Large-scale identification of viral quorum sensing systems reveal convergent evolution of density-dependent sporulation-hijacking in bacteriophages

AUTHORS

Charles Bernard [1,2,*], Yanyan Li [2], Philippe Lopez [1] and Eric Bapteste [1]

AFFILIATIONS

[1] Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d'Histoire Naturelle, EPHE, Université des Antilles, Campus Jussieu, Bâtiment A, 4eme et. Pièce 429, 75005 Paris, France

[2] Unité Molécules de Communication et Adaptation des Micro-organismes (MCAM), CNRS, Museum National d'Histoire Naturelle, CP 54, 57 rue Cuvier, 75005 Paris, France

CORRESPONDING AUTHOR

* Correspondence to Charles Bernard (ORCID Number: 0000-0002-8354-5350);

Phone: +33 (01) 44 27 34 70; E-mail address: charles.bernard@cri-paris.org

ABSTRACT

Quorum sensing systems (QSSs) are genetic systems supporting cell-cell or bacteriophage-bacteriophage communication. By regulating behavioral switches as a function of the encoding population density, QSSs shape the social dynamics of microbial communities. However, their diversity is tremendously overlooked in bacteriophages, which implies that many density-dependent behaviors likely remains to be discovered in these viruses. Here, we developed a signature-based computational method to identify novel peptide-based RRNPP QSSs in gram-positive bacteria (e.g. Firmicutes) and their mobile genetic elements. The large-scale application of this method against available genomes of Firmicutes and bacteriophages revealed 2708 candidate RRNPP-type QSSs, including 382 found in (pro)phages. These 382 viral candidate QSSs are classified into 25 different groups of homologs, of which 22 were never described before in bacteriophages. Remarkably, genomic context analyses suggest that candidate viral QSSs from 6 different families dynamically manipulate the host biology. Specifically, many viral candidate QSSs are predicted to regulate, in a density-dependent manner, adjacent (pro)phage-encoded regulator genes whose bacterial homologs are key regulators of the sporulation initiation pathway (either Rap, Spo0E, or AbrB). Consistently, we found evidence from public data that certain of our candidate (pro)phage-encoded QSSs dynamically manipulate the timing of sporulation of the bacterial host. These findings challenge the current paradigm assuming that bacteria decide to sporulate in adverse situation. Indeed, our survey highlights that bacteriophages have evolved, multiple times, genetic systems that dynamically influence this decision to their advantage, making sporulation a survival mechanism of last resort for phage-host collectives.

KEYWORDS:

Bacteriophages - Quorum sensing – Communication - Sporulation – Manipulation – RRNPP

INTRODUCTION

Quorum sensing systems (QSSs) are genetic systems primarily supporting cell-cell communication (1,2), but also plasmid-plasmid (3), or bacteriophage-bacteriophage (4,5) communication. Upon bacterial expression, a QSS enables individuals of an encoding population (bacterial chromosomes, plasmids or intracellular bacteriophage genomes) to produce a communication signal molecule that accumulates in the environment as the population grows. At a threshold concentration, reflecting a quorum of the encoding population, the signal is transduced population-wide and thereupon regulates a behavioral switch (2,6,7). QSSs thereby shape the social dynamics of microbial communities and optimize the way these communities react to changes in their environments. If QSSs are well described in bacterial chromosomes, their diversity is under-explored in mobile genetic elements (MGEs), and particularly in bacteriophages, yet by far the most abundant biological entities on Earth (8). To date, only 2 types of QSSs have been

recently described in bacteriophages: the lysogeny-regulating “arbitrium” QSSs (4,5) and the host-derived Rap-Phr QSSs (9). Expanding the diversity of bacteriophage-encoded QSSs would unravel novel decision-making processes taken by these viruses, that would have major consequences on the understanding of microbial interaction, adaptation and evolution.

Expanding the diversity of viral QSSs implies developing methods to detect novel QSS families, beyond homology searches that limits the results to representatives of already known families. Here we demonstrate that an *in silico* detectable signature is common between distinct, experimentally-characterized families of QSSs and is thus sufficiently generic to discover novel QSSs while being specific to quorum sensing. These families rely on small peptides as communication molecules, are specific to Firmicutes and their MGEs, and are grouped under the name RRNPP, which stands for the Rap, Rgg, NprR, PlcR and PrgX families of quorum sensing receptors (7,10–12)). We thus systematically queried the RRNPP signature against the NCBI database of complete genomes of Viruses but also of Firmicutes, because a bacteriophage genome can be inserted, under the form of a latent prophage, within the genome of its bacterial host. For more applied considerations, we also searched for this signature within human-associated bacteriophages from the Gut Phage Database (13). We report the identification of 382 (pro)phage-encoded candidate QSSs, classified into 25 distinct QSS families of homologs, of which 22 were never described before in bacteriophages, which may represent a 7-fold increase of the described diversity of viral QSS families.

RRNPP-type QSSs often regulate adjacent genes, which is especially true (no counterexamples yet known) for QSSs encoded by MGEs such as bacteriophages and plasmids (4,5,12,14). Consistently, we meticulously examined the genomic context of our candidate (pro)phage-encoded QSSs to predict their function. Remarkably, in many cases, we observed an unsuspected clustering of different viral QSSs with (pro)phage-encoded regulator genes (i.e *rap*, *spo0E*, or *abrB)* whose bacterial homologs are key regulators of the bacterial sporulation initiation pathway (15–18). Consistent with this observation, we next found in the literature multiple independent experimental data reporting that some of our candidate QSSs that we predict to be encoded by *Bacillus* and *Clostridium* prophages affect the timing of sporulation in their respective host. Finally, we uncovered a high abundance of *spo0E* and *abrB* genes, as well as one *rap*-based QSS in the Gut Phage Database (13), highlighting that gastrointestinal viruses regulate, within humans, the dynamics of formation of bacterial endospores specialized for host-host transmission (19).

Here, our findings challenge the sporulation paradigm, which assumes that spore-forming Firmicutes decide to sporulate in adverse situations (20). Indeed, our survey revealed that bacteriophages have evolved, multiple times, QSSs that dynamically influence the sporulation

decision-making process for their own evolutionary benefit. Importantly, as the sporulation initiation pathway can trigger a wide range of biological processes (sporulation, biofilm formation, cannibalism, toxin production or solventogenesis) (21,22), our unraveled viral candidate QSSs also likely manipulate, in a density-dependent manner, a substantially broader spectrum of the host biology than spore formation alone. Considering that endospores formed by pathogens are linked to serious health issues ranging from food-safety, bio-terrorism to infectious diseases (23–29) and that endospores formed by commensal bacteria can be leveraged to treat gastrointestinal dysbioses (30), these new insights may pave the way to major practical outcomes.

RESULTS

Large-scale query of the RRNPP-type signature reveals hundreds of candidate QSSs encoded by free bacteriophages or prophages

RRNPP-type QSSs are composed of two adjacent genes and are specific to gram-positive Firmicutes bacteria and their bacteriophages. The emitter gene encodes a small pro-peptide that is secreted, except in rare exceptions, via the SEC-translocon and matured extracellularly by exopeptidases into a mature quorum sensing peptide. This mature peptide accumulates in the medium as the emitting population grows, and is imported by the Opp permease at high concentrations, therefore at high population densities. The receptor gene encodes an intracellular protein inhibitor or a transcription factor that interacts with the imported mature peptide, via peptide-binding motifs called tetratricopeptide repeats (TPRs). Upon binding with the signal peptide, the receptor undergoes a conformational change, which translates into the subsequent induction or inhibition of target pathways at high population densities (7,10–12) (**Fig. S1**).

The detailed examination of similarities between different, functionally-validated RRNPP-type QSS families revealed a generic signature of 5 criteria that can be very effectively detected *in silico* (explained in details in **Fig. S2** and in **Materials and Methods**). In brief, detecting this signature consists first, of identifying candidate receptors, defined as proteins of 250-460aa matching Hidden Markov Models of TPRs (E-value of <1E-5, 1000x more stringent than default threshold), the structural motifs involved in the binding of small peptides (and in the case of RRNPP QSSs, of quorum sensing peptides). Second, it consists in retaining only the coding sequences of those putative receptors that are located directly adjacent to the coding sequence of a candidate communication pro-peptide, defined as a small protein of 15-65 aa predicted to be secreted via the SEC-translocon by the stringent SignalP software (**Fig. 1** and **S2, Materials and Methods**). The pubmed query '*"Tetratricopeptide" "Peptide" Secretion Firmicutes'*, despite no keywords directly linked to quorum sensing, yields 10 (out of 11) results describing RRNPP-type QSSs, highlighting the intrinsic link between this signature and quorum sensing. As this signature-based method does not rely on homology search of already known QSSs, it has the potential to detect novel candidate

RRNPP-type QSS families, and thus novel 'languages' of peptide-based biocommunication via quorum sensing. The same principle, *albeit* implemented differently, was recently applied by Voichek *et al.* in the *Paenibacillus* genus, and proved its efficiency at detecting novel, functional QSSs (31).

At first, we queried the RRNPP-type specific signature against the high-quality complete genomes of Firmicutes (3,577 genomes (chromosomes + plasmids)) and Viruses (32,327 genomes) available at the NCBI. This systematic search led to the detection of 2681 candidate QSSs. There were no false negatives for reference RRNPP-type QSSs: we identified 100% of the Rap-Phr, NprR-NprX, PlcR-PapR, TraA-Ipd1 AimP-AimR and $AimP_{like}$-$AimR_{like}$ reference QSS families in which the pro-peptide is not mentioned to be secreted otherwise than via the SEC-translocon (4,5,12) (**Table S2, Materials and Methods**). Consistent with the fact that RRNPP-type QSSs are specific to Firmicutes, only QSSs encoded by bacteriophages of Firmicutes were identified in the dataset of all available viral genomes. Here, our 2681 unraveled candidate QSSs are distributed as such: 2124 are encoded by chromosomes, 189 by plasmids (Bernard *et al.* in prep), 10 by genomes of free phages of Firmicutes while 358 were predicted by Phaster (32) and ProphageHunter (33) to belong to prophages (174 classified as intact/active prophages, 68 as questionable/ambiguous prophages and 116 as incomplete prophages) (**Table S1, Materials and Methods**). We next sought to characterize the diversity of this unprecedented, massive library of phage- and prophage-encoded candidate QSSs.

These (pro)phage-encoded candidate QSSs are distributed into 16 families, 13 of which were never described before in bacteriophages

We next classified these 2681 unraveled candidate QSSs into families, defined as groups of homologous receptors. To this end, we launched a BLASTp (34) all vs all of the 2681 receptors, and retained only pairs of receptors yielding a sequence identity >=30% over more than 80% of the lengths of the two sequences. Subsequently, the connected components of the resulting sequence similarity network were used to define QSS families (**Materials and Methods**). We thereby identified a total of 56 families of candidate QSS receptors, 16 of which included at least one candidate QSS encoded by either a phage or a predicted prophage (**Table S1**). We next focused our study on the computational characterization of the viral QSSs from these 16 families.

Homology assessment of these 16 families with reference RRNPP-type QSS receptors revealed that only 3 families had already been characterized before in phages: the Rap-Phr family shared between chromosomes, plasmids and bacteriophages (9,35), the AimR-AimP QSS family specific to (pro)phages of the *B. subtilis* group (4), and the AimR-AimP-like QSS family specific to (pro)phages of the *B. cereus* group (5) (**Table S2, Materials and Methods**). Accordingly, 13 of the

16 RRNPP-type candidate QSS families in which at least one candidate QSS is encoded by a (pro)phage had never been described before in bacteriophages and may therefore substantially expand the known diversity of viral QSSs (**Table 1**). Interestingly, 3 of these 13 families (families n°1, 2 and 3) happen to be present in both bacterial chromosomes and phages/prophages, as in the case of the Rap-Phr family (**Fig. S3**).

Table 1: novel candidate QSSs in phages and predicted prophages

QSS id	Family	Receptor NCBI Id	DNA binding motif	Pro-peptide NCBI Id	SEC-secretion likelihood	Inferred mature peptide	Intergenic distance (bp)	QSS-encoding genome	Prophage Hunter prediction	Phaster prediction
1α	1	ALA47936.1	Yes	ALA47937.1	0.81	TDNPGY	-1	Brevibacillus phage Sundance	Phage genome (not applicable)	Phage genome (not applicable)
1β	1	AIG26090.1	Yes	AIG26091.1	0.94	NADPGY	13	*Brevibacillus latero-sporus* LMG15441	Ambiguous prophage (0.53)	-
1γ	1	VEF92012.1	Yes	VEF92013.1	0.98	RVEPDW	21	*Brevibacillus brevis* NCTC2611	Ambiguous prophage (0.79)	-
2	2	VEF92631.1	Yes	VEF92630.1	0.76	THGAG	-1	*Brevibacillus brevis* NCTC2611	Active prophage (0.83)	-
3α	3	AGF56487.1	Yes	AGF56488.1	0.93	DSRDPD	68	*Clostridium saccharo-perbutylacetonicum* N1-4(HMT)	Active prophage (0.98)	-
3β	3	AGF59421.1	Yes	AGF59420.1	0.97	NTTDPY	112	*Clostridium saccharo-perbutylacetonicum* N1-4(HMT)	Ambiguous prophage (0.63)	-
3γ	3	AQR95595.1	Yes	AQR95596.1	0.94	NTLDPN	74	*Clostridium saccharo-perbutylacetonicum* N1-504	Active prophage (0.85)	Intact prophage (100)
4α	4	VEF87222.1	Yes	VEF87223.1	0.99	GPPE	15	*Brevibacillus brevis* NCTC2611	Active prophage (0.93)	Intact prophage (150)
4β	4	VEF87585.1	Yes	VEF87586.1	0.98	GPPD	25	*Brevibacillus brevis* NCTC2611	Active prophage (0.95)	Intact prophage (150)
5α	5	QIC08170.1	Yes	QIC08171.1	0.96	ITEPEW	-4	*Brevibacillus* sp. 7WMA2	Active prophage (0.83)	-
5β	5	AIG27473.1	Yes	AIG27472.1	0.89	STAPDW	1	*Brevibacillus latero-sporus* LMG15441	-	Incomplete prophage (10)
6	6	AGR47394.1	Yes	AGR47395.1	0.97		74	Brevibacillus phage Emery	Phage genome (not applicable)	Phage genome (not applicable)
7	7	ANT39976.1	Yes	ANT39977.1	0.92		120	Bacillus phage vB_BtS_BMBtp14	Phage genome (not applicable)	Phage genome (not applicable)
8	8	ADI00470.1	Yes	ADI00469.1	0.94		190	*Bacillus seleniti-reducens* MLS10	Ambiguous prophage (0.64)	Intact prophage (150)
9	9	BCB03503.1	Yes	BCB03504.1	0.98		55	*Bacillus* sp. KH172YL63	Ambiguous prophage (0.72)	-
10	10	QHQ60545.1	Yes	QHQ60546.1	0.87		99	*Anaerocolumna* sp. CBA3638	Ambiguous prophage (0.55)	-
11	11	ARU61133.1	Yes	ARU61134.1	0.91		-8	*Tumebacillus avium* AR23208	-	Incomplete prophage (10)
12	12	AGV99457.1	Yes	AGV99458.1	0.65		-59	Bacillus phage phiCM3	Phage genome (not applicable)	Phage genome (not applicable)
13	13	QCU03546.1	Yes	QCU03545.1	0.44		149	*Blautia* sp. SC05B48	Ambiguous prophage (0.55)	-

These 13 uncharacterized families include a total of 19 viral representatives, all presented in **Table 1** and named by an integer, indicative of the QSS family, followed by a greek letter in case of plural viral representatives in the family. QSSs 1α, 6, 7, 12 are encoded by genomes of free phages whereas QSSs 1β, 1γ, 2, 3α, 3β, 3γ, 4α, 4β, 5α, 5β, 8, 9, 10, 11, 13, 14 are predicted to belong to prophages (**Table 1**). Inducing prophage excision in each of the bacterial strains containing these systems will indicate whether these candidate QSSs belong to active prophages, able to re-initiate the lytic cycle after excision, or to cryptic prophages. The prediction of the activity of each of these prophages are given in **Table 1**. For each of these 20 novel viral candidate QSSs, the small, operonic intergenic distance between the receptor and the pro-peptide genes, together with the high likelihood that the pro-peptide is secreted via the SEC-translocon are excellent predictors that the genetic system is a QSS, functioning according to the canonical mechanism depicted in **Fig. S1**. The multiple sequence alignment of predicted cognate propeptides in each family of QSSs of size > 1 is shown in **Fig. S4**.

