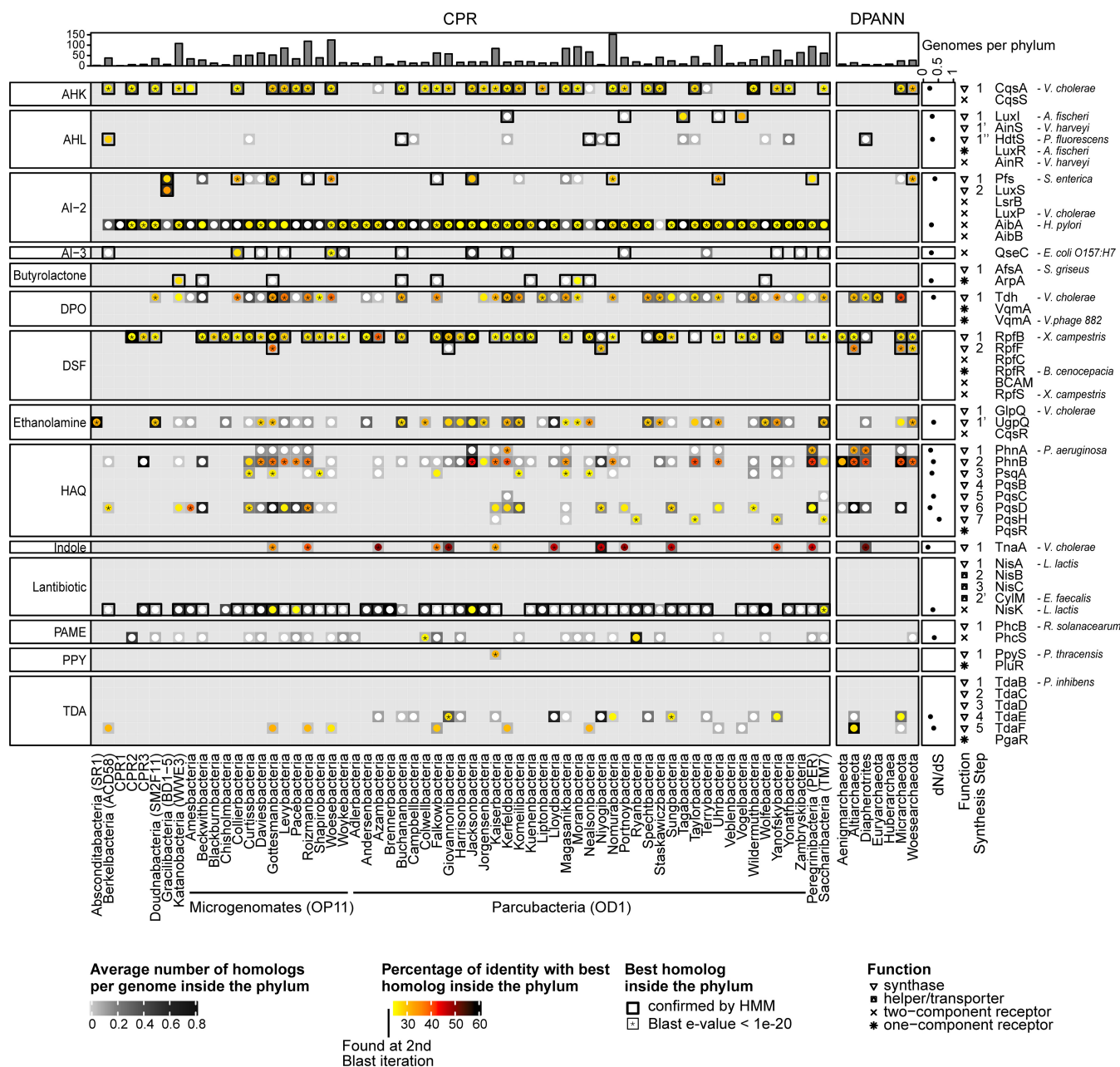
FIG 2 Multi-informative heat map of the QS synthases and receptors detected by BLAST in the CPR and DPANN phyla. Each column represents a phylum, and the histogram at the top displays the number of genomes per phylum. Rows represent query (reference) QS protein families, and the families are grouped by type of QS signals that they either produce or sense (label on the left). The species origin of the representative sequence of each reference protein family is given on the right of each row. The symbol adjacent to the name of each reference protein family indicates whether it corresponds to a QS synthase or a QS receptor (either one- or two-component system). Additionally, a special "synthesis step" label allows distinguishing families of reference synthases that are part of the same biosynthetic route from distinct families of synthases that might nonetheless catalyze the same biosynthetic step. The background in grayscale at each intersection of the heat map indicates the number of homologs of a reference protein family detected in a CPR or DPANN phylum, normalized by the number of genomes in the phylum. The color circle in the foreground displays the percentage of sequence identity between a query protein family and the best homolog detected in a CPR/DPANN phylum. An asterisk (*) indicates that this best homolog gave rise to a BLAST E value of <1e−20, a white circle indicates that it was found at a second BLAST iteration, and a bold rectangle indicates that it was detected by available HMM profiles as well. Finally, the plot on the right panel of the heat map displays the dN/dS ratio of each family of homologs.

also confirmed by available HMMs (10 families/8 signals) (Fig. 2; see also Tables S2 and S3). A total of 13 of the 14 detected QS signals were signals initially discovered in the *Proteobacteria* phylum, consistently with the wide spectrum of specificities reported for proteobacterial QS metabolites, ranging from intraspecies to interkingdom levels and suggesting that these signals might be produced or recognized by the phylogenetically

distant CPR and DPANN lineages. The 14th and last type of QS signal identified in our data set corresponds to the γ-butyrolactones produced by *Actinobacteria*.

Overall, the 24 families of CPR/DPANN homologs of reference QS proteins are distributed across 63 CPR and 6 DPANN phyla; 19 families correspond to homologs of proteins involved in the biosynthesis of QS signals whereas 5 other families correspond to homologs of reference QS receptors (Fig. 2). These results suggest that CPR and DPANN members might emit and sense communication molecules, presumably to influence and monitor the biology of their hosts and neighbors.

The vast majority of the sequences of the CPR and DPANN variants of QS proteins are divergent and under strong selective pressure. In order to better characterize the CPR/DPANN homologs of reference QS proteins, we launched a BLAST search against the protein sequences encoded by the 332 complete genomes of *Archaea* and the 1,530 representative genomes of *Bacteria* available at the NCBI database to identify their best match in reference prokaryotes (Materials and Methods; see also Table S2). For each CPR/DPANN homolog found at a first BLAST iteration, its sequence identities to its query reference QS protein and to its best match within reference prokaryotic proteomes were plotted against each other in a scatterplot (Fig. S2). It appears that the overwhelming majority of the CPR/DPANN homologs of reference QS proteins exhibit no more than 60% identity to any protein predicted from the complete genomes of 1,862 prokaryotes, thereby highlighting how divergent their sequences are from those of well-studied organisms. On another note, comparisons of the functional annotations between each reference QS protein, the best corresponding CPR/DPANN homolog, and its closest match in the reference prokaryotic proteomes did not reveal functional inconsistencies (Table S2). Together, these results reinforce the prediction that CPRs and DPANNs emit and sense communication molecules but also suggest that these molecules may differ slightly from well-characterized QS signals.

Importantly, although the CPR/DPANN QS-related homologs are divergent, we found that selection was inferred to act against changes in their protein sequence. Indeed, the dN/dS ratio (a metric relying on the ratio of the number of nonsynonymous mutations to the number of synonymous mutations along distinct coding sequences) corresponding to each family of homologs was always closer to 0 than to 1 (median = 0.3), which is indicative of strong, purifying selection acting on these genes (Materials and Methods) (the distribution of dN/dS ratios is shown in Fig. 2). This finding highlights the importance of the functions that the CPR/DPANN's homologs of QS proteins actually support.

CPR and DPANN are predicted to produce interspecies QS signals that might influence their hosts/neighbors. Having uncovered the presence of 19 families of QS synthases in the CPR and DPANN lineages, we sought to characterize which types of communication molecules these proteins might contribute to produce. We found that these 19 families of QS synthases are distributed in the biosynthetic pathways of 10 distinct types of QS signals. For 2 of these 10 types of QS signals, namely, HAQ (hydroxy-alkyl quinolines) and TDA (tropodithietic acid), no complete biosynthetic pathways were identified in CPR/DPANN genomes (Fig. 2). Conversely, the biosynthetic routes for the 8 other types of QS signals are found complete in some CPR/DPANN genomes, in which they thus likely support the production of communication molecules.

Specifically, we identified that some CPRs and DPANNs might produce (i) AHK (alpha-hydroxyketones) via homologs of the CqsA, LqsA, and JqsA ketosynthases (33); (ii) AHL (acyl-homoserinelactones) via homologs of the LuxI (13) or the HdtS (34) families; (iii) pyrazines/pyrazinoles such as AI-3 (autoinducer-3) and DPO (3,5-dimethylpyrazin-2-ol) via homologs of the Tdh threonine dehydrogenase (35, 36); (iv) DSF (diffusible small factors) via the RpfB/RpfF pathway (37); (v) ethanolamine via homologs of the UgpQ glycerophosphoryl-diester-phosphodiesterase (38); and (vi) indole via homologs of the TnaA tryptophanase (39). In addition, some members of the CPR lineage might produce (vii) AI-2 (autoinducer-2) via the Pfs/LuxS pathway (22) and

(viii) PPY (photopyrones) via homologs of the PpyS ketosynthase (40) (Fig. 2; see also Table 1).

For every case in which structural studies have been conducted on a reference QS protein family and led to the identification of residues that are essential for the biosynthesis of the QS signal(s), we assessed the level of conservation of these residues in the best homolog found within each CPR/DPANN phylum. Overall, homologs of CqsA, LuxI, HdtS, Tdh, LuxS, and RpfF were found to almost systematically exhibit the residues that are known to be important for the production of their respective type of QS signals (Fig. 3). When such was not the case, a key residue was often found to have been substituted by an amino acid of the same kind (e.g., the Trp-35 of the LuxI family was either conserved or substituted by another aromatic residue [Phe or Tyr]), hinting at alternative biosynthetic modalities in some CPR/DPANN's homologs rather than at disrupted biosynthetic capabilities (Fig. 3).

