



HAL
open science

Réseaux de communication au sein de communautés microbiennes environnementales

Charles Bernard

► **To cite this version:**

Charles Bernard. Réseaux de communication au sein de communautés microbiennes environnementales. Interactions entre organismes. Sorbonne Université, 2022. Français. NNT : 2022SORUS010 . tel-03709098

HAL Id: tel-03709098

<https://theses.hal.science/tel-03709098v1>

Submitted on 29 Jun 2022

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.



Sorbonne Université

Programme doctoral Interfaces pour le Vivant (IPV)

École doctorale Complexité du Vivant – ED515

Équipe Adaptation, Intégration, Réticulation et Évolution (AIRE) - UMR 7205 ISYEB (SU, MNHN)

Équipe Biochimie des Intéractions Microbiennes (BIM) – UMR 7245 MCAM (MNHN)

Réseaux de communication au sein de communautés microbiennes environnementales

Par Charles BERNARD

Thèse de doctorat en Bioinformatique Évolutive

Dirigée par :

Philippe LOPEZ, directeur de thèse

Yanyan LI, co-directrice de thèse

Éric BAPTESTE, co-encadrant de thèse

Présentée et soutenue publiquement le 20/01/22 devant un jury composé de:

| | |
|--|-------------|
| Dr. BAPTESTE Éric (DR CNRS, Sorbonne Université, Muséum National d'Histoire Naturelle) | Encadrant |
| Pr. DESSIMOZ Christophe (PR, Université de Lausanne) | Examineur |
| Pr. DEVAUX Frédéric (PR, Sorbonne Université) | Examineur |
| Dr. FILÉE Jonathan (CR CNRS, Paris-Saclay) | Examineur |
| Dr. LAROSE Catherine (CR CNRS, École centrale de Lyon) | Rapportrice |
| Dr. LI Yanyan (CR CNRS, Muséum National d'Histoire Naturelle) | Encadrant |
| Pr. LOPEZ Philippe (PR, Sorbonne Université, Muséum National d'Histoire Naturelle) | Encadrant |
| Dr. TOUZET Hélène (DR CNS, Université de Lille) | Rapportrice |

Remerciements

Je tiens à adresser mes premiers remerciements à Philippe Lopez, Yanyan Li et Éric Bapteste, mes trois encadrants de thèse : ma reconnaissance à votre égard est aussi immense que profondément sincère. Philippe, je souhaite te remercier pour ta bienveillance et ta gentillesse de tout instant durant ces trois années. Ma thèse aurait été tellement plus contraignante sans tes médiations, si nombreuses qu'il en deviendrait presque arrangeant de ne pouvoir les citer :-). Yanyan, je souhaite te remercier pour ta compréhension à l'égard de mes aspirations durant cette thèse, ton empathie n'a d'égal que ta gentillesse : tu avais toujours une pensée ou une attention pour Joséphine pour me le rappeler. Merci pour l'accueil toujours aussi chaleureux au MNHN, c'était à chaque fois un grand plaisir d'y venir te voir. Enfin, Éric, on dit souvent qu'il est de bon ton d'aborder les remerciements de thèse sur le ton de l'humour mais je n'ai vraiment pas le cœur au second degré pour te témoigner ma reconnaissance infinie. Je ne saurais trouver les mots pour te l'exprimer, mais j'espère que tu pourras entrevoir toute la reconnaissance que je te porte derrière ce sobre merci. Je n'oublierai jamais ton investissement pour moi, ton soutien constant, ta gentillesse, ces brainstormings fusant dans tous les sens, et ton respect de tous au labo. Je n'aurais pas pu rêver meilleure ambiance de travail pour cette thèse et je vous en suis reconnaissant à tous trois. J'espère sincèrement que nos routes se recroiseront dans le futur, ce serait un plaisir pour moi de travailler à nouveau en collaboration avec vous tous.

À l'heure de remercier mes mentors, je ne pourrais pas oublier de citer mes ex- et très estimés collègues, Fatima Mechta-Grigoriou et Yann Kieffer. Vous m'avez formé au métier de chercheur durant cette année passée ensemble à Curie et je vous en suis très reconnaissant. Merci également pour votre bienveillance à mon égard, et pour la confiance que vous m'avez accordée. Puisque j'évoque cette année passée à Curie, j'aimerais également remercier tous mes ex-collègues de l'équipe Stress et Cancer pour ces supers moments passés ensemble : Hocine, Floriane, Géraldine, Monika, Virginie et Ilaria.

Je souhaiterais maintenant remercier tous les membres de l'équipe AIRE pour cette ambiance magique au cours de ces trois années de thèse et toutes ces tranches de rigolade qui faisaient le sel de nos journées au labo. La recherche nous incite certes à trouver mais il est des équipes où cette exigence aurait pu être bien plus désagréable. Merci à Jérôme Teulière, mon acolyte indéfectible dans toutes nos excursions de gourmet, qu'il s'agisse de gastronomie japonaise, cadienne, ou turque. Merci également de m'avoir fait prendre conscience que des lunettes de geek ne seraient sans doute pas du luxe pour mes pauvres petits yeux malmenés par la pratique à haut risque de la bioinfo. Merci

à Eduardo Corel, pour tous ces moments aussi incongrus que délirants. Avec qui d'autre aurions-nous pu disserter avec tant de rigolade sur des sujets tels que l'assiette italienne du XII^{ème} siècle, la taxonomie des buffles, les fentes de Young, le guide des convenances, le pendant sucré de la poutine ou encore la reprise de « Vapeur mauve ». Merci à François Papale, notre temps passé ensemble fut bref, mais non moins mémorable. Merci de m'avoir mis Dune entre les mains et pour toutes ces séances de jeux de plateau, même si tu nous mettais la branlée à chaque fois. J'essayerai de venir te voir au Québec tantôt. Merci à Hugo Bonnefous et Duncan Sussfeld, on ne se sera pas beaucoup vus mais j'ai adoré passer du temps avec vous. J'aurais aimé pouvoir profiter davantage des jeux de mots d'Hugo et de l'humour délicieusement pince-sans-rire de Duncan, alias Mr. Idaho le végan marseillais avec un nom alsacien. Je n'oublie pas non plus les anciens membres de l'équipe, Andrew et son mythique Watson effect, Guillaume Bernard et ses boîtes de maquereaux, ainsi que Romain Lannes et la richesse de ses enseignements sur les expressions propres au ballon ovale telles que « marquer la chair » ou « châtie-le ». Merci également à Danielle pour son soutien administratif si apprécié, et ses punchs inimitables. Je souhaiterais également remercier Victor Sayous de l'équipe BIM pour sa patience et ses enseignements en microbiologie. Enfin, merci à Revolunet, mon fidèle destrier, et désolé d'avoir flingué ton disque système après trois ans de bons et loyaux services.

Merci également à tous mes amis du CRI, nous sommes sortis du master plus riches des enseignements des uns et des autres. Pour cela, je souhaiterais remercier tout particulièrement Chetan Aditya, Ellie McCallum, Louis Faure, Clovis Basier et Julie Levi-Strauss.

Je souhaiterais désormais remercier mes parents pour leur soutien indéfectible tout au long de mon parcours à l'université, et notamment dans les années d'études parisiennes particulièrement difficiles. Merci à Juliette pour l'hébergement quand nous n'avions pas encore trouvé d'appartement à Paris ainsi que pour les escapades dej' fort appréciées durant la thèse. Merci à Fred de m'avoir épargné tant de temps de constitution de garde-robe. Un grand merci aussi à Annie Rozan pour son hospitalité et sa générosité sans faille durant toutes ces années, nos week-ends à Noyon étaient notre bouée de sauvetage quand nous étions étudiants à Paris.

Un grand merci aussi au programme doctoral IPV pour cette bourse doctorale, sans qui cette belle aventure n'aurait tout simplement pas pu exister.

Je souhaite également remercier Chris Bowler, Guillaume Achaz et Olaya Rendueles-Garcia pour leurs précieux conseils au cours de mes deux comités de thèse.

Je tiens par ailleurs à adresser mes remerciements aux membres de mon jury de thèse. Je suis très reconnaissant envers Hélène Touzet et Catherine Larose, mes deux rapportrices. Un grand merci d'avoir accepté de lire et d'évaluer mon manuscrit de thèse et merci pour vos retours. Un grand merci également à Jonathan Filée, Frédéric Devaux et Christophe Dessimoz qui ont accepté de participer à mon jury en tant qu'examineurs.

Je tiens également à remercier Joséphine, pour la plus belle des surprises que tu m'as réservée l'année dernière et pour ta personnalité si bienveillante, espiègle et radieuse, qui ont contribué à rendre cette dernière année de thèse tellement moins difficile qu'elle n'aurait pu l'être. Je suis si fier d'être ton papa.

Enfin à l'heure d'écrire les derniers mots de ce manuscrit de thèse, toutes mes pensées vont à toi Magali. Avec la naissance de notre petite Jojo, écrire cette thèse aurait tout simplement été impossible sans toi. Je te réserve donc mon dernier merci. Tu m'as fait grandir en tant que personne et je n'oublierai jamais tout ce que tu as fait pour moi. Merci pour ton amour et ton soutien.

Table des matières

| | |
|---|-----|
| 1. Introduction générale sur la communication microbienne par voie de Quorum Sensing (QS)... | 1 |
| 1.1. Bref historique de la découverte du QS, un mécanisme de biocommunication densité-dépendant..... | 1 |
| 1.2. Les signaux de QS n'informent pas seulement sur la densité de population | 8 |
| 1.3. Importance du QS en écologie, en sciences appliquées et en biologie évolutive..... | 9 |
| 1.4. Limites de la recherche sur le QS et objectifs de thèse | 11 |
| 2. Propriétés des gènes et signaux de QS à considérer pour détecter, prédire les fonctions et comparer des systèmes de QS en bioinformatique..... | 14 |
| 2.1. Base génétique du QS : des gènes émetteurs aux gènes récepteurs..... | 15 |
| 2.2. Diversité des signaux de QS : les « langues » et les « dialectes »..... | 21 |
| 2.3. Spécificité des signaux de QS : des « langues » plus ou moins secrètes..... | 29 |
| 2.4. Hiérarchie intragénomique des systèmes de QS : adaptabilité des micro-organismes « polyglottes »..... | 36 |
| 2.5. Lien entre similarité de séquences entre protéines de QS et probabilité d'émission ou de reconnaissance d'une molécule de QS en commun..... | 40 |
| 2.6. Combinatoire de gènes de QS entre deux génomes distincts et modalités de communication : de la communication bilatérale à la manipulation et l'espionnage..... | 46 |
| 2.7. Coévolution bactéries-eucaryotes et signalisation inter-domaines..... | 54 |
| 3. Détection de variants de gènes de QS dont les signaux associés sont de petites molécules de QS par des méthodes classiques de recherche d'homologues..... | 54 |
| • article 1 : Rich Repertoire of Quorum Sensing Protein Coding Sequences in CPR and DPANN Associated with Interspecies and Interkingdom Communication | |
| 4. Identification systématique de systèmes de QS dont les signaux sont des oligopeptides..... | 80 |
| 4.1. Détection de variants de gènes de QS par des méthodes classiques de recherche d'homologues..... | 80 |
| • article 2 : Beyond arbitrium: identification of a second communication system in Bacillus phage phi3T that may regulate host defense mechanisms | |
| 4.2. Détection de nouveaux gènes de QS par une nouvelle méthode basée sur une signature de critères génomiques, structurels et biochimiques..... | 100 |
| • article 3 : RRNPP_detector : a tool to detect RRNPP quorum sensing systems in chromosomes, plasmids and phages of gram-positive bacteria | |

| | |
|--|-----|
| 5. De la détection de nouveaux systèmes de QS à leur inférence fonctionnelle..... | 125 |
| • article 4 : Large-scale identification of viral quorum sensing systems reveals density-dependent sporulation-hijacking mechanisms in bacteriophages | |
| 6. Génomique comparative des systèmes de QS et inférence de réseaux de communication..... | 153 |
| • chapitre de livre : Microbial Communication Networks: Sketching a Method for Analyzing the Communication of Bacteriophages Inside Environmental Communities | |
| 7. Travaux en cours | 173 |
| 7.1. Détection de variants divergents de gènes de QS associés à de petites molécules de QS par des recherches itératives d'homologues et des critères de conservation de synténie..... | 173 |
| 7.2. Vers une inférence à large échelle des fonctions régulées par QSSs dont le récepteur est à une composante | 177 |
| 8. Conclusions et perspectives..... | 181 |
| 9. References..... | 192 |
| 10. Annexes..... | 204 |
| • article 5 : Interspecific interactions that affect ageing: age-distorters manipulate host ageing to their own evolutionary benefits | |

Liste des figures

- Figure 1 : Principe du QS (page 2)
- Figure 2 : Mécanisme du système de QS « arbitrium » (page 7)
- Figure 3 : Les quatre grands types de QSSs (page 18)
- Figure 4 : La famille des AHLs (langue) et ses variants structuraux (dialectes) (page 22)
- Figure 5 : La famille de propeptides Phr et ses différents peptides matures en C-terminal (page 22)
- Figure 6 : Diversité non exhaustive des signaux de QS cartographiés sur différents clades procaryotes (page 23)
- Figure 7 : Diversité structurelle non exhaustive des petites molécules de QS (page 25)
- Figure 8 : Diversité structurelle non exhaustive des peptides de QS reconnus par des systèmes à deux composantes (page 27)
- Figure 9 : Diversité non exhaustive des réponses au quorum régulées par des QSSs de type RRNPP (page 28)
- Figure 10 : Signaux inter-espèces (page 30)
- Figure 11 : Complexité structurelle des peptides de QS (page 31)
- Figure 12 : Triche facultative densité-dépendante due à l'acquisition d'un système Rap-Phr ultra-spécifique chez les Bacilles (page 32)
- Figure 13 : Arbre phylogénétique des synthases de la famille LuxS (famille 34)
- Figure 14 : LuxS, la synthase de l'AI-2 est une enzyme clé du méthyle activé (page 35)
- Figure 15 : Signaux intra-spécifique (CAI-1) et inter-spécifique (AI-2) « coincidents » chez *V. cholerae* (page 37)
- Figure 16 : Hiérarchie des QSSs intraspécifiques chez *P. aeruginosa* (page 39)
- Figure 17 : Heatmap des différents « dialectes » moléculaires d'AHLs synthétisés par différentes espèces de Protéobactéries (page 41)
- Figure 18 : Similarité de séquences entre protéines de QS et probabilité de production/reconnaissance d'une molécule de QS en commun (page 44)
- Figure 19 : Espionnage du signal de QS de l'hôte par Vibrio phage VP882 (page 48)
- Figure 20 : Amplification du signal de QS de *V. cholerae* par *R. obeum* (page 49)
- Figure 21 : Aperçu non exhaustif de la signalisation homme-bactéries (page 50)
- Figure 22 : Espionnage d'hormones de stress humaines par une souche enterohémorragique d'*Escherichia coli* (page 51)
- Figure 23 : Création d'un profil HMM à partir d'un alignement multiple de protéines (page 56)
- Figure 24 : Réseau de similarité de séquences de la famille LuxI (page 176)

- Figure 25 : Les différentes communautés de Leiden dans le réseau de similarité de séquences de LuxI putatifs (page 177)
- Figure 26 : Réseau de co-occurrences de différentes familles de gènes au sein d'opérons *luxI* et donc inférées comme étant régulées par QS (page 179)

Liste des tables

- Table 1 : Les différentes langues et petites molécules de QS et leurs différentes familles de gènes de QS associées (page 24)
- Table 2 : [Les différents types de peptides de QS et leurs différentes familles de gènes de QS associées](#) (page 26)
- Table 3 : Combinatoire de gènes de QS et modalités de communication (page 47)

Liste des abréviations les plus courantes

- AI-2 : Autoinducer-2
- AHL : Acyl Homoserine Lactone
- BGC : Biosynthetic Gene Cluster (Cluster de gènes de biosynthèse)
- HMM : Hidden Markov Model (modèle caché de Markov)
- MAG : Metagenomics-Assembled Genome (génomomes assemblés par métagénomique)
- MGE : Mobile Genetic Element (élément génétique mobile)
- QS : Quorum Sensing (détection du quorum)
- QSS : Quorum Sensing System (système génétique de détection du quorum)
- RRNPP : Groupe de récepteurs intracellulaires Rap, Rgg, NprR, PlcR et PrgX

Résumé

Les systèmes de « Quorum Sensing » (QSSs) sont des systèmes génétiques permettant la communication entre cellules ou entre bactériophages. Cette communication est réalisée par l'émission et la détection d'une molécule signal dont la concentration extracellulaire reflète la densité de la population codant le QSS. Les QSSs ont une importance capitale dans la régulation densité-dépendante de processus biologiques clés tels que la virulence, la sporulation ou la formation de biofilms chez les bactéries, la conjugaison chez les plasmides, ou la lysogénie chez les bactériophages tempérés. Néanmoins, la diversité des QSSs demeure largement sous-explorée et il en va de même pour celle des organismes, plasmides et virus codant ces systèmes. Ainsi, beaucoup de comportements microbiens/viraux régulés de façon densité-dépendante restent probablement à découvrir, dont certains pourraient peut-être révolutionner notre vision de l'adaptation microbienne et de la coévolution entre bactéries et leurs éléments génétiques mobiles. Précisément, cette thèse de bioinformatique évolutive explore les diversités phylogénétique et fonctionnelle du processus de détection du quorum par des méthodes d'analyses de génomes et de réseaux appliquées à des éléments génétiques traditionnellement délaissés par ce champ de recherche : génomes de lignées méconnues comme les CPRs et les DPANNs, métagénomes environnementaux, génomes viraux ou encore plasmides. En particulier, cette thèse pose les bases théoriques de l'inférence de réseaux de communication au sein de communautés microbiennes environnementales et inclut le développement d'une nouvelle méthode permettant l'identification de QSSs de type RRNPP (Rap-Rgg-NprR-PlcR-PrgX) non homologues à des QSSs déjà connus. Ce travail révèle notamment les premiers bactériophages bilingues, c'est-à-dire codant deux QSSs appartenant à des familles génétiques différentes ainsi que les premiers bactériophages prédits pour manipuler de façon densité-dépendante la biologie de leur hôte bactérien.

Abstract

Quorum sensing systems (QSSs) are genetic systems supporting cell-cell or bacteriophage-bacteriophage communication via the production and the detection of a signal molecule, the extracellular concentration of which reflects the density of the QSS-encoding population. QSSs have a prime importance in the regulation of key biological processes such as virulence, sporulation or biofilm formation in bacteria, conjugation in plasmids or lysogeny in temperate bacteriophages. However, the genetic diversity of QSSs remains largely underexplored and the same holds for the diversity of organisms, plasmids and viruses encoding these systems. Hence, many bacterial and viral density-dependent behaviors likely await to be discovered, some of which could perhaps transform our views of microbial adaptation and of the co-evolution between bacteria and their mobile genetic elements. Specifically, this PhD in evolutionary bioinformatics explores the phylogenetic and functional diversity of quorum sensing using genome and network analysis methods applied to genetic elements traditionally neglected by this research field : genomes of poorly known lineages such as CPRs and DPANNs, environmental metagenomes, viral genomes or plasmids. In particular, this thesis lays the theoretical foundations for the inference of communication networks within environmental microbial communities and includes the development of a new method allowing the identification of QSSs of the RRNPP type (Rap-Rgg-NprR-PlcR-PrgX) that are non-homologous to already known QSSs. This work notably reveals the first bilingual bacteriophages, i.e. encoding two QSSs belonging to different genetic families, as well as the first bacteriophages predicted to manipulate in a density-dependent manner the biology of their bacterial host.

1. Introduction générale sur la communication microbienne par voie de Quorum Sensing (QS)

Les micro-organismes ou microbes (bactéries, archées, eucaryotes unicellulaires, auxquels nous ajouterons par commodité les virus) communiquent et sont donc capables d'organiser et de coordonner leurs actions, au même titre que les organismes multicellulaires. Mais qu'entend-on par « communication microbienne » et en quoi ce thème de recherche est-il important, aussi bien d'un point de vue fondamental qu'appliqué ?

De façon générale, la communication microbienne peut se résumer à tout type d'interaction entre cellules/virus réalisée par l'intermédiaire d'un signal (Witzany, 2011). Typiquement, un signal de biocommunication transporte un contenu informationnel, dont l'intégration par une cellule/virus entraîne une réaction comportementale (Witzany, 2011). Toutefois, cette définition générique peut donner lieu à certaines ambiguïtés, selon ce que l'on considère être ou non un signal ou un changement comportemental. En pratique, quand les microbiologistes et les évolutionnistes parlent de « communication microbienne », ils se réfèrent alors généralement à un sous-ensemble bien défini du spectre des interactions entre micro-organismes réalisées par l'intermédiaire d'un signal. Ce sous-ensemble correspond à un processus microbien nommé quorum sensing (QS), dont les contours très caractéristiques ont le mérite de faire, à quelques nuances près évoquées en **section 1.2**, consensus parmi les biologistes. Dans ce processus, le signal est une molécule extracellulaire de communication, l'information véhiculée par ce signal concerne une densité de population, et les comportements initiés en réponse à ce signal correspondent à des transitions phénotypiques très claires. Pour son caractère explicite très pratique, nous utiliserons également le terme de QS comme synonyme de communication microbienne au cours de cette thèse.

1.1 Bref historique de la découverte du QS, un mécanisme de biocommunication densité-dépendant

L'expression « QS » fut introduite pour la première fois en 1994 par Fuqua, Winans et Greenberg (Fuqua, Winans and Greenberg, 1994) pour mettre un nom sur un phénomène biologique révélé dès le début années 1970 par les travaux du laboratoire de Hastings. Dans leur étude pionnière, Nealson, Platt et Hastings observent que la bioluminescence des bactéries marines *Aliivibrio fischeri* est régulée génétiquement : le gène codant la luciférase - l'enzyme catalysant la réaction de bioluminescence - n'est transcrit qu'à partir du milieu de la phase exponentielle de croissance, après quoi son niveau d'expression augmente intensément (Nealson, Platt and Hastings, 1970). En

investiguant la cause de cette régulation, Eberhard et Nealson identifient qu'une molécule excrétée par *A. fischeri* s'accumule dans le milieu de culture à mesure que la population bactérienne croît. Lorsque la concentration de cette molécule atteint un niveau critique, elle stimule la transcription de la luciférase (Eberhard, 1972; Nealson, 1977). En 1981, Eberhard *et al.* identifient cette molécule comme étant une acyl-homoserine lactone (AHL), une petite molécule diffusible à travers les membranes cellulaires (Eberhard *et al.*, 1981). Deux ans plus tard, Engebrecht *et al.* caractérisent les gènes d'*A. fischeri* responsable de la production et de la détection de cette molécule : *luxI* code l'enzyme synthétisant l'AHL et *luxR* code son récepteur intracellulaire, un facteur de transcription activant l'expression de la luciférase suite à sa liaison à l'AHL (Engebrecht, Nealson and Silverman, 1983) (nous reviendrons plus en détails sur les bases génétique du QS en **section 2.1**). Ensemble, ces travaux ont révélé que les bactéries de l'espèce *A. fischeri* produisent et excrètent un métabolite de communication dont la concentration extracellulaire reflète la densité de population. À partir d'un certain seuil de concentration, lorsque la densité cellulaire devient importante, ce métabolite commence à être reconnu et intégré de façon robuste par chaque cellule de la population, déclenchant alors au niveau transcriptionnel une réaction de bioluminescence collective et synchronisée (**Figure 1**).

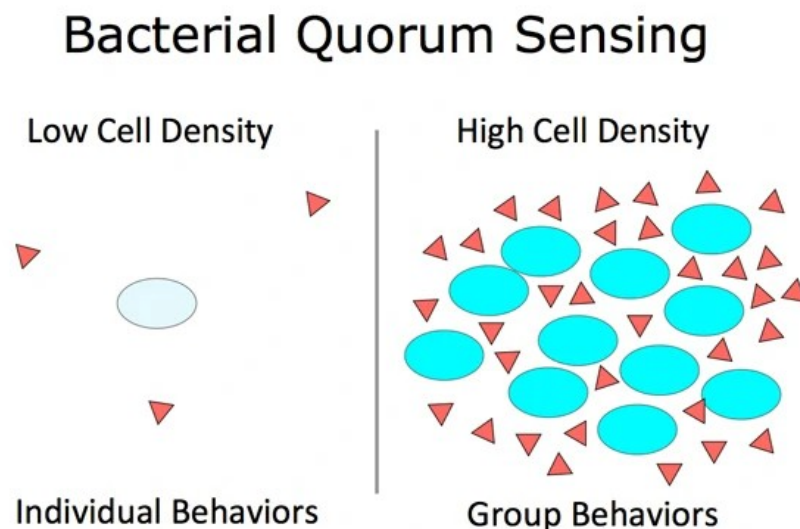


Figure 1: Principe du QS. Le triangle rouge correspond au signal de quorum sensing, les ellipses bleues aux cellules émettrices de ce signal.

Ces découvertes mécanistiques historiques laissaient toutefois une question essentielle en suspens. Quels avantages sélectifs cette régulation densité-dépendante de la bioluminescence offre-t-elle à *A. fischeri* ? Un début de réponse vint de l'observation que seule une lumière émise à haute

densité cellulaire peut être suffisamment intense pour avoir un impact visuel, et donc pour exercer un rôle écologique (Kempner and Hanson, 1968). Ainsi, la régulation densité-dépendante de la bioluminescence permettrait à *A. fischeri* de ne pas payer le coût de production de la luciférase quand cela s'avérerait écologiquement futile. Mais quel serait alors l'intérêt évolutif d'une bioluminescence de masse ? Cette réponse fut apportée par les études d'*A. fischeri* dans son milieu naturel. À l'état planctonique, les bactéries sont diluées dans l'océan (en moyenne 10^5 cellules par mL) et ne se retrouvent jamais en densité locale suffisante pour que la bioluminescence soit déclenchée. La seule circonstance où *A. fischeri* se retrouve suffisamment concentré pour émettre de la lumière (de l'ordre de 10^{10} cellules par mL) est lorsque cette bactérie colonise l'organe photophore de poissons et de seiches tels que le poney commun (*Leionathus equulus*) (Hastings, 1971) ou la sépiole (*Euprymna scolopes*) (Boettcher and Ruby, 1990). Suite à la découverte de l'association *A. fischeri* – *L. equulus*, Hastings identifie qu'à la nuit venue, la lumière émise par *A. fischeri* au niveau du ventre du poisson permet à ce dernier de camoufler son ombre projetée par la lune, et donc, de se dissimuler de ses prédateurs (Hastings, 1971). La nuit, les coûts investis par chaque partenaire dans la symbiose – partages des ressources pour *L. equulus* ; production intense de luciférase pour *A. fischeri* – sont alors largement compensés par les bénéfices de leur coopération : les chances de survie et de reproduction du poisson sont maximisées par la contre-illumination offerte par la bactérie et la bactérie bénéficie de la protection et du flux nutritif de son hôte pour se répliquer en masse. Toutefois, à la lumière du jour, la contre-illumination devient inefficace et les intérêts évolutifs du poisson ne sont plus alignés avec ceux de la bactérie. 95% des bactéries présentes dans l'organe photophore de l'hôte (dont beaucoup de cellules devenues sénescents au cours de la nuit) sont alors expulsées dans l'océan au petit matin (Millikan and Ruby, 2004). Ce grand ménage permet à l'hôte de limiter l'allocation de ressources à *A. fischeri* tout en assurant que les 5% de bactéries restantes pourront se répliquer au cours de la journée dans un organe photophore délesté d'un maximum de bactéries sénescents, gage d'une bioluminescence intense à la tombée de la nuit. Ces travaux d'écologie ont ainsi révélé que la bioluminescence de masse d'*A. fischeri* a sans doute créé une forme de dépendance chez son hôte dont la bactérie tire aujourd'hui profit pour se répliquer de façon plus efficace qu'à l'état planctonique (Visick *et al.*, 2000). L'avantage sélectif de la bioluminescence ne se manifestant toutefois pour l'hôte qu'en période nocturne, la quasi-totalité de la population d'*A. fischeri* est expulsée au matin, ce qui se traduit par une densité bactérienne moindre dans l'organe photophore de l'hôte en journée que pendant la nuit. Ainsi, les conditions de faibles densités cellulaires d'*A. fischeri* concordent avec les périodes où la production de bioluminescence ne serait ni efficace écologiquement, ni utile à la consolidation de la symbiose avec l'hôte. La communication d'*A. fischeri*, base mécanistique de la régulation densité-dépendante de la luciférase, active donc la bioluminescence uniquement lorsque cela s'avère évolutivement avantageux, optimisant ainsi la

valeur reproductive (fitness) de la bactérie au gré des fluctuations de son contexte environnemental et social.

Suite à la découverte du système génétique *luxI-luxR*, d'autres systèmes homologues de réponse transcriptionnelle à la densité de population ont été identifiés au début des années 1990. D'abord chez *Pseudomonas aeruginosa*, un pathogène humain où le système homologue à *luxI-luxR* (nommé *lasI-lasR*) régule de façon densité-dépendante la sécrétion de facteurs de virulence (Passador *et al.*, 1993). L'intérêt évolutif d'une telle régulation réside dans le fait que seule une production collective peut amener ces facteurs de virulence à la concentration requise pour affaiblir l'hôte de façon efficace, facilitant ainsi son invasion au profit de la réplication de la bactérie. Puis vint le tour d'*Erwinia carotova*, un pathogène de plantes chez qui le système de type *luxI-luxR* (*expi-expR*) n'induit l'expression d'enzymes dégradants le mur cellulaire de son hôte qu'à haute densité bactérienne (Pirhonen *et al.*, 1993). Enfin chez le pathogène de plantes *Agrobacterium tumefaciens*, un système homologue à *luxI-luxR* fut identifié (*traI-traR*), cette fois-ci, non pas sur un chromosome bactérien mais sur un plasmide (Fuqua, Winans and Greenberg, 1994). Ce plasmide contient des gènes, qui, lorsque transférés au noyau des cellules de l'hôte végétal, dirigent la synthèse et la sécrétion d'opines, alors consommées par la bactérie comme source de nutriments. Le système plasmidique *traI-traR* coordonne alors le transfert du plasmide entre bactéries de l'espèce *A. tumefaciens* uniquement à haute densité plasmidique, au sein de nodules cancéreux de l'hôte. Cette régulation de la conjugaison assure que les coûts d'expression et de réplication du plasmide ne soient payés par un maximum d'hôtes bactériens uniquement dans les conditions environnementales où le plasmide offre un maximum d'avantages à son hôte, concourant alors au succès évolutif de cet élément génétique mobile.

En faisant la revue de ces systèmes *luxI-luxR*, Fuqua, Winans et Greenberg nomment la régulation densité-dépendante d'un processus biologique par voie de communication le QS (Fuqua, Winans and Greenberg, 1994). Selon leur propres mots « certains comportements bactériens ne peuvent avoir une portée efficace que s'ils sont entrepris par une population de bactéries suffisamment grande. Nous appelons cette unité comportementale minimale un quorum de bactéries. Les systèmes génétiques *luxI-luxR* offrent aux bactéries un moyen efficace de recenser leur nombre et jouent un rôle régulateur en réponse à la densité ».

Le terme de quorum provient du vocabulaire juridique où il désigne le nombre minimal de personnes devant être présentes à une assemblée pour qu'une décision puisse être officiellement prise. Transposé au monde microbien, le QS désigne donc la détection par un microorganisme du nombre minimal d'individus émetteurs d'un signal requis pour qu'une action collective soit déclenchée. Ce terme véhicule bien l'idée que le dénombrement réalisé indirectement par les systèmes génétiques de réponse à la densité n'est pas continu, mais bien discret, et plus exactement binaire : le nombre d'individus émetteurs d'un signal est soit en dessous du quorum, soit au dessus.

De par son élégance et sa qualité descriptive, l'expression « QS » s'est répandue comme un feu de paille. La popularité de ce terme eut deux conséquences majeures. La première fut de réinterpréter toutes les études passées sur la communication microbienne à la lumière du QS. La deuxième fut de créer une véritable émulation pour la recherche de nouveaux systèmes de QS (QSSs pour Quorum Sensing Systems) et signaux de QS.

La relecture des études passées à travers le prisme du QS a permis d'étendre la diversité décrite des microorganismes utilisant le QS, initialement circonscrite, avec les systèmes *luxI-luxR*, au seul phylum des Protéobactéries à d'autres phyla bactériens : d'abord au phylum des Actinobactéries, au sein duquel il fut décrit que les espèces du genre *Streptomyces* utilisent les gamma-butyrolactones (GBLs) comme signal moléculaire pour coordonner la production collective d'antibiotiques (Horinouchi and Beppu, 1992), puis au phylum des Firmicutes, chez qui les petits oligopeptides sécrétés par les individus d'une population étaient connus pour déclencher, à haute concentration, la sporulation et/ou la compétence dans le genre *Bacillus* (Rudner *et al.*, 1991), ou pour réguler la conjugaison dans le genre *Enterococcus* (Clewell and Weaver, 1989).

Les nouvelles recherches ont, elles, permis de bousculer les paradigmes jusqu'ici associés au QS, à commencer par la notion d'intra-spécificité des signaux de QS. Jusqu'en 1994 et la définition du terme de QS, les signaux moléculaires découverts étaient en effet spécifiques d'une espèce ou d'une souche donnée et ne permettaient pas d'induire une réponse chez d'autres espèces. Par exemple, l'AHL produite par *A. fischeri*, la *N*-(3-oxohexanoyl)-*L*-homoserine-lactone, est différente de l'AHL produite par *P. aeruginosa*, la *N*-(3-oxodecanoyl)-*L*-homoserine-lactone, et ces deux signaux ne sont pas inter-fonctionnels (Fuqua, Winans and Greenberg, 1994). Il faudra attendre la fin des années 1990, avec la découverte par Bassler du signal inter-spécifique nommé Autoinducer-2 (AI-2) pour révolutionner ce cadre théorique (Bassler, Greenberg and Stevens, 1997; Surette, Miller and Bassler, 1999). Bassler et ses collègues identifient en effet que les milieux de culture de *Vibrio parahaemolyticus*, *Escherichia coli*, et *Salmonella typhimurium* entraînent le même changement transcriptionnel chez *Vibrio harveyi* que celui observé à haute densité dans son propre milieu de culture. Ce phénomène est attribuable à l'AI-2, une molécule de QS reconnue chez *V. harveyi* par le récepteur membranaire LuxP et synthétisée par toutes les espèces susnommées *via* la famille de gènes *luxS*. Suite à cette découverte, des études de génomique comparative ont révélé que la famille *luxS* est non seulement conservée parmi plusieurs phyla bactériens, parmi lesquels les Protéobactéries, les Firmicutes, les Actinobactéries, les Fusobactéries, les Spirochètes et les Bacteroidetes mais que l'AI-2 synthétisé par des représentants de chacun de ces phyla réussit à stimuler LuxP chez *V. harveyi* (Sun *et al.*, 2004; Martha Antunes *et al.*, 2005; Jang *et al.*, 2013). De façon intéressante, la large couverture phylogénétique de *luxS* n'est pas congruente avec la distribution de *luxP*, qui, elle, est spécifique à l'ordre des *Vibrionales*, suggérant une variété importante de récepteurs à l'AI-2 (suggestion plus tard confortée par les découvertes d'autres

récepteurs à l'AI-2 (Pereira *et al.*, 2009; Anderson *et al.*, 2015; Torcato *et al.*, 2019; Zhang *et al.*, 2020)). La découverte de l'AI-2 a donc révélé qu'il existe des signaux de QS plus ou moins secrets, et donc des réponses microbiennes à des densités de populations plus ou moins hétérogènes. Cette spécificité relative des QSSs est un phénomène fascinant sur lequel nous reviendrons en **sections 2.3 et 2.4.**

Par ailleurs, de nouvelles recherches ont permis d'étendre la diversité des micro-organismes utilisant le QS, jusqu'ici limitée au seul domaine des bactéries, à tous les autres domaines du vivant. Chez les eucaryotes, Hornby *et al.* rapportent en 2001 que les champignons *Candida albicans* secrètent du farnésol comme signal de QS. En réponse à ce signal, les gènes contribuant à la formation de structures filamenteuses multicellulaires sont réprimés, favorisant ainsi la forme unicellulaire de ce pathogène humain (Hornby *et al.*, 2001; Oh *et al.*, 2001). Puis en 2018, Tian *et al.* rapportent que la reproduction sexuelle du champignon *Cryptococcus neoformans* est contrôlée par un peptide de QS (Tian *et al.*, 2018). Chez les archées, Zhang *et al.* rapportent en 2012 que *Methanosaeta harundinacea* secrète une AHL carboxylée dont la détection à haute concentration déclenche la transition d'un mode unicellulaire à une organisation filamenteuse multicellulaire (Zhang *et al.*, 2012). Enfin, en 2017, Erez *et al.* font la première validation expérimentale d'un QSS porté par un virus (Erez *et al.*, 2017). Lorsqu'il est exprimé par un hôte bactérien, ce système codé par Bacillus phage phi3T dirige la production d'un petit peptide de QS nommé « arbitrium » (**Figure 2**). La concentration extracellulaire du peptide reflète alors le nombre de fois que les hôtes bactériens ont exprimé le génome du phage depuis le début d'une infection. En dessous du quorum, le cycle lytique est favorisé : les phages phi3T se répliquent massivement au détriment de leur hôte respectif. Au dessus du quorum, la poursuite du cycle lytique à l'échelle de la population virale menacerait la survie de la population hôte, et donc du phage lui-même. En réponse au quorum, le système arbitrium coordonne alors une transition synchronisée du cycle lytique au cycle lysogénique, optimisant de fait la fitness du phage. Suite à cette transition, les phages de la population insèrent leur génome dans celui de leur hôte - devenant prophages - et confèrent à l'hôte une immunité vis à vis des virions. Le collectif phage-hôte est alors protégé au compromis d'une répllication plus lente du prophage, en tant que partie du génome bactérien. Ensemble, ces études fascinantes semblent indiquer que le QS est un processus ubiquitaire chez les microorganismes et invitent à découvrir les systèmes génétiques à la base de ce processus dans les nombreux phyla microbiens encore non identifiés comme utilisateurs de QS.

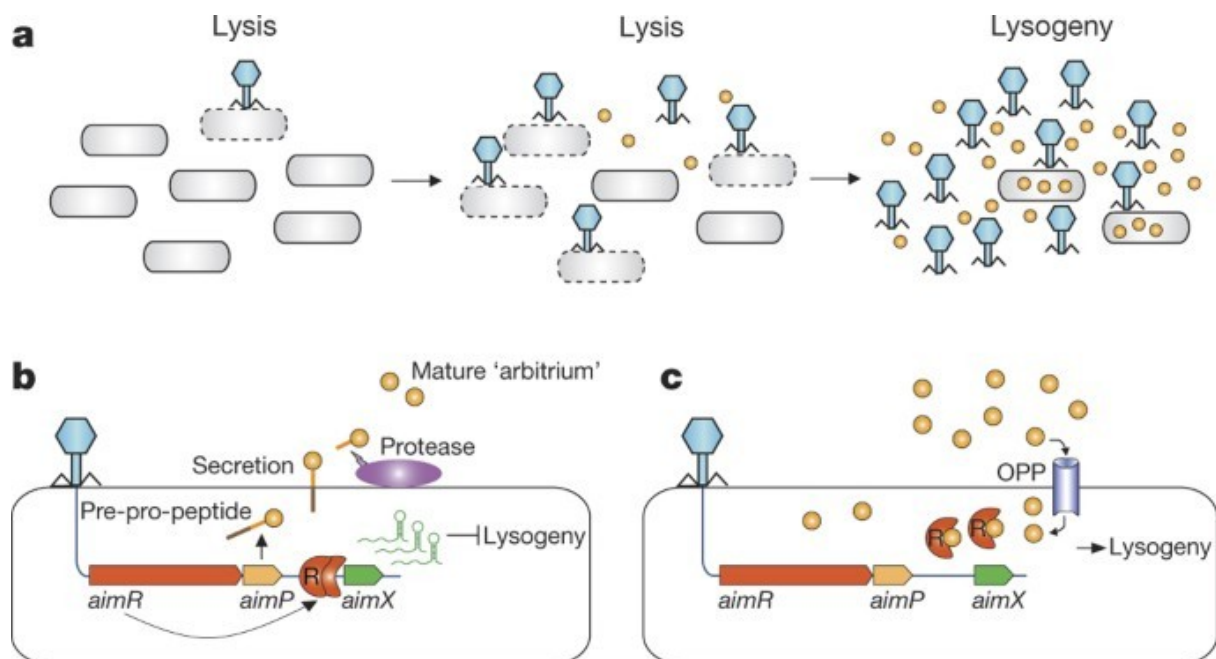


Figure 2 : Mécanisme du système de QS « arbitrium » régulant la transition lyse-lysogénie chez certains phages tempérés de Bacilles [Source : (Erez *et al.*, 2017)]

En plus d'avoir dévoilé une diversité importante d'organismes (ainsi que quelques plasmides et virus) utilisant le QS, le corpus d'études portant sur le QS a révélé, depuis les années 1970, une myriade de signaux moléculaires et de gènes de QS différents, sur lesquels nous reviendrons en détail dans la **section 2.2**. D'autre part, ces études ont permis d'identifier une grande diversité de processus microbiens régulés de façon densité-dépendante.

Parmi ces processus, la transition entre deux phases d'un cycle de vie est typique. La transition lyse-lysogénie de Bacillus phage phi3T en est un bon exemple, auquel nous pourrions rajouter, chez la bactérie pathogène *Legionella pneumophila*, la transition d'une phase parasitaire au sein d'une cellule hôte à une phase libre dite « transmissible » permettant d'infecter d'autres cellules hôtes (Hochstrasser and Hilbi, 2017). Cependant, l'exemple le plus récurrent concerne la transition d'un mode de vie unicellulaire à une organisation multicellulaire, ou vice versa. La régulation densité-dépendante de la formation ou de la dissolution de biofilms en offre une parfaite illustration chez les bactéries (Miller and Bassler, 2001; Federle and Bassler, 2003). Par ailleurs, en impliquant une sécrétion collective d'éléments constitutifs d'une matrice extracellulaire, le cas de la formation de biofilms fait le lien vers un autre processus régulé fréquemment par le QS : la production de molécules de biens communs. En effet, antibiotiques, facteurs de virulence, exozymes et autres biosurfactants sont typiquement régulés par QS (Heilmann, Krishna and Kerr, 2015). Toutefois, la production de biens communs déclenchée par QS ne constitue pas une règle absolue. En effet, la densité de population peut certes représenter une opportunité pour un comportement de groupe, mais elle peut également s'accompagner de stress environnementaux importants, dont la raréfaction en ressources

disponibles. Ainsi, le QS peut contrôler des processus d'acquisition de nutriments, et donc, de biens privés, et même coordonner des stratégies facultative de triche chez certaines sous-populations en inhibant temporairement la production de biens communs de façon densité-dépendante (Pollak *et al.*, 2016; Kalamara *et al.*, 2018). Dans cette même veine, de nombreux comportements associés à la survie sont déclenchés par le QS. La sporulation, le « cannibalisme » ou la fuite par sur-régulation de protéines liées à la motilité en sont autant de manifestations (Bischofs *et al.*, 2009; González-Pastor, 2011). Néanmoins, certains comportements altruistes peuvent aussi être déclenchés en réponse à un quorum intrinsèquement lié au stress. Le suicide cellulaire via l'expression d'une autolysine en réponse au quorum en constitue un exemple édifiant chez *Escherichia coli* et *Streptococcus mutans* (Dufour and Lévesque, 2013; Kumar and Engelberg-Kulka, 2014). Cette liste des processus régulés par QS ne saurait être exhaustive, mais ajoutons tout de même aux exemples précédents les cas bien décrits de la bioluminescence, du transfert d'ADN (par conjugaison ou compétence) ou du changement morphologique cellulaire. Tous ces exemples démontrent bien que le QS est non seulement ubiquitaire chez les microorganismes mais également central dans la régulation de processus biologiques d'importance fondamentale pour leur adaptation.

1.2. Les signaux de QS n'informent pas seulement sur la densité de population

Bien que les QSSs soient communément admis comme des systèmes de réponse à la densité de population, il convient d'indiquer que certains biologistes les considèrent avant tout comme des moyens de sonder les propriétés physico-chimiques d'un environnement. Par exemple, Redfield avance que ces systèmes sont moins des systèmes de QS mesurant et régulant des propriétés de groupe que des systèmes de diffusion sensing, permettant à une cellule de mesurer la vitesse à laquelle un signal émis s'éloigne d'elle et de décider en conséquence si la production d'exofacteurs aura un effet écologique optimal ou non (Redfield, 2002). Il est à cet égard intéressant de noter que si deux populations sont identiques en nombre par unité de volume, mais que l'une est confinée dans un milieu clos, tandis que l'autre est dans un milieu ouvert et très fluide, les deux populations ne réagiront pas au même quorum. En effet, dans le premier milieu, la probabilité qu'un signal se lie à un récepteur sera supérieure à celle du deuxième milieu, et le quorum reflété y sera donc plus petit. Néanmoins, selon West *et al.*, le QS et le diffusion sensing ne sont pas deux théories mutuellement exclusives dès lors que l'on considère que les QSSs régulent des comportements sociaux en réponse à de multiples facteurs environnementaux, dont des facteurs de densité et de diffusion (West *et al.*, 2012). D'ailleurs, il est aujourd'hui assez bien établi qu'un signal de QS véhicule de multiples informations sur un environnement (Cornforth *et al.*, 2014; Moreno-Gámez *et al.*, 2017; Chu, Groisman and Levchenko,

2019). Par exemple, la température et le pH du milieu peuvent influencer sur la stabilité du signal de QS (Horswill *et al.*, 2007) ou sur la conformation 3D de son récepteur, et donc sur son affinité vis à vis du signal (Do *et al.*, 2019). Des facteurs abiotiques peuvent donc influencer sur la chronologie de la transduction synchronisée d'un signal de QS. Une vision unificatrice de ces théories consisterait à dire que les QSSs ont évolué pour déclencher un comportement, social ou non, dans les conditions environnementales ayant prouvé par le passé offrir davantage de succès à ce comportement qu'à d'autres traits, conditions dont la densité de population est un facteur prépondérant.

1.3. Importance du QS

En écologie

Les communautés microbiennes réalisent des processus fondamentaux sur Terre. Les communautés microbiennes dans les sols et les mers participent aux cycles géochimiques et celles de nos microbiotes influencent notre santé et notre bien-être (Bertrand *et al.*, 2015; Gilbert *et al.*, 2018). Les processus propres à ces communautés émergent des interactions entre leurs différentes espèces constitutives. En tissant bon nombre de ces interactions et en coordonnant leurs réponses associées, le QS contribue fortement aux dynamiques sociales des communautés microbiennes. L'étude du QS peut alors se révéler utile à la compréhension de l'évolution et de l'émergence des processus collectifs propres aux communautés microbiennes qui façonnent notre planète.

En médecine

Le QS est aussi très intéressant selon des considérations plus anthropocentristes, notamment puisque l'humain est un holobionte, constitué d'au moins autant de cellules somatiques humaines que de cellules microbiennes (ratio estimé à 1,3 cellule bactérienne par cellule somatique) (Sender, Fuchs and Milo, 2016) sans compter les virus qu'ils lui sont associés et qui dépasseraient ce nombre de cellules microbiennes d'au moins un ordre de magnitude (Reyes *et al.*, 2010). En régulant certains comportements des microbes qui nous composent, nous influencent, et nous assistent, le QS apparaît comme une composante importante de notre biologie, bien que ce processus ne soit pas inscrit dans notre code génétique (Wu and Luo, 2021). Cette composante est d'ailleurs si importante qu'elle n'œuvre pas tout à fait en parallèle des réseaux de régulation et de signalisation strictement codés par notre propre génome. En effet, la coévolution homme-microbes a notamment donné lieu à un phénomène fascinant : l'inter-influence par voie de signaux moléculaires. Concrètement, certains signaux de QS produits par nos microbiomes interagissent directement avec certaines protéines humaines et *a contrario*, certaines de nos hormones peuvent, en tant qu'ersatz de signaux de QS, être interprétées par des récepteurs microbiens (Ismail, Valastyan and Bassler, 2016; Lee *et al.*, 2018;

Kim *et al.*, 2020; Wu and Luo, 2021) (plus de détails en **section 2.7**). Par ailleurs, en plus de pouvoir apporter un éclairage sur certains pans de notre propre biologie émergeant de l'interaction avec nos microbiotes, l'étude du QS dans les écosystèmes formés par l'humain et ses résidents microbiens pourrait mener à des applications médicales importantes. À titre d'exemple, des effets bénéfiques pour la santé ont été imputés à certaines molécules de QS dans un contexte d'immunomodulation (Landman *et al.*, 2018) ou d'équilibre du microbiote intestinal (Thompson *et al.*, 2015). *A contrario*, beaucoup de pathogènes humains deviennent virulents en réponse au QS. Ainsi, la recherche de signaux de QS enrichis spécifiquement dans des contextes infectieux pourrait aboutir à des stratégies thérapeutiques visant à court-circuiter la voie de communication par laquelle certains pathogènes résistants aux antibiotiques deviennent virulents (Jiang *et al.*, 2019). Parmi ces stratégies anti-QS, l'administration d'enzymes dites de Quorum Quenching, connues pour dégrader certains signaux de QS constituerait une option radicale mais peu prudente et onéreuse, tandis que l'administration d'analogues de signaux de QS, créant une compétition inhibitrice pour leurs récepteurs dédiés, semble offrir une option plus réaliste ayant déjà fait l'objet d'une preuve de concept chez la souris (Adiliaghdam *et al.*, 2019). Outre le QS intra-humain, le QS environnemental s'avère être aussi extrêmement intéressant d'un point de vue anthropocentrique car il régule des processus microbiens directement liés à des problématiques sociétales majeures. Par exemple, les biofilms, spores et autres voies de virulence régulés par le QS chez les pathogènes d'animaux et de plantes sont autant de facteurs à risque pour la sécurité alimentaire et les rendements agricoles. D'autre part, la conjugaison et la compétence régulées par le QS contribuent significativement à la diffusion des gènes de résistance aux antibiotiques (Von Wintersdorff *et al.*, 2016). Enfin, puisque le QS initie souvent la production collective de molécules antimicrobiennes, une détection systématique des QSSs et des gènes qu'ils régulent pourrait mener, selon une perspective plus optimiste, à l'enrichissement de notre pharmacopée (Hoover *et al.*, 2015; Brotherton, Medema and Greenberg, 2018; Rued *et al.*, 2021).

En biologie évolutive

Le quorum sensing offre enfin des questions stimulantes sur le plan de la biologie évolutive. Par exemple, comment expliquer la stabilité évolutive de systèmes complets de quorum sensing alors que des tricheurs capables d'intégrer le signal de communication sans le produire peuvent émerger d'une population ? Si le cadre théorique de la sélection de parentèle peut expliquer la pérennité de QSSs intra-spécifiques, comment expliquer la stabilité évolutive de QSSs inter-spécifiques (Diggle, Gardner, *et al.*, 2007) ? Si un signal inter-spécifique reflète la densité d'une communauté bactérienne hétérogène, comment expliquer la sélection du gène producteur de ce signal sachant que son récepteur seul permettrait à moindre coût de s'adapter aux dynamiques de population des autres espèces de la communauté ? Se pourrait-il que les réponses d'un collectif d'espèces à une haute

concentration d'un signal interspécifique puisse bénéficier à un microorganisme d'une espèce donnée au point d'exercer une pression de sélection pour sa contribution à l'accumulation de cette molécule ? Se pourrait-il, d'autre part, qu'en ne produisant pas un signal inter-spécifique, un tricheur soit désavantageusement exclu d'un collectif hétérogène mutualiste ? De façon intéressante, le quorum reflété par une haute concentration d'un signal inter-spécifique peut très bien correspondre à des représentants d'une seule espèce émettrice de ce signal tout comme à ceux d'une multitude d'espèces émettrices. Existerait-il alors des environnements dans lesquels un collectif d'espèces différentes pourrait se substituer à une fraction de la population d'une seule espèce pour le même succès évolutif du trait déclenché en réponse au quorum ? La sortie des biofilms multi-espèces ou mono-espèce à haute densité d'individus apparentés ou non semblerait en constituer un exemple (Bridges and Bassler, 2019). En outre, certaines cassettes génomiques correspondant à un QSS associé à ses gènes cibles sont mobiles et pourraient également abonder dans ce sens. Selon une perspective encore plus excitante, la co-évolution entre espèces différentes aurait-elle pu donner lieu à des réponses différentes et complémentaires à un quorum hétérogène, de sorte qu'un QSS interspécifique ne soit évolutivement stable que dans un contexte social diversifié ? Autrement dit, certains signaux de quorum sensing pourraient-ils, dans certains environnements, orchestrer une division des tâches parmi un consortium d'espèces différentes ? À ces égards, les études du laboratoire de Xavier suggérant un rôle de l'AI-2 dans l'équilibre d'un microbiote intestinal sain sont fascinantes (Thompson *et al.*, 2015). Dans un autre registre, comment expliquer la sélection d'une autolyse déclenchée par une haute densité de population ? Ou comment penser l'évolution des QSSs et des systèmes de signalisation eucaryotes vers des modalités d'inter-influence dans le cadre des co-évolutions bactéries-plante ou bactéries-animal ? Ces questions fascinantes illustrent bien que le QS constitue un cadre théorique excitant pour les évolutionnistes. L'étude approfondie de sa diversité et de ses implications écologiques dans des contextes multi-espèces apparaît alors essentielle pour penser cette évolution sur des bases concrètes solides.

1.4. Limites de la recherche sur le QS et objectifs de thèse

Ce petit tour d'horizon du QS, qui sera approfondi plus en détails dans le chapitre suivant, permet d'ores et déjà de saisir l'importance capitale du QS dans l'adaptation des micro-organismes à leur environnement social et physico-chimique, dans les interactions entre micro-organismes parents ou non, et dans la régulation de processus microbiens d'intérêt médical tels que la virulence ou la production de molécules naturelles aux propriétés antimicrobiennes. Cependant, un certain nombre de limitations, à la fois d'ordres technique et culturel sont inhérentes à la recherche sur le QS depuis les années 70 :

- la majorité des connaissances sur le QS est issue de l'étude en culture d'une poignée seulement de micro-organismes modèles (e.g. *Bacillus subtilis*, *Staphylococcus aureus*, *Streptomyces griseus*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, *Escherichia coli* ...) représentatifs de seulement 3 phyla bactériens (Firmicutes, Protéobactéries, Actinobactéries).
- la recherche du QS est particulièrement biaisée vers la biologie des pathogènes humains.
- bien qu'une transition soit en train de s'opérer depuis le milieu des années 2010 vers une étude du QS dans des contextes multi-espèces, la majorité des comportements microbiens régulés par le QS furent observés en mono-culture.
- les implications systémiques du QS dans les communautés microbiennes demeurent ainsi assez méconnues, et sont davantage abordées d'un point de vue théorique que sous le prisme de réelles données biologiques.
- enfin, l'écrasante majorité des QSSs étudiés concernent des QSSs chromosomiques. Les QSSs portés par des MGEs et notamment par des phages représentent un nouveau champ de recherche largement sous-exploré.

Face à ces limitations, la bioinformatique pourrait tirer son épingle du jeu en propageant les connaissances accumulées sur les génomes historiquement étudiés dans le cadre du QS à d'autres séquences d'ADN sous-explorées, aujourd'hui disponibles grâce à l'explosion récente de la quantité de génomes et de MGEs séquencés et grâce au développement de la métagénomique, une technique de séquençage environnemental permettant d'accéder aux génomes de micro-organismes cohabitant au sein d'une même communauté. Un premier pas vers l'enrichissement des connaissances actuelles sur le QS pourrait alors consister à développer des méthodes bioinformatiques de détection et d'inférence fonctionnelle de gènes de QS, puis de les appliquer aux génomes de lignées sous-étudiées, aux génomes environnementaux assemblés par métagénomiques (MAGs), et aux MGEs. Comme le QS est avant tout un processus d'interaction entre entités microbiennes, un deuxième pas pourrait concerner la comparaison des gènes de QS préalablement détectés entre génomes. En effet, une idée intuitive, dont la trivialité sera néanmoins éprouvée au cours du chapitre suivant, pourrait consister à dire que si deux génomes codent des QSSs dont les composantes sont très similaires en terme de séquence, il pourrait être envisageable que ces deux génomes puissent émettre et reconnaître au moins une molécule de QS en commun, justifiant alors la prédiction d'un lien de communication entre ces deux génomes. En somme, identifier qui communique comme qui pourrait permettre d'inférer qui communique avec qui, malgré de nombreuses limitations à surmonter dans ces inférences. Appliquées aux MAGs d'un même environnement, ces prédictions de liens de communication pourraient alors déverrouiller un troisième pas d'ordre écologique en permettant la reconstruction d'un réseau de communication hypothétique propre à chaque environnement, dans lequel chaque nœud (MAG) serait annoté par des labels taxonomique et d'abondance. L'analyse des topologies globale et locale de chacun des réseaux environnementaux ainsi générés, suivie de la

comparaison de ces réseaux entre eux pourrait alors permettre de gagner en compréhension sur les possibles rôles écologiques du QS.

Ce plan initial, motivant cette thèse de bioinformatique évolutive intitulée « Réseaux de communication au sein de communautés microbiennes environnementales » délimite quelques objectifs intermédiaires :

- peut-on détecter des variants (y compris divergents) de gènes connus de QS dans des lignées sous-explorées, des MGEs et des MAGs environnementaux ? (**sections 3 et 4.1** de cette thèse)
- peut-on identifier de nouvelles molécules de communication synthétisées et reconnues par ces variants ? (**section 7.1**)
- peut-on, par des méthodes de génomique comparative et d'analyse de réseaux, appréhender l'importance écologique du QS ? (**section 6**)

Toutefois, comme dans toute exploration, un plan initial est souvent sujet à modifications, au gré des impasses non anticipées et des opportunités nouvelles révélées en cours de chemin. Cette thèse n'y fait pas exception. La veille bibliographique aura notamment révélé que l'inférence de liens de communication entre génomes distincts s'avère être un problème beaucoup moins trivial qu'initialement anticipé. Selon des perspectives plus enthousiastes, cette veille bibliographique aura aussi livré son lot d'agréables surprises en révélant quelques propriétés clés de certains QSSs permettant d'appréhender par la bioinformatique deux nouvelles questions qui n'étaient pas envisagées initialement :

- peut-on identifier de nouveaux QSSs, sans lien d'homologie avec les QSSs déjà connus ? (**section 4.2**)
- peut-on inférer de nouvelles fonctions biologiques régulées de façon densité-dépendante par les QSSs préalablement détectés ? (**sections 4.2, 5 et 7.2**)

Les 5 objectifs intermédiaires définis ci-dessus participent tous de la même volonté d'explorer les diversités phylogénétique et fonctionnelle du processus de QS par des méthodes d'analyses de génomes et de réseaux, volonté fondée sur le pari que la détection de QSSs au sein d'éléments génétiques traditionnellement délaissés par ce champ de recherche pourrait mener à des découvertes à même de transformer notre vision des interactions, de l'adaptation et de l'évolution des microbes et des virus et à même de déboucher sur de possibles applications médicales ou agronomiques.

