



HAL
open science

RRNPP_detector: a tool to detect RRNPP quorum sensing systems in chromosomes, plasmids and phages of gram-positive bacteria

Charles Bernard, Yanyan Li, Eric Bapteste, Philippe Lopez

► To cite this version:

Charles Bernard, Yanyan Li, Eric Bapteste, Philippe Lopez. RRNPP_detector: a tool to detect RRNPP quorum sensing systems in chromosomes, plasmids and phages of gram-positive bacteria. 2021. hal-03384406

HAL Id: hal-03384406

<https://hal.science/hal-03384406>

Preprint submitted on 20 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **RRNPP_detector: a tool to detect RRNPP quorum sensing systems in**
2 **chromosomes, plasmids and phages of gram-positive bacteria**

3

4 **AUTHORS**

5 Charles Bernard ^{1,2}, Yanyan Li ², Eric Bapteste ¹ and Philippe Lopez ^{1,*}

6

7 **AFFILIATIONS**

8 ¹ Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum

9 National d'Histoire Naturelle, EPHE, Université des Antilles, Campus Jussieu, Bâtiment A, 4eme et.

10 Pièce 429, 75005 Paris, France

11 ² Unité Molécules de Communication et Adaptation des Micro-organismes (MCAM), CNRS,

12 Museum National d'Histoire Naturelle, CP 54, 57 rue Cuvier, 75005 Paris, France

13

14 **CORRESPONDING AUTHOR**

15 * Correspondence to Philippe Lopez;

16 Phone: +33 (01) 44 27 34 70; E-mail address: philippe.lopez@upmc.fr

17 **ABSTRACT**

18 Gram-positive bacteria (e.g. Firmicutes) and their mobile genetic elements (plasmids,
19 bacteriophages) encode peptide-based quorum sensing systems (QSSs) that regulate behavioral
20 transitions in a density-dependent manner. In their simplest form, termed “RRNPP”, these QSSs
21 are composed of two adjacent genes: a communication propeptide and its cognate intracellular
22 receptor. Despite the prime importance of RRNPP QSSs in the regulation of key biological
23 pathways such as virulence, sporulation or biofilm formation in bacteria, conjugation in plasmids or
24 lysogeny in temperate bacteriophages, no tools exist to predict their presence in target
25 genomes/mobilomes. Here, we introduce RRNPP_detector, a software to predict RRNPP QSSs in
26 chromosomes, plasmids and bacteriophages of gram-positive bacteria, available at
27 https://github.com/TeamAIRE/RRNPP_detector. RRNPP_detector does not rely on homology
28 searches but on a signature of multiple criteria, which are common between distinct families of
29 experimentally-validated RRNPP QSSs. Because this signature is generic while specific to the
30 canonical mechanism of RRNPP quorum sensing, it enables the discovery of novel RRNPP QSSs
31 and thus of novel “languages” of biocommunication. Applying RRNPP_detector against complete
32 genomes of viruses and Firmicutes available on the NCBI, we report a potential 7.5-fold expansion
33 of RRNPP QSS diversity, alternative secretion-modes for certain candidate QSS propeptides,
34 ‘bilingual’ bacteriophages and plasmids, as well as predicted chromosomal and plasmidic
35 Biosynthetic-Gene-Clusters regulated by QSSs.

36

37 **INTRODUCTION**

38 Quorum sensing is the mechanism by which microbial entities sense when their population density
39 reaches a threshold level, and thereupon typically switch from individual to group behaviors (1).
40 The population density is reflected by the extracellular concentration of a communication signal,
41 produced and secreted by individual entities. The quorum is met when this signal reaches a
42 threshold concentration, at which it starts to be robustly detected and transduced population-wide
43 by its cognate receptor module. In bacteria of the Proteobacteria or Actinobacteria phyla, these
44 communication signals typically are small molecules synthesized by enzymes (2,3) whereas in the
45 Firmicutes phylum, these are oligopeptides, matured from genetically encoded pro-peptides (4).
46 Peptide-based quorum sensing systems can be divided into two main categories: those with a
47 receptor module composed of a membrane-bound sensor coupled with an intracellular response
48 regulator (two-components system) (5), and those in which the receptor is an intracellular
49 transcription factor (or a protein inhibitor) that gets either turned-on or -off upon binding with the
50 imported communication peptide (one component system) (6,7). The latter are generally included
51 under the term RRNPP, named after the 5 first experimentally-characterized families of such
52 receptors: Rap (*Bacillus* genus), Rgg (*Streptococcus* genus), NprR (*Bacillus cereus* group), PlcR
53 (*Bacillus cereus* group) and PrgX (pCF10 plasmid of *Enterococcus faecalis*) (7–9).

54

55 The initial members of the RRNPP group of QSSs were reported to trigger key biological pathways
56 when their encoding population reach high densities: from virulence (Rgg, PlcR) to competence
57 (Rgg, Rap), necrotropism (NprR), sporulation, biofilm formation (Rap, NprR) and inhibition of
58 conjugation (PrgX) (7–9). Considering that the virulence of *Bacillus* and *Streptococcus* pathogens
59 may cause infectious diseases in humans (10,11), that the spore is the transmissible form of many
60 *Bacillus* and *Clostridium* human pathogens (12), that biofilms contribute to infections or food-
61 poisoning (13–15), and that competence and conjugation are responsible for the spread of antibiotic
62 resistance genes (16), RRNPP QSSs are directly linked to central health issues.

63

64 Interestingly, the case of the plasmidic PrgX system illustrates that RRNPP QSSs may not only be
65 present on bacterial chromosomes but also on mobile genetic elements (MGEs). Consistently, the
66 conjugation-regulating TraA-Ipd1 system have later on expanded the list of plasmidic RRNPP
67 QSSs (17). In 2017, Erez *et al.* even discovered that some temperate bacteriophages encode
68 RRNPP QSSs, namely the “arbitrium” systems that guide the lysis-lyogeny decision upon *Bacillus*
69 infection (18,19). Recently, two other chromosomal RRNPP QSSs have been experimentally
70 validated: the QsrB-QspB system that delays sporulation and solvent formation in *Clostridium*
71 *acetobutylicum* (20) and the AloR13-AloP13 sporulation-regulator in *Paenibacillus polymyxa* (21).
72 As important the experimentally-validated RRNPP QSSs are in the regulation of the biology of their
73 encoding microbial entity, they represent only the tip of the iceberg. Indeed, the diversity of RRNPP
74 QSSs is likely not fully explored, as hinted by the numerous candidate RRNPP QSS receptors in
75 *Clostridium* and *Enterococcus* species reported to harbor local similarities with regions of Rap,
76 NprR, PlcR or Rgg (20,22,23). Uncovering novel RRNPP QSS families would unveil new
77 bacterium-bacterium, plasmid-plasmid or phage-phage modes of communication but most and
78 foremost reveal novel density-dependent evolutionary strategies that could transform our views of
79 microbial interaction, adaptation and evolution. Expanding RRNPP QSS diversity would also likely
80 pave way to major practical outcomes as novel RRNPP QSSs could regulate the production of
81 novel antimicrobial compounds (24,25) or could underlie mechanisms by which some human
82 pathogens acquire virulence (26–29).

83

84 Conveniently, we noticed that the different members of the RRNPP group share a common
85 signature of 5 criteria (**Fig. 1**): i) the pro-peptide is a small protein (10-70aa), ii) the pro-peptide is
86 secreted, except in rare exceptions (e.g. Shp and PrgQ (7)), via the SEC-translocon and further
87 matured by exopeptidases into a communication peptide, iii) the receptor has a length comprised
88 between 250 and 500aa, iv) the receptor harbors tetratricopeptide repeats (TPR), which are
89 structural motifs involved in the binding of small peptides (in this case, the cognate communication
90 peptide), v) the genes encoding the pro-peptide and the receptors are directly adjacent to each

91 other. Advantageously, a large amount of reference Hidden-Markov-Models (HMMs) from the Cath-
92 Gene3D (30), Superfamily (31), SMART (32), Pfam (33) and TIGRFAM (34) databases are already
93 available to detect TPRs in protein sequences. Moreover, a tool called SignalP specifically
94 computes the likelihood that proteins harbor a signal sequence for the SEC-translocon
95 (35) (**Fig .1**). Consequently, the generic, yet specific signature of RRNPP QSSs is detectable *in*
96 *silico*, without requiring homology searches that would limit the output to representatives of already
97 known QSSs. On this basis, we have developed RRNPP_detector, a python software that detects
98 the RRNPP signature in chromosomes and MGEs of gram-positive bacteria. We present its
99 workflow, its usage and showcase practical examples of analyses against the complete genomes
100 of viruses and of Firmicutes available on the NCBI.