We also paid attention to the synteny of the coding DNA sequences (CDS) of these homologs of QS synthases, because adjacent QS genes are likely to be functionally linked and to participate together in the QS function. We report that the *pfs* (*S*-adenosylhomocysteine nucleosidase) and *luxS* (*S*-ribosylhomocysteine lyase) genes encoding putative AI-2 synthases are found adjacent to each other in "*Candidatus* Gracilibacteria" from Crystal Geyser, UT. Likewise, the CDS of the homologs of the RpfB (fatty-acid-coenzyme A [CoA]-ligase) and RpfF (enoyl-CoA-hydratase) DSF synthases are found adjacent to each other in three genomes of "*Candidatus* Giovannonibacteria." Furthermore, HMMs allowed us to identify 8 additional similar genomic contexts in "*Candidatus* Giovannonibacteria" and "*Candidatus* Harrisonbacteria" (Table S4). Of note, homologs of *rpfB* and *rpfF* are also present together, yet not adjacent, in 1 "*Candidatus* Gottesmanbacteria," 1 "*Candidatus* Niyogibacteria," 1 "*Candidatus* Altiarchaeota," and 2 "*Candidatus* Woesearchaeota" genomes. Importantly, we also found that 51 CDS of the CqsA homologs in CPRs are in synteny with the CDS of proteins matching the HMM profile of the CqsS α-hydroxyketone receptor kinases, hinting at possible complete AHK-based QS systems in certain CPR genomes. Likewise, the CDS of all the homologs of PpyS as well as of 18 homologs of LuxI were found adjacent to a sensory kinase in some CPR genomes, thereby representing other candidates for complete QS systems (Table S4).
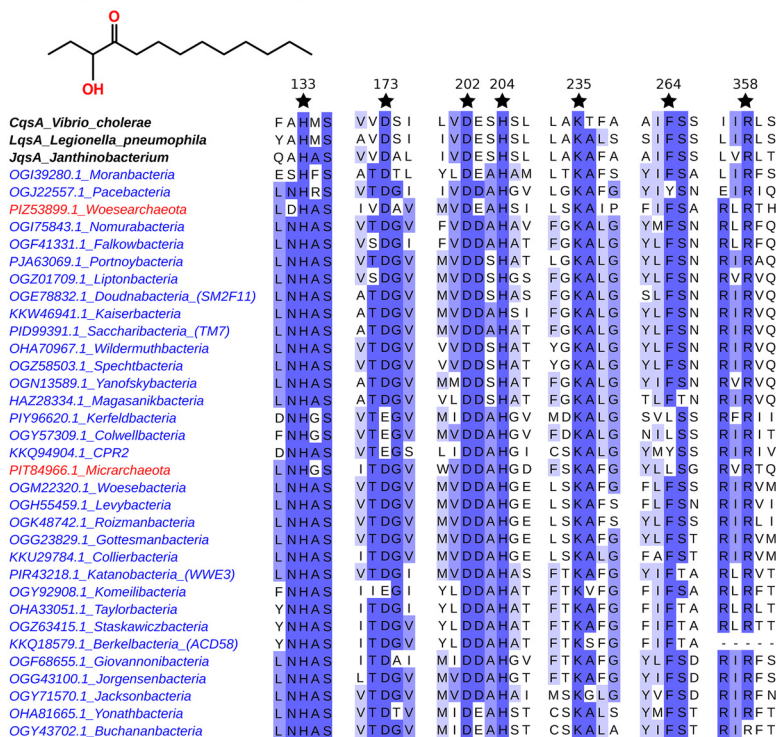
Finally, we looked at the literature to propose roles for these 8 types of communication molecules inferred to be secreted by CPRs and DPANNs. It appears that these families of molecules are known to mediate interspecies signaling (AI-2, the DPO pyrazinol, AHL, indole, DSF) and interkingdom signaling (AHK, AHL, the AI-3 pyrazine, DSF, ethanolamine) (Table 1). The secretion of these types of QS signals hence might help CPRs and DPANNs to promote and/or maintain their crucial interactions with other species. To give a few examples, the AI-2 signal emitted by some "*Candidatus* Gracilibacteria" CPR strains could facilitate their incorporation into multispecies biofilms, in agreement with the acknowledged association between the concentration of the AI-2 signal and the formation of mutualistic biofilms (41). Again, the AHK produced by CPRs and DPANNs might regulate the growth or the motility of eukaryotic hosts, to their own benefit (42–45). The possible AHK-based and AHL-based QS systems of certain CPR endobionts might control when to remain within or when to escape from the host according to the density of CPRs within the host's cytoplasm (28, 42, 46). The biosynthesis of indole might also have been selected in some CPR and DPANN members for the various effects that this molecule exerts on microbial communities, such as modulation of biofilm formation, motility, virulence, antibiotic resistance, etc. (47) (Table 1).

CPR and DPANN are predicted to detect exogenous communication molecules and to collect cues about their hosts/neighbors. In addition to the 19 families of QS synthases, 5 families of CPR/DPANN homologs correspond to reference sensors of QS molecules (Fig. 2). Two of these five families, namely, homologs of QseC and PhcS, can potentially produce false positives for recognition of QS signal(s) because no homologs of QseC exhibit the characteristic glutamate- and aspartate-rich motifs responsible for AI-3 binding (48) and no homologs of PhcS match the N-terminal sensory region of
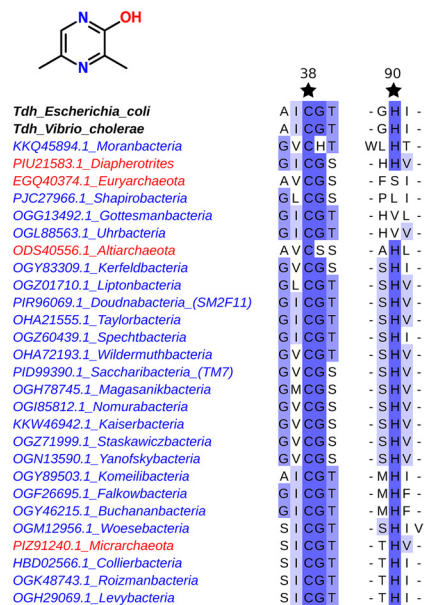
TABLE 1 Proposed roles for the homologs of reference QS proteins found within the CPR and DPANN lineages

Signal	Reference protein family	No. of mapped phyla (BLAST)	No. of homologs (BLAST)	No. of homologs (HMM)	Conserved residue(s)	No. of remarkable CDS adjacencies	Possible roles for the CPR/DPANN homologs
AHK (alpha hydroxy-ketones)	CqsA synthase (Vibrio cholerae)	35 (CPR), 2 (DPANN)	362 (CPR), 11 (DPANN)	364 (CPR), 18 (DPANN)	Fig. 3A (42)	51 CqsA$_{homolog}$/putative sensory kinase (with a CqsS sensory domain)	Switch persistence/escape from the host (as in Legionella pneumophila [28, 42]), modulation of eukaryotic motility (as in Vibrio cholerae and Legionella pneumophila [42–44]), inhibition of eukaryotic growth (as in Janthinobacterium [45]), ...
AHL (acyl homoserine lactones)	LuxI synthase (Aliivibrio fischeri)	5 (CPR)	9 (CPR)	36 (CPR)	Fig. 3D (76)	18 LuxI$_{homolog}$/putative sensory kinase	Persistence within the host (as in Brucella melitensis [46]), interspecies to interkingdom signaling (27), induction of multispecies biofilm formation (78), activation or inhibition of QS receptors in other prokaryotes, ...
	HdtS synthase (Pseudomonas fluorescens)	12 (CPR), 1 (DPANN)	23 (CPR), 1 (DPANN)	83 (CPR), 3 (DPANN)	Fig. 3F (34)		
AI-2 (auto-inducer-2)	Pfs synthase (step 1) (Saccharomyces enterica)	16 (CPR), 1 (DPANN)	84 (CPR), 2 (DPANN)	56 (CPR), 1 (DPANN)	Fig. 3C (77)	4 Pfs$_{homolog}$/LuxS$_{homolog}$	Signaling presence to other prokaryotes (22), facilitation of the incorporation into multispecies biofilm (41), ...
	LuxS synthase (step 2) (Saccharomyces enterica)	1 (CPR)	4 (CPR)	6 (CPR)			
	AibA receptor (Helicobacter pylori)	62 (CPR)	1,876 (CPR)	No reference HMMs			Estimation of the cellular density of the bacterial community (22), ...
Pyrazines/pyrazinols	Tdh synthase (Vibrio cholerae)	41 (CPR), 4 (DPANN)	326 (CPR), 21 (DPANN)	No reference HMMs	Fig. 3B (74)		Interspecies (25) to interkingdom (36) signaling, ...
DSF (diffusible small factors)	RpfB synthase (step 1) (Xanthomonas campestris)	40 (CPR), 5 (DPANN)	433 (CPR), 22 (DPANN)	619 (CPR), 34 (DPANN)		11 RpfB$_{homolog}$/RpfF$_{homolog}$	Interspecies to interkingdom signaling, regulation of motility, biofilm formation, iron uptake, virulence in other prokaryotes, elicitation of the innate immunity of plants, induction of the stringent response, and siderophore production in other bacteria (as in Burkholderia cenocepacia) (37), ...
	RpfF synthase (step 2) (Xanthomonas campestris)	3 (CPR), 3 (DPANN)	6 (CPR), 7 (DPANN)	67 (CPR), 8 (DPANN)	Fig. 3E (77)		
Ethanolamine	UgpQ synthase (Vibrio cholerae)	37 (CPR), 7 (DPANN)	247 (CPR), 7 (DPANN)	No reference HMMs			Interkingdom signaling (38), ...
Indole	TnaA synthase (Vibrio cholerae)	12 (CPR), 1 (DPANN)	61 (CPR), 2 (DPANN)	No reference HMMs			Inhibition of QS receptors in other prokaryotes (79), modulation of biofilm formation, motility and antibiotic resistance within microbial communities (47), ...
PPY (photopyrones)	PpyS synthase (Photorhabdus thracensis)	1 (CPR)	3 (CPR)	No reference HMMs		3 PpyS$_{homolog}$/putative sensory kinase	Cell clumping (as in Photorhabdus thracensis [40]), ...
γ-Butyro-lactones	ArpA receptor (Streptomyces griseus)	10 (CPR)	17 (CPR)	63 (CPR)	Fig. 3G (49)		Eavesdropping on host QS molecules (53), ...
Lantibiotics	NisK receptor (Lactococcus lactis)	51 (CPR)	454 (CPR)	71 (CPR), 2 (DPANN)			Efflux/resistance to exogenous lantibiotics (52), QS (50), ...