Rap-Phr QSSs that delay the timing of sporulation are found in many, diverse *Bacillus* bacteriophages

Among the already characterized QSS families that are matched by (pro)phage-encoded candidate QSS, the Rap-Phr is especially interestingly. Indeed, the Rap-Phr QSS family has long thought to be specific to genomes of *Bacillus* bacteria (36,37). In the *Bacillus* genus, bacterial Rap-Phr QSSs tend to be subpopulation-specific and regulate the last-resort sporulation initiation pathway in a density-dependent manner (35). In Firmicutes, the sporulation program leads to the formation of especially resistant endospores, able to resist extreme environmental stresses for prolonged periods (sometimes several thousand of years (38)) and to resume vegetative growth in response to favorable changes in environmental conditions (20). The sporulation pathway is initiated when transmembrane kinases sense stress stimuli, and thereupon transfer their phosphate, either directly (*Clostridium*) or via phosphorelay (*Bacillus, Brevibacillus*) to Spo0A, the master regulator of sporulation (39,40). The regulatory regions of developmental genes enacting the irrevocable entry into spore formation have a low affinity for the active Spo0A-P transcriptional regulator, implying that only high Spo0A-P concentrations, and therefore intense stresses, can commit a cell to sporulate (41). The research on the sporulation initiation pathway contributed to build the following paradigm: in adverse circumstances, a bacterium senses environmental stress factors, processes these input signals via an elaborated decision-making network of bacterial genes/proteins and undergoes spore formation only if the Spo0A-P concentration outputted by this regulatory circuit meets a certain threshold (16–18,42).

Notably, a Rap-Phr QSS ensures that Spo0A-P only accumulates when the Rap-Phr encoding subpopulation reaches high densities (42). Thus, the Rap-Phr QSS has been proposed as a

means for a *Bacillus* cell to delay a costly commitment to sporulation as long as the ratio of available food per kin-cell is compatible with individual survival in periods of nutrient limitation (42), in line with the paradigm posing that the decision to sporulate is essentially a bacterial biological process.

However, the Rap-Phr QSS family was recently shown to be mobile (35), and we previously demonstrated that it can be found on plasmids and (pro)phages, in addition to bacterial chromosomes (9). Accordingly, the delay in the timing of sporulation observed in a bacterium expressing a Rap-Phr QSS can find its source in a non bacterial, third party genetic element, and can therefore be dependent on the density of this genetic element entrapped within bacteria rather than on the actual bacterial cell density. For example, we showed that a functionally validated Rap-Phr system, the RapBL5-PhrBL5 system (NCBI IDs AAU41846.1 and AAU41847.1) of *B. licheniformis* (35), initially thought to be encoded by bacterial genes, was actually assessed by Phaster to belong to an intact prophage region (9). Consequently, the delay in Spo0A-P accumulation shown to be controlled by RapBL5-PhrBL5, was in fact governed by a viral genetic system. In the discussion section of this manuscript, we attempt to explain what evolutionary advantages may underlie the selection of such manipulative Rap-Phr QSSs in bacteriophages.

Here, we identified 1753 chromosomal, 179 plasmidic, 324 prophage-encoded and 1 phage-encoded *rap-phr* genetic systems in the complete genomes of Viruses and Firmicutes available at the NCBI, unraveling an unsuspected massive use of Rap-Phr QSSs by bacteriophages (**Table S1**, **Materials and Methods**). To further appreciate the diversity of these viral QSSs and to better understand how Rap-Phr travel onto different kinds of genetic supports (chromosomes, plasmids, phages), we inferred the maximum-likelihood phylogeny of these Rap quorum sensing receptors. On the resulting, mid-rooted phylogenetic tree, we colored each leaf according to the type of genetic element encoding the Rap-Phr QSS: blue for chromosomes, orange for plasmids and purple for bacteriophages (**Fig. 2**). This unprecedented mapping reveals a high diversity of viral Rap-Phr QSSs, in both prophages of the *Bacillus subtilis* and *Bacillus cereus* groups, because these viral Rap-Phr were not monophyletic but distributed into at least 6 groups, interspaced between bacterial clades. This polyphily of viral Rap-Phr QSSs suggest multiple, independent acquisitions of Rap-Phr in bacteriophages, and thus multiple acquisitions of potential sporulation-hijacking genetic systems. On another note, this phylogenetic tree highlights, for the first time, that frequent transfers of communication systems can occur between bacterial chromosomes and MGEs.

Bacteriophages have evolved many different genetic systems predicted to dynamically modulate the bacterial sporulation initiation pathway via quorum sensing

We next focused our study on the computational characterization of the 19 novel candidate QSSs described in **Table 1,** for which the function remains unknown. To infer what biological processes these candidate QSSs might regulate, we took advantage of the following characteristic of functionally-validated RRNPP-type QSSs encoded by MGEs: when the intracellular receptor is a transcription factor that gets activated/deactivated upon binding with its cognate communication peptide, the genes regulated by the QSS were found to be located in its vicinity (**Fig. S2**) (4,5,12,14). Querying the HMMs of DNA binding domains found within functionally characterized RRNPP receptors, we found that these 19 QSS receptors all harbor a DNA binding domain and thus likely regulate the transcription of adjacent target genes (**Tables 1 and S1, Materials and Methods**). Accordingly, we analyzed the genomic neighborhood of these QSS receptors to predict their function.

Remarkably, we noticed that the two main regulators of Spo0A-P other than *rap*, *i.e.* the *spo0E* dephosphorylator of Spo0A-P and the *abrB* regulator of the transition state from vegative growth to sporulation (16–18), are often found in the genomic neighborhood of viral candidate QSSs. Specifically, we identified *spo0E* directly adjacent to QSS1α (Brevibacillus phage Sundance), two copies of *spo0E* directly adjacent with QSS3β (predicted prophage of *Clostridium saccharoperbutylacetonicum*), and *abrB* in the genomic vicinity of QSS4β (predicted prophage of *Brevibacillus brevis*) and QSS5α (predicted prophage of *Brevibacillus brevis* sp. 7WMA2) (**Fig. 3A**). These results especially make sense in light of a recently identified chromosomal RRNPP-type QSS, shown to regulate the expression of its adjacent *spo0E* gene in a density dependent manner (31). The functions of the other (pro)phage-encoded candidate QSSs were difficult to predict from their genomic contexts and would require further exciting functional studies to characterize which biological processes they might regulate in a (pro)phage-density dependent manner.

At this stage of analysis, we found that QSSs 1α, 3β, 4β and 5α represent 21% of the predicted novel viral candidate QSSs, which, added to the viral Rap-Phr QSSs, suggest a remarkable functional association between quorum sensing and the regulation of sporulation in bacteriophages. In addition to Rap-Phr of *Bacillus* phages, these results suggest that some phages/prophages of the *Brevibacillus* and *Clostridium* genera likely rely on other QSS families to communicate in order to keep track of their respective population density and regulate the expression of the (pro)phage-encoded *spo0E* or *abrB* gene accordingly (**Fig. 4**). Consistently, by influencing the total concentration of the Spo0E or AbrB regulator within bacterial hosts in a (pro)phage-density dependent manner, these viral genetic systems might influence the dynamics of Spo0A-P accumulation and thereby modulate the target pathways of the sporulation initiation program. From an evolutionary viewpoint, the facts that Rap and the receptors of QSSs 1α, 3β, 4β

and 5α belong to distinct protein families, and are encoded by bacteriophages from different hosts, suggest a remarkable convergent evolution, in bacteriophages, of the functional association between viral quorum sensing and the manipulation of the bacterial sporulation initiation pathway.

Furthermore, the AbrB and Rap proteins have not only been reported to regulate Spo0A-P accumulation but also to inhibit the competence pathway in *Bacillus*. Indeed, AbrB represses ComK, the transcription factor of late competence genes, whereas Rap may inhibit, in addition or in place of Spo0F-P, the ComA-P regulator of early competence genes (18). Accordingly, the Rap-Phr and AbrB-regulating QSSs encoded by bacteriophages could also modulate the host competence pathway, in addition to the sporulation initation pathway (**Fig. 5**). The RapBL5-PhrBL5 QSS of *B. licheniformis* prophage has even been experimentally demonstrated to delay both the sporulation and the competence pathways (35). Altogether, our genomic analyses suggest that different bacteriophages use different quorum sensing systems to dynamically manipulate a wide range of host biological processes, spanning from competence to the phenotypes controlled by Spo0A-P such as sporulation, biofilm formation, cannibalism, toxin production or solventogenesis (21,22,43).

Experimental evidence supporting the prediction that prophage-encoded QSSs influence the dynamics of sporulation in the *Clostridium* genus

As experimental data in *B. licheniformis* already support the prediction that viral Rap-Phr delay the *Bacillus* sporulation program as a function of (pro)phage densities (9), we next tried to identify whether publicly available biological data would substantiate our prediction that QSSs 1α, 3β, 4β and 5α regulate the expression of (pro)phage-encoded *spo0E* or *abrB* genes, and thereby dynamically manipulate the host sporulation initiation pathway. If we did not find experimental data in *Brevibacillus* bacteria to test this hypothesis for QSSs 1α, 4β and 5α, we noticed that two recent studies focuses on the functional characterization of putative RRNPP-type QSSs in solventogenic *Clostridium* species, the type of hosts of the predicted QSS3βR-encoding prophage.

The first study investigated the functions of the 5 RRNPP-type QSSs predicted in the genome of *Clostridium saccharoperbutylacetonicum* str. N1-4(HMT), the lysogenized host of the predicted QSS3α- and QSS3β-encoding prophages (44). In this study, the functions of the QSS3αR (locus Cspa_c27220) and QSS3βR (locus Cspa_c56960) receptors were assessed although it was then unknown that these two QSSs might actually correspond to two prophage regions. The results of this study indicate that QSS3βR likely represses its two downstream *spo0E* genes, in line with our prediction (**Fig. 3**). Consistent with the fact that Spo0E dephosphorylates Spo0A-P, the deletion of QSS3βR, expected to alleviate *spo0E* repression, was shown to result in decreased Spo0A-P levels and decreased sporulation efficiency. The same decrease in sporulation efficiency was observed when QSS3αR is deleted, despite no sporulation regulators in the genomic

neighborhood of QSS3α, suggesting that the monophyletic QSS3βR and QSS3αR (**Fig. S3**) may bind the same DNA motifs and thus repress a common set of target genes, including *spo0E.* Further, the authors of the study showed that QSS3αR and QSS3βR overexpression, expected to over-repress the *spo0E* inhibitor of Spo0A-P, each resulted in increased sporulation efficiency as compared to the wild type phenotype, with basal QSS3βR expression. As the overexpression of a QSS receptor should yield more free receptors than receptor-peptide complexes, this phenotype is expected to reflect the function of a receptor below the quorum mediated by high concentration of its cognate quorum sensing peptide. Hence, these results highlight that QSSs predicted to belong to *C. saccharoperbutylacetonicum* prophages antagonize the host sporulation initiation pathway in a density-dependent manner (**Fig. 4**).

In the second study, 8 RRNPP-type QSSs in the genome of *Clostridium acetobutylicum* ATCC 824 have been studied (45). The authors mentioned that the open reading frames of 7 of the 8 QSS pro-peptides were not present in the annotation file of the genome deposited on the NCBI. Our algorithm *a priori* captured only 1 QSS (QSSf) in this genome but succeeded at identifying 3 additional RRNPP-type QSSs when all these small ORFs were taken into account (QSSb, QSSg and QSSh) (**Table S3**). Among the 8 QSSs, QSSf (locus CA_C1214) and QSSg (locus CA_C1949) were identified by ProphageHunter as belonging to active prophages (likelihoods of 0.95 and 0.94, respectively) and QSSg was also predicted by Phaster to be encoded by an intact prophage (score of 150). Importantly, we found the *abrB* gene in the vicinity of QSSg, adding the latter to our initial list of prophage-encoded QSS inferred to regulate the sporulation initiation pathway of the host (**Fig. 3A**). In line with our prediction, the study showed that the QSSgR mutant was the only one of the 8 QSS receptor-mutants that resulted in a significant reduction in the number of endospores as compared to wild type after 7 days of culture (3-fold, p-value = 0.03).

The results from these two independent studies show that when certain QSS receptors detected as prophages-encoded are deleted, the sporulation pathway of *Clostridium* is antagonized. However, a QSS receptor is itself activated or inactivated upon binding with its cognate mature peptide, whose concentration reflects the density of the QSS-encoding population. Therefore, the QSSs of (pro)phages of solventogenic *Clostridium* might ensure that the inhibition of the host sporulation initiation pathway by the QSS receptors is not constitutive but only happens at high (pro)phage densities. As solventogenic *Clostridium* species acidify their medium as they grow, this mechanism could perhaps enable (pro)phages to coerce their hosts to maintain Spo0A-P levels that favors the costly alkalizing solventengenesis pathway over sporulation in response to medium acidification, for the benefit of (pro)phage replication (**Fig. 4**). These data, coupled with evidence from the RapBL5-PhrBL5 of *B. licheniformis* highlight that some (pro)phage-encoded QSSs manipulate the host sporulation initiation pathway in a density-dependent manner.

Metagenomic evidence that intestinal bacteriophages regulate sporulation in the human gut microbiota, via a rich repertoire of *spo0E* and *abrB* regulators

If our identification of multiple families of (pro)phage-encoded candidate QSSs predicted to mediate sporulation-hijacking is interesting from a fundamental viewpoint, notably in microbiology and evolution, it may also be of interest for more practical fields such as medicine. Indeed, as sporulation enables bacteria to resist various harsh environmental conditions, it represents a route for bacteria to travel between environments, and notably to end up within human bodies. Consequently, the endospore is the infectious form of many pathogens, among which *Bacillus anthracis* (26), the causative agent of anthrax or *Clostridium* (reclassified as *"Clostridioides"*) *difficile*, an emergent pathogen responsible for almost 223,900 hospitalizations and at least 12,800 US deaths in 2017 alone (23). It is notably well known that differentiating into an endospore allows anaerobic bacteria to resist air exposure and thus transmit between humans. Sporulation therefore participates in the dynamics of exchange of gastrointestinal bacteria between humans, which may cause outbreaks of nosocomial infections in the case of pathogenic species (24,29). In a recent study, Browne et al estimated that at least 50-60% of the bacterial genera from the intestinal microbiota of a healthy individual produce resilient spores, specialized for host-to-host transmission (19). These fascinating observations prompted us to wonder whether bacteriophages can influence the dynamics of spore formation in the human gut microbiota and therefore influence the dynamics of host-to-host transmission of intestinal bacteria.