101

102 **MATERIALS AND METHODS**

103 **Overview of the RRNPP_detector workflow**

104 Given a fasta of protein sequences and the coordinates of their coding sequences in target
105 genome(s), Metagenomics-Assembled-Genome(s) or contig(s), RRNPP_detector first reduces the
106 search space by retaining only proteins that have a length compatible with RRNPP receptors (250-
107 500aa) and pro-peptides (10-70aa). Next, only the 250-500aa proteins whose coding sequences
108 (CDSs) are found adjacent with the CDS of a 10-70aa protein are retained, and vice versa. Then,
109 HMMs of TPRs found in reference RRNPP receptors (PFAM: PF13181, PF13424; Gene 3D:
110 1.25.40.400, 1.25.40.10; TIGRFAM: TIGR01716; Superfamily: 48452, SMART: SM00028) are
111 queried with hmmsearch (36) against the remaining 250-500aa proteins. Only the proteins
112 matched by at least one HMM of TPRs (E-value < 1E-5) are retained and designated as potential
113 receptors. These potential receptors next undergo a computational characterization. First, a
114 BLASTp search (37) of reference RRNPP receptors (Rap, Rgg, NprR, PlcR, PrgX, TraA, AimR of
115 Bacillus phage phi3T, AimR of Bacillus phage Waukesha, QsrB and AloR13) against these
116 potential receptors are launched to identify the homologs of already known QSSs (by default: E-
117 value <= 1E-5, identity >= 20%, and alignment coverage >= 60% of the length of both the query
118 and the target sequences). Then, an HMMsearch of DNA-binding domains found in reference
119 receptors (PFAM: PF01381; Gene3D: 1.10.260.40; Superfamily: 47413; SMART: SM00530) is
120 launched against the potential receptors to identify those that may be transcription factors. After
121 this characterization step, the 10-70aa proteins are further reduced to those with a CDS adjacent
122 to the one of a potential receptor. At this stage, each tandem of remaining adjacent pro-peptide and
123 receptor are considered forming a potential QSS and will eventually be outputted. However, either a
124 'permissive' or a 'conservative' label will be assigned to these candidate QSSs according to the
125 next filters. First, any QSS with an intergenic distance > 400 base pairs between the two CDSs will
126 be considered as 'permissive' because low intergenic distance tends to correlate with functional

127 association (38). Second, only the remaining QSSs with a pro-peptide predicted by SignalP (35) to
128 harbor a signal for the SEC-translocon under the '-org gram+' option will be considered as
129 'conservative' (**Fig. 2**). Because a few experimentally-characterized pro-peptides are known to be
130 secreted via other systems than the SEC-translocon (e.g. Shp2 and PrgQ (7)), we thought that it
131 would be relevant to keep the QSSs in which the pro-peptides are not predicted to be secreted via
132 the SEC-translocon in the 'permissive' output.

133

134 **Usage, input and output of RRNPP_detector**

135 RRNPP_detector was designed to take two kinds of input. The first option is to provide a
136 nucleotide sequence (fna) of either one or multiple genetic elements. RRNPP_detector will then
137 call Prodigal (39) to detect the CDSs in these nucleotide sequences and produce a fasta of the
138 predicted protein sequences (faa) and an annotation file of the coordinates of their CDSs (gff).
139 Alternatively, the user can submit a fasta of the protein sequences from one or multiple
140 genomes/MAGs/contigs (faa) along with an annotation file of these genetic elements (in one
141 concatenated gff or in one concatenated NCBI Assembly feature table). Option 2 has the
142 advantage to integrate pre-computed annotations and therefore to keep reference IDs of proteins
143 along the analysis. RRNPP_detector was designed to be launched in two modes: a default fast
144 mode that processes all the target genomes/MAGs/contigs together in a single run; and a
145 'preserve_ram' mode designed to minimize RAM usage at the expense of speed by processing
146 multiple genomes one by one in a for loop. RRNPP_detector outputs a fasta directory with the
147 protein sequences of the 'permissive' pro-peptides, the 'permissive' receptors, the 'conservative'
148 pro-peptides and the 'conservative' receptors. In addition, two 'permissive' and 'conservative'
149 summary tabular files with various information regarding each candidate QSS are produced. A
150 summary file typically displays, for each QSS, the ID of the encoding genetic element, the
151 intergenic distance between the pro-peptide and the receptor genes, the genomic orientation of
152 these two genes, the IDs and the genomic coordinates of the receptor and the propeptide, the best
153 HMM of TPRs and DNA-binding domains found in the sequence of the receptor (with associated E-
154 value), the BLASTp results in case of an homology of the candidate receptor with a reference
155 receptor, and the SignalP prediction of the secretion mode of the pro-peptide (with associated
156 likelihood score).

157

158 **RESULTS**

159 **RRNPP_detector reveals thousands of candidate RRNPP QSSs in complete bacterial** 160 **chromosomes, plasmids and phage genomes**

161 To illustrate RRNPP_detector, we launched it against the 4974 complete genomes of Firmicutes
162 and the 32919 complete genomes of Viruses Firmicutes available on the NCBI Assembly Database

163 **(Fig. 2).** We describe how such analyses can be easily done with the practical example of Viruses
164 in the readme file of RRNPP_detector:
165 https://github.com/TeamAIRE/RRNPP_detector/readme.md. We report the identification of 6467
166 'permissive' candidate RRNPP QSSs, distributed across 1074 different taxa (**Table S1**). Some
167 receptors of these 'permissive' candidate QSSs notably include the RopB regulator of virulence in
168 *Streptococcus pyogenes* (27), as well as additional members of the AimR, AloR13, NprR, PlcR,
169 PrgX, QsrB, RapA and Rgg2 experimentally-validated families that did not pass the 'conservative'
170 thresholds of RRNPP_detector (**Table S1**). Interestingly, 105 'permissive' RRNPP QSSs have
171 been detected in 97 isolated genomes of viruses, of which 89 bacteriophages and 8 giant
172 eukaryotic viruses of the *Mimiviridae* taxonomic family. Accordingly, some bacteriophage genomes
173 are found to encode multiple 'permissive' candidate QSSs (up to 4 in Brevibacillus phage
174 SecTim467) (**Table S1**). Last but not least, 435 and 47 'permissive' candidate QSSs include a
175 propeptide predicted to be secreted via the LIPO (Sec/SplI) and TAT (TaT/SPI) secretion systems,
176 respectively, pointing at alternative secretion modes than SEC (Sec/Spl) for some candidate
177 RRNPP propeptides (**Table S1**). In addition, we report the identification of 3529 'conservative'
178 candidate RRNPP QSSs that we hereafter characterize in more depth (**Table S2**). We first used
179 the Phaster API (40) for each chromosome or plasmid carrying at least one QSS to identify
180 whether the genomic coordinates of the QSS(s) fell within prophage regions (genomes of lysogenic
181 phages inserted within host genomes) and should be thus considered as viral instead of bacterial.
182 Of the 3529 conservative candidate RRNPP QSSs, we found that 2793 are chromosomal (on 941
183 different chromosomes), 222 are plasmidic (on 172 different plasmids), 505 are predicted to be
184 prophage-encoded (on 219 predicted intact prophages, 102 questionable prophages and 176
185 incomplete prophages) and 9 are encoded by genomes isolated from free temperate
186 bacteriophages (on 8 distinct bacteriophages) (**Table S2**). Although the presence of multiple QSSs
187 on single chromosomes is not rare (8,21) due to the selective pressure that may exist for the
188 acquisition of additional subpopulation-specific QSSs (41,42), only 1 MGE (Bacillus phage phi3T)
189 has been previously reported to encode two RRNPP QSSs (43). Here, we report the identification
190 of 30 plasmids and 8 (pro)phages encoding multiple candidate RRNPP QSSs (up to 5 in the
191 pBMB400 plasmid of *Bacillus thuringiensis* serovar kurstaki str. YBR-1520) (**Table S3**).
192 Interestingly, multiple QSSs might enable a MGE to produce distinct communication peptides that
193 accumulate differentially in the medium, hence to sense and react to multiple population density
194 levels (43).

195

196 **RRNPP_detector massively expands RRNPP QSS diversity**

197 In order to classify the 3529 conservative candidate RRNPP QSSs, the sequences of the
198 candidate receptors were BLASTed against each other to identify groups of homologs, defined as
199 connected components in the resulting sequence similarity network (wherein two nodes (proteins)

200 are connected if they show a sequence identity $\geq 30\%$ over more than 80% of the lengths of both
201 sequences, as in (44) (**Fig. 3**). We report the identification of 31 groups of homologous receptors
202 (with at least 2 homologous sequences in it) and of 29 singletons (any receptor with no clear
203 homology to others) (**Table S2** and **Fig. 3**). If a singleton or at least one sequence in a group of
204 homologous receptors was matched by an experimentally-validated receptor in a BLASTp search
205 (same thresholds as above), we considered this group as already known. Of the 60 different types
206 of receptors detected, 8 were hence considered as already known: the Rap + NprR group
207 (N=2991), PlcR (N=279), AimR of *B. subtilis* phages (N=20), AloR13 (N=16), AimR of *B. cereus*
208 phages (N=10), TraA (N=5), Rgg2 (N=4) and QsrB (N=2) (**Fig. 3** and **Fig. 4**). Our results thus
209 increase the diversity of RRNPP QSSs by a factor of 7.5 and paves the way to the identification of
210 multiple novel peptide-based biocommunication “languages” (**Fig. 4**). Importantly, these different
211 groups of QSSs cover a substantial phylogenetic diversity. Indeed, they are distributed across 21
212 bacterial taxonomic families (**Fig. 4**), some of which include notorious pathogenicity-associated
213 genera like *Clostridium*, *Bacillus*, *Enterococcus* or *Streptococcus* (**Table S2**). Functional
214 investigations of the candidate RRNPP QSSs encoded by these genera could hence potentially
215 unearth novel density-dependent mechanisms linked to pathogenicity. Moreover, this analysis
216 reveals that some of the previously identified poly-encoding QSSs MGEs (**Table S3**) encode
217 candidate RRNPP QSSs that belong to distinct groups/gene families (e.g. Rap and AimR in
218 bacteriophages of *B. subtilis*; AloR13 and group5 in pBRLA3 plasmid of *B. laterosporus* (**Fig. 4**))
219 and could be thus considered as ‘bilingual’. From a host-MGE coevolution perspective, it is also
220 interesting to note that besides the two lysis-lysogeny regulating AimR groups that should be 100%
221 viral (Phaster failed to predict a prophage origin for 8 QSSs (unlike Prophage Hunter)), 6 groups of
222 homologous QSSs were found to be shared between chromosomes and MGEs (**Fig. 3**). This
223 suggests that some QSSs may be externalized between chromosomes, plasmids, and phages
224 (45), with the possibility that such QSSs interfere with the regulation of genes in different genomes/
225 genetic elements. Typically, QSS families shared by MGEs and chromosomes can give rise to
226 density-dependent manipulations of bacterial hosts by MGEs, as exemplified in our previous study
227 on the host-derived Rap-Phr QSS of *Bacillus* phage phi3T (43) or in the study of Silpe and Bassler
228 on the host-derived VqmA quorum sensing receptor of *Vibrio* phage VP882 (46).