A α-hydroxyketones (AHK)

	133	173	202 204	235	264	358
CqsA_Vibrio_cholerae	FAHMS	VVDSI	LVDESHSL	LAKTFA	AIFSS	IIRLS
LqsA_Legionella_pneumophila	YAHMS	AVDSI	IVDESHSL	LAKALS	SIFSS	LIRFS
JqsA_Janthinobacterium	QAHAS	VVDAL	IVDESHSL	LAKAFA	AIFSS	LVRLT
OGI39280.1_Moranbacteria	ESHFS	ATDTL	YLDEAHAM	LTKAFS	YIFSA	LIRFS
OGJ22557.1_Pacebacteria	LNHRS	VTDGI	IVDDAHGV	LGKAFG	YIYSN	EIRIQ
PIZ53899.1_Woesearchaeota	LDHAS	IVDAV	MVDEAHSI	LSKAIP	FIFSA	RLRTH
OGI75843.1_Nomurabacteria	LNHAS	VTDGV	FVDDAHAV	FGKALG	YMFSN	RLRFQ
OGF41331.1_Falkowbacteria	LNHAS	VSDGI	FVDDAHAT	FGKALG	YLFSN	RLRFQ
PJA63069.1_Portnoybacteria	LNHAS	VTDGV	MVDDSHAT	LGKALG	YLFSN	RIRAQ
OGZ01709.1_Liptonbacteria	LNHAS	VSDGV	MVDDSHGS	FGKALG	YLFSN	RVRVQ
OGE78832.1_Doudnabacteria_(SM2F11)	LNHAS	ATDGV	MVDDSHAT	FGKALG	SLFSN	RIRVQ
KKW46941.1_Kaiserbacteria	LNHAS	ATDGV	MVDDAHSI	FGKALG	YLFSN	RIRVQ
PID99391.1_Saccharibacteria_(TM7)	LNHAS	ATDGV	MVDDAHAT	FGKALG	YLFSN	RIRVQ
OHA70967.1_Wildermuthbacteria	LNHAS	VTDGV	VVDDSHAT	YGKALG	YLFSN	RIRVQ
OGZ58503.1_Spechtbacteria	LNHAS	VTDGV	VVDDSHAT	YGKALG	YLFSN	RIRVQ
OGN13589.1_Yanofskybacteria	LNHAS	ATDGV	MMDDSHAT	FGKALG	YIFSN	RVRVQ
HAZ28334.1_Magasanikbacteria	LNHAS	ATDGV	VLDDSHAT	FGKALG	YLFSN	RIRVQ
PIY96620.1_Kerfeldbacteria	DNHGS	VTEGV	MIDDAHGV	MDKALG	SVLSS	RFRII
OGY57309.1_Colwellbacteria	FNHGS	VTEGV	MVDDAHGV	FDKALG	NILSS	RIRIT
KKQ94904.1_CPR2	DNHAS	VTEGS	LIDDAHGI	CSKALG	YMYSS	RIRIV
PIT84966.1_Micrarchaeota	LNHGS	IITDGV	WVDDAHGD	FSKAFG	YLLSG	RVRTQ
OGM22320.1_Woesebacteria	LNHAS	VTDGV	MVDDAHGE	LSKAFG	FLFSS	RIRVM
OGH55459.1_Levybacteria	LNHAS	VTDGV	MVDDAHGE	LSKAFS	FLFSN	RIRVI
OGK48742.1_Roizmanbacteria	LNHAS	VTDGV	MVDDAHGE	LSKAFS	YLFSS	RIRLI
OGG23829.1_Gottesmanbacteria	LNHAS	VTDGV	MVDDAHGE	LSKAFG	YLFST	RIRVM
KKU29784.1_Collierbacteria	LNHAS	IITDGV	MVDDAHGE	LSKALG	FAFST	RIRVM
PIR43218.1_Katanobacteria_(WWE3)	LNHAS	VTDGI	MVDDAHAS	FTKAFG	YIFTA	RLRVT
OGY92908.1_Komeilibacteria	FNHAS	IIEGI	YLDDAHAT	FTKVFG	FIFSA	RLRFT
OHA33051.1_Taylorbacteria	YNHAS	IITDGV	YLDDAHAT	FTKAFG	FIFTA	RLRLT
OGZ63415.1_Staskawiczbacteria	YNHAS	IITDGV	YLDDAHAT	FTKAFG	FIFTA	RLRTT
KKQ18579.1_Berkelbacteria_(ACD58)	YNHAS	IITDGV	YLDDAHAT	FTKSFG	FIFTA	------
OGF68655.1_Giovannonibacteria	LNHAS	IITDAI	MIDDAHGV	FTKAFG	YLFSD	RIRFS
OGG43100.1_Jorgensenbacteria	LNHAS	LTDGV	MVDDAHGT	FTKAFG	YIFSD	RIRFS
OGY71570.1_Jacksonbacteria	LNHAS	VTDGV	MVDDAHAI	MSKGLG	YVFSD	RIRFN
OHA81665.1_Yonathbacteria	LNHAS	VTDTV	MIDEAHST	CSKALS	YMFST	RIRFT
OGY43702.1_Buchananbacteria	LNHAS	VTDGV	MIDEAHST	CSKALA	YIFST	RIRFT

B Pyrazinols (DPO)

	38	90
Tdh_Escherichia_coli	AICGT	-GHI-
Tdh_Vibrio_cholerae	AICGT	-GHI-
KKQ45894.1_Moranbacteria	GVCHT	WLHT-
PIU21583.1_Diapherotrites	GICGS	-HHV-
EGQ40374.1_Euryarchaeota	AVCGS	-FSI-
PJC27966.1_Shapirobacteria	GLCGS	-PLI-
OGG13492.1_Gottesmanbacteria	GICGT	-HVL-
OGL88563.1_Uhrbacteria	GICGT	-HVV-
ODS40556.1_Altiarchaeota	AVCGS	-AHL-
OGY83309.1_Kerfeldbacteria	GVCGS	-SHI-
OGZ01710.1_Liptonbacteria	GLCGT	-SHV-
PIR96069.1_Doudnabacteria_(SM2F11)	GICGT	-SHV-
OHA21555.1_Taylorbacteria	GICGT	-SHV-
OGZ60439.1_Spechtbacteria	GICGT	-SHI-
OHA72193.1_Wildermuthbacteria	GVCGT	-SHV-
PID99390.1_Saccharibacteria_(TM7)	GVCGS	-SHV-
OGH78745.1_Magasanikbacteria	GMCGS	-SHV-
OGI85812.1_Nomurabacteria	GVCGS	-SHV-
KKW46942.1_Kaiserbacteria	GVCGS	-SHV-
OGZ71999.1_Staskawiczbacteria	GVCGS	-SHV-
OGN13590.1_Yanofskybacteria	GVCGS	-SHV-
OGY89503.1_Komeilibacteria	AICGT	-MHI-
OGF26695.1_Falkowbacteria	GICGT	-MHF-
OGY46215.1_Buchananbacteria	GICGT	-MHF-
OGM12956.1_Woesebacteria	SICGT	-SHV-
PIZ91240.1_Micrarchaeota	SICGT	-THV-
HBD02566.1_Collierbacteria	SICGT	-THI-
OGK48743.1_Roizmanbacteria	SICGT	-THI-
OGH29069.1_Levybacteria	SICGT	-THI-

C Autoinducer-2 (AI-2)

	57	85
LuxS_Salmonella_enterica	TLEHL	MGCRT
LuxS_Vibrio_cholerae	TLEHL	MGCRT
PIQ41187.1_Gracilibacteria	AIEHL	MGCRT

D Acyl-homoserine-lactones (AHL)

	25	29	35	46	49	70	101	104
LuxI_Aliivibrio_fischeri	SLRYQVF	KRRLEW--		ESDEYDNS		CWRLL	IVELSRFA	
CviI_Chromobacterium_violaceum	AFRHKIF	RENLRW--		DRDEYDAI		SVRFT	SAEISRFA	
PpuI_Pseudomonas_putida	KLRARIF	KVKKMW--		EIDGYDAL		CWRIL	IWELSRFA	
CmrI_Pseudomonas_sp_CMR12a	SYRYKVF	VEKLGW--		ELDEFDRD		TARLL	VWELSRFA	
LuxI_Nitrobacter_winogradskyi	RLRHNVF	VDEMGWD-		EIDQFDDG		YQRML	IWEWTRYC	
LuxI_Sinorhizobium_fredii_GR64	QLRARVF	SERLGW--		ESDGFDSL		CARLL	MVESSRFC	
PIP60426.1_Uhrbacteria	ALRYREY	CRIGSFQ-		EKDGYDET		SIRVC	TCELSRLV	
OGY84289.1_Kerfeldbacteria	KIRYRVY	CEEYNYIS		EKDEWDSH		TVRMI	SVNLSKAA	
PIR44755.1_Vogelbacteria	KLRHEVF	VKKLAW--		ERDSYDEY		TARFT	AVELSRLG	
OGZ33548.1_Portnoybacteria	KLRKVIY	SDLLKW--		ETDEYDKY		YCRLI	TAEISRLT	
PIS13373.1_Tagabacteria	QLRRKVF	VKELGW--		EKDIYDTH		YVRLI	AAEISRLV	

E Diffusible small factors (DSF)