We hence queried the HMMs of Rap (PFAM PF18801), Spo0E (PFAM PF09388) and AbrB (SMART SM00966) against the protein sequences predicted from all the MAGs of bacteriophages present in the Gut Phage Database (13). This HMMsearch revealed 1 match for Rap, 172 for Spo0E and 861 for AbrB (E-value < 1E-5), hinting at likely phage-mediated sporulation regulations in the human gut microbiota (**Table S4**). The RRNPP-type signature furthermore led to the identification of 17 candidate QSSs in MAGs of intestinal bacteriophages, distributed in 10 families, of which only 2 were previously described in this study: family QSS14 (Prophage of *Blautia*) and the Rap-Phr family (**Fig. 3B** and **Table S5**). Altogether, these computational results suggest, for the first time, that intestinal bacteriophages interfere with the sporulation of intestinal bacteria and thereby influence the dynamics of transmissibility of bacteria between humans.

DISCUSSION

If bacterial quorum sensing was discovered in 1970 (46), the first characterization of a functional QSS in a bacteriophage only dates back to 2017, where its was shown to coordinate the lysis-lysogeny transition as a function of phage densities (4). Evidence has emerged only recently that bacteriophages may use or exploit quorum sensing mechanisms to interfere, for their evolutionary

394 benefit, with the biology of bacterial hosts (9,47). These findings open fascinating perspectives
395 that will significantly enrich extant models or bacteria-phages co-evolution. However, the diversity
396 of viral QSSs remains tremendously overlooked, with only two “arbitrium” families (4,5) and the
397 rap-phr family (9). Here, using a signature-based computational approach, we were able to identify
398 22 families of candidate RRNPP-type QSSs that were never described before in bacteriophages:
399 14 in reference genomes from the NCBI (**Tables S1** and **S3**) and 8 additional in MAGs of interstinal
400 bacteriophages from the Gut Phage Database (13) (**Table S5**). Altogether, our results might thus
401 expand the known diversity of viral QSS families by 7-fold. Our computational results therefore
402 pave the way to an exciting research on the characterization of novel density-dependent social
403 processes in bacteriophages, which would unravel unknown decision-making processes in viruses.
404
405 Analyzing the genomic context of our viral candidate RRNPP-type QSSs to predict their
406 function, we found that the regulation of sporulation by a modulation of Spo0A phosphorylation is
407 well represented. In a recent study, we reported that the Rap-Phr RRNPP-type QSS family, known
408 to regulate the competence and sporulation initiation pathways in *Bacillus* bacteria, can actually be
409 carried by (pro)phages (9). Building on this previous work, we now unraveled a massive unknown
410 abundance and diversity of viral Rap-Phr QSSs, in both (pro)phages of the *Bacillus subtilis* and
411 *Bacillus cereus* groups (**Fig. 2**, **Table S1**). Furthermore, we discovered that (pro)phage-encoded
412 QSSs can dynamically manipulate the host biology beyond the sole *Bacillus* genus, and beyond
413 the sole Rap-Phr mechanism. Indeed, we identified 7 (pro)phage-encoded candidate RRNPP-type
414 QSSs (coined QSSs 1α, 3α, 3β, 3γ, 4β, 5α and g) predicted to regulate the expression of a
415 (pro)phage-encoded sporulation regulator (*spo0E* or *abrB*) (**Fig. 3**). Moreover, we found in the
416 literature experimental data reporting that QSSs 3α, 3β, 3γ and g affect the timing of sporulation in
417 their respective host. Because the receptors of the Rap-Phr, 1α, 3α, 3β, 3γ, 4β, 5α and g QSSs are
418 distributed into 6 different gene families, and are encoded by (pro)phages of different hosts, our
419 results highlight, for the first time in bacteriophages, a remarkable convergent evolution of density-
420 dependent mechanisms of manipulation of a substantial spectrum of the bacterial biology: from
421 competence (18) (**Fig. 5**) to Spo0A-P target pathways (sporulation, biofilm formation, toxin
422 production or solventogenesis (21,22,43)) (**Fig. 4**).
423
424 These findings would have major implications, both fundamental and applied. For instance,
425 these sporulation- and competence-modulating QSSs in bacteriophages of *Bacillus*, *Clostridium*
426 and *Brevibacillus* bacteria could shed some new light on the molecular mechanisms underlying
427 antibiotic-resistance and host-to-host transmission of bacteria, with potential practical applications.
428 Indeed, bacterial competence is well known to contribute to the spread of antibiotic resistance
429 genes whereas sporulation is a developmental program through which many bacteria become
430 transmissible and resistant to a wide range of chemical products, including antibiotics.

Consistently, the sporulation of pathogenic bacteria represents a serious threat for human health. For instance, it is under the form of endospores that *B. cereus* causes anthrax and food poisoning and that *C. botulinum*, *C. perfringens* and *C. difficile* cause food poisoning, wood infection and intestinal diarrhea, respectively (27,28). Hence, understanding which and how bacteriophages dynamically modulate the competence and sporulation initiation pathways in Firmicutes bacteria might open fascinating perspectives in microbiology, medicine and food industry.

Our results also have a fundamental implication. They challenge the sporulation paradigm, which assumes that bacteria sporulate in adverse circumstances and implies that only bacterial genes govern the sporulation decision-making process. Indeed, our computational survey invites to reconsider the sporulation decision-making process as a biological process falling under the scope of the (pro)phage-host collective, rather than a strict bacterial process of last resort. With this regard, it is interesting to note that non-sporulating bacteria have been observed to form spores when they are lysogenized by “spore-converting bacteriophages” (48–50). The converse case, namely, the impairment of the capacity to sporulate caused by a prophages has also been observed (51). Either way, these previously described activation or inhibition of the host sporulation pathway by prophages happened to be constitutive and were therefore not the result of a decision-making process, unlike the dynamical modulation of the sporulation pathway described in this study, which is indeed predicted to be function of (pro)phage densities. Adopting an evolutionary perspective provides original explanations on why (pro)phages may dynamically manipulate bacterial sporulation. *Erez et al.* brilliantly demonstrated that the viral “arbitrium” QSS coordinates the transition from the lytic cycle to the host-protective lysogenic cycle at high concentration of arbitrium peptide (*i.e.* high phage densities), when a lot of host cells have been lysed and the phage-host collective likely needs to be protected (4). On this basis, we can predict that the manipulative phage-encoded candidate QSSs described in our study function according to the same principle and optimize the trade-off between the replication of the phage and the protection of the phage-host collective. Specifically, they could i) hijack the host sporulation/competence pathway when densities of intracellular phage genomes reflect best timings for phages to maximize their fitness irrespective of the fitness of their hosts, and ii) alleviate this manipulation when phage densities reflect a benefit in letting hosts enact the survival/adaptive sporulation/competence pathways. For instance, at low densities of free phages, when only a few bacteria are lysed and the host population is not yet endangered, we can hypothesize that it might be beneficial for phages to inhibit the sporulation/competence pathways for the following reasons. First, a phage genome that is inside a sporulating will not be replicated by the cell (52). Second, the sporulation initiation pathway can trigger cannibalistic behaviors that may kill neighbor cells and thereby reduce opportunities for phages to replicate their genome (22,43). Third, the competence pathway is proposed to enable a bacterium to pick up from the environmental pangenome the CRISPR-cas

system that specifically targets the phage DNA (53). However, when phages have well replicated at the expense of their hosts, and when the survival of the phage-host collective is thus likely compromised, it might then be best for phages to alleviate the manipulation of host survival mechanisms. Indeed, under harsh environmental conditions, a time might eventually come when it would be more advantageous for intracellular phage genomes to be protected inside bacterial endospores rather than to keep promoting phage replication through sporulation inhibitions. In addition to, or in place of their effects during the lytic cycle, we could also consider that the viral QSSs described in this study may have been selected because of their beneficiary effects during the lysogenic cycle. In light of this perspective, these viral QSSs could be considered as adaptive genes for the host, conferring an evolutionary advantage upon the prophage-host collective relative to non-lysogenized bacteria (54,55). For instance, as medium levels of Spo0A-P can enact the biofilm pathway, prophage-encoded QSSs that antagonize Spo0A-P accumulation could provide the lysogenized subpopulation with a means to temporarily delay the production of biofilm molecules, hence temporarily increasing the fitness of the prophage-host collective (55).

CONCLUSION

In light of the density-dependent host-hijacking mechanisms discussed in this study, many (pro)phage-encoded QSSs are likely to be extremely sophisticated regulatory systems, that can subtly modulate the biology of a (pro)phage-host collective. Their in-silico identification constitutes a fundamental step towards refining models of phage-host co-evolution, discovering novel decision-making processes in bacteriophages, and foremost understanding the fundamental molecular mechanisms underlying bacterial sporulation and competence, with major theoretical and practical outcomes. The next step will naturally be to experimentally characterize the viral candidate QSSs described in this study. Accordingly, we provided all the NCBI identifiers of the pro-peptide and receptor proteins of our candidate (pro)phage-encoded QSSs in the main and supplementary tables. We designed this survey to make it as easy as possible for experimentalists to build on further functional studies, as we believe that this work has the potential to open many fascinating perspectives in many different areas of biology.

METHODS

Construction of the RRNPP-type signature

We carefully mined the literature to identify all experimentally characterized RRNPP-QSSs from different families, fetch their representative sequence on the NCBI (56), visualize their genomic context, and analyse their similarities to delineate decision rules for the detection of candidate RRNPP-type QSSs. This dataset was composed of the following reference QSSs: *rapA-phrA*, *nprR-nprX*, *plcR-papR*, *rgg2-shp2*, *aimR-aimP*, *prgX-prgQ* and *traA-iPD1*. The extreme values in the lengths of these experimentally validated receptors and pro-peptides (**Fig. S2**) were used as

references to define ranges of acceptable lengths for candidate receptors and pro-peptides. PlcR being the shortest receptor (285aa) and NprR being the longest (423aa), we established **rule n°1** that candidate receptors must have a length comprised between 250aa and 460aa. Likewise, on the basis that Shp2 is the shortest pro-peptide (21aa) and AimP is the longest (49aa), **rule n°2** poses that candidate pro-peptides must have a length comprised between 15aa and 65aa. Because the genes encoding the receptor and the propeptide are always directly adjacent to each other in reference RRNPP QSSs (**Fig. S2**), **rule n°3** poses that the two genes of RRNPP-type candidate QSSs must be direct neighbors. Next, using InterProScan version *5.36-75.0* (57), the protein sequences of reference RRNPP-type receptors were queried against the InterPro database of structural motifs to identify HMMs of tetratricopeptide repeats (TPRs) and DNA binding domains that are characteristic of these proteins. These HMMs (displayed in **Fig. S2**) were further retrieved and compiled in two distinct libraries, using hmmpress from the HMMER suite version *3.2.1* (58). This allowed defining **rule n°4** that a candidate receptor must be matched by at least one HMM of the library of TPRs found within reference receptors (E-value<1E-5, 1000x times more stringent than default inclusion threshold). The HMM library of DNA binding domains was designed to predict whether a candidate receptor might function as an intracellular transcription factor. Finally, SignalP version *5.0b Linux x86_64* was run with the option '-org gram+' against the reference RRNPP-type pro-peptides to illustrate the reliability of this software to predict the SEC-dependent excretion of small quorum sensing peptides (59). Indeed, only the PrgQ and Shp reference pro-peptides were not predicted by SignalP to harbor a N-terminal signal sequence addressed to the SEC-translocon (**Fig. S2**), consistent with the fact that they are the only RRNPP-type pro-peptides mentioned to be exported via another secretion system, namely the ABC-type transporter PptAB (12). This legitimized the use of SignalP to establish **rule n°5** that a candidate pro-peptide must be predicted by SignalP to be secreted via the SEC-translocon.

Construction of the target datasets

The complete genomes of Viruses and Firmicutes were queried from the NCBI 'Assembly' database (56), as of 28/04/2020 and 10/04/2020, respectively. The features tables (annotations) and the encoded protein sequences of these genomes were downloaded using 'GenBank' as source database. The Gut Phage Database (13) was downloaded as of 29/10/2020, from the following url: http://ftp.ebi.ac.uk/pub/databases/metagenomics/genome_sets/gut_phage_database/

Detection of RRNPP-type candidate QSSs

We launched the systematical search of the RRNPP-type signature independently against i) the complete genomes of Viruses and Firmicutes available on the NCBI and ii) the MAGs of bacteriophages from the Gut Phage Database. **Step n°1** consisted in reducing the search space by sub-setting all the protein sequences of a dataset into two libraries: a library 'potential receptors'

containing the protein sequences of length between 250aa and 460aa and a library 'potential pro-peptides' containing the protein sequences of length between 15 and 65aa. **Step n°2** further reduced the search space by filtering all the proteins from one library whose coding sequence was not directly adjacent with the coding sequence of a protein from the other library. In **step n°3**, an HMMsearch of the HMM library of TPRs was launched with HMMER version *3.2.1* against the remaining 'potential receptors' and only the sequences matched by a HMM with an E-value < 1E-5 (1000x times more stringent than default inclusion threshold) were conserved. Another coding sequence adjacency filter was applied in **step n°4** to reduce the search space in the 'potential pro-peptides' library. **Step n°5** filtered out all the remaining 'potential pro-peptides' that were not predicted by SignalP to be secreted via the SEC-translocon. At last, the two libraries were intersected to define candidate QSSs based on coding sequence adjacency. If a candidate receptor happened to be flanked on both sides by two pro-peptides (or vice-versa), therefore if a protein happened to be assigned to two distinct QSSs, only the QSS with the smallest intergenic distance between the two genes was retained. Eventually, QSSs with intergenic distance >600 bp were filtered out. Of the total of 2718 QSSs detected after the intersection of the two libraries, only 19 have been discarded by these ultimate filtering criteria. As a post-processing step, an HMMsearch of the HMM library of DNA binding domains was launched against the candidate RRNPP-type receptors to identify the receptors that are susceptible to be transcriptional regulators (E-value < 1E-5).

Classification of the candidate QSSs into families

Because quorum sensing pro-peptides offer few amino acids to compare, are versatile and subjected to intragenic duplication (35), we classified the QSSs based on sequence homology of the receptors. We launched a BLASTp (34) All vs All of the receptors of the 2681 candidate QSSs identified in the complete genomes of Viruses and Firmicutes. The output of BLASTp was filtered to retain only the pairs of receptors giving rise to at least 30% sequence identity over more than 80% of the length of the two proteins. These pairs were used to build a sequence similarity network and the families were defined based on the connected components of the graph (mean clustering coefficient of connected components=0.97).

Identification of already known families

A BLASTp search was launched using as queries the RapA (NP_389125.1), NprR (WP_001187960.1), PlcR (WP_000542912.1), Rgg2 (WP_002990747.1), AimR (APD21232.1), AimR-like (AID50226.1), PrgX (WP_002366018.1), TraA (BAA11197.1) reference receptors, and as a target database, the 2681 candidate QSS receptors found in complete genomes of Viruses and Firmicutes. If the best hit of a reference RRNPP-type receptor gave rise to a sequence identity

>= 30% over more than 80% mutual coverage, then the family to which this best hit belongs is considered as an already known family (**Table S2**).

Prophage detection

All the NCBI ids of the genomic accessions of chromosomes or plasmids of Firmicutes encoding one or several candidate QSSs were retrieved and automatically submitted to the Phaster webtool (32). Eventually, each QSS was defined as viral if its genomic coordinates on a given chromosome/plasmid fell within a region predicted by Phaster to belong to a prophage (qualified as either 'intact', 'questionable' or 'incomplete' prophage). Phaster results were complemented by ProphageHunter (33), a webtool that computes the likelihood that a prophage is active (able to reinitiate the lytic cycle by excision). Because ProphageHunter cannot be automatically queried, we only called upon this webtool for chromosomes/plasmids which encode QSSs that are not part of the biggest families, namely Rap-Phr (2257 candidate QSSs) and PlcR-PapR (223 candidate QSSs). Likewise, coordinates of candidate QSSs were eventually intersected with predicted prophage regions to detect potential prophage-encoded candidate QSSs that could have been missed by Phaster (**Table 1**).