229

230 **Post-processing functional prediction of candidate RRNPP QSS reveals 50 putative QSS-** 231 **regulated biosynthetic gene clusters**

232 QSSs with a one component receptor module tend to be located nearby their target genes (47–
233 49) and RRNPP QSSs are no exceptions to this trend, especially not MGE-encoded RRNPP QSSs
234 (as shown in **Fig. 1**). The genomic neighborhood of a RRNPP QSS can thus provide insights on
235 the potential function(s) regulated in a density-dependent by the QSS, as suggested in our
236 previous study on the functional inference of viral RRNPP QSSs

237 (<https://www.biorxiv.org/content/10.1101/2021.07.15.452460v1>). For instance, one function
238 regulated by quorum sensing typically is the production of public good metabolites, e.g.
239 antimicrobial compounds, because only a collective production may bring such molecules to the
240 concentration levels required to exert a significant effect on the microbial community (50).
241 Consistently, many natural product biosynthetic gene clusters (BGCs) have been demonstrated to
242 be controlled by a QSS located in their genomic vicinity, be it a molecule-based QSS (48,49) or a
243 peptide-based QSS (**Fig. 5A**) (24,25). To illustrate how the output of RRNPP_detector can be
244 coupled with functional insights, we chose to focus on this practical example of putative QSS-
245 regulated BGCs. Indeed, as a major challenge in the field of natural-product discovery is that many
246 BGCs are not expressed under laboratory growth conditions (51), QSS-regulated BGCs are
247 promising: they may represent a source of novel natural products for which the link to population
248 density provides some understanding about how to elicit production. Hence, to identify candidate
249 QSS-regulated BGCs, we first searched for BGCs with antiSMASH standalone version 6.0.0
250 (default parameters) (52) in the 474 genetic elements encoding at least one candidate RRNPP
251 QSS with a receptor detected as a transcription factor (harboring a DNA binding domain). We then
252 intersected the list of the 3153 BGCs detected by antiSMASH with our list of candidate RRNPP
253 QSSs on the basis of the inclusion of the QSS region (from the start codon of the first gene to the
254 stop codon of the second gene) within the region of a BGC defined by antiSMASH. This resulted in
255 a subset of 46 candidate BGCs inferred to be regulated by RRNPP QSSs (**Table S4** and **Fig. 5**).
256 Among these 46 putative QSS-regulated BGCs, 3 are plasmidic and are inferred by antiSMASH to
257 produce antimicrobial peptides (**Fig. 5B**). As these plasmidic RRNPP QSSs likely enact the
258 production of defense metabolites only when the quorum of plasmids is met, they might create a
259 selective pressure for the acquisition of the plasmid by host cells at high plasmid densities,
260 supporting complex scenarios of co-evolution. Interestingly, the 46 putative QSS-regulated BGCs
261 produce major classes of natural products (**Table S4** and **Fig. 5C**), including ribosomally
262 synthesized and posttranslationally-modified peptides (RiPPs), non-ribosomal peptides and
263 polyketides. Worth to note, RiPPs are of most frequent occurrence, likely reflecting the important
264 roles of RiPPs in bacterial physiology (53). Last but not least, the inference that these BGCs are
265 expressed at high population densities hints at important ecological roles for their biosynthesized
266 natural products, among which could be novel defense metabolites of medical relevance.

267

268 **CONCLUSION AND FUTURE PERSPECTIVES**

269 RRNPP_detector is able to predict a wide range of candidate RRNPP QSSs in chromosomes or in
270 mobile genetic elements (e.g. plasmids, phages) of gram-positive bacteria (e.g. Firmicutes).
271 Because RRNPP_detector does not rely on sequence homology to known QSS components to
272 detect candidate QSS gene families, it can detect novel families of RRNPP QSSs, as exemplified
273 in this article. Moreover, as RRNPP QSSs tend to regulate adjacent genes in a density-dependent

274 manner, post-processing genomic context analyses can couple RRNPP_detector output with
275 functional insights, as shown in this article with the specific example of putative QSS-regulated
276 BGCs.

277

278 RRNPP_detector presents a first step towards the large-scale identification of peptide-based
279 QSSs. However, this software is dedicated to the sole detection RRNPP QSSs and we already
280 anticipate to expand its usage to peptide-based QSSs with two-component systems. We believe
281 that our work will unlock new biological knowledge regarding peptide-based biocommunication and
282 will reveal many novel density-dependent decision-making processes in bacteria, plasmids and
283 bacteriophages that could likely transform our views of microbial adaptation and bacteria-MGE co-
284 evolution. Importantly, Firmicutes represent with Bacteroidetes the most prevalent phyla in human
285 gut microbiomes (54), which makes the use of RRNPP_detector particularly relevant against
286 human-associated metagenomics-assembled genomes or mobile genetic elements (e.g. from the
287 human MGE database (55) or the Gut Phage Database (56)) to infer density-dependent behaviors
288 that may take place within human intestinal microbiomes, plasmidomes and viromes.

289

290 **DATA AVAILABILITY**

291 RRNPP_detector is freely available on github at https://github.com/TeamAIRE/RRNPP_detector

292

293 **FUNDING**

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

297

298 **REFERENCES**

- 299 1. Mukherjee S, Bassler BL. Bacterial quorum sensing in complex and dynamically changing
300 environments. *Nat Rev Microbiol* [Internet]. 2019 Jun 3 [cited 2019 Jun 12];17(6):371–82.
301 Available from: <http://www.nature.com/articles/s41579-019-0186-5>
- 302 2. Papenfort K, Bassler BL. Quorum sensing signal-response systems in Gram-negative
303 bacteria. *Nat Rev Microbiol* [Internet]. 2016 [cited 2019 May 11];14(9):576–88. Available
304 from: <http://www.ncbi.nlm.nih.gov/pubmed/27510864>
- 305 3. Polkade A V., Mantri SS, Patwekar UJ, Jangid K. Quorum sensing: An under-explored
306 phenomenon in the phylum Actinobacteria. Vol. 7, *Frontiers in Microbiology*. Frontiers Media
307 S.A.; 2016.
- 308 4. Bhatt VS. Quorum sensing mechanisms in gram positive bacteria. In: *Implication of Quorum*
309 *Sensing System in Biofilm Formation and Virulence*. Springer Singapore; 2019. p. 297–311.

- 310 5. Sturme MHJ, Kleerebezem M, Nakayama J, Akkermans ADL, Vaughan EE, De Vos WM.
311 Cell to cell communication by autoinducing peptides in gram-positive bacteria. *Antonie van*
312 *Leeuwenhoek*, Int J Gen Mol Microbiol. 2002;81(1–4):233–43.
- 313 6. Rocha-Estrada J, Aceves-Diez AE, Guarneros G, de la Torre M. The RNPP family of
314 quorum-sensing proteins in Gram-positive bacteria. *Appl Microbiol Biotechnol* [Internet].
315 2010 Jul 26 [cited 2019 Oct 16];87(3):913–23. Available from:
316 <http://link.springer.com/10.1007/s00253-010-2651-y>
- 317 7. Neiditch MB, Capodagli GC, Prehna G, Federle MJ. Genetic and Structural Analyses of
318 RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria [Internet]. Vol. 51, Annual
319 Review of Genetics. Annual Reviews Inc.; 2017 [cited 2020 Nov 19]. p. 311–33. Available
320 from: <https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-120116-023507>
- 321 8. Do H, Kumaraswami M. Structural Mechanisms of Peptide Recognition and Allosteric
322 Modulation of Gene Regulation by the RRNPP Family of Quorum-Sensing Regulators. *J Mol*
323 *Biol* [Internet]. 2016 Jul 17 [cited 2019 Oct 16];428(14):2793–804. Available from:
324 <http://www.ncbi.nlm.nih.gov/pubmed/27283781>
- 325 9. Perez-Pascual D, Monnet V, Gardan R. Bacterial Cell-Cell Communication in the Host via
326 RRNPP Peptide-Binding Regulators. *Front Microbiol* [Internet]. 2016 [cited 2019 Oct
327 16];7:706. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27242728>
- 328 10. Baldwin VM. You Can't *B. cereus* – A Review of *Bacillus cereus* Strains That Cause Anthrax-
329 Like Disease. Vol. 11, *Frontiers in Microbiology*. Frontiers Media S.A.; 2020. p. 1731.
- 330 11. Lannes-Costa PS, de Oliveira JSS, da Silva Santos G, Nagao PE. A current review of
331 pathogenicity determinants of *Streptococcus* sp. [Internet]. *Journal of Applied Microbiology*.
332 Blackwell Publishing Ltd; 2021 [cited 2021 Jul 26]. Available from:
333 <https://pubmed.ncbi.nlm.nih.gov/33772968/>
- 334 12. Mallozzi M, Viswanathan VK, Vedantam G. Spore-forming Bacilli and Clostridia in human
335 disease [Internet]. Vol. 5, *Future Microbiology*. Future Medicine Ltd.; 2010 [cited 2021 Mar
336 12]. p. 1109–23. Available from: <https://pubmed.ncbi.nlm.nih.gov/20632809/>
- 337 13. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: A common cause of persistent
338 infections [Internet]. Vol. 284, *Science*. American Association for the Advancement of
339 Science; 1999 [cited 2021 Jul 26]. p. 1318–22. Available from:
340 <https://science.sciencemag.org/content/284/5418/1318>
- 341 14. Høiby N, Ciofu O, Johansen HK, Song ZJ, Moser C, Jensen PØ, et al. The clinical impact of
342 bacterial biofilms. In: *International Journal of Oral Science* [Internet]. *Int J Oral Sci*; 2011
343 [cited 2021 Jul 26]. p. 55–65. Available from: <https://pubmed.ncbi.nlm.nih.gov/21485309/>
- 344 15. Galié S, García-Gutiérrez C, Miguélez EM, Villar CJ, Lombó F. Biofilms in the food industry:
345 Health aspects and control methods [Internet]. Vol. 9, *Frontiers in Microbiology*. Frontiers
346 Media S.A.; 2018 [cited 2021 Jul 26]. Available from: [/pmc/articles/PMC5949339/](https://pmc/articles/PMC5949339/)
- 347 16. Von Wintersdorff CJH, Penders J, Van Niekerk JM, Mills ND, Majumder S, Van Alphen LB,
348 et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal
349 gene transfer [Internet]. Vol. 7, *Frontiers in Microbiology*. Frontiers Media S.A.; 2016 [cited
350 2021 Jul 26]. p. 173. Available from: [/pmc/articles/PMC4759269/](https://pmc/articles/PMC4759269/)