2. Propriétés des gènes et signaux de QS à considérer pour détecter, prédire les fonctions et comparer des systèmes de QS en bioinformatique

Le développement de tout outil bioinformatique de détection et/ou de comparaison de gènes ou de système de gènes d'intérêt nécessite une étude approfondie des propriétés connues de ces gènes, afin de distinguer un ensemble de caractéristiques qui les démarquent des autres et informent sur leurs fonctions. La détection de systèmes plus ou moins complets de gènes de QS n'échappe pas à cette règle ; ce chapitre aura alors pour vocation de faire un inventaire des différentes propriétés connues des QSSs dans le but d'identifier certaines caractéristiques qui pourraient être exploitées pour la détection *in silico* de gènes de QS, pour la prédiction des fonctions qu'ils régulent, et pour leur comparaison à travers les génomes à des fins d'inférence de liens de communication. Par exemple, une description des propriétés des quatre grands types de QSSs (**section 2.1**), des familles de gènes qu'ils regroupent et de leurs signaux associés (**section 2.2**) apparaît comme un prérequis indispensable au développement d'outils bioinformatiques de détection de gènes de QS, notamment par des méthodes de recherche basées sur l'homologie de séquences tels que l'aligneur BLAST (Altschul *et al.*, 1990) ou le détecteur de profils HMM (Hidden Markov Model) *hmmsearch* (Eddy, 2011a). De nombreuses intuitions sur les caractéristiques propres à chaque signal de QS et sur la combinaison de QSSs différents au sein d'un même génome seront également apportés en **sections 2.2, 2.3 et 2.4** afin de guider au mieux l'interprétation des résultats. Par ailleurs, pour comparer des gènes de QS entre eux, il est important de présenter les notions de lien entre similarité de séquences et probabilité de production/reconnaissance d'une molécule de QS en commun et d'analogie structurelle de signaux moléculaires distincts (**sections 2.5**). Dans cette même veine, les différentes modalités de communication émergeant de la combinatoire des modules génétiques de QSS entre deux génomes distincts (communication bilatérale, amplification du signal d'autrui, espionnage etc.) seront traitées en **section 2.6** car elles pourraient servir à l'orientation des arêtes d'un réseau hypothétique de communication entre entités microbiennes. Enfin, les notions de signalisation inter-domaines, en particulier bactéries-eucaryotes seront évoquées pour apporter une dimension supplémentaire à l'interprétation de la sélection de gènes de QS à travers les génomes microbiens (**section 2.7**).

2.1. Base génétique du QS : des gènes émetteurs aux gènes récepteurs

Principe général

Les systèmes génétiques de QS (QSSs pour Quorum Sensing Systems) sont composés de deux modules : l'un dit « émetteur », codant une ou plusieurs protéines impliquées dans la production du signal de QS, et l'autre dit « récepteur », codant une ou plusieurs protéines impliquées dans la transduction du signal. Nous reviendrons sur la composition précise de ces modules un peu plus bas.

Les QSSs peuvent être portés par des chromosomes, ainsi que par des éléments génétiques mobiles (MGEs) comme des plasmides, ou des génomes viraux pouvant s'intégrer dans le génome de leur hôte (*e.g.* bactériophages tempérés). Les QSSs codés par des phages doivent alors être exprimés et traités par une cellule hôte pour exercer leurs effets. Ce dernier cas de figure illustre que la concentration d'un signal de QS reflète le nombre d'éléments génétiques codant le QSS ayant été exprimé par une cellule hôte. Ce nombre correspond à :

- un quorum de cellules dans le cas d'un QSS chromosomique ;
- un quorum de plasmides intracellulaires dans le cas d'un QSS plasmidique ;
- un quorum de génomes de phages ayant été injectés par les virions dans leur cellule hôte (reflétant ainsi un nombre de cellules lysées) dans le cas d'un QSS de phage ;
- Un quorum de génomes de phages ayant été insérés dans un génome bactérien (reflétant ainsi un nombre de cellules lysogénisées) dans le cas d'un QSS de prophage.

Lorsque le signal atteint une concentration importante, reflétant un des quorums mentionnés précédemment, sa fréquence de liaison aux récepteurs membranaires ou intracellulaires des individus de la population codant le QSS (ou codant le récepteur seul dans les cas d'espions ou de tricheurs) devient importante. Suite à sa liaison physique avec le signal, le récepteur change de conformation, et va alors réguler (ou cesser de réguler), avec plus ou moins d'intermédiaires protéiques, la transcription de gènes cibles ou l'activité de protéines cibles. Ces régulations en réponse au quorum sont à la base de transitions comportementales observées à l'échelle de la population.

Les gènes émetteurs : enzymes de synthèse et propeptides

Dans le monde microbien, on distingue deux grands types de signaux de QS.

Le premier correspond à de petites molécules, souvent dérivées du métabolisme des acides gras, et produites par une enzyme ou un cluster d'enzymes de synthèse (appelée(s) synthase(s)) à partir d'un ou plusieurs substrat(s)/précurseur(s) moléculaire(s) (Papenfort and Bassler, 2016; Tobias *et al.*, 2019).

Le second correspond à des peptides à la structure chimique plus complexe, maturés post-traductionnellement à partir d'un propeptide codé génétiquement chez les bactéries à gram positif

(Sturme *et al.*, 2002; Neiditch *et al.*, 2017; Bhatt, 2019). À la rare exception des peptides qualifiés de « leaderless » (Do *et al.*, 2017), les propeptides présentent une séquence dite « signal » pour l'adressage à un complexe protéique de sécrétion, séquence qui sera clivée lors de la translocation du peptide vers le milieu extracellulaire. Les maturations post-traductionnelles des propeptides peuvent avoir lieu dans le cytosol ou à la surface externe de la membrane cellulaire et peuvent aller du simple clivage protéique à des phénomènes de cyclisation ou de modification d'acides aminés, conduisant ainsi soit à des peptides linéaires composés d'acides aminés naturels, soit à des peptides aux structures chimiques complexes. D'un point de vue bioinformatique, l'identification d'homologues de propeptides de QS connus peut s'avérer difficile car ces protéines évoluent rapidement (voir **section 4**). Néanmoins, l'identification de propeptides est moins sujette aux faux positifs que les synthèses de QS car les propeptides codent directement une partie du signal de QS, propriété qui sera d'ailleurs exploitée au cours de cette thèse. Il existe toutefois une exception à cette propriété avec le cas des diketopiperazines, des peptides de deux amino-acides cycliques et synthétisés par des enzymes, des CDPS (cyclodopeptide synthases) (Gondry *et al.*, 2018) ou NRPS (Non-Ribosomal Peptides Synthetase) (González *et al.*, 2017).

De façon générale, comme nous le verrons en **section 2.3**, plus un signal de QS a une structure chimique complexe, plus il est probable qu'il soit spécifique d'une espèce voire d'une souche donnée. Ainsi, les peptides de QS réalisent généralement une communication secrète intra-souche ou intra-espèce tandis que les petites molécules de QS, plus simples structurellement, sous-tendent des modalités de communication plus publiques, allant de l'intra-spécifique à l'inter-spécifique. Bien qu'il existe des exceptions, la littérature sur le QS attribue généralement la synthèse des peptides de QS au phylum des Firmicutes (Neiditch *et al.*, 2017; Bhatt, 2019) et celle des petites molécules de QS aux phyla des Protéobactéries et des Actinobactéries (Papenfort and Bassler, 2016; Polkade *et al.*, 2016). Enfin, il est à noter que les signaux de QS n'ont pas toujours pour unique vocation de servir de canal de biocommunication. En effet, certains peptides de QS à l'activité antimicrobienne (lantibiotiques) peuvent à la fois servir des fins de communication et de compétition (Kleerebezem, 2004) quand d'autres molécules de QS de la famille des AHLs et des Hydroxy-Alkyl-Quinolines peuvent par ailleurs fixer le fer (Fe^{3+}) et en faciliter l'absorption (Kaufmann *et al.*, 2005; Diggle, Matthijs, *et al.*, 2007).

Les gènes récepteurs : systèmes à une composante et systèmes à deux composantes

Quels qu'en soient les signaux associés, petites molécules ou peptides, les modules de gènes récepteurs peuvent être classés en deux grandes catégories.

La première correspond à des récepteurs intracellulaires présentant deux domaines protéiques principaux : l'un de liaison au signal de QS et l'autre de régulation, par exemple un domaine de liaison à l'ADN ou d'interaction protéique. Lorsque le signal de QS rentre dans une cellule, par simple diffusion ou via une perméase, sa liaison au récepteur va induire chez ce dernier un changement

conformationnel qui va alors soit activer, soit désactiver son domaine régulateur. Ainsi, si le récepteur est un facteur de transcription, une activation du domaine régulateur entraînera typiquement la transcription ou la répression de gènes cibles alors qu'une désactivation de ce domaine lèvera toute régulation transcriptionnelle réalisée antérieurement par le récepteur à l'état libre. Si le récepteur est un inhibiteur protéique, le même principe s'applique à la nuance près que les régulations sont alors engagées au niveau protéique et non plus transcriptionnel. Ces récepteurs intracellulaires, réalisant à la fois la détection et la transduction du signal sont appelés « systèmes à une composante ».

La deuxième catégorie de modules récepteurs concerne « les systèmes à deux composantes ». Dans ces systèmes, la détection et la transduction du signal de QS sont assurées par deux protéines distinctes, un « récepteur » transmembranaire (une histidine kinase) et son « régulateur de réponse » intracellulaire compagnon. Contrairement aux systèmes à une composante, les signaux de QS reconnus par des systèmes à deux composantes sont détectés à la surface d'une cellule, et ne nécessitent donc pas d'être importés pour être transduits. Suite au changement conformationnel d'un récepteur transmembranaire reconnaissant un signal de QS, son domaine cytosolique va transférer un phosphate au régulateur de réponse intracellulaire, qui devient alors actif et va se détacher du récepteur pour initier la réponse transcriptionnelle de QS. Une liste détaillée des différentes familles de gènes émetteurs et récepteurs de QS en fonction de la nature et des modes de production et de détection de leur signal associé est disponible en **section 2.2**.

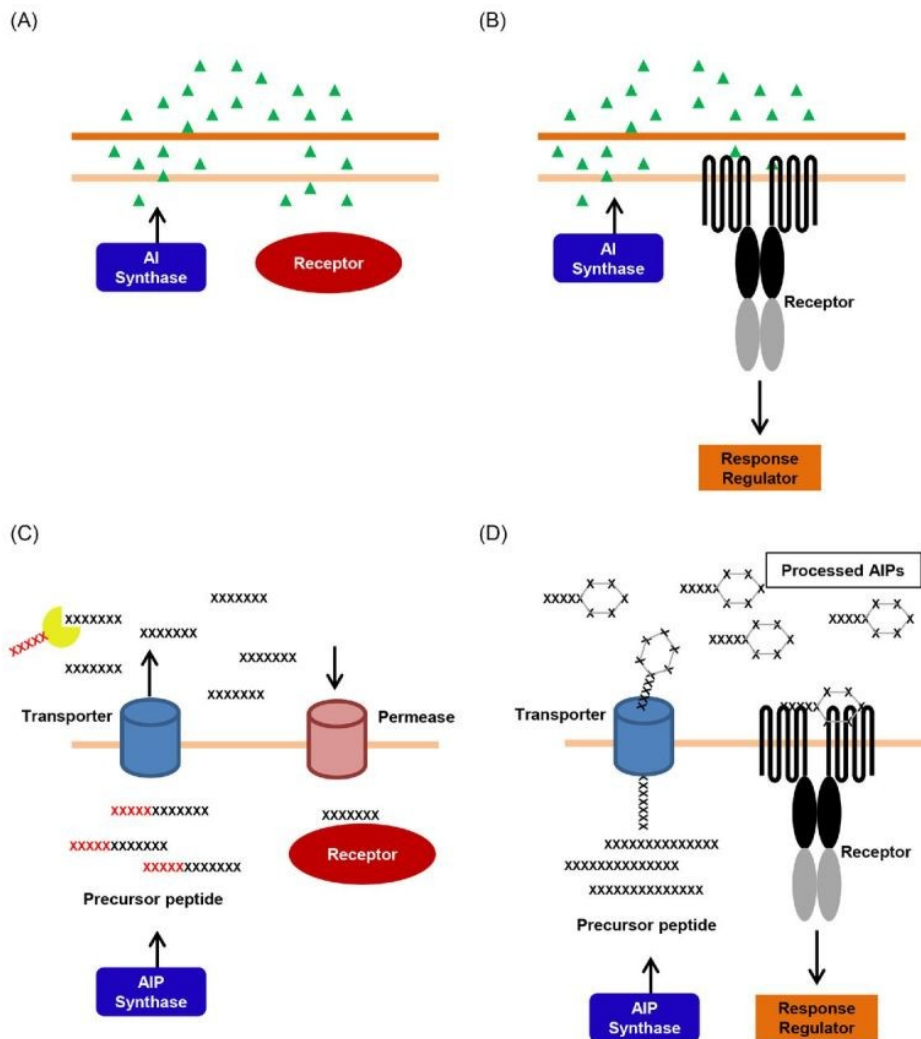


Figure 3: Les quatre grands types de QSSs. A) Système composé d'une ou plusieurs enzyme(s) de synthèse et d'un récepteur intracellulaire à une composante. Le module de synthèse produit une petite molécule de QS (ou une gamme de molécules) à partir de précurseur(s) moléculaire(s). Ce type de QSS est particulièrement prévalent chez les Protéobactéries et les Actinobactéries, peut être autoinductible et réalise généralement une communication intra-espèce. B) Système composé d'une ou plusieurs enzyme(s) de synthèse et d'un module de détection à deux composantes, formé par un récepteur membranaire et son régulateur de réponse intracellulaire compagne. Ce type de QSS associé à une ou des petite(s) molécule(s) de QS est prévalent chez les Protéobactéries, mais peut être retrouvé parmi d'autres phyla bactériens dans le cas de signaux inter-espèces tels que l'AI-2. C) Système qualifié de RRNPP, composé d'un propeptide et d'un récepteur intracellulaire à une composante. Le propeptide est clivé post-traductionnellement en un peptide de QS par l'entremise d'exopeptidase(s). Le peptide mature est importé de façon active et irréversible. Ce type de QSS réalise une communication de courte portée dont la spécificité peut aller de l'intra-souche à l'intra-espèce. Ces QSSs sont codés par les chromosomes, plasmides et phages de Firmicutes. D) Système composé d'un propeptide et d'un module de détection à deux composantes. Les maturations post-traductionnelles du propeptide sont plus complexes, et peuvent aboutir à des peptides cycliques composés d'acides aminés modifiés. La spécificité de ces signaux peut aller de l'intra-souche à l'intra-genre. Ces QSSs réalisent une communication de longue portée et sont retrouvés chez les Firmicutes. [Source : cette thèse et (Hawver, Jung and Ng, 2016)]

Organisation génétique des systèmes de QS

Bien que ce ne soit pas une règle absolue, les gènes émetteurs et récepteurs des signaux de QS ont tendance à être groupés ensemble dans les génomes. En effet, les réarrangements génétiques résultant en une colocalisation des gènes participant à un même processus biologique semblent être favorisés par la sélection naturelle. Lorsque groupés ensemble, les modules émetteurs et récepteurs sont souvent retrouvés dans un contexte co-directionnel (sur même brin d'ADN et généralement en opéron $\rightarrow \rightarrow$) ou dans un contexte de divergence (sur brins d'ADN opposés, dans des directions opposés $\leftarrow \rightarrow$). En effet, un contexte co-directionnel permet à moindre coût, *i.e.* en un seul passage de l'ARN polymérase, de cotranscrire le(s) gène(s) émetteur(s) et le(s) gène(s) récepteur(s) lorsqu'un système de QS a évolué vers une modalité de co-expression de ses composantes. Les contextes divergents sont davantage spécifiques aux QSSs dont le récepteur est intracellulaire bien que ces QSSs soient aussi fréquemment retrouvés dans un contexte co-directionnel. Dans un contexte divergent, l'expression du gène émetteur est découplée de celle du récepteur mais la régulation du gène émetteur est souvent sous contrôle du récepteur. La colocalisation de ces deux gènes permet alors au récepteur, une fois traduit, de parcourir un chemin minimal pour réguler la transcription du gène émetteur. Selon ce même principe d'économie cellulaire, les gènes cibles régulés par les systèmes de QS dont le récepteur est à une composante ont tendance à se trouver dans le voisinage génomique du système de QS (Engbrecht, Neilson and Silverman, 1983; Hoover *et al.*, 2015; Erez *et al.*, 2017; Brotherton, Medema and Greenberg, 2018; Stokar-Avihail *et al.*, 2019; Rued *et al.*, 2021). Au cours de cette thèse, cette propriété de proximité spatiale des récepteurs de QS à une composante avec les gènes qu'ils régulent sera d'ailleurs exploitée à des fins d'inférence fonctionnelle des QSSs. Lorsque le module de détection est à deux composantes, le récepteur et le régulateur de réponse sont adressés à la membrane et la pression de sélection est donc probablement moins forte pour une proximité génomique des gènes récepteurs et des gènes cibles. De façon très intéressante, la tendance à la colocalisation des gènes composant un QSS peut être exploitée pour la détection *in silico* de QSSs, notamment dans le cas de variants divergents naturellement plus sujets aux faux positifs.

Communication courte-portée et longue-portée

Dans le cas des peptides de QS, il fut brillamment démontré que lorsque le module récepteur est à une composante, la communication réalisée par le signal de QS est de courte portée tandis que lorsque le module récepteur est à deux composantes, la communication est de longue portée (Co *et al.*, 2020; van Gestel *et al.*, 2021). Dans le premier cas, le signal est en effet irréversiblement absorbé par les cellules voisines alors que dans le deuxième cas, le signal est relâché dans l'environnement extracellulaire suite à sa détection par un récepteur de surface et peut donc « ricocher » de cellules en cellules. Au regard de la tendance à la colocalisation des entités apparentées génétiquement dans les

communautés microbiennes (notamment du fait de la division cellulaire), il est intéressant de noter que les signaux de communication longue portée sont plutôt population- et espèce-spécifiques alors que les signaux de communications courte portée sont plutôt sous-population- et souche-spécifiques. Le lien de causalité entre communication courte-portée et absorption irréversible du signal n'est toutefois pas valable pour les petites molécules de QS. En effet, alors que les propeptides de QS sont sécrétés, clivés par des exopeptidases et que le peptide mature est importé de façon active et irréversible, les petites molécules de QS reconnues par des récepteurs à une composante diffusent à travers les membranes cellulaires et peuvent donc recirculer dans l'environnement extracellulaire suite à leur absorption.

Auto-induction et accumulation exponentielle du signal de QS

On ne pourrait présenter les bases génétiques du QS sans aborder le phénomène d'autoinduction propre à certains systèmes de QS, dont le fameux système *luxI-luxR*. En effet, dans certains cas, un complexe protéique récepteur-signal, par exemple LuxR-AHL, va activer la transcription du gène émetteur voire du QSS complet. Dans de tels cas d'autoinduction, les signaux de QS, qualifiés alors d'« autoinducteurs », ne s'accumulent pas linéairement dans l'environnement à mesure que la population émettrice croît mais bien exponentiellement. Les cas d'autoinduction sont intéressants d'un point de vue bioinformatique/génomique car ils permettent d'inférer à partir du promoteur de l'opéron du système de QS, le motif d'ADN susceptible d'être reconnu par le récepteur et donc, d'inférer de potentiels gènes cibles régulés par le système de QS.

Paramètres génétiques influant sur le quorum

Le cas de l'autoinduction introduit la notion que le quorum associé à un QSS donné dépend, en plus de paramètres environnementaux, de paramètres génétiques. Deux grands types de paramètres génétiques peuvent alors être identifiés : ceux influant sur le taux de production des émetteurs et des récepteurs, et ceux influant sur l'affinité du récepteur vis à vis du signal. En jouant sur le taux de transcription et de traduction des gènes du système de QS (e.g. fréquence d'utilisation des codons, force de promoteur, affinité du ribosome pour le ribosome-binding site du transcrit, boucle d'autoinduction) ou en optimisant/réduisant l'affinité du récepteur pour son signal par évolution dirigée, il devient alors possible en biologie synthétique d'ajuster le quorum associé à un système de QS (Maire and Youk, 2015; Tang *et al.*, 2020). Par ailleurs, ces possibilités d'optimisation peuvent aussi être exploitées dans des contextes naturels. Par exemple, il fut reporté que l'interaction d'un episymbiote obligatoire (*Candidatus saccharibacteria* TM7) avec son hôte (*Actinomyces odontolyticus*) induit chez ce dernier une surproduction d'un récepteur de QS, précipitant ainsi sa réponse au QS. Comme cette réponse correspond à la formation de biofilm, cette manipulation du circuit de QS de

l'hôte par le symbiote pourrait renforcer, à l'avantage du symbiote, la stabilité de son association avec son hôte (Bedree *et al.*, 2018; Tang *et al.*, 2020).

Gènes de quorum quenching et d'inhibition du QS

Si le quorum sensing permet à un microorganisme d'optimiser sa croissance et ses chances de survie en fonction des fluctuations dans son environnement social et physico-chimique, il n'est pas surprenant que la compétition entre espèces ait pu faire émerger des gènes à même de saboter les canaux de communication d'autres espèces. De tels gènes peuvent être classés en deux catégories : ceux dits de quorum quenching, codant des enzymes modifiant/dégradant des signaux de QS, et ceux dits de QS inhibition, codant des synthèses d'analogues de signaux de QS entrant en compétition inhibitrice avec les signaux de QS pour leur récepteur dédié chez d'autres espèces (Grandclément *et al.*, 2015).

2.2. Diversité des signaux de QS : les « langues » et les « dialectes »

Après cette présentation des deux grands types de signaux (petites molécules et peptides de QS) et des propriétés générales associées aux quatre classes de QSS (**Figure 3**), il convient d'aborder la diversité structurelle des différentes petites molécules et différents peptides de QS et de présenter les différentes familles de gènes émetteurs et récepteurs qui leur sont associées. Ces connaissances seront en effet nécessaires au développement d'outils bio-informatiques de détection de gènes de QS. Pour structurer l'exploration de cette diversité, les signaux seront regroupés en familles structurelles (langues), elles même partitionnées en variants structurels (dialectes).

Langues et dialectes moléculaires

Les paragraphes précédents ont commencé à introduire la notion de multiplicités des signaux de QS. Ces signaux peuvent être classés en différentes familles, selon leur structure chimique. Par exemple, les acyl-homosérine lactones (AHLs) abordées précédemment forment une telle famille de signaux. Les familles de petites molécules de QS regroupent un ensemble de molécules partageant une structure générale commune mais présentant néanmoins quelques variations structurelles. Dans le cas des AHLs, cette partie commune comprend un cycle lactone associé à une chaîne acyle et les variations structurelles observées concernent la longueur, la saturation, l'hydroxylation, la carbonylation, la phénylation et/ou la phenolation de la chaîne carbonée liée au groupement acyle (Davis, Muller and Haynes, 2015) (**Figure 4**). Selon le champ lexical de la communication, le signal générique AHL pourrait à ce titre être qualifié de « langue », et chacun de ses variants structuraux de « dialecte ». De telles notions de « langues » et de « dialectes » seraient applicables à toutes les

petites molécules de QS bien que pour les familles de signaux inter-espèces tels que l'AI-2 ou l'indole à la structure chimique très simple, les variations structurales possibles sont très limitées (**Figures 10 et 14**). Les signaux inter-espèces constitueraient alors des sortes de « langues universelles » avec peu ou pas de dialecte(s) existant, d'où leur inter-spécificité (Sun *et al.*, 2004) (**section 2.3**). Dans le cas des peptides de QS, on pourrait considérer une langue comme un ensemble de peptides matures dont les propeptides sont homologues. Les différents « dialectes » correspondraient alors à des peptides matures se différenciant les uns des autres par des substitutions, des insertions ou des délétions d'acides aminés (**Figure 5**). Les peptides de QS se différenciant par une simple substitution d'un acide aminé de même type chimique pourraient être qualifiés de dialectes analogues (e.g. SRKAT et SRNAT dans la **figure 5**).

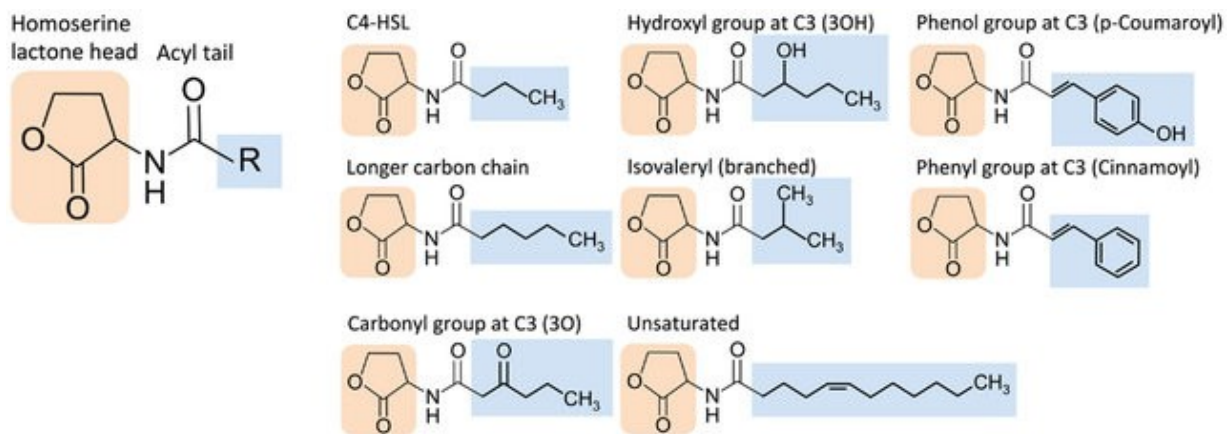


Figure 4: La famille des acyl-homoserine lactones (langue) et ses variants structuraux (dialectes) [Source : (Davis, Muller and Haynes, 2015)]

Table 1. Sequence of *Bacillus subtilis* Phr proteins

| Phr Protein | Signal Sequence Domain ¹ | | | Secreted Domain ² | |
|------------------------|-------------------------------------|---------------------|------------------|------------------------------|-----------------------------|
| | N | H | C | Pre-peptide | Mature Peptide ³ |
| PhrA | MKSK | WMSGLLLAVGFSFTQVM | VHA | GETANTEGK TFHIA | ARNQT |
| PhrC | MKLKSK | LFVICLAAAIFTAAGVAS | ANA | EAL DFHV T | ERGMT |
| PhrE | MKSK | LFISLSAVLIGLAFFGS | MYN ⁴ | GEMKEA | SRNVT ⁵ |
| PhrF | MKLKSK | LLSCLALSTVGVATTI | ANA | PTH QIEVA | QRGMI |
| PhrG | MKR | FLIGAGVAAVILSGW | FIA ⁴ | DHQTHSQ EMKVA | EKMIG |
| PhrI | MKISR | ILLAAVILSSVFSI | TYL ⁴ | QSDHTE EIKVAA | DRVGA |
| PhrK | MKK | LVLCVSILAVILSGVAL | TQL ⁴ | STDSPS NIQVA | ERPVG ⁵ |
| Phr-pLS20 | MKKIN | GWIVALLAVTTVGA | AAA | IQYTNADSPG QFQVA | QKGMV |
| Phr-pTA1040 | MKFN | ALLLLIVCASLLIVSGSSF | VIQ ⁴ | QDS NVSVA | SRKAT |
| Phr- | MTFKK | IMAAVLILAVTVAPVY | GLA ⁴ | TQDNS SVSVA | SRNAT |
| Phr-pTA1060 | MKFK | GLFSAVLIVSLLVGAGY | SFV ⁴ | HHDE EVSVA | SRNAT |
| Phr-pPL10 | MSIKK | SFALAGLAITMLVGGSY | LNH ⁴ | TET SMELA | IRFVT |
| Consensus ⁶ | 2-3 basic | 14-19 hydrophobic | | <u>a</u> hcVA p p | b |

Figure 5: La famille génétique de propeptides Phr (langue) et ses différents peptides matures en C-terminal (dialectes) [Source : (Pottathil and Lazazzera, 2003)]

Diversité des signaux de QS et de leurs gènes de QS associés

Un petit aperçu non exhaustif de la diversité des signaux de QS, à la fois représentatif des petites molécules et des peptides de QS, et capturant les notions de « langues » et « dialectes » moléculaires est donnée en **figure 6**. Pour explorer cette diversité plus en détails et présenter les différentes familles de gènes de QS associées à chaque langue, une dichotomie sera appliquée ci-après entre la petites molécules de QS et peptides de QS. À noter que chacun de ces deux grands types de signaux fait l'objet d'une base de données publique de référence : Sigmol pour les petites molécules (Rajput, Kaur and Kumar, 2016), Quorumpeps pour les peptides (Wynendaele *et al.*, 2013).

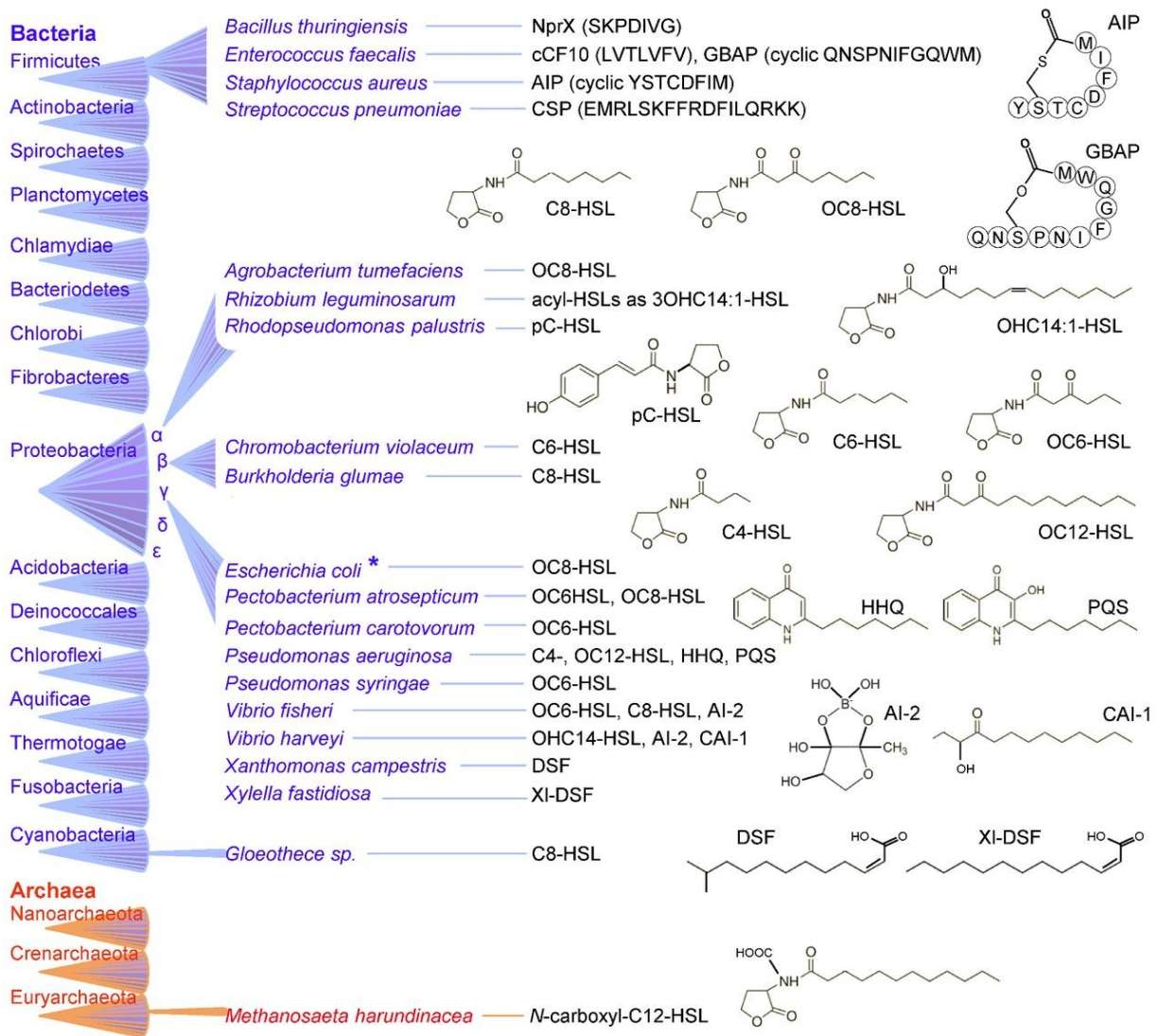


Figure 6 : Diversité non exhaustive des signaux de QS cartographiés sur différents clades procaryotes [Source : (Grandclément *et al.*, 2015)]

Petites molécules de QS

Table 1 : Les différentes langues de petites molécules de QS et leurs différentes familles de gènes de QS associées [Source : cette thèse]

| Famille de signaux (langue) | Quelques molécules représentatives (dialectes) | Famille(s) de synthèses | Famille(s) de récepteurs 1c=one-component, 2c=two-components | Taxa représentatifs | References Pubmed ID |
|--|--|-------------------------------------|--|---|--|
| AHKs (Alpha-Hydroxy-Ketones) | CAI-1, LAI-1, JAI-1 | CqsA | CqsS (2c) | <i>V. cholerae</i> , <i>L. pneumophila</i> , <i>Janthinobacterium</i> | Sigmol, 17614967, 23405110 |
| AHLs (Acyl-Homosérine-Lactones) | Voir figure 3 | LuxI Hdts AinS | LuxR (1c) AinR (2c) | <i>A. fischeri</i> , <i>P. aeruginosa</i> , <i>B. cenocepacia</i> , <i>R. leguminosarum</i> <i>P. fluorescens</i> , <i>A. ferrooxidans</i> <i>A. fischeri</i> , <i>V. anguillarum</i> | Sigmol, 24273537 29977398, 29967162 Sigmol, 21125993 Sigmol |
| cAHLs (carboxyAHLs) | HOOC12-AHL | Fill | FilR (1c) | <i>M. harundinacea</i> (archée) | Sigmol |
| AI-2 (autoinducer-2) | R-THMF-borate S-THMF | LuxS | LsrB (2c), LuxP (2c), AibA (1c), AibB (1c), PctA (2c), PlcQ (2c), KinD (2c) etc... | <i>V. cholerae</i> , <i>E. coli</i> , <i>H. pilori</i> , <i>P. aeruginosa</i> , <i>B. subtilis</i> | Sigmol, 26152582 33097715 |
| DAR/CHD (Dialkylresorcinol/Cyclohexanedione) | DAR, CHD | DarA + DarB (cluster of enzymes) | PauR (1c) | <i>P. asymbiotica</i> | Sigmol |
| Diketopiperazines | cPY, cPV, cPL | PchE + PchF | LasR (1c) | <i>P. aeruginosa</i> | Sigmol, 27900439, 21482761 |
| DSFs (Diffusible Small Factors) | DSF, BDSF, CDSF | RpfB + RpfF (cluster) | RpfC (2c), BCAM0027 (2c), RpfS (2c), RpfR (1c) | <i>X. campestris</i> , <i>X. fastidiosa</i> , <i>S. maltophilia</i> , <i>B. cenocepacia</i> | Sigmol, 27979499 |
| Ethanolamine | | GlpQ, UgpQ | CqsR (2c) | <i>V. cholerae</i> | 32059031 |
| Farnesol | | DPP3 | ? | <i>C. albicans</i> (eucaryote) | 12954333 |
| GBLs (Gamma-Butyrolactones) | A-factor | AfsA | ArpA (1c) | <i>S. griseus</i> | 26904007 |
| HAQs (Hydroxy Alkyl Quinolines) | HHQ, PQS | PqsD, PqsH | PqsR (2c) | <i>P. aeruginosa</i> | Sigmol |
| Indole | | TnaA | ? | <i>Vibrio</i> genus | 22282601 |
| IQS (Hydroxyphenyl thiazole carbaldehyde) | IQS | AmbB + AmbC + AmbD + AmbE (cluster) | ? | <i>P. aeruginosa</i> | 23542643 |
| Méthylesters | PAME, MAME | PhcB | PhcR (2c) | <i>R. solanacearum</i> | 28642776 |
| PPYs (Photopyrones) | PPY | PpyS | PauR (1c) | <i>P. thracensis</i> | Sigmol |
| Pyrazines/Pyrazynols | DPO AI-3 | Tdh | Vqma (1c) QseC (2c) | <i>V. cholerae</i> <i>E. coli</i> | 28319101 32123737 |

Un aperçu des molécules synthétisées et reconnues par ces protéines est donnée en **figure 7**.

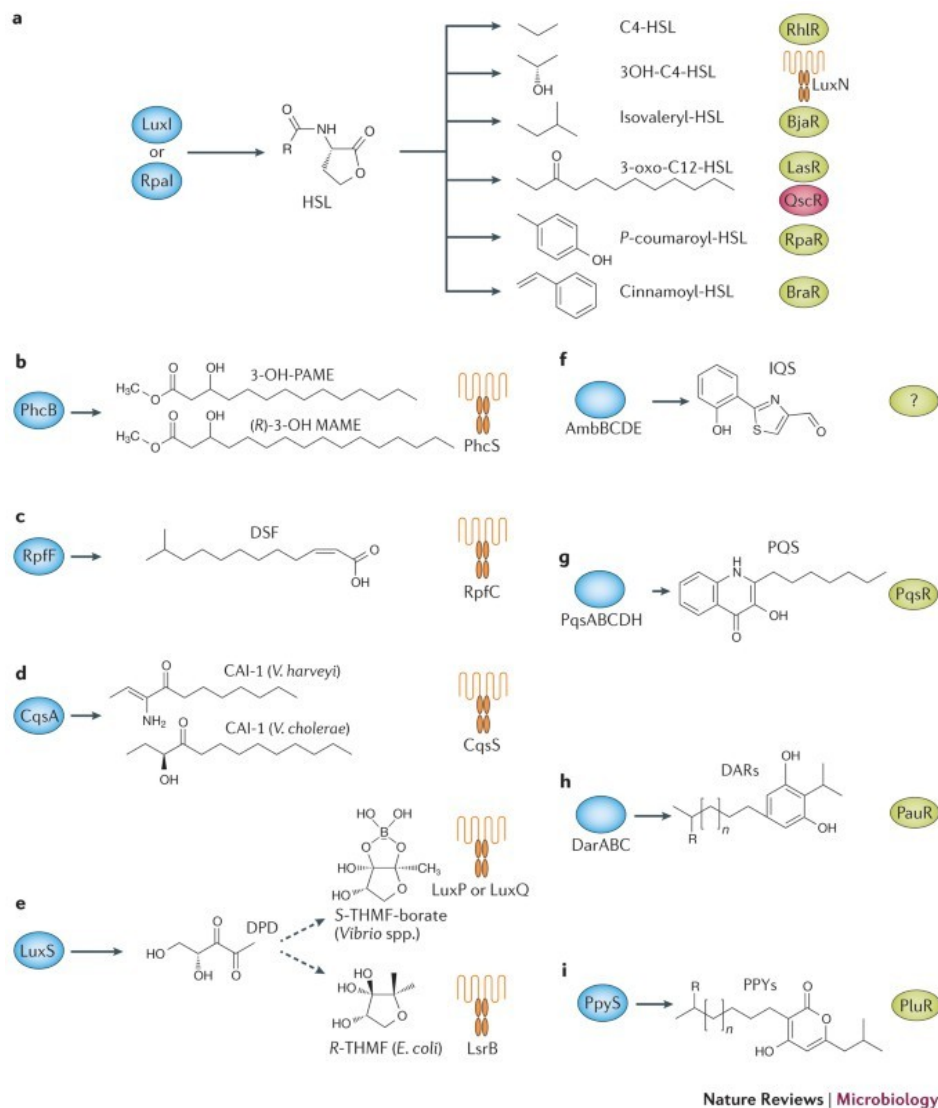


Figure 7: Diversité structurale non exhaustive des petites molécules de QS. Ellipse bleue : synthase ou cluster de synthases ; ellipse verte : récepteur intracellulaire (constituant un système à une composante) ; protéine orange : récepteur transmembranaire (faisant partie d'un système à deux composantes) [Source : (Papenfort and Bassler, 2016)]

Peptides de QS

Les systèmes de QS basés sur les peptides, particulièrement prévalent dans le phylum des Firmicutes, peuvent être classés en 2 grandes classes : ceux dont le module récepteur est à deux composantes (**Figure 8**), et ceux dont le récepteur est à une composante (**Figure 9**). Les systèmes de QS peptidique dont le récepteur est à une composante sont regroupés sous le terme RRNPP, un acronyme correspondant aux initiales des 5 premières familles découvertes de ces récepteurs : Rap

(genre *Bacillus*), Rgg (genre *Streptococcus*), NprR (groupe *Bacillus cereus*), PlcR (genre *Bacillus*) et PrgX (plasmides d'*Enterococcus faecalis*) (Neiditch *et al.*, 2017). Un aperçu des processus que ces systèmes de QSSs régulent est donné en **figure 9**. Bien que les systèmes RRNPP soient génétiquement divers, ils fonctionnent néanmoins selon un même mécanisme canonique (**Figure 9**), ce qui sous-tend des propriétés communes entre les différentes familles de QSSs du type RRNPP. Par exemple, les récepteurs RRNPP interagissent toujours avec leur peptide cognat via des motifs structurels appelés « tetratricopeptide repeats » (TPRs) quand les propeptides RRNPP arborent quasiment toujours, à de rares exceptions près (PrgQ et Shp), une séquence signale en N-terminal pour l'adressage au SEC translocon (Neiditch *et al.*, 2017). En raison de leurs propriétés communes remarquables, les systèmes de QS qualifiés de « RRNPP » recevront une attention particulière au cours de cette thèse.

Table 2 : Les différents types de peptides de QS et leurs différentes familles de gènes de QS associées (Les QSSs qualifiés de RRNPP dont le récepteur est intracellulaire sont grisés)
[Source :cette thèse]

| Famille de signaux | Quelques peptides représentatifs | Famille(s) de propeptides | Famille(s) de récepteurs transmembranaires | Taxa représentatifs | References Pubmed ID |
|--|--|--|---|---|----------------------|
| AIP (Autoinducer peptides) et Bactéricocines de classe 2 | YSTCDFIM thiolacton linkage between C4 and M8, QNSPNIFGQWM lacton linkage between S3 and M11, EMRISRILDFLFLRKK, ADPITRQW(Farnesyl)GD) ... | AgrD (inclut AgrD, FsrD, ComC, EntF, PltA ...), ComX, MrsA | AgrC (inclut AgrC, FsrC, ComD, EntK, PltK ...), ComP, MrsK2, VirS | <i>S. aureus</i> , <i>E. faecalis</i> , <i>S. pneumoniae</i> , <i>E. faecium</i> , <i>L. plantarum</i> <i>B. subtilis</i> <i>B. subtilis</i> <i>C. perfringens</i> | Quorumpeps |
| Bacteriocines de classe 1 (lantibiotiques) | Nisin: I (Dhb)AI (Dha)LA (Abu)PGAK (Abu) GALMGANMK (Abu)A (Abu)AHASIHV (Dha)K, cyclisation between (Ala3-S-Ala7), (Abu8-S-Ala11), (Abu13-S-Ala19), (Abu23-S-Ala26) and (Abu25-S-Ala28) | NisA (inclut NisA, SpaA, SrtA ...) | NisK (inclut NisK, SpaK, SrtK ...) | <i>L. lactis</i> , <i>B. subtilis</i> , <i>S. pyogenes</i> | Quorumpeps |
| TM0504 | PWASPSAARVFALGRDLPR AGCPLPQPS | TM0504 | ? | <i>T. maritima</i> | Quorumpeps |
| Phr | ARNQT, ERGMT | Phr | Rap | <i>Bacillus</i> genus | Quorumpeps |
| Shp | MDIIIIVGG | Shp | Rgg | <i>Streptococcus</i> genus | Quorumpeps |
| NprX | SKPDIVG | NprX | NprR | <i>Bacillus cereus</i> group | Quorumpeps |
| PapR | LPFEF | PapR | PlcR | <i>Bacillus cereus</i> group | Quorumpeps |
| iCF10, cCF10 | AITLIFI | PrgQ | PrgX | <i>Enterococcus faecalis</i> plasmids | Quorumpeps |
| IPD1 | ALILTLVS | lpd1 | TraA | <i>Enterococcus faecalis</i> plasmids | Quorumpeps |
| Arbitrium | SAIRGA | AimP | AimR | <i>Bacillus subtilis</i> phages | Quorumpeps |
| Arbitrium-like | MMSEPGGGGW | AimP-like | AimR-like | <i>Bacillus cereus</i> phages | 31071296 |

| | | | | | |
|------|----------|------|------|-----------------------------------|----------|
| AloP | SHGRGG | AloP | AloR | <i>Paenibacillus polymyxa</i> | 32764104 |
| QspB | LGAEPTWG | QspB | QsrB | <i>Clostridium acetobutylicum</i> | 32375981 |

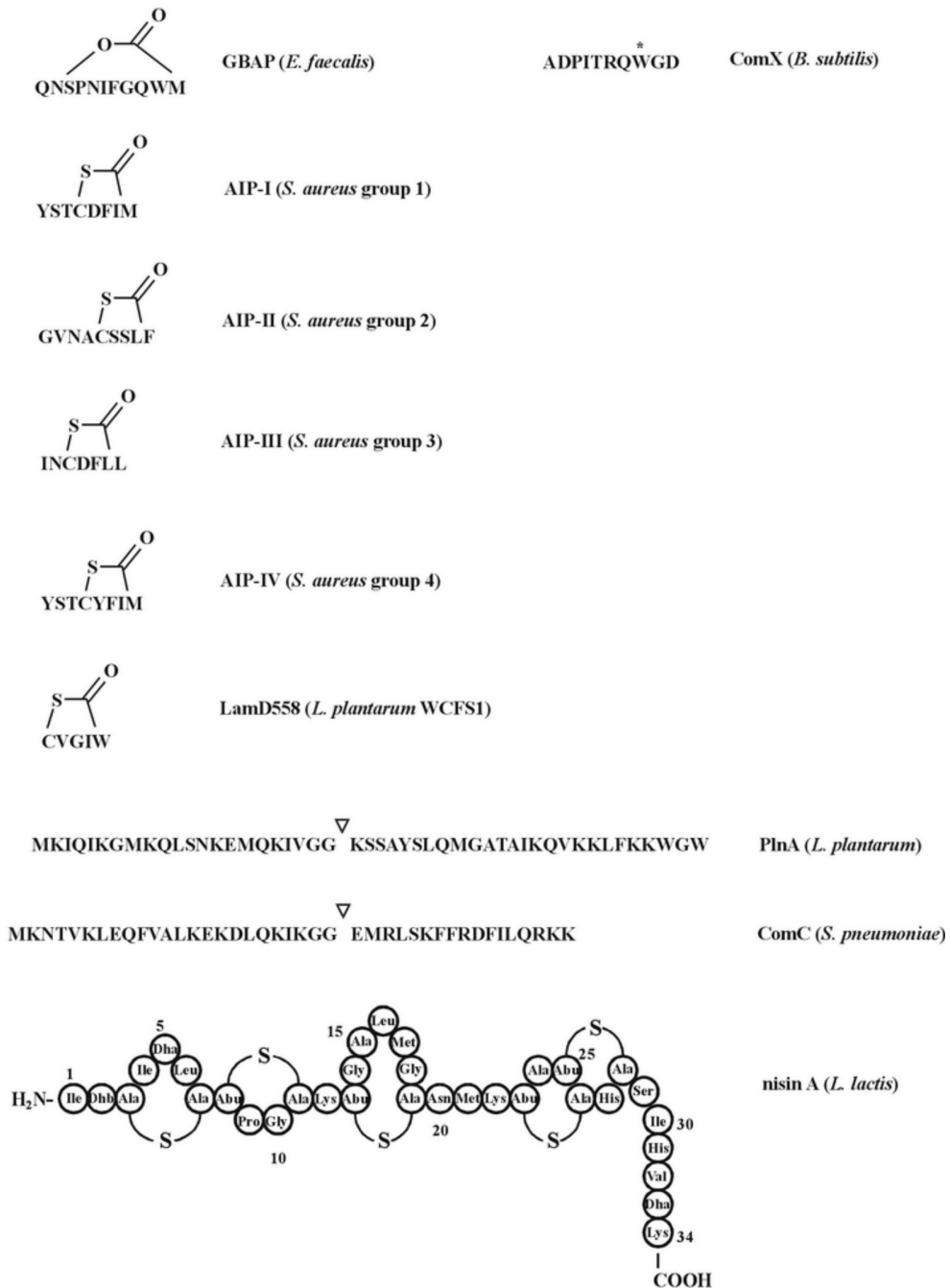


Figure 8: Diversité structurelle non exhaustive des peptides de QS reconnus par des systèmes à deux composants [Source (Sturme, 2005)]

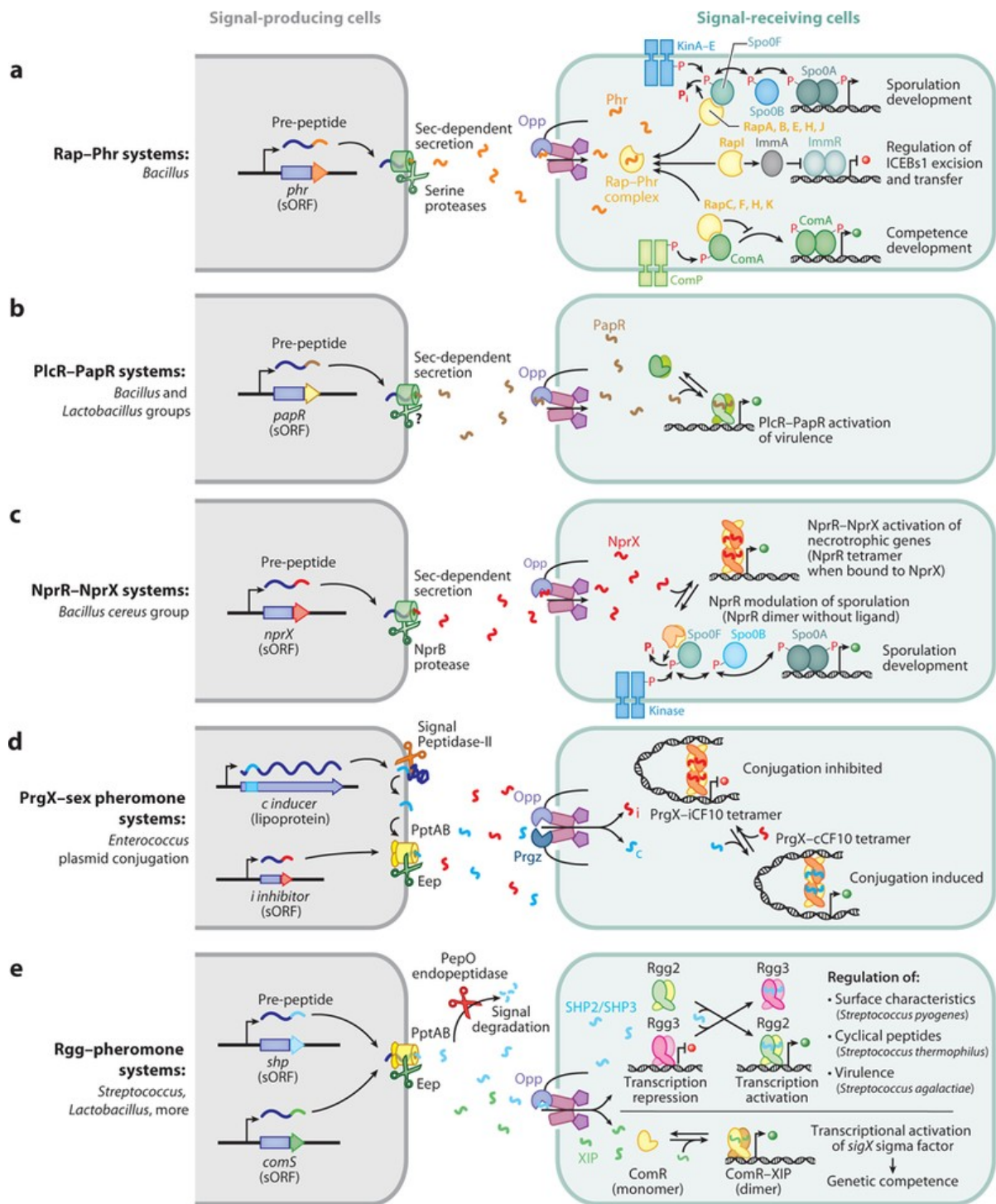


Figure 9: Diversité non exhaustive des réponses au quorum régulées par des QSSs de type RRNP [Source (Neiditch *et al.*, 2017)]

Liste des séquences protéiques et des profils HMMs de domaines clés de synthèse ou de reconnaissance de signaux regroupés par familles de gènes

La classification de cette diversité de signaux et de familles de gènes de QS représente un premier pas vers l'identification d'homologues de QSSs déjà connus au sein de génomes d'intérêt. Un travail minutieux de recensement fut par la suite nécessaire à la récupération des séquences protéiques validées expérimentalement pour chaque famille de synthase, propeptide, récepteur membranaire ou récepteur intracellulaire d'un signal de QS connu et des HMMs de motifs clés spécifiques à ces familles. Les identifiants de ces protéines et de ces HMMs sont disponibles dans les informations supplémentaires de l'article présenté en **section 3.1**. Les fichiers fastas et HMMs sont disponibles sur demande en attendant leur mise en ligne prochaine sur un dépôt GitLab.

2.3. Spécificité des signaux de QS : des « langues » plus ou moins secrètes

Les sections précédentes ont commencé à aborder la notion de spécificité d'un signal de QS, soit sa tendance à refléter un quorum homogène. Un signal de QS est peu spécifique d'une part s'il est émis par un nombre important d'espèces différentes et d'autres part s'il est reconnu par un nombre important d'espèces, notamment via une grande diversité de récepteurs, dont certains pouvant être dédiés à la reconnaissance d'autres molécules. Cette notion de spécificité des signaux de QS est importante pour appréhender la portée écologique des QSSs détectés au sein de génomes d'intérêt. Cette section dressera un catalogue des différents signaux selon leur niveau de spécificité et tentera d'identifier quelques facteurs déterminants de la spécificité d'un signal de QS.

Signaux inter-espèces

Comme évoqué précédemment dans le bref historique du QS, la découverte du premier signal synthétisé et reconnu par une multitude d'espèce différentes date de la fin des années 1990 (Bassler, Greenberg and Stevens, 1997; Surette, Miller and Bassler, 1999). Ce signal moléculaire nommé autoinducer-2 (AI-2) régule une vaste gamme de comportements chez différentes bactéries : de la virulence, à l'acquisition du fer en passant par la motilité, la bioluminescence, la production d'antibiotique ou la division cellulaire (Federle and Bassler, 2003). En 2010, l'indole (Lee and Lee, 2010) s'ajoute à cette liste de signaux inter-espèces pour son émission quasi-ubiquitaire chez les bactéries et son large spectre d'influence sur les phénotypes bactériens : de la sporulation, à la virulence en passant par la formation de biofilms, la stabilité des plasmides et la résistances aux métabolites antimicrobiens. Enfin, la biosynthèse des molécules de QS avec un cycle pyrazine tels que le DPO de *V. cholerae* et l'AI-3 d'*E. coli* fut récemment identifiée comme impliquant la thréonine

déshydrogénase (Tdh) (Papenfort *et al.*, 2017a; Kim *et al.*, 2020), une enzyme conservée chez les bactéries, archées et les eucaryotes, suggérant alors une possible production de tels signaux de QS par une large diversité d'espèces microbiennes (Shi and Bode, 2017).

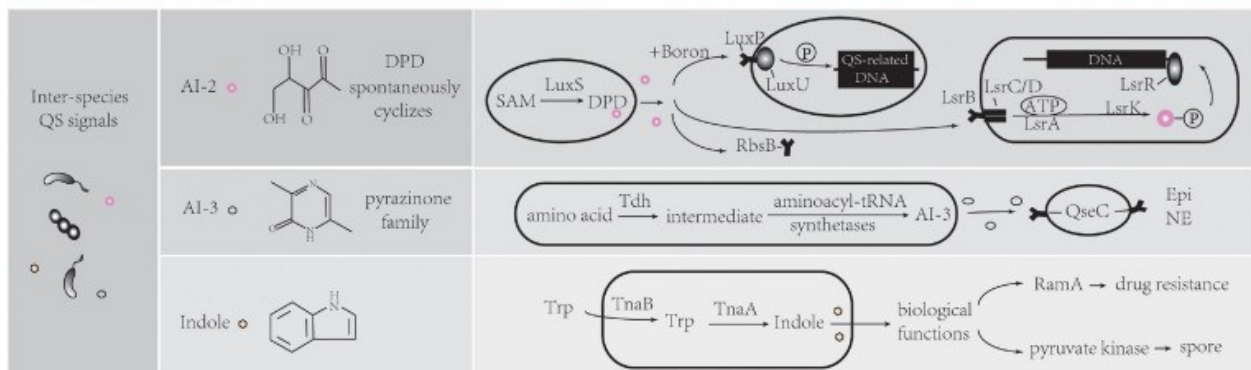


Figure 10: Signaux inter-espèces : Autoinducer-2 (AI-2) ; famille des pyrazinones (e.g AI-3 et DPO) ; Indole. La voie de biosynthèse de ces différents types de signaux ainsi que leur mode de reconnaissance chez certaines bactéries modèles (*V. cholerae* (LuxP), *E. coli* (LsrB, QseC), *S. enterica* (RamA), *S. auriantaca* (pyruvate kinase)) sont données à droite de la figure. [Source : (Wu and Luo, 2021)]

Signaux intra-espèce voire intra-souche

Les signaux les plus spécifiques correspondent aux peptides de QS, notamment du fait de leur complexité structurale imposant d'importantes contraintes physico-chimiques à leur reconnaissance par une protéine (**Figure 11**). De façon intéressante, chez les bacilles, les peptides de QS intra-espèce associés aux QSSs de type ComX-ComP ont tendance à coordonner des comportements de groupe tels que la production de biens communs à haute densité de population hétérogène de bacilles tandis que les peptides de QS intra-souche associés aux QSSs de type Rap-Phr permettent d'inhiber temporairement cette production de biens communs jusqu'à ce que la sous-population de bacilles codant le QSS atteigne une haute densité cellulaire (Even-Tov *et al.*, 2016; Pollak *et al.*, 2016; Kalamara *et al.*, 2018). Le délai de triche réalisé par un système Rap-Phr sera alors d'autant plus long qu'un faible nombre d'individus de la population codera ce système. Autrement dit, les bénéfices des régulations densité-dépendantes d'un système Rap-Phr seront d'autant plus importants que son signal de QS réalisera une communication ultraconfidentielle parmi un faible nombre d'individus. Une pression de sélection peut alors exister chez un bacille d'une part pour l'acquisition par transfert horizontal d'un nouveau système Rap-Phr de triche facultative (certains bacilles pouvant coder plus d'une dizaine de Rap-Phr différents) et d'autre part pour l'évolution rapide d'un système Rap-Phr préexistant dans un génome bactérien vers une ultra-spécificité du signal pour ce génome (**Figure 12**). Cet exemple illustre bien que la tendance d'une famille de signaux de QS à être souche-spécifique peut être due à une pression de sélection pour la divergence rapide de leur QSS associé.

Les exemples précédents (système ComX-ComP altruiste, chromosomique, de longue portée et évoluant plutôt lentement ; systèmes Rap-Phr de triche facultative, plutôt mobiles, de courte portée et évoluant rapidement) semblent alors indiquer une possible corrélation négative entre la pression de sélection pour la divergence d'un QSS et le niveau d'altruisme des comportements contrôlés par un QSS à faible densité de population.