Prediction of the mature quorum sensing peptides

For each uncharacterized families of candidate receptors of size >1 with at least one (pro)phage-encoded member, the cognate pro-peptides were aligned in a multiple sequence alignment (MSA) using MUSCLE *version 3.8.31* (60). Each MSA was visualized with Jalview *version 1.8.0_201* under the ClustalX color scheme which colors amino acids based on residue type conservation (61). The region of RRNPP-type pro-peptides encoding the mature quorum sensing peptide usually corresponds to a small sequence (5-6aa), located in the C-terminal of the pro-peptide, with conserved amino-acids types in at least 3 positions (4,9,37,45). Based on the amino-acid profile of C-terminal residues in each MSA, putative mature quorum sensing peptides were manually determined (**Fig. S4**).

Phylogenetic trees of the Rap, QSS1R, QSS2R and QSS3R families

For each family shared between chromosomes and (pro)phages, a multiple sequence alignment (MSA) of the protein sequences of the receptors was performed using MUSCLE *version 3.8.31* (60). Each MSA was then trimmed using trimmal *version 1.4.rev2*2 with the option '-automated1', optimized for maximum likelihood phylogenetic tree reconstruction (62). Each trimmed MSA was then given as input to IQ-TREE *version multicore 1.6.10* to infer the maximum likelihood phylogenetic tree of the corresponding family under the LG+G model with 1000 ultrafast bootstraps (63). Each tree was further edited via the Interactive Tree Of Life (ITOL) online tool (**Fig. 2 and S3**) (64).

Analysis of the genomic context of QSSs

The genomic context of the 20 (pro)phage-encoded candidate QSSs from unknown families were visualized using the nucleotide graphic report on the NCBI. We systematically retrieved the functional annotation of adjacent genes, and analyzed their sequences with a “Conserved Domains” search as well as a BLASTp search against the NR (non-redundant) protein database maintained by the NCBI. The genomic contexts of predicted sporulation-regulating QSSs are shown in **Fig. 3**.

Identification of rap, spo0E and AbrB genes in the Gut Phage Database

With HMMER, we launched an HMM search of reference HMMs of Rap (PFAM PF18801), Spo0E (PFAM PF09388) and AbrB (SMART SM00966) against all the protein sequences predicted from the ORFs of the MAGs from the Gut Phage Database. The hits were retained only if they gave rise to an E-value < 1E-5 (**Table S4**).

ABBREVIATIONS:

- HMMs: Hidden Markov Models
- MAGs: Metagenomics-Assembled-Genomes
- MGEs: Mobile Genetic Elements
- NCBI: National Center for Biotechnology Information
- Phages: Bacteriophages
- QSSs: Quorum Sensing Systems
- RRNPP: Rap, Rgg, NprR, PlcR and PrgX families of QSS receptors
- TPRs: TetratricoPeptide Repeats

AUTHOR CONTRIBUTIONS

C.B, Y.L, E.B and P.L conceived the study. C.B performed the analyses. C.B, Y.L and E.B wrote the manuscript with input from all authors. All documents were edited and approved by all authors.

DECLARATIONS

Ethis approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All the NCBI or Gut Phage Database IDs of the proteins discussed in this manuscript are available in the supplementary tables.

Competing Interests

The authors of this manuscript have no competing interests to disclose.

Fundings

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. C. Bernard was supported by a PhD grant from the Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.

Authors' Contribution

C.B, Y.L, E.B and P.L conceived the study. C.B performed the analyses. C.B, Y.L and E.B wrote the manuscript with input from all authors. All documents were edited and approved by all authors.

Acknowledgments

We would like to thank Dr. A. K. Watson for critical reading and discussion.

REFERENCES

1. Papenfort K, Bassler BL. Quorum sensing signal-response systems in Gram-negative bacteria. Nat Rev Microbiol [Internet]. 2016 [cited 2019 May 11];14(9):576–88. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27510864

2. Bhatt VS. Quorum sensing mechanisms in gram positive bacteria. In: Implication of Quorum Sensing System in Biofilm Formation and Virulence. Springer Singapore; 2019. p. 297–311.

3. Banderas A, Carcano Id A, Sia Id E, Li S, Lindnerid AB. Ratiometric quorum sensing governs the trade-off between bacterial vertical and horizontal antibiotic resistance propagation. 2020 [cited 2021 Feb 15]; Available from: https://doi.org/10.1371/journal.pbio.3000814

4. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, et al. Communication between viruses guides lysis-lysogeny decisions. Nature [Internet]. 2017 [cited 2019 Jul 4];541(7638):488–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28099413

5. Stokar-Avihail A, Tal N, Erez Z, Lopatina A, Sorek R. Widespread Utilization of Peptide Communication in Phages Infecting Soil and Pathogenic Bacteria. Cell Host Microbe [Internet]. 2019 May 8 [cited 2019 Nov 24];25(5):746-755.e5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31071296

6. Fuqua WC, Winans SC, Greenberg EP. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J Bacteriol [Internet]. 1994 Jan [cited 2019 Sep 25];176(2):269–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8288518

7. Perez-Pascual D, Monnet V, Gardan R. Bacterial Cell-Cell Communication in the Host via RRNPP Peptide-Binding Regulators. Front Microbiol [Internet]. 2016 [cited 2019 Oct 16];7:706. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27242728

8. Clokie MRJ, Millard AD, Letarov A V., Heaphy S. Phages in nature. Bacteriophage [Internet]. 2011 Jan 22 [cited 2019 Dec 19];1(1):31–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21687533

9. Bernard C, Li Y, Lopez P, Bapteste E. Beyond arbitrium: identification of a second communication system in Bacillus phage phi3T that may regulate host defense mechanisms. ISME J [Internet]. 2020; Available from: http://dx.doi.org/10.1038/s41396-020-00795-9

10. Rocha-Estrada J, Aceves-Diez AE, Guarneros G, de la Torre M. The RNPP family of quorum-sensing proteins in Gram-positive bacteria. Appl Microbiol Biotechnol [Internet]. 2010 Jul 26 [cited 2019 Oct 16];87(3):913–23. Available from: http://link.springer.com/10.1007/s00253-010-2651-y

11. 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 [Internet]. 2016 Jul 17 [cited 2019 Oct 16];428(14):2793–804. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27283781

12. Neiditch MB, Capodagli GC, Prehna G, Federle MJ. Genetic and Structural Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria [Internet]. Vol. 51, Annual Review of Genetics. Annual Reviews Inc.; 2017 [cited 2020 Nov 19]. p. 311–33. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-120116-023507

13. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. Cell [Internet]. 2021 Feb [cited 2021 Mar 12];184(4):1098-1109.e9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867421000726

14. Kohler V, Keller W, Grohmann E. Regulation of gram-positive conjugation [Internet]. Vol. 10, Frontiers in Microbiology. Frontiers Media S.A.; 2019 [cited 2021 Mar 12]. p. 1134. Available from: www.frontiersin.org

15. Bischofs IB, Hug JA, Liu AW, Wolf DM, Arkin AP. Complexity in bacterial cell- cell communication: Quorum signal integration and subpopulation signaling in the Bacillus subtilis phosphorelay. Proc Natl Acad Sci U S A. 2009;106(16):6459–64.

16. Shafikhani SH, Leighton T. AbrB and Spo0E Control the Proper Timing of Sporulation in Bacillus subtilis. Curr Microbiol. 2004;48(4):262–9.

17. Schultz D, Wolynes PG, Jacob E Ben, Onuchic JN. Deciding fate in adverse times: Sporulation and competence in Bacillus subtilis. Proc Natl Acad Sci U S A. 2009;106(50):21027–34.

18. Schultz D, Lu M, Stavropoulos T, Onuchic J, Ben-Jacob E. Turning oscillations into opportunities: Lessons from a bacterial decision gate. Sci Rep. 2013;3.

19. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing of “unculturable” human microbiota reveals novel taxa and extensive sporulation. Nature [Internet]. 2016 May 4 [cited 2020 Oct 6];533(7604):543–6. Available from: https://www.nature.com/articles/nature17645

20. Galperin MY. Genome Diversity of Spore-Forming Firmicutes. Microbiol Spectr [Internet]. 2013 Dec 27 [cited 2020 Nov 6];1(2):TBS-0015-2012. Available from: /pmc/articles/PMC4306282/?report=abstract

21. Dürre P, Böhringer M, Nakotte S, Schaffer S, Thormann K, Zickner B. Transcriptional regulation of solventogenesis in Clostridium acetobutylicum. In: Journal of Molecular Microbiology and Biotechnology [Internet]. 2002 [cited 2021 Feb 12]. p. 295–300. Available from: https://europepmc.org/article/med/11931561

22. González-Pastor JE, Hobbs EC, Losick R. Cannibalism by Sporulating Bacteria. Science (80-). 2003;301(July):510–3.

23. Centers for Disease Control U. Antibiotic Resistance Threats in the United States, 2019. [cited 2021 Feb 12]; Available from: http://dx.doi.org/10.15620/cdc:82532.

24. Wilcox MH, Fawley WN. Hospital disinfectants and spore formation by Clostridium difficile. Lancet. 2000 Oct 14;356(9238):1324.

25. Schuch R, Nelson D, Fischetti VA. A bacteriolytic agent that detects and kills Bacillus anthracis. Nature [Internet]. 2002 Aug 22 [cited 2021 Mar 12];418(6900):884–9. Available from: https://www.nature.com/articles/nature01026

26. Anthrax in Humans and Animals [Internet]. Anthrax in Humans and Animals. World Health Organization; 2008 [cited 2020 Nov 19]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26269867

27. Mallozzi M, Viswanathan VK, Vedantam G. Spore-forming Bacilli and Clostridia in human disease [Internet]. Vol. 5, Future Microbiology. Future Medicine Ltd.; 2010 [cited 2021 Apr 23]. p. 1109–23. Available from: https://pubmed.ncbi.nlm.nih.gov/20632809/

28. Postollec F, Mathot AG, Bernard M, Divanac'h ML, Pavan S, Sohier D. Tracking spore-forming bacteria in food: From natural biodiversity to selection by processes. Int J Food Microbiol [Internet]. 2012 Aug 1 [cited 2021 Mar 12];158(1):1–8. Available from: https://pubmed.ncbi.nlm.nih.gov/22795797/

29. Swick MC, Koehler TM, Driks A. Surviving Between Hosts: Sporulation and Transmission. Microbiol Spectr. 2016 Aug 18;4(4).

30. Khanna S, Pardi DS, Kelly CR, Kraft CS, Dhere T, Henn MR, et al. A Novel Microbiome Therapeutic Increases Gut Microbial Diversity and Prevents Recurrent Clostridium difficile Infection. J Infect Dis [Internet]. 2016 Jul 15 [cited 2020 Nov 9];214(2):173–81. Available from: https://academic.oup.com/jid/article/214/2/173/2572105

31. Voichek M, Maaß S, Kroniger T, Becher D, Sorek R. Peptide-based quorum sensing systems in Paenibacillus polymyxa. Life Sci Alliance [Internet]. 2020 Oct 1 [cited 2021 Apr 22];3(10). Available from: https://doi.org/10.26508/lsa.202000847

32. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res [Internet]. 2016 [cited 2020 Aug 4];44(Web Server issue):W16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987931/

33. Song W, Sun HX, Zhang C, Cheng L, Peng Y, Deng Z, et al. Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Res [Internet]. 2019 Jul 1 [cited 2021 Mar 15];47(W1):W74–80. Available from: https://pro-hunter.

34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol [Internet]. 1990 Oct 5 [cited 2019 Oct 16];215(3):403–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2231712

35. 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. Gore J, editor. PLOS Biol [Internet]. 2016 Dec 29 [cited 2019 Oct 17];14(12):e2000330. Available from: http://dx.plos.org/10.1371/journal.pbio.2000330

36. Reizer J, Reizer A, Perego M, Saier MH. Characterization of a Family of Bacterial Response Regulator Aspartyl-Phosphate (RAP) Phosphatases. Microb Comp Genomics [Internet]. 1997 Jan [cited 2019 Oct 16];2(2):103–11. Available from: http://www.liebertpub.com/doi/10.1089/omi.1.1997.2.103

37. Pottathil M, Lazazzera BA. The extracellular PHR peptide-Rap phosphatase signaling circuit of bacillus subtilis. Front Biosci [Internet]. 2003 Jan 1 [cited 2019 Oct 16];8(4):913. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12456319

38. Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P. Resistance of Bacillus Endospores to Extreme Terrestrial and Extraterrestrial Environments. Microbiol Mol Biol Rev [Internet]. 2000 Sep 1 [cited 2020 Oct 6];64(3):548–72. Available from: /pmc/articles/PMC99004/?report=abstract

39. Tan IS, Ramamurthi KS. Spore formation in Bacillus subtilis [Internet]. Vol. 6, Environmental Microbiology Reports. Wiley-Blackwell; 2014 [cited 2020 Nov 19]. p. 212–25. Available from: /pmc/articles/PMC4078662/?report=abstract

40. Al-Hinai MA, Jones SW, Papoutsakis ET. The Clostridium Sporulation Programs: Diversity and Preservation of Endospore Differentiation. Microbiol Mol Biol Rev [Internet]. 2015 Mar 1 [cited 2020 Nov 6];79(1):19–37. Available from: http://mmbr.asm.org/

41. Fujita M, Losick R. Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. Genes Dev. 2005 Sep 15;19(18):2236–44.

42. Bischofs IB, Hug JA, Liu AW, Wolf DM, Arkin AP. Complexity in bacterial cell-cell communication: quorum signal integration and subpopulation signaling in the Bacillus subtilis phosphorelay. Proc Natl Acad Sci U S A [Internet]. 2009 Apr 21 [cited 2019 Oct 20];106(16):6459–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19380751

43. González-Pastor JE. Cannibalism: A social behavior in sporulating Bacillus subtilis. FEMS Microbiol Rev. 2011;35(3):415–24.

44. Feng J, Zong W, Wang P, Zhang ZT, Gu Y, Dougherty M, et al. RRNPP-Type quorum-sensing systems regulate solvent formation, sporulation and cell motility in Clostridium saccharoperbutylacetonicum. Biotechnol Biofuels [Internet]. 2020;13(1):1–16. Available from: https://doi.org/10.1186/s13068-020-01723-x

45. Kotte AK, Severn O, Bean Z, Schwarz K, Minton NP, Winzer K. RRNPP-type quorum sensing affects solvent formation and sporulation in clostridium acetobutylicum. Microbiol (United Kingdom). 2020;166(6):579–92.

46. Nealson KH, Platt T, Hastings JW. Cellular control of the synthesis and activity of the bacterial luminescent system. J Bacteriol [Internet]. 1970 Oct [cited 2019 Sep 25];104(1):313–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/5473898

47. Silpe JE, Bassler BL. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. Cell [Internet]. 2019 Jan 10 [cited 2019 Jun 12];176(1–2):268-280.e13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30554875

48. Boudreaux DP, Srinivasan VR. Bacteriophage-induced Sporulation in Bacillus cereus T. Journal of General Microbiology.