- 351 17. Kohler V, Keller W, Grohmann E. Regulation of gram-positive conjugation [Internet]. Vol. 10,
352 Frontiers in Microbiology. Frontiers Media S.A.; 2019 [cited 2021 Mar 12]. p. 1134. Available
353 from: www.frontiersin.org
- 354 18. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, et al.
355 Communication between viruses guides lysis-lysogeny decisions. Nature [Internet]. 2017
356 [cited 2019 Jul 4];541(7638):488–93. Available from:
357 <http://www.ncbi.nlm.nih.gov/pubmed/28099413>
- 358 19. Stokar-Avihail A, Tal N, Erez Z, Lopatina A, Sorek R. Widespread Utilization of Peptide
359 Communication in Phages Infecting Soil and Pathogenic Bacteria. Cell Host Microbe
360 [Internet]. 2019 May 8 [cited 2019 Nov 24];25(5):746–755.e5. Available from:
361 <http://www.ncbi.nlm.nih.gov/pubmed/31071296>
- 362 20. Kotte AK, Severn O, Bean Z, Schwarz K, Minton NP, Winzer K. RRNPP-type quorum
363 sensing affects solvent formation and sporulation in clostridium acetobutylicum. Microbiol
364 (United Kingdom). 2020;166(6):579–92.
- 365 21. Voichek M, Maaß S, Kroniger T, Becher D, Sorek R. Peptide-based quorum sensing
366 systems in Paenibacillus polymyxa. Life Sci Alliance [Internet]. 2020 Oct 1 [cited 2021 Apr
367 22];3(10). Available from: <https://doi.org/10.26508/lsa.202000847>
- 368 22. Feng J, Zong W, Wang P, Zhang ZT, Gu Y, Dougherty M, et al. RRNPP-Type quorum-
369 sensing systems regulate solvent formation, sporulation and cell motility in Clostridium
370 saccharoperbutylacetonicum. Biotechnol Biofuels [Internet]. 2020;13(1):1–16. Available
371 from: <https://doi.org/10.1186/s13068-020-01723-x>
- 372 23. Parthasarathy S, Jordan LD, Schwarting N, Woods MA, Abdullahi Z, Varahan S, et al.
373 Involvement of chromosomally encoded homologs of the RRNPP protein family in
374 enterococcus faecalis biofilm formation and urinary tract infection pathogenesis. J Bacteriol
375 [Internet]. 2020 Sep 1 [cited 2021 Jul 18];202(17). Available from: <https://doi.org/10.1128/JB>
- 376 24. Hoover SE, Perez AJ, Tsui HCT, Sinha D, Smiley DL, Dimarchi RD, et al. A new quorum-
377 sensing system (TprA/PhrA) for Streptococcus pneumoniae D39 that regulates a lantibiotic
378 biosynthesis gene cluster. Mol Microbiol [Internet]. 2015 Jul 1 [cited 2021 Jul 18];97(2):229–
379 43. Available from: [/pmc/articles/PMC4676566/](http://pmc/articles/PMC4676566/)
- 380 25. Rued BE, Covington BC, Bushin LB, Szewczyk G, Laczko V, Seyedsayamdost MR, et al.
381 Quorum sensing in streptococcus mutans regulates production of tryglysin, a novel ras-ripp
382 antimicrobial compound. MBio. 2021 Mar 1;12(2):1–25.
- 383 26. Edwards AN, Tamayo R, McBride SM. A novel regulator controls Clostridium difficile
384 sporulation, motility and toxin production. Mol Microbiol [Internet]. 2016 Jun 1 [cited 2021 Jul
385 26];100(6):954–71. Available from: [/pmc/articles/PMC4899322/](http://pmc/articles/PMC4899322/)
- 386 27. Do H, Makthal N, VanderWal AR, Rettel M, Savitski MM, Peschek N, et al. Leaderless
387 secreted peptide signaling molecule alters global gene expression and increases virulence
388 of a human bacterial pathogen. Proc Natl Acad Sci U S A [Internet]. 2017 Oct 3 [cited 2021
389 Jul 26];114(40):E8498–507. Available from:
390 www.pnas.org/cgi/doi/10.1073/pnas.1705972114

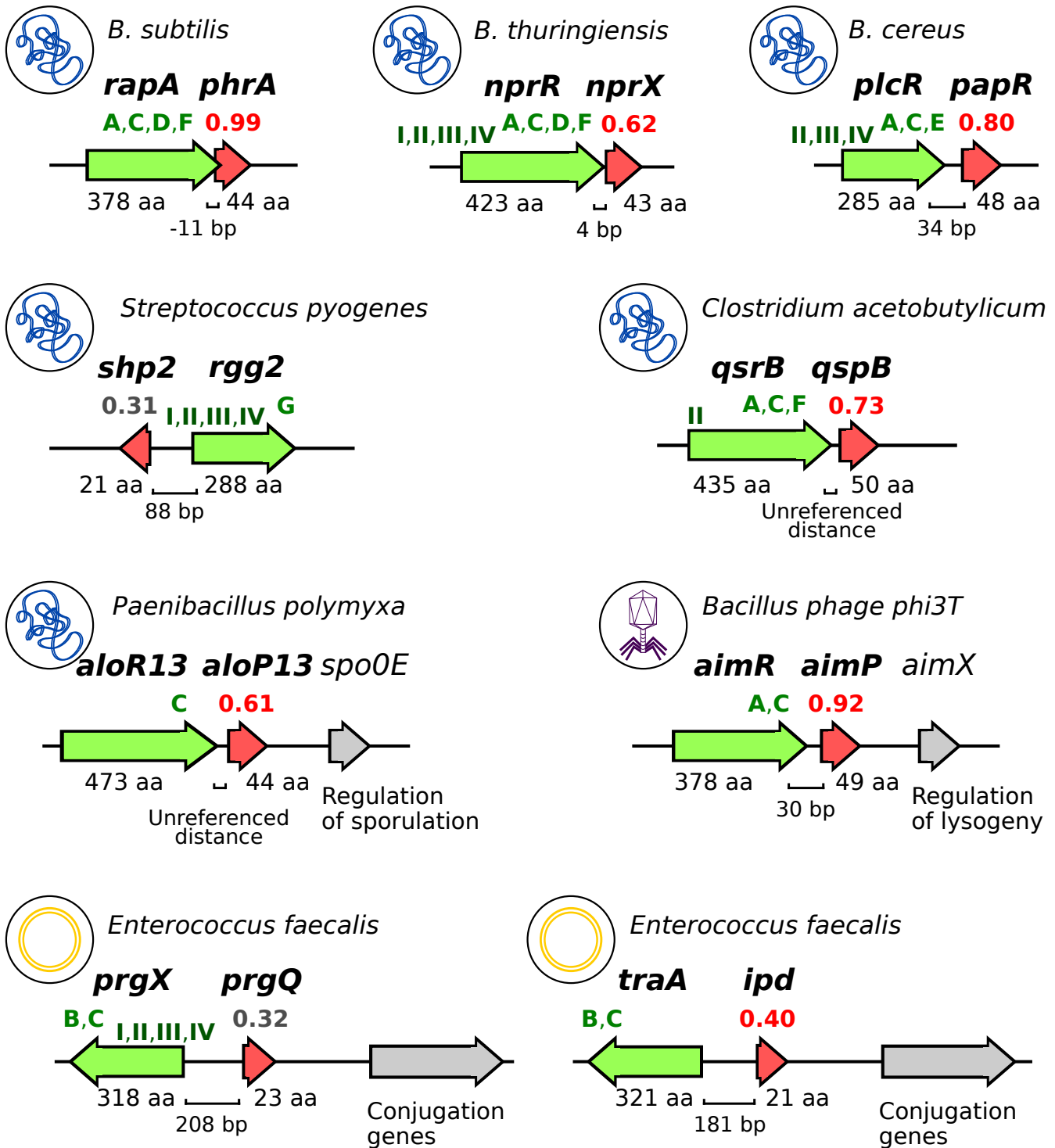
- 391 28. Edwards AN, Anjuwon-Foster BR, McBride SM. Rsta is a major regulator of clostridioides
392 difficile toxin production and motility. *MBio* [Internet]. 2019 [cited 2021 Jul 26];10(2):1–21.
393 Available from: <https://doi.org/10>
- 394 29. Do H, Makthal N, VanderWal AR, Saavedra MO, Olsen RJ, Musser JM, et al. Environmental
395 pH and peptide signaling control virulence of *Streptococcus pyogenes* via a quorum-sensing
396 pathway. *Nat Commun* [Internet]. 2019 Dec 1 [cited 2021 Jul 26];10(1):1–14. Available from:
397 <https://doi.org/10.1038/s41467-019-10556-8>
- 398 30. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, et al. CATH: Increased
399 structural coverage of functional space. *Nucleic Acids Res* [Internet]. 2021 Jan 8 [cited 2021
400 Jul 26];49(D1):D266–73. Available from: [/pmc/articles/PMC7778904/](https://pubmed.ncbi.nlm.nih.gov/37778904/)
- 401 31. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY -
402 Sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic
403 Acids Res* [Internet]. 2009 Jan 1 [cited 2021 Jul 26];37(SUPPL. 1):D380–6. Available from:
404 https://academic.oup.com/nar/article/37/suppl_1/D380/1009184
- 405 32. Letunic I, Khedkar S, Bork P. SMART: Recent updates, new developments and status in
406 2020. *Nucleic Acids Res* [Internet]. 2021 Jan 8 [cited 2021 Jul 26];49(D1):D458–60.
407 Available from: <https://academic.oup.com/nar/article/49/D1/D458/5940513>
- 408 33. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al.
409 Pfam: The protein families database in 2021. *Nucleic Acids Res* [Internet]. 2021 Jan 8 [cited
410 2021 Jul 26];49(D1):D412–9. Available from: <https://covid-19>.
- 411 34. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and genome
412 properties in 2013. *Nucleic Acids Res* [Internet]. 2013 Jan 1 [cited 2021 Jul 26];41(D1).
413 Available from: <https://pubmed.ncbi.nlm.nih.gov/23197656/>
- 414 35. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et
415 al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat
416 Biotechnol* [Internet]. 2019 Apr 18 [cited 2019 Oct 16];37(4):420–3. Available from:
417 <http://www.nature.com/articles/s41587-019-0036-z>
- 418 36. Eddy SR. Accelerated Profile HMM Searches. Pearson WR, editor. *PLoS Comput Biol*
419 [Internet]. 2011 Oct 20 [cited 2019 May 28];7(10):e1002195. Available from:
420 <https://dx.plos.org/10.1371/journal.pcbi.1002195>
- 421 37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J
422 Mol Biol* [Internet]. 1990 Oct 5 [cited 2019 Oct 16];215(3):403–10. Available from:
423 <http://www.ncbi.nlm.nih.gov/pubmed/2231712>
- 424 38. Junier I, Unal EB, Yus E, Lloréns-Rico V, Serrano L. Insights into the Mechanisms of Basal
425 Coordination of Transcription Using a Genome-Reduced Bacterium. *Cell Syst*. 2016 Jun
426 22;2(6):391–401.
- 427 39. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic
428 gene recognition and translation initiation site identification. *BMC Bioinformatics* [Internet].
429 2010 Mar 8 [cited 2021 Jul 21];11. Available from:
430 <https://pubmed.ncbi.nlm.nih.gov/20211023/>