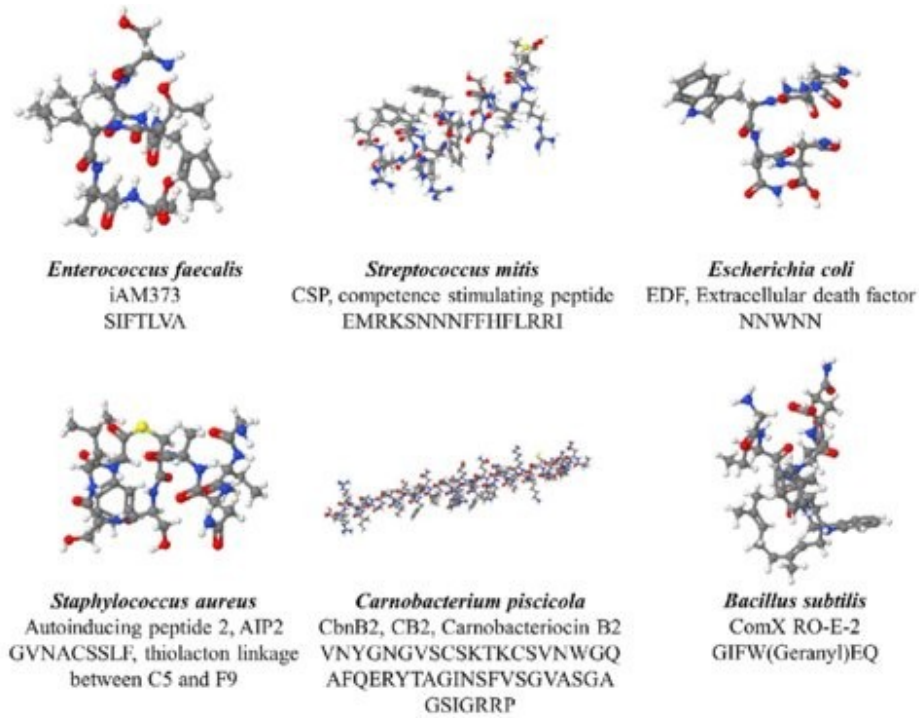


Figure 11 : Complexité structurelle des peptides de QS. Les signaux de QS les plus spécifiques sont les peptides de QS, notamment du fait de leur importante complexité structurelle imposant d'importantes contraintes physico-chimiques à leur reconnaissance par une protéine [Source : (Verbeke *et al.*, 2017)]

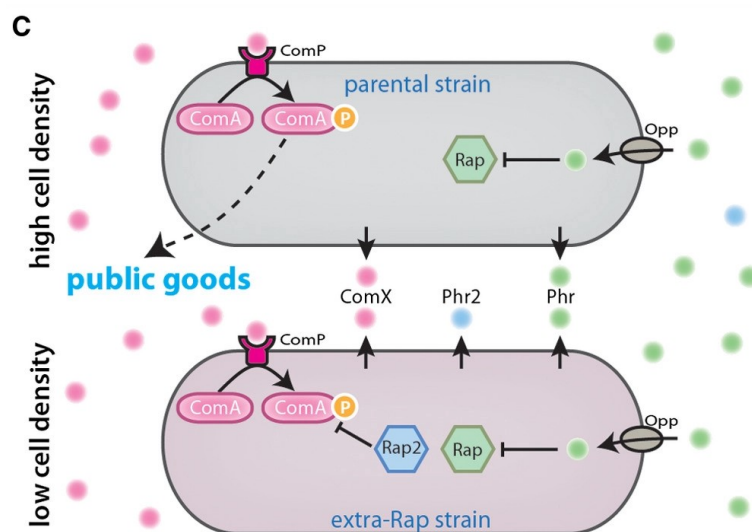


Figure 12: Triche facultative densité-dépendante due à l'acquisition d'un système Rap-Phr ultra-spécifique. Lors d'un quorum hétérogène de bacilles reflété par une haute concentration du peptide intra-espèce ComX (en rouge), la production de biens communs est déclenchée. Un système Rap-Phr spécifique d'une sous-population permet de retarder la participation à cette production en fonction de la densité de cette sous-population. L'acquisition d'un nouveau système Rap-Phr ultra-spécifique (en bleu) par une bactérie permet alors à cette dernière de bénéficier d'un délai de triche plus long (nouveau quorum bleu non atteint) que chez la bactérie ancestrale (quorum vert atteint). [Source : (Kalamara *et al.*, 2018)]

Signaux dont la spécificité varie entre l'intra-espèce et l'inter-espèces

Les signaux de QS dont la spécificité se trouve entre ces deux extrêmes concernent souvent des petites molécules de QS appartenant à une famille comprenant de multiples dialectes, de type AHLs, AHKs, GBDs, DSFs, Dikétopiperazines et Méthyl esthers (**Table 1**). Leur structure chimique est de complexité intermédiaire par rapport aux signaux inter-espèces très simples et aux peptides ultra-spécifiques complexes.

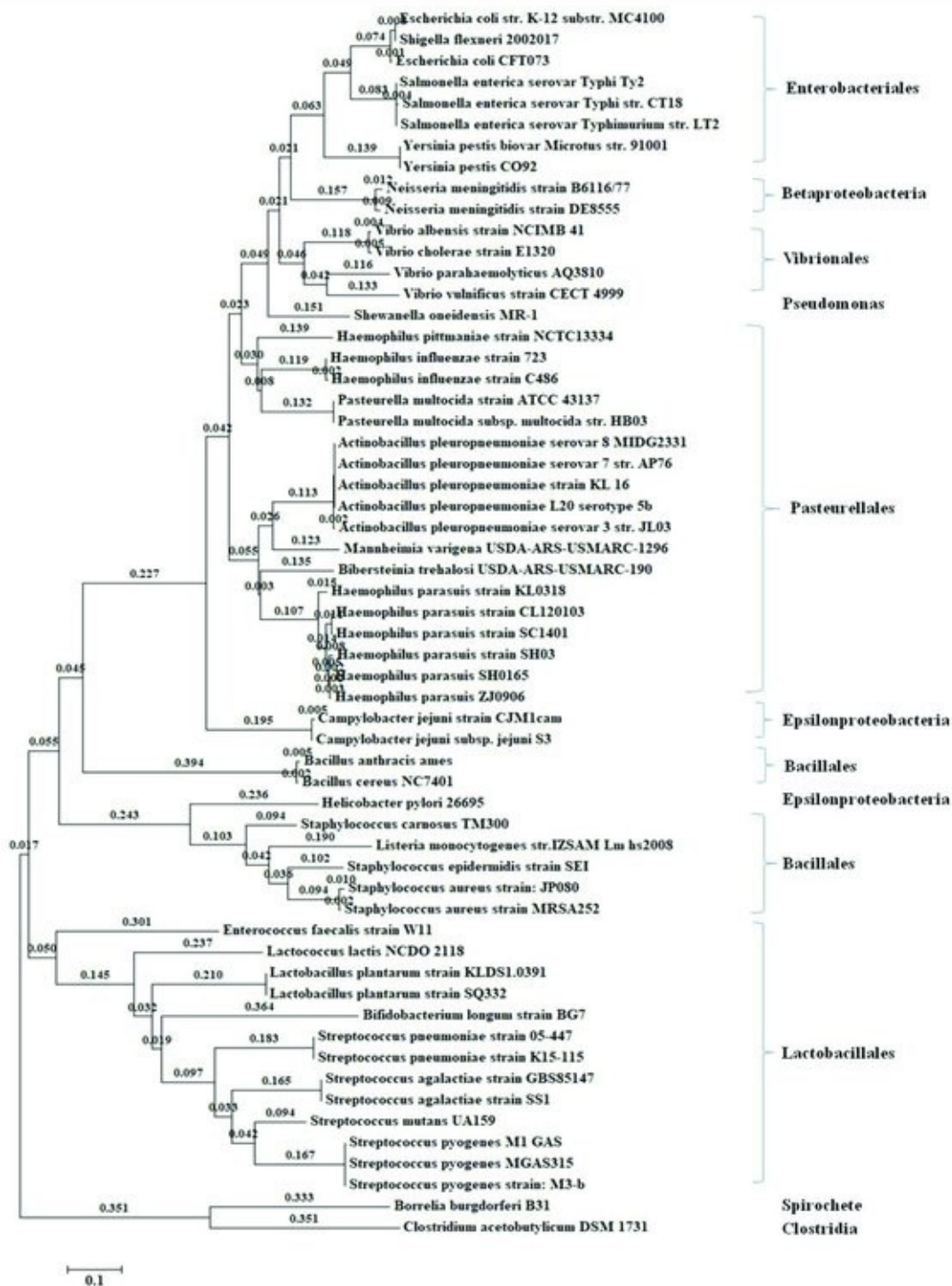
Déterminant de la spécificité taxonomique des signaux de QS

L'existence de molécules de QS inter-espèces et intra-souche questionne inéluctablement les facteurs déterminants de la spécificité taxonomique d'un signal de QS. Une tentative d'identifier quelques uns de ces facteurs, tous interconnectés, est exposée ci après.

- Comme évoqué plus haut, la pression de sélection pour une divergence rapide des QSSs vers une nouvelle sélectivité récepteur-signal en constitue un premier.
- Un deuxième facteur peut concerner la portée du signal de QS. Comme les bactéries ont tendance à être colocalisées selon leur niveau de parenté génétique dans les communautés microbiennes (du fait notamment de la division cellulaire), un QSS réalisant une

communication de courte portée aura d'autant plus de chances d'être spécifique d'un même génome (Co *et al.*, 2020; van Gestel *et al.*, 2021).

- Comme évoqué précédemment, un troisième facteur pourrait correspondre à la complexité de la structure chimique d'un signal de QS. Les signaux les plus simples structurellement ont en effet tendance à réaliser des instances de communication inter-espèces (**Figure 10**) alors que les signaux les plus complexes tendent à être espèce- voire souche-spécifique (**Figure 11**). En effet, plus une molécule de QS sera petite et simple structurellement, moins elle imposera de contraintes physico-chimiques à un récepteur pour sa reconnaissance. Ainsi, plus un signal sera petit et simple structurellement, plus ce signal aura de chances de s'insinuer dans les sites actifs de protéines et plus la gamme de récepteurs et donc d'organismes capable de détecter cette molécule pourra être vaste. Par exemple, une grande diversité de récepteurs et d'espèces capable de répondre à l'AI-2 a d'ores et déjà été caractérisée (Pereira *et al.*, 2009; Anderson *et al.*, 2015; Torcato *et al.*, 2019; Zhang *et al.*, 2020)
- Un quatrième facteur peut concerner la couverture phylogénétique de la famille d'une synthase de QS. Plus la distribution taxonomique d'une famille de synthases sera vaste, plus la probabilité que deux espèces différentes codent une synthase de cette famille sera grande, et plus la probabilité que deux espèces différentes produisent une même molécule de QS sera importante. À titre d'exemple, un arbre phylogénétique de la famille LuxS (la synthase de l'AI-2) est donnée en **figure 13** et couvre plusieurs phyla bactériens. À l'opposé, les signaux les plus secrets/spécifiques sont sécrétés à partir de propeptides retrouvés uniquement au niveau taxonomique de l'espèce.



rp

Figure 13: arbre phylogénétique des synthases LuxS codées par les génomes complets de la base de données KEGG. Arbre inféré par maximum de vraisemblance avec 1000 bootstraps [Source : (Zhang *et al.*, 2019). Selon une étude à plus large échelle, la famille LuxS serait distribuée dans au moins 7 phyla bactériens différents : Protéobactéries, Firmicutes, Actinobactéries, Spirochètes, Bacteroidetes, Fusobactéries, Dienococcus-Thermales (Rao, Pasha and Sowdhamini, 2016)

- Une synthase sera d'autant plus conservée évolutivement qu'elle sera impliquée dans des processus ubiquitaires dans le vivant. Cette notion introduit le cinquième facteur correspondant à l'implication d'une synthase de QS dans une voie importante de métabolisme. En effet, les signaux inter-espèces que sont l'indole et les pyrazines sont respectivement synthétisés par la tryptophanase (TnaA) et la thréonine déshydrogénase (Tdh), deux enzymes directement impliquées dans la dégradation d'acides aminés naturels (**Figure 10**). Enfin, LuxS, la synthase de l'AI-2, est une composante principale du cycle du méthyle activé (AMC) qui permet de recycler la méthionine (Vendeville *et al.*, 2005) (**Figure 14**). À ces égards, un signal de QS inter-espèce peut être considéré comme un produit dérivé du métabolisme primaire. À l'extrême opposé, les peptides ultrasécifiques associés aux systèmes RRNPP sont maturés à partir de propeptides sans autres fonctions que celle de signalisation.

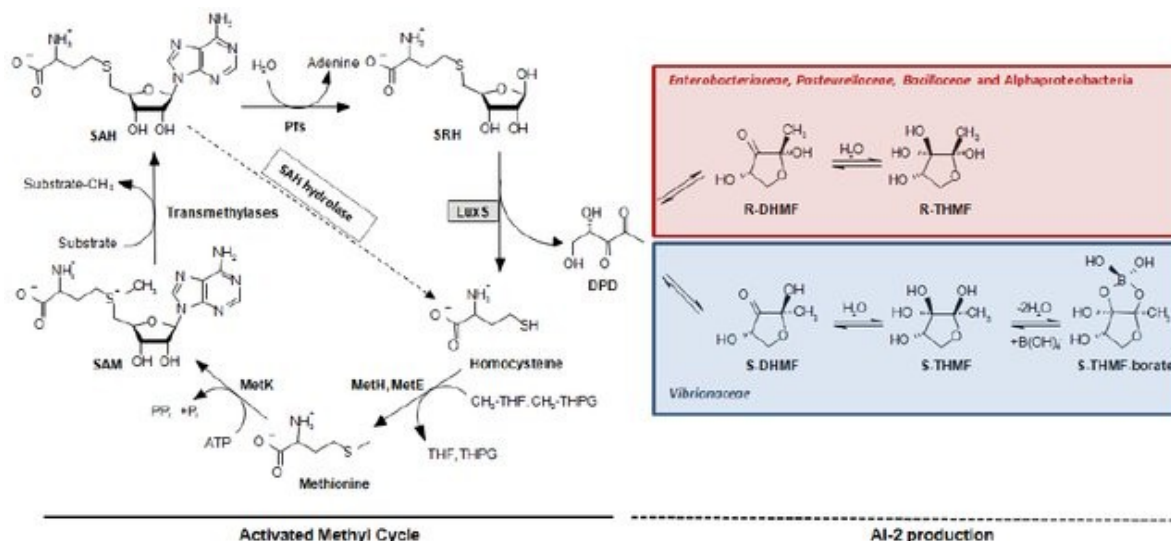


Figure 14: LuxS, la synthase de l'AI-2 est une enzyme clé du cycle du méthyle activé [Source : (Rezzonico, Smits and Duffy, 2012)]

- Un sixième facteur concerne le nombre de variations structurelles observées pour une famille de signaux de QS normalisées par le nombre total d'occurrences de cette famille parmi différentes espèces. Par exemple, bien que la synthase LuxS soit retrouvée dans plusieurs phyla bactériens (**Figure 13**), seuls 2 variants structurels de l'AI-2 sont connus (le R-THMF et le S-THMF-borate) et résultent tous deux de la cyclisation d'un même composé : le DPD (**Figure 14**).
- Cette notion introduit le septième facteur probable, et concerne l'invariabilité structurelle du substrat incorporé par une famille de synthase. En effet, l'utilisation d'un substrat unique par un ensemble de synthases homologues à travers différentes espèces limite les possibilités de variations structurelles de leur produit de biosynthèse. À cet égard, les substrats de Tna, Tdh

et LuxS sont invariables structurellement (tryptophane, thréonine et DPD, respectivement) et les molécules de QS produites par ces différentes synthèses sont toutes inter-spécifiques.

2.4. Hiérarchie intragénomique des systèmes de QS : adaptabilité des micro-organismes « polyglottes »

Comme illustré précédemment avec les bacilles codant à la fois les systèmes ComX-ComP et Rap-Phr, un génome microbien peut coder de multiples QSSs. Pour poursuivre avec cet exemple, ces QSSs peuvent à la fois appartenir à des familles différentes (e.g. ComX-ComP et Rap-Phr) comme représenter de nombreuses occurrences d'une même famille de QSS (e.g. multiples systèmes Rap-Phr dans les génomes de bacilles). Cet exemple illustre également qu'un génome peut coder plusieurs QSSs aux spécificités différentes (ComX-ComP intra-espèce et Rap-Phr intra-souche). La multiplicité et la diversité de QSSs au sein d'un même génome questionne alors les avantages sélectifs du « polyglottisme » microbien. Pour aborder ce thème nous présenterons d'abord l'intérêt évolutif de coder des QSSs aux spécificités différentes, puis celle de coder plusieurs QSSs de même spécificité. Ces notions sont importantes pour l'interprétation des résultats propres à l'identification de multiples QSSs au sein d'un même génome d'intérêt.

Présence de QSSs aux spécificités différentes au sein d'une même génome : estimation ratiométrique de plusieurs densités de populations

La combinaison d'un QSS intra-spécifique et d'un QSS inter-spécifique peut notamment permettre à un microorganisme de contextualiser sa propre densité de population par rapport à celle de sa communauté. Par exemple, une étude fonctionnelle a récemment investigué la synergie de deux QSSs chez *Vibrio cholerae* : l'un intra-espèce voire intra-genre basé sur le signal CAI-1 et l'autre inter-spécifique basé sur le signal AI-2 (Bridges and Bassler, 2019). De façon intéressante, ces deux signaux sont interprétés par la même voie de transduction, voie qui n'est alors activée de façon robuste qu'en présence des deux signaux. La stimulation de cette voie, et donc la réponse non pas au quorum mais aux quorums résulte en une inhibition de la virulence et/ou en une dissolution des biofilms. Un nombre minimal d'individus producteur de CAI-1 (*V. cholerae*) et d'individus producteurs d'AI-2 (*V. cholerae* et, optionnellement, d'autres espèces) sont alors requis pour la dissolution des biofilms par *V. cholerae*. De façon intéressante, en mono-culture, le CAI-1 intraspécifique s'accumule beaucoup plus rapidement que l'AI-2, ce dernier signal pouvant alors être considéré comme le facteur limitant de la dissolution des biofilms. En monoculture, la sortie des biofilms intervient donc seulement à partir d'un nombre important d'individus de l'espèce *V. cholerae*, nombre nécessaire à l'accumulation de l'AI-2 vers une concentration critique. En co-culture, l'histoire est tout autre ; l'AI-2

interspécifique s'accumule plus rapidement du fait de la contribution des autres espèces à sa production. La sortie des biofilms intervient alors à partir d'un nombre beaucoup plus petit d'individus de l'espèce *V. cholerae*. Ces deux QSSs aux spécificités différentes permettent ainsi à *V. cholerae* de sortir des biofilms dès lors qu'un nombre total d'individus apparentés ou non est atteint dans l'environnement local (reflété par l'AI-2) mais uniquement si la condition d'un nombre minimal d'individus de l'espèce *V. cholerae* (reflété par le CAI-1) est respectée (**Figure 15**).

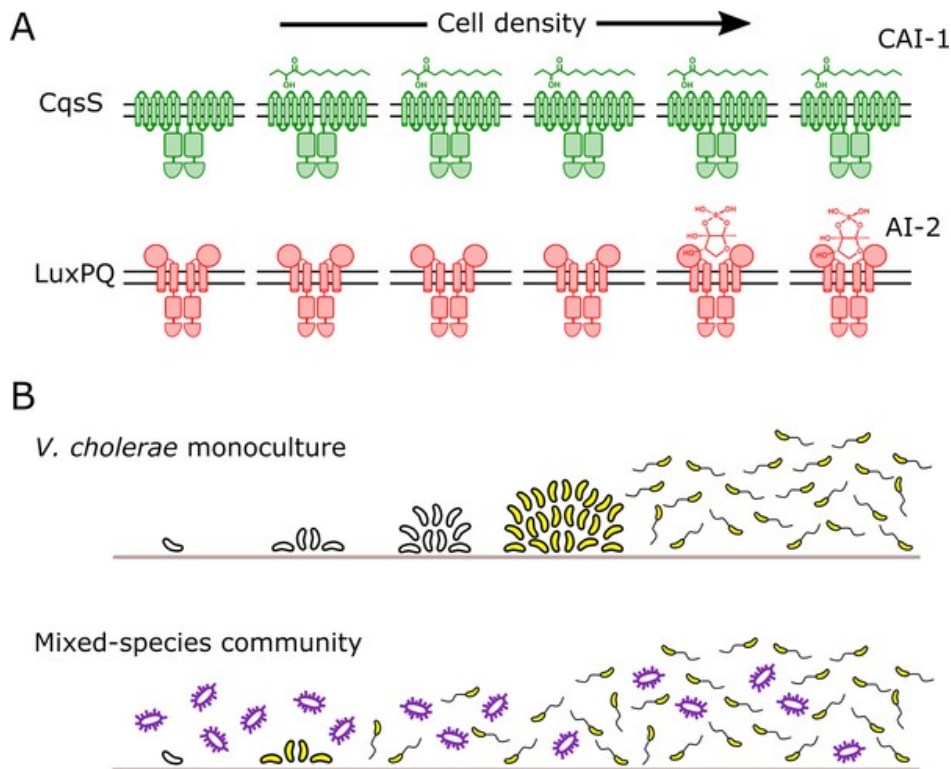


Figure 15: Signaux intra-spécifique (CAI-1) et inter-spécifique (AI-2) « coincidents » chez *V. cholerae*. Ces deux QSSs aux spécificités différentes permettent à la bactérie de sortir des biofilms dès lors qu'un nombre total d'individus apparentés ou non est atteint dans l'environnement local (reflété par l'AI-2) mais uniquement si la condition d'un nombre minimal d'individus de l'espèce *V. cholerae* (reflété par le CAI-1) est respectée [Source : (Bridges and Bassler, 2019)]

Dans l'exemple précédent, la même voie de transduction est stimulée par deux signaux de QS aux spécificités différentes. Ces deux signaux sont alors qualifiés de facteurs coincidents. L'initiation d'un comportement requérant que soient atteints deux quorums caractérisés par des signaux aux spécificités différentes peut néanmoins émerger d'une architecture différente des circuits de QS. Cette architecture alternative correspond au cas de figure où la voie de transduction d'un signal est verrouillée jusqu'à ce que l'autre signal soit transduit, sans que la réciproque ne soit vraie. La combinaison des systèmes ComX-ComP intra-espèce et Rap-Phr intra-souche évoquée précédemment en fournit une parfaite illustration chez les bacilles (**Figure 12**). Lors d'un quorum hétérogène de bacilles reflété par une haute concentration de ComX, ComP détecte ComX et va alors

transférer son phosphate à ComA, le régulateur de réponse activant la transcription de gènes de compétence et de production de biens communs. Néanmoins, ComA-P est dephosphorylé par la forme libre du récepteur intracellulaire Rap. Autrement dit, à haute densité hétérogène de bacilles mais à faible densité d'une sous-population codant un système Rap-Phr, ComA-P est verrouillé et ne peut activer ses gènes cibles chez cette sous-population. Une triche est alors promue par Rap. Cette triche s'avère toutefois temporaire car à haute densité de la sous-population reflétée par une haute concentration de Phr, Phr est importé par les cellules et séquestre Rap, levant alors l'inhibition de ComA-P et donc, de production de biens communs. L'intérêt évolutif du caractère facultatif de cette triche réside sans doute dans le fait que lorsqu'une sous-population codant un système Rap-Phr atteint une densité importante et devient prédominante localement, la non production de biens communs pourrait lui être davantage défavorable que profitable en terme de fitness (Pollak *et al.*, 2016; Kalamara *et al.*, 2018).

Les architectures de QS présentées ci-avant, bien que faisant appel à des mécanismes différents, résultent en une même porte logique « AND » : $\text{if}(\text{quorum}_{\text{intraspecific}} \text{ AND } \text{quorum}_{\text{interspecific}})$; then initiate response. Elles représentent alors un parfait exemple d'évolution convergente de l'intrication de la transduction de deux signaux aux spécificités différentes comme base mécanistique de la contextualisation d'une densité de population homogène par rapport à celle d'une population hétérogène. À noter que d'autres types de portes logiques pourraient théoriquement être implémentées par des architectures alternatives de circuits de QS aux spécificités différentes. En l'occurrence la porte $\text{if}(\text{quorum}_{\text{interspecific}} \text{ AND NOT } \text{quorum}_{\text{intraspecific}})$ pourrait s'avérer particulièrement utile pour maximiser les chances de survie d'un microorganisme dont la population de parents se trouve en minorité au sein d'une communauté.

Hiérarchie de QSSs intra-spécifiques dont les signaux s'accumulent différemment : adaptation à de multiples seuils de densité de population

Outre la combinaison de QSSs aux spécificités différentes au sein d'un même génome, la multiplicité de QSSs distincts de même spécificité dans un génome donné est assez conservée dans le vivant (Papenfort and Bassler, 2016). Quels avantages peuvent alors bien représenter la possession de multiples QSSs intraspécifiques ? Un début de réponse vient du fait qu'un QSS unique ne permet de mesurer une densité de population que de façon binaire : en dessous ou au dessus du quorum. Face à cette limite, plusieurs QSSs dont les signaux respectifs s'accumulent différemment dans l'environnement peuvent alors offrir un recours efficace (bien que coûteux) pour permettre la prise de plusieurs mesures discrètes de la densité de population, mesures correspondant aux différents quorums reflétés par le seuil de concentration du signal associé à chaque QSS. La combinaison de QSSs distincts est particulièrement étudiée chez l'organisme modèle *Pseudomonas aeruginosa* et révèle une hiérarchie entre ses différents QSSs, la réponse

transcriptionnelle à un quorum supérieur régulant l'expression du QSS correspondant à un quorum inférieur (**Figure 16**). Cette hiérarchie permet à *P. aeruginosa* d'estimer finement son environnement social et physico-chimique et d'adopter en conséquence la stratégie évolutive qui lui sera la plus avantageuse (Cornforth *et al.*, 2014; Hawver, Jung and Ng, 2016; Papenfort and Bassler, 2016). Par ailleurs, la possible inhibition systématique d'une réponse à un quorum inférieur par la réponse à un quorum supérieur pourrait permettre à un micro-organisme d'orchestrer une sorte de programme développemental multicellulaire en fonction des différents régimes de sa densité de la population (Mehta *et al.*, 2009).

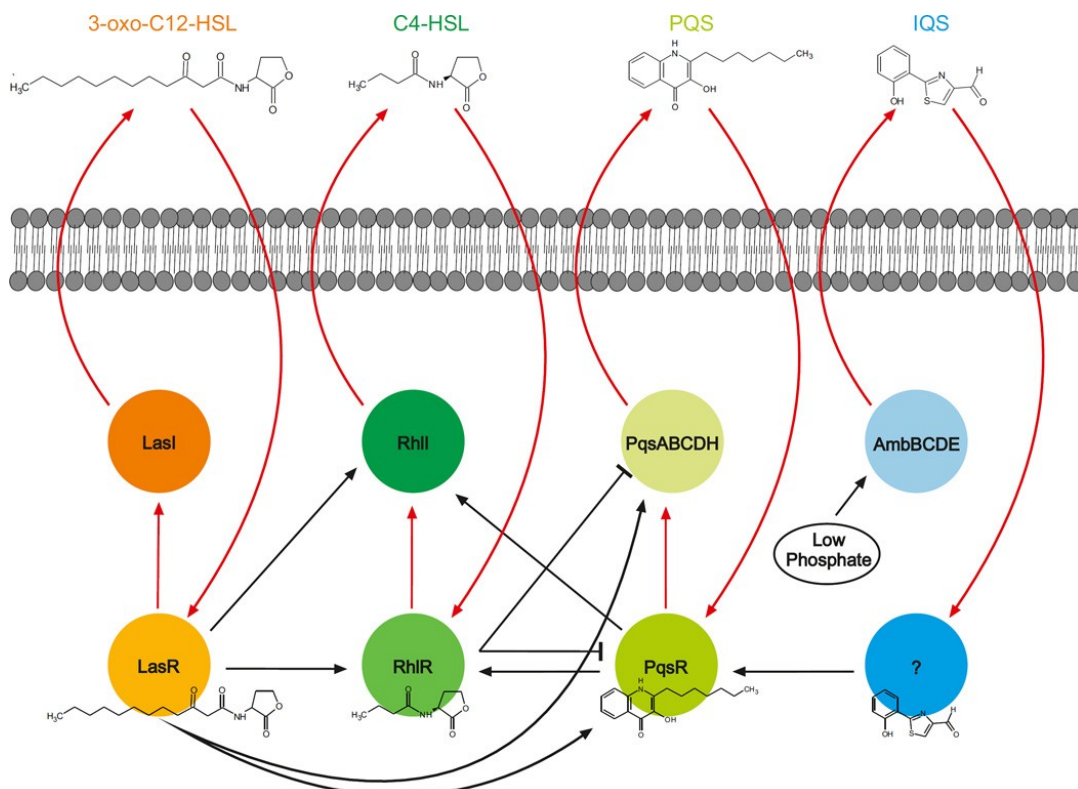


Figure 16: Hiérarchie des QSSs intraspécifiques chez *P. aeruginosa*. Les flèches rouges décrivent les autoinductions des différents signaux de QS [Source : (Papenfort and Bassler, 2016)]

2.5. Lien entre similarité de séquences entre protéines de QS et probabilité d'émission ou de reconnaissance d'une molécule de QS en commun

Les espèces produisant une molécule de QS en commun peuvent être considérées comme « parlant le même dialecte » et donc comme capables d'inter-influences. En guise d'illustration, la distribution taxonomique des différents « dialectes » d'AHLs est représentée en **figure 17** (Rajput, Kaur and Kumar, 2016). Une question émerge alors de ces observations : peut-on inférer quelles espèces sont susceptibles de parler le même dialecte ? Cette question amène à se demander s'il existe un lien entre similarité de séquences parmi des synthases/récepteurs codés par des génomes différents et probabilité de production/reconnaissance d'une molécule en commun ? Par ailleurs, puisqu'il existe des signaux plus ou moins spécifiques, une autre question sous-jacente consisterait à se demander si ce lien est variable, voire valide, d'une famille de gènes à l'autre. Les réponses à ces questions apparaissent essentielles pour comparer des gènes de QS entre génomes en vue d'inférer de potentiels liens de communication.

Avant de répondre à ces questions, il apparaît important de préciser qu'un lien de communication peut toutefois exister entre deux génomes sans production commune d'une même molécule de QS. En effet, cette inter-influence peut aussi émerger de différents dialectes d'une même langue. La probabilité que deux « dialectes » différents soient inter-fonctionnels entre deux espèces distinctes sera alors d'autant plus grande que ces « dialectes » se ressembleront d'un point de vue structural. Cette inter-fonctionnalité peut toutefois ne pas être strictement équivalente car un récepteur a souvent une affinité préférentielle pour un signal en particulier, et des affinités relatives pour d'autres signaux analogues (Schaefer *et al.*, 1996). Cependant, une étude récente de systématique sur la sélectivité des récepteurs à l'AHL vis à vis de différentes sortes d'AHLs a pu démontrer que la majorité de ces récepteurs peuvent répondre avec sélectivité et intensité à au moins une AHL étrangère (non produite par l'espèce codant le récepteur), suggérant alors un potentiel important d'inter-influence de ces systèmes de QS (Wellington and Peter Greenberg, 2019). Au regard de ces observations, la question pourrait être reformulée ainsi : existe-t-il un lien entre similarité de séquences parmi des synthèses/récepteurs codés par des génomes différents et probabilité de production/reconnaissance d'une molécule exogène, identique ou analogue à une molécule endogène.

Pour tenter de répondre à cette question, il est intéressant de noter en premier lieu que dans certains cas, une famille de signaux / « langue » donnée n'est pas nécessairement synthétisée par une seule famille de gènes émetteurs. Par exemple, on recense actuellement 3 familles génétiques différentes d'AHL synthèses : *luxI*, *ainS* et *hdts* (Waters and Bassler, 2005) (**Table 1**). D'autre part, un nombre important de familles différentes de récepteurs peuvent être associés à une même langue. La grande diversité des récepteurs de l'AI-2 en offre une parfaite illustration (Pereira *et al.*, 2009; Anderson *et al.*, 2015; Torcato *et al.*, 2019; Zhang *et al.*, 2020). Malgré tout, un lien pourrait toutefois exister entre similarité de séquence des protéines à l'intérieur d'une même famille de synthèses/récepteurs et similarité de structures des molécules produites/reconnues par ces différentes synthèses/récepteurs. Le cas échéant, une famille de gènes de QS pourrait théoriquement être partitionnée en sous-familles représentatives de différents dialectes moléculaires.

Un argument à l'encontre de cette partition est qu'une synthase identique (par exemple clonée) peut ne pas toujours produire le même signal (ou la même gamme de signaux) d'une espèce à l'autre car les précurseurs moléculaires disponibles pour cette synthase peuvent être espèce-dépendants (Churchill and Chen, 2011). C'est notamment le cas pour les AHL synthèses de la famille LuxI qui nécessitent deux substrats en entrée, une SAM (S-adénosylméthionine) à la structure invariable et une ACP (protéine porteuse d'acyl) ou une Acyl-CoA (acyl coenzyme A) dont les chaînes carbonées peuvent justement être variables, espèce-dépendantes et déterminer la structure finale de l'AHL biosynthétisée. Son contre-argument est que les synthèses de signaux inter-espèces incorporent toujours le même substrat unique en entrée et produisent généralement la même molécule de QS en dépit de divergences génétiques importantes.

Une tentative de répondre à cette question sur des bases moins qualitatives et plus statistiques pourrait émerger d'analyses de données quantitatives liant pour chaque séquence d'une famille de synthases ou de récepteurs de QS donnée, la gamme de molécules synthétisée ou reconnue par cette séquence. Avantageusement, la base de données Sigmol (Rajput, Kaur and Kumar, 2016) associe, pour deux langues de signaux (les AHLs et l'AI-2), un nombre important de séquences protéiques des familles LuxI (synthase de l'AHL), LuxR (récepteur de l'AHL), LuxS (synthase de l'AHL) et LsrB (récepteur de l'AHL) aux structures moléculaires des signaux produits/reconnus par ces séquences. Puisque les AHLs sont des signaux plutôt intra-spécifiques tandis que l'AI-2 est un signal inter-spécifique par excellence, les liens entre similarité de séquences au sein des différentes familles susnommées et probabilité de synthèse/reconnaissance d'une molécule en commun peuvent alors être explorés sous le prisme de la spécificité relative de ces deux langues (**Figure 18**).

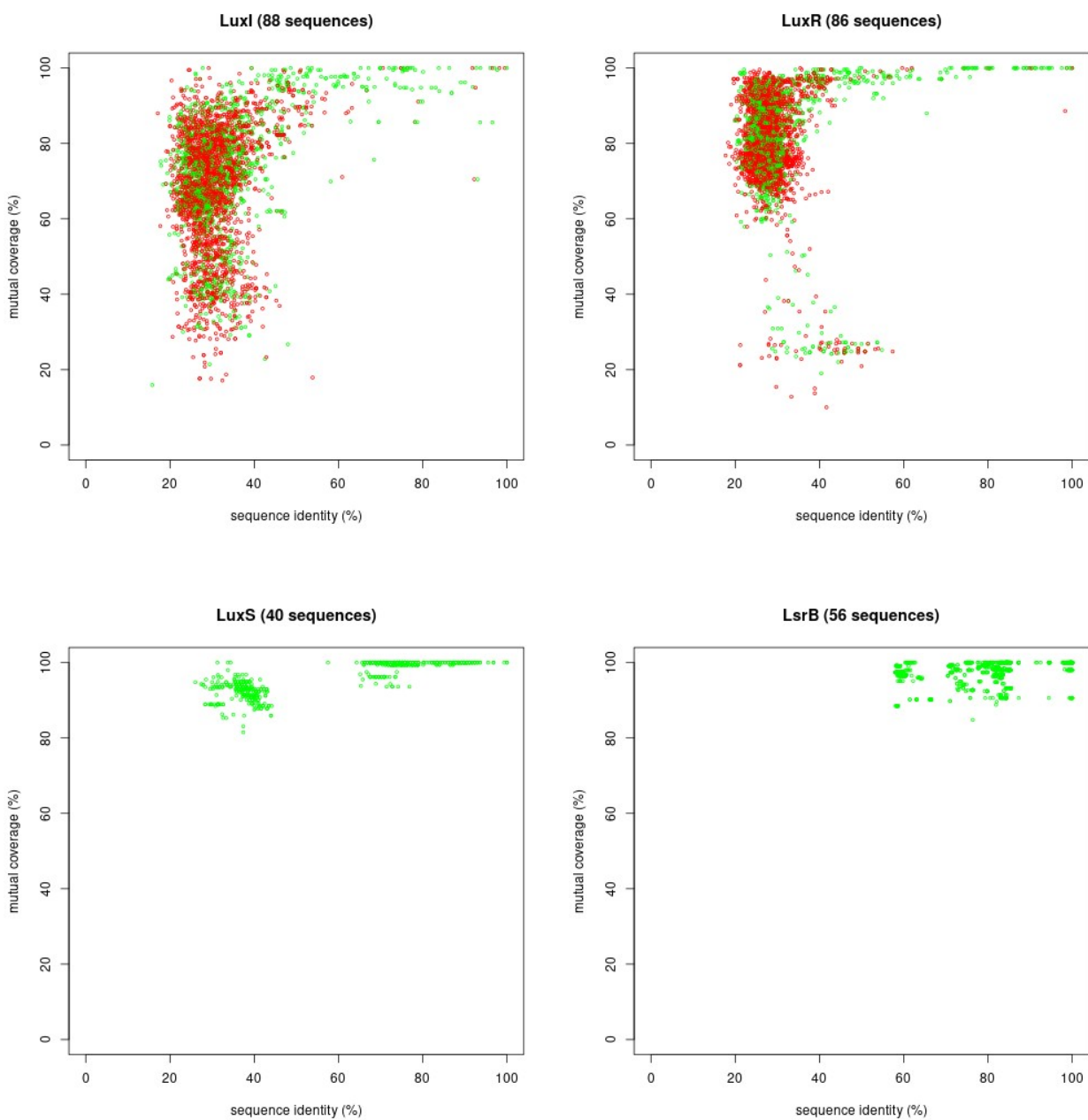


Figure 18: Scatterplots de la couverture mutuelle (%) et de l'identité de séquence (%) entre chaque paire de séquences au sein des familles LuxI (synthases d' AHLs), LuxR (récepteurs d' AHLs), LuxS (synthases de l' AI-2) et LsrB (une des familles de récepteurs de l' AI-2). Les paires en rouge correspondent à des synthases/récepteurs ne produisant/reconnaissant aucune molécule en commun, celles en vert à des synthases/récepteurs produisant/reconnaissant au moins une molécule en commun selon la littérature. Pour les familles LuxI et LuxR, des seuils d'identité > 50% et de couverture mutuelle > 80% permettrait de maximiser la différence « nombre de vrai positifs – nombre de faux positifs ». Dans la famille LuxS, associée au signal inter-espèce AI-2, les synthases de séquences peuvent être assez divergentes entre elles (< 40 % identité) et produire tout de même le même signal [Source : cette thèse, données de (Rajput, Kaur and Kumar, 2016)]


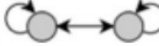


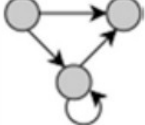
La **figure 18** révèle que les seuils d'identité de séquence et de couverture mutuelle (sur toute la longueur ou sur certaines régions clés) entre protéines de QS homologues sur lesquels se reposer pour définir ces liens de communication seraient variables d'une famille de protéines à l'autre.

De façon générale, il y a premièrement différents modes de production de signaux de QS (e.g. peptides maturés à partir de propeptides et petites molécules synthétisées par des synthases à partir de précurseurs) et les propeptides doivent être quasiment identiques dans leur partie C-terminale codant le peptide mature quand les enzymes pourraient bénéficier de seuils moins stringents. Deuxièmement, une forte similarité de séquences entre deux récepteurs de QS peut être biaisée par des aminoacides non impliqués dans la réception du signal (e.g. aminoacides impliqués dans l'incorporation à la membrane pour des récepteurs membranaires, aminoacides impliqués dans la liaison à l'ADN pour des récepteurs intracellulaires ...). Enfin, il existe des signaux plus ou moins spécifiques ; les synthases de signaux inter-spécifiques peuvent être assez divergentes et produire tout de même le même signal alors que des synthases de signaux intra-spécifiques peuvent être assez similaires sans produire de molécules en commun. Il est donc primordial d'identifier pour chaque protéine de QS, les régions protéiques et les résidus clés liés directement à la production ou à la reconnaissance du signal. Dans le meilleur des cas, ces régions correspondent à un enchaînement linéaire d'acides aminés consécutifs mais malheureusement, ces résidus clés peuvent également être localisés à différentes positions sur la protéine. De ces observations naissent un constat, il paraît essentiel d'obtenir des données liant séquence et gamme de molécules produites ou reconnues par cette séquence pour chaque famille de protéine de QS afin d'identifier les résidus déterminants dans la sélectivité des protéines de QS envers leurs signaux associés (Wellington Miranda *et al.*, 2021) et de justifier des seuils d'identité entre régions de protéines de QS homologues sur des bases statistiques solides. Cette définition de seuils d'identité pour aussi valoir dans le cadre d'alignement de séquences 3D prédites par des programmes tels qu'AlphaFold (Jumper *et al.*, 2021) pour estimer jusqu'à quel point des protéines d'une même famille doivent se ressembler structurellement pour être susceptibles de produire/reconnaître une molécule de QS en commun. Enfin, et de façon moins supervisée, une sous-partie de ces données pourraient permettre d'entraîner des modèles d'apprentissage profond à inférer des probabilités de production/reconnaissance de structures moléculaires associées à chaque synthase/récepteurs tandis que l'autre sous-partie serait dédié au test de ces modèles. Les synthases/récepteurs aux probabilités importantes de production/reconnaissance d'une molécule en commun pour alors permettre d'inférer des liens de communication. Cependant, ces données ne sont aujourd'hui disponibles que pour très peu de familles de protéines de QS. Au regard de toutes ces subtilités et prérequis méthodologiques, l'inférence de réseaux de communication apparaît alors comme un problème scientifique beaucoup moins trivial qu'il n'y paraît, bien que théoriquement réalisable.

2.6. Modes de bio-communication : entre communication, manipulation et espionnage

De façon intéressante, un lien de communication peut émerger d'autres cas de figure que celui d'un lien d'homologie entre des QSSs complets codés par deux génomes distincts, c'est à dire lorsque non seulement les deux récepteurs sont homologues mais que les deux synthases le sont aussi. En effet, un récepteur seul peut permettre de s'adapter à des populations strictement étrangères quand un module émetteur peut permettre de manipuler la réponse au quorum d'une autre population au bénéfice du manipulateur. Ces différentes modalités de communication sont dictées par la combinatoire des modules de gènes de QS entre deux génomes distincts (**Table 3**). Ces notions sont importantes pour l'orientation éventuelle de réseaux hypothétiques de communication, comme illustré dans la **table 3**.

Table 3 : Combinatoire de gènes de QS et modalités de communication. Le nœud 1 correspond à l'entité microbienne hébergeant le génome 1 et le nœud 2 à celle hébergeant le génome 2. Une flèche partant d'un nœud correspond à une émission d'un signal de QS, et une flèche touchant un nœud correspond à une réception d'un signal de QS [Source : cette thèse]

| Communication modality | Node 1 | Node 2 | Genome 1 | Genome 2 |
|--|---|--------|-----------------------------------|---|
| Canonical QS (intra-entity communication) |  | | – QS synthase(s) – QS receptor | |
| Inter-entity communication |  | | – QS synthase(s) – QS receptor | – QS synthase(s) homolog – QS receptor homolog |
| Eavesdropping |  | | – QS receptor | – QS synthase(s) – QS receptor homolog |
| Signal amplification |  | | – QS synthase(s) | – QS synthase(s) homolog – QS receptor |
| Unilateral signaling by functional complementation |  | | – QS synthase(s) | – QS receptor |

Comme la communication intraspécifique (entre parents) et interspécifique (avec des parents comme avec d'autres entités microbiennes) correspondant aux deux premières lignes du tableau a largement été abordée précédemment, seules les autres modalités de communication seront abordées à l'aide d'un exemple représentatif dans les lignes suivantes.

Exemple d'espionnage

Un exemple d'espionnage fascinant met en scène *V. cholerae* et *Vibrio* phage 882 (Silpe and Bassler, 2019a). Ce dernier code un homologue du récepteur intracellulaire VqmA de son hôte. Chez *V. cholerae*, le signal de QS reconnu par VqmA est le DPO, et la transduction du DPO par VqmA à haute densité cellulaire déclenche une dissolution des biofilms, à l'instar des signaux AI-2 et CAI-1 évoqués précédemment (Papenfort *et al.*, 2017b). Lors de l'infection de *V. cholerae* par *Vibrio* phage 882, le cycle lytique du phage est réprimé à faible concentration de DPO, et donc à faible densité d'hôtes. Mais lorsque le quorum d'hôtes reflété par une forte concentration de DPO est atteint, le complexe moléculaire DPO-VqmA_{phage} initie le cycle lytique de VP882 (**Figure 19**). La survie du phage étant dépendante de celle de la population d'hôtes, ce cas d'espionnage du signal de QS de l'hôte permet d'optimiser la fitness du virus en évitant le scénario catastrophe d'un cycle lytique déclenché à

faible densité d'hôtes. De façon intéressante, alors que $VqmA_{\text{phage}}$ peut réguler chez *V. cholerae* le régulon cible de $VqmA$, la réciproque n'est pas vraie.

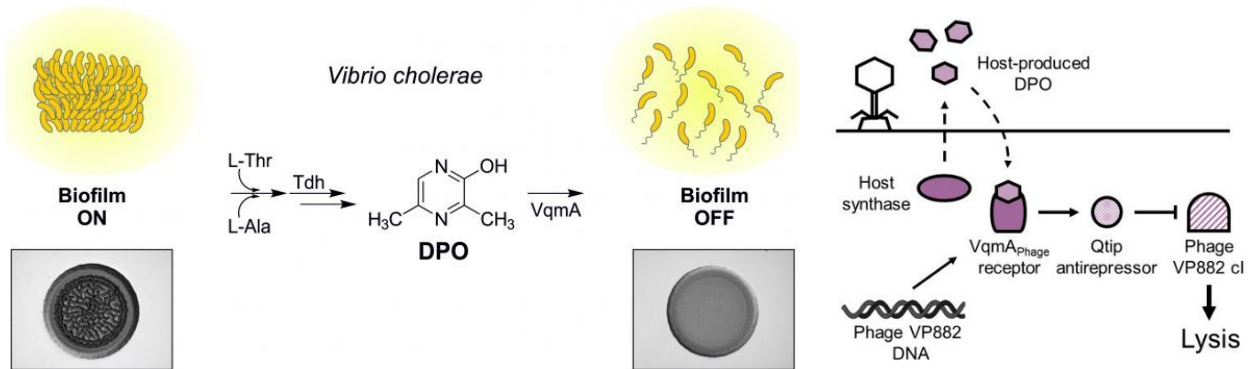


Figure 19: Espionnage du signal de QS de l'hôte par *Vibrio* phage VP882 [Source : (Silpe and Bassler, 2019a)]

Selon une perspective plus large, une étude bioinformatique à grande échelle sur le récepteur de QS LuxR a révélé que 75% des *luxR* détectés n'étaient pas trouvés adjacents à un gène *luxI* codant une AHL synthase, suggérant alors une possible prévalence de l'espionnage de molécules de QS exogènes dans le vivant (Hudaiberdiev *et al.*, 2015). Pour faire le lien avec l'exemple précédent, un *luxR* orphelin a d'ailleurs été trouvé dans un génome de phage et fut identifié comme capable de détecter l'AHL synthétisée par son hôte (Silpe and Bassler, 2019b) .

Exemple d'amplification du signal

Un exemple de manipulation fait intervenir cette fois *Ruminococcus obeum* et *V. cholerae*. Dans une étude menée au Bangladesh sur le choléra, Hsiao *et al.* ont identifié un enrichissement de la bactérie *R. obeum* dans le microbiote intestinal de patients ayant guéri du choléra (Hsiao *et al.*, 2014). En investiguant les possibles causes de l'antagonisme soupçonné de *V. cholerae* par *R. obeum*, Hsiao *et al.* ont identifié un processus de coercition lié au QS. Comme évoqué précédemment, une haute concentration d'AI-2 déclenche chez *V. cholerae* une dissolution des biofilms et une répression de la virulence, deux facteurs importants de colonisation. Et de façon fascinante, en présence de *V. cholerae*, le gène *luxS* de *R. obeum* est surexprimé, résultant en une surproduction d'AI-2 par rapport aux conditions de monoculture de *R. obeum* (**Figure 20**). Comme prédit par la **table 3**, ce cas de figure correspond à l'amplification du signal de QS d'une espèce manipulée par une espèce manipulatrice et résulte en une précipitation de la réponse au quorum de la première au bénéfice de la seconde. En amplifiant le signal de QS de répression des facteurs de colonisation de *V. cholerae*, *R. obeum* protège en effet sa niche écologique de l'invasion du pathogène. D'un point de vue

anthropocentrique, la bactérie commensale *R. obeum* peut ainsi être considérée comme un garde-fou vis à vis de pathogènes entéro-hémorragiques.

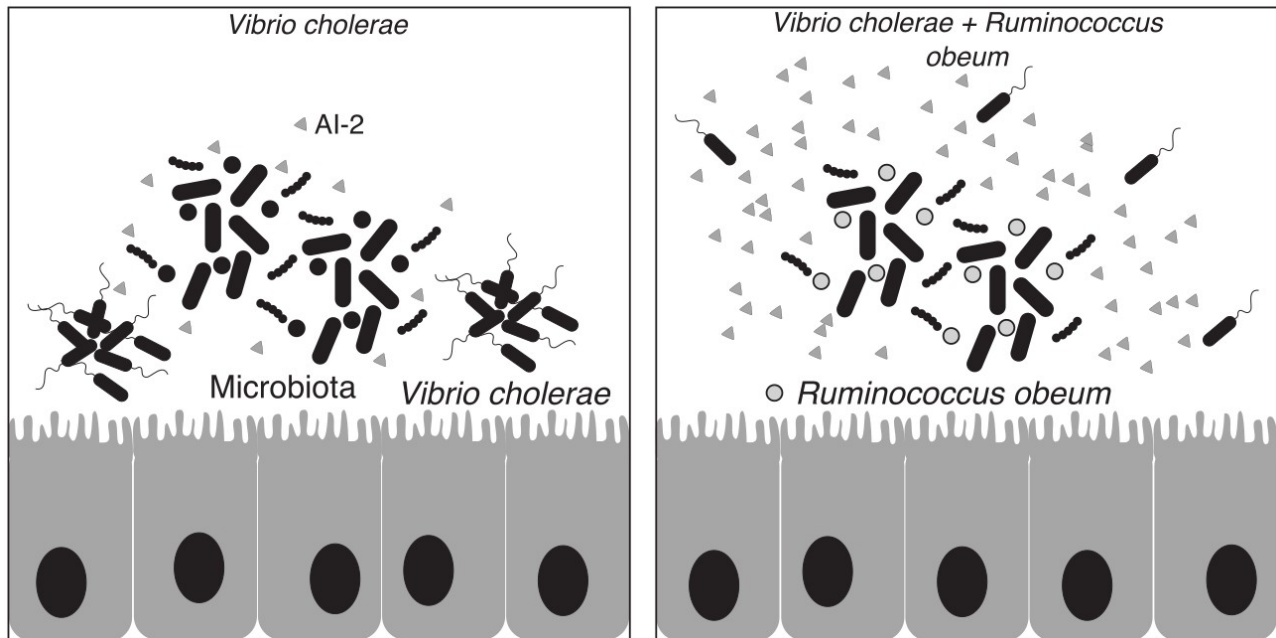


Figure 20: Amplification du signal de QS de *V. cholerae* par *R. obeum*. Le signal AI-2 réprime la virulence et l'association en biofilm chez *V. cholerae*. En présence de *V. cholerae*, *R. obeum* surproduit de l'AI-2 et limite ainsi la colonisation de *V. cholerae* [Sources : (Hsiao et al., 2014; Jimenez and Sperandio, 2019)]

Signalisation unilatérale par complémentation fonctionnelle

Cette modalité de communication via des gènes strictement de QS n'a pas, à notre connaissance, encore été identifiée à ce jour dans un contexte naturel, malgré un cas s'en rapprochant (Pietschke et al., 2017). Ce mode de signalisation a toutefois été implémentée par biologie synthétique dans une étude de biologie des systèmes (Youk and Lim, 2014). Dans la nature, un tel scénario évolutif de signalisation unilatérale par complémentation fonctionnelle de gènes de QS ne serait sûrement crédible que dans un contexte d'interaction obligatoire entre deux espèces (e.g. un endosymbionte et son hôte).

2.7. Coévolution bactérie-eucaryote et signalisation inter-règles

La section précédente a introduit les notions d'espionnage et de coercition par l'intermédiaire de signaux de QS. Ces cas de hacking du circuit de QS d'une tierce espèce avaient néanmoins tous en commun d'être supportés par un gène de QS (un récepteur ou une synthase homologue à celui ou celle de l'espèce hackée). Or, comme évoqué en **section 2.3** sur les déterminants de la spécificité

des signaux de QS, les molécules de QS les plus simples structurellement imposent peu de contraintes physico-chimiques à leur fixation par une tierce protéine. Ainsi, certaines molécules de QS peuvent être fixées par des protéines non codées par des gènes de QS. À l'inverse, des analogues structurels exogènes d'une molécule de QS peuvent être reconnus par un récepteur de QS. Certains de ces analogues peuvent alors être synthétisés par des protéines non affiliées au processus de QS. Ces propriétés posent les bases de possibles signalisations moléculaires entre deux entités dont l'une ne code pas de gènes de QS. De ces propriétés peuvent alors émerger un phénomène fascinant : l'inter-influence des microbes et des organismes eucaryotes multicellulaires par l'intermédiaire de signaux moléculaires. Ce phénomène, appelé signalisation inter-règnes, sera présenté dans les lignes suivantes dans le contexte des coévolutions homme-bactéries, plante-bactéries, et nématodes-bactéries.

Signalisation homme-bactéries

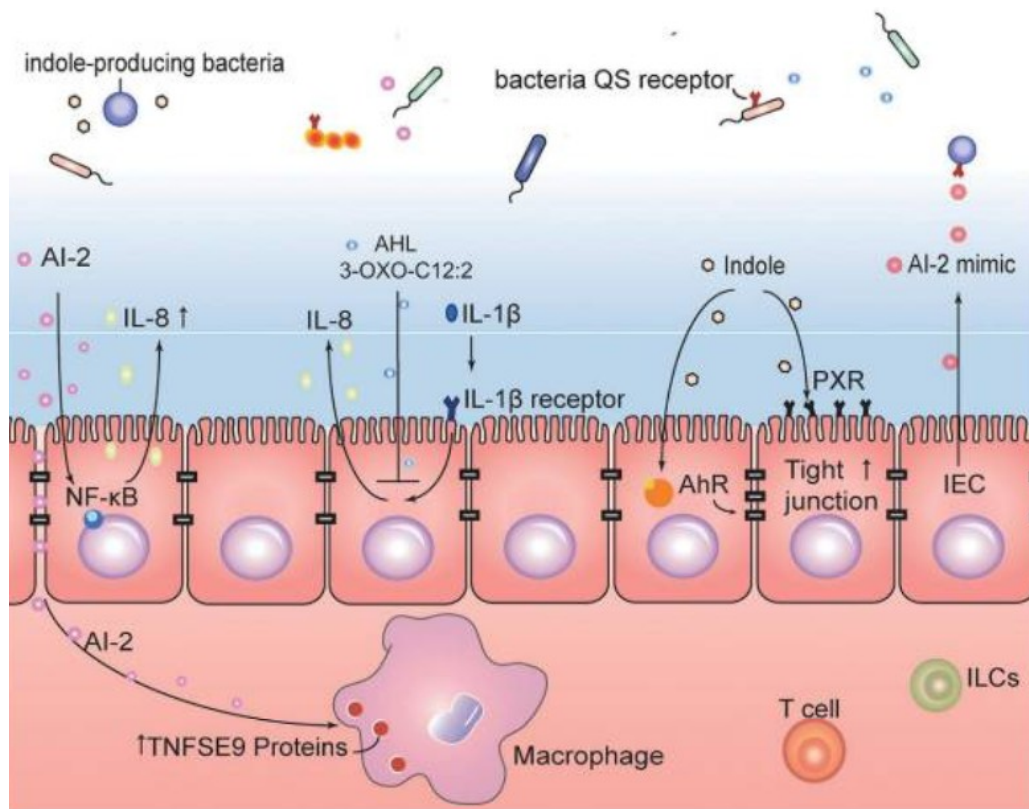


Figure 21: Aperçu non exhaustif de la signalisation homme-bactéries. Les signaux de QS bactériens que sont l'AI-2, l'AHL 3-oxo-12:2 et l'Indole peuvent influencer sur la biologie de l'homme (e.g. immunomodulation, barrière épithéliale) et l'ersatz de l'AI-2 synthétisé par les cellules épithéliales est reconnu par les récepteurs à l'AI-2 des bactéries [Source : (Wu and Luo, 2021)]

Molécules signalétiques humaines interprétées par des récepteurs de QS microbiens

Un premier exemple fascinant fait intervenir la synthèse d'un analogue de l'AI-2 par les cellules épithéliales de l'intestin humain (**Figure 21**) (Ismail, Valastyan and Bassler, 2016). Cette synthèse survient en réponse à la disruption des jonctions épithéliales ou en réponse à la présence de bactéries. Il fut démontré que l'analogue humain de l'AI-2 active la réponse bactérienne au QS via les récepteurs à l'AI-2 des familles LuxP ou LsrB. Le fait que cet ersatz de l'AI-2 soit notamment sécrété quand l'épithélium est exposé directement ou indirectement à des populations bactériennes suggère que sa synthèse est stimulée par une molécule émise par un pathogène bactérien. Par exemple, il fut démontré que l'aérolysine, une toxine connue pour rompre les jonctions épithéliales, suffit à induire la production de cet ersatz d'AI-2. Comme l'analogue humain de l'AI-2 est sécrété dans la lumière intestinale et comme les bactéries commensales sont préférentiellement associées à l'apex des cellules épithéliales quand les pathogènes ont un tropisme par leur partie baso-latérale, il fut proposé que ce mécanisme permette à l'intestin de recruter l'assistance de bactéries commensales dans la lutte contre l'invasion des pathogènes.

Un autre exemple concerne cette fois-ci l'exploitation d'hormones de stress humaines par une souche enterohémorragique d'*E. coli* (Kim *et al.*, 2020). Chez ce pathogène, la virulence est activée par le système à deux composantes QseC-QseB suite à la reconnaissance d'une haute concentration de la molécule endogène de QS AI-3. Mais de façon intéressante, ce senseur déclenche aussi la virulence de la bactérie suite à sa liaison à l'épinéphrine et la norépinéphrine, deux hormones de stress humaines structurellement proches de l'AI-3 (**Figure 22**). L'implication de l'espionnage de ces hormones de stress dans la régulation de processus bactériens de virulence pourrait expliquer en partie le lien entre stress et susceptibilité aux infections chez l'homme.