	141	161
RpfF_Xanthomonas_campestris	GFEAA	LPEVL
RpfF_Xylella_fastidiosa	GFEAA	FPEVL
RpfF_Stenotrophomonas_maltophilia	GFEAA	LPEVL
BCAM0581_Burkholderia_cenocepacia	GFEAA	FPEIA
DspI_Pseudomonas_aeruginosa	GLECA	LPEAA
OGF62448.1_Giovannibacteria	GLELA	FPECK
PIR69299.1_Niyogibacteria	GLELA	FPEVT
PIO05160.1_Woesearchaeota	GNELS	QPETK
OJI06719.1_Micrarchaeota	GNELA	QPEVK
ODS40711.1_Altiarchaeota	GCELA	LPEVT
KKR30078.1_Gottesmanbacteria	GTEIA	LPEVG

F Acyl-homoserine-lactones (AHL)

	67	81 82 84	126	153 156	184
HdtS_Pseudomonas_fluorescens	VKGAE	QSN-HQSTW	IDRDN	LIFPEGTR	VLPIA
Act_Acidithiobacillus_ferrooxidans	VSGLA	VAN-HQSAL	IDRNG	VIFPEGTR	LLPIA
Dqsl-2_Deinococcus_radiodurans	VTG--	APN-HQSWW	FLRRV	VLFPEGAL	LIPVG
Dqsl-1_Deinococcus_radiodurans	VHGRE	AAN-HQSAL	VDRES	GIFPQGTR	ILPAG
Jqsl-2_Candidatus_Jettenia_caeni	ILGID	ASN-HLSYI	IQRND	VVFPEGTR	VIPMA
Jqsl-1_Candidatus_Jettenia_caeni	VKGIE	VSN-HSSMY	VNRQN	VIFPEGRR	VVPTL
TSC92749.1_Berkelbacteria_(ACD58)	TSGLE	ISN-HQSMM	INRQK	LVFAEGTR	IVPVS
PIS05939.1_Buchananbacteria	RLIIN	VPN-HESLF	IDRKN	IIFPEGTR	IVPVG
OHA13958.1_Taylorbacteria	VYGLE	ISN-HKHML	VIRGE	LIHPEGRV	ILPIA
OHA13958.1_Tagabacteria	VYGLE	ISN-HKHML	VIRGE	LIHPEGRV	ILPIA
PIS40439.1_Kerfeldbacteria	VDGLE	VAN-HVGLQ	LLT--	LIYPEGKV	VIPIG
PIU22553.1_Diapherotrites	VKGLE	AAN-HQSHL	LNRKS	ILYPEGTR	ILPVG
PIW92454.1_Nealsonbacteria	IKGKE	VSN-HQSDL	LNRKD	IIYPEGRR	ILPTG
OGD67499.1_Campbellbacteria	VKGLK	AVN-HSSEL	AYTGK	CIFPEGRK	VVPVR
OIO30597.1_Nomurabacteria	VQGRG	AVN-HLSDF	VLIGT	GIFPEGRR	VVPVA
HBB44449.1_Yonathbacteria	VLGRE	AVN-HSSQF	VTVGT	CIFPEGGK	VLPVV
OGZ32756.1_Niyogibacteria	IKGGE	AAN-HASYL	VVQGI	GVFPEGKI	VLPVN
KKR55681.1_Curtissbacteria	VKGLE	AAAIHANYI	LEKDV	GVFVEGRL	VLPVG
KKR89021.1_Wolfebacteria	VKGLE	AAAIHANYI	LEKDV	GVFVEGRL	VLPVG
OHA23948.1_Taylorbacteria	IRGRH	ISN-HIAFY	IVPGQ	VIYPEGEM	VLPLS

G γ-butyrolactones

	119
ArpA_Streptomyces_griseus	GDWID
JadR2_Streptomyces_venezuelae	QEWQS
BarB_Streptomyces_virginiae	EHWRQ
BarA_Streptomyces_virginae	GEWGD
ScbR_Streptomyces_coelicolor	RRWHE
FarA_Streptomyces_sp_FRI-5	RNWVD
SngR_Streptomyces_natalensis	TPWIA
SrrA_Streptomyces_rochei	PTWIA
TarA_Streptomyces_tendae	PAWSR
OGC61252.1_Katanobacteria_(WWE3)	EMWTG
OGY55708.1_Buchananbacteria	TEWAE
OGG01799.1_Gottesmanbacteria	AVLGG
PIT86894.1_Magasanikbacteria	KQIAD
OGF41499.1_Falkowbacteria	NQLYK
PIS17238.1_Nealsonbacteria	KQVVN
PIU98784.1_Wolfebacteria	ERVAN
HBT45860.1_Moranbacteria	EQVAD
PKL72708.1_Kuenenbacteria	EQVAN
OIN88169.1_Beckwithbacteria	QNLIK

FIG 3 Partial multiple-sequence alignments of CPR/DPANN homologs with reference QS proteins to study the conservation of key residues. Wherever important residues (black stars) for the QS function have been identified in the sequences of reference QS proteins (names in bold), their level of

PhcS that binds QS signals of methyl esters such as palmitic acid methyl ester (PAME) or myristic acid methyl ester (MAME). Conversely, the 3 other families, detected only in the CPR lineage, are likely to support recognition of QS signal(s).

Specifically, CPR members might sense (i) the γ-butyrolactones produced by potential actinobacterial hosts via homologs of ArpA (49); (ii) endogenous or exogenous peptide lantibiotics such as nisin, subtilin, and mersacidin via homologs of the NisK kinase (50); and (iii) the AI-2 QS molecule via homologs of the AibA QS receptor, recently identified in *Helicobacter pylori* (51) (Fig. 2; see also Table 1).

ArpA is a transcription factor with a DNA binding domain (C-terminal HTH motif) that becomes activated when its N-terminal sensory domain binds with γ-butyrolactone, notably via a key tryptophan residue at position 119 (49). All 17 homologs of ArpA harbor a canonical HTH domain, but the Trp119 residue is present only in the "*Candidatus* Katanobacteria" (WW3) and "*Candidatus* Bachananbacteria" phyla (Fig. 3G). Hence, besides some members of the "*Candidatus* Katanobacteria" and "*Candidatus* Bachananbacteria" phyla which are highly likely to sense γ-butyrolactones, the other ArpA-encoding genomes of the CPR radiation might either bind with γ-butyrolactones despite the absence of Trp119 or bind with other types of communication molecules.

The examination of the genomic context of the CDS of the only homolog of NisK with an E value of $<1e-20$ (protein RYC73835.1 of a "*Candidatus* Saccharibacteria" strain) reveals that it forms a two-component system with an adjacent response regulator, flanked on both sides by tandem ABC transporters from the BceA-BceB family. This genomic context is the signature of a resistance/efflux system for exogenous bacteriocins (52).

To sum up, the CPR homologs of the ArpA family are predicted to support eavesdropping on host-produced, likely actinobacterial QS molecules, presumably as a means to collect cues about the density status of hosts and to trigger pathways accordingly. This prediction is consistent with the relation of epibiosis between a CPR member (the TM7x "*Candidatus* Saccharibacteria") and an actinobacterium (*Actinomyces odontolyticus*) as previously reported (8). As a matter of fact, our HMM of the ArpA protein family actually matches a protein encoded by TM7 oral taxon 349 (protein TWP21250.1, E value = $2e-18$, Table S3). Homologs of NisK might enable some CPRs to sense and pump out exogenous peptide lantibiotics, but they could as well be specific to endogenous lantibiotics and mediate intraspecies QS as in *Lactococcus lactis* or *Bacillus subtilis* (48). Last but not least, the 1,876 CPR homologs of AibA were detected by BLAST in all CPR phyla, except in the "*Candidatus* Absconditabacteria" phylum (Fig. 2; see also Table S2). Considering that the AI-2 QS signal is produced by bacteria from many different phyla, the functional validation of all the AibA homologs as AI-2 sensors would imply that almost all CPR members are able to estimate the cellular density of their bacterial community.

Finally, in addition to the homologs of ArpA, NisK, and AibA identified by BLAST, we found interesting the match of 28 CPR proteins (distributed across the "*Candidatus* Kerfeldbacteria" and "*Candidatus* Wolfebacteria" CPR phyla) with the HMM profile built from the multiple-sequence alignment of the PAS4 domain of 35 receptors of the VqmA family (DPO sensing) (Table S3). Indeed, PAS4 domains are assumed to bind with hormone-like molecules, notably from eukaryotic hosts, and therefore to be involved in both QS and host sensing (40).

Signal and sense tripartite network. To summarize the results of our analysis, we introduce "Signal and Sense Tripartite Networks" representing a novel framework that

FIG 3 Legend (Continued)

conservation and the representative homologs of each CPR phylum (names in blue) and DPANN phylum (names in red) are displayed. (A) Ketosynthases (CqsA family): synthesis of α-hydroxyketones (structural study [42]). (B) Threonine dehydrogenases (Tdh family): synthesis of pyrazines and pyrazinols (structural study [74]). (C) S-Ribosylhomocysteine lyases (LuxS family): synthesis of AI-2 (structural study [75]). (D) AHL synthases (LuxI family): synthesis of acyl-homoserine lactones (structural study [76]). When not conserved, the aromatic F-29 and W-35 residues are substituted by other aromatic amino acids. (E) Enoyl-CoA-hydratases (RpfF family): synthesis of diffusible small factors (structural study [77]). (F) AHL synthases (HdtS family): synthesis of acyl-homoserine lactones (structural study [34]). (G) One-component receptors (ArpA family): sensing of γ-butyrolactones (structural study [49]).

describes, for each genome (central nodes), which type of QS signal(s) it is predicted to produce (left nodes), and which type of QS signal(s) it is predicted to sense (right nodes). Specifically, we partitioned each relevant family of CPR/DPANN homologs of either QS synthases or QS receptors into subfamilies using a clustering threshold at 90% sequence identity (Table S5). Hence, whenever two nodes of the central layer "Genome" are connected to the same node of the left layer "Synthase," it means that these two genomes are likely to produce the same QS molecule and thus to be cosignalers and to speak the same "dialect" of a broader "language," each "language" corresponding to a family of QS molecules. Likewise, genomes connected to the same node of the right layer "Receptor" highlight likely cosensors. This depiction allows appreciating which and how many QS signals the different CPR and DPANN members might secrete and/or eavesdrop on at the genome level. The "Signal and Sense Tripartite Network" of the DPANN genomes is given in Fig. 4, whereas the network of the CPR genomes, much bigger, is given in Fig. S3.