- 431 40. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster
432 version of the PHAST phage search tool. *Nucleic Acids Res* [Internet]. 2016 [cited 2020 Aug
433 4];44(Web Server issue):W16. Available from:
434 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987931/>
- 435 41. Even-Tov E, Omer Bendori S, Pollak S, Eldar A. Transient Duplication-Dependent
436 Divergence and Horizontal Transfer Underlie the Evolutionary Dynamics of Bacterial Cell–
437 Cell Signaling. Gore J, editor. *PLOS Biol* [Internet]. 2016 Dec 29 [cited 2019 Oct
438 17];14(12):e2000330. Available from: <http://dx.plos.org/10.1371/journal.pbio.2000330>
- 439 42. Kalamara M, Spacapan M, Mandic-Mulec I, Stanley-Wall NR. Social behaviours by *Bacillus*
440 *subtilis*: quorum sensing, kin discrimination and beyond. *Mol Microbiol* [Internet]. 2018 [cited
441 2019 Oct 16];110(6):863. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30218468>
- 442 43. Bernard C, Li Y, Lopez P, Bapteste E. Beyond arbitrium: identification of a second
443 communication system in *Bacillus* phage phi3T that may regulate host defense
444 mechanisms. *ISME J* [Internet]. 2020; Available from: [http://dx.doi.org/10.1038/s41396-020-](http://dx.doi.org/10.1038/s41396-020-00795-9)
445 [00795-9](http://dx.doi.org/10.1038/s41396-020-00795-9)
- 446 44. Méheust R, Watson AK, Lapointe FJ, Papke RT, Lopez P, Bapteste E. Hundreds of novel
447 composite genes and chimeric genes with bacterial origins contributed to haloarchaeal
448 evolution. *Genome Biol*. 2018 Jun 7;19(1).
- 449 45. Corel E, Méheust R, Watson AK, McInerney JO, Lopez P, Bapteste E. Bipartite Network
450 Analysis of Gene Sharings in the Microbial World. *Mol Biol Evol* [Internet]. 2018 [cited 2019
451 Nov 24];35(4):899–913. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29346651>
- 452 46. Silpe JE, Bassler BL. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage
453 Lysis-Lysogeny Decision. *Cell* [Internet]. 2019 Jan 10 [cited 2019 Jun 12];176(1–2):268–
454 280.e13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30554875>
- 455 47. Engebrecht J, Nealson K, Silverman M. Bacterial bioluminescence: Isolation and genetic
456 analysis of functions from *Vibrio fischeri*. *Cell* [Internet]. 1983 [cited 2021 Jul 28];32(3):773–
457 81. Available from: <https://pubmed.ncbi.nlm.nih.gov/6831560/>
- 458 48. Brotherton CA, Medema MH, Greenberg EP. luxR Homolog-Linked Biosynthetic Gene
459 Clusters in Proteobacteria . *mSystems* [Internet]. 2018 Jun 26 [cited 2021 Jul 18];3(3).
460 Available from: </pmc/articles/PMC5872303/>
- 461 49. He H, Ye L, Li C, Wang H, Guo X, Wang X, et al. SbbR/SbbA, an Important ArpA/AfsA-Like
462 System, Regulates Milbemycin Production in *Streptomyces bingchenggensis*. *Front*
463 *Microbiol* [Internet]. 2018 May 23 [cited 2021 Jul 28];9(MAY):1064. Available from:
464 <https://www.frontiersin.org/article/10.3389/fmicb.2018.01064/full>
- 465 50. Heilmann S, Krishna S, Kerr B. Why do bacteria regulate public goods by quorum sensing?–
466 How the shapes of cost and benefit functions determine the form of optimal regulation. *Front*
467 *Microbiol* [Internet]. 2015 [cited 2021 Jul 28];6(AUG):767. Available from:
468 </pmc/articles/PMC4517451/>
- 469 51. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent
470 biosynthetic gene clusters [Internet]. Vol. 13, *Nature Reviews Microbiology*. Nature




- 471 Publishing Group; 2015 [cited 2021 Aug 17]. p. 509–23. Available from:
472 <https://pubmed.ncbi.nlm.nih.gov/26119570/>
- 473 52. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al.
474 antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res*
475 [Internet]. 2021 Jul 2 [cited 2021 Jul 28];49(W1):W29–35. Available from:
476 <https://doi.org/10.1093/nar/gkab335>
- 477 53. Li Y, Rebuffat S. The manifold roles of microbial ribosomal peptide-based natural products in
478 physiology and ecology [Internet]. Vol. 295, *Journal of Biological Chemistry*. American
479 Society for Biochemistry and Molecular Biology Inc.; 2020 [cited 2021 Aug 17]. p. 34–54.
480 Available from: [/pmc/articles/PMC6952617/](https://pubmed.ncbi.nlm.nih.gov/36952617/)
- 481 54. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, et al. Health and disease
482 markers correlate with gut microbiome composition across thousands of people. *Nat*
483 *Commun* [Internet]. 2020 Dec 1 [cited 2021 Jul 18];11(1):1–12. Available from:
484 <https://doi.org/10.1038/s41467-020-18871-1>
- 485 55. Lai S, Jia L, Subramanian B, Pan S, Zhang J, Dong Y, et al. mMGE: A database for human
486 metagenomic extrachromosomal mobile genetic elements. *Nucleic Acids Res* [Internet].
487 2021 Jan 8 [cited 2021 Jul 18];49(D1):D783–91. Available from:
488 <https://academic.oup.com/nar/article/49/D1/D783/5930394>
- 489 56. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive
490 expansion of human gut bacteriophage diversity. *Cell* [Internet]. 2021 Feb [cited 2021 Mar
491 12];184(4):1098-1109.e9. Available from:
492 <https://linkinghub.elsevier.com/retrieve/pii/S0092867421000726>
493

494 **SUPPLEMENTARY MATERIALS:**




- 495 • **Table S1:** 6467 ‘permissive’ candidate RRNPP QSSs
496 • **Table S2:** 3529 ‘conservative’ candidate RRNPP QSSs
497 • **Table S3:** 38 MGEs encoding multiple ‘conservative’ candidate RRNPP QSSs
498 • **Table S4:** 46 putative QSS-regulated BGCs