49. Bramucci MG, Keggins KM, Lovett PS. Bacteriophage conversion of spore-negative mutants to spore-positive in Bacillus pumilus. J Virol [Internet]. 1977 [cited 2021 Apr 23];22(1):194–202. Available from: /pmc/articles/PMC515700/?report=abstract

50. Silver-Mysliwiec TH, Bramucci MG. Bacteriophage-enhanced sporulation: Comparison of spore-converting bacteriophages PMB12 and SP10. J Bacteriol [Internet]. 1990 [cited 2021 Apr 23];172(4):1948–53. Available from: /pmc/articles/PMC208690/?report=abstract

51. Schuch R, Fischetti VA. The secret life of the anthrax agent Bacillus anthracis: bacteriophage-mediated ecological adaptations. PLoS One [Internet]. 2009 Aug 12 [cited 2019 Dec 4];4(8):e6532. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19672290

52. Meijer WJ, Castilla-Llorente V, Villar L, Murray H, Errington J, Salas M. Molecular basis for the exploitation of spore formation as survival mechanism by virulent phage φ29. EMBO J [Internet]. 2005 Oct 19 [cited 2019 Oct 21];24(20):3647–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16193065

53. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource [Internet]. 2020 [cited 2021 Mar 15]. Available from: www.nature.com/nrmicro

54. Gallegos-Monterrosa R, Christensen MN, Barchewitz T, Koppenhöfer S, Priyadarshini B, Bálint B, et al. Impact of Rap-Phr system abundance on adaptation of Bacillus subtilis. Commun Biol [Internet]. 2021 Dec [cited 2021 Apr 24];4(1). Available from: https://pubmed.ncbi.nlm.nih.gov/33850233/

55. Kalamara M, Spacapan M, Mandic-Mulec I, Stanley-Wall NR. Social behaviours by Bacillus subtilis: quorum sensing, kin discrimination and beyond. Mol Microbiol [Internet]. 2018 [cited 2019 Oct 16];110(6):863. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30218468

56. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res [Internet]. 2016 Jan 4 [cited 2019 May 28];44(D1):D7–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26615191

57. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics [Internet]. 2014 May 1 [cited 2021 Mar 25];30(9):1236–40. Available from: /pmc/articles/PMC3998142/

58. Eddy SR. Accelerated Profile HMM Searches. Pearson WR, editor. PLoS Comput Biol [Internet]. 2011 Oct 20 [cited 2019 Oct 16];7(10):e1002195. Available from: https://dx.plos.org/10.1371/journal.pcbi.1002195

59. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol [Internet]. 2019 Apr 18 [cited 2019 Oct 16];37(4):420–3. Available from: http://www.nature.com/articles/s41587-019-0036-z

60. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res [Internet]. 2004 [cited 2019 May 28];32(5):1792–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15034147

61. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics [Internet]. 2009 May 1 [cited 2019 May 28];25(9):1189–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19151095

62. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics [Internet]. 2009 Aug 1 [cited 2019 May 28];25(15):1972–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19505945

63. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol [Internet]. 2015 Jan 1 [cited 2019 May 28];32(1):268–74. Available from: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300

64. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res [Internet]. 2019 Jul 1 [cited 2021 Mar 26];47(W1):W256–9. Available from: https://academic.oup.com/nar/article/47/W1/W256/5424068

65. Perchat S, Talagas A, Zouhir S, Poncet S, Bouillaut L, Nessler S, et al. NprR, a moonlighting quorum sensor shifting from a phosphatase activity to a transcriptional activator. Microb Cell [Internet]. 2016 Nov 1 [cited 2021 Apr 24];3(11):573–5. Available from: https://pubmed.ncbi.nlm.nih.gov/28357327/

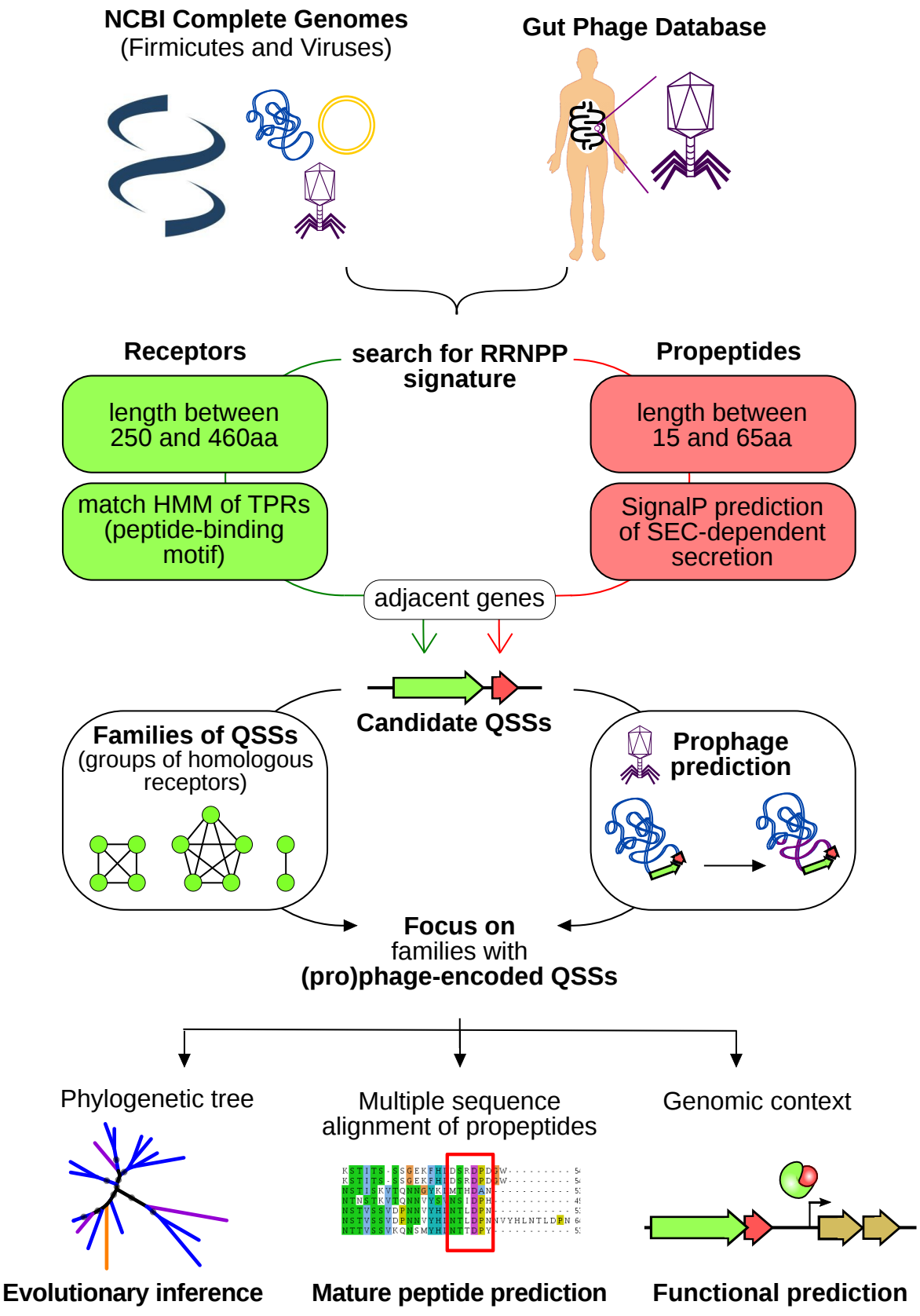

NCBI Complete Genomes
(Firmicutes and Viruses)
Gut Phage Database
search for RRNPP signature
Receptors
length between 250 and 460aa
match HMM of TPRs (peptide-binding motif)
Propeptides
length between 15 and 65aa
SignalP prediction of SEC-dependent secretion
adjacent genes
Candidate QSSs
Families of QSSs
(groups of homologous receptors)
Prophage prediction
Focus on
families with
(pro)phage-encoded QSSs
Phylogenetic tree
Multiple sequence alignment of propeptides
Genomic context
Evolutionary inference
Mature peptide prediction
Functional prediction

FIGURE LEGENDS

Figure 1: Study design.

The RRNPP-type signature was independently queried against the complete genomes of Firmicutes and Viruses from the NCBI and against the Gut Phage Database. This consisted of only retaining pairs of adjacent genes encoding a medium-length protein matched by HMM models of TPRs (the candidate receptor) and a small protein predicted by SignalP to harbor a N-terminal signal sequence for the SEC-translocon (the candidate pro-peptide), respectively. The candidate QSSs were further classified into families, defined as groups of homologous receptors in a BLASTp all vs all. This study further focused on families in which at least one QSS is encoded by a phage or a genomic region predicted by Phaster and/or ProphageHunter to belong to a prophage inserted within a bacterial genome. Subsequently, each QSS family with viral representatives was computationally characterized. Protein families of receptors shared between bacterial genomes and phage genomes were aligned, trimmed and given as input to IQ-TREE to construct phylogenetic trees in order to visualize if and how QSSs travel onto different kinds of genetic supports (chromosomes, plasmids, phage genomes) rather than to stay in their hosts lineages. In QSS families comprising more than one QSS, the propeptides were also aligned and visualized with Jalview to predict the sequence of each mature peptide. Finally, as RRNPP-type receptors that are transcription factors tend to regulate adjacent genes, the genomic neighborhood of each (pro)phage-encoded receptor with a detected DNA binding domain was analyzed to predict the functions regulated by the QSS.

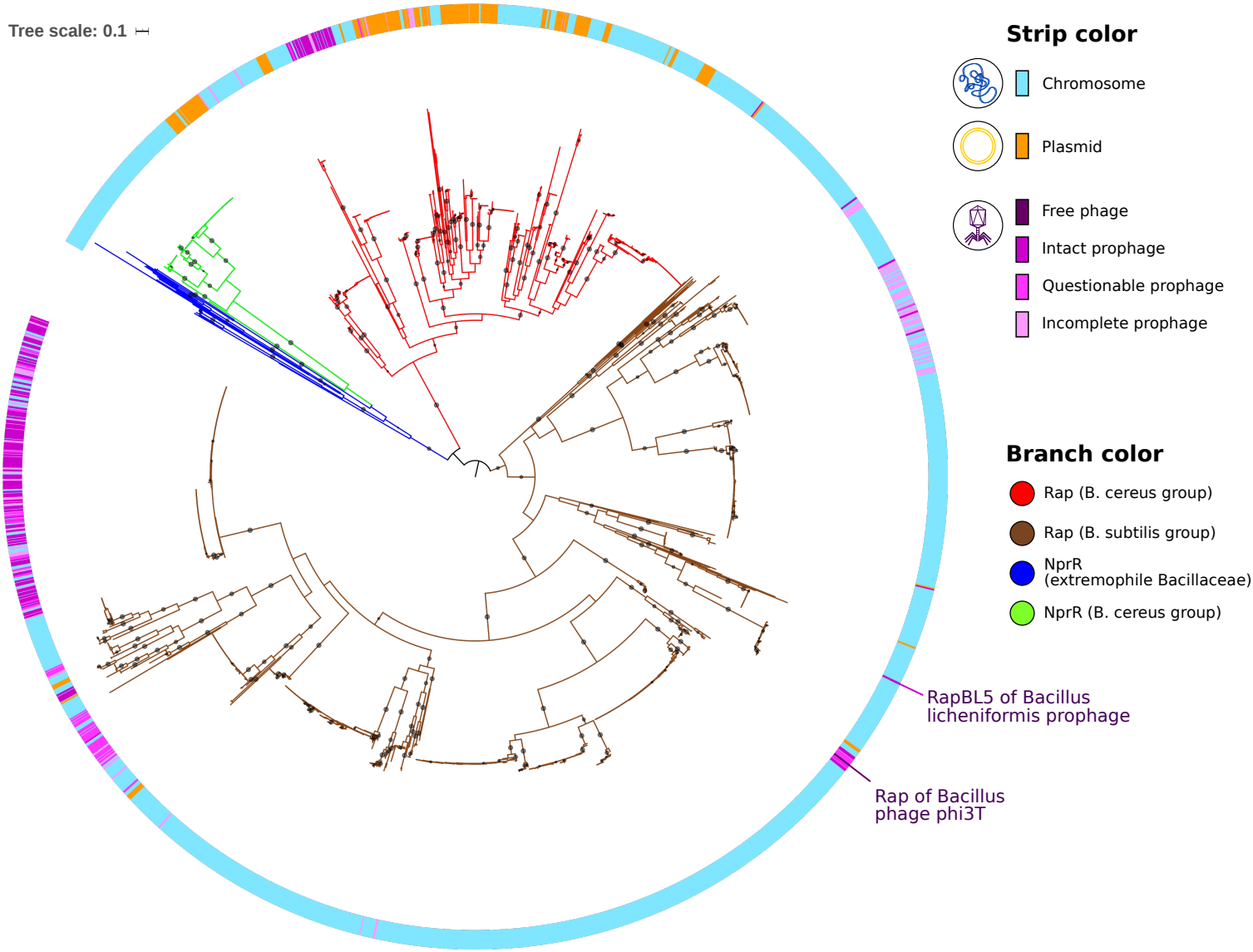

Tree scale: 0.1
Strip color
Chromosome
Plasmid
Free phage
Intact prophage
Questionable prophage
Incomplete prophage
Branch color
Rap (B. cereus group)
Rap (B. subtilis group)
NprR
(extremophile Bacillaceae)
NprR (B. cereus group)
RapBL5 of Bacillus
licheniformis prophage
Rap of Bacillus
phage phi3T

Figure 2: Polyphily of viral Rap-Phr QSSs linked to sporulation regulation.

The figure displays the maximum-likelihood phylogenetic tree of the family comprising the Rap (no DNA binding domain) and the NprR (DNA binding domain) receptors that are part of a detected RRNPP-type QSS. The clustering of Rap and NprR into the same protein family is consistent with the common phylogenetic origin proposed for these receptors (65). The tree was midpoint rooted and a small black circle at the middle of a branch indicates that the branch is supported by 90% of the 1000 ultrafast bootstraps performed. Branch colors are indicative of the type of receptor (Rap or NprR) and of the bacterial group that either directly encodes the QSS or hosts a (pro)phage that encodes the QSS. The colorstrip surrounding the phylogenetic tree assigns a color to each leaf based on the type of genetic support that encodes the QSS: blue for chromosomes, orange for plasmids, dark purple for free phage genomes, different levels of purple for Phaster-predicted intact, questionable and incomplete prophages. The Rap receptors of Bacillus phage phi3T (only Rap found in a free phage genome) and of *B. licheniformis* intact prophage (viral Rap shown to modulate the sporulation and competence pathways of its host) are outlined.