DISCUSSION

Our survey results show a rich repertoire of homologs of QS proteins encoded by 2,205 CPR and 66 DPANN genomes. We displayed the distribution of these proteins across the different CPR and DPANN phyla (Fig. 2) and found that they are associated with 14 different QS signals, of 25 tested. Since the CPR and DPANN lineages account for a substantial amount of phylogenetic diversity, the functional characterization of these QS proteins could significantly expand our knowledge of microbial communication across the tree of life. Specifically, QS signals have been characterized so far, to the best of our knowledge, in at least 10 microbial phyla, distributed across *Bacteria* (14, 22, 53, 54), *Archaea* (55, 56), *Eukarya* (57), and viruses (17), and our study on the CPR and DPANN lineages suggests that this spectrum could potentially expand to up to 69 additional prokaryotic phyla.

Interestingly, the majority of the sequences of the CPR/DPANN variants of reference QS proteins are divergent from those of proteins encoded by reference genomes (see Fig. S2 and Table S2 in the supplemental material). In this regard, heterologous expression of CPR/DPANN putative QS synthases in model prokaryotes would be particularly interesting because it could lead to the discovery of new communication molecules. Given that they may act as antagonists of known QS receptors in pathogens, this could lead to new anti-infective strategies.

From an evolutionary viewpoint, considering that CPR and DPANN genomes are suggested to have undergone genome reduction, the persistence of QS genes which would allow them to signal their presence, influence their neighbors, and collect social cues might underline the prime importance of the role played by social traits in their survival. Our predictions could hence convey the strong message that social traits can sometimes be more critical for the survival of certain species than many traits that are considered essential in most microbes and yet are absent in most CPR and DPANN members, such as, for example, the *de novo* synthesis of nucleotides, lipids, or amino acids.

Our results also pave the way to functional studies that could help decipher the underlying functions supporting the inferred symbiotic/parasitic lifestyles of CPR and DPANN members. Indeed, our survey results suggest that many partial or complete QS systems in CPR/DPANN genomes support interorganism and cross-kingdom communication (Table 1), which could be determinant in the promotion and/or the maintenance of the critical interactions of CPR and DPANN with other species. Specifically, depending on the different combinations of QS components (QS synthase[s] and/or QS receptor) identified in the CPR and DPANN genomes, we predict three types of communication modalities for these microorganisms: quorum sensing, presence signaling (or manipulation), and eavesdropping on exogenous signals.

Density-dependent mechanisms of QS are typically supported by complete QS systems (QS synthase[s] plus QS receptor). In this respect, we identified genes of QS synthases in CPR genomes that are adjacent to transmembrane kinases harboring a
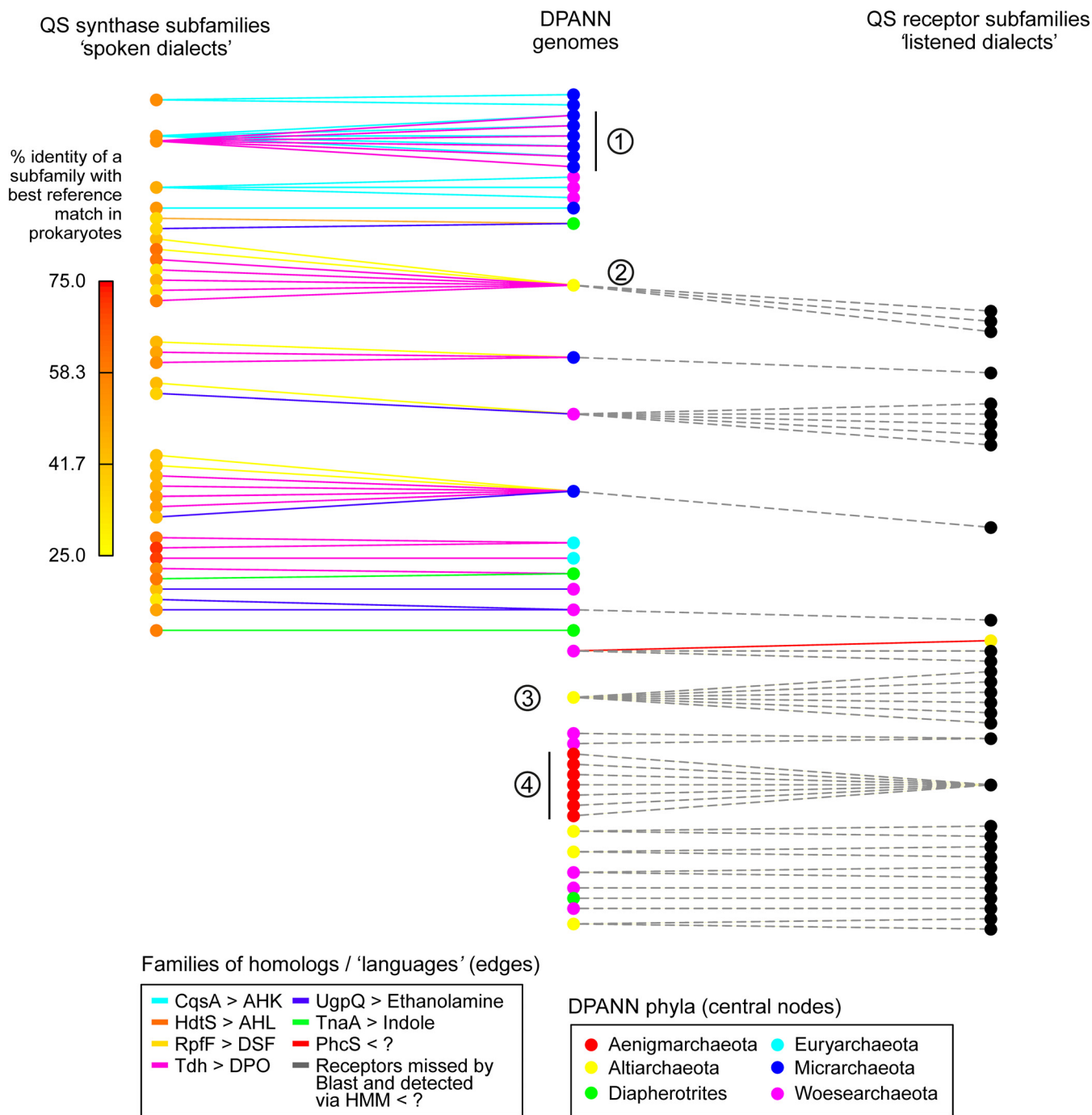
FIG 4 Signal and sense tripartite network in the DPANN lineage. This representation allows appreciating which and how many QS signals the different DPANN members might secrete and/or eavesdrop on at the genome level. The central layer of nodes corresponds to DPANN genomes. The left layer of nodes corresponds to subfamilies of homologs of QS synthases (family of homologs subpartitioned at 90% sequence identity), and the right layer of nodes to subfamilies of homologs of QS receptors. Hence, wherever two "Genome" nodes are connected to a left node, it means that they encode highly similar variants of QS synthases and are thus likely cosignalers. Likewise, wherever two "Genome" nodes are connected to a right node, it means that they are likely cosensors. The color of the edges allows identifying the initial family from which a subfamily has been further defined. Hence, left and right nodes could depict "spoken" and "listened" "dialects," respectively, whereas the color of the edges could depict the "languages" from which these dialects derive. Of note, dashed edges correspond to putative QS receptors identified by HMM and missed by BLAST. The color of the central nodes indicates the phylum of each genome. The color of the subfamilies identified by BLAST indicates their respective average percentages of identity with the best matches in reference prokaryotes; yellow nodes thus highlight sequences that are more divergent than those highlighted by red nodes. Examples: (1) genomes predicted to "speak" the same "dialects" of the AHK and DPO languages; (2) genome predicted to "speak" several dialects of the DPO and DSF languages and to sense several signals; (3) genome predicted to "listen to" several dialects; (4) genomes predicted to "listen to" the same dialect. The tripartite network of the CPRs is displayed in Fig. S3.

binding domain for the QS signal predicted to be synthesized (Table S4). This genomic context is typical of a complete QS system, since QS synthases and their cognate QS receptors are often colocalized in microbial genomes. This being said, the selection of a complete QS system of genes (synthase[s] plus receptor) to regulate biological processes in a cell density-dependent manner presupposes that the microbial population expressing the QS system occasionally encounters a high cellular density. Consistently, the prediction of complete QS systems in CPR genomes is reinforced by several studies reporting that some CPR species are found to be abundant under some conditions. For example, "*Candidatus* Sonnebornia yantaiensis" is usually found at low abundance in freshwater but is sometimes present by the thousands in the cytoplasm of paramecia (7). Again, it has been reported that the relative abundance of "*Candidatus* Saccharibacteria" is about ~1% in healthy human oral cavities but can increase to up to 21% of the whole microbial community in cases of periodontal diseases (58, 59). Together, these findings suggest that the putative QS mechanisms of certain CPRs might be triggered when they reach a critical density within the cytoplasm of their hosts or within multispecies biofilms.

Apart from these predicted complete QS genetic systems, we also identified variants of QS synthases and QS receptors with no detected cognate QS component in some genomes. Although "not detected" does not necessarily imply "not present," we further discuss different scenarios that could explain the selection of uncoupled QS synthases and uncoupled QS receptors. Orphan QS synthases could be selected for the advantageous influence that the produced QS signal(s) would exert on the behavior of other species, to the benefit of the emitting population. For instance, the QS signals that they produce could act as antagonists of QS receptors upregulating microbial defense mechanisms in other species. Also, since most of the homologs of QS synthases in the genomes of CPR and DPANN are associated with interspecies and interkingdom signaling, the QS signal(s) that they would produce could mediate a wide range of influences within microbial communities; α-hydroxyketones could modulate the motility of eukaryotic cells, AI-2 could coerce other cells to produce multispecies biofilms, indole could modulate many biological processes inside a community of microorganisms (motility, virulence, antibiotic resistance, etc.). The converse case, namely, the selection of orphan QS receptors in CPR genomes (notably for AI-2 and γ-butyrolactones), is not rare in nature (24, 25, 27) and is explained by the fact that these receptors can enable eavesdropping on the density status of hosts or neighbors.