Genetic element

-  Chromosome
-  Plasmid
-  Phage genome

Genes

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

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

Matched N-terminal DNA binding domains (HMM)

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

499 **FIGURE LEGENDS**

500 **Figure 1: Common features between experimentally validated RRNPP-type QSSs.**

501 Each genomic context corresponds to the representative QSS of an experimentally characterized
502 RRNPP-type QSS family. The icon at the left of each context indicates the genetic element that
503 encodes the QSS (bacterial chromosome, phage genome or plasmid) and the associated label
504 indicates the taxon to which this genetic element belongs. The green gene corresponds to the
505 quorum sensing receptor and the red gene to its cognate propeptide, whereas a grey gene
506 indicates an adjacent, target gene (or set of genes) demonstrated to be regulated by the QSS. The
507 intergenic distance between the receptor and the propeptide genes is given in number of base
508 pairs and the length of receptors and propeptides is given in number of aminoacids. The number
509 above each pro-peptide corresponds to the likelihood, computed by SignalP, that the propeptide
510 harbors a N-terminal signal sequence for the SEC-translocon. A likelihood score colored in red
511 means that the propeptide is predicted by SignalP to be secreted via the SEC-translocon, whereas
512 a score colored in grey means that it is predicted to be secreted otherwise. Shp2 and PrgQ are the
513 only two propeptides predicted to be secreted otherwise, consistent with the fact that they are the
514 only ones known to be secreted via another secretion system than Sec(SEC/Spl), namely the
515 PptAB complex (7). The green letters above the C-terminal encoding region of each receptor
516 indicate the names of the HMM (PFAM, SMART, TIGR) or of the HMM family (CATH, SuperFamily)
517 of Tetratricopeptide repeats (TPRs) that are found within the sequence of the translated protein.
518 The roman numbers above the N-terminal encoding region of each receptor indicate the names of
519 the HMM or of the HMM family of DNA binding domains found in the sequence of the translated
520 protein.

15,901,185 PROTEIN SEQUENCES

4,974 COMPLETE GENOMES | 14,628,025 ENCODED BY FIRMICUTES | 1,273,160 ENCODED BY VIRUSES | 32,919 COMPLETE GENOMES

Filter by length
(250-500 aa)

Filter by length
(10-70 aa)

6,130,631
PROTEIN SEQUENCES

Filter by coding
sequence adjacency

1,149,358
PROTEIN SEQUENCES

602,708
PROTEIN SEQUENCES

520,653
PROTEIN SEQUENCES

Filter by TPR
(HMMsearch)

Inclusion threshold



9,354
PUTATIVE RECEPTORS

9,986
PUTATIVE PROPEPTIDES

Identification of homologs of
reference receptors (BLASTp)
Identification of transcription
factors (HMMsearch)

Filter by SP(Sec/SPI)
dependent secretion
according to SignalP

3,521
CONSERVATIVE CANDIDATE
RECEPTORS

3,525
CONSERVATIVE CANDIDATE
PROPEPTIDES

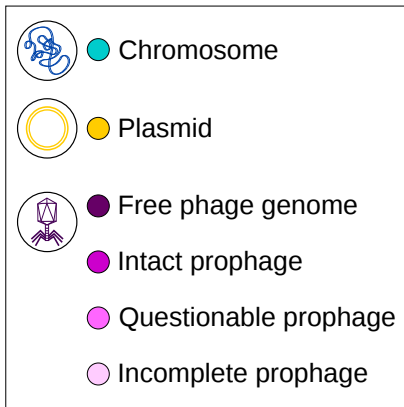
Filter by intergenic distance (<400bp)

3,529 Candidate RRNPP QSSs

521 **Figure 2: Workflow of RRNPP_detector illustrated with real data from complete genomes of**
522 **Firmicutes and viruses**

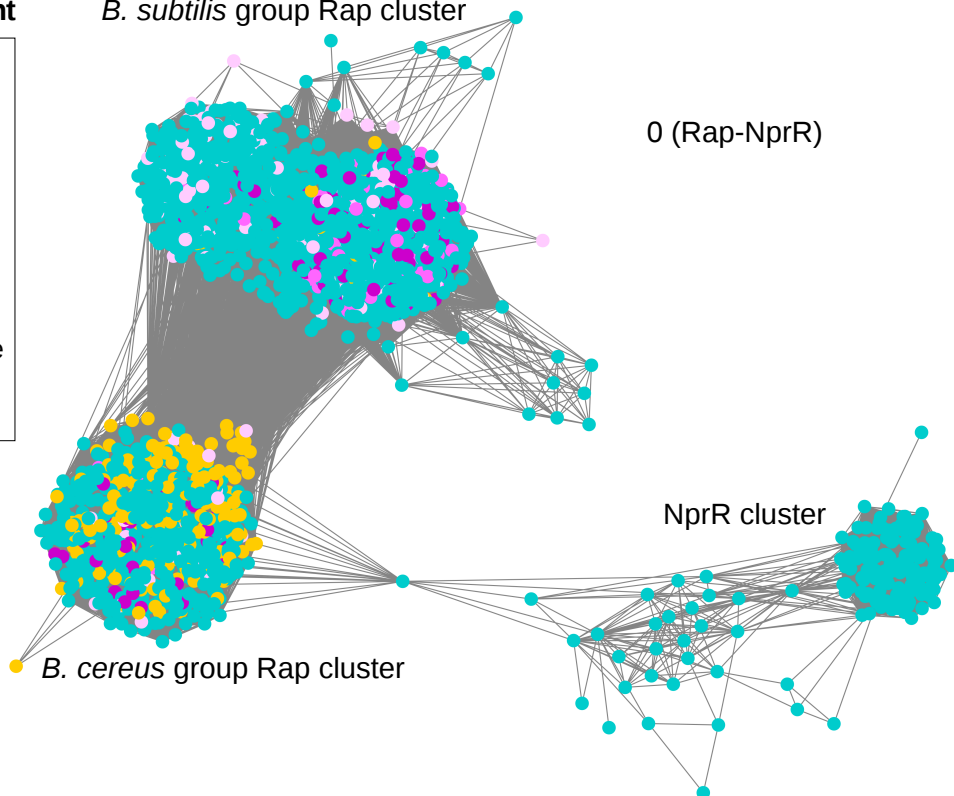
523 RRNPP_detector defines candidate RRNPP-type quorum sensing systems as tandems of adjacent
524 CDSs encoding a candidate receptor (250-500aa protein matching HMMs of peptide-binding
525 tetratricopeptide repeats (TPRs)) and a candidate pro-peptide (10-70aa protein predicted to be
526 excreted via the SEC-translocon). Each green and red rectangle represents a step towards the
527 final identification of candidate receptors and candidate propeptides, respectively. Each arrow
528 represents a filtering operation to narrow down the search space to 'conservative' candidate
529 RRNPP QSSs. The horizontal red line indicates the inclusion threshold after which all remaining
530 pairs of adjacent receptors and propeptides will be outputted: pairs that pass all the filters below the
531 inclusion threshold will be considered as 'conservative' candidate RRNPP QSSs, whereas the
532 other pairs will be considered as 'permissive' candidates. RRNPP_detector identified 3529
533 'conservative' candidate RRNPP QSSs from 3521 candidate receptors and 3525 candidate
534 propeptides because a receptor can be flanked on both side by a candidate propeptide and vice
535 versa.

QSS-encoding genetic element



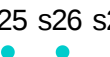
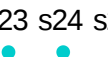
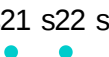
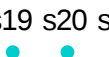
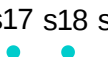
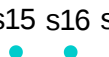
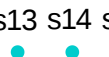
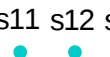
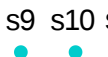
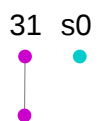
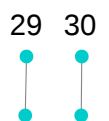
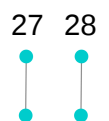
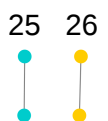
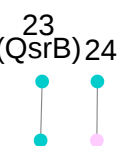
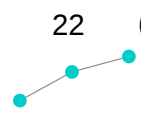
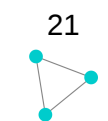
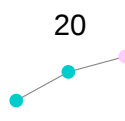
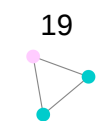
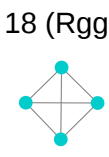
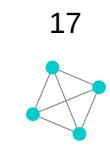
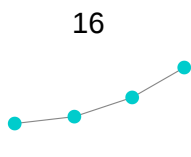
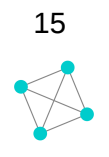
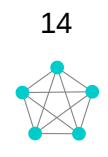
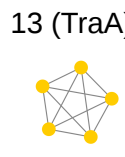
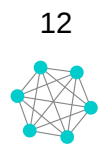
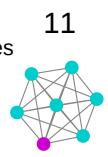
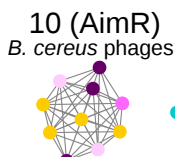
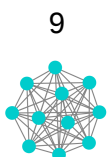
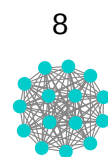
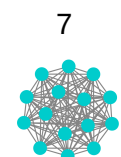
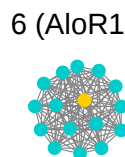
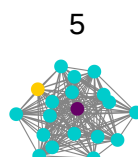
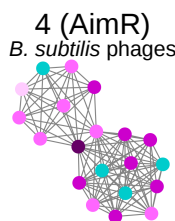
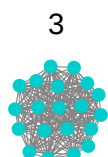
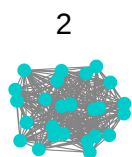
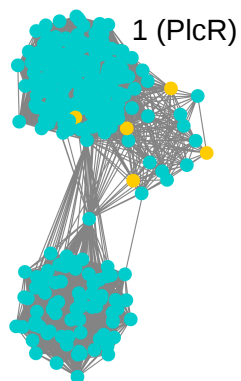
B. subtilis group Rap cluster