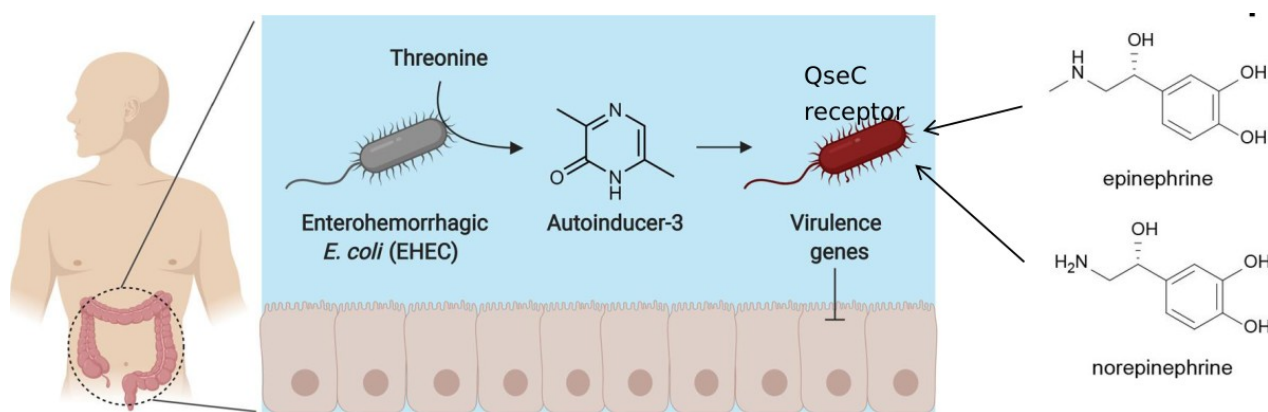


Figure 22 : Espionnage d'hormones de stress humaines par une souche enterohémorragique d'*Escherichia coli* [Source : (Kim *et al.*, 2020)]

Signaux de QS microbiens interagissant avec des protéines humaines

Comme montrée en figure 21, des processus d'immuno-modulation chez l'humain furent directement imputés à certaines molécules de QS affiliées aux langues des AHLs et de l'AI-2 (Zargar *et al.*, 2015; Landman *et al.*, 2018; Li *et al.*, 2019). Par ailleurs, il fut démontré que l'indole peut être reconnu par le récepteur AHR et se lier au récepteur PXR des cellules épithéliales intestinales (Hubbard, Murray and Perdew, 2015). Cette détection de l'indole serait impliquée dans le resserrement des jonctions épithéliales et dans la régulation de la différenciation de cellules souches des cryptes intestinales en cellules épithéliales (**Figure 21**) (Shimada *et al.*, 2013; Venkatesh *et al.*, 2014). Enfin, selon des perspectives moins optimistes, il fut reporté que la molécule de QS cFP du pathogène *Vibrio vulnificus* inhibe le récepteur humain RIG-1 impliqué dans la réponse immunitaire innée antivirale. Ce mécanisme suggère alors une possible promotion d'un collectif infectieux *V. vulnificus* – virus réalisée par la molécule de QS de *Vibrio vulnificus* au bénéfice de la facilitation de l'invasion de ce pathogène (Lee *et al.*, 2018).

Signalisation plante-bactéries

Molécules signalétiques végétales interprétées par des récepteurs de QS microbiens

Un corpus d'études a mis en évidence qu'un nombre important de bactéries associées aux plantes (genres *Rhizobia*, *Pseudomonas* et *Xanthomonas*) codent un LuxR orphelin (sans LuxI associé) dont la partie N-terminal de liaison au signal ne présente pas les résidus essentiels à la reconnaissance des AHLs. Ces études ont démontré que ces LuxR orphelins sont impliqués dans l'espionnage de molécules végétales et régulent des processus bactériens d'importance capitale dans l'interaction plante-bactéries : de la virulence à l'adhésion avec la plante chez les pathogènes et de la nodulation à la production d'antimicrobiens ou de chitinase chez les symbiotes (González and Venturi, 2013). Chez l'endophyte racinaire du peuplier *Pseudomonas* sp. GM79, Coutinho *et al.* ont même pu identifier que la molécule végétale espionnée par la bactérie est une N-(2-hydroxyethyl)-2-(2-hydroxyethylamino) acetamide, un dérivé de l'éthanolamine (Coutinho *et al.*, 2018).

Signalisation nématode-bactéries

Signaux de QS microbiens interagissant avec des protéines de C. elegans

Les exemples précédents de signalisation inter-règnes faisaient intervenir de petites molécules de QS. Dans le contexte d'une co-évolution *C. elegans* – *B. subtilis*, une exception est à noter. Malgré une complexité structurale importante, le peptide mature PhrC (ERGMT) de *B. subtilis* induit chez son hôte une activation des voies métaboliques de survie et de longévité, participant ainsi à un allongement de la durée de vie de *C. elegans* (Donato *et al.*, 2017). L'action pro-longévité du peptide chez *C. elegans* n'est pas indirecte (e.g via à un processus de *B. subtilis* favorisant la longévité de l'hôte qui serait déclenché en réponse à ce peptide) mais bien directe car l'administration seule du

peptide suffit à accroître la longévité de *C. elegans*. Ce peptide est en effet importé par les cellules épithéliales de l'intestin de *C. elegans* via la protéine OCTN2, un transporteur membranaire de cations présent à la surface apicale de ces cellules. Cet exemple fait par ailleurs le lien avec un projet auxiliaire de cette thèse : l'identification de distorceurs d'âges, soit tout virus, microorganisme ou organisme multicellulaire modulant à son avantage les processus de vieillissement et de longévité de son hôte (**Section 10**).

La signalisation nématode-bactérie peut aussi concerner le chimiotactisme de *C. elegans* vers des molécules de QS. Le chimiotactisme est un processus essentiel pour localiser la nourriture, trouver des hôtes et éviter les prédateurs. Dans ce contexte, Werner et al. ont identifié que la molécule de QS CAI-1 de *V. cholerae* influence le chimiotactisme du nématode puisque ce dernier peine à chasser les mutants de *V. cholerae* déficients pour la production de CAI-1. Werner et al. ont alors démontré que la détection du CAI-1 est réalisée par le neurone aphide sensoriel AWC^{on} de *C. elegans* (Werner et al., 2014).

3. Détection de variants de gènes de QS dont les signaux associés sont de petites molécules de QS par des méthodes classiques de recherche d'homologues

Ce chapitre est dédié à la détection de QSSs dont le signal est synthétisé par un ou plusieurs enzymes de synthèse. Les QSSs dont le signal associé est de nature peptidique seront abordés dans le chapitre suivant. La recherche d'homologues de QSSs déjà validés expérimentalement peut permettre d'étendre les connaissances acquises sur le QS à des séquences génétiques traditionnellement délaissées par ce champ de recherche. De ces propagations de connaissance peuvent émerger des intuitions nouvelles sur les interactions sociales et les régulations densité-dépendantes d'entités microbiennes méconnues.

De façon assez instinctive, une manière d'identifier des QSSs au sein de génomes d'intérêt est de détecter parmi leur protéome prédit/annoté, des homologues de protéines codées par les gènes émetteurs et récepteurs de QSSs déjà connus et validés expérimentalement.

Les alignements deux à deux d'une séquence cible avec une séquence de QS de référence offrent alors une première possibilité pour identifier des homologues de protéines de QS. L'outil le plus utilisé pour ce genre de recherche d'homologues est BLASTp pour « Basic Local Alignment Sequence Tool », car ce programme est reconnu pour sa spécificité (paramètre minimisant les faux positifs) et son heuristique assez rapide et sensible (paramètre maximisant les détections) (Altschul *et al.*, 1990). Néanmoins, des alternatives moins sensibles mais beaucoup plus rapides telles que DIAMOND (Buchfink, Xie and Huson, 2015) ont été développées pour tirer leur épingle du jeu dans la recherche d'homologues au sein d'énormes jeux de données telles que la base de données NR (la base de données de protéines non-redondantes du NCBI (NCBI Resource Coordinators, 2016) ~500 millions de séquences uniques) ou MGnify (~300 millions de séquences représentatives de clusters à 90% d'identité) (Mitchell *et al.*, 2020) pour lesquelles BLASTp devient typiquement limitant au delà du millier de requêtes. Dans une recherche BLASTp ou DIAMOND, les séquences de chaque famille de protéine de QS référencées **en section 2.2** peuvent alors servir d'autant de requêtes pour interroger un jeu de données cible sur la présence d'homologues de ces protéines. Pour s'assurer de la pertinence des séquences identifiées, des filtres peuvent ensuite être appliqués sur la sortie de BLASTp/DIAMOND pour ne garder que les hits les plus significatifs, c'est à dire ceux correspondant aux protéines cibles les plus susceptibles d'être des homologues des séquences utilisées comme requêtes.

Une des limitations des méthodes d'alignement deux à deux est que la requête pour interroger une base de données cible concerne une séquence protéique unique, définie par un seul acide aminé de référence par position. Or, dans une même famille de protéines, certaines régions d'acides aminés peuvent être plus ou moins sujettes à sélection, et donc plus ou moins variables. Une méthode aujourd'hui très plébiscitée pour prendre en compte la conservation relative des acides aminés à différentes positions au sein d'une famille de séquences de référence est la recherche de profils HMMs « Hidden Markov Models » (Eddy, 2011a). Un modèle HMM est construit à partir d'un alignement multiple de séquences de référence et est sollicité pour calculer la probabilité qu'une séquence cible ait été générée par le modèle (**Figure 23**). Plus cette probabilité est grande, plus cette séquence aura de chances d'appartenir à la famille de protéines constitutives de l'alignement multiple de départ. Un HMM est décrit par des probabilités d'émission, c'est à dire des probabilités d'observer un acide aminé pour une position donnée et des probabilités de transition, c'est à dire des probabilités de passer d'une position à une autre. Contrairement aux matrices de position de type PSI-BLAST (Altschul *et al.*, 1997), les notions d'insertion et de délétion d'acides aminés sont donc prises en compte dans un HMM. Par ailleurs, en intégrant une gamme de possibilités d'acides aminés pour chaque position d'une séquence cible, la recherche de profils HMM exerce moins de contraintes que BLASTp (un seul acide aminé par position) pour la détection d'homologues. Cette sensibilité offre alors l'avantage de pouvoir identifier des homologues de séquences d'intérêt assez divergents. Des HMMs propres aux différentes familles de protéines de QS peuvent être construits grâce à l'outil *hmmbuild* du package HMMER (Eddy, 2011a) à partir de l'alignement multiple des séquences de chacune de ces familles, produit par exemple via *MUSCLE* (Edgar, 2004) ou *MAFFT* (Kuraku *et al.*, 2013) et éventuellement rogné avec *trimAl* (Capella-Gutierrez, Silla-Martinez and Gabaldon, 2009) pour ne considérer que certaines régions/domaines d'intérêt. Par ailleurs, et de façon très avantageuse, il existe des bases de données publiques recensant des HMMs de référence pour de nombreuses familles de protéines ou de régions protéiques (Wilson *et al.*, 2009; Haft *et al.*, 2013; Letunic, Khedkar and Bork, 2021; Mistry *et al.*, 2021; Sillitoe *et al.*, 2021). L'outil appelé *Interproscan* a alors été développé dans le but de pouvoir interroger ces bases de données avec une séquence protéique d'intérêt afin d'identifier les domaines de cette protéine décrits par des HMMs de référence (Jones *et al.*, 2014). L'application d'*Interproscan* aux différentes protéines de QS validées expérimentalement peut ainsi révéler des modèles HMM de référence associés, voire même idéalement spécifiques à ces protéines. Une fois identifiés, ces HMMs spécifiques peuvent alors être recherchés dans une base de données de protéines de cibles, par exemple via l'outil de référence *hmmsearch* du package HMMER, afin d'identifier des homologues de protéines de QS. Là, encore, la sortie de *hmmsearch* peut être filtrée pour ne garder que les protéines qui ont la plus grande probabilité d'avoir été générée par un HMM utilisé comme requête.

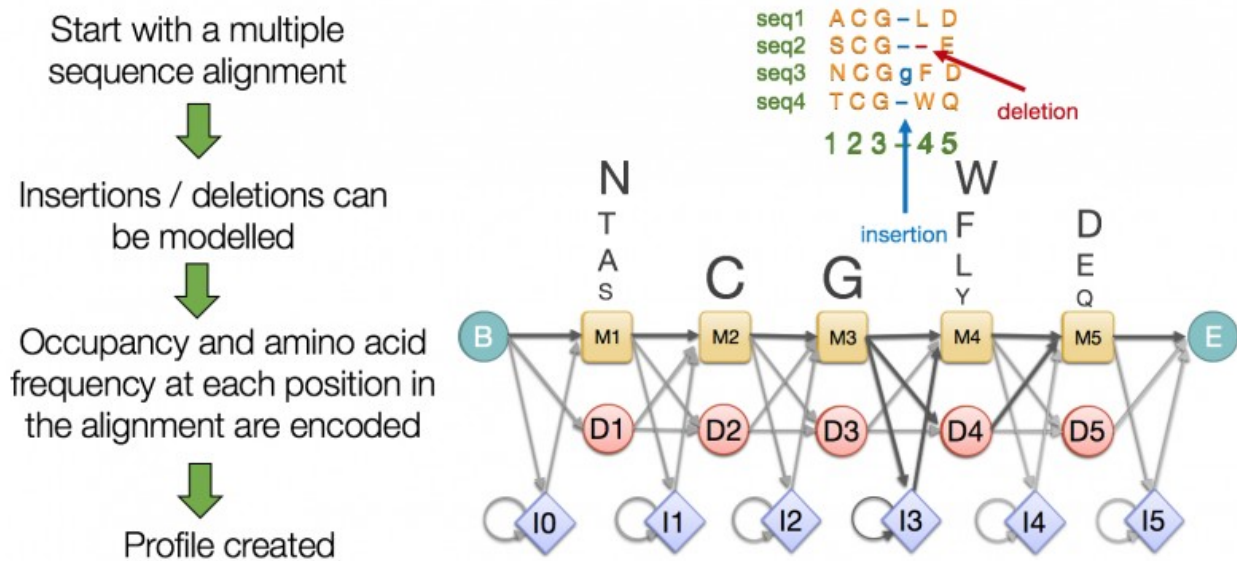


Figure 23 : Création d'un profile HMM à partir d'un alignement multiple de protéines [Source : (Eddy, 2011a)]

L'article suivant publié dans le journal *mSystems* fournit une illustration de la combinaison de BLASTP et hmmsearch pour la recherche d'homologues de protéines de QS au sein de métagénomés de CPRs et de DPANNs, deux lignées procaryotes récemment découvertes à la biologie aussi mystérieuse que fascinante (Castelle and Banfield, 2018; Castelle *et al.*, 2018). Puisque ces lignées sont très différentes des lignées procaryotes cultivées et connues auparavant, leur découverte a permis une expansion massive de la diversité décrite des organismes vivant actuellement sur Terre (Hug *et al.*, 2016). Néanmoins, peu de représentants de CPRs et de DPANNs sont cultivables en laboratoire et la biologie des membres de ces différentes lignées reste encore très peu connue à ce jour. Parce que la réduction de leur répertoire de gènes fut vraisemblablement une force importante de l'évolution de leur génome, les membres de ces lignées sont souvent décrits comme dépendants d'autres espèces pour survivre (Castelle *et al.*, 2018; Dombrowski *et al.*, 2019), ce qui pose la question de savoir comment ces CPRs et DPANNs interagissent avec d'autres espèces. Comme les signaux de QS sont d'importants médiateurs d'interactions entre espèces, cet article teste l'hypothèse selon laquelle les génomes de ces lignées puissent coder des gènes de QS facilitateurs d'interactions inter-espèces.



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

 Charles Bernard,^{a,b} Romain Lannes,^{a*} Yanyan Li,^b Éric Bapteste,^a  Philippe Lopez^a

^aInstitut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d'Histoire Naturelle, Paris, France

^bUnité Molécules de Communication et Adaptation des Micro-organismes (MCAM), CNRS, Museum National d'Histoire Naturelle, Paris, France

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

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

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

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

Citation Bernard C, Lannes R, Li Y, Bapteste É, Lopez P. 2020. Rich repertoire of quorum sensing protein coding sequences in CPR and DPANN associated with interspecies and interkingdom communication. *mSystems* 5:e00414-20. <https://doi.org/10.1128/mSystems.00414-20>.

Editor Robert G. Beiko, Dalhousie University

Copyright © 2020 Bernard et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Philippe Lopez, philippe.lopez@upmc.fr.

* Present address: Romain Lannes, Institut Universitaire du Cancer de Toulouse, Toulouse, France.

Received 12 May 2020

Accepted 24 September 2020

Published 13 October 2020

prokaryotic lineages, such as the *de novo* synthesis of nucleotides, amino acids or lipids (3, 5). These unusual traits suggest that the majority of these ultrasmall species might depend on other organisms to survive (6) or might even be obligate symbionts (5), a suggestion supported by the few endobiotic (7) and epibiotic (8, 9) relationships uncovered between a CPR or a DPANN and another microorganism(s). Nevertheless, the biological mechanisms underlying these interactions are currently poorly understood.

However, two recent studies suggest that interspecies communication via quorum sensing (QS) signals could be a new avenue of investigation to study how CPR and DPANN members may promote and/or maintain their interactions with other species (10, 11). The first study shows that the CPR member "*Candidatus Saccharibacteria*" TM7x HMT-952 interacts with and modulates the QS circuit of its obligate epibiont, *Actinomyces odontolyticus* XH001, which upregulates the biofilm formation pathway in this host and results into an increased stability of the TM7x-XH001 collective within dual-species biofilms (10). The second study hints at the presence of putative QS receptors in two DPANN genomes, "*Candidatus Woesarchaeota*" and "*Candidatus Micrarchaeota*" (11).

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

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

RESULTS

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

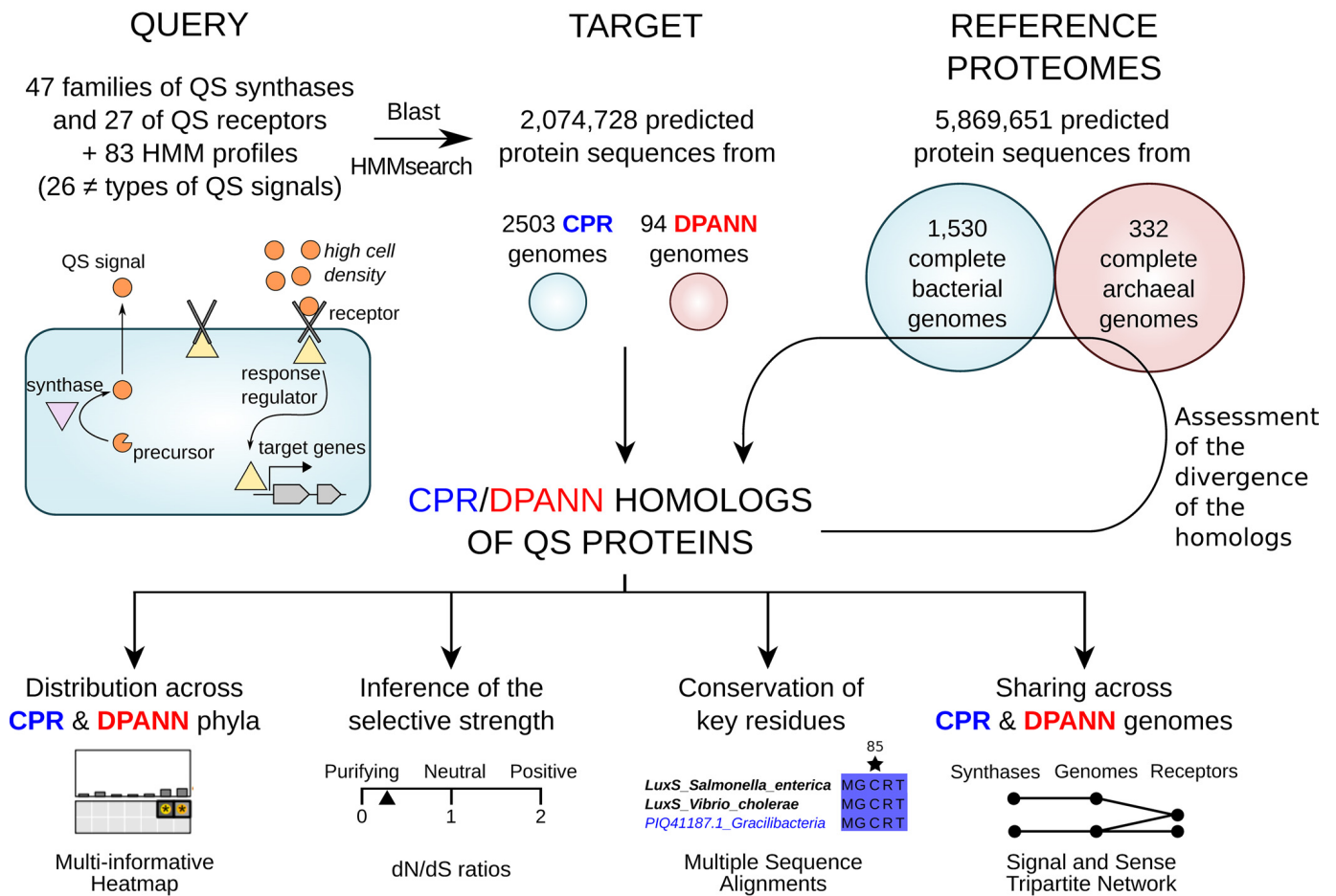


FIG 1 Study design. The first step consists in the identification of variants of QS synthases and receptors in the CPR and DPANN genomes available on the NCBI, using both BLAST queries and hidden Markov models (HMM) corresponding to reference, experimentally characterized QS proteins. Then, using BLAST,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

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

We detected a total of 5,693 homologs, matching 24 query reference QS protein families and associated with 14 distinct QS signals; 2,003 proteins were found at the first BLAST iteration (24 families/14 signals), including 1,508 hits characterized by an E value of $<1e-20$ (18 families/12 signals), whereas 1,568 proteins identified by BLAST were

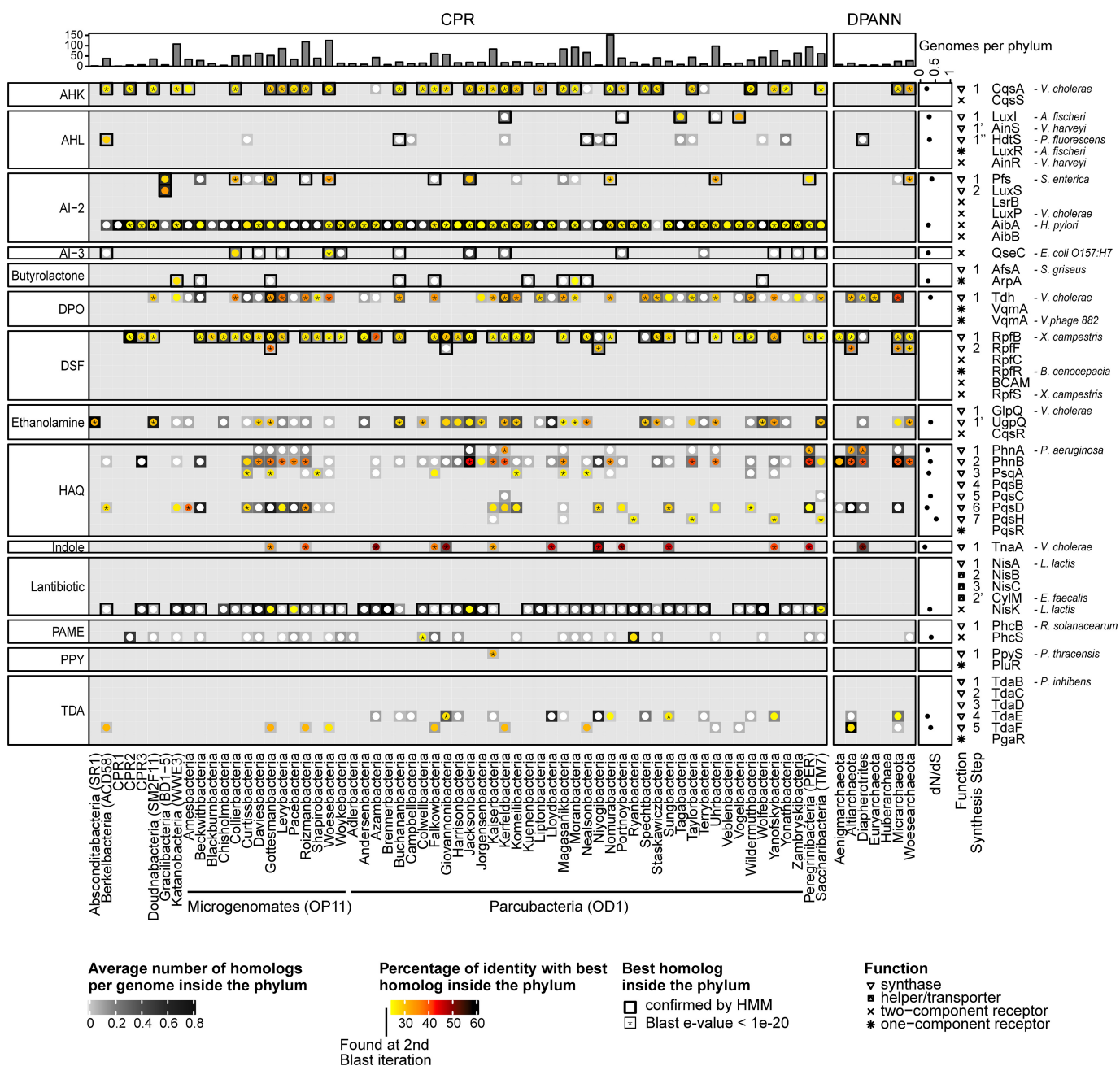


FIG 2 Multi-informative heat map of the QS synthases and receptors detected by BLAST in the CPR and DPANN phyla. Each column represents a phylum, and the histogram at the top displays the number of genomes per phylum. Rows represent query (reference) QS protein families, and the families are grouped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

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

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

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

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

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

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

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

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

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

We also paid attention to the synteny of the coding DNA sequences (CDS) of these homologs of QS synthases, because adjacent QS genes are likely to be functionally linked and to participate together in the QS function. We report that the *pfs* (*S*-adenosylhomocysteine nucleosidase) and *luxS* (*S*-ribosylhomocysteine lyase) genes encoding putative AI-2 synthases are found adjacent to each other in "*Candidatus* Gracilibacteria" from Crystal Geyser, UT. Likewise, the CDS of the homologs of the RpfB (fatty-acid-coenzyme A [CoA]-ligase) and RpfF (enoyl-CoA-hydratase) DSF synthases are found adjacent to each other in three genomes of "*Candidatus* Giovannonibacteria." Furthermore, HMMs allowed us to identify 8 additional similar genomic contexts in "*Candidatus* Giovannonibacteria" and "*Candidatus* Harrisonbacteria" (Table S4). Of note, homologs of *rpfB* and *rpfF* are also present together, yet not adjacent, in 1 "*Candidatus* Gottesmanbacteria," 1 "*Candidatus* Niyogibacteria," 1 "*Candidatus* Altiaarchaeota," and 2 "*Candidatus* Woesearchaeota" genomes. Importantly, we also found that 51 CDS of the CqsA homologs in CPRs are in synteny with the CDS of proteins matching the HMM profile of the CqsS α -hydroxyketone receptor kinases, hinting at possible complete AHK-based QS systems in certain CPR genomes. Likewise, the CDS of all the homologs of PpyS as well as of 18 homologs of LuxI were found adjacent to a sensory kinase in some CPR genomes, thereby representing other candidates for complete QS systems (Table S4).

Finally, we looked at the literature to propose roles for these 8 types of communication molecules inferred to be secreted by CPRs and DPANNs. It appears that these families of molecules are known to mediate interspecies signaling (AI-2, the DPO pyrazinol, AHL, indole, DSF) and interkingdom signaling (AHK, AHL, the AI-3 pyrazine, DSF, ethanolamine) (Table 1). The secretion of these types of QS signals hence might help CPRs and DPANNs to promote and/or maintain their crucial interactions with other species. To give a few examples, the AI-2 signal emitted by some "*Candidatus* Gracilibacteria" CPR strains could facilitate their incorporation into multispecies biofilms, in agreement with the acknowledged association between the concentration of the AI-2 signal and the formation of mutualistic biofilms (41). Again, the AHK produced by CPRs and DPANNs might regulate the growth or the motility of eukaryotic hosts, to their own benefit (42–45). The possible AHK-based and AHL-based QS systems of certain CPR endobionts might control when to remain within or when to escape from the host according to the density of CPRs within the host's cytoplasm (28, 42, 46). The biosynthesis of indole might also have been selected in some CPR and DPANN members for the various effects that this molecule exerts on microbial communities, such as modulation of biofilm formation, motility, virulence, antibiotic resistance, etc. (47) (Table 1).

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

TABLE 1 Proposed roles for the homologs of reference QS proteins found within the CPR and DPANN lineages

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

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

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

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

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

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

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

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

FIG 3 Legend (Continued)

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

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

DISCUSSION

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

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

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

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

Density-dependent mechanisms of QS are typically supported by complete QS systems (QS synthase[s] plus QS receptor). In this respect, we identified genes of QS synthases in CPR genomes that are adjacent to transmembrane kinases harboring a

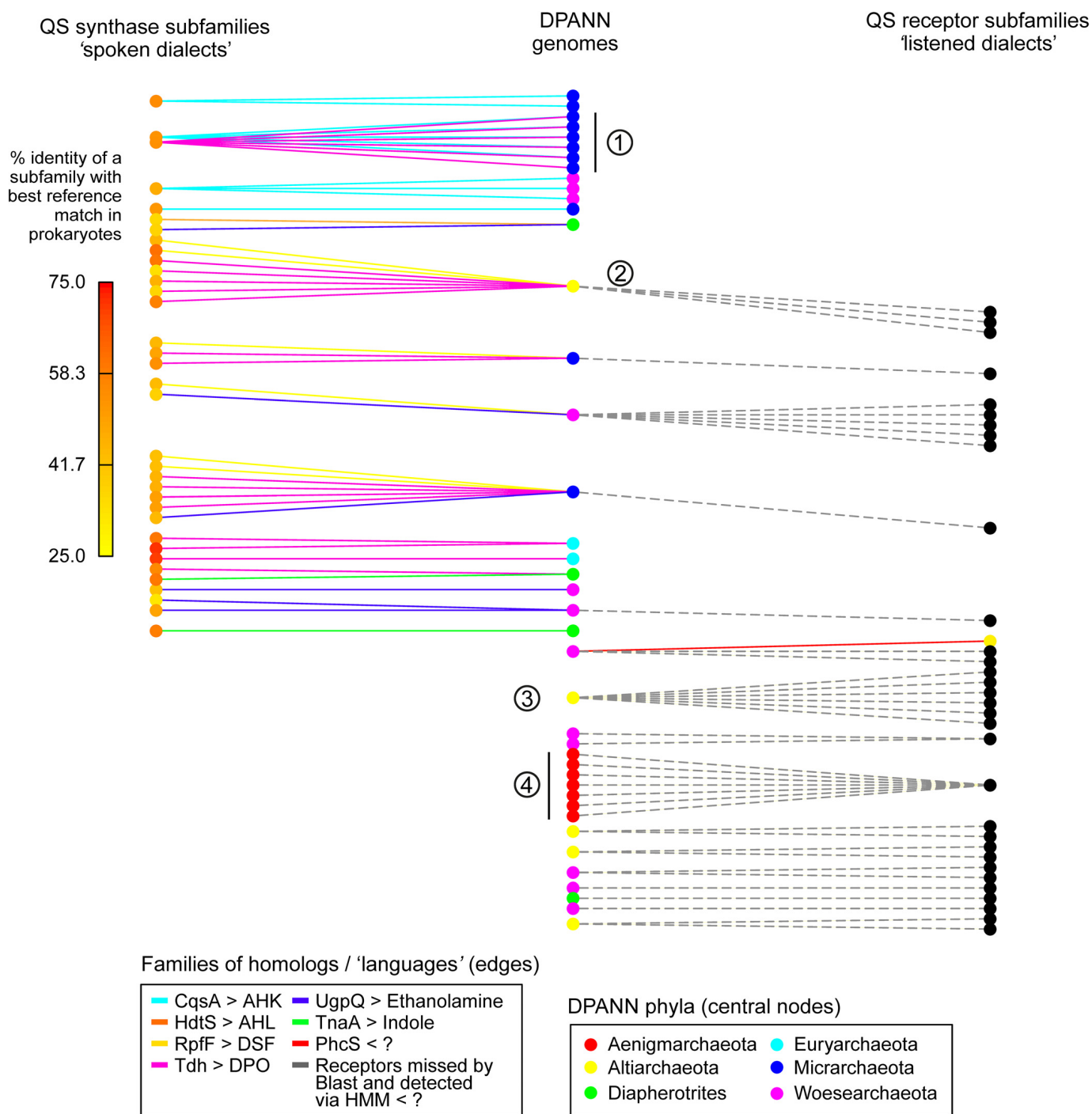


FIG 4 Signal and sense tripartite network in the DPANN lineage. This representation allows appreciating which and how many QS signals the different DPANN members might secrete and/or eavesdrop on at the genome level. The central layer of nodes corresponds to DPANN genomes. The left layer of nodes corresponds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

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

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

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

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

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

MATERIALS AND METHODS

Construction of a reference database of sequences and HMMs of QS synthases and receptors.

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

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

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

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

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

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

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

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

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

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

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.7 MB.

FIG S2, PDF file, 2.2 MB.

FIG S3, PDF file, 0.6 MB.

TABLE S1, XLSX file, 0.04 MB.

TABLE S2, XLSX file, 0.7 MB.

TABLE S3, XLSX file, 1.5 MB.

TABLE S4, XLSX file, 0.1 MB.

TABLE S5, XLSX file, 0.4 MB.

TABLE S6, XLSX file, 0.2 MB.

ACKNOWLEDGMENTS

C.B., Y.L., E.B., and P.L. conceived the study. R.L. constructed the data set of CPR/DPANN genomes/proteomes. C.B. and Y.L. constructed the database of QS systems. C.B. performed the analyses. C.B., Y.L., and E.B. wrote the manuscript with input from all of us. All documents were edited and approved by all of us.

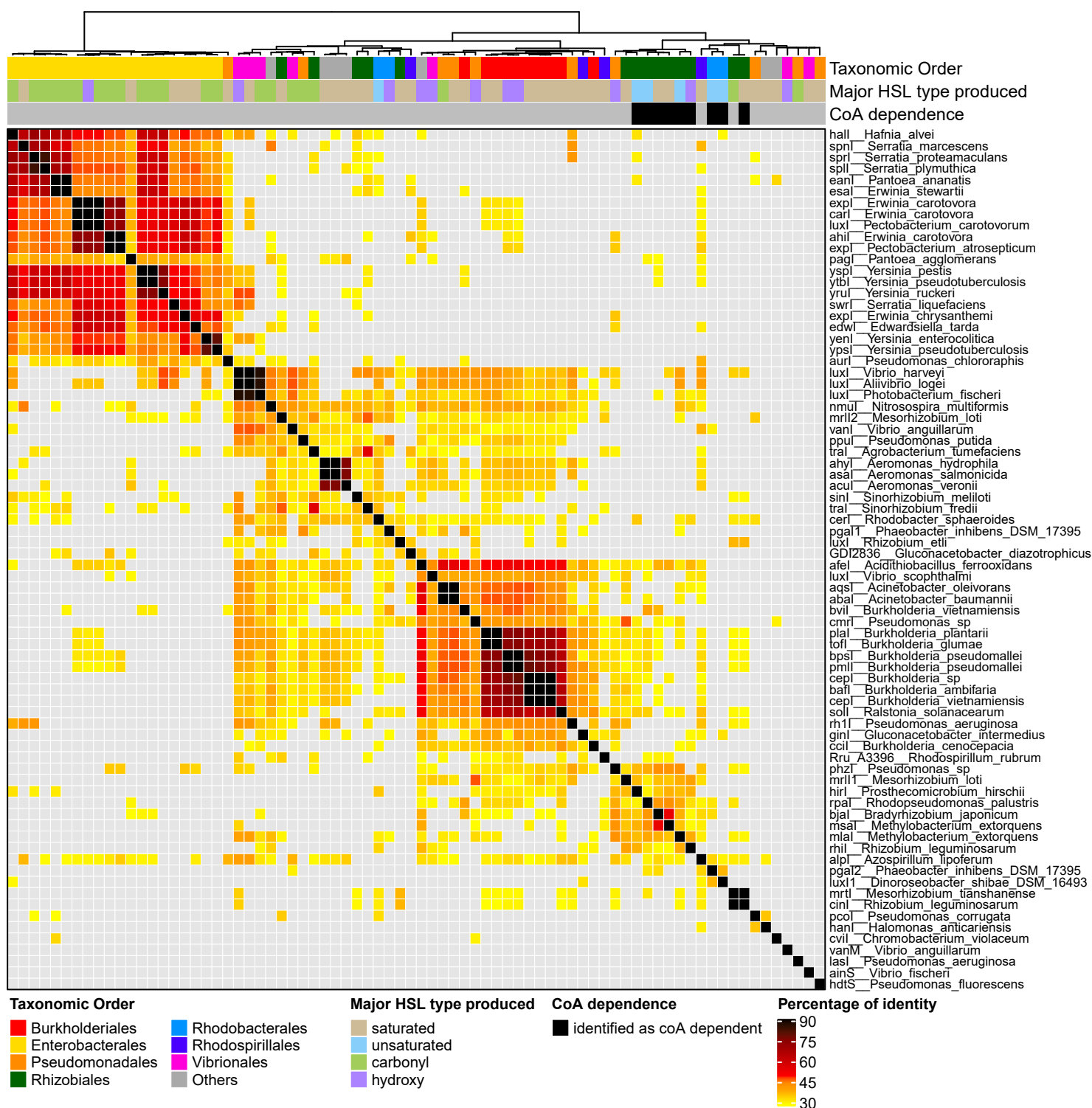
REFERENCES

- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hemsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. *Nat Microbiol* 1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>.
- Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
- Castelle CJ, Banfield JF. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 172: 1181–1197. <https://doi.org/10.1016/j.cell.2018.02.016>.
- Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, Singh A, Williams KH, Siegerist CE, Tringe SG, Downing KH, Comolli LR, Banfield JF. 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun* 6:6372. <https://doi.org/10.1038/ncomms7372>.
- Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. 2018. Biosynthetic capacity, metabolic variety and unusual biology in

- the CPR and DPANN radiations. *Nat Rev Microbiol* 16:629–645. <https://doi.org/10.1038/s41579-018-0076-2>.
6. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, Thomas BC, Banfield JF. 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *mBio* 4:e00708–13. <https://doi.org/10.1128/mBio.00708-13>.
 7. Gong J, Qing Y, Guo X, Warren A. 2014. “Candidatus *Sonnebornia yantaiensis*”, a member of candidate division OD1, as intracellular bacteria of the ciliated protist *Paramecium bursaria* (Ciliophora, Oligohymenophorea). *Syst Appl Microbiol* 37:35–41. <https://doi.org/10.1016/j.syam.2013.08.007>.
 8. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu S-Y, Dorrestein PC, Esquenazi E, Hunter RC, Cheng G, Nelson KE, Lux R, Shi W. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc Natl Acad Sci U S A* 112:244–249. <https://doi.org/10.1073/pnas.1419038112>.
 9. Dombrowski N, Lee J-H, Williams TA, Offre P, Spang A. 2019. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol Lett* 366:fnz008. <https://doi.org/10.1093/femsle/fnz008>.
 10. Bedree JK, Bor B, Cen L, Edlund A, Lux R, McLean JS, Shi W, He X. 2018. Quorum sensing modulates the epibiotic-parasitic relationship between *Actinomyces odontolyticus* and its *Saccharibacteria* epibiont, a *Nanosynbacter lyticus* strain, TM7x. *Front Microbiol* 9:2049. <https://doi.org/10.3389/fmicb.2018.02049>.
 11. Rajput A, Kumar M. 2017. Computational exploration of putative LuxR solos in Archaea and their functional implications in quorum sensing. *Front Microbiol* 8:798. <https://doi.org/10.3389/fmicb.2017.00798>.
 12. Nealson KH, Platt T, Hastings JW. 1970. Cellular control of the synthesis and activity of the bacterial luminescent system. *J Bacteriol* 104:313–322. <https://doi.org/10.1128/JB.104.1.313-322.1970>.
 13. Fuqua WC, Winans SC, Greenberg EP. 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* 176:269–275. <https://doi.org/10.1128/jb.176.2.269-275.1994>.
 14. Miller MB, Bassler BL. 2001. Quorum sensing in bacteria. *Annu Rev Microbiol* 55:165–199. <https://doi.org/10.1146/annurev.micro.55.1.165>.
 15. Verma SC, Miyashiro T. 2013. Quorum sensing in the squid-Vibrio symbiosis. *Int J Mol Sci* 14:16386–16401. <https://doi.org/10.3390/ijms140816386>.
 16. Pérez-Montaño F, Jiménez-Guerrero I, Del Cerro P, Baena-Ropero I, López-Baena FJ, Ollero FJ, Bellogín R, Lloret J, Espuny R. 2014. The symbiotic biofilm of *Sinorhizobium fredii* SMH12, necessary for successful colonization and symbiosis of *Glycine max* cv Osumi, is regulated by quorum sensing systems and inducing flavonoids via NodD1. *PLoS One* 9:e105901. <https://doi.org/10.1371/journal.pone.0105901>.
 17. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, Melamed S, Leavitt A, Savidor A, Albeck S, Amitai G, Sorek R. 2017. Communication between viruses guides lysis-lysogeny decisions. *Nature* 541:488–493. <https://doi.org/10.1038/nature21049>.
 18. Kai K. 2018. Bacterial quorum sensing in symbiotic and pathogenic relationships with hosts. *Biosci Biotechnol Biochem* 82:363–371. <https://doi.org/10.1080/09168451.2018.1433992>.
 19. Smith RS, Iglewski BH. 2003. P aeruginosa quorum-sensing systems and virulence. *Curr Opin Microbiol* 6:56–60. [https://doi.org/10.1016/s1369-5274\(03\)00008-0](https://doi.org/10.1016/s1369-5274(03)00008-0).
 20. Jung SA, Chapman CA, Ng W-L. 2015. Quadruple quorum-sensing inputs control *Vibrio cholerae* virulence and maintain system robustness. *PLoS Pathog* 11:e1004837. <https://doi.org/10.1371/journal.ppat.1004837>.
 21. Sztajer H, Szafranski SP, Tomasz J, Reck M, Nimtz M, Rohde M, Wagner-Döbler I. 2014. Cross-feeding and interkingdom communication in dual-species biofilms of *Streptococcus* mutants and *Candida albicans*. *ISME J* 8:2256–2271. <https://doi.org/10.1038/ismej.2014.73>.
 22. Sun J, Daniel R, Wagner-Döbler I, Zeng A-P. 2004. Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. *BMC Evol Biol* 4:36. <https://doi.org/10.1186/1471-2148-4-36>.
 23. Mukherjee S, Bassler BL. 2019. Bacterial quorum sensing in complex and dynamically changing environments. *Nat Rev Microbiol* 17:371–382. <https://doi.org/10.1038/s41579-019-0186-5>.
 24. Coutinho BG, Mevers E, Schaefer AL, Pelletier DA, Harwood CS, Clardy J, Greenberg EP. 2018. A plant-responsive bacterial-signaling system senses an ethanolamine derivative. *Proc Natl Acad Sci U S A* 115:9785–9790. <https://doi.org/10.1073/pnas.1809611115>.
 25. Silpe JE, Bassler BL. 2019. A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision. *Cell* 176:268–280.e13. <https://doi.org/10.1016/j.cell.2018.10.059>.
 26. Karavolos MH, Winzer K, Williams P, Khan CMA. 2013. Pathogen espionage: multiple bacterial adrenergic sensors eavesdrop on host communication systems. *Mol Microbiol* 87:455–465. <https://doi.org/10.1111/mmi.12110>.
 27. González JF, Venturi V. 2013. A novel widespread interkingdom signaling circuit. *Trends Plant Sci* 18:167–174. <https://doi.org/10.1016/j.tplants.2012.09.007>.
 28. Hochstrasser R, Hilbi H. 2017. Intra-species and inter-kingdom signaling of *Legionella pneumophila*. *Front Microbiol* 8:79. <https://doi.org/10.3389/fmicb.2017.00079>.
 29. Lee W, Lee S-H, Kim M, Moon J-S, Kim G-W, Jung H-G, Kim IH, Oh JE, Jung HE, Lee HK, Ku KB, Ahn D-G, Kim S-J, Kim K-S, Oh J-W. 2018. *Vibrio vulnificus* quorum-sensing molecule cyclo(Phe-Pro) inhibits RIG-I-mediated antiviral innate immunity. *Nat Commun* 9:1606. <https://doi.org/10.1038/s41467-018-04075-1>.
 30. Haggerty LS, Jachiet PA, Hanage WP, Fitzpatrick DA, Lopez P, O’Connell MJ, Pisani D, Wilkinson M, Baptiste E, McInerney JO. 2014. A pluralistic account of homology: adapting the models to the data. *Mol Biol Evol* 31:501–516. <https://doi.org/10.1093/molbev/mst228>.
 31. Lannes R, Olsson-Francis K, Lopez P, Baptiste E. 2019. Carbon fixation by marine ultrasmall prokaryotes. *Genome Biol Evol* 11:1166–1177. <https://doi.org/10.1093/gbe/evz050>.
 32. Rajput A, Kaur K, Kumar M. 2016. SigMol: repertoire of quorum sensing signaling molecules in prokaryotes. *Nucleic Acids Res* 44:D634–D639. <https://doi.org/10.1093/nar/gkv1076>.
 33. Hornung C, Pöehlein A, Haack FS, Schmidt M, Dierking K, Pohlen A, Schulenburg H, Blokesch M, Plener L, Jung K, Bonge A, Krohn-Molt I, Utpatel C, Timmermann G, Spieck E, Pommerening-Röser A, Bode E, Bode HB, Daniel R, Schmeisser C, Streit WR. 2013. The *Janthinobacterium* sp. HH01 genome encodes a homologue of the *V. cholerae* CqsA and *L. pneumophila* LqsA autoinducer synthases. *PLoS One* 8:e55045. <https://doi.org/10.1371/journal.pone.0055045>.
 34. Tang X, Guo Y, Zhu T, Tao H, Liu S. 2019. Identification of quorum sensing signal AHLs synthases in *Candidatus* *Jettenia caeni* and their roles in anammox activity. *Chemosphere* 225:608–617. <https://doi.org/10.1016/j.chemosphere.2019.02.192>.
 35. Papenfort K, Silpe JE, Schramma KR, Cong JP, Seyedsayamdost MR, Bassler BL. 2017. A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation. *Nat Chem Biol* 13:551–557. <https://doi.org/10.1038/nchembio.2336>.
 36. Kim CS, Gatsios A, Cuesta S, Lam YC, Wei Z, Chen H, Russell RM, Shine EE, Wang R, Wyche TP, Piizzi G, Flavell RA, Palm NW, Sperandio V, Crawford JM. 2020. Characterization of autoinducer-3 structure and biosynthesis in *E. coli*. *ACS Cent Sci* 6:197–206. <https://doi.org/10.1021/acscentsci.9b01076>.
 37. Zhou L, Zhang LH, Cámara M, He YW. 2017. The DSF family of quorum sensing signals: diversity, biosynthesis, and turnover. *Trends Microbiol* 25:293–303. <https://doi.org/10.1016/j.tim.2016.11.013>.
 38. Watve S, Barrasso K, Jung SA, Davis KJ, Hawver LA, Khataokar A, Palaganas RG, Neiditch MB, Perez LJ, Ng W-L. 2020. Parallel quorum-sensing system in *Vibrio cholerae* prevents signal interference inside the host. *PLoS Pathog* 16:e1008313. <https://doi.org/10.1371/journal.ppat.1008313>.
 39. Howard MF, Renee Bina X, Bina JE. 2019. Indole inhibits ToxR regulon expression in *Vibrio cholerae*. *Infect Immun* 87:e00776-18. <https://doi.org/10.1128/IAI.00776-18>.
 40. Brameyer S, Heermann R. 2017. Quorum sensing and LuxR solos in *Photobacterium*. *Curr Top Microbiol Immunol* 402:103–119. https://doi.org/10.1007/82_2016_28.
 41. Rickard AH, Palmer RJ, Bleher DS, Campagna SR, Semmelhack MF, Eglund PG, Bassler BL, Kolenbrander PE. 2006. Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth. *Mol Microbiol* 60:1446–1456. <https://doi.org/10.1111/j.1365-2958.2006.05202.x>.
 42. Spirig T, Tiaden A, Kiefer P, Buchrieser C, Vorholt JA, Hilbi H. 2008. The *Legionella* autoinducer synthase LqsA produces an alpha-hydroxyketone signaling molecule. *J Biol Chem* 283:18113–18123. <https://doi.org/10.1074/jbc.M801929200>.
 43. Simon S, Schell U, Heuer N, Hager D, Albers MF, Matthias J, Fahrnbauer F, Trauner D, Eichinger L, Hedberg C, Hilbi H. 2015. Inter-kingdom signaling by the *Legionella* quorum sensing molecule LAI-1 modulates cell migration through an IQGAP1-Cdc42-ARHGEF9-

- dependent pathway. *PLoS Pathog* 11:e1005307. <https://doi.org/10.1371/journal.ppat.1005307>.
44. Werner KM, Perez LJ, Ghosh R, Semmelhack MF, Bassler BL. 2014. *Caenorhabditis elegans* recognizes a bacterial quorum-sensing signal molecule through the AWCON neuron. *J Biol Chem* 289:26566–26573. <https://doi.org/10.1074/jbc.M114.573832>.
 45. Haack FS, Poehlein A, Kröger C, Voigt CA, Piepenbring M, Bode HB, Daniel R, Schäfer W, Streit WR. 2016. Molecular keys to the Janthinobacterium and Duganella spp. interaction with the plant pathogen Fusarium graminearum. *Front Microbiol* 7:1668. <https://doi.org/10.3389/fmicb.2016.01668>.
 46. Brambila-Tapia AJL, Pérez-Rueda E. 2014. A functional and phylogenetic comparison of quorum sensing related genes in *Brucella melitensis* 16M. *J Microbiol* 52:709–715. <https://doi.org/10.1007/s12275-014-3570-x>.
 47. Kim J, Park W. 2015. Indole: a signaling molecule or a mere metabolic byproduct that alters bacterial physiology at a high concentration? *J Microbiol* 53:421–428. <https://doi.org/10.1007/s12275-015-5273-3>.
 48. Weigel WA, Demuth DR. 2016. QseBC, a two-component bacterial adrenergic receptor and global regulator of virulence in Enterobacteriaceae and Pasteurellaceae. *Mol Oral Microbiol* 31:379–397. <https://doi.org/10.1111/omi.12138>.
 49. Arakawa K, Mochizuki S, Yamada K, Noma T, Kinashi H. 2007. gamma-Butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in *Streptomyces rochei*. *Microbiology (Reading)* 153:1817–1827. <https://doi.org/10.1099/mic.0.2006/002170-0>.
 50. Kleerebezem M. 2004. Quorum sensing control of lantibiotic production; nisin and subtilin autoregulate their own biosynthesis. *Peptides* 25: 1405–1414. <https://doi.org/10.1016/j.peptides.2003.10.021>.
 51. Anderson JK, Huang JY, Wreden C, Sweeney EG, Goers J, Remington SJ, Guillemin K. 2015. Chemorepulsion from the quorum signal autoinducer-2 promotes *Helicobacter pylori* biofilm dispersal. *mBio* 6:e00379-15. <https://doi.org/10.1128/mBio.00379-15>.
 52. Dintner S, Staron A, Berchtold E, Petri T, Mascher T, Gebhard S. 2011. Coevolution of ABC transporters and two-component regulatory systems as resistance modules against antimicrobial peptides in Firmicutes bacteria. *J Bacteriol* 193:3851–3862. <https://doi.org/10.1128/JB.05175-11>.
 53. Polkade AV, Mantri SS, Patwekar UJ, Jangid K. 2016. Quorum sensing: an under-explored phenomenon in the phylum Actinobacteria. *Front Microbiol* 7:131. <https://doi.org/10.3389/fmicb.2016.00131>.
 54. Sharif DI, Gallon J, Smith CJ, Dudley E. 2008. Quorum sensing in Cyanobacteria: N-octanoyl-homoserine lactone release and response, by the epilithic colonial cyanobacterium *Gloeotheca PCC6909*. *ISME J* 2:1171–1182. <https://doi.org/10.1038/ismej.2008.68>.
 55. Zhang G, Zhang F, Ding G, Li J, Guo X, Zhu J, Zhou L, Cai S, Liu X, Luo Y, Zhang G, Shi W, Dong X. 2012. Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon. *ISME J* 6:1336–1344. <https://doi.org/10.1038/ismej.2011.203>.
 56. Paggi RA, Martone CB, Fuqua C, Castro RE. 2003. Detection of quorum sensing signals in the haloalkaliphilic archaeon *Natronococcus occultus*. *FEMS Microbiol Lett* 221:49–52. [https://doi.org/10.1016/S0378-1097\(03\)00174-5](https://doi.org/10.1016/S0378-1097(03)00174-5).
 57. Mehmood A, Liu G, Wang X, Meng G, Wang C, Liu Y. 2019. Fungal quorum-sensing molecules and inhibitors with potential antifungal activity: a review. *Molecules* 24:1950. <https://doi.org/10.3390/molecules24101950>.
 58. Liu B, Faller LL, Klitgord N, Mazumdar V, Ghodsi M, Sommer DD, Gibbons TR, Treangen TJ, Chang Y-C, Li S, Stine OC, Hasturk H, Kasif S, Segrè D, Pop M, Amar S. 2012. Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PLoS One* 7:e37919. <https://doi.org/10.1371/journal.pone.0037919>.
 59. Rylev M, Bek-Thomsen M, Reinholdt J, Ennibi O-K, Kilian M. 2011. Microbiological and immunological characteristics of young Moroccan patients with aggressive periodontitis with and without detectable Aggregatibacter actinomycetemcomitans JP2 infection. *Mol Oral Microbiol* 26:35–51. <https://doi.org/10.1111/j.2041-1014.2010.00593.x>.
 60. Jang Y-J, Choi Y-J, Lee S-H, Jun H-K, Choi B-K. 2013. Autoinducer 2 of *Fusobacterium nucleatum* as a target molecule to inhibit biofilm formation of periodontopathogens. *Arch Oral Biol* 58:17–27. <https://doi.org/10.1016/j.archoralbio.2012.04.016>.
 61. Wynendaele E, Bronselaer A, Nielandt J, D'Hondt M, Stalmans S, Bracke N, Verbeke F, Van De Wiele C, De Tré G, De Spiegeleer B. 2013. Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. *Nucleic Acids Res* 41:D655–D659. <https://doi.org/10.1093/nar/gks1137>.
 62. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>.
 63. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA, Sillitoe I. 2017. CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* 45:D289–D295. <https://doi.org/10.1093/nar/gkw1098>.
 64. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
 65. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
 66. NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 44:D7–D19. <https://doi.org/10.1093/nar/gkv1290>.
 67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 68. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
 69. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
 70. Korber-Irrgang B. 2000. HIV signature and sequence variation analysis. Computational analysis of HIV molecular sequences. <https://www.scienceopen.com/document?vid=f054582b-15a6-4ccc-be34-99d0e9d09004>.
 71. Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32: 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>.
 72. Csardi G, Nepusz T. 2005. The igraph software package for complex network research. https://www.researchgate.net/publication/221995787_The_Igraph_Software_Package_for_Complex_Network_Research.
 73. Nikolov NS. 2016. Sugiyama algorithm, p 2162–2166. *In* Encyclopedia of algorithms. Springer New York, New York, NY.
 74. Johnson AR, Dekker EE. 1998. Site-directed mutagenesis of histidine-90 in *Escherichia coli* L-threonine dehydrogenase alters its substrate specificity. *Arch Biochem Biophys* 351:8–16. <https://doi.org/10.1006/abbi.1997.0501>.
 75. Wang Y, Yi L, Wang S, Fan H, Ding C, Mao X, Lu C. 2015. Crystal structure and identification of two key amino acids involved in AI-2 production and biofilm formation in *Streptococcus suis* LuxS. *PLoS One* 10: e0138826. <https://doi.org/10.1371/journal.pone.0138826>.
 76. Churchill MEA, Chen L. 2011. Structural basis of acyl-homoserine lactone-dependent signaling. *Chem Rev* 111:68–85. <https://doi.org/10.1021/cr1000817>.
 77. Liu L, Li T, Cheng XJ, Peng CT, Li CC, He LH, Ju SM, Wang NY, Ye TH, Lian M, Xiao QJ, Song YJ, Zhu YB, Yu LT, Wang ZL, Bao R. 2018. Structural and functional studies on *Pseudomonas aeruginosa* DspI: implications for its role in DSF biosynthesis. *Sci Rep* 8:3928. <https://doi.org/10.1038/s41598-018-22300-1>.
 78. Wang J, Liu Q, Li X, Ma S, Hu H, Wu B, Zhang X-X, Ren H. 2020. In-situ monitoring AHL-mediated quorum-sensing regulation of the initial phase of wastewater biofilm formation. *Environ Int* 135:105326. <https://doi.org/10.1016/j.envint.2019.105326>.
 79. Kim J, Park W. 2013. Indole inhibits bacterial quorum sensing signal transmission by interfering with quorum sensing regulator folding. *Microbiology (Reading)* 159:2616–2625. <https://doi.org/10.1099/mic.0.070615-0>.