A recent functional study provided insights into the way in which CPR could make use of such orphan receptors to regulate their interactions with other microorganisms (10). That study showed that *Actinomyces odontolyticus* XH001, the obligate epibiont of the "*Candidatus* Saccharibacteria" TM7x HMT-952 CPR, relies collectively on the emission (via the LuxS$_{homolog}$ QS synthase) and perception (via the LsrB$_{homolog}$ transmembrane QS receptor) of the AI-2 QS signal to orchestrate biofilm formation in a density-dependent manner. Those authors further report three fascinating results: (i) upon association with the TM7x CPR, the most highly induced gene in XH001 was the *lsrB* QS receptor; (ii) upon deletion of the *luxS* QS synthase in XH001, TM7x cells no longer induced significant upregulation of *lsrB;* (iii) wild-type XH001 cocultured with TM7x displayed significantly increased biofilm formation compared to XH001 alone and to XH001$_{\Delta luxS}$-TM7x and XH001$_{\Delta lsrB}$-TM7x cocultures. A parsimonious hypothesis to account for these observations could be that upon detection of the host-produced AI-2 QS signal, perhaps via a homolog of AibA (Fig. 2), TM7x cells would perceive when the XH001 population reached a sufficient cellular density compatible with interspecies biofilm formation. Thereupon, they would coerce XH001 cells, through *lsrB* upregulation, to be themselves more sensitive to the AI-2 signal, misleading these hosts to overestimate their quorum and to subsequently precipitate the induction of the biofilm formation pathway to the benefit of the stability of the TM7x-XH001 collective. This hypothesis is consistent with the previously reported observations that the AI-2 QS signal promotes interspecies interactions between periodontopathogens via enhanced expression of biofilm molecules (60) and that the relative abundance of "*Candidatus*

Saccharibacteria" in the oral microbiome is about 1% in healthy human cavities but can increase to 21% in case of periodontal diseases (58, 59).

To conclude, our *in silico* analysis opens the possibility of exciting perspectives in CPR/DPANN biology and in prokaryotic QS research in the foreseeable future. In the long term, deciphering QS processes in the CPR and DPANN novel lineages will be an important milestone toward understanding how these microorganisms achieve their critical interactions with other species. The high number of variants of QS synthases identified in CPR and DPANN genomes also offers a promising reservoir for the discovery of new molecules of communication and suggests that the multiplicity of "dialects" in the microbial world is probably larger than previously appreciated.

MATERIALS AND METHODS

Construction of a reference database of sequences and HMMs of QS synthases and receptors. We carefully mined the literature related to QS to establish a list of QS systems. We notably relied on the Sigmol (32) and the Quorum Peps (61) databases of QS metabolites and QS peptides, respectively. This reference list of experimentally characterized QS proteins is available as a tabular file (see sheet 1 of Table S1 in the supplemental material) and summarizes for each protein its function and its NCBI or Uniprot identifier (ID) as well as the QS signal with which it is associated. When available, HMMs of the domains that are specific for QS proteins were retrieved from Pfam (62) and Cath-Gene3d (63) and were further assembled in a library using the hmmpress tool from HMMER suite version 3.2.1 (64) (Table S1, sheet 2). We also constructed our own HMM profiles from the query protein families that comprised more than 5 validated QS proteins (e.g., LuxS, LuxI, etc.) and mined the literature of structural studies related to well-studied QS families in order to construct HMM profiles of domains that are specific to QS proteins (AHL-binding domain of LuxR, periplasmic domain of QseC, etc.) (Table S1, sheet 2). Each HMM profile was built as follows: the multiple-sequence alignment of either full sequences or domains of QS proteins was built using MUSCLE version 3.8.31 (65) and was given as input to the hmmbuild tool (the HMMER suite version 3.2.1; 64). All resulting HMMs were compiled together in a library using hmmpress.

Retrieval of the protein sequences of the CPR and DPANN genomes. All the protein sequences of the CPR and DPANN genomes were downloaded from the NCBI assembly database as of 17 December 2019 (66). At the time of writing, these CPR genomes/proteomes were able to be retrieved from the lineages corresponding to the following taxonomic IDs: txid74243, txid95818, txid221235, txid363464, txid422282, txid1618330, txid1618338, txid1618339, txid1618340, txid1619053, txid1794810, txid1794811, and txid1817799. The DPANN genomes/proteomes can be retrieved from the lineages corresponding to the following taxonomic IDs: txid1462430, txid1783276, and txid1803511. All the taxonomic information on these genomes is given in Table S6.

Detection of homologs of reference QS proteins in CPR and DPANN. The homologs of reference, experimentally validated QS proteins in CPR and DPANN, were identified with BLASTp version 2.2.31+ (67), using as queries all the protein sequences composing the reference QS protein families (Table S1) and as targets all the retrieved protein sequences corresponding to the publicly available CPR and DPANN genomes. Homology was classically assessed according to the following thresholds: sequence identity, $\geq$25%; E value, $<$1e$-$5; mutual alignment coverage, $\geq$75% (30, 31). Homologs identified with an E value of $<$1e$-$20 were assigned a specific "significance" label in the corresponding heat map (Fig. 2). Homologs found at the first BLAST iteration were used as queries in a second BLAST search, using the same thresholds but imposing the restriction that the remote homologs must exhibit $\geq$75% mutual coverage with the initial, reference QS proteins. Independently, the HMMs of QS proteins were searched against the library of CPR and DPANN protein sequences using hmmsearch from the HMMER suite (64). Only the hits yielding an E value of $<$1e$-$4 were retained, representing a threshold which is 100 times more stringent than the default inclusion threshold. Homologs identified by both BLAST and hmmsearch were highlighted in the corresponding heat map (Fig. 2).

Detection of homologs of the CPR and DPANN variants of QS proteins in the complete genomes of *Bacteria* and Archaea. The protein sequences corresponding to all available complete genomes of *Archaea* and representative *Bacteria* were retrieved from the NCBI assembly database 16 July 2019. The best homolog of each CPR/DPANN variant of a QS protein was identified using the same method as described above, albeit with no second BLAST iteration. The comprehensive list of all query QS proteins and the CPR/DPANN variants and their respective best matches in reference prokaryotes is given in Table S2.

Multiple-sequence alignment (MSA) of each family of homologs. All the homologs identified by a query family of QS proteins were aligned together to further compute their associated *dN/dS* ratio. For visualization purposes, the MSAs displayed in Fig. 3 were built only from the best homolog (lowest E value) identified in each CPR or DPANN phylum. Jalview was then used to display the conservation level of the important residues (68). In both cases, the MSAs were built using Muscle version 3.2.1 with the option "-maxiters 50" (65).

***dN/dS* ratio of each family of homologs.** The MSA of each family of homologs identified in the CPR and DPANN lineages was trimmed by the use of trimAl version 1.4.rev22 with the options "-gt 0.8" and "colnumbering" to retain only the positions with less than 20% of gaps (69). The NCBI IDs of the protein sequences in the alignment were used to fetch the corresponding coding DNA sequences (CDS) from the NCBI, via the "efetch -db protein -format fasta_cds_na -id $<$protein_id$>$" E-utilities command line. We generated the MSA of CDS based on the protein following MSA: gaps were substituted by the "—"

characters and amino acids by their corresponding codons. Accordingly, the stop codons were not represented in the MSA of CDS. Then, we introduced a slight modification in the source code of the SNAP perl script, a tool to compute synonymous and nonsynonymous values for a MSA of CDS (70). Namely, in the dictionary which associates an amino acid letter to a cognate codon, we substituted the dummy "Z" character associated with the opal "UGA" stop codon by a "G," to account for the alternative genetic code of "*Candidatus* Gracilibacteria" and "*Candidatus* Absconditabacteria," in which the opal codon encodes a glycine (NCBI:transl_table = 25). Of note, the other DPANN and CPR phyla were not reported to rely on an alternative genetic code. The *dN/dS* ratio of each family of homologs was then given as output by the thus-modified SNAP tool as the average of the *dN/dS* ratios computed for each possible pairwise comparison of sequences in the alignments.

Distribution of the homologs in the CPR and DPANN phyla. The multi-informative heat map (Fig. 2) was generated using the ComplexHeatmap R package version 1.99.5 (71). Two matrices were superimposed; the first one corresponds to the number of homologs identified in each phylum, normalized by the number of genomes in the phylum (grayscale background), whereas the second corresponds to the percentage of identity that is representative of the best match (lowest E value) between the proteins of a query family and their homologs identified in a phylum (colorscale foreground).

Secrete and sense tripartite network. The protein sequences of each family of homologs were processed in a BLAST All versus All search (serving as both queries and targets) (67). Only the pairs of proteins showing a percentage of identity of ≥90% over 75% mutual coverage were retained. Based on these pairs of highly similar proteins, a network was built to identify the nodes (proteins) that are connected to each other, thereby forming connected components (a group of highly similar proteins). The subfamilies were defined according to the distinct connected components identified in the network (Table S5, sheet 1). Each subfamily was further assigned a "divergence" indicator, defined as the average of the percentages of sequence identity between the sequences of the CPR/DPANN subfamilies and their respective best match in the reference prokaryotes. Subsequently, an "Edge" table was constructed to reference the membership of each genome in the different subfamilies (Table S5, sheet 2). A "Type" table was also constructed to assign a level to each node of the "Edge" table as follows: level 1 for the subfamilies of putative QS synthases, level 2 for the genomes, and level 3 for the subfamilies of putative QS receptors (Table S5, sheet 3). All the edges were considered in the DPANN tripartite network whereas only the connected components of size >3 were retained in the CPR network. These two tripartite networks were plotted with the igraph R package version 1.2.4 (72), using the Sugiyama algorithm to display the three levels of nodes in layers and to minimize the entanglement of the edges in the two-dimensional (2D) projection of each graph (73).