0 (Rap-NprR)



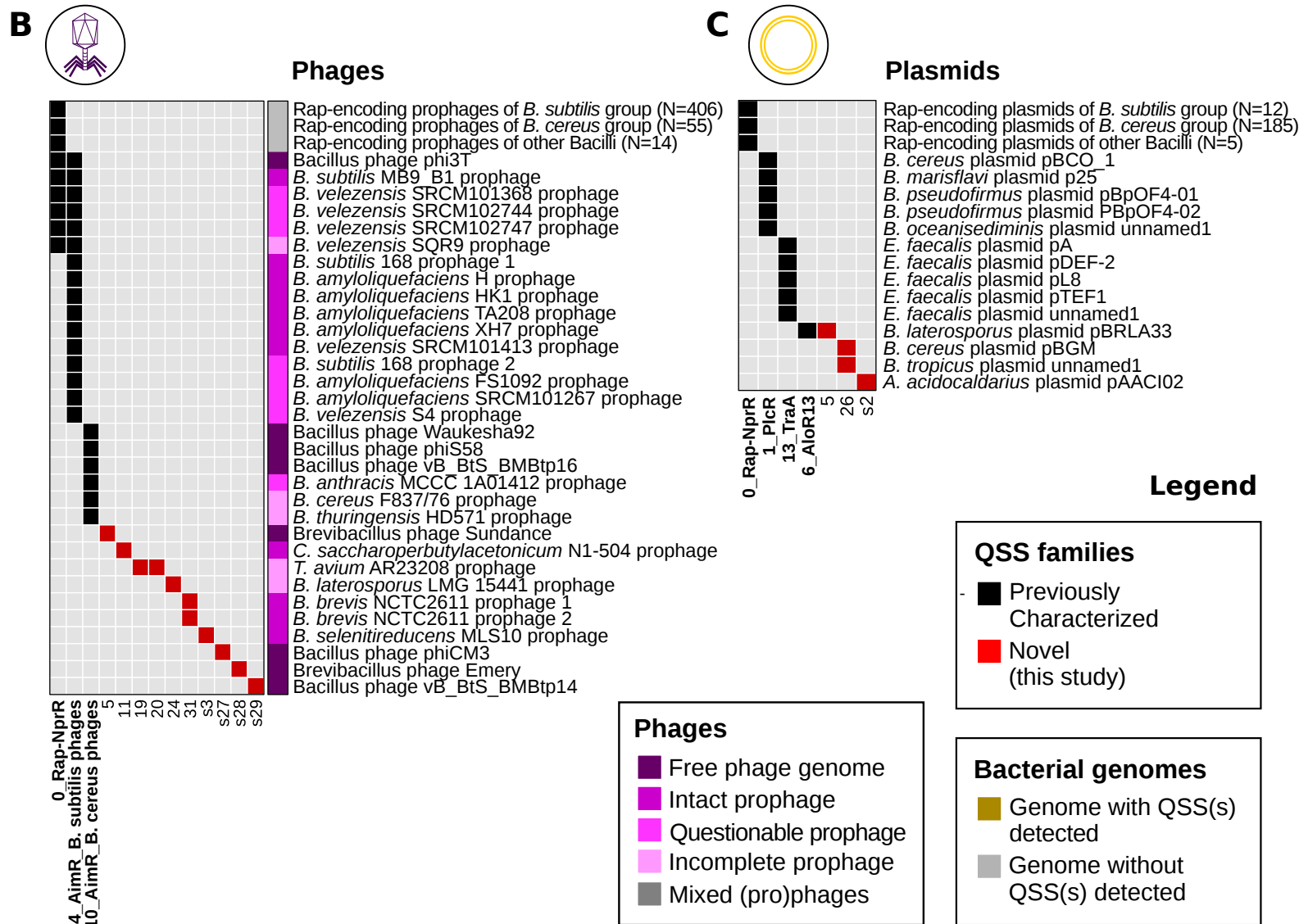
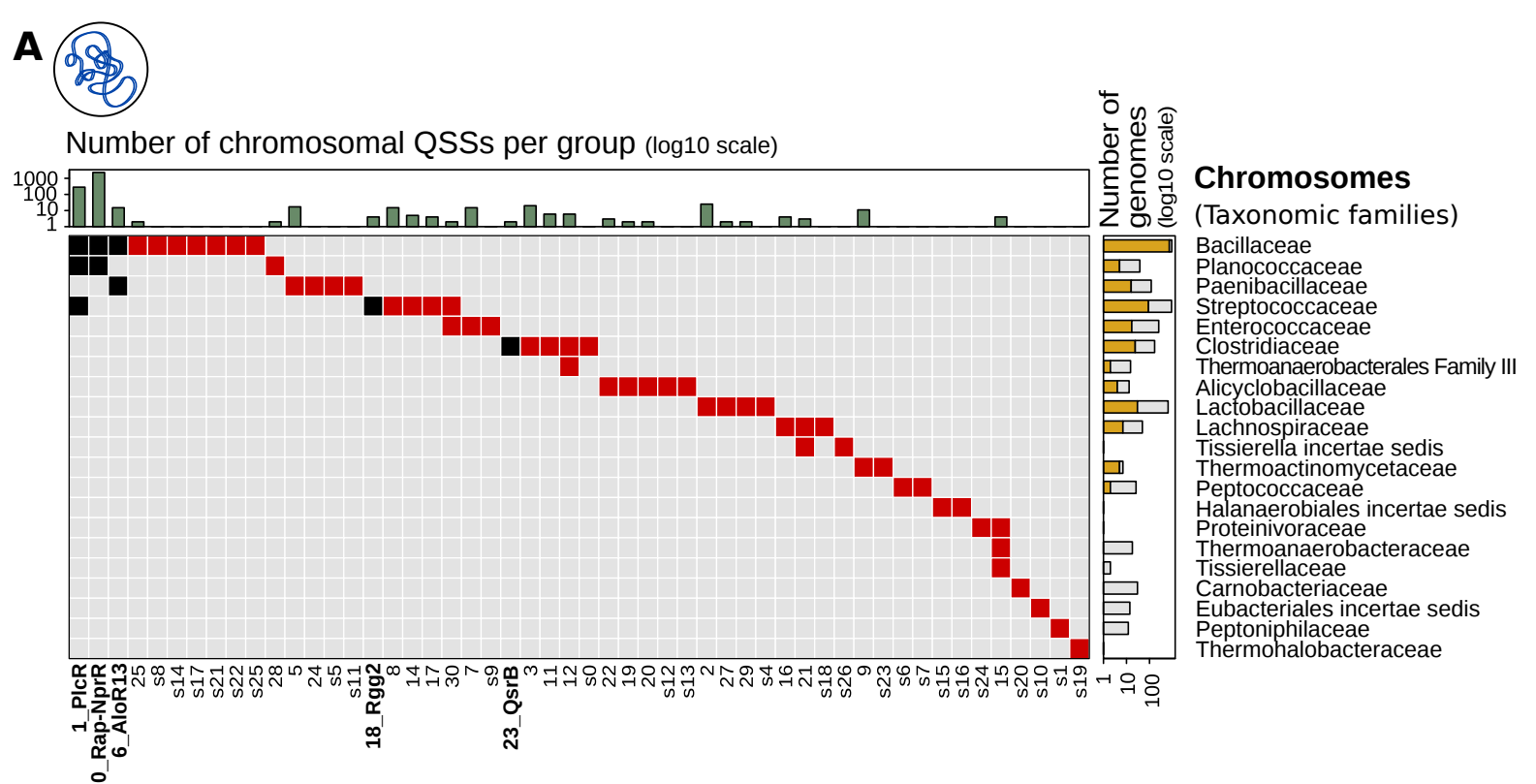
NprR cluster

B. cereus group Rap cluster



536 **Figure 3: Sequence similarity network of the receptors forming a ‘conservative’ candidate**
537 **RRNPP QSS with a cognate propeptide.**

538 Each node corresponds to a receptor sequence found adjacent to a candidate pro-peptide and is
539 colored according to the type of genetic element encoding the QSS, as displayed in the legend.
540 Each edge corresponds to a similarity link between two receptors defined according to the
541 following thresholds: percentage identity $\geq 30\%$, alignment coverage $\geq 80\%$ of the lengths of
542 both receptors, E-value $\leq 1E-5$. Each connected component of the graph thereby defines groups
543 of homologous receptors, as in (44). The groups are ordered from the largest to the smallest.
544 Labels that begin with an integer correspond to groups of homologous receptors (more than 1
545 sequence) whereas labels that begin with the ‘s’ character correspond to singletons (1 sequence
546 only). A label followed by the name of a reference receptor in brackets means that at least one
547 sequence in the group of homologous receptors is found to be homolog with a reference receptor
548 according to the above thresholds.