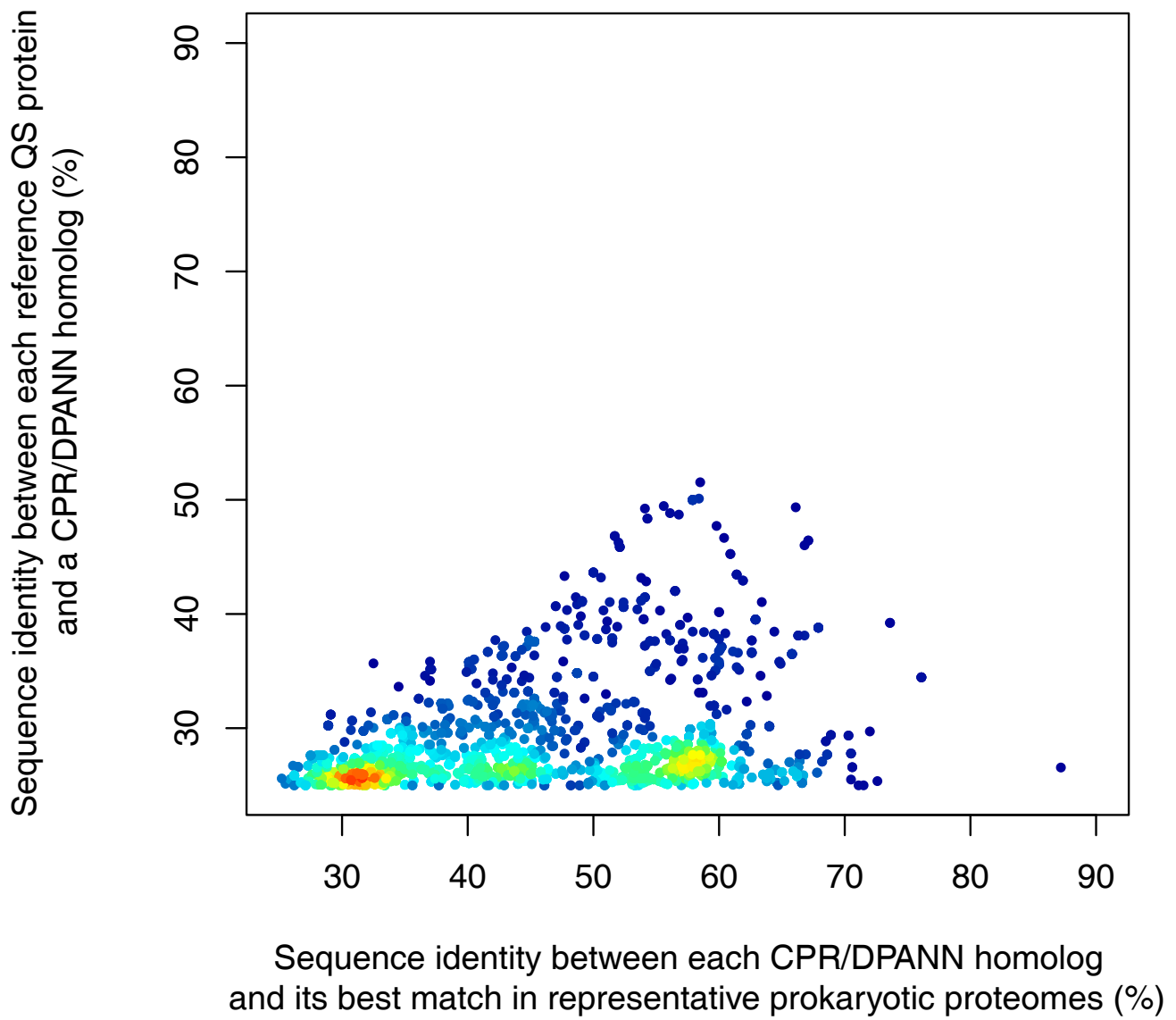
Figure S1



Supplementary Figure 1

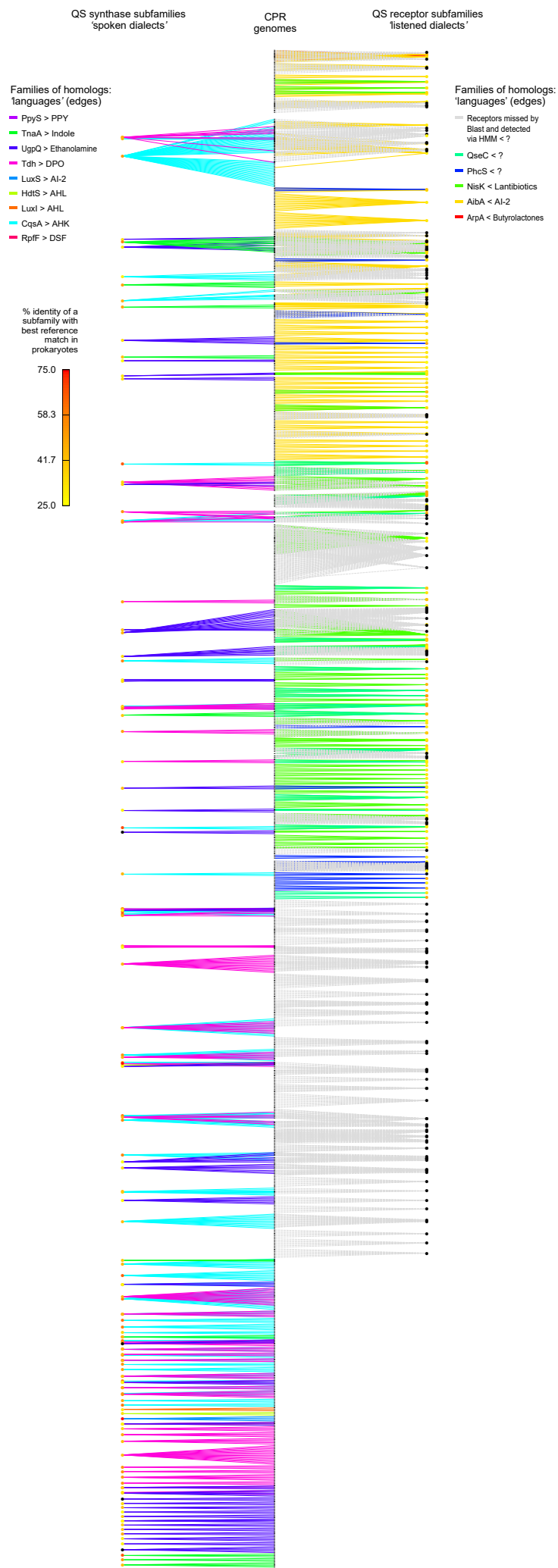
Heatmap representing the mutual identity of 76 experimentally validated acyl/aryl-homoserine lactone synthases retrieved from the Sigmol database. The matrix is symmetrical and each row/column represents an AHL synthase. Each row is labelled according to the name of a synthase, followed by the species within which it has been characterized. The color intensity of the heatmap represents the percentage of identity between two synthases over more than 75% mutual coverage. A grey color means that the two synthases did not pass our thresholds to assess a homology. Top annotation of the heatmap displays three pieces of information for each synthase: i) the taxonomic order of the species that encodes the synthase, ii) the major chemical type of acyl-homoserine lactones it produces according to Sigmol, iii) whether it is known to preferentially use coenzyme A (CoA) over acyl carrier protein (ACP) for the biosynthesis of acyl/aryl-homoserine lactones.

Figure S2



Supplementary Figure 2

Scatterplot of the sequence identities of each CPR/DPANN homolog with: its corresponding reference QS-related protein (Y axis), its closest homolog identified in the predicted proteomes of reference prokaryotes (X axis). The color gradient is representative of the local density of data points.



Supplementary Figure 3

Signal and Sense Tripartite Network in the CPR lineage. This representation allows appreciating which and how many QS signals the different CPR members might secrete and/or eavesdrop on at the genome-level. The central layer of nodes corresponds to CPR genomes. The left layer of nodes corresponds to subfamilies of homologs of QS synthases (family of homologs sub-partitioned at 90% sequence identity), and the right layer of nodes to subfamilies of homologs of QS receptors. Hence, whenever two “Genome” Nodes are connected to a left node, it means that they encode highly similar variants of QS synthases, and are thus likely co-signalers. Likewise, two “Genome” Nodes connected to a right node means that they are likely co-sensors. The color of edges allows identifying the initial family from which a subfamily has been further defined. Hence, left and right nodes could depict “spoken” and “listened” “dialects”, respectively, whereas the color of edges could depict the “languages” from which these dialects derive. Of note, dashed edges correspond to putative QS receptors identified by HMM and missed by Blast. The color of the subfamilies identified by Blast indicates their respective average percentage of identity with the best matches in reference prokaryotes: yellow nodes thus highlight more divergent sequences than red nodes.

Les séquences codantes des AHL synthases de la famille LuxI (la plus étudiée dans le QS) identifiées dans l'article précédent au sein de génomes de CPRs ont été commandées et optimisées pour l'utilisation de codons d'*Escherichia coli* en vue de les cloner dans un plasmide d'expression par lequel *E. coli* sera transformé. La comparaison des spectres de masse des milieux de culture d'*E. coli* sans et avec un homologue de *luxI* inséré dans le plasmide permettra ainsi d'identifier si chacune des synthases produit effectivement un métabolite. Le cas échéant, ce métabolite sera purifié, sa structure sera caractérisée et son appartenance à la famille des AHLs pourra donc être confirmée ou infirmée. La qualité de synthèse d'AHL des homologues de LuxI identifiés chez les CPRs pourra ainsi être testée fonctionnellement.

4. Identification systématique de systèmes de QS dont les signaux sont des peptides de QS

À la différence du chapitre précédent consacré aux QSSs dont le signal associé est une petite molécule de QS synthétisée par une enzyme, ce chapitre porte sur l'identification de QSSs candidats dont le signal est un peptide, mûri à partir d'un propeptide codé génétiquement. En outre, ce chapitre ne sera pas consacré à la seule identification de QSSs par des méthodes de recherche d'homologues. En effet, une nouvelle méthode d'identification de QSSs candidats non homologues à des QSSs déjà connus sera présentée. Cette méthode permet d'étendre la diversité génétique décrite des QSSs peptidiques, notamment chez les virus, un objet de recherche nouveau et sous-exploré dans le champ disciplinaire du QS.

4.1. Détection de variants de gènes de QS par des méthodes classiques de recherche d'homologues

Les propeptides de QS, secrétés et mûris en un petit peptide de QS peuvent évoluer très vite, notamment car la partie N-terminale d'adressage à un complexe protéique de sécrétion n'est pas soumise à une grande pression de purification (des substitutions d'acides aminés du même type, hydrophobes ou basiques, ne sont pas discriminantes pour l'adressage au système d'export) et car une pression de sélection peut exister pour la divergence concomitante de leur partie C-terminale codant le peptide mature avec celle des résidus du récepteur impliqués dans la sélectivité du peptide reconnu (Even-Tov *et al.*, 2016). Cette pression de sélection pour l'émergence d'une nouvelle spécificité récepteur-peptide chez certains QSSs peut par ailleurs expliquer les événements de duplication intragénique suivis de la divergence de parties dupliquées observées chez certaines familles de propeptides (Even-Tov *et al.*, 2016). Ainsi, les séquences de propeptides homologues ont tendance à être très versatiles, aussi bien en terme de longueur que de composition en acides aminés. Du fait de cette versatilité et de la petite taille des propeptides (<70aa) n'offrant que peu d'acides aminés à comparer, la détection de QSSs dont le signal associé est peptidique consiste avant tout à identifier un homologue de récepteur déjà connu, puis d'analyser son voisinage génomique pour détecter un éventuel candidat au statut de propeptide de QS.

Puisque le QS des virus est assez méconnu et représente un champ de recherche nouveau (Erez *et al.*, 2017; Stokar-Avihail *et al.*, 2019), une analyse fut lancée en cours de thèse pour chercher d'éventuels homologues de protéines de QS (aussi bien associées aux petites molécules qu'aux

peptides de QS) au sein des génomes de virus du NCBI. Seules 5 familles de récepteurs eurent au moins un match significatif. Parmi ces hits, les familles VqmA et LuxR de récepteurs de petites molécules de QS étaient déjà décrites comme impliquées dans l'espionnage du quorum sensing de l'hôte par des bactériophages (Silpe and Bassler, 2019a, 2019b). En outre, les familles AimR et AimR-like correspondent aux récepteurs des peptides de QS « arbitrium » déjà caractérisés chez les phages (Erez *et al.*, 2017; Stokar-Avihail *et al.*, 2019). En revanche, le 5ème hit correspondant à la famille Rap de récepteurs peptidiques de QS caractérisée chez les Bacilles (Pottathil and Lazazzera, 2003) n'avait jamais été décrit auparavant chez les virus. Qui plus est, ce hit viral de *rap* présent dans le génome de Bacillus phage phi3T était le voisin direct d'un gène codant une petite protéine correspondant exactement au profil des propeptides de la famille Phr associés à Rap chez les Bacilles (Pottathil and Lazazzera, 2003). L'article suivant publié dans le journal *ISMEj* décrit cette découverte, quantifie la prévalence des systèmes Rap-Phr chez les phages/prophages tempérés de Bacilles et discute des éventuelles fonctions que ces systèmes Rap-Phr viraux pourraient assurer. Parmi ces fonctions, la manipulation densité-dépendante de la biologie de l'hôte est pressentie, ce qui représenterait alors une contribution majeure dans le champ du QS viral car seuls des processus viraux de types transition lyse-lysogénie étaient jusqu'alors décrits comme régulés par QS chez les phages (Stokar-Avihail *et al.*, 2019). Enfin, cet article présente le premier génome de bactériophage identifié comme codant de multiples QSSs, qui plus est de familles génétiques différentes, ce qui apporte un nouvel éclairage sur le degré de complexité que peut revêtir la communication entre virus.



Beyond arbitrium: identification of a second communication system in *Bacillus* phage phi3T that may regulate host defense mechanisms

Charles Bernard^{1,2} · Yanyan Li² · Philippe Lopez¹ · Eric Bapteste¹

Received: 2 April 2020 / Revised: 17 September 2020 / Accepted: 24 September 2020
© The Author(s) 2020. This article is published with open access

Abstract

The evolutionary stability of temperate bacteriophages at low abundance of susceptible bacterial hosts lies in the trade-off between the maximization of phage replication, performed by the host-destructive lytic cycle, and the protection of the phage-host collective, enacted by lysogeny. Upon *Bacillus* infection, *Bacillus* phages phi3T rely on the “arbitrium” quorum sensing (QS) system to communicate on their population density in order to orchestrate the lysis-to-lysogeny transition. At high phage densities, where there may be limited host cells to infect, lysogeny is induced to preserve chances of phage survival. Here, we report the presence of an additional, host-derived QS system in the phi3T genome, making it the first known virus with two communication systems. Specifically, this additional system, coined “Rap ϕ -Phr ϕ ”, is predicted to downregulate host defense mechanisms during the viral infection, but only upon stress or high abundance of *Bacillus* cells and at low density of population of the phi3T phages. Post-lysogenization, Rap ϕ -Phr ϕ is also predicted to provide the lysogenized bacteria with an immediate fitness advantage: delaying the costly production of public goods while nonetheless benefiting from the public goods produced by other non-lysogenized *Bacillus* bacteria. The discovered “Rap ϕ -Phr ϕ ” QS system hence provides novel mechanistic insights into how phage communication systems could contribute to the phage-host evolutionary stability.

Introduction

Temperate bacteriophages are viruses that can infect their bacterial hosts either through the host-destructive lytic cycle or the non-destructive lysogenic cycle. During the latter stage, the phage replicates as part of the bacterial genome as a “prophage” and confers immunity to a second infection upon

the lysogenized bacterium [1, 2]. Lysogeny, therefore, is a strategy that protects the phage-host collective [3].

Remarkably, Erez et al. observed that upon *Bacillus subtilis* infection, *Bacillus* phage phi3T (phi3T) relies on an endogenous quorum sensing (QS) system called “arbitrium” to orchestrate the lysis-to-lysogeny transition, when phages become abundant and the susceptible hosts left to infect are likely to be few [4].

Here, we report the discovery of a second communication system, termed “Rap ϕ -Phr ϕ ”, in the genome of phi3T, now the first known virus with two communication systems. Rap ϕ -Phr ϕ is the first known phage-encoded QS system predicted to regulate bacterial mechanisms, and in both a phage- and a host-density dependent manner.

These authors contributed equally: Philippe Lopez, Eric Bapteste

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-00795-9>) contains supplementary material, which is available to authorized users.

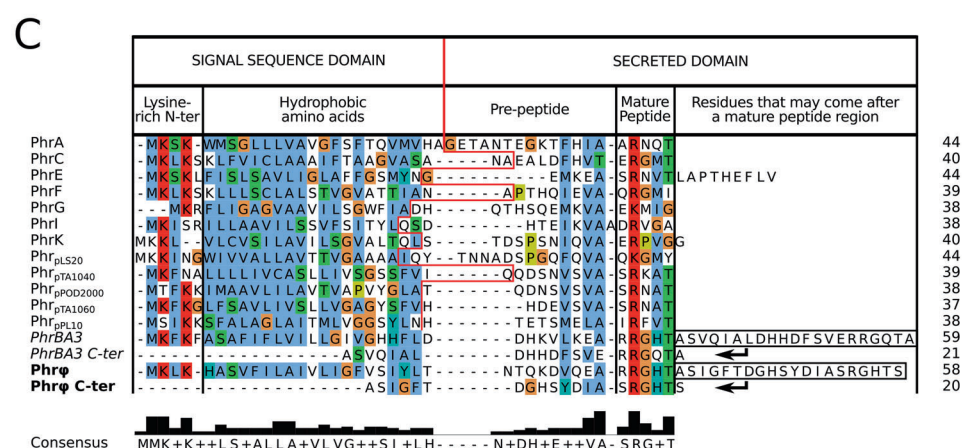
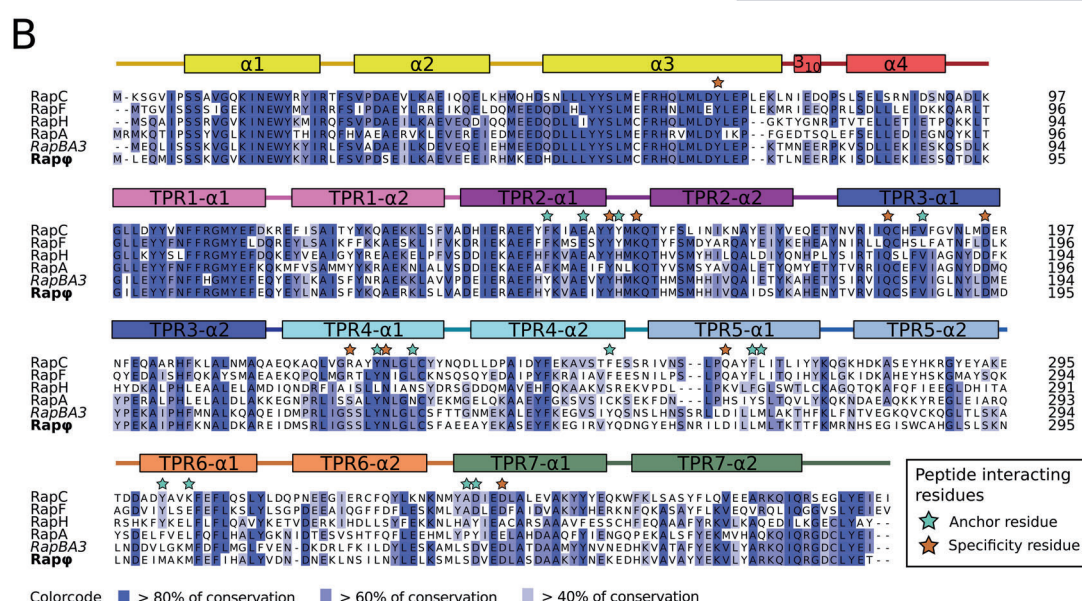
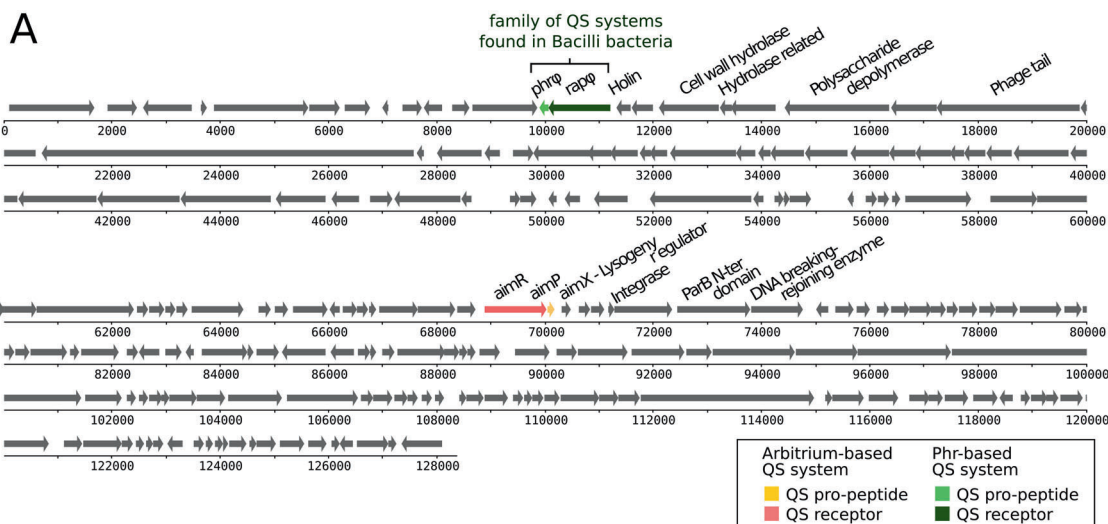
✉ Charles Bernard
charles.bernard@cri-paris.org

¹ Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Muséum National d’Histoire Naturelle, Campus Jussieu, Bâtiment A, 4eme et. Pièce 429, 75005 Paris, France

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

Results and discussion

In addition to arbitrium, we report that *Bacillus* phage phi3T encodes a second QS system, coined “Rap ϕ -Phr ϕ ”, from the Rap-Phr family of *Bacillus* bacteria [5, 6] (Fig. 1a). The Rap ϕ intracellular QS receptor (NCBI accession APD21157.1) is a 379aa long protein that shows 48%



sequence identity over its entire length with RapC, a well-characterized member of the Rap protein family [7, 8] (Fig. 1b). Rapφ harbors trichopeptide repeats (TPR

(Pfam PF13181.6 and PF13424.6 HMM profiles), typically involved in the binding of QS peptides [9] (Fig. 1b). Phrφ (NCBI accession APD21156.1) is a 58aa long protein that

Fig. 1 Characterization of the Rap ϕ -Phr ϕ QS system in *Bacillus* phage phi3T. **a Position of the *rap ϕ -phr ϕ* and of the arbitrium (*aimR-aimP*) QS systems in the genome of phi3T. On this genomic map, arrows pointing towards the right correspond to genes on the “+” DNA strand whereas arrows pointing towards the left correspond to genes on the “-” strand. The position of each gene in the genome of phi3T is indicated by the horizontal axis (number of base pairs). The genes belonging to a QS system are in color. The *aimX* regulator of lysogeny is located downstream from *aimP*. The genes in the vicinity of each QS system are functionally annotated, whenever it was possible. **b** Multiple sequences alignment of the RapA, C, F and H proteins (*B. subtilis* str. 168) for which the structure has been resolved in Gallego del Sol and Marina [7] alongside RapBA3 (prophage of *B. velezensis* FZB42) and Rap ϕ (*Bacillus* phage phi3T). The N-terminal region of the Rap proteins is involved in the interaction with the ComA-P and/or Spo0F-P response regulators whereas the C-terminal region (TPR repeats) is involved in the interaction with the Phr_{mature} QS peptide. Residues are colored according to the “% identity” colorcode. The residues that have been identified by Gallego del Sol and Marina as involved in the interaction with the QS peptide are annotated by stars, and brown stars specifically highlight the residues responsible for the specificity of Rap towards its cognate Phr_{mature} peptide. **c** Multiple sequences alignment of the canonical Phr proteins considered in Pottathil et al [10] alongside the full PhrBA3 protein, the C-terminus of PhrBA3 that follows the first PhrBA3_{mature} region (RRGHT), the full Phr ϕ protein, and the C-terminus of Phr ϕ that follows the first Phr ϕ _{mature} region (RRGHT). Residues are colored according to the “Clustal” colorcode (see: <http://www.jalview.org/help/html/colourSchemes/clustal.html>). The red line in the middle of the alignment separates the N-terminal signal sequence domain from the C-terminal secreted domain of Phr proteins, as annotated by Pottathil et al. The sequences of the different QS pentapeptides are indicated by the “Mature Peptide” column.**

exhibits all the characteristics of Rap cognate Phr pre-peptides [10], both in terms of sequence composition (Fig. 1c) and genetic organization, with the *phr* gene directly downstream of *rap*, on the same DNA strand (Fig. 1a). Phr ϕ is predicted to be secreted via the SEC-translocon by SignalP-5.0 [11] (likelihood of 0.76) and harbors two characteristic regions that might be processed by extracellular proteases into the mature R-R-G-H-T and/or S-R-G-H-T QS signal pentapeptide(s) (Fig. 1c). Although non-canonical, the presence of two regions that are eventually matured into QS peptides was reported in many other functional Phr proteins [5]. The *rap ϕ -phr ϕ* locus forms an operon, directly flanked downstream and upstream by intrinsic terminators (Table S1). The mapping of RNA-seq reads sequenced during the infection of *Bacillus* by phi3T [4] to the phi3T genome indicated that *rap ϕ* and *phr ϕ* are expressed upon *Bacillus* infection (Fig. S1).

In the *Bacillus* genus, Rap-Phr systems are diverse, often strain specific and subject to horizontal transfers [5, 6]. We identified 2360 homologs of the Rap ϕ -Phr ϕ QS systems in the complete genomes of Firmicutes available on the NCBI, including 333 predicted by Phaster [12] to belong to prophages, suggesting that *rap-phr* QS systems are widely used and horizontally transferred by phages (Table S2 and Fig. S2). Two of the QS systems carried by prophages (RapBA3-PhrBA3 in *B. velezensis* FZB42 and RapBL5-

PhrBL5 in *B. licheniformis* ATCC14580) have previously been functionally tested via a dedicated synthetic construct in *B. subtilis* strain PY79 (although it was then unknown that these systems belong to prophages) [5].

At low densities of the *Bacillus* subpopulation that either expresses the cloned *rapBA3* or *rapBA5* receptors, the Rap protein was experimentally demonstrated to strongly inhibit the Spo0F-P and ComA-P response regulators [5]. The Spo0F-P pathway leads to biofilm formation or sporulation [13, 14], shown to elicit the lytic cycle of certain bacteriophages [15, 16]. The ComA-P pathway triggers competence and the production of antimicrobials [17] that might contribute to the defense against phages. As a Rap-Phr expressing *Bacillus* subpopulation grows, the secreted Phr_{mature} peptide accumulates in the medium. At high concentrations, PhrBA3_{mature} and PhrBL5_{mature} are internalized by the Opp permease and demonstrably sequester their cognate Rap receptor, which, in turn, alleviates the inhibition of ComA-P, Spo0F-P and their target pathways [5]. Therefore, there is evidence that Rap-Phr systems of prophages can delay, in *Bacillus*, defense mechanisms (antimicrobials, sporulation) and the production of public goods (biofilm molecules, antimicrobials) until the *Bacillus* subpopulation expressing the prophage QS system reaches a substantial cellular density (Figs. 2a and S3). Importantly, Rap ϕ -Phr ϕ is highly similar to RapBA3-PhrBA3. Rap ϕ and RapBA3 yield 77% sequence identity at full mutual coverage and harbor the exact same residues that account for the QS peptide specificity (Fig. 1b). Consistently, Phr ϕ presents a mature peptide region (R-R-G-H-T), identical to the PhrBA3_{mature} QS peptide shown to inhibit RapBA3 [5] (Fig. 1c).

On these bases, we predict that Rap ϕ -Phr ϕ is likely to be functional and that Rap ϕ can thus downregulate host defense mechanisms via Spo0F-P and ComA-P inhibition, enhancing the efficiency of the phi3T lytic cycle. Importantly, Rap ϕ regulation of host processes is predicted to occur only at low concentrations of Phr ϕ _{mature} and when Spo0F is phosphorylated upon stress or ComA during the QS response of the entire, mixed population of *Bacillus* cells [13, 18, 19] (see refs. [20–23] for the potential advantages for a phage to inhibit the host’s QS response) (Figs. 2 and S4). Hence, when defense pathways are triggered at high densities of *Bacillus* bacteria, phi3T phages might “disarm” their host only at low densities of phages, when the phage-host collective is not yet endangered. Meanwhile, the arbitrium system of phi3T ensures a fast-replication through the lytic cycle when phages are rare. When phages become abundant, it triggers a population-wide, host-protective lysogeny [4]. Particularly, if the quorum mediated by arbitrium were reached before that of Phr ϕ _{mature}, the Rap ϕ protein of phi3T prophages could delay the production of public goods in their freshly lysogenized hosts until the lysogenized subpopulation, likely afflicted by the previous lytic cycle, reaches a substantial density [19].

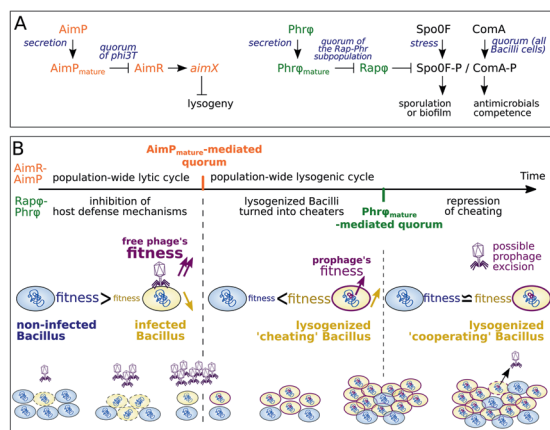


Fig. 2 Hypotheses on the mechanism of Rap Φ -Phr Φ , on its synergy with the arbitrium system, and on the evolutionary advantages for phi3T to own both systems. **a** Wiring diagrams of the characterized mechanism of the arbitrium system (left) and of the predicted mechanism of Rap Φ -Phr Φ (right). Upon *Bacillus* infection, AimP and Phr Φ are predicted to be exported in the medium by the SEC-translocon and cleaved by extracellular peptidases into mature QS peptides that accumulate in the medium as phi3T replicates. At low densities of phi3T, the unbound AimR QS receptor activates the expression of *aimX* encoding the negative regulator of lysogeny, which ensures a population-wide lytic cycle. At high densities of phi3T, AimP_{mature} reaches a threshold extracellular concentration and is consistently imported by neighboring cells via the Opp permease. Within infected *Bacillus* cells, AimP_{mature} inhibits AimR and alleviates lysogeny inhibition. A population-wide lysogeny is then triggered. At low densities of phi3T, the unbound Rap Φ QS receptor is predicted to inhibit the ComA-P and Spo0F-P response regulators of *Bacillus* (if the host is stressed or if the entire *Bacillus* population is at high density), which might downregulate host defense mechanisms. Either at high densities of phi3T or at high densities of the lysogenized subpopulation, Phr_{mature} is predicted to inhibit Rap Φ , which might reactivate the ComA-P and Spo0F-P pathways of the host. **b** A timeline is drawn on the top of the panel; the orange and the green vertical marks indicate when the quorums mediated by AimP_{mature} and by Phr_{mature} are proposed to intervene. The different mechanisms regulated by the arbitrium QS system over time are indicated on the topside of the timeline whereas the ones predicted to be regulated by Rap Φ -Phr Φ are indicated on the bottom side. The predicted variations of density of the different subpopulations (phages phi3T, non-infected, infected and lysogenized *Bacillus* bacteria) over time are shown in the bottom of the panel. Their relative predicted fitness over time is shown in the middle of the panel. Before the AimP_{mature}-mediated quorum, phi3T quickly replicates at the expense of *Bacillus* during the lytic cycle. Meanwhile, Rap Φ might enhance the efficiency of the lytic cycle by downregulating host defense mechanism when *Bacillus* cells are abundant. When phages become abundant and when there may be limited host cells to infect, the AimP_{mature}-mediated quorum triggers the host-protective lysogenic cycle and phi3T phages are internalized as prophages. The arbitrium QS system thus maximizes the time spent in the lytic cycle before it becomes a threat for the phage-host collective. After lysogeny, the Rap Φ protein of phi3T prophages might enhance the fitness of the prophage-host collective, by conferring an immediate fitness advantage on the freshly lysogenized subpopulation, likely afflicted by the previous lytic cycle. Indeed, whenever the entire *Bacillus* population would be at high densities and would transduce the signal for the production of public goods, the ComA-P and Spo0F-P inhibitions via Rap Φ might provide the lysogenized subpopulation with the means to cheat (delay the costly production of biofilm molecules and antimicrobials) until this lysogenized subpopulation reaches again a substantial cellular density, as reflected by a threshold concentration of Phr_{mature}.

Rap Φ -Phr Φ could hence momentarily turn lysogenized *Bacillus* bacteria into cheaters, enhancing both the fitness of these cells and of the phi3T prophages. Interestingly, one prophage of *B. velezensis* SQR9 carried both the arbitrium and the Rap-Phr QS systems, as in phi3T (Fig. S2B), further supporting that owning both communication systems might be advantageous for a temperate bacteriophage.

Data availability

All data generated or analyzed during this study are included in this published article and its Supplementary information files. The genome of *Bacillus* phage phi3T can be accessed from the NCBI “Assembly” database with the “ASM260144v1” accession ID. The accession IDs for the protein sequences of Rap Φ and Phr Φ are “APD21157.1” and “APD21156.1”, respectively.

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

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 paper with input from all authors. All documents were edited and approved by all authors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

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

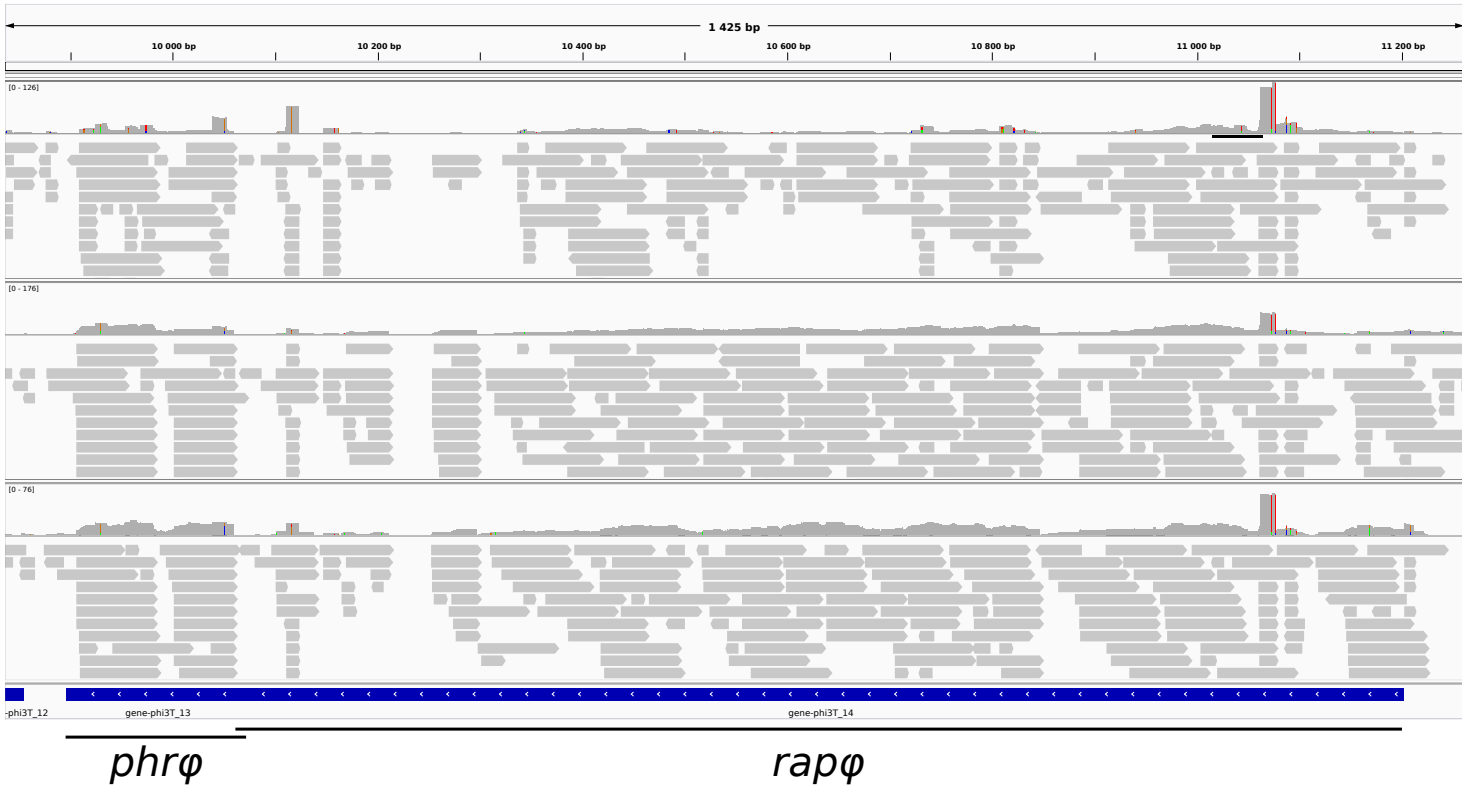
References

1. Openheim AB, Kobiler O, Stavans J, Court DL, Adhya S. Switches in Bacteriophage Lambda Development. *Annu Rev Genet.* 2005;39:409–29.
2. Abedon ST. Bacteriophage secondary infection. *Virology.* 2015; 30:3–10.
3. Wahl LM, Betti MI, Dick DW, Pattenden T, Puccini AJ. Evolutionary stability of the lysis-lysogeny decision: why be virulent? *Evolution (N.Y.)*. 2018;73:evo.13648.

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*. 2017;541:488–93.
5. Even-Tov E, Omer Bendori S, Pollak S, Eldar A. Transient duplication-dependent divergence and horizontal transfer underlie the evolutionary dynamics of bacterial cell–cell signaling. *PLoS Biol*. 2016;14:e2000330.
6. Cardoso P, de F, Perchat S, Vilas-Boas LA, Lereclus D, Vilas-Bôas GT. Diversity of the Rap–Phr quorum-sensing systems in the *Bacillus cereus* group. *Curr Genet*. 2019;65:1367–81.
7. Gallego del Sol F, Marina A. Structural basis of Rap phosphatase inhibition by Phr peptides. *PLoS Biol*. 2013;11:e1001511.
8. Reizer J, Reizer A, Perego M, Saier MH. Characterization of a family of bacterial response regulator aspartyl-phosphate (RAP) phosphatases. *Micro Comp Genomics*. 1997;2:103–11.
9. 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*. 2016;428:2793–804.
10. Pottathil M, Lazazzera BA. The extracellular PHR peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. *Front Biosci*. 2003;8:913.
11. 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*. 2019;37:420–3.
12. 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*. 2016;44:W16.
13. Trach K, Burbulys D, Strauch M, Wu JJ, Dhillon N, Jonas R, et al. Control of the initiation of sporulation in *Bacillus subtilis* by a phosphorelay. *Res Microbiol*. 1991;142:815–23.
14. Molle V, Fujita M, Jensen ST, Eichenberger P, González-Pastor JE, Liu JS, et al. The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol*. 2003;50:1683–701.
15. Osburne MS, Sonenshein AL. Behavior of a temperate bacteriophage in differentiating cells of *Bacillus subtilis*. *J Virol*. 1976;19:26–35.
16. 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 ϕ 29. *EMBO J*. 2005;24:3647–57.
17. Comella N, Grossman AD. Conservation of genes and processes controlled by the quorum response in bacteria: characterization of genes controlled by the quorum-sensing transcription factor ComA in *Bacillus subtilis*. *Mol Microbiol*. 2005;57:1159–74.
18. 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 USA*. 2009;106:6459–64.
19. Kalamara M, Spacapan M, Mandic-Mulec I, Stanley-Wall NR. Social behaviours by *Bacillus subtilis*: quorum sensing, kin discrimination and beyond. *Mol Microbiol*. 2018;110:863.
20. Tan D, Hansen MF, de Carvalho LN, Røder HL, Burmølle M, Middelboe M, et al. High cell densities favor lysogeny: induction of an H20 prophage is repressed by quorum sensing and enhances biofilm formation in *Vibrio anguillarum*. *ISME J*. 2020;14:1731–42.
21. Patterson AG, Jackson SA, Taylor C, Evans GB, Salmond GPC, Przybilski R, et al. Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Mol Cell*. 2016;64:1102–8.
22. Høyland-Kroghsbo NM, Maerkedahl RB, Svenningsen SLO. A quorum-sensing-induced bacteriophage defense mechanism. *MBio*. 2013;4:e00362–12.
23. Høyland-Kroghsbo NM, Paczkowski J, Mukherjee S, Broniewski J, Westra E, Bondy-Denomy J, et al. Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci USA*. 2017;114:131–5.

Figure S1

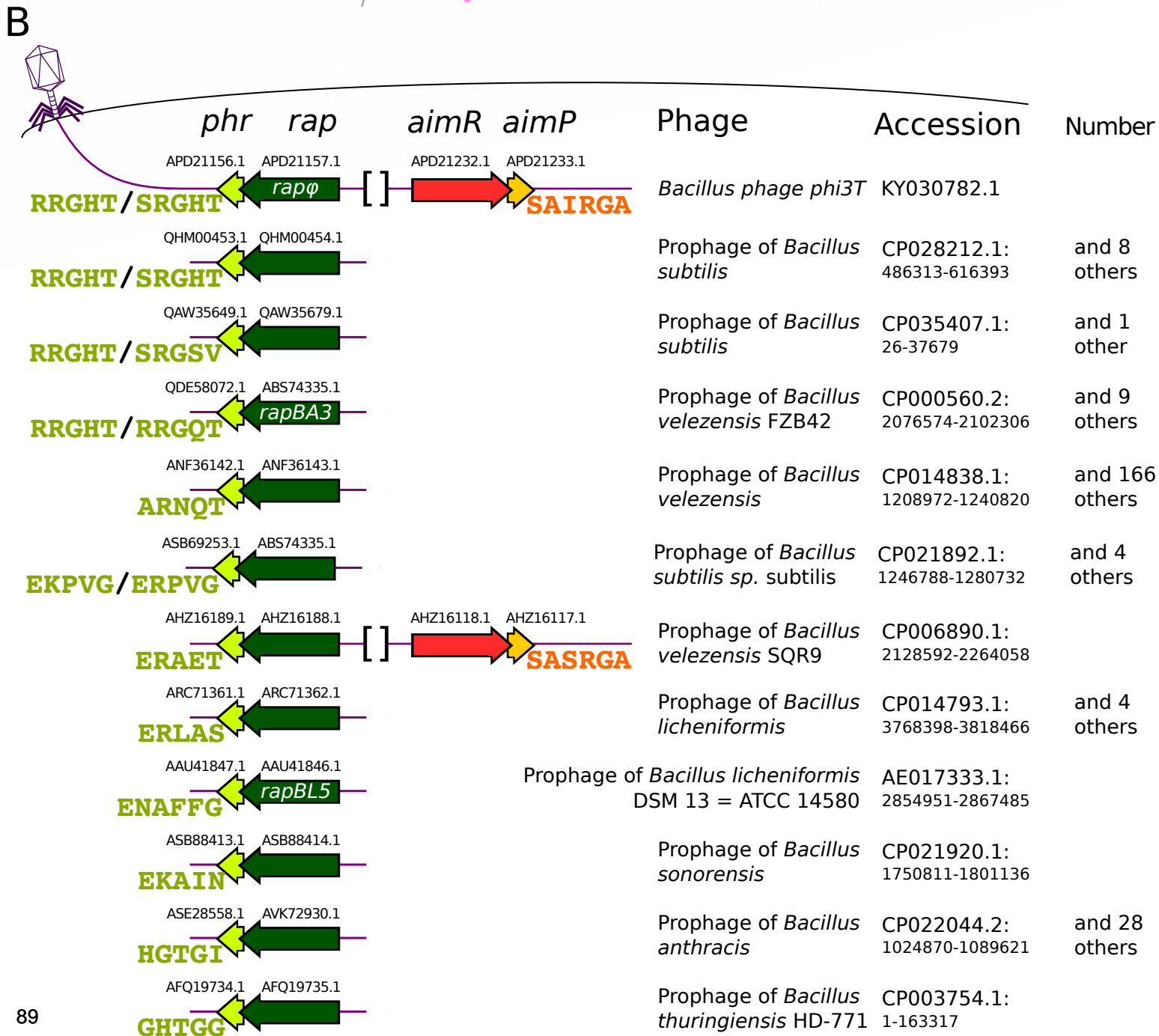
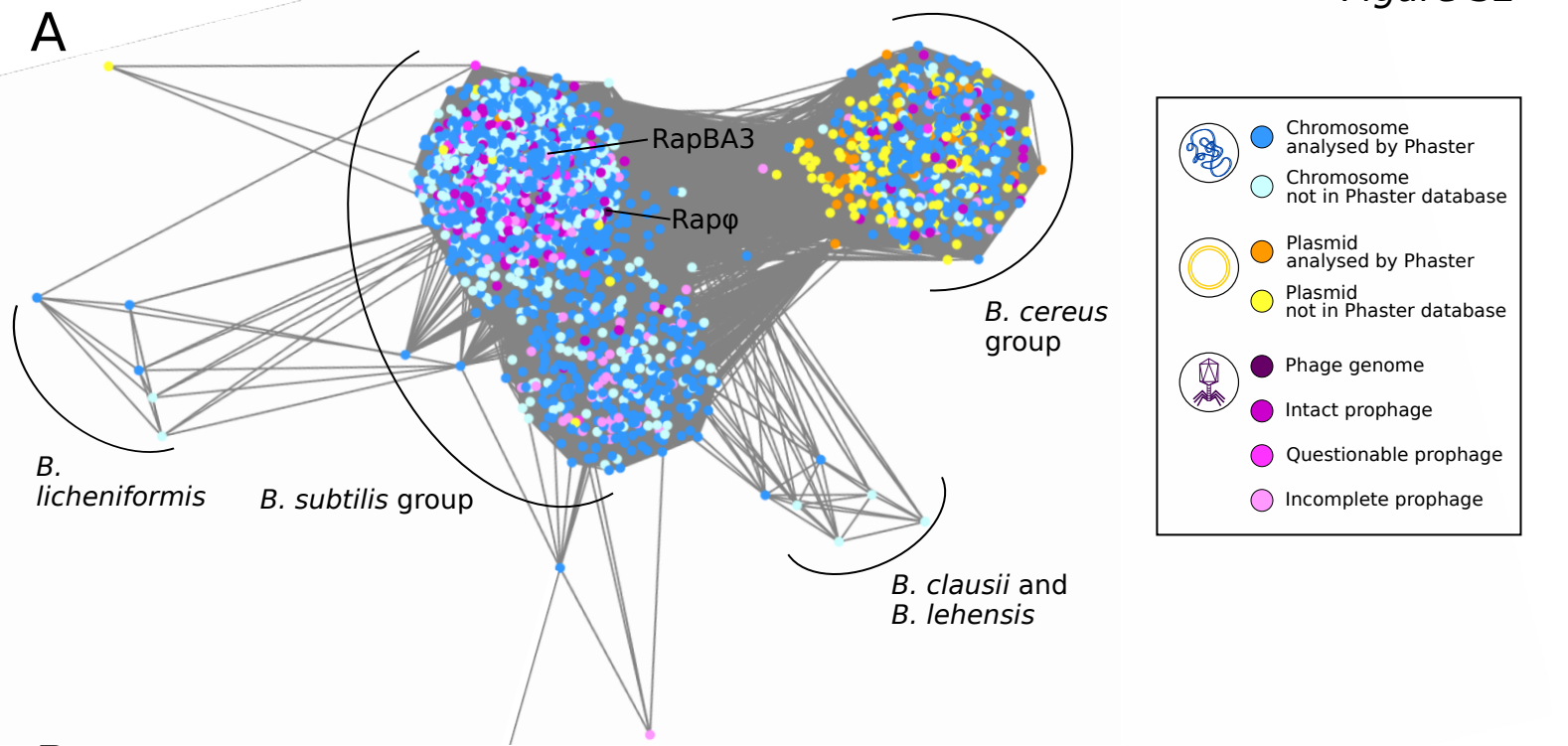
Replicate 1
Replicate 2
Replicate 3



1 **Figure S1: Coverage profile and alignments of RNA-Seq reads to the *rapφ-phrφ* locus.**

2 Each horizontal box corresponds to a replicate of RNA-seq reads, sequenced by Erez et al.
3 (1) after 20min of infection of *Bacillus* str. BEST7007 by phage phi3T. For each replicate, the
4 coverage profile is displayed at the top of the box, and the first 11 lines of the alignment at
5 the bottom. The two segments at the bottom of the figure delineate the overlapping *phrφ* and
6 *rapφ* genes.

Figure S2



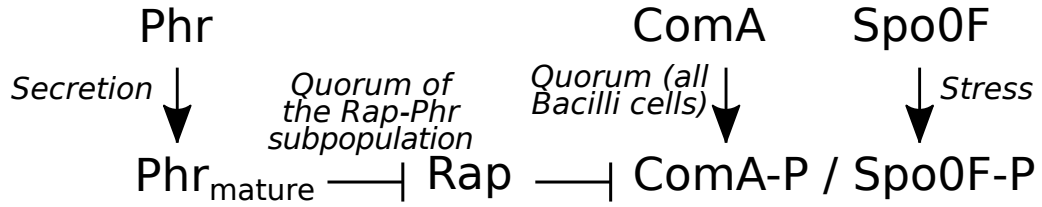
7 **Figure S2: Abundance and diversity of Rap-Phr systems in Bacillus' prophages.**

8 **A.** Sequence similarity network of the Rap proteins that form a Rap-Phr system in the
9 complete genomes of Firmicutes. Each node corresponds to a Rap protein and each edge
10 between two nodes (two Raps) is supported by a sequence identity > 30% and a mutual
11 coverage > 80%. The nodes are colored according to the type of DNA that encodes the Rap-
12 Phr system, as displayed by the legend. Specifically, there are 1849 chromosomal Rap-Phr
13 systems (including 1330 asserted by Phaster as not belonging to a prophage), 178 plasmidic
14 Rap-Phr systems (including 43 asserted by Phaster as not belonging to a prophage), 1 Rap-
15 Phr system in a phage genome (Rap ϕ -Phr ϕ) and 333 Rap-Phr systems predicted by Phaster
16 as belonging to prophages (160 to intact prophages, 59 to questionable prophages and 114
17 to incomplete prophages).

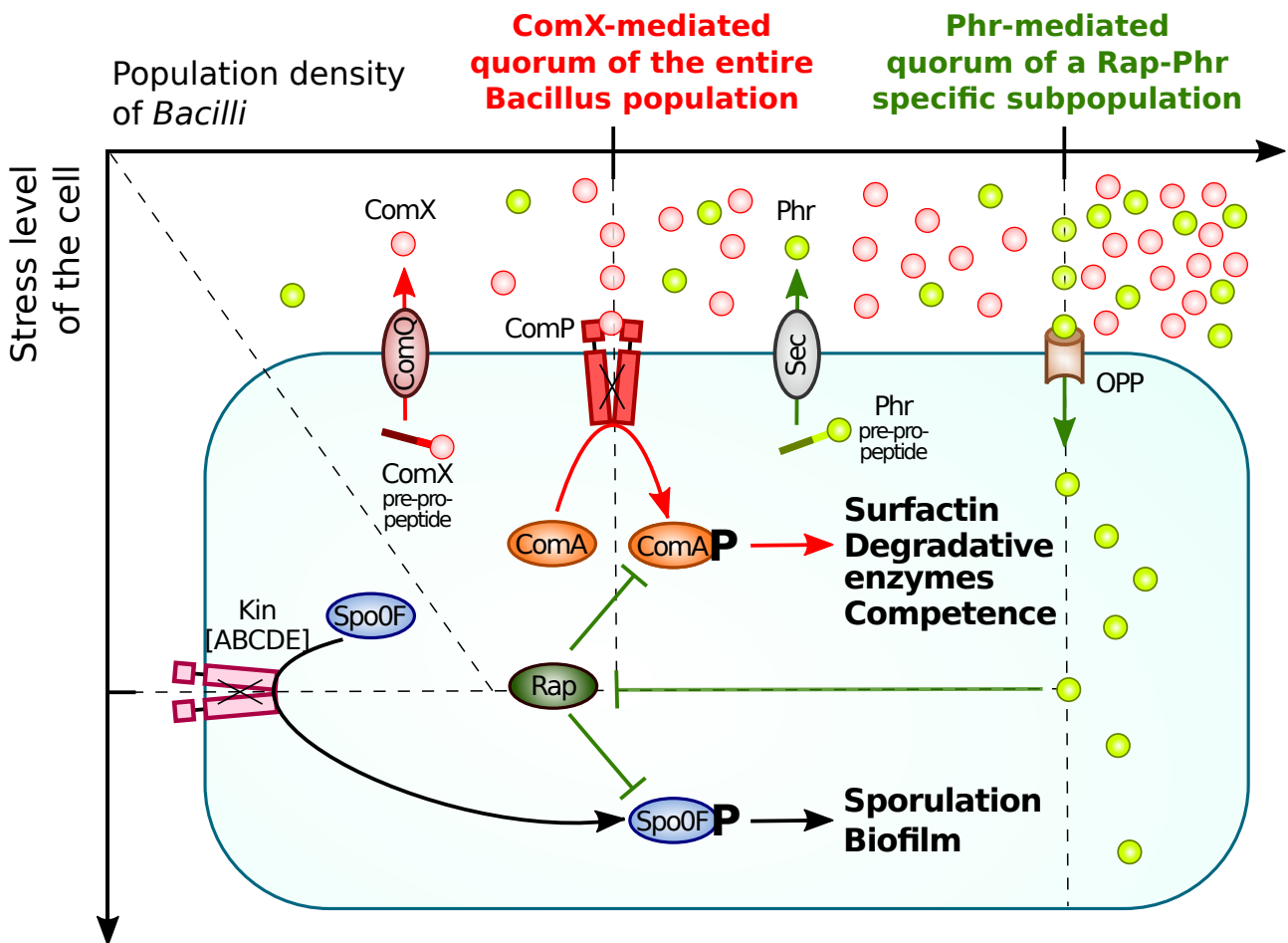
18 **B.** Overview of the diversity of the Rap-Phr systems carried by prophages. Each line
19 corresponds to the DNA of a phage or a prophage that encodes a Rap-Phr QS system and
20 that may as well encode an arbitrium QS system. Each shown prophage is representative of
21 a given type of Rap-Phr system, as determined by the sequence(s) of the predicted Phr_{mature}
22 QS peptide(s). The number on the right side of each line gives the number of prophages that
23 a displayed prophage is representative of. For each prophage, the figure shows the predicted
24 sequence(s) of the Phr_{mature} and AimP_{mature} QS peptide(s), the NCBI ID of each QS protein,
25 the host bacterial genome and the coordinates of the prophage in this bacterial genome
26 (predicted by Phaster). The RapBA3-PhrBA3 and the RapBL5-PhrBL5 QS systems that have
27 previously been functionally validated by Even-Tov et al (2) are also indicated.

Figure S3

A



B

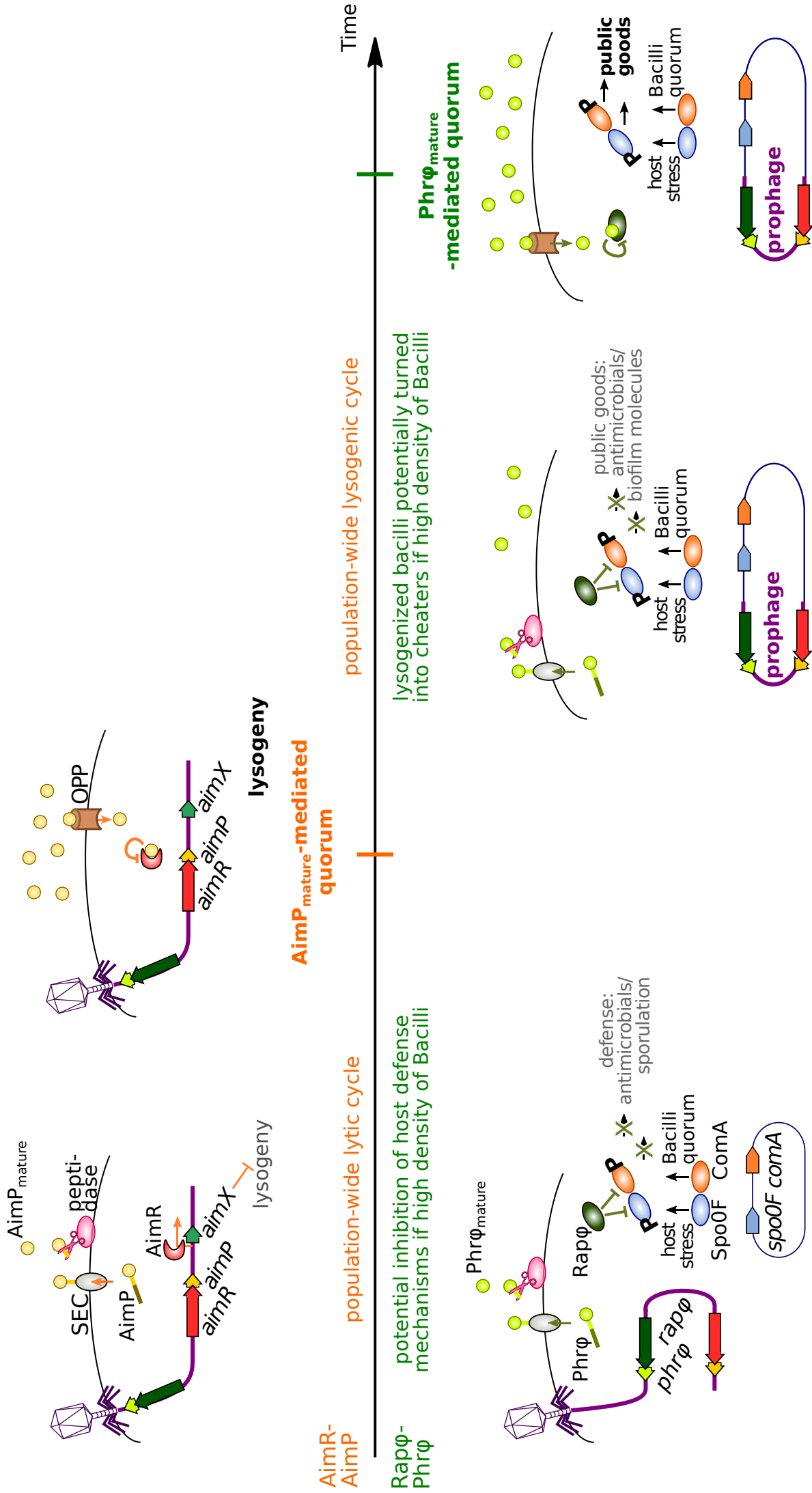


28 **Figure S3: Typical mechanism of a Rap-Phr QS system in *Bacillus* bacteria**

29 **A.** Wiring diagram of the molecular mechanism of a Rap-Phr system.

30 **B.** Illustration of the molecular mechanism of a Rap-Phr system within a *Bacillus* cell. On the
31 top hemisphere of the cell and from the left to the center, is schematized the ComX-based
32 Quorum Sensing mechanism in *Bacillus* bacteria. The different steps of this mechanism are
33 depicted by red arrows. The ComX pre-pro-peptide is exported in the medium and processed
34 by ComQ to mature to ComX_{mature} QS peptide. When the concentration of ComX_{mature}
35 reaches a threshold and reflects the quorum of the mixed population of Bacilli, it robustly
36 binds to the ComP histidine kinase, which then transfers a phosphate to ComA, the response
37 regulator of the ComX-based QS response. Once phosphorylated, ComA-P triggers the
38 expression of genes related to competence and to the production of antimicrobials and
39 degradative enzymes. From the topleft to the bottom of the cell is schematized the typical
40 mechanism of the stress response. Upon receiving a stress stimulus, stress kinases
41 phosphorylate the Spo0F protein, which thereupon initiates a phosphorelay chain through
42 several intermediate proteins that eventually triggers sporulation. From the center to the right
43 of the cell is schematized the typical mechanism of a Rap-Phr QS system. Green arrows
44 highlight the different steps of this mechanism. Below the Phr-quorum, the intracellular Rap
45 receptor antagonizes either or both of the ComA-P and Spo0F-P response regulators, either
46 through dephosphorylation or physical sequestration. Meanwhile, the Phr protein is exported
47 by the Sec secretion system and is cleaved by exopeptidases to produce a mature QS
48 peptide. When the Phr_{mature} QS peptide reaches a threshold concentration, it is imported by
49 the OPP permease, and further sequesters the Rap intracellular receptors, thereby releasing
50 the inhibition of the ComA-P and Spo0F-P regulators. The quorum mediated by Phr_{mature} often
51 intervenes after the quorum mediated by ComX_{mature} because a Rap-Phr system is often
52 expressed by a subpopulation of cells, unlike the ComX-ComP system which is expressed by
53 all Bacilli.