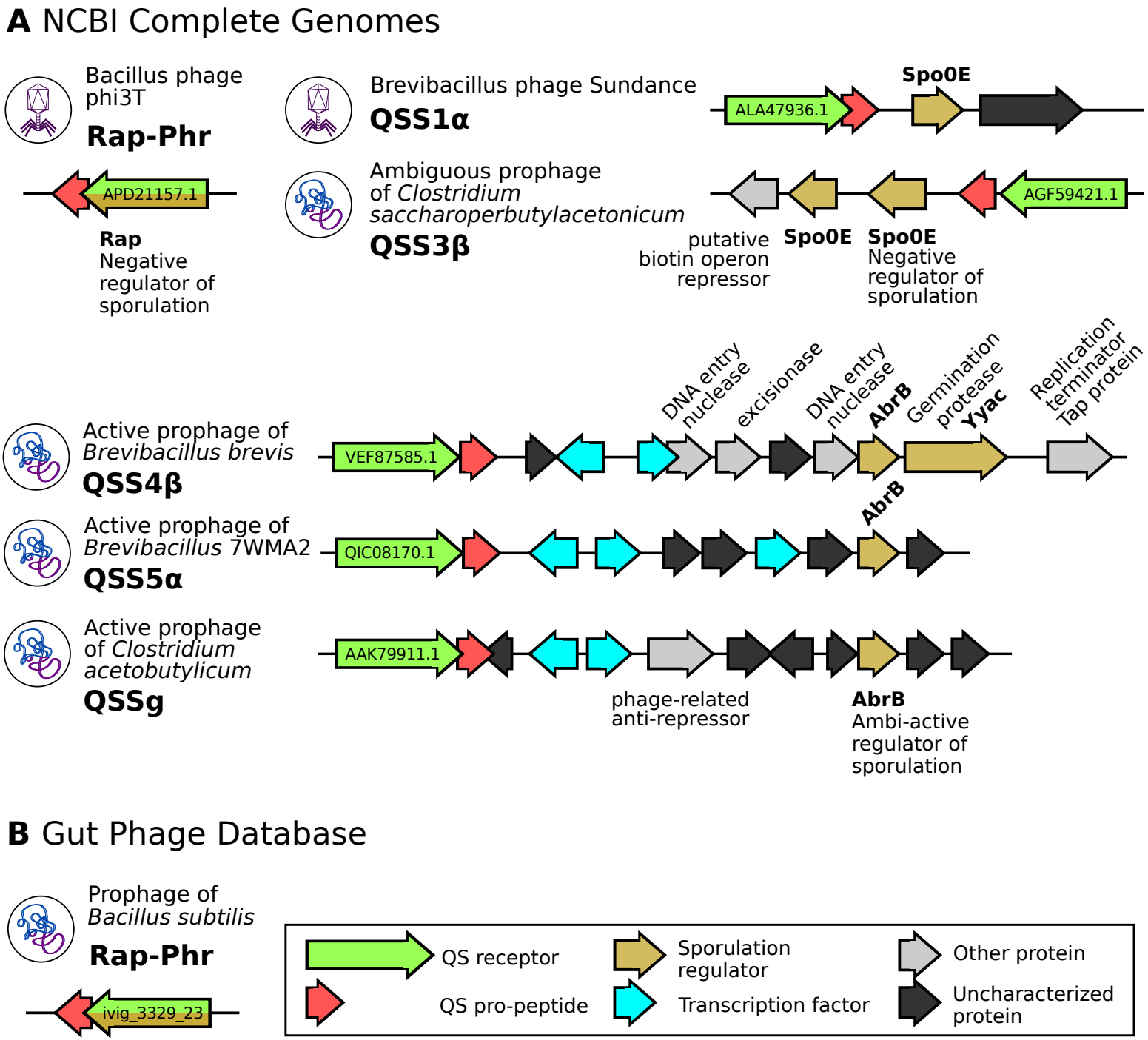

A NCBI Complete Genomes
Bacillus phage phi3T
Rap-Phr
APD21157.1
Rap
Negative regulator of sporulation
Brevibacillus phage Sundance
QSS1α
ALA47936.1
Spo0E
Ambiguous prophage of Clostridium saccharoperbutylacetonicum
QSS3β
AGF59421.1
putative biotin operon repressor
Spo0E
Spo0E
Negative regulator of sporulation
Active prophage of Brevibacillus brevis
QSS4β
VEF87585.1
DNA entry nuclease
excisionase
DNA entry nuclease
AbrB
Germination protease
Yyac
Replication terminator Tap protein
Active prophage of Brevibacillus 7WMA2
QSS5α
QIC08170.1
AbrB
Active prophage of Clostridium acetobutylicum
QSSg
AAK79911.1
phage-related anti-repressor
AbrB
Ambi-active regulator of sporulation
B Gut Phage Database
Prophage of Bacillus subtilis
Rap-Phr
ivig_3329_23
QS receptor
QS pro-peptide
Sporulation regulator
Transcription factor
Other protein
Uncharacterized protein

Figure 3: Key sporulation regulators in the genomic neighborhood of (pro)phage-encoded QSSs.

The genomic contexts of viral QSSs that are adjacent with homologs of regulators of the sporulation initiation pathway are displayed. Panel A corresponds to QSSs found inside complete genomes of the NCBI whereas panel B corresponds to QSSs found within MAGs of intestinal bacteriophages from the Gut Phage Database. Arrow sizes and distances between arrows are approximately proportional to gene lengths and to intergenic distances, respectively. Genes are colored according to their functional roles, as displayed in the legend. The *rap* gene is colored in both green and brown because it functions both as a QSS receptor and as a potential inhibitor of the sporulation initiation pathway. The text inside quorum sensing receptor genes correspond to the NCBI or Gut Phage Database identifier of the related protein. The taxonomic label of each genomic context refers to the name of the genome that encodes the QSS. The Rap-Phr operon of phage phi3T displayed in the top left is representative of all the other 324 prophage-encoded Rap-Phr found inside *Bacillus* genomes.

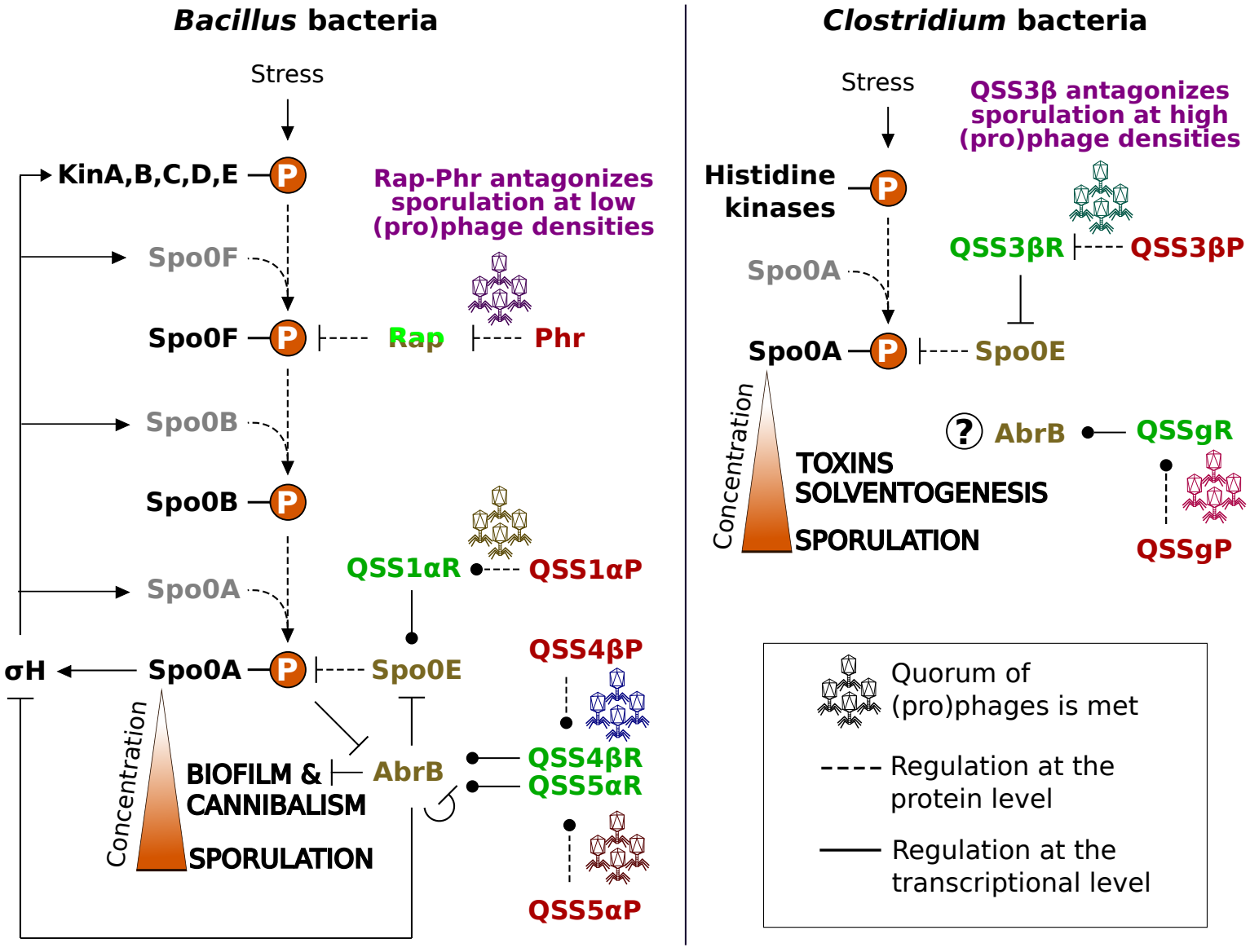

Bacillus bacteria
Stress
KinA,B,C,D,E
P
Rap-Phr antagonizes sporulation at low (pro)phage densities
Spo0F
Spo0F
P
Rap
Phr
Spo0B
Spo0B
P
Spo0A
QSS1αR
QSS1αP
σH
Spo0A
P
Spo0E
QSS4βP
Concentration
BIOFILM & CANNIBALISM
AbrB
QSS4βR
QSS5αR
SPORULATION
QSS5αP
Clostridium bacteria
Stress
QSS3β antagonizes sporulation at high (pro)phage densities
Histidine kinases
P
Spo0A
QSS3βR
QSS3βP
Spo0A
P
Spo0E
Concentration
?
AbrB
QSSgR
TOXINS
SOLVENTOGENESIS
SPORULATION
QSSgP
Quorum of (pro)phages is met
Regulation at the protein level
Regulation at the transcriptional level

Figure 4: Predicted modulations of the sporulation initiation pathway mediated by (pro)phage-encoded QSSs

The left and the right panels display the sporulation initiation pathway of the *Bacillus* genus and of the *Clostridium* genus, respectively. Transcriptional regulations are depicted by plain lines, whereas regulations at protein levels are depicted by dashed lines. At the end of each line, an arrow depicts an activation, a “T” symbol depicts an inhibition while a circle depicts an unknown direction of regulation. Written in grey are the inactive forms of sporulation proteins whereas their active, phosphorylated forms are written in black. The gradient of concentration starting from Spo0A-P indicates that sporulation is triggered by high levels of the master Spo0A-P regulator. Lower concentrations of Spo0A-P can trigger other bacterial processes than sporulation, as they may relieve a specific environmental stress and thus prevent, through alleviation of Spo0A phosphorylation, from a costly commitment to spore formation. The brown proteins (Rap, Spo0E and AbrB) depict regulators of Spo0A-P accumulation that are encoded by both bacteria and (pro)phages. The expression of (pro)phage-encoded *rap*, *spo0E* or *abrB* thus likely amplifies, by additive effect, the step of the host pathway controlled by each corresponding bacterial homolog. Red and green proteins depict the mature peptide and the receptor of a (pro)phage-encoded QSS inferred to regulate (pro)phage-encoded Rap, Spo0E or AbrB. An icon of grouped phages signifies that the regulation from the mature peptide to the receptor is expected to happen only at high (pro)phage densities. Each icon has its own color to highlight that the QSS genetic systems are encoded by different (pro)phages. These mechanisms are proposed to enable some bacteriophages to modulate the host sporulation initiation pathway in a density-dependent manner.

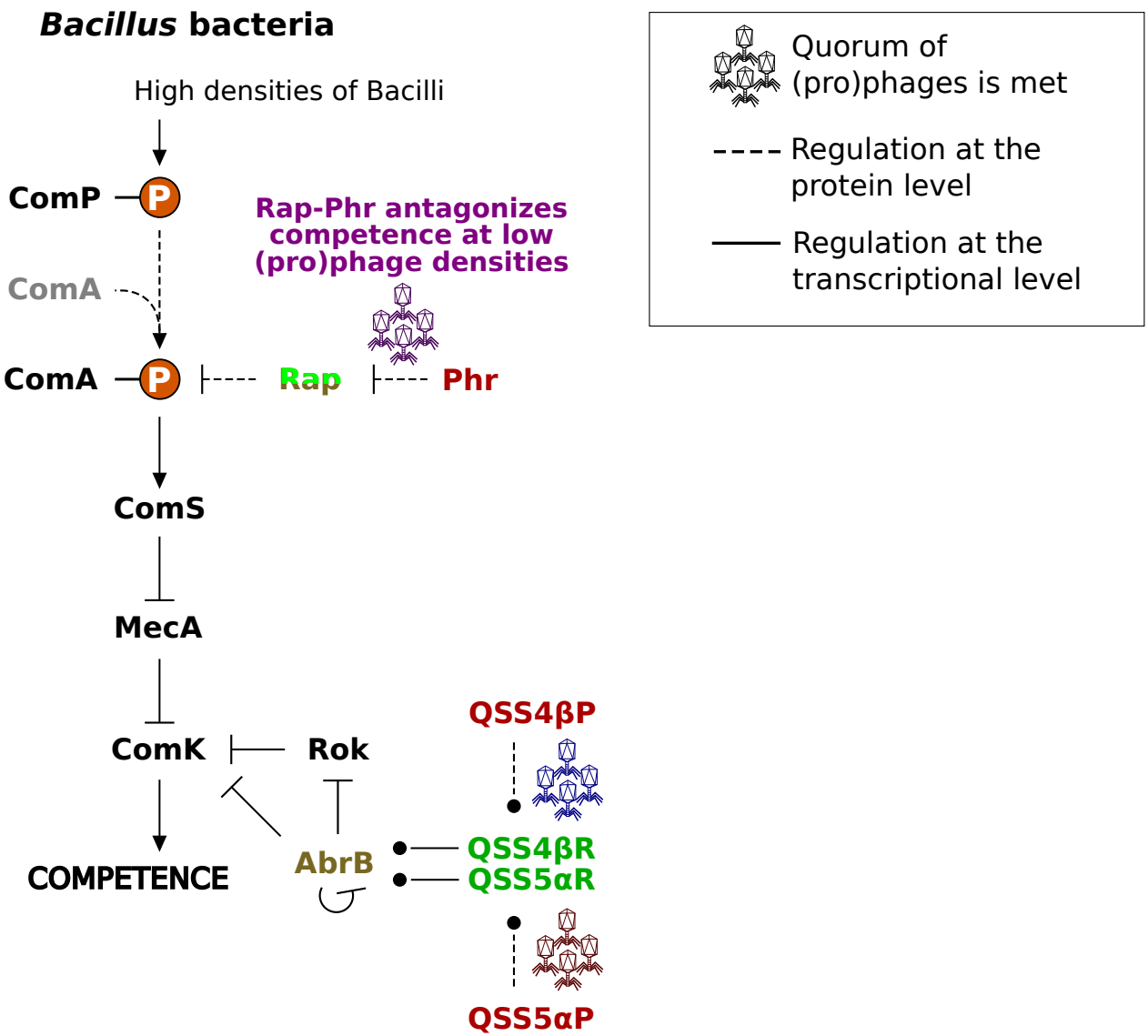
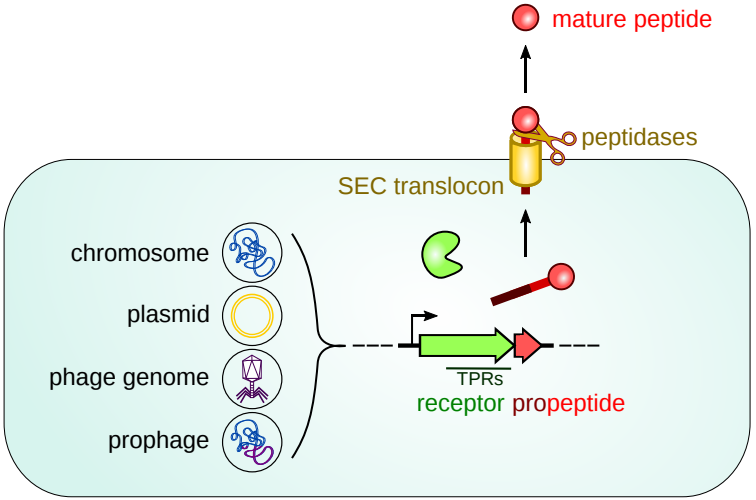

Bacillus bacteria
High densities of Bacilli
ComP
P
ComA
ComA
P
Rap-Phr antagonizes competence at low (pro)phage densities
Rap
Phr
ComS
MecA
ComK
Rok
COMPETENCE
AbrB
QSS4βP
QSS4βR
QSS5αR
QSS5αP
Quorum of (pro)phages is met
Regulation at the protein level
Regulation at the transcriptional level

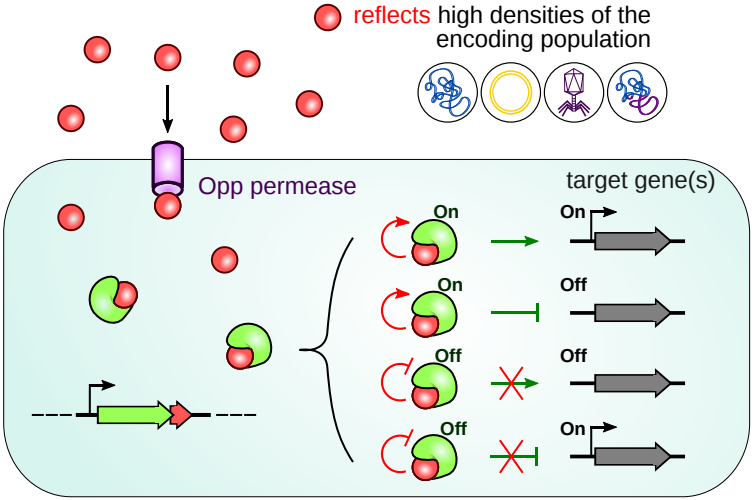
Figure 5: Predicted modulations of the competence pathway mediated by (pro)phage-encoded QSSs

This figure displays the competence pathway of the *Bacillus* genus and is built according to the same codes as figure 5. The modulations of AbrB and Rap total concentrations mediated by (pro)phage-encoded QSSs are proposed to enable encoding bacteriophages to interfere with the host competence pathway in a density-dependent manner.