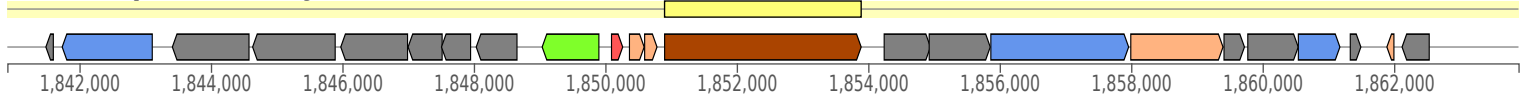
549 **Figure 4: Distribution of groups of ‘conservative’ candidate RRNPP QSSs across taxonomic**
550 **families (chromosomal QSSs) or MGEs (plasmidic and (pro)phage-encoded QSSs).**
551 **A.** Distribution of groups of RRNPP QSSs encoded by chromosomes (columns) across taxonomic
552 families of the Firmicutes phylum (rows). Grey squares mean that no QSS of a given group has
553 been detected in a given taxonomic family. Conversely, black and red squares mean that a given
554 group has been detected in at least one chromosome of a given taxonomic family. Specifically,
555 black squares highlight groups of RRNPP QSSs previously experimentally validated, whereas red
556 squares represent candidate novel groups of RRNPP QSSs. The number of chromosomal
557 representatives of each group of RRNPP QSSs is given in the histogram at the top of the heatmap
558 (log₁₀ scale). The number of chromosomes in each taxonomic family is given by the histogram at
559 the right of the heatmap (log₁₀scale): the gold area corresponds to the number of chromosomes in
560 which at least one ‘conservative’ RRNPP QSS has been identified, whereas the grey area depicts
561 the remaining chromosomes without any RRNPP QSS detected. **B.** Distribution of groups of
562 RRNPP QSSs across isolated genomes of temperate bacteriophages or prophages detected by
563 Phaster within genomes of Firmicutes. Each row represents a single phage genome or a single
564 prophage, to the exceptions of the first three rows that depict groups of prophages in which a
565 single RRNPP QSS of the Rap family has been detected. The color of each (pro)phage indicates
566 the quality of the prediction that the QSS-encoding entity is a phage: dark purple corresponds to
567 sequenced phage genomes (100% sure), whereas lighter shades of purple correspond to different
568 prophage qualities assessed by Phaster. **C.** Distribution of groups of RRNPP QSSs across
569 plasmids sequenced in Firmicutes.

A



Grp: 1_PlcR (AVN86787.1) *Streptococcus pneumoniae* D39V (CP027540.1)

Experimentally validated Class II lanthipeptide cluster like mutacin II (U40620) [13 detected similar contexts]

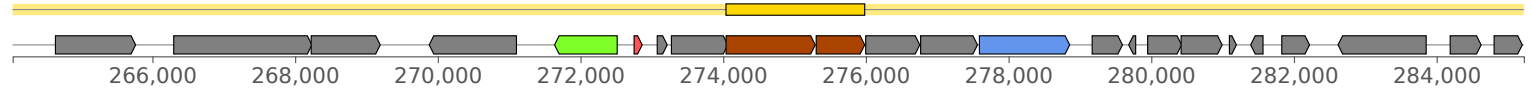


B



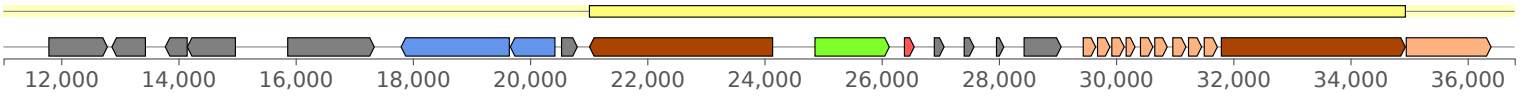
Grp: 1_PlcR (ADC52264.1) *Bacillus pseudofirmus* OF4 - plasmid pBpOF4-01 (CP001879.1)

Linear azol(in)e-containing peptides



Grp: 26 (QKH59184.1) *Bacillus cereus* ATCC 4342 - plasmid pBGM (CP053956.1)

Class II lanthipeptide cluster like mutacin II (U40620) [1 detected similar context]

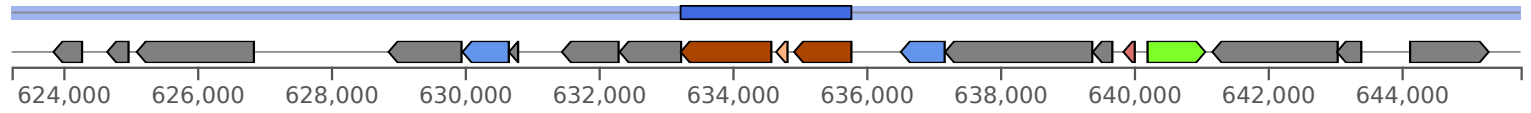


C



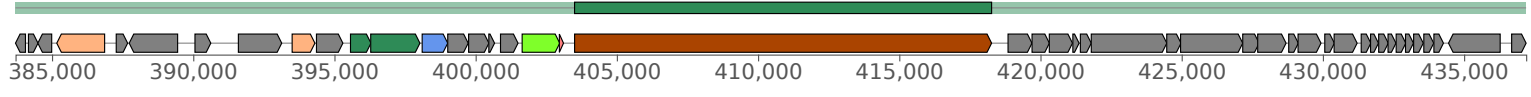
Grp: 1_PlcR (AGY39829.1) *Streptococcus* sp. I-G2 (CP006805.1)

Streptide-like thioether-bond ribosomally synthesised and post-translationally modified peptide product (RIPP) cluster



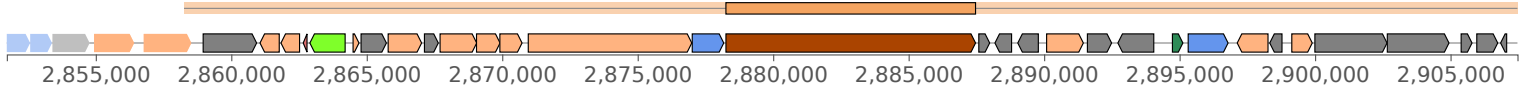
Grp: 9 (QKG83360.1) *Kroppenstedtia pulmonis* W9323 (CP048104.1)

Non-ribosomal peptide synthetase cluster



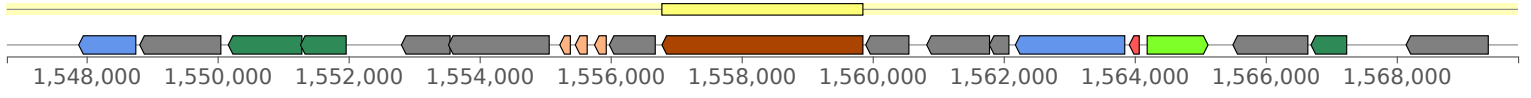
Grp: 12 (AQR95344.1) *Clostridium saccharoperbutylacetonicum* N1-504 (CP016087.1)

Type I PKS (Polyketide synthase)



Grp: 15 (QQY79072.1) *Keratinibaculum paraultunense* KD-1 (CP068564.1)

Class II lanthipeptide clusters like mutacin II (U40620)



Grp: 20 (ASS75455.1) *Tumebacillus algifaecis* THMBR28 (CP022657.1)

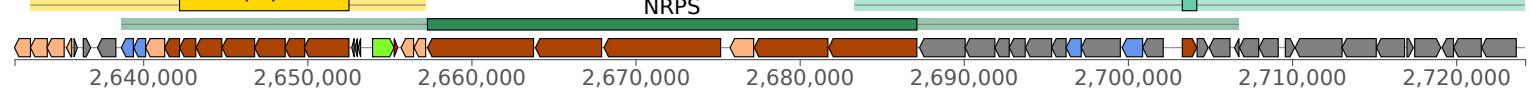
LAP

T1PKS

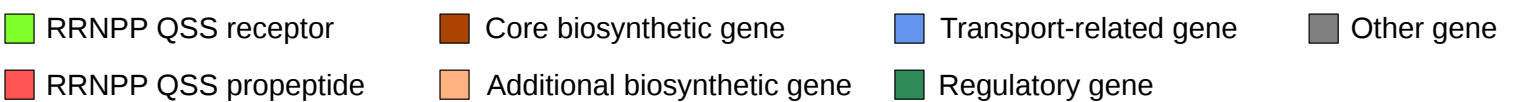
thiopeptide

NRPS

phosphonate



Types of genes



570 **Figure 5: Selection of BGCs outputed by antiSMASH that are likely regulated by candidate**
571 **RRNPP QSSs.**

572 Each genomic context corresponds to a BGC region outputed by antiSMASH within which a
573 'conservative' candidate RRNPP QSSs has been detected by RRNPP_detector. The genes are
574 colored according to the legend displayed at the bottom of the figure. For each genomic context, a
575 first line of information is composed of a thumbnail indicative of the genetic element encoding the
576 BGC region (chromosome or plasmid). Then, a 'receptor' label indicates the group of homologous
577 receptors to which the candidate RRNPP QSS receptor belongs and displays the NCBI ID of this
578 receptor in brackets. Finally, the name of the genome along with its NCBI accession are given. The
579 second line indicates the biosynthesis mode of the BGC as classified by antiSMASH, together with
580 the number of similar genomic contexts identified in the taxon. The tickmarks at the bottom of
581 each BGC correspond to genomic coordinates, given in number of base pairs. **A.** Proof of concept:
582 BGC demonstrated to be regulated by a RRNPP QSS in (24) and captured by our method. **B.**
583 Plasmidic BGCs inferred to be regulated by a candidate RRNPP QSS. The context displayed for
584 plasmid pBGM is conserved in plasmid unnamed 1 of *B. tropicus*. **C.** Selection of chromosomal
585 BGCs inferred to be regulated by candidate RRNPP QSSs to give an overview of putative QSS-
586 regulated BGC diversity.