Data availability. All the CPR/DPANN homologs of QS proteins, as well as the taxonomic information on the genomes discussed here, are given in the supplemental material.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.7 MB.
FIG S2, PDF file, 2.2 MB.
FIG S3, PDF file, 0.6 MB.
TABLE S1, XLSX file, 0.04 MB.
TABLE S2, XLSX file, 0.7 MB.
TABLE S3, XLSX file, 1.5 MB.
TABLE S4, XLSX file, 0.1 MB.
TABLE S5, XLSX file, 0.4 MB.
TABLE S6, XLSX file, 0.2 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48.

2. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542. https://doi.org/10.1038/s41564-017-0012-7.

3. Castelle CJ, Banfield JF. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell 172:1181–1197. https://doi.org/10.1016/j.cell.2018.02.016.

4. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, Singh A, Williams KH, Siegerist CE, Tringe SG, Downing KH, Comolli LR, Banfield JF. 2015. Diverse uncultivated ultra-small bacterial cells in ground-water. Nat Commun 6:6372. https://doi.org/10.1038/ncomms7372.

5. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. 2018. Biosynthetic capacity, metabolic variety and unusual biology in

the CPR and DPANN radiations. Nat Rev Microbiol 16:629–645. https://doi.org/10.1038/s41579-018-0076-2.

6. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, Thomas BC, Banfield JF. 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. mBio 4:e00708-13. https://doi.org/10.1128/mBio.00708-13.

7. Gong J, Qing Y, Guo X, Warren A. 2014. "Candidatus Sonnebornia yantaiensis", a member of candidate division OD1, as intracellular bacteria of the ciliated protist Paramecium bursaria (Ciliophora, Oligohymenophorea). Syst Appl Microbiol 37:35–41. https://doi.org/10.1016/j.syapm.2013.08.007.

8. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu S-Y, Dorrestein PC, Esquenazi E, Hunter RC, Cheng G, Nelson KE, Lux R, Shi W. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci U S A 112:244–249. https://doi.org/10.1073/pnas.1419038112.

9. Dombrowski N, Lee J-H, Williams TA, Offre P, Spang A. 2019. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiol Lett 366:fnz008. https://doi.org/10.1093/femsle/fnz008.

10. Bedree JK, Bor B, Cen L, Edlund A, Lux R, McLean JS, Shi W, He X. 2018. Quorum sensing modulates the epibiotic-parasitic relationship between Actinomyces odontolyticus and its Saccharibacteria epibiont, a Nanosynbacter lyticus strain, TM7x. Front Microbiol 9:2049. https://doi.org/10.3389/fmicb.2018.02049.

11. Rajput A, Kumar M. 2017. Computational exploration of putative LuxR solos in Archaea and their functional implications in quorum sensing. Front Microbiol 8:798. https://doi.org/10.3389/fmicb.2017.00798.

12. Nealson KH, Platt T, Hastings JW. 1970. Cellular control of the synthesis and activity of the bacterial luminescent system. J Bacteriol 104:313–322. https://doi.org/10.1128/JB.104.1.313-322.1970.

13. Fuqua WC, Winans SC, Greenberg EP. 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J Bacteriol 176:269–275. https://doi.org/10.1128/jb.176.2.269-275.1994.

14. Miller MB, Bassler BL. 2001. Quorum sensing in bacteria. Annu Rev Microbiol 55:165–199. https://doi.org/10.1146/annurev.micro.55.1.165.

15. Verma SC, Miyashiro T. 2013. Quorum sensing in the squid-Vibrio symbiosis. Int J Mol Sci 14:16386–16401. https://doi.org/10.3390/ijms140816386.

16. Pérez-Montaño F, Jiménez-Guerrero I, Del Cerro P, Baena-Ropero I, López-Baena FJ, Ollero FJ, Bellogín R, Lloret J, Espuny MR. 2014. The symbiotic biofilm of Sinorhizobium fredii SMH12, necessary for successful colonization and symbiosis of Glycine max cv Osumi, is regulated by quorum sensing systems and inducing flavonoids via NodD1. PLoS One 9:e105901. https://doi.org/10.1371/journal.pone.0105901.

17. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, Melamed S, Leavitt A, Savidor A, Albeck S, Amitai G, Sorek R. 2017. Communication between viruses guides lysis-lysogeny decisions. Nature 541:488–493. https://doi.org/10.1038/nature21049.

18. Kai K. 2018. Bacterial quorum sensing in symbiotic and pathogenic relationships with hosts. Biosci Biotechnol Biochem 82:363–371. https://doi.org/10.1080/09168451.2018.1433992.

19. Smith RS, Iglewski BH. 2003. P aeruginosa quorum-sensing systems and virulence. Curr Opin Microbiol 6:56–60. https://doi.org/10.1016/s1369-5274(03)00008-0.

20. Jung SA, Chapman CA, Ng W-L. 2015. Quadruple quorum-sensing inputs control Vibrio cholerae virulence and maintain signal robustness. PLoS Pathog 11:e1004837. https://doi.org/10.1371/journal.ppat.1004837.

21. Sztajer H, Szafranski SP, Tomasch J, Reck M, Nimtz M, Rohde M, Wagner-Döbler I. 2014. Cross-feeding and interkingdom communication in dual-species biofilms of Streptococcus mutans and Candida albicans. ISME J 8:2256–2271. https://doi.org/10.1038/ismej.2014.73.

22. Sun J, Daniel R, Wagner-Döbler I, Zeng A-P. 2004. Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. BMC Evol Biol 4:36. https://doi.org/10.1186/1471-2148-4-36.

23. Mukherjee S, Bassler BL. 2019. Bacterial quorum sensing in complex and dynamically changing environments. Nat Rev Microbiol 17:371–382. https://doi.org/10.1038/s41579-019-0186-5.

24. Coutinho BG, Mevers E, Schaefer AL, Pelletier DA, Harwood CS, Clardy J, Greenberg EP. 2018. A plant-responsive bacterial-signaling system senses an ethanolamine derivative. Proc Natl Acad Sci U S A 115:9785–9790. https://doi.org/10.1073/pnas.1809611115.

25. Silpe JE, Bassler BL. 2019. A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision. Cell 176:268–280.e13. https://doi.org/10.1016/j.cell.2018.10.059.

26. Karavolos MH, Winzer K, Williams P, Khan CMA. 2013. Pathogen espionage: multiple bacterial adrenergic sensors eavesdrop on host communication systems. Mol Microbiol 87:455–465. https://doi.org/10.1111/mmi.12110.

27. González JF, Venturi V. 2013. A novel widespread interkingdom signaling circuit. Trends Plant Sci 18:167–174. https://doi.org/10.1016/j.tplants.2012.09.007.

28. Hochstrasser R, Hilbi H. 2017. Intra-species and inter-kingdom signaling of Legionella pneumophila. Front Microbiol 8:79. https://doi.org/10.3389/fmicb.2017.00079.

29. Lee W, Lee S-H, Kim M, Moon J-S, Kim G-W, Jung H-G, Kim IH, Oh JE, Jung HE, Lee HK, Ku KB, Ahn D-G, Kim S-J, Kim K-S, Oh J-W. 2018. Vibrio vulnificus quorum-sensing molecule cyclo(Phe-Pro) inhibits RIG-I-mediated antiviral innate immunity. Nat Commun 9:1606. https://doi.org/10.1038/s41467-018-04075-1.

30. Haggerty LS, Jachiet PA, Hanage WP, Fitzpatrick DA, Lopez P, O'Connell MJ, Pisani D, Wilkinson M, Bapteste E, McInerney JO. 2014. A pluralistic account of homology: adapting the models to the data. Mol Biol Evol 31:501–516. https://doi.org/10.1093/molbev/mst228.

31. Lannes R, Olsson-Francis K, Lopez P, Bapteste E. 2019. Carbon fixation by marine ultrasmall prokaryotes. Genome Biol Evol 11:1166–1177. https://doi.org/10.1093/gbe/evz050.

32. Rajput A, Kaur K, Kumar M. 2016. SigMol: repertoire of quorum sensing signaling molecules in prokaryotes. Nucleic Acids Res 44:D634–D639. https://doi.org/10.1093/nar/gkv1076.

33. Hornung C, Poehlein A, Haack FS, Schmidt M, Dierking K, Pohlen A, Schulenburg H, Blokesch M, Plener L, Jung K, Bonge A, Krohn-Molt I, Utpatel C, Timmermann G, Spieck E, Pommerening-Röser A, Bode E, Bode HB, Daniel R, Schmeisser C, Streit WR. 2013. The Janthinobacterium sp. HH01 genome encodes a homologue of the V. cholerae CqsA and L. pneumophila LqsA autoinducer synthases. PLoS One 8:e55045. https://doi.org/10.1371/journal.pone.0055045.

34. Tang X, Guo Y, Zhu T, Tao H, Liu S. 2019. Identification of quorum sensing signal AHLs synthases in Candidatus Jettenia caeni and their roles in anammox activity. Chemosphere 225:608–617. https://doi.org/10.1016/j.chemosphere.2019.02.192.

35. Papenfort K, Silpe JE, Schramma KR, Cong JP, Seyedsayamdost MR, Bassler BL. 2017. A Vibrio cholerae autoinducer-receptor pair that controls biofilm formation. Nat Chem Biol 13:551–557. https://doi.org/10.1038/nchembio.2336.

36. Kim CS, Gatsios A, Cuesta S, Lam YC, Wei Z, Chen H, Russell RM, Shine EE, Wang R, Wyche TP, Piizzi G, Flavell RA, Palm NW, Sperandio V, Crawford JM. 2020. Characterization of autoinducer-3 structure and biosynthesis in E. coli. ACS Cent Sci 6:197–206. https://doi.org/10.1021/acscentsci.9b01076.