SUPPLEMENTAL INFORMATION

- **Fig. S1**: canonical mechanism of RRNPP-type QSSs
- **Fig. S2**: common features between experimentally validated RRNPP-type QSSs
- **Fig. S2**: phylogenetic trees of candidate receptor families shared between (pro)phages and bacterial genomes
- **Fig. S3**: multiple sequence alignments of QSS families cognate pro-peptides
- **Table S1**: Candidate QSSs of the 16 families with at least 1 (pro)phage-encoded candidate QSS detected in NCBI complete genomes
- **Table S2**: Candidate QSSs matching already known QSS families
- **Table S3**: QSSs in the genome of *Clostridium acetobutylicum* ATCC 824
- **Table S4**: Hits of Rap, Spo0e and AbrB HMMs in the Gut Phage Database
- **Table S5**: Candidate QSSs found in the Gut Phage Database

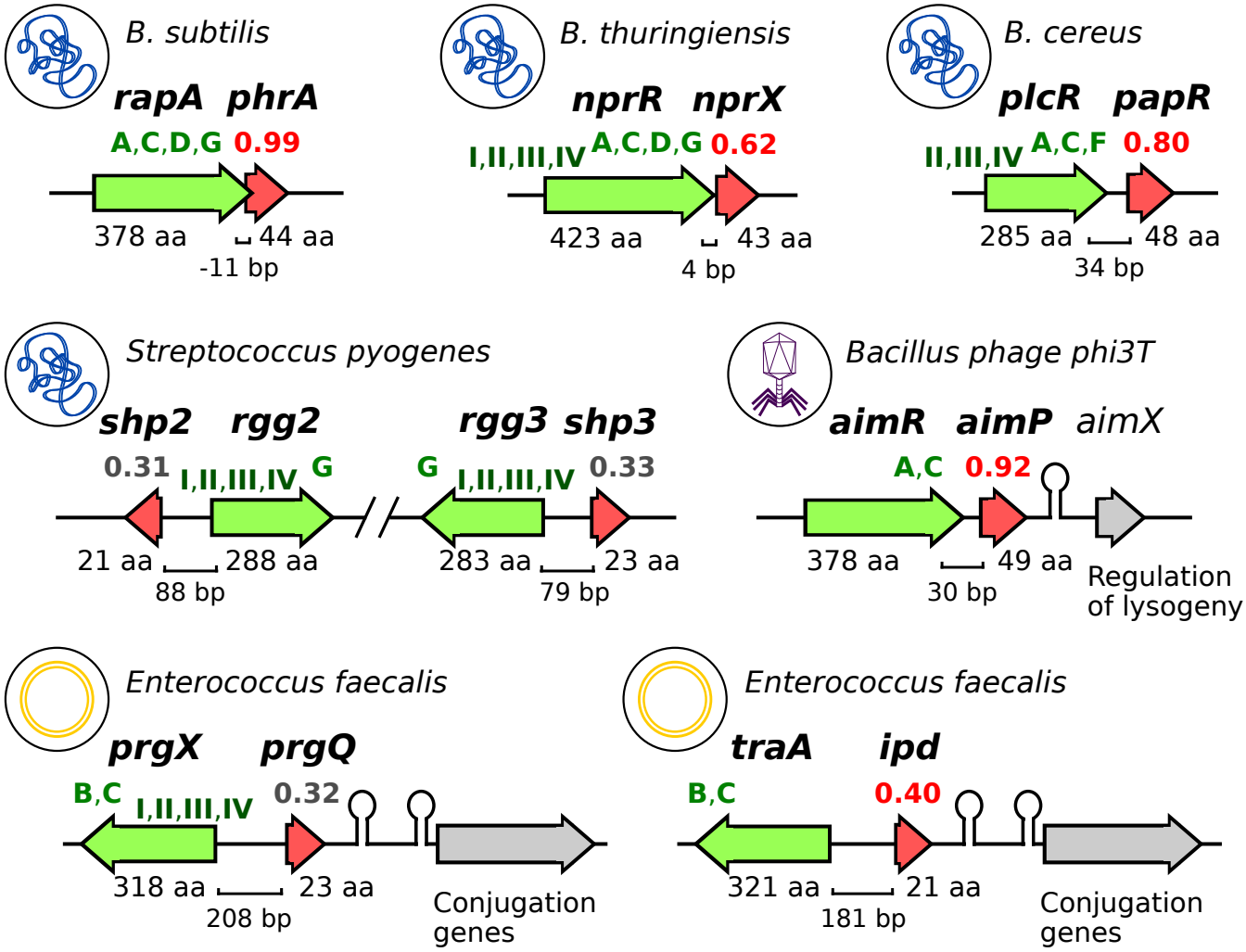

mature peptide
peptidases
SEC translocon
chromosome
plasmid
phage genome
prophage
TPRs
receptor propeptide


reflects high densities of the encoding population
Opp permease
target gene(s)
On
On
Off
Off
Off
On

Supplementary figures legend

Figure S1: **canonical mechanism of RRNPP-type QSSs**.

The left panel shows the behavior of a RRNPP-type QSS at low population densities of its encoding DNA molecule: a bacterial chromosome, a plasmid, a free phage genome, or a prophage inserted into the bacterial genome. Upon bacterial expression of the QSS, an intracellular receptor (in green) and a pro-peptide (in red) are produced. The pro-peptide contains a N-terminal signal sequence (in dark red) that tags the protein for transport through the cell membrane, typically via the SEC-translocon. Upon secretion, the propeptide is cleaved by exopeptidases, which releases a small mature quorum sensing peptide into the extracellular medium. The right panel shows the behavioral switch that is triggered when high concentrations of the peptide are reached, reflecting high densities of the encoding population. The peptide is robustly imported by bacteria and within QSS-expressing cells, binds with the tetratricopeptide repeats (TPRs) of its cognate receptor. Upon binding with the peptide, the receptor undergoes a conformational change and gets either turned-on or -off. This results in the subsequent downregulation or upregulation of target gene(s) according to the four displayed scenarios, depending on whether the receptor acts as a repressor or as an activator. Of note, such regulations can also happen at the protein level if the receptor is a protein regulator rather than a transcription factor. This quorum sensing mechanism allows a RRNPP-type QSS-encoding population to coordinate behavioral transitions in a density-dependent manner.



DNA molecules

- Chromosome
- Plasmid
- Phage genome

Genes

- QS receptor
- QS pro-peptide
- Adjacent target genes

Matched N-terminal DNA binding domains (HMM)

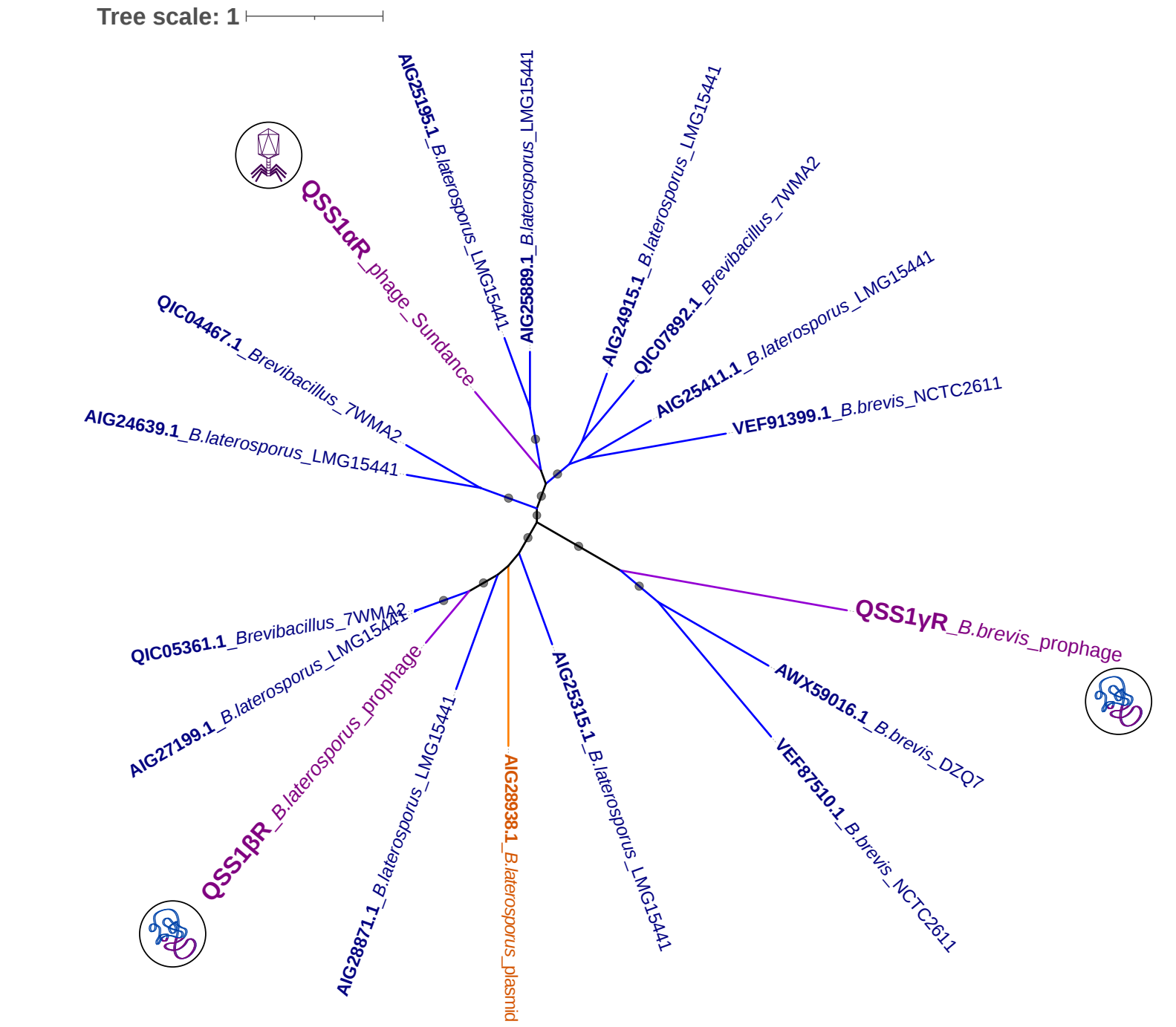
I	CATH	1.10.260.40	**III**	PFAM	PF01381
II	SF	SSF47413	**IV**	SMART	SM00530

Matched C-terminal TPR repeats (HMM)

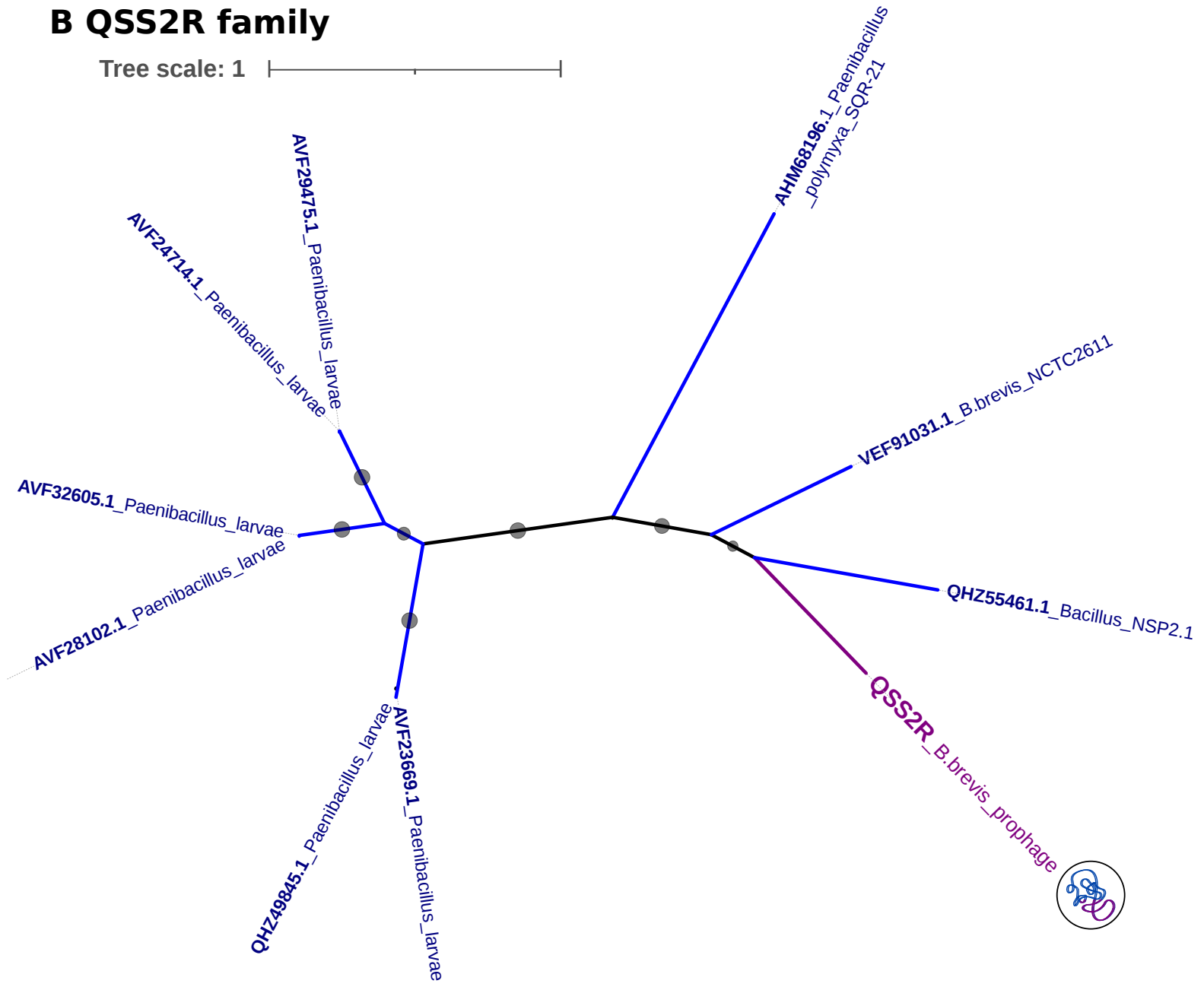
A	CATH	1.25.40.10
B	CATH	1.25.40.400
C	SF	SSF48452
D	PFAM	PF13424
E	PFAM	PF18768
F	SMART	SM00028
G	TIGR	TIGR01716