Figure S4



54 **Figure S4: Mechanism of the arbitrium QS system and proposed mechanism of the**
55 **Rap ϕ -Phr ϕ QS system in Bacillus phage phi3T.**

56 On the top is displayed the characterized mechanism of the arbitrium QS system of phi3T
57 (with the different steps controlled by the system colored in orange) (1). At low density of
58 phages, unbound AimR indirectly inhibits lysogeny while the AimP_{mature} peptide is secreted
59 and accumulates in the medium. At high density of phages, AimP_{mature} is imported by the Opp
60 permease, binds to AimR, and alleviates lysogeny inhibition. Thereby, the arbitrium QS
61 system ensures that the lytic cycle continues as long as the phages are few and the phage-
62 host collective is not critically threatened by the phage infection. On the bottom is displayed
63 the predicted mechanism of the Rap ϕ -Phr ϕ QS system of phi3T (with the different steps
64 controlled by the system colored in green), in the proposed scenario where the quorum
65 mediated by AimP_{mature} is reached before the quorum mediated by Phr ϕ _{mature}. At low density
66 of phages, Rap ϕ is predicted to enhance the efficiency of the phi3T lytic cycle. Indeed, Rap ϕ
67 is predicted to inhibit ComA and Spo0F and their target pathways linked to defense
68 mechanisms in the host cell, but only when these two response regulators are activated by
69 phosphorylation, upon stress and high cellular density of the entire population of Bacilli.
70 When phages are abundant, the lysogeny is triggered by the arbitrium system across the
71 population of phages. Afterwards, phi3T is internalized under the form of the prophage. The
72 Rap ϕ protein of the prophage might then confer an immediate advantage upon the
73 lysogenized bacterium, enhancing the fitness of both the host and the prophage. Indeed, the
74 inhibition of ComA-P and Spo0F-P as well as their target pathways linked to the production
75 of antimicrobials and biofilm molecules might turn the lysogenized bacteria into cheaters. At
76 high density of the lysogenized subpopulation, Phr ϕ _{mature} is predicted to inhibit Rap ϕ , which,
77 in turn, might alleviate ComA-P and Spo0F-P inhibition and repress the cheating phenotype.

78 MATERIALS AND METHODS

79 Detection of the Rap ϕ -Phr ϕ QS system

80 The QS receptor, Rap ϕ , was detected with Blastp (version 2.2.3+) (3) using, as a query,
81 the protein sequences of reference RRNPP QS receptors and, as a target, all the protein
82 sequences from the complete genomes of viruses that were available on the NCBI as of
83 18/07/19. The query was constructed based on the entries corresponding to RRNPP QS
84 systems on the Quorumpeps database of QS signaling peptides (4) (NCBI accessions
85 ANC26365.1 (ComR), AVD29695.1 (NprR), AAD41788.1 (PlcR), AAA65845.1 (PrgX),
86 BAA11197.1 (TraA), ABQ95981.1 (Rgg), ABJ66531.1 (Rgg1358), NP_38125.1,
87 NP_391550.1, NP_388259.1, NP_391519.1, NP_390460.2, NP_391626.1, NP_391910.1,
88 NP_388565.2, NP_388382.1, NP_388164.1 (RapA to RapJ)). Rap ϕ was also independently
89 detected with hmmsearch from the HMMER suite (version 3.2.1) (5), using, as query, a
90 concatenation of the canonical HMM profiles of tetratricopeptide repeats that match the
91 members of the RRNPP family of QS receptors (PF13181, PF13424, PF18710, PF18768
92 (Pfam models), SM00028, SM00530 (Smart models), 1.25.40.10 and 1.25.40.400 (CATH
93 models), SSF48452 (Supfam models)). The QS pre-pro-peptide, Phr ϕ , has been detected as
94 a signal peptide containing protein by SignalP (version 5.0b Linux x86_64) (6) with the key
95 option '-org gram+', using as input the protein sequences encoded by the upstream and
96 downstream genes of *rap ϕ* . Phr ϕ was further characterized as a Phr protein on the basis of
97 the quality of its alignment with canonical Phr proteins (see below).

98 Evaluation of the genomic context of the *rap ϕ -phr ϕ* tandem of genes

99 The map of the Phi3T genome was produced with DNAplotter (release 1.11) (7), using as
100 input the Genbank annotation file of the phi3T genome retrieved from the NCBI. The genes
101 of phi3T highlighted as 'similar in SPbeta' were annotated as 'similar to SPBeta' in the
102 annotation file. The functional annotation of the surrounding genes of the QS systems in
103 Figure 1 relied either on the annotations present in the annotation file whenever they were

104 available, or were deduced from their constitutive domains, as detected by interproscan
105 (version 5.36-75.0). The positions of intrinsic (or rho-independent) terminators in the genome
106 of *Bacillus phage phi3T* were predicted by RNIE (8), configured with infernal (version 1.02)
107 and launched independently with the '-genome' specific and the '-gene' sensitive modes and
108 with a bitscore threshold of 14.

109 **Multiple sequence alignment (MSA) of Rap and Phr proteins**

110 The protein sequence of Rap ϕ was aligned with RapBA3 as well as with the Rap proteins
111 from *B. subtilis strain 168* for which the structure has been resolved in the study of Gallego
112 del Sol and Marina (9). This MSA was done with Muscle (version 3.8.31) (10) and visualized
113 with Jalview version (2.10.5) (11) using the « % identity» colorcode to color the residues. The
114 top annotation of the residues strictly corresponds to the one displayed by Gallego del Sol
115 and Marina in their structure-guided MSA of *B. subtilis* Rap proteins. The whole protein
116 sequence of Phr ϕ as well as its duplicated C-terminus have been aligned with PhrBA3, the
117 duplicated C-terminus of PhrBA3, and all the canonical Phr proteins considered in Pottathil et
118 al (12). The MSA was performed with Muscle, further manually adjusted, and visualized with
119 Jalview (11) under the «Clustal» colorcode for residues (See
120 <http://www.jalview.org/help/html/colourSchemes/clustal.html>). The mature sequences of Phr
121 were inferred manually from this MSA.

122 **Mapping of RNA-Seq reads to the phi3T genome**

123 Three replicates of RNA-Seq reads sequenced by Erez et al. (1) after 20min of infection of
124 *Bacillus subtilis* strain BEST7003 by phi3T were retrieved (Sample accessions
125 SAMEA24223918, SAMEA24226918 and SAMEA24229918) from the European Nucleotide
126 Database. For each replicate, the reads were trimmed using fastx_clipper with default
127 parameters (13). Then, we classically aligned the reads to the phi3T genome (NCBI
128 accession KY030782) with Bowtie (14), allowing only one nucleotide misalignment and
129 considering only the best alignment of each mapped read. The Integrative Genomics Viewer

130 was used to visualize both the coverage profile and the alignment of the reads to the *rapφ-*
131 *phrφ* locus (15).

132 **Identification of Rap-Phr systems in Bacillus' prophages**

133 First, homologs of Rapφ were searched with Blastp (version 2.2.3+) (3) against all the
134 protein sequences encoded by the complete genomes of Firmicutes available on the NCBI. A
135 homolog was defined as a protein with no less than 30% sequence identity with a query
136 protein, over more than 80% mutual coverage. The Rap homologs found by Blastp at each
137 iteration were allowed to serve as queries in a next iteration, as long as the remote homologs
138 exhibit 80% mutual coverage with Rapφ. Then, very stringent thresholds were applied to
139 retain the Rap homologs that (i) harbor TPR repeats (E-value < 10⁻⁵ for at least one of the
140 TPR HMM models cited above), (ii) do not have a DNA binding domain since Rap is not a
141 transcription factor (E-value < 10⁻⁵ for any of the CATH 1.10.260.40, PFAM PF011381,
142 SupFam SSF47413 and SMART SM00530 HMM models), (iii) are followed, on the same
143 DNA strand and with an intergenic distance < 150bp, by a gene encoding a Phr protein, as
144 defined by a protein of length between 15aa and 65aa and yielding a SEC-secretion
145 likelihood score > 0.25 according to SignalP-5.0 (6). Next, for all the genomes identified as
146 encoding at least one Rap-Phr system, the Phaster database (16) was queried to identify
147 whether these genomes are composed of known prophage regions. At the time of the
148 analysis, it was impossible to submit new queries to Phaster due to maintenance of the
149 server, but this information was available for the genomes that had previously been queried
150 by Phaster's users. If the coordinates of a *rap-phr* system fell within the coordinates of a
151 prophage predicted by Phaster, the QS system was assessed as belonging to a prophage.
152 The detailed results (with scores and NCBI IDs) of this analysis are given in Table S2.

153 **Sequence similarity network of the Rap proteins**

154 The sequence similarity network was built based on the output of the Blastp "all versus all"
155 of the 2361 Rap proteins identified as forming a Rap-Phr QS system. An edge between two

156 nodes was drawn if the sequence identity between two Rap proteins was > 30% over more
157 than 80% mutual coverage. The network was visualized with the Cytoscape software (17).

158 REFERENCES

- 159 1. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, et al.
160 Communication between viruses guides lysis-lysogeny decisions. *Nature*.
161 2017;541(7638):488–93.
- 162 2. Even-Tov E, Omer Bendori S, Pollak S, Eldar A. Transient Duplication-Dependent Divergence
163 and Horizontal Transfer Underlie the Evolutionary Dynamics of Bacterial Cell–Cell Signaling.
164 Gore J, editor. *PLoS Biol*. 2016 Dec 29;14(12):e2000330.
- 165 3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol*
166 *Biol*. 1990 Oct 5;215(3):403–10.
- 167 4. Wynendaele E, Bronselaer A, Nielandt J, D’Hondt M, Stalmans S, Bracke N, et al.
168 Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing
169 peptides. *Nucleic Acids Res*. 2013 Jan 1;41(D1):D655–9.
- 170 5. Eddy SR. Accelerated Profile HMM Searches. Pearson WR, editor. *PLoS Comput Biol*. 2011
171 Oct 20;7(10):e1002195.
- 172 6. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al.
173 SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*.
174 2019 Apr 18;37(4):420–3.
- 175 7. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear
176 interactive genome visualization. *Bioinformatics*. 2009 Jan 1;25(1):119–20.
- 177 8. Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z. RNIE: genome-wide prediction
178 of bacterial intrinsic terminators. *Nucleic Acids Res*. 2011 Aug 1;39(14):5845–52.
- 179 9. Gallego del Sol F, Marina A. Structural Basis of Rap Phosphatase Inhibition by Phr Peptides.
180 Laub MT, editor. *PLoS Biol*. 2013 Mar 19;11(3):e1001511.
- 181 10. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
182 *Nucleic Acids Res*. 2004;32(5):1792–7.
- 183 11. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple
184 sequence alignment editor and analysis workbench. *Bioinformatics*. 2009 May 1;25(9):1189–
185 91.
- 186 12. Pottathil M, Lazazzera BA. The extracellular PHR peptide-Rap phosphatase signaling circuit of
187 *bacillus subtilis*. *Front Biosci*. 2003 Jan 1;8(4):913.
- 188 13. FASTX-Toolkit [Internet]. [cited 2020 Aug 14]. Available from:
189 http://hannonlab.cshl.edu/fastx_toolkit/
- 190 14. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of
191 short DNA sequences to the human genome. *Genome Biol*. 2009 Mar 4;10(3):R25.
- 192 15. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative
193 genomics viewer. *Nat Biotechnol*. 2011 Jan 1;29(1):24–6.
- 194 16. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version

- 195 of the PHAST phage search tool. *Nucleic Acids Res.* 2016;44(Web Server issue):W16.
- 196 17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software
197 environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003
198 Nov;13(11):2498–504.
- 199

4.2. Détection de nouveaux gènes de QS par une nouvelle méthode basée sur une signature de critères génomiques, structurels et biochimiques

Jusqu'alors, les gènes de QS identifiés au sein de génomes d'intérêt étaient des homologues de composantes de QS déjà connues. Néanmoins, l'article précédent a commencé à présenter avec le système Rap-Phr quelques caractéristiques génériques propres aux différents systèmes de type RRNPP. En effet, la présence de TPRs impliqués dans l'interaction avec de petits peptides au sein du récepteur Rap, la présence d'une partie N-terminale d'adressage au translocon SEC au sein du propeptide Phr ainsi que l'adjacence des gènes *rap* et *phr* sont des propriétés qui sont communes à la majorité des systèmes RRNPP (Erez *et al.*, 2017; Neiditch *et al.*, 2017; Voichek *et al.*, 2020). Ces propriétés partagées par les différents QSSs de type RRNPP en dépit de leur diversité génétique importante sont inhérentes au mécanisme moléculaire commun à tous ces systèmes. Les différents propeptides de type RRNPP sont en effet tous sécrétés, à part de rares exceptions (e.g. Shp et PrgX) (**Figure 9**), par le même système de sécrétion SEC, sont réimportés par la perméase OPP, et interagissent avec les TPRs de leur récepteur associé, induisant alors chez ce dernier un changement conformationnel à l'origine de la réponse au QS (Neiditch *et al.*, 2017). Ces propriétés génériques mais néanmoins spécifiques au mécanisme de QS de type RRNPP peuvent alors être exploitées pour détecter de nouveaux QSSs, non homologues à des QSSs déjà connus. L'article suivant en fait la démonstration, et présente le programme RRNPP_detector développé pour détecter des systèmes RRNPP encore inconnus au sein de chromosomes, plasmides et phages de Firmicutes. Cette article en instance de soumission présente également les découvertes révélées par l'application à grande échelle de RRNPP_detector contre tous les génomes complets de Firmicutes et de virus disponibles sur le NCBI.

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 Applying RRNPP_detector against complete genomes of viruses and Firmicutes of the NCBI, we
32 report a potential 7.5-fold expansion of RRNPP QSS diversity, alternative secretion-modes for
33 certain candidate QSS propeptides, ‘bilingual’ bacteriophages and plasmids, as well as predicted
34 chromosomal and plasmidic Biosynthetic-Gene-Clusters regulated by quorum sensing.

35

36 **INTRODUCTION**

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

53

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

62
63 Interestingly, the case of the plasmidic PrgX system illustrates that RRNPP QSSs may not only be
64 present on bacterial chromosomes but also on mobile genetic elements (MGEs). Consistently, the
65 conjugation-regulating TraA-Ipd1 system have later on expanded the list of plasmidic RRNPP
66 QSSs (17). In 2017, Erez *et al.* even discovered that some temperate bacteriophages encode
67 RRNPP QSSs, namely the “arbitrium” systems that guide the lysis-lyogeny decision upon *Bacillus*
68 infection (18,19). Recently, two other chromosomal RRNPP QSSs have been experimentally
69 validated: the QsrB-QspB system that delays sporulation and solvent formation in *Clostridium*
70 *acetobutylicum* (20) and the AloR13-AloP13 sporulation-regulator in *Paenibacillus polymyxa* (21).
71 As important the experimentally-validated RRNPP QSSs are in the regulation of the biology of their
72 encoding microbial entity, they represent only the tip of the iceberg. Indeed, the diversity of RRNPP
73 QSSs is likely not fully explored, as hinted by the numerous candidate RRNPP QSS receptors in
74 *Clostridium* and *Enterococcus* species reported to harbor local similarities with regions of Rap,
75 NprR, PlcR or Rgg (20,22,23). Uncovering novel RRNPP QSS families would unveil new
76 bacterium-bacterium, plasmid-plasmid or phage-phage modes of communication but most and
77 foremost reveal novel density-dependent evolutionary strategies that could transform our views of
78 microbial interaction, adaptation and evolution. Expanding RRNPP QSS diversity would also likely
79 pave way to major practical outcomes as novel RRNPP QSSs could regulate the production of
80 novel antimicrobial compounds (24,25) or could underlie mechanisms by which some human
81 pathogens acquire virulence (26–29).

82
83 Conveniently, we noticed that the different members of the RRNPP group share a common
84 signature of 5 criteria (**Fig. 1**): i) the pro-peptide is a small protein (10-70aa), ii) the pro-peptide is
85 secreted, except in rare exceptions (e.g. Shp and PrgQ (7)), via the SEC-translocon and further
86 matured by exopeptidases into a communication peptide, iii) the receptor has a length comprised
87 between 250 and 500aa, iv) the receptor harbors tetratricopeptide repeats (TPR), which are
88 structural motifs involved in the binding of small peptides (in this case, the cognate communication
89 peptide), v) the genes encoding the pro-peptide and the receptors are directly adjacent to each
90 other. Advantageously, a large amount of reference Hidden-Markov-Models (HMMs) from the Cath-

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

100

101 **MATERIALS AND METHODS**

102 **Overview of the RRNPP_detector workflow**

103 Given a fasta file of protein sequences and the coordinates of their coding sequences in target
104 genome(s), Metagenomics-Assembled-Genome(s) or contig(s), RRNPP_detector first reduces the
105 search space by retaining only proteins that have a length compatible with RRNPP receptors (250-
106 500aa) and pro-peptides (10-70aa). Next, only the 250-500aa proteins whose coding sequences
107 (CDSs) are found adjacent with the CDS of a 10-70aa protein are retained, and vice versa. Then,
108 HMMs of TPRs found in reference RRNPP receptors (PFAM: PF13181, PF13424; Gene 3D:
109 1.25.40.400, 1.25.40.10; TIGRFAM: TIGR01716; Superfamily: 48452, SMART: SM00028) are
110 queried with hmmsearch (36) against the remaining 250-500aa proteins. Only the proteins
111 matched by at least one HMM of TPRs (E-value < 1E-5) are retained and designated as potential
112 receptors. These potential receptors next undergo a computational characterization. First, a
113 BLASTp search (37) of reference RRNPP receptors (Rap, Rgg, NprR, PlcR, PrgX, TraA, AimR of
114 Bacillus phage phi3T, AimR of Bacillus phage Waukesha, QsrB and A1oR13) against these
115 potential receptors are launched to identify the homologs of already known QSSs (by default: E-
116 value <= 1E-5, identity >= 20%, and alignment coverage >= 60% of the length of both the query
117 and the target sequences). Then, an HMMsearch of DNA-binding domains found in reference
118 receptors (PFAM: PF01381; Gene3D: 1.10.260.40; Superfamily: 47413; SMART: SM00530) is
119 launched against the potential receptors to identify those that may be transcription factors. After
120 this characterization step, the 10-70aa proteins are further reduced to those with a CDS adjacent
121 to the one of a potential receptor. At this stage, each tandem of remaining adjacent pro-peptide and
122 receptor are considered forming a potential QSS and will eventually be outputted. However, either a
123 'permissive' or a 'conservative' label will be assigned to these candidate QSSs according to the
124 next filters. First, any QSS with an intergenic distance > 400 base pairs between the two CDSs will
125 be considered as 'permissive' because low intergenic distance tends to correlate with functional
126 association (38). Second, only the remaining QSSs with a pro-peptide predicted by SignalP (35) to

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

132

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

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

155

156 **RESULTS**

157 **RRNPP_detector reveals thousands of candidate RRNPP QSSs in complete bacterial** 158 **chromosomes, plasmids and phage genomes**

159 To illustrate RRNPP_detector, we launched it against the 4974 complete genomes of Firmicutes
160 and the 32919 complete genomes of Viruses Firmicutes available on the NCBI Assembly Database
161 (**Fig. 2**). We describe how such analyses can be easily done with the practical example of Viruses
162 in the readme file of RRNPP_detector:

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

193

194 **RRNPP_detector massively expands RRNPP QSS diversity**

195 In order to classify the 3529 conservative candidate RRNPP QSSs, the sequences of the
196 candidate receptors were BLASTed against each other to identify groups of homologs, defined as
197 connected components in the resulting sequence similarity network (wherein two nodes (proteins)
198 are connected if they show a sequence identity $\geq 30\%$ over more than 80% of the lengths of both
199 sequences, as in (44)) (**Fig. 3**). We report the identification of 31 groups of homologous receptors

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

227

228 **Post-processing functional prediction of candidate RRNPP QSS reveals 50 putative QSS-** 229 **regulated biosynthetic gene clusters**

230 QSSs with a one component receptor module tend to be located nearby their target genes (47–
231 49) and RRNPP QSSs are no exceptions to this trend, especially not MGE-encoded RRNPP QSSs
232 (as shown in **Fig. 1**). The genomic neighborhood of a RRNPP QSS can thus provide insights on
233 the potential function(s) regulated in a density-dependent by the QSS, as suggested in our
234 previous study on the functional inference of viral RRNPP QSSs
235 (<https://www.biorxiv.org/content/10.1101/2021.07.15.452460v1>). For instance, one function
236 regulated by quorum sensing typically is the production of public good metabolites, e.g.

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

266

267 **CONCLUSION AND FUTURE PERSPECTIVES**

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

274 functional insights, as shown in this article with the specific example of putative QSS-regulated
275 BGCs.

276

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

288

289 **DATA AVAILABILITY**

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

291

292 **FUNDING**

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

296

297 **REFERENCES**

- 298 1. Mukherjee S, Bassler BL. Bacterial quorum sensing in complex and dynamically changing
299 environments. *Nat Rev Microbiol* [Internet]. 2019 Jun 3 [cited 2019 Jun 12];17(6):371–82.
300 Available from: <http://www.nature.com/articles/s41579-019-0186-5>
- 301 2. Papenfort K, Bassler BL. Quorum sensing signal-response systems in Gram-negative
302 bacteria. *Nat Rev Microbiol* [Internet]. 2016 [cited 2019 May 11];14(9):576–88. Available
303 from: <http://www.ncbi.nlm.nih.gov/pubmed/27510864>
- 304 3. Polkade A V., Mantri SS, Patwekar UJ, Jangid K. Quorum sensing: An under-explored
305 phenomenon in the phylum Actinobacteria. Vol. 7, *Frontiers in Microbiology*. Frontiers Media
306 S.A.; 2016.
- 307 4. Bhatt VS. Quorum sensing mechanisms in gram positive bacteria. In: *Implication of Quorum*
308 *Sensing System in Biofilm Formation and Virulence*. Springer Singapore; 2019. p. 297–311.
- 309 5. Sturme MHJ, Kleerebezem M, Nakayama J, Akkermans ADL, Vaughan EE, De Vos WM.
310 Cell to cell communication by autoinducing peptides in gram-positive bacteria. *Antonie van*
311 *Leeuwenhoek*, *Int J Gen Mol Microbiol*. 2002;81(1–4):233–43.

- 312 6. Rocha-Estrada J, Aceves-Diez AE, Guarneros G, de la Torre M. The RNPP family of
 313 quorum-sensing proteins in Gram-positive bacteria. *Appl Microbiol Biotechnol* [Internet].
 314 2010 Jul 26 [cited 2019 Oct 16];87(3):913–23. Available from:
 315 <http://link.springer.com/10.1007/s00253-010-2651-y>
- 316 7. Neiditch MB, Capodagli GC, Prehna G, Federle MJ. Genetic and Structural Analyses of
 317 RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria [Internet]. Vol. 51, *Annual*
 318 *Review of Genetics*. Annual Reviews Inc.; 2017 [cited 2020 Nov 19]. p. 311–33. Available
 319 from: <https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-120116-023507>
- 320 8. Do H, Kumaraswami M. Structural Mechanisms of Peptide Recognition and Allosteric
 321 Modulation of Gene Regulation by the RRNPP Family of Quorum-Sensing Regulators. *J Mol*
 322 *Biol* [Internet]. 2016 Jul 17 [cited 2019 Oct 16];428(14):2793–804. Available from:
 323 <http://www.ncbi.nlm.nih.gov/pubmed/27283781>
- 324 9. Perez-Pascual D, Monnet V, Gardan R. Bacterial Cell-Cell Communication in the Host via
 325 RRNPP Peptide-Binding Regulators. *Front Microbiol* [Internet]. 2016 [cited 2019 Oct
 326 16];7:706. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27242728>
- 327 10. Baldwin VM. You Can't B. cereus – A Review of Bacillus cereus Strains That Cause Anthrax-
 328 Like Disease. Vol. 11, *Frontiers in Microbiology*. Frontiers Media S.A.; 2020. p. 1731.
- 329 11. Lannes-Costa PS, de Oliveira JSS, da Silva Santos G, Nagao PE. A current review of
 330 pathogenicity determinants of Streptococcus sp. [Internet]. *Journal of Applied Microbiology*.
 331 Blackwell Publishing Ltd; 2021 [cited 2021 Jul 26]. Available from:
 332 <https://pubmed.ncbi.nlm.nih.gov/33772968/>
- 333 12. Mallozzi M, Viswanathan VK, Vedantam G. Spore-forming Bacilli and Clostridia in human
 334 disease [Internet]. Vol. 5, *Future Microbiology*. Future Medicine Ltd.; 2010 [cited 2021 Mar
 335 12]. p. 1109–23. Available from: <https://pubmed.ncbi.nlm.nih.gov/20632809/>
- 336 13. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: A common cause of persistent
 337 infections [Internet]. Vol. 284, *Science*. American Association for the Advancement of
 338 Science; 1999 [cited 2021 Jul 26]. p. 1318–22. Available from:
 339 <https://science.sciencemag.org/content/284/5418/1318>
- 340 14. Høiby N, Ciofu O, Johansen HK, Song ZJ, Moser C, Jensen PØ, et al. The clinical impact of
 341 bacterial biofilms. In: *International Journal of Oral Science* [Internet]. *Int J Oral Sci*; 2011
 342 [cited 2021 Jul 26]. p. 55–65. Available from: <https://pubmed.ncbi.nlm.nih.gov/21485309/>
- 343 15. Galié S, García-Gutiérrez C, Miguélez EM, Villar CJ, Lombó F. Biofilms in the food industry:
 344 Health aspects and control methods [Internet]. Vol. 9, *Frontiers in Microbiology*. Frontiers
 345 Media S.A.; 2018 [cited 2021 Jul 26]. Available from: [/pmc/articles/PMC5949339/](https://pubmed.ncbi.nlm.nih.gov/31549339/)
- 346 16. Von Wintersdorff CJH, Penders J, Van Niekerk JM, Mills ND, Majumder S, Van Alphen LB,
 347 et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal
 348 gene transfer [Internet]. Vol. 7, *Frontiers in Microbiology*. Frontiers Media S.A.; 2016 [cited
 349 2021 Jul 26]. p. 173. Available from: [/pmc/articles/PMC4759269/](https://pubmed.ncbi.nlm.nih.gov/2759269/)
- 350 17. Kohler V, Keller W, Grohmann E. Regulation of gram-positive conjugation [Internet]. Vol. 10,
 351 *Frontiers in Microbiology*. Frontiers Media S.A.; 2019 [cited 2021 Mar 12]. p. 1134. Available
 352 from: www.frontiersin.org

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

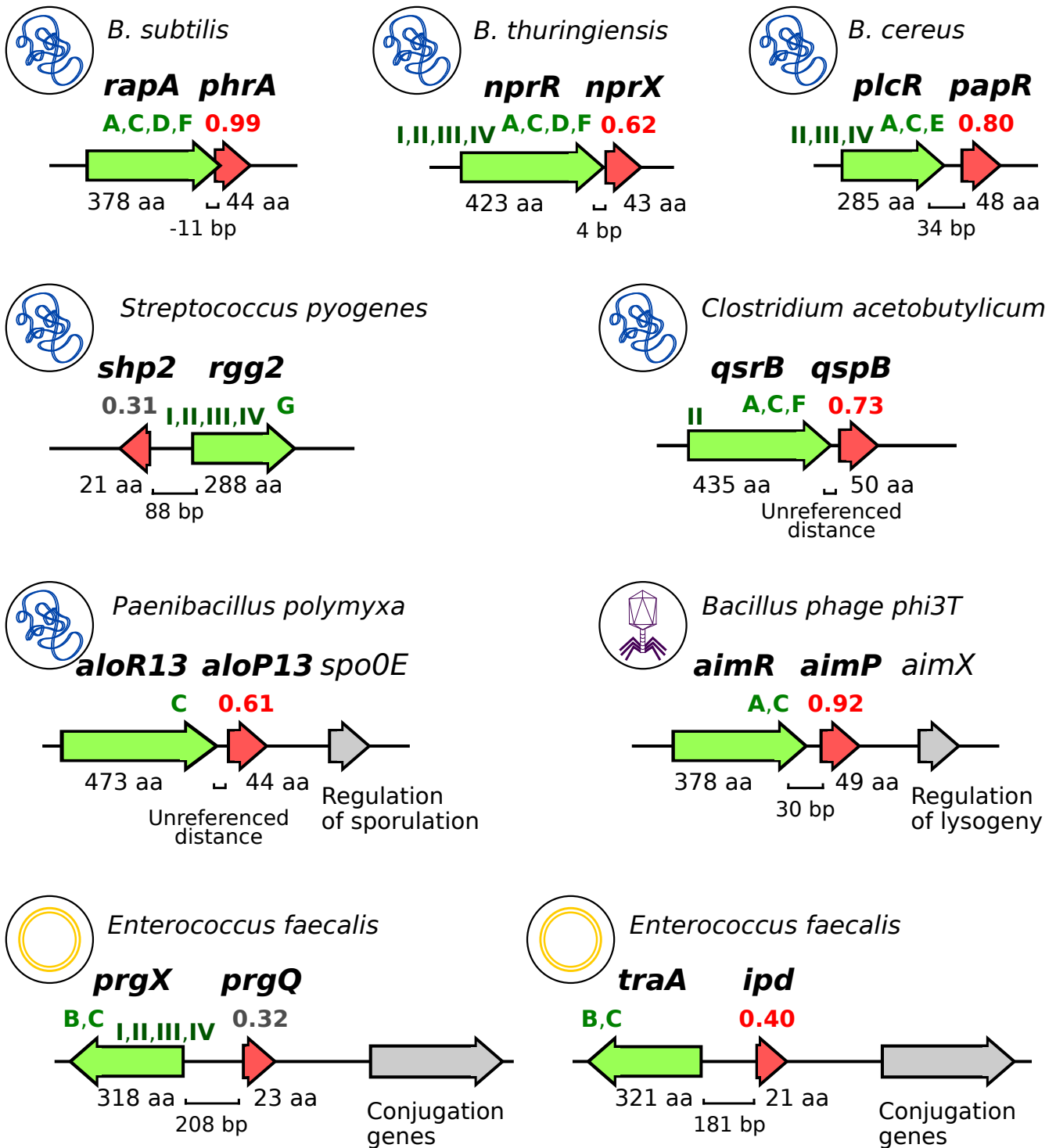
- 393 29. Do H, Makthal N, VanderWal AR, Saavedra MO, Olsen RJ, Musser JM, et al. Environmental
394 pH and peptide signaling control virulence of *Streptococcus pyogenes* via a quorum-sensing
395 pathway. *Nat Commun* [Internet]. 2019 Dec 1 [cited 2021 Jul 26];10(1):1–14. Available from:
396 <https://doi.org/10.1038/s41467-019-10556-8>
- 397 30. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, et al. CATH: Increased
398 structural coverage of functional space. *Nucleic Acids Res* [Internet]. 2021 Jan 8 [cited 2021
399 Jul 26];49(D1):D266–73. Available from: [/pmc/articles/PMC7778904/](https://pubmed.ncbi.nlm.nih.gov/3541467/)
- 400 31. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY -
401 Sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic*
402 *Acids Res* [Internet]. 2009 Jan 1 [cited 2021 Jul 26];37(SUPPL. 1):D380–6. Available from:
403 https://academic.oup.com/nar/article/37/suppl_1/D380/1009184
- 404 32. Letunic I, Khedkar S, Bork P. SMART: Recent updates, new developments and status in
405 2020. *Nucleic Acids Res* [Internet]. 2021 Jan 8 [cited 2021 Jul 26];49(D1):D458–60.
406 Available from: <https://academic.oup.com/nar/article/49/D1/D458/5940513>
- 407 33. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al.
408 Pfam: The protein families database in 2021. *Nucleic Acids Res* [Internet]. 2021 Jan 8 [cited
409 2021 Jul 26];49(D1):D412–9. Available from: <https://covid-19>.
- 410 34. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and genome
411 properties in 2013. *Nucleic Acids Res* [Internet]. 2013 Jan 1 [cited 2021 Jul 26];41(D1).
412 Available from: <https://pubmed.ncbi.nlm.nih.gov/23197656/>
- 413 35. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et
414 al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat*
415 *Biotechnol* [Internet]. 2019 Apr 18 [cited 2019 Oct 16];37(4):420–3. Available from:
416 <http://www.nature.com/articles/s41587-019-0036-z>
- 417 36. Eddy SR. Accelerated Profile HMM Searches. Pearson WR, editor. *PLoS Comput Biol*
418 [Internet]. 2011 Oct 20 [cited 2019 May 28];7(10):e1002195. Available from:
419 <https://dx.plos.org/10.1371/journal.pcbi.1002195>
- 420 37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J*
421 *Mol Biol* [Internet]. 1990 Oct 5 [cited 2019 Oct 16];215(3):403–10. Available from:
422 <http://www.ncbi.nlm.nih.gov/pubmed/2231712>
- 423 38. Junier I, Unal EB, Yus E, Lloréns-Rico V, Serrano L. Insights into the Mechanisms of Basal
424 Coordination of Transcription Using a Genome-Reduced Bacterium. *Cell Syst*. 2016 Jun
425 22;2(6):391–401.
- 426 39. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic
427 gene recognition and translation initiation site identification. *BMC Bioinformatics* [Internet].
428 2010 Mar 8 [cited 2021 Jul 21];11. Available from:
429 <https://pubmed.ncbi.nlm.nih.gov/20211023/>
- 430 40. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster
431 version of the PHAST phage search tool. *Nucleic Acids Res* [Internet]. 2016 [cited 2020 Aug
432 4];44(Web Server issue):W16. Available from:
433 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987931/>

- 434 41. Even-Tov E, Omer Bendori S, Pollak S, Eldar A. Transient Duplication-Dependent
435 Divergence and Horizontal Transfer Underlie the Evolutionary Dynamics of Bacterial Cell–
436 Cell Signaling. Gore J, editor. PLOS Biol [Internet]. 2016 Dec 29 [cited 2019 Oct
437 17];14(12):e2000330. Available from: <http://dx.plos.org/10.1371/journal.pbio.2000330>
- 438 42. Kalamara M, Spacapan M, Mandic-Mulec I, Stanley-Wall NR. Social behaviours by *Bacillus*
439 *subtilis*: quorum sensing, kin discrimination and beyond. Mol Microbiol [Internet]. 2018 [cited
440 2019 Oct 16];110(6):863. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30218468>
- 441 43. Bernard C, Li Y, Lopez P, Bapteste E. Beyond arbitrium: identification of a second
442 communication system in *Bacillus* phage phi3T that may regulate host defense
443 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)
444 00795-9
- 445 44. Méheust R, Watson AK, Lapointe FJ, Papke RT, Lopez P, Bapteste E. Hundreds of novel
446 composite genes and chimeric genes with bacterial origins contributed to haloarchaeal
447 evolution. Genome Biol. 2018 Jun 7;19(1).
- 448 45. Corel E, Méheust R, Watson AK, McInerney JO, Lopez P, Bapteste E. Bipartite Network
449 Analysis of Gene Sharings in the Microbial World. Mol Biol Evol [Internet]. 2018 [cited 2019
450 Nov 24];35(4):899–913. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29346651>
- 451 46. Silpe JE, Bassler BL. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage
452 Lysis-Lysogeny Decision. Cell [Internet]. 2019 Jan 10 [cited 2019 Jun 12];176(1–2):268-
453 280.e13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30554875>
- 454 47. Engebrecht J, Nealson K, Silverman M. Bacterial bioluminescence: Isolation and genetic
455 analysis of functions from *Vibrio fischeri*. Cell [Internet]. 1983 [cited 2021 Jul 28];32(3):773–
456 81. Available from: <https://pubmed.ncbi.nlm.nih.gov/6831560/>
- 457 48. Brotherton CA, Medema MH, Greenberg EP. luxR Homolog-Linked Biosynthetic Gene
458 Clusters in Proteobacteria . mSystems [Internet]. 2018 Jun 26 [cited 2021 Jul 18];3(3).
459 Available from: </pmc/articles/PMC5872303/>
- 460 49. He H, Ye L, Li C, Wang H, Guo X, Wang X, et al. SbbR/SbbA, an Important ArpA/AfsA-Like
461 System, Regulates Milbemycin Production in *Streptomyces bingchenggensis*. Front
462 Microbiol [Internet]. 2018 May 23 [cited 2021 Jul 28];9(MAY):1064. Available from:
463 <https://www.frontiersin.org/article/10.3389/fmicb.2018.01064/full>
- 464 50. Heilmann S, Krishna S, Kerr B. Why do bacteria regulate public goods by quorum sensing?–
465 How the shapes of cost and benefit functions determine the form of optimal regulation. Front
466 Microbiol [Internet]. 2015 [cited 2021 Jul 28];6(AUG):767. Available from:
467 </pmc/articles/PMC4517451/>
- 468 51. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent
469 biosynthetic gene clusters [Internet]. Vol. 13, Nature Reviews Microbiology. Nature
470 Publishing Group; 2015 [cited 2021 Aug 17]. p. 509–23. Available from:
471 <https://pubmed.ncbi.nlm.nih.gov/26119570/>
- 472 52. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al.
473 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/3452617/)
- 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

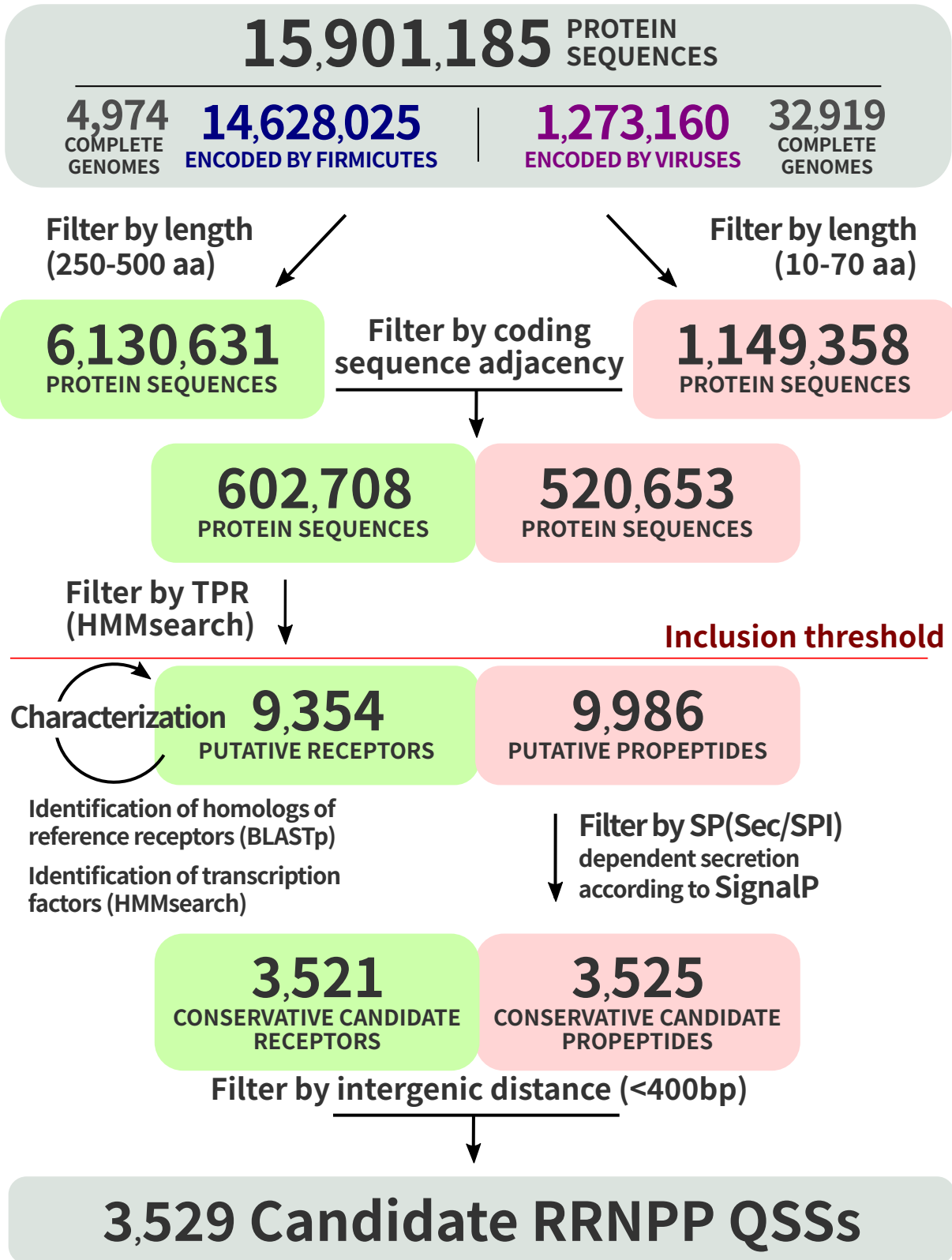
Matched N-terminal DNA binding domains (HMM)

- 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.**

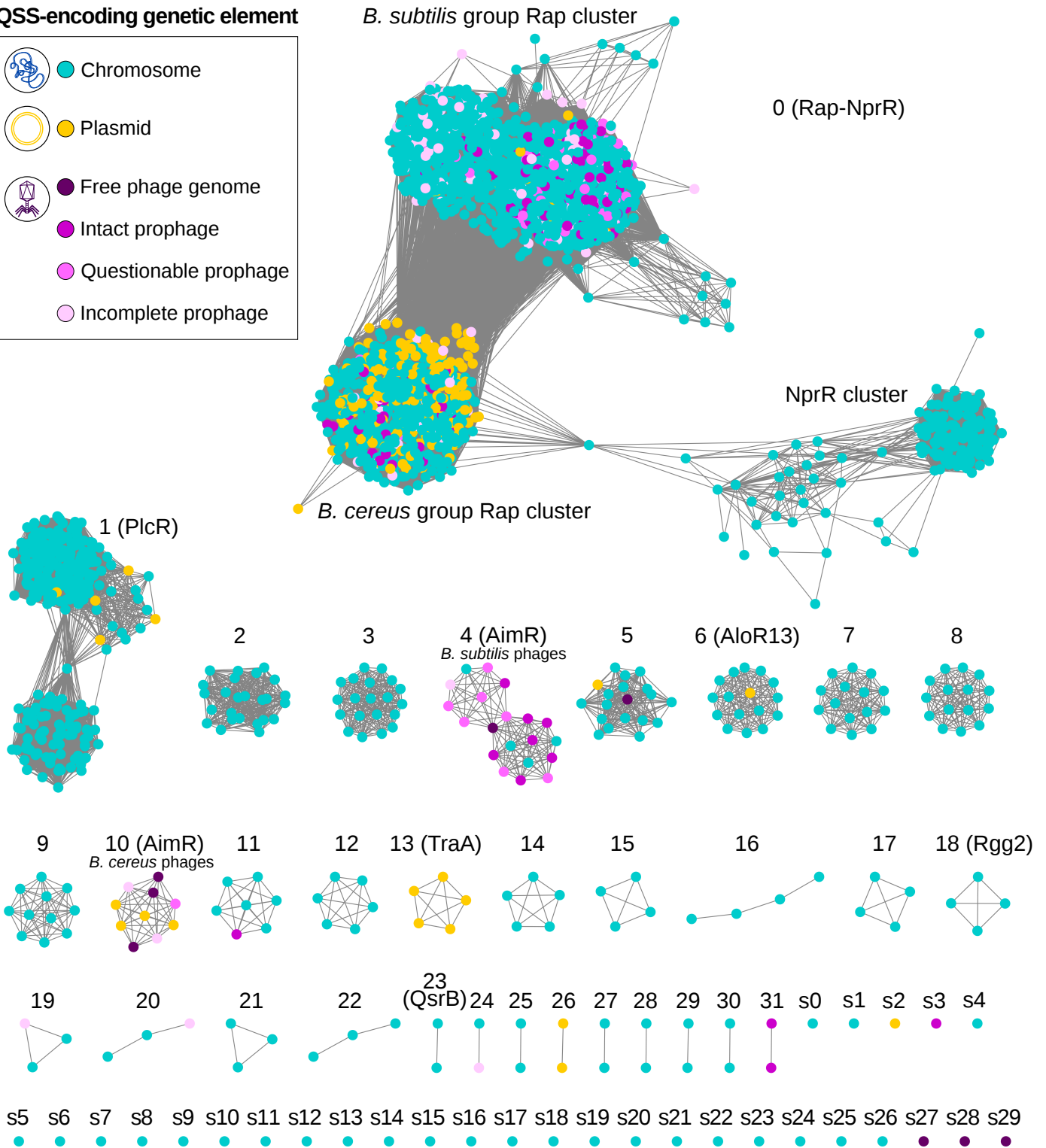
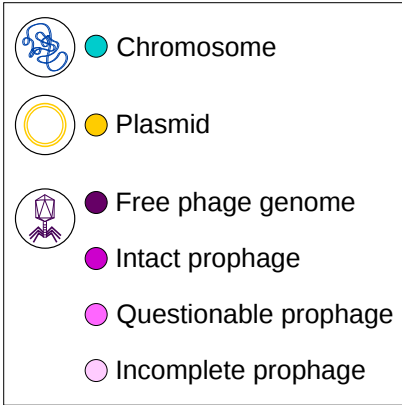
501 Each genomic context corresponds to the representative QSS of an experimentally characterized
502 RRNPP-type QSS family. The icon at the left of each context indicates the genetic element that
503 encodes the QSS (bacterial chromosome, phage genome or plasmid) and the associated label
504 indicates 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.



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



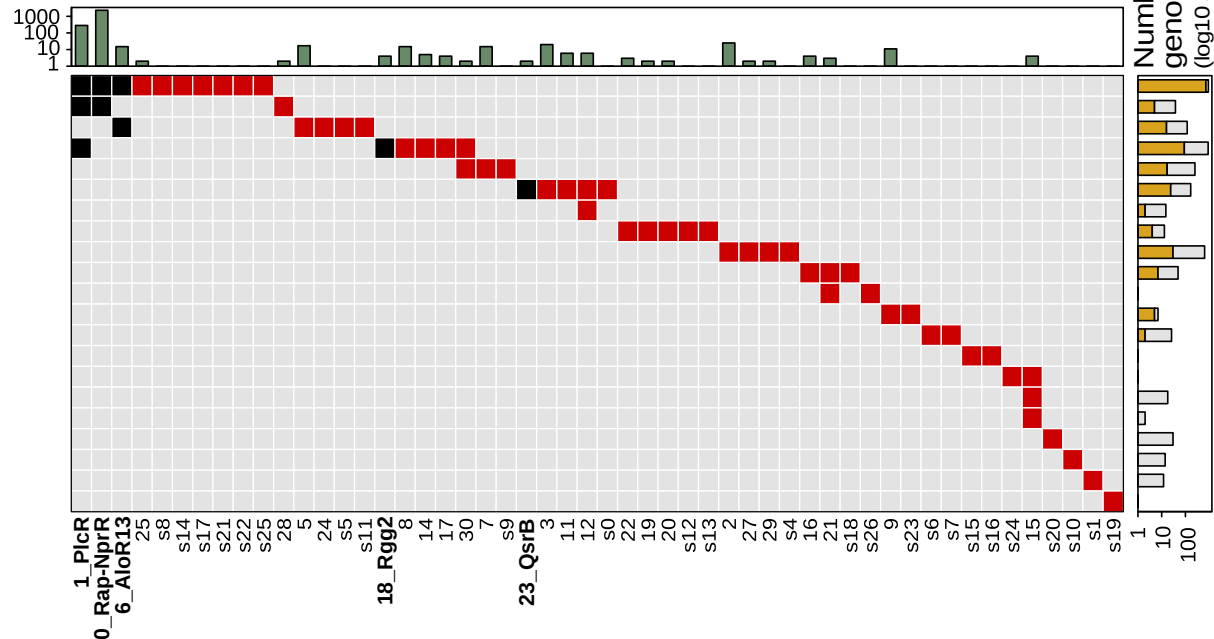
536 **Figure 3: Sequence similarity network of the receptors forming a 'conservative' candidate**
537 **RRNPP QSS with a cognate propeptide.**

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.

A



Number of chromosomal QSSs per group (log10 scale)



Chromosomes

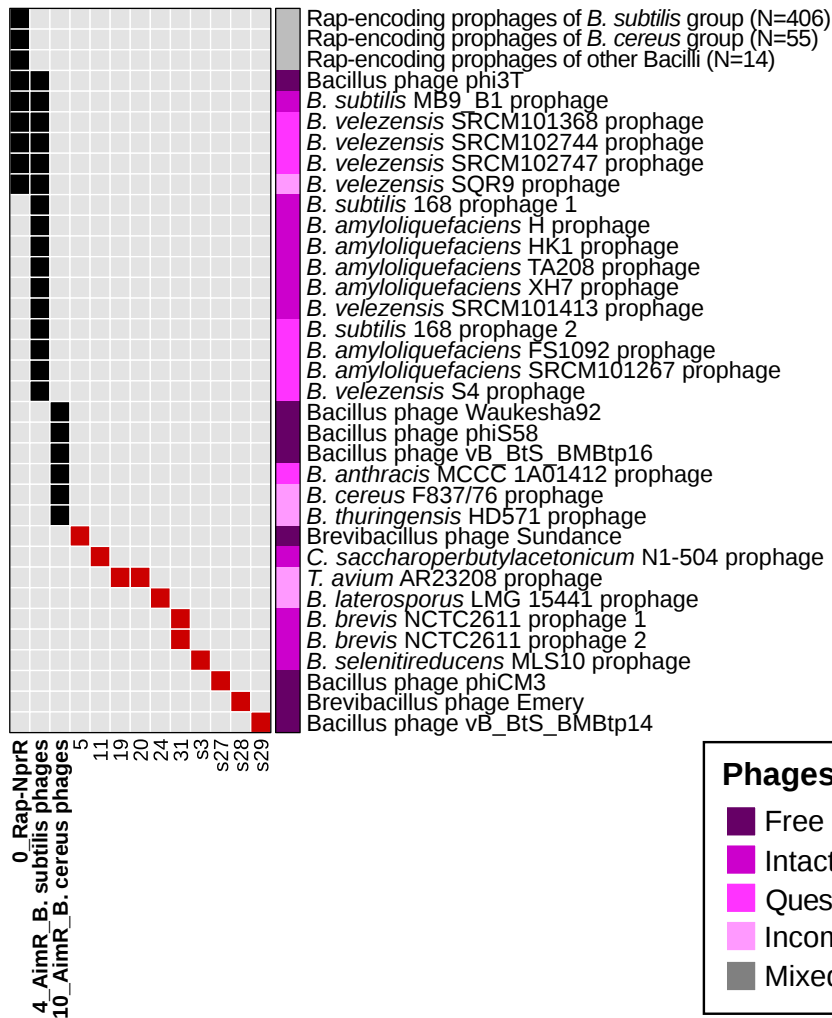
(Taxonomic families)

- Bacillaceae
- Planococcaceae
- Paenibacillaceae
- Streptococcaceae
- Enterococcaceae
- Clostridiaceae
- Thermoanaerobacterales Family III
- Alicyclobacillaceae
- Lactobacillaceae
- Lachnospiraceae
- Tissierella incertae sedis
- Thermoactinomycetaceae
- Peptococcaceae
- Halanaerobiales incertae sedis
- Proteinivoraceae
- Thermoanaerobacteraceae
- Tissierellaceae
- Carnobacteriaceae
- Eubacteriales incertae sedis
- Peptoniphilaceae
- Thermohalobacteraceae

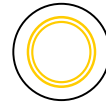
B



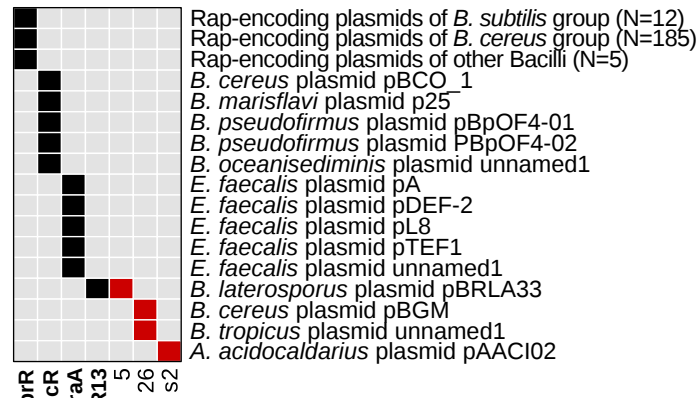
Phages



C



Plasmids



Legend

QSS families

- Previously Characterized
- Novel (this study)

Phages

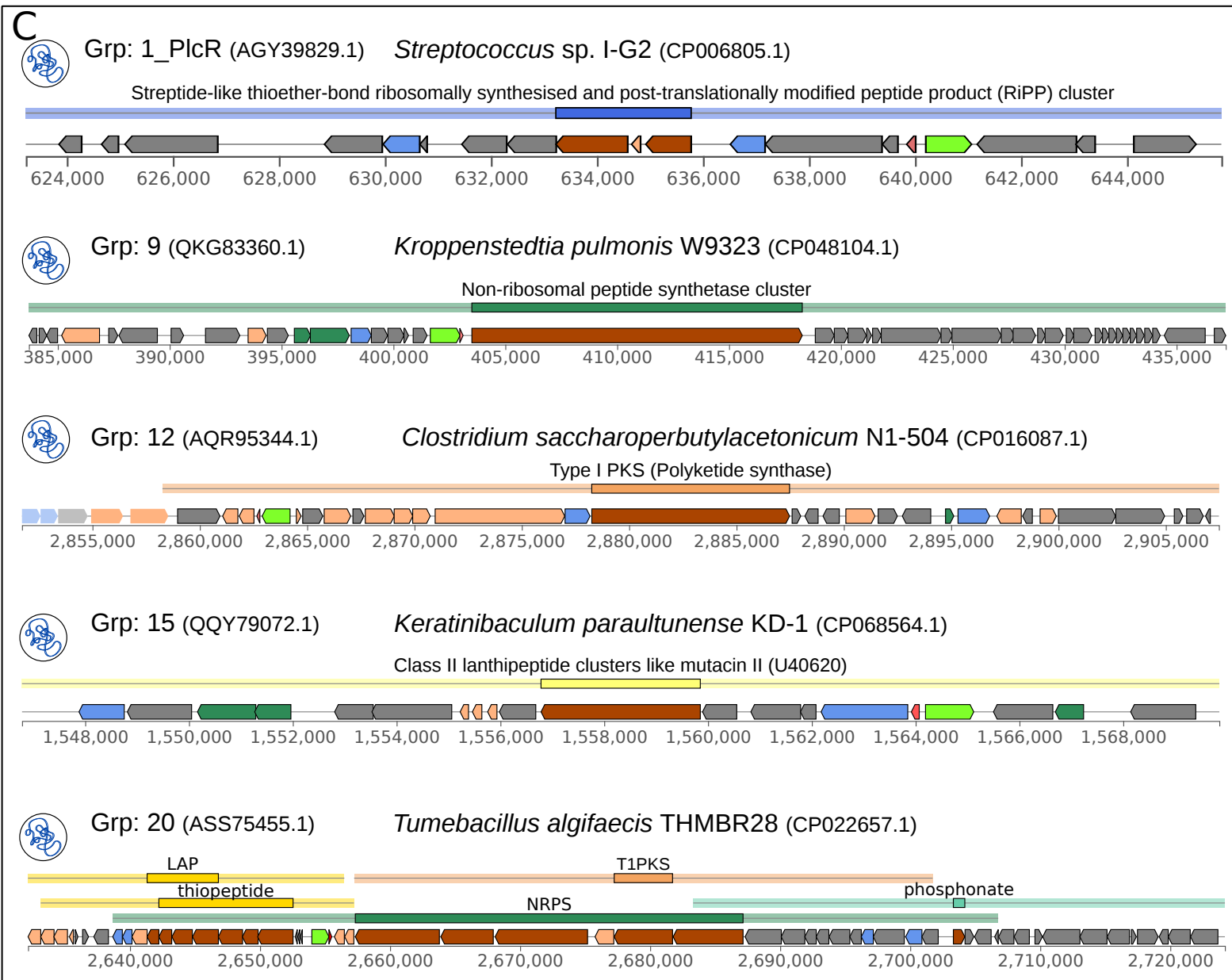
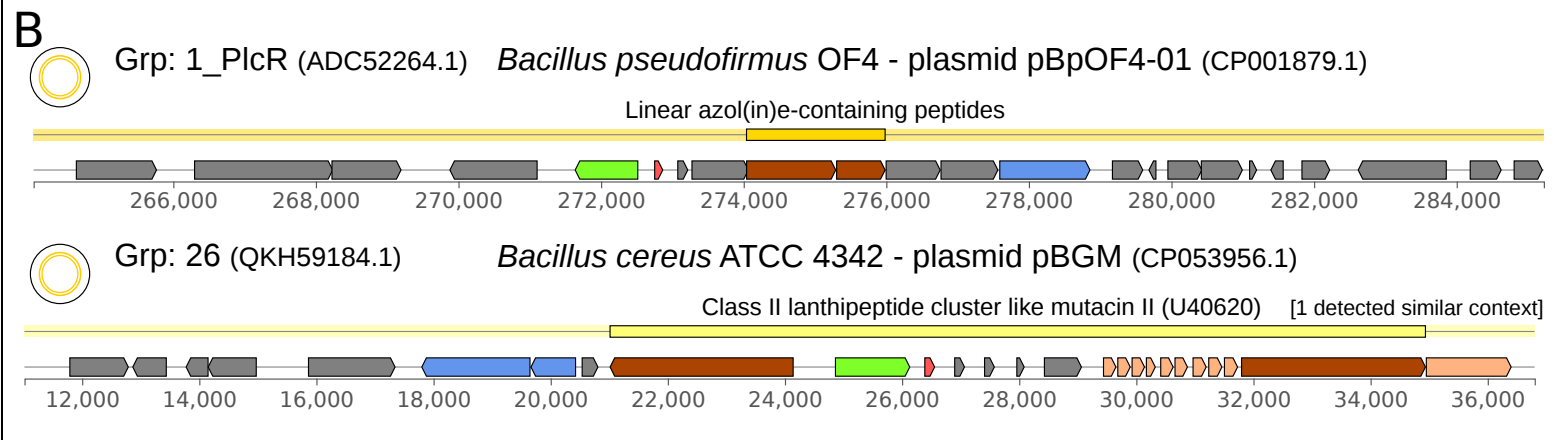
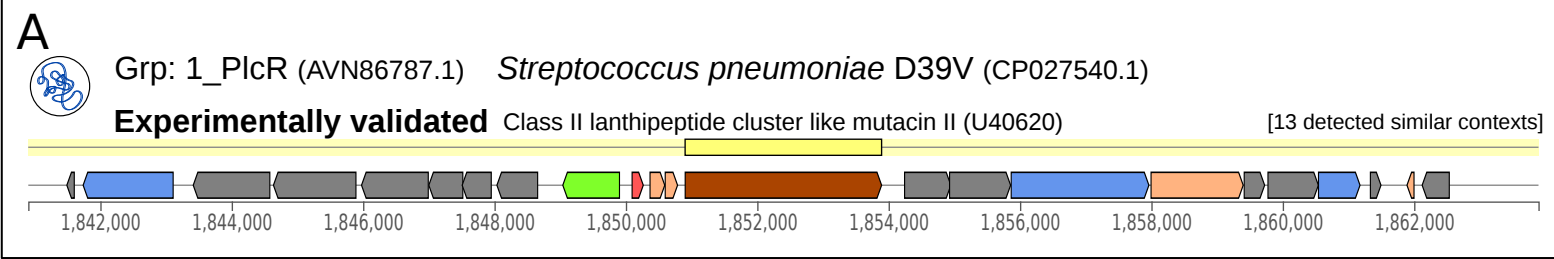
- Free phage genome
- Intact prophage
- Questionable prophage
- Incomplete prophage
- Mixed (pro)phages

Bacterial genomes

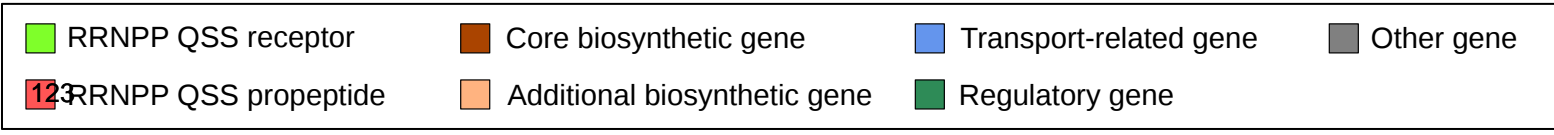
- Genome with QSS(s) detected
- Genome without QSS(s) detected

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.