37. Zhou L, Zhang LH, Cámara M, He YW. 2017. The DSF family of quorum sensing signals: diversity, biosynthesis, and turnover. Trends Microbiol 25:293–303. https://doi.org/10.1016/j.tim.2016.11.013.

38. Watve S, Barrasso K, Jung SA, Davis KJ, Hawver LA, Khataokar A, Palaganas RG, Neiditch MB, Perez LJ, Ng W-L. 2020. Parallel quorum-sensing system in Vibrio cholerae prevents signal interference inside the host. PLoS Pathog 16:e1008313. https://doi.org/10.1371/journal.ppat.1008313.

39. Howard MF, Renee Bina X, Bina JE. 2019. Indole inhibits ToxR regulon expression in Vibrio cholerae. Infect Immun 87:e00776-18. https://doi.org/10.1128/IAI.00776-18.

40. Brameyer S, Heermann R. 2017. Quorum sensing and LuxR solos in Photorhabdus. Curr Top Microbiol Immunol 402:103–119. https://doi.org/10.1007/82_2016_28.

41. Rickard AH, Palmer RJ, Blehert DS, Campagna SR, Semmelhack MF, Egland PG, Bassler BL, Kolenbrander PE. 2006. Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth. Mol Microbiol 60:1446–1456. https://doi.org/10.1111/j.1365-2958.2006.05202.x.

42. Spirig T, Tiaden A, Kiefer P, Buchrieser C, Vorholt JA, Hilbi H. 2008. The Legionella autoinducer synthase LqsA produces an alpha-hydroxyketone signaling molecule. J Biol Chem 283:18113–18123. https://doi.org/10.1074/jbc.M801929200.

43. Simon S, Schell U, Heuer N, Hager D, Albers MF, Matthias J, Fahrnbauer F, Trauner D, Eichinger L, Hedberg C, Hilbi H. 2015. Interkingdom signaling by the Legionella quorum sensing molecule LAI-1 modulates cell migration through an IQGAP1-Cdc42-ARHGEF9-

dependent pathway. PLoS Pathog 11:e1005307. https://doi.org/10.1371/journal.ppat.1005307.

44. Werner KM, Perez LJ, Ghosh R, Semmelhack MF, Bassler BL. 2014. Caenorhabditis elegans recognizes a bacterial quorum-sensing signal molecule through the AWCON neuron. J Biol Chem 289:26566–26573. https://doi.org/10.1074/jbc.M114.573832.

45. Haack FS, Poehlein A, Kröger C, Voigt CA, Piepenbring M, Bode HB, Daniel R, Schäfer W, Streit WR. 2016. Molecular keys to the Janthinobacterium and Duganella spp. interaction with the plant pathogen Fusarium graminearum. Front Microbiol 7:1668. https://doi.org/10.3389/fmicb.2016.01668.

46. Brambila-Tapia AJL, Pérez-Rueda E. 2014. A functional and phylogenetic comparison of quorum sensing related genes in Brucella melitensis 16M. J Microbiol 52:709–715. https://doi.org/10.1007/s12275-014-3570-x.

47. Kim J, Park W. 2015. Indole: a signaling molecule or a mere metabolic byproduct that alters bacterial physiology at a high concentration? J Microbiol 53:421–428. https://doi.org/10.1007/s12275-015-5273-3.

48. Weigel WA, Demuth DR. 2016. QseBC, a two-component bacterial adrenergic receptor and global regulator of virulence in Enterobacteriaceae and Pasteurellaceae. Mol Oral Microbiol 31:379–397. https://doi.org/10.1111/omi.12138.

49. Arakawa K, Mochizuki S, Yamada K, Noma T, Kinashi H. 2007. gamma-Butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in Streptomyces rochei. Microbiology (Reading) 153:1817–1827. https://doi.org/10.1099/mic.0.2006/002170-0.

50. Kleerebezem M. 2004. Quorum sensing control of lantibiotic production; nisin and subtilin autoregulate their own biosynthesis. Peptides 25:1405–1414. https://doi.org/10.1016/j.peptides.2003.10.021.

51. Anderson JK, Huang JY, Wreden C, Sweeney EG, Goers J, Remington SJ, Guillemin K. 2015. Chemorepulsion from the quorum signal autoinducer-2 promotes Helicobacter pylori biofilm dispersal. mBio 6:e00379-15. https://doi.org/10.1128/mBio.00379-15.

52. Dintner S, Staron A, Berchtold E, Petri T, Mascher T, Gebhard S. 2011. Coevolution of ABC transporters and two-component regulatory systems as resistance modules against antimicrobial peptides in Firmicutes bacteria. J Bacteriol 193:3851–3862. https://doi.org/10.1128/JB.05175-11.

53. Polkade AV, Mantri SS, Patwekar UJ, Jangid K. 2016. Quorum sensing: an under-explored phenomenon in the phylum Actinobacteria. Front Microbiol 7:131. https://doi.org/10.3389/fmicb.2016.00131.

54. Sharif DI, Gallon J, Smith CJ, Dudley E. 2008. Quorum sensing in Cyanobacteria: N-octanoyl-homoserine lactone release and response, by the epilithic colonial cyanobacterium Gloeothece PCC6909. ISME J 2:1171–1182. https://doi.org/10.1038/ismej.2008.68.

55. Zhang G, Zhang F, Ding G, Li J, Guo X, Zhu J, Zhou L, Cai S, Liu X, Luo Y, Zhang G, Shi W, Dong X. 2012. Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon. ISME J 6:1336–1344. https://doi.org/10.1038/ismej.2011.203.

56. Paggi RA, Martone CB, Fuqua C, Castro RE. 2003. Detection of quorum sensing signals in the haloalkaliphilic archaeon Natronococcus occultus. FEMS Microbiol Lett 221:49–52. https://doi.org/10.1016/S0378-1097(03)00174-5.

57. Mehmood A, Liu G, Wang X, Meng G, Wang C, Liu Y. 2019. Fungal quorum-sensing molecules and inhibitors with potential antifungal activity: a review. Molecules 24:1950. https://doi.org/10.3390/molecules24101950.

58. Liu B, Faller LL, Klitgord N, Mazumdar V, Ghodsi M, Sommer DD, Gibbons TR, Treangen TJ, Chang Y-C, Li S, Stine OC, Hasturk H, Kasif S, Segrè D, Pop M, Amar S. 2012. Deep sequencing of the oral microbiome reveals signatures of periodontal disease. PLoS One 7:e37919. https://doi.org/10.1371/journal.pone.0037919.

59. Rylev M, Bek-Thomsen M, Reinholdt J, Ennibi O-K, Kilian M. 2011. Microbiological and immunological characteristics of young Moroccan patients with aggressive periodontitis with and without detectable Aggregatibacter actinomycetemcomitans JP2 infection. Mol Oral Microbiol 26:35–51. https://doi.org/10.1111/j.2041-1014.2010.00593.x.

60. Jang Y-J, Choi Y-J, Lee S-H, Jun H-K, Choi B-K. 2013. Autoinducer 2 of Fusobacterium nucleatum as a target molecule to inhibit biofilm formation of periodontopathogens. Arch Oral Biol 58:17–27. https://doi.org/10.1016/j.archoralbio.2012.04.016.

61. Wynendaele E, Bronselaer A, Nielandt J, D'Hondt M, Stalmans S, Bracke N, Verbeke F, Van De Wiele C, De Tré G, De Spiegeleer B. 2013. Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. Nucleic Acids Res 41:D655–D659. https://doi.org/10.1093/nar/gks1137.

62. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432. https://doi.org/10.1093/nar/gky995.

63. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA, Sillitoe I. 2017. CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res 45:D289–D295. https://doi.org/10.1093/nar/gkw1098.

64. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

65. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.

66. NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44:D7–D19. https://doi.org/10.1093/nar/gkv1290.

67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

68. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

69. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

70. Korber-Irrgang B. 2000. HIV signature and sequence variation analysis. Computational analysis of HIV molecular sequences. https://www.scienceopen.com/document?vid=f054582b-15a6-4ccc-be34-99d0e9d09004.

71. Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32:2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

72. Csardi G, Nepusz T. 2005. The igraph software package for complex network research. https://www.researchgate.net/publication/221995787_The_Igraph_Software_Package_for_Complex_Network_Research.

73. Nikolov NS. 2016. Sugiyama algorithm, p 2162–2166. In Encyclopedia of algorithms. Springer New York, New York, NY.

74. Johnson AR, Dekker EE. 1998. Site-directed mutagenesis of histidine-90 in Escherichia coli L-threonine dehydrogenase alters its substrate specificity. Arch Biochem Biophys 351:8–16. https://doi.org/10.1006/abbi.1997.0501.

75. Wang Y, Yi L, Wang S, Fan H, Ding C, Mao X, Lu C. 2015. Crystal structure and identification of two key amino acids involved in AI-2 production and biofilm formation in Streptococcus suis LuxS. PLoS One 10:e0138826. https://doi.org/10.1371/journal.pone.0138826.

76. Churchill MEA, Chen L. 2011. Structural basis of acyl-homoserine lactone-dependent signaling. Chem Rev 111:68–85. https://doi.org/10.1021/cr1000817.

77. Liu L, Li T, Cheng XJ, Peng CT, Li CC, He LH, Ju SM, Wang NY, Ye TH, Lian M, Xiao QJ, Song YJ, Zhu YB, Yu LT, Wang ZL, Bao R. 2018. Structural and functional studies on Pseudomonas aeruginosa DspI: implications for its role in DSF biosynthesis. Sci Rep 8:3928. https://doi.org/10.1038/s41598-018-22300-1.

78. Wang J, Liu Q, Li X, Ma S, Hu H, Wu B, Zhang X-X, Ren H. 2020. In-situ monitoring AHL-mediated quorum-sensing regulation of the initial phase of wastewater biofilm formation. Environ Int 135:105326. https://doi.org/10.1016/j.envint.2019.105326.

79. Kim J, Park W. 2013. Indole inhibits bacterial quorum sensing signal transmission by interfering with quorum sensing regulator folding. Microbiology (Reading) 159:2616–2625. https://doi.org/10.1099/mic.0.070615-0.