Figure S2: **Common features between experimentally validated RRNPP-type QSSs**. Each genomic context corresponds to the representative QSS of an experimentally-characterized RRNPP-type QSS family: rapA-phrA (loci BSU_12430 and BSU_12440 in B. subtilis genome NP_389125.1), nprR-nprX (loci BTHUR0002_RS02765 and BTHUR0002_RS32155 in B. thuringensis genome NZ_CM000747.1), plcR-papR (loci EJ379_RS27345 and EJ379_RS27340 in B. cereus genome NZ_CP034551.1), rgg2-shp2 (loci SD90_RS02145 and SD90_RS09265 in S. pyogenes genome NZ_CP010450.1), aimR-aimP (loci phi3T_89 and phi3T_90 in B. phage phi3T genome KY030782.1), prgX-prgQ (genes prgX and prgQ in E. faecalis plasmid pCF10 AY855841.2) and traA-iPD1 (genes traA and iPD1 in E. faecalis plasmid pPD1 D78016.1). The icon at the left of each context indicates the genetic element that encodes the QSS (bacterial chromosome, phage genome or plasmid) and the associated label indicates the genome to which this genetic element belongs. The green gene corresponds to the quorum sensing receptor and the red gene to its cognate propeptide. The intergenic distance between the two genes is given in number of base pairs. The length of each gene is given by the number of amino acids in the translated protein. The hairpin symbol depicts an intrinsic terminator and a grey gene indicates an adjacent, target gene regulated by the QSS. The number above each pro-peptide corresponds to the likelihood, computed by SignalP, that the propeptide harbors a N-terminal signal sequence for the SEC-translocon. A likelihood score colored in red means that the propeptide is predicted by SignalP to be secreted via the SEC-translocon whereas a score colored in grey means that it is predicted to be secreted otherwise. The green letters above the C-terminal encoding region of each receptor indicate the names of the HMM (PFAM, SMART, TIGR) or of the HMM family (CATH, SuperFamily) of Tetratricopeptide repeats (TPRs) that are found within the sequence of the translated protein. The roman numbers above the N-terminal encoding region of each receptor indicate the names of the HMM or of the HMM family of DNA binding domains found in the sequence of the translated protein.

A QSS1R family



B QSS2R family



C QSS3R family

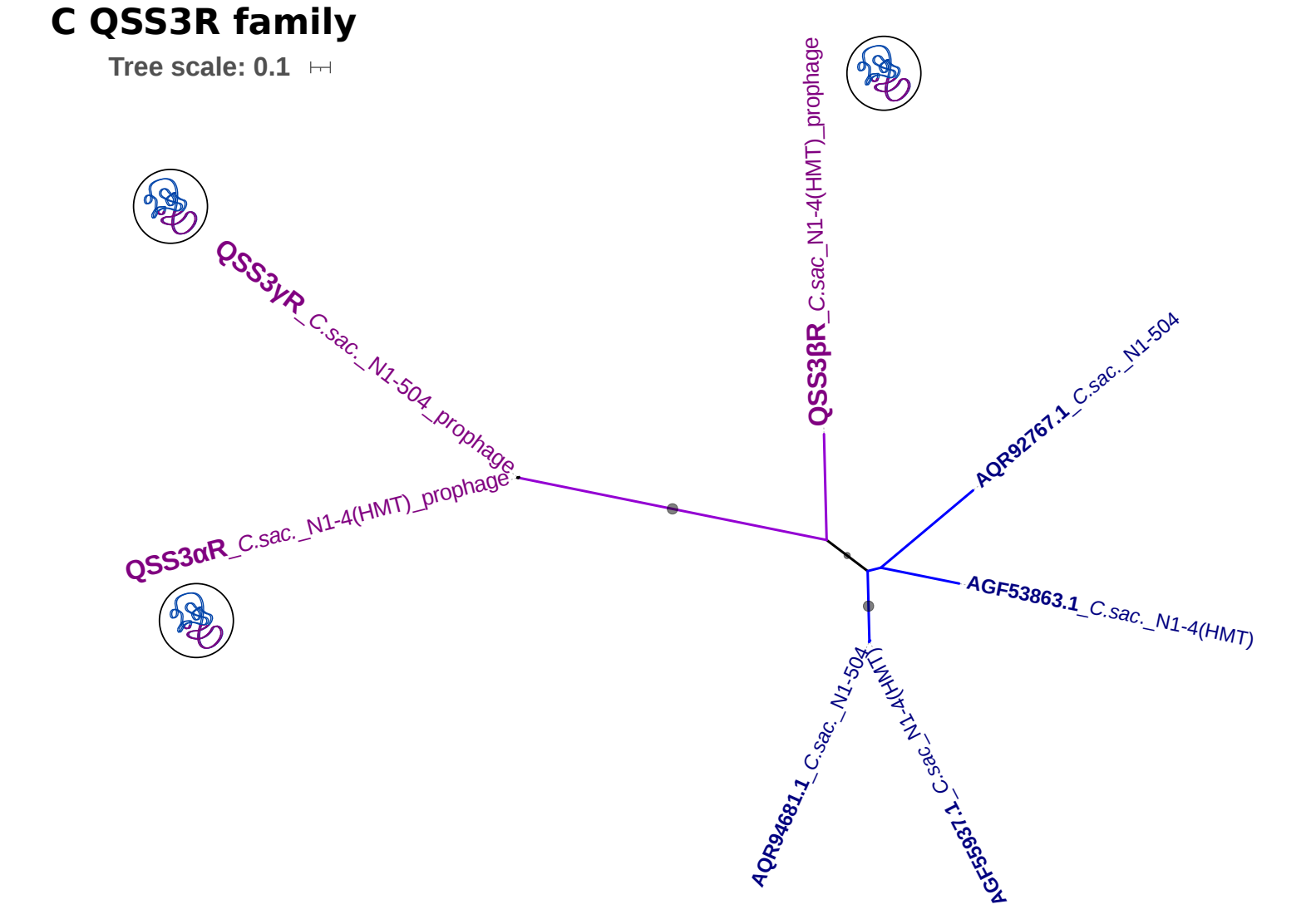


Figure S3: Phylogenetic trees of candidate receptor families shared between (pro)phages and bacterial genomes

Each maximum-likelihood phylogenetic tree is unrooted. Black dots indicate that the branch is supported by 90% of the 1000 ultrafast bootstraps performed. The color of the leaves indicate the genetic element encoding the QSS: blue for chromosomes, orange for plasmids, purple for phage genomes and predicted prophages.

QSS1 family

```
VEF92013.1__15653/1-36   1 -MQKLLLVFALVVFFSSPVVSEAYDP-----------------------TSEPYRVEPDW   36
VEF87511.1__15653/1-40   1 MKKWILSLCMLTAFVTLPVLVEQPAK--------------------QTNSIIQYGWPEPW   40
AIG26091.1__1534/1-45    1 -MRKKIVLMLLTLGLITNMGFALVKTDEANNL--------------NTFSMRANNADPGY   45
AWX59017.1__4003/1-59    1 -MKKKASFLLVALSMVLGSFLAIDVPYKNSNNDSLLAVDVPYGISKEKSNSPQRVPDPGY   59
ALA47937.1__5003/1-48    1 -MKRIFWGVSVFLTMFSFSLLALENDKTEHSTNKI-----------QHYAITVYTDNPGY   48
AIG25196.1__1534/1-46    1 MMKKIVISTIIFSTFLSLTFIVYDKFMGVYSE--------------KYSASKVYNEDPGH   46
AIG25890.1__1534/1-46    1 -MKKLVIVSVVFSSILSSFVFIDMDNCSEKEYA-------------SYQSIQVFFEDPGH   46
AIG24640.1__1534/1-45    1 -MKKIVVATSIILASLSLTFLFSIERQTENQI--------------VSGHIKVYTGDPGH   45
QIC04468.1__14625/1-47   1 MKKFVLTASVLLLSLSLSTPFLSVDK--PIEKRLF-----------STNTFKPYTGDGGH   47
VEF91398.1__15653/1-46   1 MRKFLAVLSRVSLVIACFLVVVSTGANSMSLP--------------KKPPIIMYSDNPEG   46
AIG25410.1__1534/1-46    1 MKKRLTVLMTASLVLCSSLFAVAIFSNSEDHV--------------VYSSIKVYTDNPGH   46
AIG24914.1__1534/1-46    1 MKKYLALITTVSLVLLSSLFTMEITGKFEVNN--------------VASSMKPYTDDPQG   46
QIC07891.1__14625/1-45   1 -MKKYLTIIATVSFVICSLFTVASFLDFEKYG--------------EQQSAKTFHDDPQG   45
AIG25316.1__1534/1-45    1 -MKKTVVATVLSLTLLVAMNVVPLYKTTDSNI--------------TAFKVIPTVVDPGH   45
AIG28939.1__1534/1-44    1 -MKKVIGMLLMSVTMLVTVTLTDSVVPTPSG---------------TSLQVQPSDYDPGY   44
AIG27198.1__1534/1-48    1 -MKKLISILTLALAIICFISTESPTSDSNATTKKI-----------GLYTIQQNHYDPGY   48
QIC05360.1__14625/1-48   1 -MKKLISILTLALAIICFISTESPTSDSNATTQKI-----------GLYTIQQNHYDPGY   48
AIG28870.1__1534/1-44    1 -MKKAIITIFMTLALVTSIGSIDSFSHVAEA---------------SKPHAIQYVYDPGF   44
```

QSS2 family

```
AVF28101.1__8996/1-49    1 -MKI-KNWILCITVAITLIGGSSAFSTTTQATASN-----------ATIALS-DHGVGGDYAS   49
AVF32604.1__8997/1-49    1 -MKI-KNWILCITVAITLIGGSSAFSTTTQATASN-----------ATIALS-DHGVGGDYAS   49
AVF23670.1__8994/1-46    1 -MKI-KKYLVFLILTFSLLYVVNGINPDKTHEAAK-----------QSVIAYAENGVGE----   46
QHZ49846.1__14754/1-46   1 -MKI-KKYLVFLILTFSLLYVVNGINPDKTHEAAK-----------QSVIAYAENGVGE----   46
AVF24715.1__8996/1-58    1 MKKF-KKLLLSVVLAAAILGGGIALNLPQTSEDSAVAALHGTGQAPANMSVAALHGTGQ----   58
AVF29476.1__8997/1-58    1 MKKF-KKLLLSVVLAAAILGGGIALNLPQTSEDSAVAALHGTGQAPANMSVAALHGTGQ----   58
QHZ55460.1__2483/1-47    1 MKKFVKGFVLSLCLVLGLGFGINVHIADGKVITIK-----------PPIIQY-DHGVGG----   47
AHM68195.1__2742/1-33    1 -MQF-KNLKYMLIIG---LFGFLELS--------------------HTITVF-GHGTGA----   33
VEF92630.1__15653/1-42   1 ---M-RSMILLLTVLFGLSHGIV-VPDHQEVLSSH-----------HVVTTY-THGAGG----   42
VEF91032.1__15653/1-44   1 -MKL-KMFSLSFALFFSLCIGIVNSETSSSNTMSR-----------PNVTTY-THGMG-----   44
```

QSS3 family

```
AGF56488.1__2211/1-54    1 MKKLLLTMVACVVLLSNVAFAATKTTSTTTDTKSTITS-SSGEKFHLDSRDPDGW---------   54
AQR95596.1__6751/1-54    1 MKKLLLTMVACVVLLSNVAFAATKTTNTTTDTKSTITS-SSGEKFHLDSRDPDGW---------   54
AQR92768.1__6751/1-53    1 MKKIVLAVLATSMLLTNVAFASTETTHTKSTQNSTISKVTQNNGYKLMTHDAN-----------   53
AGF53864.1__2211/1-49    1 MKKIILAVIATSMLLTNVAFASTKTTH----KNTNSTKVTQNNVYSVNSIDPH-----------   49
AGF55936.1__2211/1-53    1 MKKIILAVMATSMLLTNVAFASTKTTHTKITKNSTVSSVDPNNVYHLNTLDPN-----------   53
AQR94680.1__6751/1-64    1 MKKIILAVMATSMLLTNVAFASTKTTHTKITKNSTVSSVDPNNVYHLNTLDPNNVYHLNTLDPN   64
AGF59420.1__2211/1-53    1 MKKIILAVIATSMLLTNVAFASTKTTHTNTGKNTTVSSVKQNSMYHLNTTDPY-----------   53
```

QSS4 family

```
VEF87223.1__15653/1-63   1 -MKRILFATFAALMIGLVTPSTGVFAGPPENDKVAP-IAIDHSDGVEKFPRQKPPIVVYGPPEND   63
VEF87586.1__15653/1-63   1 MKKKLIFIPSLALLIGLFTANEVSLAGPPDGWGIAPETKVIQTGDSEPTTWGVSPLG--GPPDGW   63
```

QSS5 family

```
AIG27472.1__1534/1-48    1 -MKKLFASSLVLFAFLLMSSSMVSTEESVQVTAEKATAHFIQYSTAPDW   48
QIC08171.1__14625/1-46   1 MKKKLVAFSLALVTLVGI-SSVVGAEKSIQNVHSKPSA--IQYITEPEW   46
```

QSSg family

```
QspH/1-38        1 ----MNSRVKKLLFLVLCSFII---IGYLSVNHISVLASYPGWSW-------   38
AAK81614.1/1-41  1 -MGNMNSRVKKLLFLVLCSFII---IGYLSVNHISVLASYPGWSW-------   41
QspB/1-50        1 --MLPKKSNKFILVLILCAIVFVSAFALNTSGTRSLLGAEPTWGWNISKLLF   50
QspF/1-39        1 ---MFFKRYKKLLFIAGILAAV---LTINLGLNGNLLSGLPLMVW-------   39
QspD/1-41        1 --MFFLKKHKKFLSMLVLTLVI--GVLVTTKGGISILSMFPDAPW-------   41
QspA/1-41        1 -MITVLKKHRFLISVLTLGVLF---LCIVFTNSSYILSGAPCIPW-------   41
QspC/1-38        1 ---MFLKRHKLIFSILLLGTLI---LGVTFFKS-SLLAFYPGMPW-------   38
QspE/1-41        1 MRIVFFKRHKFLLTVLALGILF---LGATFFGN-SLLSAFPDLPW-------   41
QspG/1-39        1 ---MLLKKHKLAFSVLALTLLF---LGVNFLGNNVGLSVSPMTPW-------   39
```

Figure S4: Multiple sequence alignments of QSS families cognate pro-peptides

The figure displays the multiple sequence alignment of cognate propeptides for each receptor family of size > 1 that includes at least one (pro)phage-encoded QSS. A purple circle at the left of each protein identifier of the propeptide indicates that the QSS was found in a free phage genome whereas a purple circle indicates that the QSS was found in a predicted prophage region. The residues are colored according to the ClustalX colorscheme (http://www.jalview.org/help/html/colourSchemes/clustal.html), which colors amino acids based on residue type conservation (hydrophobic, positively charged, negatively charged, polar etc…). Pro-peptides are characterized by a N-terminal region composed of positively charged amino acids (R, K), followed by a hydrophobic region. The mature peptide (typically 5 to 6 aminoacids) is usually encoded by a C-terminal region of the propeptide and is characterized in the alignment by the entanglement of conserved and variable positions.