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 unamed 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.

5. De la détection de nouveaux systèmes de QS à leur inférence fonctionnelle

L'article précédent a introduit la notion que les systèmes de QSSs dont le récepteur est intracellulaire (à une composante) ont tendance à réguler l'expression de gènes adjacents (**section 2.1**). Cette propriété fut d'ailleurs exploitée dans l'article précédent pour la détection de clusters biosynthétiques de gènes (BGCs) dont l'expression est prédite pour être contrôlée par un QSS adjacent, prédiction dont plusieurs études confirment le bien-fondé (Hoover *et al.*, 2015; Brotherton, Medema and Greenberg, 2018; Rued *et al.*, 2021). Puisque RRNPP_detector a révélé de nouveaux systèmes de QSSs candidats chez les bactériophages et puisque le QS des virus est un thème de recherche aussi nouveau que sous-exploré, l'article suivant en préparation se propose de caractériser les gènes adjacents aux QSSs de phages révélés par RRNPP_detector en vue d'inférer de potentielles fonctions inédites régulées de façon densité-dépendante chez ces virus. Parmi ces fonctions, la manipulation dynamique des voies de sporulation de l'hôte bactérien apparaît comme particulièrement conservée. Ces prédictions apportent un nouvel éclairage sur la façon dont la communication entre virus pourrait subtilement optimiser leur valeur reproductive en fonction de leur contexte social. Bien entendu, des tests fonctionnels seront toutefois nécessaires à la validation de ces prédictions.

1 **Large-scale identification of viral quorum sensing systems reveals density-dependent**
2 **sporulation-hijacking mechanisms in bacteriophages**

3

4 **AUTHORS**

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

6

7 **AFFILIATIONS**

8 ¹Institut 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, Museum
12 National d'Histoire Naturelle, CP 54, 57 rue Cuvier, 75005 Paris, France

13

14 **CORRESPONDING AUTHOR**

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

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

17

18 **SIGNIFICANCE**

19 Communication between viruses supported by quorum sensing systems (QSSs) is a brand new
20 research area that has transformed our views of viral adaptation and virus-host co-evolution. The viral
21 QSSs discovered so far were found to guide the lysis-lysogeny decision in temperate bacteriophages
22 as a function of phage density. Here, we identified that quorum sensing-mediated communication
23 between phages can not only guide the regulation of viral processes but also the manipulation of the
24 bacterial sporulation pathway. Our finding introduces the new view that not only bacteria decide when
25 it is time to sporulate, some bacteriophages are also key stakeholders in this dynamical decision-
26 making process. Considering that spores are the transmissive form of many pathogens, these new
27 insights have important applied implications.

28

29

30

31 **ABSTRACT**

32 Communication between viruses supported by quorum sensing systems (QSSs) were found to
33 optimize the fitness of temperate bacteriophages of Bacilli by guiding the transition from the host-
34 destructive lytic cycle to the host-protective lysogenic cycle in a density-dependent manner. All known
35 phage-encoded QSSs consist of a communication propeptide and a cognate intracellular receptor that
36 regulates the expression of adjacent target genes upon recognition of the matured peptide, a
37 signature known as RRNPP and found in chromosomes, plasmids and phages of Firmicutes bacteria.
38 Recently, we have introduced the RRNPP_detector software to detect novel genetic systems matching
39 the RRNPP signature, which unearthed many novel phage-encoded candidate QSSs. Here, by
40 looking at the adjacent genes likely regulated by these viral candidate QSSs, we identified an
41 unsuspected clustering of viral QSSs with viral genes whose bacterial homologs are key regulators of
42 the last-resort bacterial sporulation initiation pathway (*rap*, *spo0E* or *abrB*). Consistently, we found
43 evidence in published data that certain of these QSSs encoded by prophages (phage genomes
44 inserted within a bacterial genome) dynamically manipulate the timing of sporulation in the host.
45 Because these viral QSSs are genetically diverse and are found associated with different sporulation
46 regulators, this suggests a convergent evolution in bacteriophages of density-dependent sporulation-
47 hijacking mechanisms.

48

49 **INTRODUCTION**

50 If cell-cell communication via quorum sensing was discovered in 1970 in bacteria, the first
51 characterization of a functional viral quorum sensing systems (QSSs) only dates back to 2017 (1).
52 Erez *et al.* discovered that certain Bacillus phages encode a communication propeptide, which upon
53 expression, secretion, and maturation by the host cellular machinery, accumulates in the extracellular
54 environment. Accordingly, during the lytic cycle, the extracellular concentration of this phage-encoded
55 peptide reflects the number of host cells that have expressed the QSS-encoding genome of the
56 phage, which correlates with the number of hosts that have been lysed by the phage (**Fig1**). At high
57 concentrations of this peptide, when lots of hosts have been lysed and the survival of the phage-host
58 collective can be endangered, the imported-peptide is transduced by its cognate phage-encoded
59 intracellular receptor, which coordinates a population-wide transition from the host-destructive lytic
60 cycle to the host-protective lysogenic cycle. Indeed, during the lysogenic cycle, the phage genome is
61 inserted within the genome of the host, replicates as part of it, and often confers upon the host an
62 immunity towards free virions This viral QSS, coined arbitrium, thereby optimizes the fitness of a
63 phage with respect to its social context.

64 Arbitrium QSSs are classified as RRNPP QSSs, which consist of a communication propeptide
65 (secreted and matured into a quorum sensing peptide) coupled with an intracellular receptor, turned-
66 on or -off upon binding with the imported mature peptide (**Fig1**) (2). RRNPP receptors often are

67 transcription factors that dynamically regulate the expression of adjacent genes. As genetically diverse
68 as RRNPP QSSs are in chromosomes, plasmids and phages of Firmicutes, they however function
69 according to the same canonical mechanism (**Fig1**), except with very few exceptions in the mode of
70 secretion (Shp and PrgQ peptides) (2). We recently identified that this shared mechanism between
71 different RRNPP QSSs underlies a common signature that can be detected in-silico (3). On this basis,
72 we have developed RRNPP_detector, a tool designed to identify novel candidate QSSs matching the
73 RRNPP signature (3).

74 In a large-scale application of RRNPP_detector, we identified many novel candidate RRNPP QSSs
75 in sequenced genomes of temperate bacteriophages and in prophages inserted within bacterial
76 genomes (3). During the development of RRNPP_detector, we noticed an unsuspected clustering of
77 viral candidate QSSs with adjacent viral genes whose bacterial homologs are key regulators of the
78 bacterial sporulation initiation pathway. Because RRNPP QSSs tend to dynamically regulate adjacent
79 genes (1, 2, 4, 5), this hinted at density-dependent manipulations of bacterial sporulation by phages.

80 In Firmicutes, sporulation leads to the formation of endospores, able to resist extreme
81 environmental stresses for prolonged periods and to resume vegetative growth in response to
82 favorable changes in environmental conditions (6). The sporulation pathway is initiated when
83 transmembrane kinases sense stress stimuli, and thereupon transfer their phosphate, either directly
84 (*Clostridium*) or via phosphorelay (*Bacillus*, *Brevibacillus*) to Spo0A, the master regulator of
85 sporulation (7, 8). However, only high Spo0A-P concentrations, and therefore intense stresses, can
86 commit a cell to sporulate (9). Because sporulation is costly, Spo0A-P accumulation is subjected to
87 multiple regulative check points by the Rap, Spo0E and AbrB proteins (10, 11). In adverse
88 circumstances, Rap, Spo0E and AbrB thereby form a decision-making regulatory circuit that controls
89 the timing of sporulation in Bacilli (11).

90 Here, we present in more details our results regarding the association of sporulation regulators
91 (either *rap*, *spo0E* or *abrB*) with candidate QSSs found in temperate bacteriophages of Firmicutes,
92 their mechanistic consequence on the host biology, and their fundamental and applied implications.

93

94 **RESULTS AND DISCUSSION**

95 **Identification of 384 viral candidate RRNPP QSSs, distributed into 26 families, of which only 6** 96 **were previously known**

97 To detect viral QSSs, we followed the study design displayed in **FigS1**. We applied
98 RRNPP_detector (15-65aa and 250-460aa length thresholds for propeptides and receptors) against
99 the Gut Phage Database (12), 32,327 NCBI complete genomes of Viruses, and 3,577 NCBI complete
100 genomes of Firmicutes (because phage genomes can be inserted within bacterial genomes). This led
101 to the identification of 16 candidate RRNPP QSSs on intestinal phage genomes, 10 on sequenced
102 genomes of temperate phages and 2671 on bacterial genomes, respectively. Prophage regions within

103 bacterial genomes were subsequently detected by Phaster (13) and Prophage Hunter (14) to
104 distinguish between genuine bacterial QSSs and prophage-encoded QSSs inserted within bacterial
105 chromosomes or plasmids. This enabled identifying 358 additional candidate viral QSSs: 174 on
106 intact/active prophages, 68 on questionable/ambiguous prophages, 116 on incomplete prophages.
107 The genomic and taxonomic details of these 384 viral candidate RRNPP QSSs can be found in
108 **TableS1**.

109 In a blast all vs all, the receptors of all the 2697 candidate QSSs were further classified into 64
110 groups of homologs / QSS families (sequence identity $\geq 30\%$; mutual length coverage of the
111 alignment $\geq 80\%$), as in (3). Of these 64 QSS families, 26 families comprised at least one
112 phage/prophage-encoded QSS and served as the focus of this study (metadata of the 26 families in
113 **TableS1**; distribution of the 384 viral candidate RRNPP QSSs into these 26 families in **Fig2**). 6 of
114 these 26 QSS families had already been described prior to our study: Rap-Npr of *Bacilli* (2, 15), AimR
115 of *B. subtilis* phages (1), AloR of *Paenibacilli* (16), AimR of *B. cereus* phages (5), the family of QSSs
116 characterized by Feng *et al.* in *Clostridium saccharoperbutylacetonicum* (17), the family of QSSs
117 characterized by Kotte *et al.* in *Clostridium acetobutylicum* (18) (**Fig2**). Accordingly, the 20 other
118 families show great promise to expand the known diversity of viral RRNPP QSSs. As expected, the
119 arbitrium systems of *B. subtilis* (N=15) and *B. cereus* (N=9) phages form two 100% viral families
120 (**Fig2B**). The two biggest families, Rap-Npr of Bacilli (N=2258) and the novel candidate QSS1 family
121 of Brevibacilli (N=18) are found to be shared between bacterial chromosomes, plasmids and
122 phages/prophages (**Fig2**). The Rap-Phr family notably is the family in which (pro)phage-encoded
123 QSSs are the most prevalent (326 viral Rap-Phr QSSs (**TableS1 and Fig2A**)).

124

125 **Rap-Phr QSSs that delay the timing of sporulation are found in many, diverse Bacillus** 126 **bacteriophages**

127 Importantly, bacterial Rap-Phr QSSs are known to regulate the competence, the sporulation and/or
128 the production of public goods in *Bacilli*. Notably, these communication systems ensure that Spo0A-P
129 only accumulates when the Rap-Phr encoding subpopulation reaches high densities (19). On this
130 basis, Rap-Phr QSSs have been proposed as a means for a cell to delay a costly commitment to
131 sporulation as long as the ratio of available food per kin-cell is compatible with individual survival in
132 periods of nutrient limitation (19). However, the fact that Rap-Phr QSSs are found on phages implies
133 that this delay in the timing of sporulation can sometimes be dependent on the intracellular density of
134 phage genomes or prophages rather than on actual cell densities, presumably for the evolutionary
135 benefit of the phage or the prophage-host collective (20).

136 Here, we inferred the maximum-likelihood phylogeny of the 2258 detected Rap receptors (**Fig2A**)
137 and identified that the 326 viral Rap-Phr QSSs are genetically diverse and polyphyletic (scattered

138 between bacterial leaves). Remarkably, this suggests that bacteria and phages frequently exchange
139 these communication systems.

140 Indeed, from a phage's perspective, a dynamical modulation of sporulation can be advantageous
141 because on the one hand, sporulation can elicit the lytic cycle (21) and trigger cannibalistic behaviors
142 that may reduce the amount of potential hosts (22) and on the other hand, spores can protect the
143 phage genome under unfavorable environmental conditions. From a lysogenized host perspective, the
144 *rap-phr* QSS acquired from the prophage can be considered as adaptive because they might support
145 a cell-cell communication between lysogenized bacteria (since prophage density correlates with host
146 density) that enables to cheat (delay production of public goods) and delay the costly sporulation
147 program in a density-dependent manner, to the benefit of the prophage-host collective (23). A delay of
148 sporulation might notably offer some additional time to resume growth if a peculiar stress happens to
149 be relieved from the environment (24) or to benefit from the last nutrients present in the environment
150 to replicate before sporulating, hence maximizing the number of representatives in the future
151 heterogeneous population of spores.

152

153 **Multiple occurrences of the quorum sensing-mediated sporulation-hijacking genomic** 154 **signature in temperate bacteriophages**

155 Remarkably, the phage-encoded Rap-Phr QSSs were not the only potential host-hijacking viral
156 QSSs. Indeed, we identified 5 additional candidate phage/prophage-encoded QSSs, distributed into 5
157 different QSSs families, predicted to manipulate the host sporulation initiation pathway in a density-
158 dependent manner (**Fig3**). This prediction lies on the observation that their receptor harbors a DNA
159 binding domain (**TableS1**) and thus likely regulates the expression of adjacent genes (like the
160 arbitrium system), and that either the *spo0E* or *arbB* sporulation regulator is found in the vicinity of the
161 QSSs on the phage/prophage DNA (**Fig3**). The same genomic context, albeit not encoded by a
162 phage, was shown to be linked to sporulation regulation in *Paenibacillus* bacteria (16). By either
163 activating or repressing the viral *spo0E* or *abrB* genes once a quorum of phages/prophages is
164 reached, our candidate viral QSSs could influence the total concentration of Spo0E or AbrB within
165 hosts in a density-dependent-manner, thus influencing the dynamics of Spo0A-P accumulation in
166 these hosts and thereby modulating the sporulation initiation pathway to the benefit of the phage or
167 the prophage (**FigS2**). Importantly, as the sporulation initiation pathway can trigger a wide range of
168 biological processes (sporulation, biofilm formation, cannibalism, toxin production and solventogenesis
169 (22, 25)), these viral candidate QSSs would also manipulate, in a density-dependent manner, a
170 substantially broader spectrum of the host biology than spore formation alone (**FigS2**). Because these
171 viral QSSs belong to distinct QSS families and are carried by phages infecting different hosts, this
172 suggests a convergent evolution of density-dependent sporulation-hijacking in bacteriophages.

173 We next found published data that support our prediction of a dynamical sporulation-hijacking
174 mediated by phage-encoded QSSs. Indeed, the RapBL5-PhrBL5 of *Bacillus licheniformis* (26), the
175 QssR5-QssP5 of *Clostridium saccharoperbutylacetonicum* (17), and the QsrG-QsrG of *Clostridium*
176 *acetobutylicum* (18) were previously found to dynamically regulate sporulation and we identified that
177 these QSSs are predicted by Prophage Hunter to be encoded by prophages inserted within bacterial
178 chromosomes (**Fig4**). Moreover, 2 of these 3 prophages are predicted by Phaster to be intact and by
179 Prophage Hunter to be active and thus able to re-initiate the lytic cycle upon excision while the other
180 probably was domesticated by its host (27). These observations support the prediction that some
181 genomes of temperate bacteriophages encode a communication system that guides the manipulation
182 of bacterial sporulation.

183

184 **Concluding remarks**

185 Here, we computationally characterized distinct candidate QSSs presenting a genomic signature
186 for a density-dependent hijacking of the bacterial sporulation initiation pathway in sequenced genomes
187 of temperate bacteriophages (*Bacillus* phage phi3T, *Brevibacillus* phage Sundance) and in latent
188 prophages. Moreover, we found published data that supports the validity of this signature. Because
189 these QSSs are genetically diverse and belong to phages/prophages infecting different species of
190 bacteria, this suggests multiple independent acquisitions of quorum sensing-mediated sporulation-
191 hijacking genetic systems in phages. Accordingly, phage-encoded QSSs would not only dynamically
192 regulate viral processes (as in the arbitrium QSSs) but also bacterial processes, for the benefit of the
193 phage or the prophage-host collective. Our study also highlights that if bacteria decide when it's time
194 to sporulate, some phages also are stakeholders in the decision-making process. Some
195 bacteriophages had already been reported to be either spore-restricting or spore-converting (28–31),
196 but either way, these activations/impairments of sporulation were not the result of a decision making-
197 process, unlike the density-dependent modulations described here.

198 From an applied viewpoint, these findings are also important because as sporulation enables
199 bacteria to resist harsh environmental conditions, it represents a route for bacteria to travel between
200 environments, and notably to end up within human bodies. Consequently, endospores are the
201 transmissive form of many bacteria, be they commensal or pathogen (32–34). Furthermore, 50-60% of
202 the bacterial genera from the human gut are estimated to produce resilient spores, specialized for
203 host-to-host transmission (35). Accordingly, by dynamically interfering with sporulation, phage-
204 encoded QSSs could influence the dynamics of transmission of bacteria in humans. With this respect,
205 we also report a high prevalence of genes in the human gut phage database matching the HMM
206 models of Spo0E and AbrB (**TableS2**), strengthening the case of a regulative effect of bacteriophages
207 on the host-to-host transmission of gastrointestinal bacteria.

208

209 **METHODS**

210 **Construction of the target data sets**

211 The complete genomes of Viruses and Firmicutes were queried from the NCBI 'Assembly'
212 database (36), as of 28/04/2020 and 10/04/2020, respectively. The features tables (annotations) and
213 the encoded protein sequences of these genomes were downloaded using 'GenBank' as source
214 database. The Gut Phage Database (12) was downloaded as of 29/10/2020, from the following url:
215 http://ftp.ebi.ac.uk/pub/databases/metagenomics/genome_sets/gut_phage_database/

216

217 **Detection of candidate RRNPP QSSs**

218 We launched a systematical search of the RRNPP-type signature using RRNPP_detector
219 (3) against the complete genomes of Viruses and Firmicutes available on the NCBI and the MAGs of
220 bacteriophages from the Gut Phage Database. RRNPP_detector defines candidate RRNPP-type
221 quorum sensing systems as tandems of adjacent ORFs encoding a candidate receptor (protein
222 matching HMMs of peptide-binding tetratricopeptide repeats (TPRs)) and a candidate pro-peptide
223 (small protein predicted to be excreted via the SEC-translocon), consistent with the genetic features
224 that are specific to the RRNPP quorum sensing mechanism and that are common between different
225 experimentally-validated RRNPP QSSs (3). We specified a length range of 15-65aa for candidate
226 propeptides and 250-460aa for candidate receptors.

227

228 **Classification of the candidate RRNPP QSSs into families**

229 Because quorum sensing pro-peptides offer few amino acids to compare, are versatile and
230 subjected to intragenic duplication (26), we classified the QSSs based on sequence homology of the
231 receptors as in (3). We launched a BLASTp (37) All vs All of the receptors of the 2697 candidate QSSs
232 identified in the complete genomes of Viruses and Firmicutes. The output of BLASTp was filtered to
233 retain only the pairs of receptors giving rise to at least 30% sequence identity over more than 80% of
234 the length of the two proteins. These pairs were used to build a sequence similarity network and the
235 families were defined based on the connected components of the graph.

236

237 **Identification of already known QSS families**

238 A BLASTp search was launched using as queries the RapA (NP_389125.1), NprR
239 (WP_001187960.1), PlcR (WP_000542912.1), Rgg2 (WP_002990747.1), AimR (APD21232.1), AimR-
240 like (AID50226.1), PrgX (WP_002366018.1), TraA (BAA11197.1), AloR13 (IMG: 2547357582), QsrB
241 (AAK78305.1), Qss5R (AGF59421.1) reference receptors, and as a target database, the 2697
242 candidate QSS receptors previously identified. If the best hit of a reference RRNPP-type receptor

243 gave rise to a sequence identity $\geq 30\%$ over more than 80% mutual coverage, then the family to
244 which this best hit belongs was considered as an already known family.

245

246 **Prophage detection**

247 All the NCBI ids of the genomic accessions of chromosomes or plasmids of Firmicutes encoding
248 one or several candidate QSSs were retrieved and automatically submitted to the Phaster webtool
249 (13). Eventually, each QSS was defined as viral if its genomic coordinates on a given
250 chromosome/plasmid fell within a region predicted by Phaster to belong to a prophage (qualified as
251 either 'intact', 'questionable' or 'incomplete' prophage). Phaster results were complemented by
252 ProphageHunter (14), a webtool that computes the likelihood that a prophage is active (able to
253 reinstate the lytic cycle by excision). Because ProphageHunter cannot be automatically queried, we
254 only called upon this webtool for chromosomes/plasmids which encode QSSs that are not part of the
255 two biggest of the 64 detected QSS families, namely Rap-Phr (2258 candidate QSSs) and PlcR-PapR
256 (223 candidate QSSs, data not shown). Likewise, coordinates of candidate QSSs were eventually
257 intersected with predicted prophage regions to detect potential prophage-encoded candidate QSSs
258 that could have been missed by Phaster (**TableS1**). Based on the results of Phaster and Prophages,
259 the focus of the study were further restricted to the 26 families comprising at least one phage or
260 prophage-encoded QSSs.

261

262 **Prediction of the mature quorum sensing peptides**

263 For each uncharacterized families of candidate receptors of size >1 with at least one (pro)phage-
264 encoded member referenced in the NCBI, the cognate pro-peptides were aligned in a multiple
265 sequence alignment (MSA) using MUSCLE *version 3.8.31* (38). Each MSA was visualized with
266 Jalview *version 1.8.0_201* under the ClustalX color scheme which colors amino acids based on
267 residue type conservation (39). The region of RRNPP-type pro-peptides encoding the mature quorum
268 sensing peptide usually corresponds to a small sequence (5-6aa), located in the C-terminal of the pro-
269 peptide, with conserved amino-acids types in at least 3 positions (1, 18, 20, 40). Based on the amino-
270 acid profile of C-terminal residues in each MSA, putative mature quorum sensing peptides were
271 manually determined (**FigS3**).

272

273 **Phylogenetic tree of Rap**

274 A multiple sequence alignment (MSA) of the protein sequences of the Rap receptors forming a
275 candidate Rap-Phr QSS was performed using MUSCLE *version 3.8.31* (38). The MSA was then
276 trimmed using trimmal *version 1.4.rev22* with the option '-automated1', optimized for maximum
277 likelihood phylogenetic tree reconstruction (41). The trimmed MSA was then given as input to IQ-
278 TREE *version multicore 1.6.10* to infer a maximum likelihood phylogenetic tree under the LG+G model

279 with 1000 ultrafast bootstraps (42). The tree was further edited via the Interactive Tree Of Life (ITOL)
280 online tool (43).

281

282 **Analysis of the genomic context of candidate QSSs**

283 The genomic context of the (pro)phage-encoded candidate QSSs that are not part of the arbitrium
284 families (functions already known in phages) and the Rap-Phr family (Rap is a protein inhibitor and not
285 a transcription factor) and whose receptors matched the DNA binding domain profiles of RRNPP
286 transcription factors (Hidden Markov Models PFAM PF01381, Superfamily SSF47413, SMART
287 SM00530, CATH 1.10.260.40) were investigated by analyzing the functional annotation of their
288 adjacent protein-coding genes, or when missing, by launching a “Conserved Domains” search within
289 their sequence and a BLASTp search of their sequence against the NR (non-redundant) protein
290 database maintained by the NCBI.

291

292 **Identification of *rap*, *spo0E* and *abrB* genes in the Gut Phage Database**

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

297

298 **DECLARATIONS**

299 **Availability of data and materials**

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

302 **Competing Interests**

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

304 **Fundings**

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

308 **Authors' Contribution**

309 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
310 manuscript with input from all authors. All documents were edited and approved by all authors.

311 **Acknowledgments**

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

313

314 **REFERENCES**

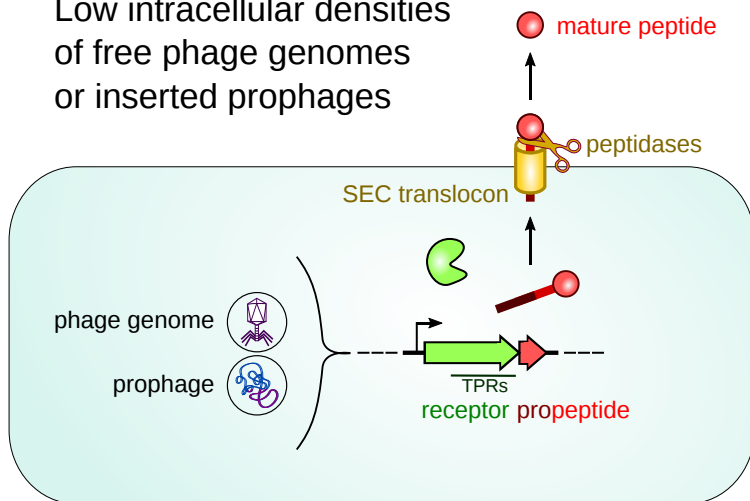
- 315 1. Z. Erez, *et al.*, Communication between viruses guides lysis-lysogeny decisions. *Nature* **541**,
316 488–493 (2017).
- 317 2. M. B. Neiditch, G. C. Capodagli, G. Prehna, M. J. Federle, Genetic and Structural Analyses of
318 RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria. *Annu. Rev. Genet.* **51**, 311–
319 333 (2017).
- 320 3. C. Bernard, Y. Li, E. Bapteste, P. Lopez, RRNPP_detector: a tool to detect RRNPP quorum
321 sensing systems in chromosomes, plasmids and phages of gram-positive bacteria AUTHORS.
322 *bioRxiv*, 2021.08.18.456871 (2021).
- 323 4. V. Kohler, W. Keller, E. Grohmann, Regulation of gram-positive conjugation. *Front. Microbiol.*
324 **10**, 1134 (2019).
- 325 5. A. Stokar-Avihail, N. Tal, Z. Erez, A. Lopatina, R. Sorek, Widespread Utilization of Peptide
326 Communication in Phages Infecting Soil and Pathogenic Bacteria. *Cell Host Microbe* **25**, 746-
327 755.e5 (2019).
- 328 6. M. Y. Galperin, Genome Diversity of Spore-Forming Firmicutes. *Microbiol. Spectr.* **1**, TBS-0015-
329 2012 (2013).
- 330 7. I. S. Tan, K. S. Ramamurthi, Spore formation in *Bacillus subtilis*. *Environ. Microbiol. Rep.* **6**,
331 212–225 (2014).
- 332 8. M. A. Al-Hinai, S. W. Jones, E. T. Papoutsakis, The *Clostridium* Sporulation Programs: Diversity
333 and Preservation of Endospore Differentiation. *Microbiol. Mol. Biol. Rev.* **79**, 19–37 (2015).
- 334 9. M. Fujita, R. Losick, Evidence that entry into sporulation in *Bacillus subtilis* is governed by a
335 gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev.* **19**, 2236–
336 2244 (2005).
- 337 10. S. H. Shafikhani, T. Leighton, AbrB and Spo0E Control the Proper Timing of Sporulation in
338 *Bacillus subtilis*. *Curr. Microbiol.* **48**, 262–269 (2004).
- 339 11. D. Schultz, M. Lu, T. Stavropoulos, J. Onuchic, E. Ben-Jacob, Turning oscillations into
340 opportunities: Lessons from a bacterial decision gate. *Sci. Rep.* **3** (2013).
- 341 12. L. F. Camarillo-Guerrero, A. Almeida, G. Rangel-Pineros, R. D. Finn, T. D. Lawley, Massive
342 expansion of human gut bacteriophage diversity. *Cell* **184**, 1098-1109.e9 (2021).
- 343 13. D. Arndt, *et al.*, PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic*
344 *Acids Res.* **44**, W16 (2016).
- 345 14. W. Song, *et al.*, Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic*
346 *Acids Res.* **47**, W74–W80 (2019).
- 347 15. S. Perchat, *et al.*, NprR, a moonlighting quorum sensor shifting from a phosphatase activity to a
348 transcriptional activator. *Microb. Cell* **3**, 573–575 (2016).
- 349 16. M. Voichek, S. Maaß, T. Kroniger, D. Becher, R. Sorek, Peptide-based quorum sensing systems
350 in *Paenibacillus polymyxa*. *Life Sci. Alliance* **3** (2020).

- 351 17. J. Feng, *et al.*, RRNPP-Type quorum-sensing systems regulate solvent formation, sporulation
352 and cell motility in *Clostridium saccharoperbutylacetonicum*. *Biotechnol. Biofuels* **13**, 1–16
353 (2020).
- 354 18. A. K. Kotte, *et al.*, RRNPP-type quorum sensing affects solvent formation and sporulation in
355 *clostridium acetobutylicum*. *Microbiol. (United Kingdom)* **166**, 579–592 (2020).
- 356 19. I. B. Bischofs, J. A. Hug, A. W. Liu, D. M. Wolf, A. P. Arkin, Complexity in bacterial cell-cell
357 communication: quorum signal integration and subpopulation signaling in the *Bacillus subtilis*
358 phosphorelay. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6459–64 (2009).
- 359 20. C. Bernard, Y. Li, P. Lopez, E. Bapteste, Beyond arbitrium: identification of a second
360 communication system in *Bacillus* phage phi3T that may regulate host defense mechanisms.
361 *ISME J.* (2020) <https://doi.org/10.1038/s41396-020-00795-9>.
- 362 21. W. J. Meijer, *et al.*, Molecular basis for the exploitation of spore formation as survival
363 mechanism by virulent phage ϕ 29. *EMBO J.* **24**, 3647–3657 (2005).
- 364 22. J. E. González-Pastor, Cannibalism: A social behavior in sporulating *Bacillus subtilis*. *FEMS*
365 *Microbiol. Rev.* **35**, 415–424 (2011).
- 366 23. M. Kalamara, M. Spacapan, I. Mandic-Mulec, N. R. Stanley-Wall, Social behaviours by *Bacillus*
367 *subtilis*: quorum sensing, kin discrimination and beyond. *Mol. Microbiol.* **110**, 863 (2018).
- 368 24. J. E. González-Pastor, E. C. Hobbs, R. Losick, Cannibalism by Sporulating Bacteria. *Science*
369 (80-). **301**, 510–513 (2003).
- 370 25. P. Dürre, *et al.*, Transcriptional regulation of solventogenesis in *Clostridium acetobutylicum* in
371 *Journal of Molecular Microbiology and Biotechnology*, (2002), pp. 295–300.
- 372 26. E. Even-Tov, S. Omer Bendori, S. Pollak, A. Eldar, Transient Duplication-Dependent Divergence
373 and Horizontal Transfer Underlie the Evolutionary Dynamics of Bacterial Cell–Cell Signaling.
374 *PLOS Biol.* **14**, e2000330 (2016).
- 375 27. L. M. Bobay, M. Touchon, E. P. C. Rocha, Pervasive domestication of defective prophages by
376 bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12127–12132 (2014).
- 377 28. D. P. Boudreaux, V. R. Srinivasan, “Bacteriophage-induced Sporulation in *Bacillus cereus* T.”
- 378 29. M. G. Bramucci, K. M. Keggins, P. S. Lovett, Bacteriophage conversion of spore-negative
379 mutants to spore-positive in *Bacillus pumilus*. *J. Virol.* **22**, 194–202 (1977).
- 380 30. T. H. Silver-Mysliwiec, M. G. Bramucci, Bacteriophage-enhanced sporulation: Comparison of
381 spore-converting bacteriophages PMB12 and SP10. *J. Bacteriol.* **172**, 1948–1953 (1990).
- 382 31. R. Schuch, V. A. Fischetti, The secret life of the anthrax agent *Bacillus anthracis*:
383 bacteriophage-mediated ecological adaptations. *PLoS One* **4**, e6532 (2009).
- 384 32. M. Mallozzi, V. K. Viswanathan, G. Vedantam, Spore-forming Bacilli and Clostridia in human
385 disease. *Future Microbiol.* **5**, 1109–1123 (2010).
- 386 33. F. Postollec, *et al.*, Tracking spore-forming bacteria in food: From natural biodiversity to
387 selection by processes. *Int. J. Food Microbiol.* **158**, 1–8 (2012).

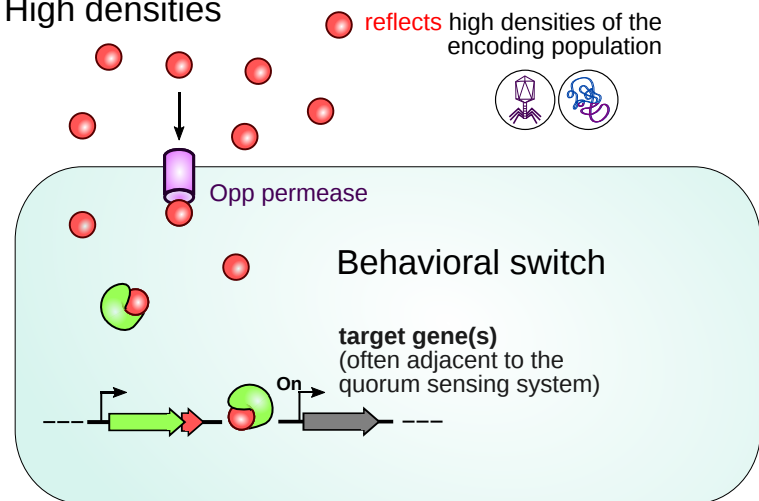
- 388 34. M. C. Swick, T. M. Koehler, A. Driks, Surviving Between Hosts: Sporulation and Transmission.
389 *Microbiol. Spectr.* **4** (2016).
- 390 35. H. P. Browne, *et al.*, Culturing of “unculturable” human microbiota reveals novel taxa and
391 extensive sporulation. *Nature* **533**, 543–546 (2016).

- 392 36. NCBI Resource Coordinators, Database resources of the National Center for Biotechnology
393 Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
- 394 37. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool.
395 *J. Mol. Biol.* **215**, 403–410 (1990).
- 396 38. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput.
397 *Nucleic Acids Res.* **32**, 1792–7 (2004).
- 398 39. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2--a
399 multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191
400 (2009).
- 401 40. M. Pottathil, B. A. Lazizzera, The extracellular PHR peptide-Rap phosphatase signaling circuit
402 of bacillus subtilis. *Front. Biosci.* **8**, 913 (2003).
- 403 41. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment
404 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
- 405 42. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective
406 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–
407 274 (2015).
- 408 43. I. Letunic, P. Bork, Interactive Tree of Life (iTOL) v4: Recent updates and new developments.
409 *Nucleic Acids Res.* **47**, W256–W259 (2019).
- 410
- 411

Low intracellular densities of free phage genomes or inserted prophages

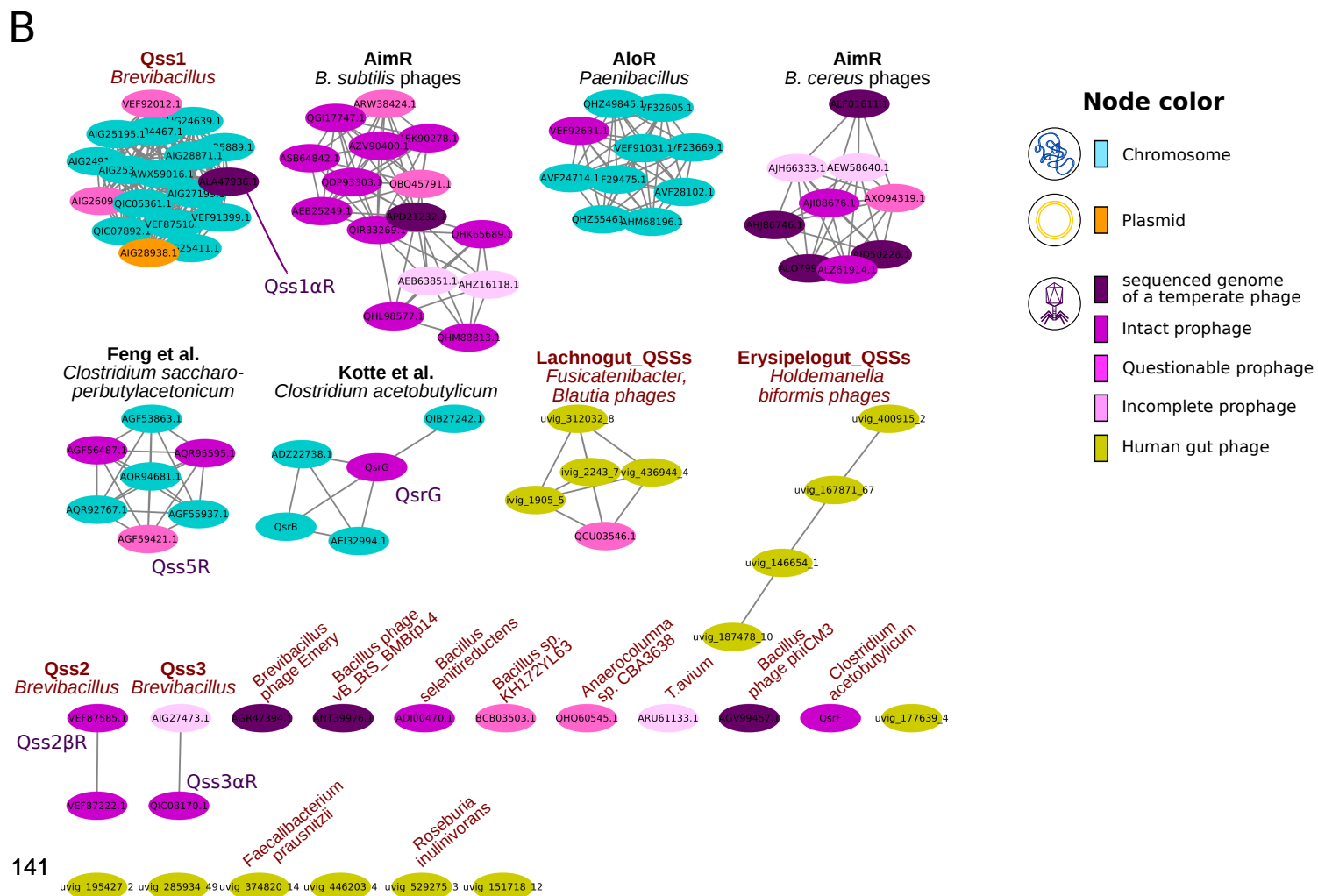
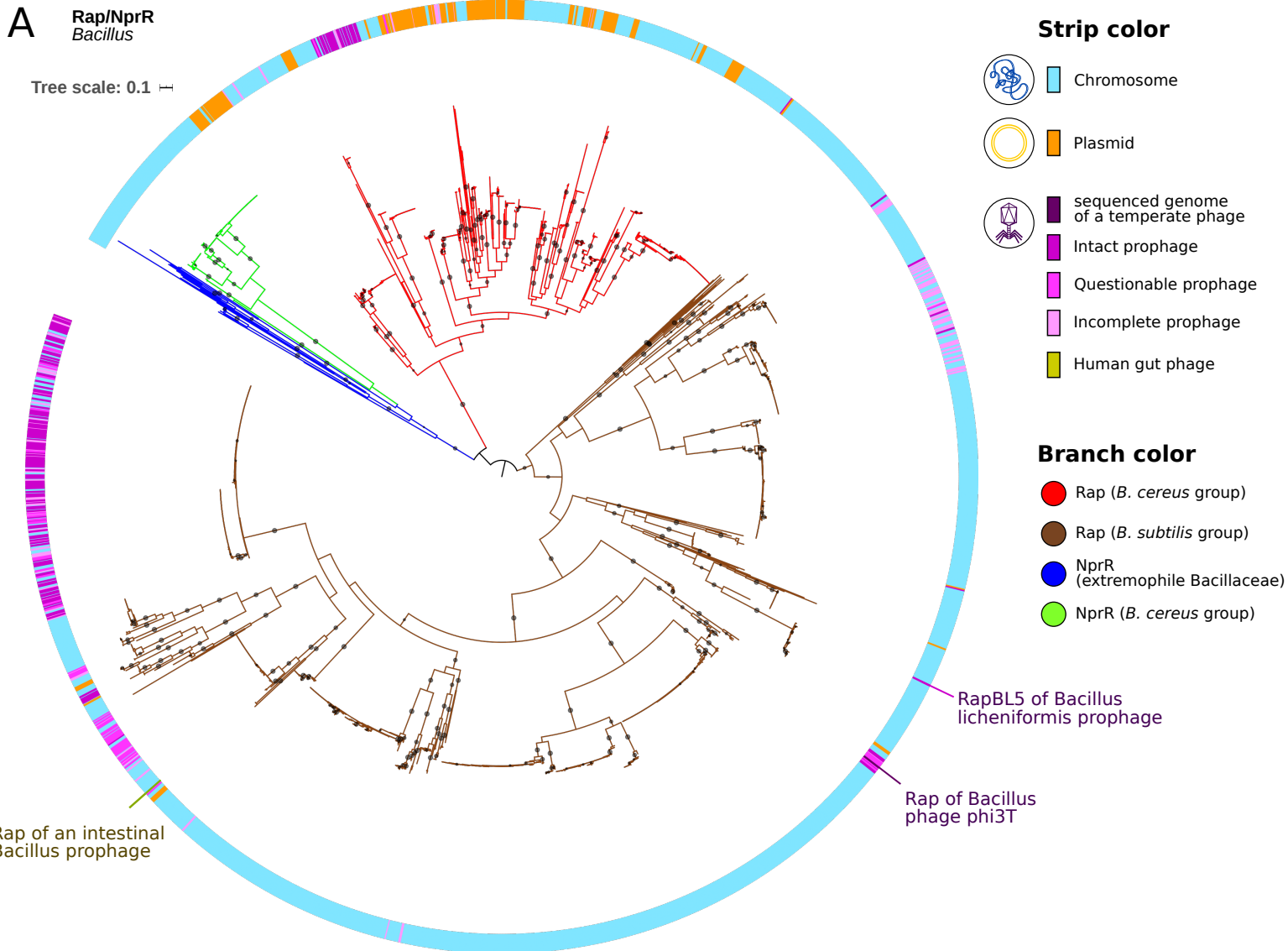


High densities



412 **Figure1**

413 **Mechanism of RRNPP QSSs in phages/prophages.** The receptor and the communication
414 propeptide of the phage-encoded QSS are in green and red, respectively. Upon bacterial expression,
415 secretion and maturation, the concentration of the communication peptide reflects the intracellular
416 density of phage genomes or prophages, which correlates with the number of lysed cells during the
417 lytic cycle and the number of lysogenized cells during the lysogenic cycle. At high concentrations of
418 the peptide, reflecting a quorum of phages/prophages, the receptor binds the imported mature peptide
419 via TPRs motifs and gets either turned-on or -off as a transcription factor or a protein inhibitor, which is
420 at the basis of density-dependent regulations of target genes (often adjacent to the QSS) or proteins.
421 These regulations thereupon coordinates a behavioral transition at the scale of the entire
422 phage/prophage population.

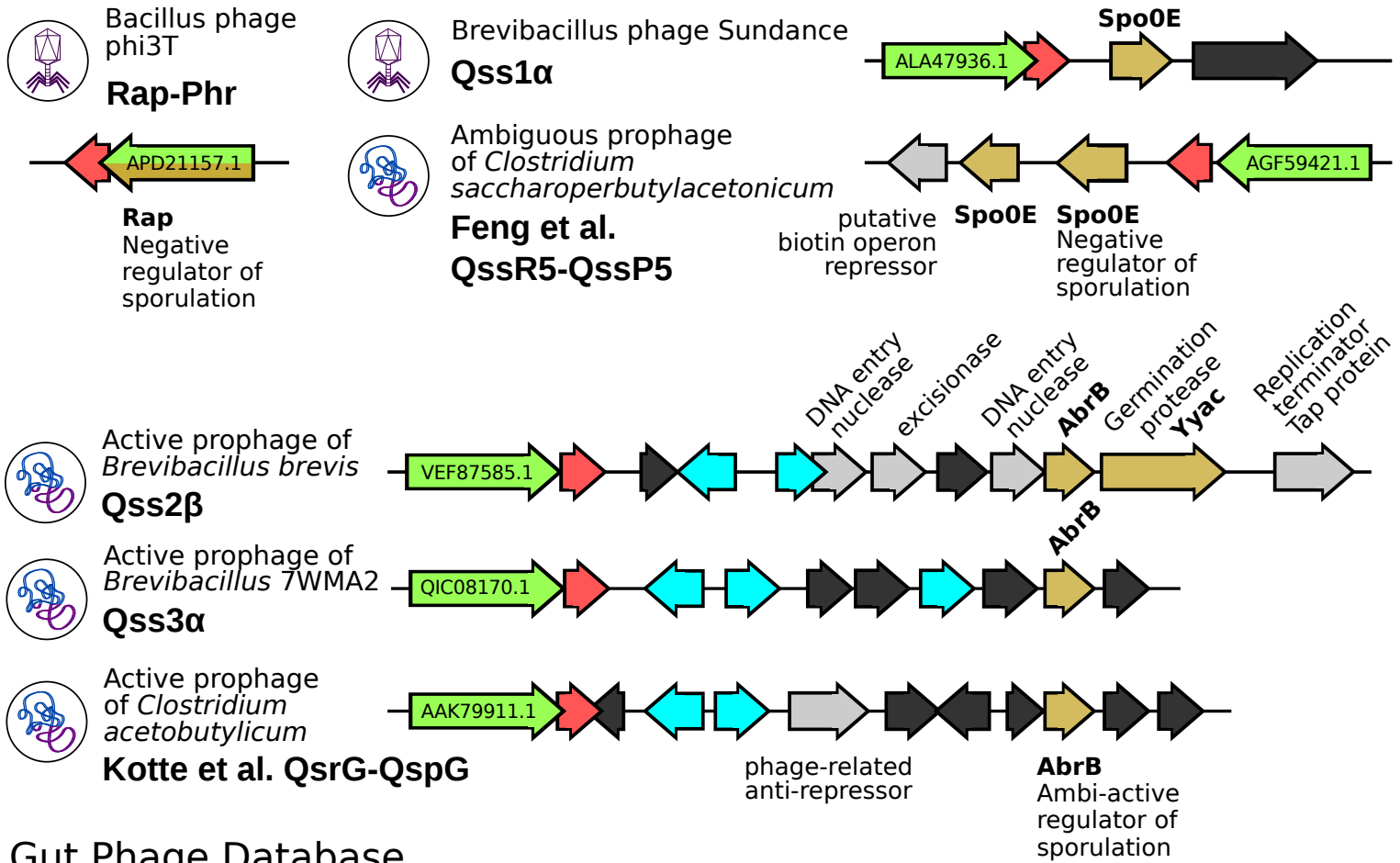


423 **Figure 2**

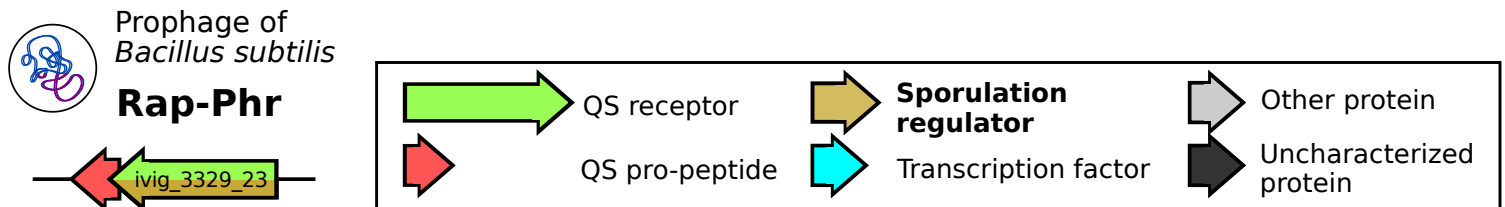
424 **QSS families comprising at least one phage/prophage-encoded candidate QSS**

425 **A. Maximum-likelihood phylogeny of receptors of the Rap/NprR family.** The figure displays the
426 maximum-likelihood phylogenetic tree of the family comprising the Rap (no DNA binding domain) and
427 the NprR (DNA binding domain) receptors that are part of a detected RRNPP-type QSS. The
428 clustering of Rap and NprR into the same protein family is consistent with the common phylogenetic
429 origin proposed for these receptors (S. Perchat, *et al.*, *Microb. Cell* 2016)). The tree was midpoint
430 rooted and a small black circle at the middle of a branch indicates that the branch is supported by 90%
431 of the 1000 ultrafast bootstraps performed. Branch colors are indicative of the type of receptor (Rap or
432 NprR) and of the bacterial group that either directly encodes the QSS or can be lysogenized by a
433 (pro)phage that encodes the QSS. The colorstrip surrounding the phylogenetic tree assigns a color to
434 each leaf based on the type of genetic support that encodes the QSS: blue for chromosomes, orange
435 for plasmids, dark purple for sequenced genomes of temperate bacteriophages, different levels of
436 purple for Phaster-predicted intact, questionable and incomplete prophages. The Rap receptors of
437 *Bacillus* phage phi3T (only Rap found in a sequenced genome of a temperate phage) and of *B.*
438 *licheniformis* intact prophage (viral Rap inserted into the host chromosome and shown to modulate the
439 bacterial sporulation and competence pathways) are outlined. **B. Sequence similarity network of the**
440 **other candidate QSS families comprising at least one phage/prophage-encoded QSS.** Each
441 node corresponds to a receptor sequence found adjacent to a candidate pro-peptide and is colored
442 according to the type of genetic element encoding the QSS, as displayed in the legend. The label of
443 each node indicates the NCBI id or a Gut Phage Database id of the candidate receptor. Each edge
444 corresponds to a similarity link between two receptors defined according to the following thresholds:
445 percentage identity $\geq 30\%$, alignment coverage $\geq 80\%$ of the lengths of both receptors, E-value \leq
446 $1E-5$. Each connected component of the graph thereby defines groups of homologous receptors and
447 is considered as a QSS family. The families are ordered from the largest to the smallest. Families with
448 a black label were already described before (but not necessarily in phages) while families with a red
449 label are novel. The most prevalent encoding-taxon in the family is displayed on top of it. The nodes of
450 the receptors that are part of a predicted sporulation-hijacking QSS are characterized by an additional
451 label.

NCBI Complete Genomes



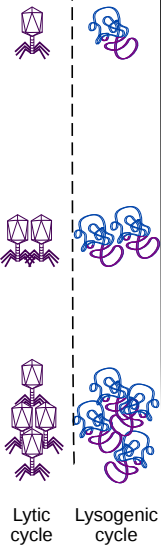
Gut Phage Database



452 **Figure 3.**

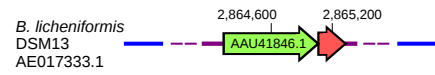
453 **Clustering of sporulation initiation regulators with QSSs found in sequenced genomes of**
454 **phages** (virion icon) **or prophages** (lysogenized chromosome icon). The name and the genome of
455 each QSS is written aside from each genomic context. Genes are colored according to their functional
456 roles, as displayed in the legend. Within the QSS receptor gene, the NCBI id of the protein is shown.
457 Because RRNPP QSSs tend to regulate adjacent genes, these genomic contexts hint at density-
458 dependent sporulation-hijacking mechanisms.

phage or prophage density

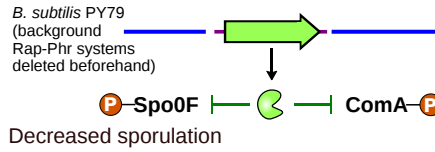


RapBL5-PhrBL5

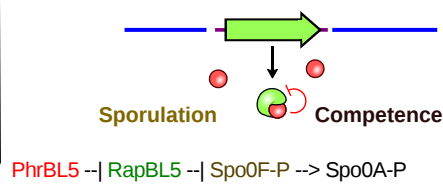
Phaster: intact prophage (2,854,951-2,867,485)
 ProphageHunter: active prophage (2,855,587-2,866,209)



rapBL5 cloned without phrBL5

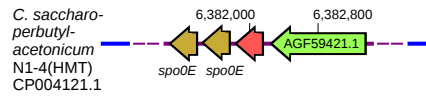


Addition of PhrBL5_{mature}

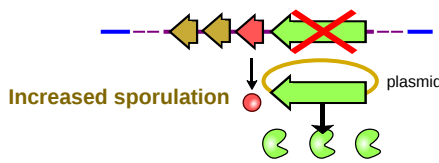


Qss5R-Qss5P

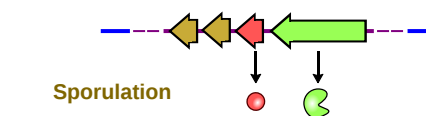
Phaster: incomplete prophage (6,390,448-6,416,714)
 ProphageHunter: ambiguous prophage (6,375,466-6,398,632)



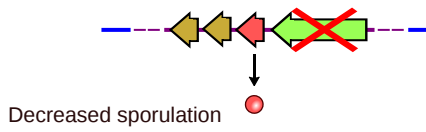
Overexpression of qss5R



Basal expression of the qss



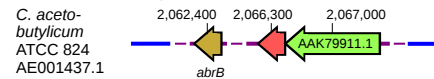
Mutation of qss5R



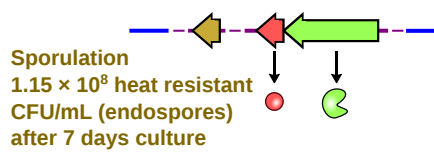
Hypothesis: Qss5P --| Qss5R --| Spo0E --| Spo0A-P

QsrG-QspG

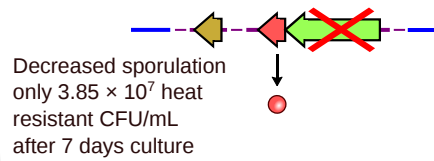
Phaster: intact prophage (2,017,737-2,083,827)
 ProphageHunter: active prophage (2,050,451-2,078,632)



Basal expression of the qss



Mutation of qsrG



459 **Figure 4.**

460 **Summary of the evidences that some viral QSSs dynamically influence bacterial sporulation.**

461 The functions of the three QSSs (one column each) were investigated in (17, 18, 26) as displayed in
462 this figure. The prediction that each QSS is encoded by a prophage is shown in purple at the top of
463 each column. A prophage qualified as active by Prophage Hunter means that it is predicted to be able
464 to reinitiate the lytic cycle upon excision from the bacterial chromosome. The interaction between the
465 quorum sensing peptide was only investigated for RapBL5-PhrBL5 and is still lacking for Qss5R-
466 Qss5P and QsrG-QspG. However, there is evidence in all QSSs that the receptor regulates the
467 sporulation pathway, via Spo0F-P inhibition for the Rap protein inhibitor, and likely via transcriptional
468 regulation of the adjacent *spo0E* and *abrB* genes for the Qss5R and QsrG transcription factors.

469

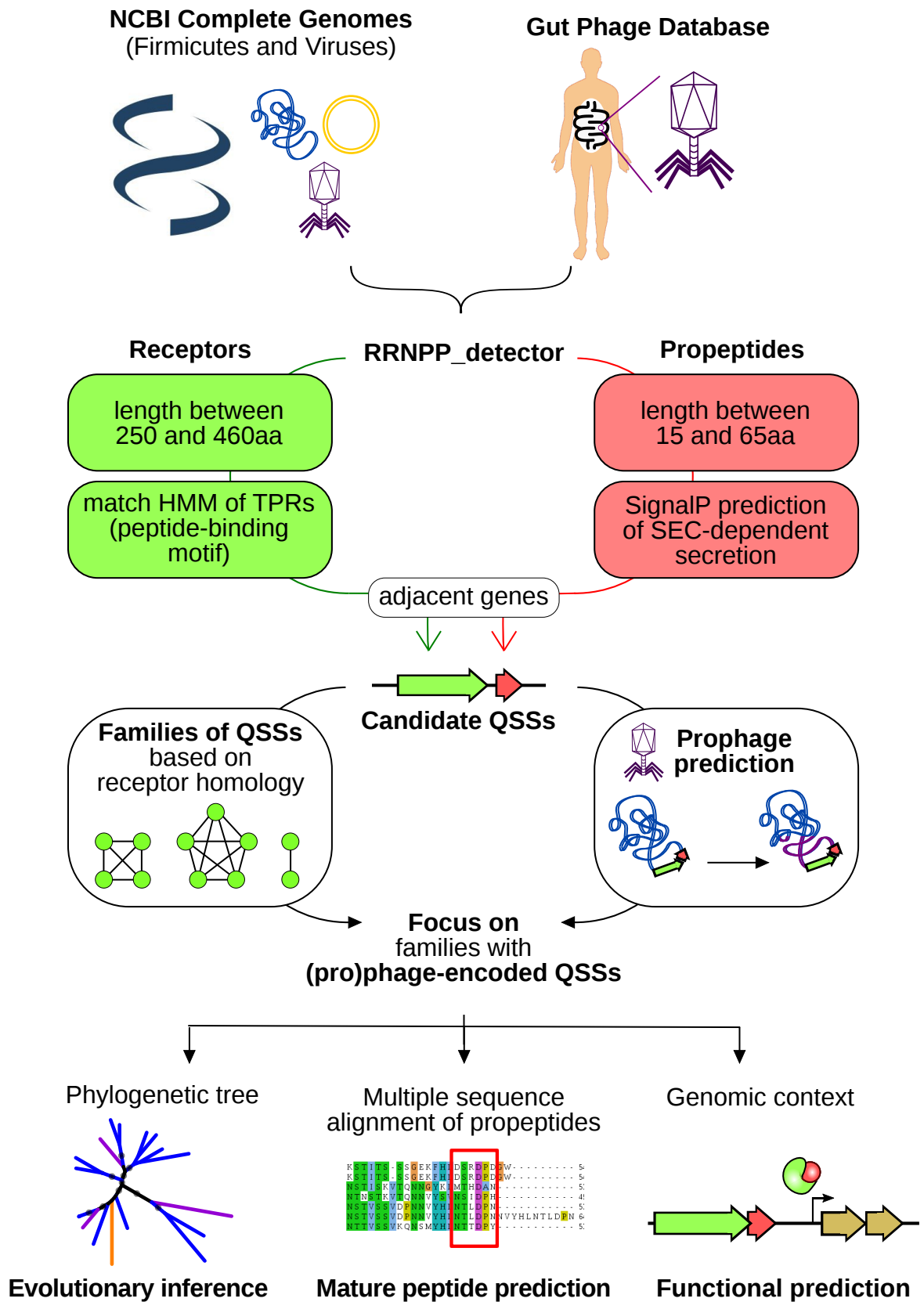
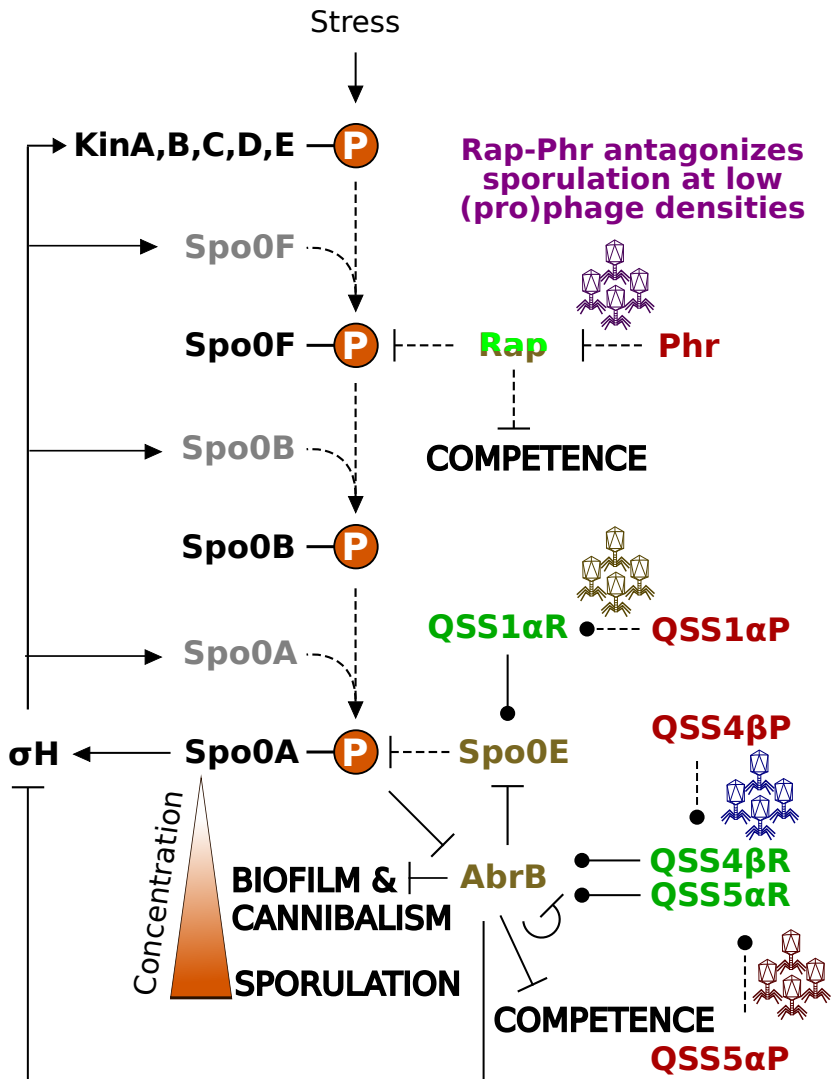


Figure S1: Study design.

RRNPP-detector was applied against complete genomes of Firmicutes and Viruses from the NCBI and against the Gut Phage Database. This algorithm defines candidate RRNPP QSSs as tandems of adjacent ORFs encoding a candidate receptor (250-460aa protein matching HMMs of peptide-binding tetratricopeptide repeats (TPRs)) and a candidate pro-peptide (15-65aa protein predicted to be excreted via the SEC-translocon by SignalP). 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.

Bacillus bacteria



Clostridium bacteria

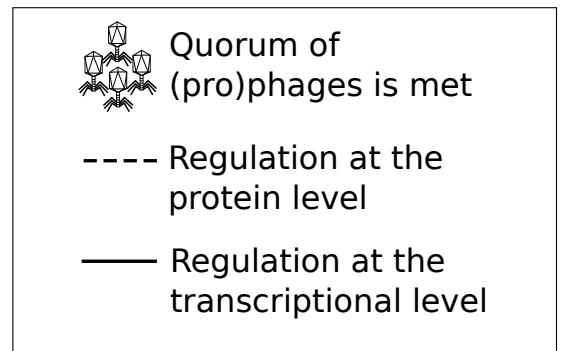
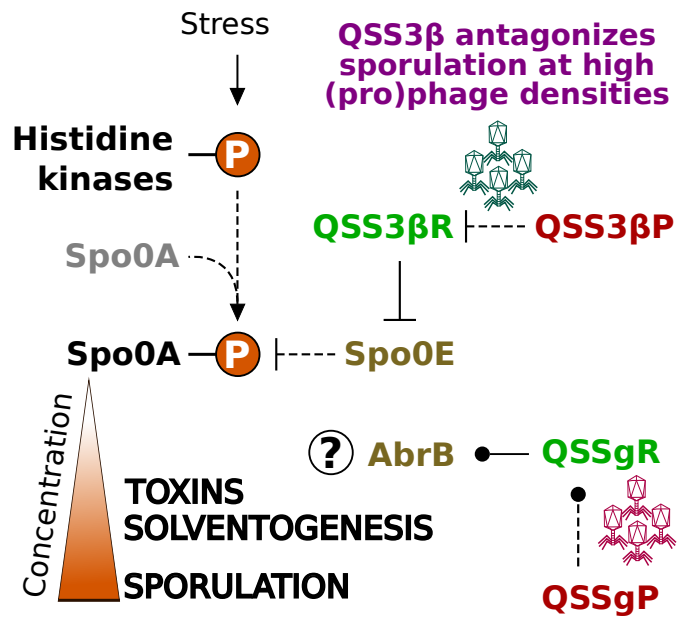


Figure S2: Predicted modulations of the sporulation initiation pathway mediated by phage/prophage-encoded QSSs

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 and the competence pathway in a density-dependent manner.

QSS1 family

| | | |
|--------------------------|--|----|
| ● VEF92013.1__15653/1-36 | 1 - MQKLLLVFALVVFSSPVVSEAYDP-----TSEPYRVPEPDW | 36 |
| VEF87511.1__15653/1-40 | 1 MKKWILSLCMLTAFVTL PVLVEQPAK-----QTNSIIQYGWPEPW | 40 |
| ● AIG26091.1__1534/1-45 | 1 - MRKKIVLMLLTLGLITNMGFALVKTDEANL-----NTFSMRANNADPGY | 45 |
| AWX59017.1__4003/1-59 | 1 - MKKKA SFLVALSMVLGSFLAIDVPYKNSNND SLLAVDVPYGI SKEKSN SPQRVPDPGY | 59 |
| ● ALA47937.1__5003/1-48 | 1 - MKRIFWGVSVFLTMFSF SLLALENDKTEHSTNKI-----QHAYAITVYTDNPGY | 48 |
| AIG25196.1__1534/1-46 | 1 MMKKIVISTII FSTFLSLTFIVYDKFMGVYSE-----KYSASKVYNEDPGH | 46 |
| AIG25890.1__1534/1-46 | 1 - MKKLVIVSVVFS SLSLSSVFV IDMNCSEKEYA-----SYQS IQVFFEDPGH | 46 |
| AIG24640.1__1534/1-45 | 1 - MKKIVVATSIILASLSLTF LFSIERQTENQI-----VSGHIKVTGDPGH | 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 MKKRLTVLMTASLVLCSSLFAVAIFSNSEHDV-----VYSSIKVYTDNPGH | 46 |
| AIG24914.1__1534/1-46 | 1 MKKYLALITTVSLVLLSFLTMEITGKFEVNN-----VASSMKPYTDDPQG | 46 |
| QIC07891.1__14625/1-45 | 1 - MKKYLTIIATVSVFVICS LFTVASFLDFEKYG-----EQSSAKTFHDDPQG | 45 |
| AIG25316.1__1534/1-45 | 1 - MKKTVVATVLSLTLVAMNVVPLYKTTDSNI-----TAFKVIPTVVDPGH | 45 |
| AIG28939.1__1534/1-44 | 1 - MKKVIGMLMSVTMLVTVTLTDSVVP TSG-----TSLOVQPSDYDPPGY | 44 |
| AIG27198.1__1534/1-48 | 1 - MKKLSILTLALAIICFISTESPTSDSNATTKKI-----GLYTIQQNHYPGY | 48 |
| QIC05360.1__14625/1-48 | 1 - MKKLSILTLALAIICFISTESPTSDSNATQKI-----GLYTIQQNHYPGY | 48 |
| AIG28870.1__1534/1-44 | 1 - MKKAIITIFMTLALVTSIGSIDSF SHVAEA-----SKPHAIQYVYDPPGF | 44 |

AloR family

| | | |
|--------------------------|---|----|
| AVF28101.1__8996/1-49 | 1 - MKI - KNWILCITVAITLIGGSSAFSTTTQATASN-----ATI ALS-DHGVGGDYAS | 49 |
| AVF32604.1__8997/1-49 | 1 - MKI - KNWILCITVAITLIGGSSAFSTTTQATASN-----ATI ALS-DHGVGGDYAS | 49 |
| AVF23670.1__8994/1-46 | 1 - MKI - KKYLVFLITLFSLLYVVGINPDKTHEAAK-----QSVIAYAENGVGEE---- | 46 |
| QHZ49846.1__14754/1-166 | 1 - MKI - KKYLVFLITLFSLLYVVGINPDKTHEAAK-----QSVIAYAENGVGEE---- | 46 |
| AVF24715.1__8996/1-58 | 1 MKKF - KKL LLSVVLAAAAILGGGIALNLPQTSSEDSAVAALHGTGGQAPANMSVAALHGTGQ---- | 58 |
| AVF29476.1__8997/1-58 | 1 MKKF - KKL LLSVVLAAAAILGGGIALNLPQTSSEDSAVAALHGTGGQAPANMSVAALHGTGQ---- | 58 |
| QHZ55460.1__2483/1-47 | 1 MKKFVKG FVLSLCLV LGLGFGINVHIADGKVITIK-----PPIIQY-DHGVGG----- | 47 |
| AHM68195.1__2742/1-33 | 1 - MGF - KNLKYMLIIG--LFGFLELS-----HTITVF-GHGTGA----- | 33 |
| ● VEF92630.1__15653/1-42 | 1 - - - M - RSMILLLTIVLFGLSHGIV-VPDHQEVLSH-----HVVTTY-THGAGG----- | 42 |
| VEF91032.1__15653/1-44 | 1 - MKL - KMFSLSFALFFSLCIGIVNS ETSSSNTMSR-----PNMVTY-THGGM----- | 44 |

Feng et al. family

| | | |
|-------------------------|--|----|
| ● AGF56488.1__2211/1-54 | 1 MKKLLLTMVACVLLSNVFAAATKTTSTTTDTKSTITS-SSGKFFHLD SRDPDGW----- | 54 |
| ● AQR95596.1__6751/1-54 | 1 MKKLLLTMVACVLLSNVFAAATKTTNTTTDTKSTITS-SSGKFFHLD SRDPDGW----- | 54 |
| AQR92768.1__6751/1-53 | 1 MKKIVLAVLATSMLLTNVAFASTETHTTKSTONSTISKVTQNGGYKLMTHDAN----- | 53 |
| AGF53864.1__2211/1-49 | 1 MKKII LAVIATSMLLTNVAFASTKTHH---KNTNSTKVTQNNVYSVNSIDPH----- | 49 |
| AGF55936.1__2211/1-53 | 1 MKKII LAVMATSMLLTNVAFASTKTHTKITKNSTVSSVDPNNVYHLNLTLDPN----- | 53 |
| AQR94680.1__6751/1-64 | 1 MKKII LAVMATSMLLTNVAFASTKTHTKITKNSTVSSVDPNNVYHLNLTLDPN----- | 64 |
| ● AGF59420.1__2211/1-53 | 1 MKKII LAVIATSMLLTNVAFASTKTHHTNTGKNTTVSSVKQNSMYHLNLTDPY----- | 53 |

QSS2 family

| | | |
|--------------------------|---|----|
| ● VEF87223.1__15653/1-63 | 1 - MKRILFATFAALMIGLVT PSTGVFAGPPENDKVAP - IAI DHS DGV EKF PRQK PPI IVVY GPPEND | 63 |
| ● VEF87586.1__15653/1-63 | 1 MKKKLIFIPSLALLIGLFTANEVSLAGPPDGWGIAPETKVIQIGDSEPTTWGVSPLG--GPPDGW | 63 |

QSS3 family

| | | |
|--------------------------|---|----|
| ● AIG27472.1__1534/1-48 | 1 - MKKLFAS S LVLFAFL LMS SSMVSTEESVQVTAEKATAHF IOYSTAPDW | 48 |
| ● QIC08171.1__14625/1-46 | 1 MKK KLVAF S LALVTLVGI - SSVVGAEKSIQNVH SKPSA--IOYITEPEW | 46 |

Kotte et al. family

| | | |
|-----------------|--|----|
| QspH/1-38 | 1 ---MNSRVK KLLFVLVCSFII---IGYLSVNHISV LASYPGWSW----- | 38 |
| AAK81614.1/1-41 | 1 - MGNMNSRVK KLLFVLVCSFII---IGYLSVNHISV LASYPGWSW----- | 41 |
| QspB/1-50 | 1 ---MLPKKSNK F ILVLI CAIVFVS AFALNTSGTR SLLGA EPTWGWNI SKLLF | 50 |
| ● QspF/1-39 | 1 ---MFFKRYK KLLFIAG ILAAV---LTINLGLNGNLLSGLPLMVW----- | 39 |
| QspD/1-41 | 1 ---MFFLKKHK KFLSMLVLT LVI---GVLVTTKGGI SILSMFPDAPW----- | 41 |
| QspA/1-41 | 1 - MI T V L K K H R F L I S V L T L G V L F --- L C I V F T N S S Y I L S G A P C I P W ----- | 41 |
| QspC/1-38 | 1 ---MFLKRHK L I F S I L L L G T L I --- L G V T F F K S - S L L A F Y P G M P W ----- | 38 |
| QspE/1-41 | 1 MR I V F F K R H K F L L T V L A L G I L F --- L G A T F F G N - S L L S A F P D L P W ----- | 41 |
| ● QspG/1-39 | 1 ---M L L K K H K L A F S V L A L T L L F --- L G V N F L G N N V G L S V S P M T P W ----- | 39 |

Figure S3: 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 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.

6. Génomique comparative des systèmes de QS et inférence de réseaux de communication

Comme évoqué en **section 2.5**, la prédiction de liens de communication entre deux génomes distincts sur la base de similarité de séquences entre composantes de leur QSS respectif est un problème complexe. Toutefois, si des méthodes robustes de prédiction pouvaient voir le jour, cela permettrait de reconstruire des réseaux hypothétiques de communication microbienne au sein de communautés environnementales à partir de MAGs. L'analyse de tels réseaux permettrait de mener des investigations sous le prisme de l'écologie. Le chapitre suivant du livre « Biocommunication of phages » publié chez *Springer*, décrit les idées à la base de ces inférences. Néanmoins, comme ce chapitre fut écrit en tout début de thèse, les subtilités décrites en **section 2.5** n'avaient pas encore été assimilées, et la complexité de la tâche ne fut pas vraiment appréhendée dans ces lignes. Néanmoins, ce chapitre offre l'intérêt de présenter quelques exemples d'analyses de tels réseaux de communication, et fournit ainsi quelques perspectives à cette problématique de recherche.

Chapter 8

Microbial Communication Networks: Sketching a Method for Analyzing the Communication of Bacteriophages Inside Environmental Communities



Charles Bernard, Philippe Lopez, and Eric Bapteste

Abstract Recent functional studies have shed light into how the combinatorics of genes associated with quorum sensing (QS) – often described as an entity-entity communication mechanism – may support different communication modalities in bacteriophages. Specifically, only systems of QS genes for phage to phage communication and eavesdropping on bacterial communication molecules have been characterized so far, which represents only a fraction of the spectrum of all the possible communication modalities predicted by this combinatorial logic. However, we argue that computational methods are already available to systematically mine the genomes of viruses and other microorganisms for QS genes, to compare these genes together across genomes, and to infer many novel links and types of communication between microbiological entities. All these putative communication links could be conveniently represented together in the form of a network, which would summarize which virus is suspected to interact with which microbiological entity, *via* which QS signaling molecule and foremost, under which communication modality. Besides, with the recent advents of metagenomics and metaviromics that enable accessing to genomes sequenced from the same environmental site, it would be theoretically possible to use this methodology to partially infer the big picture of communication inside a real community of viruses and cellular organisms. Finally, we discuss how the systematic analysis of such predicted microbial communication networks could provide insights into the many forms and “social” consequences that the biocommunication of viruses may imply.

C. Bernard

Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d’Histoire Naturelle, Paris, France

Unité Molécules de Communication et Adaptation des Micro-organismes (MCAM), CNRS, Museum National d’Histoire Naturelle, Paris, France

P. Lopez · E. Bapteste (✉)

Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d’Histoire Naturelle, Paris, France

e-mail: eric.bapteste@upmc.fr

© Springer Nature Switzerland AG 2020
G. Witzany (ed.), *Biocommunication of Phages*,
https://doi.org/10.1007/978-3-030-45885-0_8

163

8.1 Microbial Communication Via Quorum Sensing (QS) Involving Bacteriophages

Temperate bacteriophages are viruses that can insert all or part of their genetic information into their bacterial host genomes in a process named lysogeny, as an alternative to the host-destructive lytic cycle, through which they replicate and spread. When a temperate bacteriophage lysogenizes a bacterium, it becomes dormant under the form of a “prophage” and confers the lysogenized bacterium immunity to further infection by the bacteriophage free form (virion) (Oppenheim et al. 2005; Abedon 2015). Lysogeny can therefore be seen as a means to preserve the phage-host collective, notably when the lytic cycle starts becoming a threat for the survival of the host population, upon which the own phage population survival is conditioned. Consistently, when the number of host cells happens to be limited, fast growing fully-lytic phages may quickly end up eradicating their hosts (and therefore themselves) whereas temperate bacteriophages are able to maintain a slow but continuous growth through the back and forth switch between lysis and lysogeny, a strategy shown to be evolutionary stable (Wahl et al. 2018). Specifically, in order for this switch to be an optimal evolutionary strategy, it should ideally be synchronized within the viral population, and occur sufficiently late during the infection to ensure that the phages have produced a sufficiently high number of progeny at the expense of the host population (maximization of the phage replication) but also sufficiently early to ensure that the population of hosts is still composed of enough individuals (protection of the phage-host collective). In other words, from the phage perspective, the switch from lysis to lysogeny should ideally intervene in a coordinated manner, and be dependent on both of the phage and the host population densities.

One well-known microbial mechanism that orchestrates the emergence of a collective behavior in a density dependent manner is called “quorum sensing” (QS) (Fuqua et al. 1994). *Via* this mechanism, initially identified in bacteria and further extended to all superkingdoms of evolving entities (bacteria (Miller and Bassler 2001), archaea (Paggi et al. 2003; Johnson et al. 2005; Zhang et al. 2012), eukaryota (Mehmood et al. 2019) and viruses (Erez et al. 2017)), individual microorganisms “signal” their presence by secreting a small diffusible molecule, the concentration of which indirectly reflects the density of the emitting population. The quorum, or the population density threshold above which the coordinated behavior is triggered, is next materialized by the concentration at which the signaling molecule starts to be effectively sensed by its cognate receptors within (or at the surface of) each individuals, and transduced to regulate the response at the basis of the emergent behavior. Mechanistically, a canonical QS mechanism then relies on a system of genes composed of three components: a synthase (or a cluster of synthases), a receptor and a response regulator, responsible for the signal production, perception, and downstream transcriptional response, respectively. Interestingly, because a shared complete QS system within a population of microorganisms allow, among other ends (Cornforth et al. 2014), one individual to produce a signal that can influence the behavior of another individual (and *vice et versa*), QS has often been described

as a cell-cell communication mechanism (Waters and Bassler 2005). Besides, the specificity, or secrecy, of a QS signaling molecule may range from intra-species to inter-kingdom levels (Sztajer et al. 2014; Mukherjee and Bassler 2019) and may thus entail relationships of co-influence between distinct species of a same microbial community. Last but not least, the functional complementation of individual QS components of QS system(s) across different genomes may give rise to other forms of relationships than intra or inter-entity communication, such as, for example, eavesdropping (perception of exogenous communication molecules) (González and Venturi 2013; Karavolos et al. 2013; Silpe and Bassler 2019a) or manipulation (emission of a QS signaling molecule that happens to impair certain biological functions in another organism) (Hochstrasser and Hilbi 2017; Lee et al. 2018).

Surprisingly, although the density-dependent behavioral transitions mediated by QS could have appeared as an elegant candidate mechanism to control the lysis-lysogeny transition in temperate bacteriophages, 47 years passed between the first characterization of QS in the proteobacterium *Alivibrio fischerii* in 1970 (Nealson et al. 1970) and the first description of a functional QS system of genes inside viral genomes (in *Bacillus phages phi3T* and *SPbeta*), which was furthermore guiding the lysis-lysogeny decision upon *Bacillus* infection (Erez et al. 2017). The groundbreaking discovery of this first phage-phage communication through QS possibly gave rise to emulation, because two years after its characterization, another form of communication modality has been characterized in bacteriophages, namely, eavesdropping on host QS signals. The next subsections will briefly describe all these fascinating discoveries.

8.1.1 Phage to Phage “Arbitrium” Communication System

Although, QS components had already been found in genomes of bacteriophages at that time (Hargreaves et al. 2014), the first experimental characterization of a phage-phage communication mediated by a viral QS system dates from 2017 (Erez et al. 2017). Much like the QS systems of firmicutes (Bhatt 2018), the uncovered viral QS system supports a highly secretive, intra-specific mode of communication, relying on the so-called “arbitrium” short signaling peptide (SAIRGA in *phi3T*, GMPRGA in *SPbeta*), rather than on a hormone-like signaling molecule of variable specificity, as observed for QS in proteobacteria. Upon *Bacillus* infection, the arbitrium peptide is released in the medium by cleavage of the AimP propeptide, accumulates as the phage replicates, and is sensed at high concentration by the AimR transcription factor, which then ceases to transactivate the AimX negative regulator of lysogeny. The arbitrium-based QS system thereby enables a population of infecting phages to synchronously lysogenize when it becomes critically large. Since then, hundreds of homologous QS systems have been found in the genomes of other phages and prophages and demonstrates that QS through intra-specific communication molecules is actually widespread in bacteriophages (Stokar-Avihail et al. 2019).

8.1.2 *Host Eavesdropping Enabled by the Externalization of a QS Receptor from the Host to the Phage*

Two years after the characterization of the arbitrium QS, a *Vibrio* phage encoded protein that is homologous to the VqmA sensor of the 3,5-dimethylpyrazin-2-ol (DPO) QS molecule secreted by *Vibrio cholerae* was uncovered (Silpe and Bassler 2019a). The Vqma_{phage} encoding gene thus probably results from an externalization, *i.e.*, the copy of genetic material from the host to the phage (Corel et al. 2018). Unlike the arbitrium peptide, the host-produced DPO signal is typical of the quorum sensing molecules secreted by proteobacteria, namely, small diffusible secondary metabolites biosynthesized by enzymes, with a spectrum of specificity that can range from intra-species to inter-kingdom levels of recognition (Papenfort and Bassler 2016). Consistently, the authors showed that DPO is produced in part by the threonine dehydrogenase (Tdh) of *V. cholerae*, and binds to both of the VqmA sensor of *V. cholerae* and of its homolog in *Vibrio phage VP882*. Thereby, Silpe and Bassler unraveled the first instance of a unidirectional communication link from a host bacterium towards a bacteriophage. Precisely, the phage-encoded VqmA transcription factor has been shown to confer the *Vibrio phage VP882* the ability to eavesdrop on the density status of its host *via* DPO sensing and to enact the lytic program accordingly.

Shortly after their discovery, Silpe and Bassler uncovered a putative analogous phenomenon of eavesdropping involving *Aeromonas phage phiARM811d* and the *Aeromonas* bacterial host genus (Silpe and Bassler 2019b). The authors noticed that *phiARM811d* encodes a QS receptor of the LuxR family, which detects Acyl-Homoserine Lactones (HSL), a type of QS molecule produced by many *Proteobacteria*, including *Aeromonas* species. Specifically, the viral LuxR protein of *phiARM811d* presents 60% sequence identity with the bacterial LuxR protein of *Aeromonas popoffii* CIP 105493 and exhibits high affinity towards C4-HSL, the only N-Acyl-Homoserine Lactone that is produced ubiquitously within the *Aeromonas* genus according to the Sigmol database of QS (Rajput et al. 2016). On the basis of these results, Silpe and Bassler speculated that the LuxR sensor of *Aeromonas phage phiARM811d* could enable eavesdropping on the host-produced HSL QS signals, presumably to optimize the timing of execution of particular steps during the viral infection.


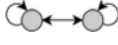

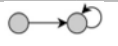
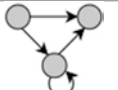
8.1.3 *State of the Art Knowledge of Communication Via QS in Bacteriophages*

In a nutshell, hundreds of homologous QS systems in bacteriophages have been so far characterized as a means for a phage to communicate on its own population density, and two distinct externalized QS receptors as a means of eavesdropping on

the QS signals of bacterial hosts. So far, the cues gathered by the characterized phage QS components on either the phage or the host population densities have been shown to serve as basis to trigger behavioral transitions during a viral infection. These transitions initiated by QS have all been shown to sophisticatedly balance the trade-off between phage replication (done at the expense of the hosts) and the preservation of the phage-host collective (upon which the phage survival is conditioned), both essential for the evolutionary stability of the temperate lifestyle.

Finally, the espionage on host QS signals mediated by the externalization of the Vqma or of the LuxR QS sensors is particularly interesting from an evolutionary viewpoint, as it shows that there is not necessarily such thing as a phylogenetic limit to microbial communication. Indeed, it demonstrates that as long as two evolving entities happen to be in interaction together in a feedback loop and form thereby a collective (in such case, the phage-host collective), they might, in some cases, eventually develop a link of communication through the externalization of QS components, may these evolving entities have a common phylogenetic origin or not. This opens many perspectives in the field of biocommunication of viruses as it allows speculating on many other communication links, involving many other species, from perhaps different superkingdoms. Moreover, beyond the externalization of QS sensors, it enables one to speculate on the potential externalization of host QS synthases or full QS systems that would support other types of communication modalities than phage-phage communication or eavesdropping on bacterial QS signals: for instance, host manipulation or even phage-host bi-directional communication (Table 8.1). In this chapter, we will propose a methodology to infer and analyze microbial communication network involving viruses, which could likely shed light into the many forms that biocommunication of viruses, including bacteriophages, may take.

Table 8.1 Type of edges in a microbial communication network

| Communication modality | Node 1 | Node 2 | Genome 1 | Genome 2 |
|--|---|--------|------------------|---|
| Canonical QS (intra-entity communication) |  | | - QS synthase(s) | |
| | | | - QS receptor | |
| Inter-entity communication |  | | - QS synthase(s) | - QS synthase(s) <small>homolog</small> |
| | | | - QS receptor | - QS receptor <small>homolog</small> |
| Eavesdropping |  | | - QS receptor | - QS synthase(s) <small>homolog</small> - QS receptor <small>homolog</small> |
| Signal amplification |  | | - QS synthase(s) | - QS synthase(s) <small>homolog</small> - QS receptor |
| Unilateral signaling by functional complementation |  | | - QS synthase(s) | - QS receptor |

8.2 Knowledge-Based Communication Network Involving Bacteriophages: An Introduction to Microbial Communication Networks

As we have seen in the previous section, the number of full or partial QS systems characterized inside complete genomes of bacteriophages has rapidly risen since 2017 and these recent discoveries will surely emulate many others in the near future. Because these viral QS systems involve multiple genes of different functions, and may serve as support for multi-partners, multi-kingdoms or even multi-systems communication, their qualitative descriptions are already, and will become more and more limiting to represent the whole picture of communication involving bacteriophages, other viruses and their hosts in a comprehensive way without computational methods. In this section, we will introduce the notion of “communication network”, a framework particularly suited for algorithms, by drawing a simple graph from the previous qualitative descriptions of viral communication modalities. In the next section, we will take advantage of this reference knowledge-based network to attempt justifying each of its edges by an underlying rule of construction and imagine other rules for other types of edges. Indeed, a complete dictionary of formal rules may be used to substantially expand this first knowledge-based communication network and to provide thereby a rich and flexible analytical framework to study viral and microbial chemical communication *via* QS.

In the knowledge-based communication network, nodes could represent microbial genomes, and edges between nodes could be directed in order to capture the nature of the communication link between interacting entities. Basically, the nodes of this network would then correspond to *Bacillus phage phi3T*, *Bacillus virus SPbeta*, *Vibrio phage VP882*, *Vibrio cholerae*, *Aeromonas phage phiARM81ld* and *Aeromonas popoffii CIP 105493*. The first type of edges to be drawn could be loops, namely arrows that connect a node to itself, for entities that use QS to estimate their own population density and therefore “speak” and “listen” to their own kind. Hence, loops would be drawn for *phage phi3T* (SAIRGA arbitrium-based QS), *phage SPbeta* (GMPRGA arbitrium-based QS), *V. cholerae* (DPO-based QS) and *A. popoffii* (HSL-based QS). Additionally, the direction of an edge could depict the direction of the QS signal between two partners of communication. Hence, arrows that start from a host bacterium and points towards a phage would capture the concept of eavesdropping on host QS signals. Such arrows would then be drawn from *V. cholerae* towards *Vibrio phage VP882* (DPO eavesdropping), and from *A. popoffii* towards *Aeromonas phage phiARM81ld* (HSL eavesdropping). In the next sections, we will see that other types of edges could actually be drawn on such a network. Here, because we imposed that this communication network should strictly depict the actual knowledge of QS involving bacteriophages, these additional types of edges cannot yet be drawn because there is no experimental evidence to support them. Finally, from a visualization point of view, one could make this network even more analytically tractable by labelling the nodes according to qualitative attributes of genomes; for instance, blue for bacterial species and purple for viruses. Second,

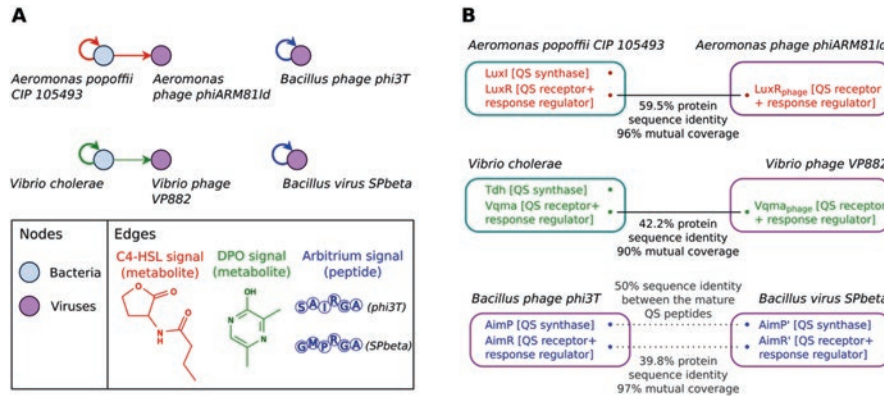


Fig. 8.1 Knowledge-based communication network involving bacteriophages. (a) Network depicting the state of the art knowledge of communication via QS molecules involving bacteriophages. (b) Genome content in QS genes of each node of the network and sequence identities between homologous QS components across genomes

one could color the edges according to the QS signal that support the interaction: blue for the arbitrium peptides, green for DPO and red for HSL. Ultimately, the resulting knowledge-based microbial communication network would look like as in Fig. 8.1a.

However, it should be said that *Bacillus phages*, *V. cholerae* and *A. popoffii* CIP 105493 are only representative users of the arbitrium-based, DPO-based and HSL-based QS, respectively. As a result, the number of new nodes and new edges would considerably grow if one desires to picture by hand all the possible communication partners of the phage nodes. Moreover, it should be duly noted that edges of this network are limited exclusively to QS signals produced and/or recognized by bacteriophages. One should keep in mind, for instance, that *Vibrio cholerae* uses at least three other QS signals than DPO: Cholera Autoinducer-1 (CAI-1), Autoinducer-2 (AI-2) and Ethanolamine, responsible for intra-species, inter-species, and inter-kingdom communication, respectively (Jung et al. 2015; Watve et al. 2019). These observations introduce the issue that, if comprehensiveness is sought to describe communication in the microbial world, the manual construction of such a network based on the literature would become an overwhelming task. However, this expansion could be easily automated with computational methods, on the basis of formal rules. To facilitate their introduction, we purposely indicated the QS protein component(s) encoded by the genome of each interacting species in Fig. 8.1b. As we will see in greater details in Sect. 8.4, the advantages of inferring and analyzing a comprehensive communication network with computational methods are many, including the possibility to quantify the different types of communication modalities that a bacteriophage may support (intra-specific communication, inter-entities communication, eavesdropping, signal amplification ...), to identify new putative phage-host associations, to assess the confidentiality of each phage QS signals, etc.

8.3 Genome-Based Underlying Logical Rules to Automate the Construction of a Microbial Communication Network

The construction of the knowledge-based network of communication involving bacteriophages relies on qualitative descriptions but we argue that all the edges that connect the nodes of this network can actually be justified with implicit formal rules. Specifically, we propose to define these rules according to characteristics of genomes that can be mined easily with bioinformatics. We will therefore establish them based first on the genome content in QS component genes to detect canonical QS mechanisms (loops), and second, on the sequence identity between QS components across genomes to infer instances of bi- and uni-directional communication between entities (other edges) (Fig. 8.1b). Before describing all these rules in detail, an overview of their basic principles as well as all the types of interactions that they might capture are given in Table 8.1.

One may notice that we took only into consideration QS synthases and receptors (including receptors with a response regulator activity (one-component systems)) but not response regulators. The reason for this is that their sequence similarity with transcription factors unrelated to QS would possibly lead to too many situations whereby a transcription factor inside a genome would be incorrectly detected as a QS response regulator.

8.3.1 Loops for Canonical QS (Intra-entity Communication)

- If the synthase(s) and the receptor of the same QS signal are together present inside a genome, then a loop is drawn for the node representing this genome (*e.g.* presence of the AimP QS synthase and of the AimR QS receptor/response regulator in the genome of *Bacillus phage phi3T*).

8.3.2 Bi-directional Arrows for Inter-entities Communication

If two taxonomically different entities both encode a complete QS system relying on the same QS signal, they could theoretically communicate together. However, it is not trivial to predict whether two distinct entities will rely on the exact same QS signal(s) and could thus “speak” and “listen” likewise based on the presence of homologous QS systems inside their genomes. Such prediction would be pretty straightforward for different entities that share an identical QS system but this best case scenario is not to be expected quite often. Then, it feels like some thresholds of sequence identities between the components of a shared QS system across two distinct genomes could be set to decide whether the homologous QS systems are

similar enough to likely produce at least one QS signal in common and support inter-genomes communication, or not. However, we advertise that such thresholds should be specific to the mode of production of the QS signal. Indeed, as we have seen earlier, the likelihood that a QS signal will serve as support for intra-entity or inter-entities communication is largely determined by the mode of production of the signal. When the QS signal is a peptide, and if this peptide (or its pro-peptide) is ribosomally synthesized as in the firmicutes phylum, its sequence is carried by a corresponding encoding region on a gene, and one non-synonymous mutation in this region is enough to alter the final conformation of the QS signal and therefore its recognition by the cognate receptor (*e.g.* SAIRGA and GMPRGA in *Phi3T* and *SPbeta*). On the other hand, when the QS signal is a metabolite or a non-ribosomally synthesized peptide, whose synthesis are catalyzed by one or several enzymes as in the proteobacteria or actinobacteria phyla, there is less chance that one substitution in the sequence of this or these enzyme(s) will drastically alter the structure of the produced signaling molecule. Accordingly, one could set a condition on the nature of the QS signal to formalize the rules for drawing a bi-directional edge between two entities:

- If the QS signal is a peptide: if the two genomes encode an identical QS peptide and if they both share a homologous cognate receptor as defined by arbitrary stringent thresholds (*e.g.* sequence identity >70% over more than 90% of mutual coverage) then a bi-directional arrow between the nodes representing these two genomes is drawn.
- If the QS signal is a metabolite: if both of the synthase(s) and of the receptor of this signal are homologous across two genomes, as defined by more relaxed thresholds (*e.g.* identity >40% over more than 80% of mutual coverage) then a bi-directional arrow between the nodes representing these two genomes is drawn.

Here these thresholds are empirical but are nonetheless supported by the sequence identities between QS components that are known to give rise or not to an asserted communication link between two genomes (Fig. 8.1b).

8.3.3 *Uni-Directional Arrow That Starts from a Node with a Loop for Eavesdropping*

- If genome A has a full QS system (loop) and if genome B encodes a protein that is homologous to that of the QS receptor of A (see thresholds in 8.3.2) but does not possess the biosynthetic pathway of the associated QS signal, then an uni-directional arrow from the node of A to the node of B is drawn (*e.g.* homology between the QS receptors Vqma of *V. cholerae* and VqmA_{phage} of *Vibrio phage VP882* but absence of homolog of the Tdh QS synthase of *V. cholerae* in *Vibrio phage VP882*).

8.3.4 *Uni-Directional Arrow That Reaches a Node with a Loop for Signal Amplification*

This case is the converse of the eavesdropping scenario, and could correspond to “signal amplification”. It would imply the situation where entity A and entity B would conjointly produce the same QS signal but only B would perceive it, and would perhaps be led (or misled) to interpret the density of the collective as its own population density.

- If genome A possesses the QS synthase(s) but not the cognate receptor of the QS system of genome B (see thresholds in 8.3.2), then draw an uni-directional arrow from the node of A to the node of B

8.3.5 *Uni-Directional Arrow That Reaches a Node Without a Loop for Unilateral Signaling*

This case would be the one where neither genome A nor genome B would have a complete QS system but each of them would possess the QS component that would functionally complement the one of the other. From an evolutionary viewpoint, this situation can be envisaged only if the unit of selection of these entities would not correspond to the organismic level but to the collective level (*e.g.* relation of interdependence between A and B). Nevertheless, this case implies that A and B would have no QS genes in common to create a mean of comparison, and therefore a bridge between the two genomes. However, the presence in the network of a third genome, say C, could solve this issue as long as it would have the complete QS system. The rule for this second case would then be the following:

- If genome C has a complete QS system, and if genome A only possesses the QS synthase(s) of this system whereas genome B only the QS receptor, then draw uni-directional arrows from A to C, from C to B and from A to B

8.3.6 *Pre and Post Processing of These Rules*

All the rules of construction of edges between entities in a communication network rely on the comparison of QS components across genomes. Therefore, a tool to automatically detect full or partial QS systems of genes inside genomes and compare their sequence identities together is a prerequisite preprocessing step. Such a tool should therefore rely on a database of protein sequences corresponding to various QS systems of reference. On another note, in order to circumscribe the resulting microbial communication network to the study of biocommunication involving viruses, one could optionally set an ultimate filter: any node corresponding to a

prokaryote or a eukaryote that is not connected to a viral node should be deleted from the network. Also, a last post processing operation could be to weigh each edge of a network according to the percentage of sequence identity between homologous QS components in order to quantify the likelihood of any inferred communication link.

8.3.7 Should Genes of Prophages Belong to Viral or Bacterial Nodes?

Besides, a tool to detect prophages inside bacterial genomes could be required to consider prophages as viral nodes or at least to distinguish between (i) genuine links of communication between phages and hosts due to externalization of QS genes and (ii) possible artefactual links of communication between phages and hosts originating from the sharing of QS genes between phages and prophages. However, one may also envisage that QS genes resulting from the internalization of a prophage inside a bacterial genome could possibly be expressed by the lysogenized bacterium and confer it a new functional QS system of genes. Indeed, it has been proposed that prophages may contribute to bacterial host phenotypes by providing adaptive genes (Schuch and Fischetti 2009; Oliver et al. 2009). Therefore, assigning a special annotation to prophages could be interesting if one would like to study the possible role of “social vector” exerted by temperate phages through the transfer of QS genes across bacterial genomes of the same community.

8.4 Microbial Communication Network: Examples of Perspectives That This Rich and Flexible Analytical Framework May Open

Now that we have introduced how to systematize the construction of a communication network from the protein sequences encoded by publicly available reference genomes, we will discuss which kind of perspectives its analysis may open. The goal of this section is of course not to be exhaustive but to give few examples of analyses that might be performed on such a network, and which kind of insights these might bring.

At the global level, the examination of the number of nodes and number of edges of this network could for instance allow to appreciate how much more viruses than the ones already characterized as QS users are suspected to communicate, and how much more microbial organisms are predicted to be involved in a communication link with viruses. Also, quantifying each type of edge involving a viral node (inter-entity communication, eavesdropping, signal amplification ...) could tell us which communication modalities are predicted to be widespread in viruses and which are

the ones that seem rather exceptional. Additionally, counting the number of edges that are supported by genes of prophages could allow assessing to which extent bacteriophages may contribute to the diffusion of QS genes across bacterial genomes.

On another note, filtering a whole communication network by the labels of edges would produce as many subnetworks as there are type of QS signals predicted to be used by viruses and could therefore allow quantifying how many and which different “languages” are suspected to be employed by viruses. Conveniently, the number of occurrences of a node in each of these subnetworks would also directly tell us how many distinct (complete or partial) putative QS systems are present inside its corresponding genome. Indeed, from a biological viewpoint, the identification of viruses that are suspected to be “multilingual” would be particularly interesting. Specifically, if a unique QS system allows a phage to assess its population density in a binary way (basically, below or above a given quorum), the accumulation of distinct QS systems could enable it to resolve its density with more refinement, through a succession of discrete measurements, corresponding to the distinct quorum associated with each QS signal. It has been demonstrated that these QS-based distinct steps could enable a population of growing proteobacteria to trigger different actions at different timing, much like a developmental program (Mehta et al. 2009). Hence, the prediction of plural QS systems in viral genomes (nodes with loops corresponding to distinct labels and thus distinct QS signals) could pave the way to the discovery of an analogous phenomenon in bacteriophages, namely that a hierarchy between distinct QS systems inside a same viral genome could control the transition between different stages of an infectious program. On another hand, even more exciting would be the discovery of plural QS systems with different levels of both completeness and specificity inside viral genomes. For instance, if a bacteriophage had a complete and intra-specific QS system to mediate phage-phage communication in addition to a QS receptor to eavesdrop on the density status of its host (node with a loop of one color and an edge directed towards it of another color), these QS components could theoretically enable it to contextualize its own density with respect to the one of its host. For instance, such a configuration could enable a bacteriophage not only to trigger lysis when it senses that its susceptible hosts are many (*e.g* DPO eavesdropping of *Vibrio phage VP882*) or trigger lysogeny when it senses that its own population becomes critically large (*e.g* arbitrium phage-phage communication system), but to switch between lysis and lysogeny at an optimal ratio of susceptible hosts available per virion.

Another example of analysis would be linked to the identification of connected components (CCs) inside these subnetworks. Basically, a CC designates a set of nodes within which each node is at least connected to another. Distinct CCs are then clearly identifiable as they have no connections between them. Specifically, the analysis of the composition of a CC may allow assessing the level of confidentiality of any communication molecule that a virus is suspected to rely on. For instance, small CCs, that may be composed of a single node, could depict an entity that rely on a very secretive QS signal (*e.g* the SAIRGA arbitrium peptide (Fig. 8.1)) whereas

large CCs which harbor a substantial phylogenetic diversity could encompass entities that may “speak” and/or “listen to” a less confidential “language”, such as, for example, the autoinducer-2 QS signal is suspected to be in bacteria (Sun et al. 2004). To sum up, the distribution of the viral nodes inside different classes of CCs (defined by a combination of their number of nodes, number of edges and intra phylogenetic diversity) may allow quantifying the confidentiality of each QS signal predicted to be used by viruses.

However, it should be said that a CC does not always depict a group of co-communicating species. For instance, two nodes may not be directly connected inside the same CC if their respective QS components (yet belonging to the same family of QS systems) were too divergent from each other but if “intermediate” QS components from other genomes nonetheless made the bridge between these distant nodes. This being said, a useful property of a network could be used to specifically detect preferential “discussion” groups inside CCs, namely, its modularity. The modularity of a CC or of a network is a scale that measures the density of edges inside communities of nodes to edges outside communities. Certain algorithms, such as the “Louvain method for community detection” aims at identifying set of nodes that optimize this value in order to identify community of nodes, or in other words, clusters of densely connected nodes (Blondel et al. 2008). Such a cluster of nodes, wherein most nodes are connected to each other, could therefore depict microbial and/or viral entities that may all inter-influence each other and form a preferential “discussion” group. From a broader point of view, identification of such groups could perhaps help discovering potential collective of entities. To follow up on this idea, the identification of a community of nodes that would include bacteriophages with many other bacteria (e.g *Vibrio phage VP882* and many other *Vibrio species*) could perhaps help identifying, in an unsupervised way, the spectrum of possible hosts for a given bacteriophage. Also, identifying “discussion” groups wherein the nodes are taxonomically different and the weight of edges are maximal (close to 100% sequence identity between QS components) would perhaps give more support to the hypothesis according to which bacteriophages may exert the role of “social vectors” in microbial community by facilitating the transfer of communication genes across phylogenetically diverse organisms. Besides, considering that many bacteria become virulent through QS, the identification of “eavesdropping” bacteriophages inside the same community of nodes as pathogenic bacteria may be of peculiar interest, especially if the viral function inferred to be triggered by QS corresponds to lysis. The *in-silico* identification of “eavesdropping” bacteriophages could indeed be of particular interest for phagotherapy as they could potentially prevent some pathogenic bacteria from spreading and switching to a virulent phenotype.

At a more local level, some analytical tools from the graph theory may help identifying influent microbial entities inside a communication network. Indeed, centrality indicators are measures that allow identifying important nodes within a network, according to different criteria. For instance, the “degree” centrality accounts for the number of edges that connects a node to another (loops are counted twice) (Shaw 1954; Freeman 1978) and may thus help identifying microbial or viral

entity that tends to be more “social” than others, especially if their predicted “partners” of communication correspond to phylogenetically diverse entities. Again, the “betweenness” centrality accounts for the number of times that a node acts as a bridge along the shortest path between two other nodes (Freeman 1978; Newman 2005). Hence, the higher the betweenness of a node, the more it acts as a connector between distinct communities of nodes. A node with a high betweenness could therefore depict entities that could be adapted to plural communication contexts, or that could perhaps represent an intermediate state before the specialization in a given “discussion” group. Another example could arise from the “eigenvector” centrality, a measure of the influence that a node has in the network (Bonacich 1972). In a nutshell, a high eigenvector score means that a node is connected to many nodes who themselves have high scores. Specifically, the identification of a cluster of nodes with an associated average eigenvector centrality that is high could therefore highlight a potential “source” of externalization of QS components inside a microbial community. A final example could be given by the “eccentricity” centrality which attributes a high score to nodes that lie at the extremities of connected components (Bouttier et al. 2003). This indicator could precisely point at microbial entities that rely on one or several divergent quorum system(s) of genes that might thus be responsible for the production and/or the recognition of communication molecules that exhibit uncharacterized chemical structures.

Of course, all these examples are not exhaustive but illustrate nonetheless that a communication network is an analytical framework that is rich, flexible and suited for many analyses, both at the global and local levels.

8.5 On the Peculiar Interest of Inferring a Communication Network from Metaviromes and Metagenomes Sequenced in the Same Environmental Site

If building a network of communication based on reference genomes available on the public databases may open many interesting perspectives, it also exhibits limitations, inherent to the fact that most of the publicly available reference genomes have been sequenced independently from each other. Indeed, reference genomes are usually the results of the following protocol: isolation from their respective environments, cultivation in laboratory and sequencing. Hence, it is almost always impossible to access to the social context that these well-studied microorganisms experienced in their environment of extraction. It follows that the inferences made by our proposed method on the communication links between reference genomes would always suffer from the lack of proof that they are in contact with each other, in the very same environment. However, with the recent advent of metagenomics and metaviromics, which allow to reconstruct (more or less completely and confidently) genomes of unicellulars and viruses from environmental DNA sequences (metagenomics-assembled genomes) (Breitwieser et al. 2019), this barrier can

nowadays be somewhat overcome. Using metagenomes and metaviromes from the same environmental site as input to our algorithm would indeed ensure that connected nodes in the resulting communication network always represent entities that are part of a same community.

Moreover, environmental DNA sequences that cannot be mapped to reference genomes, could capture the genetic information of viruses or unicellulars for which there is no isolated representatives, and may thus allow to access to a new diversity of biological systems (Albertsen et al. 2013). With this respect, it would be interesting to decipher the contribution of these uncharacterized genomes to the nodes of the communication network, as it may highlight how important these understudied microorganisms may be for the social dynamics of an environmental community. Finally, the information on the abundance of each metagenome and metavirome in a given environment could be mapped to the nodes of the corresponding communication network and could allow to study whether there is a correlation or not between the abundance of a given entity in its environment and its detected ability to communicate.

8.6 Key Perspectives from the Comparison of Several Environmental Networks

If the generation of a microbial communication network from one environment is already informative, we believe that the comparison of networks from different environments could be even more interesting.

For instance, the comparison of the abundances of each entity across distinct environments could be used to detect entities that are often found together, and at equivalent ratio of abundances in distinct communities. Indeed, the set of all co-occurrence observations between entities can be resumed in the form of a co-occurrence network (Yang et al. 2019) that can be analyzed conjointly with a communication network. For instance, deciphering whether there are areas of overlap between co-occurrence and communication networks could highlight how important communication links are to promote and/or reinforce interactions between entities (Fig. 8.2). With this respect, two genomes that would be linked by both a co-occurrence and a communication edges could further reinforce the prognosis of a suspected link of symbiosis. In this latter case, a metagenomics dataset that reflects the composition of a community at different time steps could be of particular interest, as tracking the abundances of genomes that vary likewise along time sequences could allow to infer symbiotic relationships with a higher confidence (Probst et al. 2018)

On another note, the comparison of different environmental metagenomics datasets could also unravel whether some microbial entities that harbor the same taxonomic label (Luo et al. 2015) but that thrive in distinct environments may possess different QS systems. In this regard, it could teach us to which extent the presence of QS systems inside the genome of a microbial species is determined by the habitat, and to which extent it is rather determined by their evolutionary paths. If the

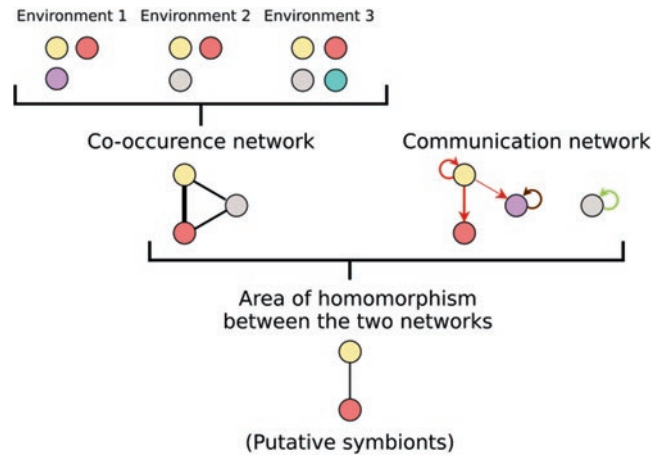


Fig. 8.2 Joint analysis of co-occurrence and communication networks to reinforce the prognosis of putative symbiotic relationships. Each node depicts a microbial entity, each directed edge a link of communication and each undirected edge a link of co-occurrence

habitat could determine the acquisition of QS systems, this would reinforce the hypothesis according to viruses and especially bacteriophages could play the role of “social” vectors within microbial communities, by transferring communication genes across genomes. With this respect, this putative role could perhaps be unraveled by a strong edge of communication between a virus and any genome that would harbor a QS system only in a given environment.

8.7 Concluding Remarks

In this chapter, we gave a brief overview of how quorum sensing genes mediate biocommunication in bacteriophages. We then argue that technical solutions are already available to mine genomes of viruses and other microorganisms for quorum sensing genes, upon which many entities rely to interact with others. We further proposed a methodology to use the results of this analysis to automatically construct networks of communication involving bacteriophages. Although such a network can be built from reference genomes sequenced independently from each other, we rather encourage inferring communication networks from metagenomes and metaviromes, sequenced altogether in the same environmental sites. Indeed, we speculate that the information that can be extracted from these networks could open many fascinating perspectives in the fields of biocommunication (*e.g.* discovery of new QS circuits in bacteriophages), ecology (*e.g.* discovery of new phage-host associations), evolution (*e.g.* examination of introgressive acquisitions of QS genes), medicine (*e.g.* discovery of bacteriophages that could antagonize pathogenic bacteria), and biochemistry (*e.g.* discovery of new communication molecules).

References

- Abedon ST (2015) Bacteriophage secondary infection. *Virology* 30(1):3–10. <https://doi.org/10.1007/s12250-014-3547-2>
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 31(6):533–538. <https://doi.org/10.1038/nbt.2579>
- Bhatt VS (2018) Quorum sensing mechanisms in gram positive bacteria, in implication of quorum sensing system in biofilm formation and virulence. Singapore, Springer Singapore, pp 297–311. https://doi.org/10.1007/978-981-13-2429-1_20
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp* 2008(10):P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. *J Math Sociol* 2(1):113–120. <https://doi.org/10.1080/0022250X.1972.9989806>
- Bouttier J, Di Francesco P, Guitter E (2003) Geodesic distance in planar graphs. *Nucl Phys B* 663(3):535–567. [https://doi.org/10.1016/S0550-3213\(03\)00355-9](https://doi.org/10.1016/S0550-3213(03)00355-9)
- Breitwieser FP, Lu J, Salzberg SL (2019) A review of methods and databases for metagenomic classification and assembly. *Brief Bioinform* 20(4):1125–1136. <https://doi.org/10.1093/bib/bbx120>
- Corel E, Méheust R, Watson AK, McInerney JO, Lopez P, Bapteste E (2018) Bipartite network analysis of gene sharings in the microbial world. *Mol Biol Evol* 35(4):899–913. <https://doi.org/10.1093/molbev/msy001>
- Cornforth DM, Popat R, McNally L, Gurney J, Scott-Phillips TC, Ivens A, Diggle SP, Brown SP (2014) Combinatorial quorum sensing allows bacteria to resolve their social and physical environment. *Proc Natl Acad Sci U S A* 111(11):4280–4284. <https://doi.org/10.1073/pnas.1319175111>
- Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, Melamed S et al (2017) Communication between viruses guides lysis-lysogeny decisions. *Nature* 541(7638):488–493. <https://doi.org/10.1038/nature21049>
- Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Networks* 1(3):215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* 176(2):269–275. <https://doi.org/10.1128/jb.176.2.269-275.1994>
- González JF, Venturi V (2013) A novel widespread interkingdom signaling circuit. *Trends Plant Sci* 18(3):167–174. <https://doi.org/10.1016/j.tplants.2012.09.007>
- Hargreaves KR, Kropinski AM, Clokie MRJ (2014) “What does the talking?: quorum sensing signalling genes discovered in a bacteriophage genome.” edited by Gunnar F. Kaufmann. *PLoS One* 9(1):e85131. <https://doi.org/10.1371/journal.pone.0085131>
- Hochstrasser R, Hilbi H (2017) Intra-species and inter-kingdom signaling of *Legionella Pneumophila*. *Front Microbiol* 8:79. <https://doi.org/10.3389/FMICB.2017.00079>
- Johnson MR, Montero CI, Connors SB, Shockley KR, Bridger SL, Kelly RM (2005) Population density-dependent regulation of exopolysaccharide formation in the hyperthermophilic bacterium *Thermotoga Maritima*. *Mol Microbiol* 55(3):664–674. <https://doi.org/10.1111/j.1365-2958.2004.04419.x>
- Jung SA, Chapman CA, Ng W-L (2015) Quadruple quorum-sensing inputs control *Vibrio cholerae* virulence and maintain system robustness. *PLoS Pathog* 11(4):e1004837. <https://doi.org/10.1371/journal.ppat.1004837>
- Karavolos MH, Winzer K, Williams P, Anjam Khan CM (2013) Pathogen espionage: multiple bacterial adrenergic sensors eavesdrop on host communication systems. *Mol Microbiol* 87(3):455–465. <https://doi.org/10.1111/mmi.12110>

- Lee W, Lee S-H, Kim M, Moon J-S, Kim G-W, Jung H-G, Kim IH et al (2018) Vibrio Vulnificus quorum-sensing molecule cyclo(Phe-Pro) inhibits RIG-I-mediated antiviral innate immunity. *Nat Commun* 9(1):1606. <https://doi.org/10.1038/s41467-018-04075-1>
- Luo C, Knight R, Siljander H, Knip M, Xavier RJ, Gevers D (2015) ConStrains identifies microbial strains in metagenomic datasets. *Nat Biotechnol* 33(10):1045–1052. <https://doi.org/10.1038/nbt.3319>
- Mehmood A, Liu G, Wang X, Meng G, Wang C, Liu Y (2019) Fungal quorum-sensing molecules and inhibitors with potential antifungal activity: a review. *Molecules (Basel, Switzerland)* 24(10). <https://doi.org/10.3390/molecules24101950>
- Mehta P, Goyal S, Long T, Bassler BL, Wingreen NS (2009) Information processing and signal integration in bacterial quorum sensing. *Mol Syst Biol* 5(1):325. <https://doi.org/10.1038/msb.2009.79>
- Miller MB, Bassler BL (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* 55(1):165–199. <https://doi.org/10.1146/annurev.micro.55.1.165>
- Mukherjee S, Bassler BL (2019) Bacterial quorum sensing in complex and dynamically changing environments. *Nat Rev Microbiol* 17(6):371–382. <https://doi.org/10.1038/s41579-019-0186-5>
- Nealson KH, Platt T, Hastings JW (1970) Cellular control of the synthesis and activity of the bacterial luminescent system. *J Bacteriol* 104(1):313–322
- Newman MEJ (2005) A measure of betweenness centrality based on random walks. *Soc Networks* 27(1):39–54. <https://doi.org/10.1016/J.SOCNET.2004.11.009>
- Oliver KM, Degnan PH, Hunter MS, Moran NA (2009) Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science (New York, NY)* 325(5943):992–994. <https://doi.org/10.1126/science.1174463>
- Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S (2005) Switches in bacteriophage lambda development. *Annu Rev Genet* 39(1):409–429. <https://doi.org/10.1146/annurev.genet.39.073003.113656>
- Paggi RA, Martone CB, Fuqua C, Castro RE (2003) Detection of quorum sensing signals in the haloalkaliphilic archaeon *Natronococcus Occultus*. *FEMS Microbiol Lett* 221(1):49–52. [https://doi.org/10.1016/S0378-1097\(03\)00174-5](https://doi.org/10.1016/S0378-1097(03)00174-5)
- Papenfort K, Bassler BL (2016) Quorum sensing signal-response systems in gram-negative bacteria. *Nat Rev Microbiol* 14(9):576–588. <https://doi.org/10.1038/nrmicro.2016.89>
- Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K et al (2018) Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol* 3(3):328–336. <https://doi.org/10.1038/s41564-017-0098-y>
- Rajput A, Kaur K, Kumar M (2016) SigMol: repertoire of quorum sensing signaling molecules in prokaryotes. *Nucleic Acids Res* 44(D1):D634–D639. <https://doi.org/10.1093/nar/gkv1076>
- Schuch R, Fischetti VA (2009) The secret life of the anthrax agent *Bacillus Anthracis*: bacteriophage-mediated ecological adaptations. *PLoS One* 4(8):e6532. <https://doi.org/10.1371/journal.pone.0006532>
- Shaw ME (1954) Group structure and the behavior of individuals in small groups. *J Psychol* 38(1):139–149. <https://doi.org/10.1080/00223980.1954.9712925>
- Silpe JE, Bassler BL (2019a) A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision. *Cell* 176(1–2):268–280.e13. <https://doi.org/10.1016/J.CELL.2018.10.059>
- Silpe JE, Bassler BL (2019b) Phage-encoded LuxR-type receptors responsive to host-produced bacterial quorum-sensing autoinducers. *BioRxiv March*:577726. <https://doi.org/10.1101/577726>
- Stokar-Avihail A, Tal N, Erez Z, Lopatina A, Sorek R (2019) Widespread utilization of peptide communication in phages infecting soil and pathogenic bacteria. *Cell Host Microbe* 25(5):746–755.e5. <https://doi.org/10.1016/j.chom.2019.03.017>
- Sun J, Daniel R, Wagner-Döbler I, Zeng A-P (2004) Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. *BMC Evol Biol* 4(1):36. <https://doi.org/10.1186/1471-2148-4-36>

- Sztajner H, Szafranski SP, Tomasch J, Reck M, Nimtz M, Rohde M, Wagner-Döbler I (2014) Cross-feeding and interkingdom communication in dual-species biofilms of *Streptococcus Mutans* and *Candida Albicans*. *ISME J* 8(11):2256–2271. <https://doi.org/10.1038/ismej.2014.73>
- Wahl LM, Betti MI, Dick DW, Pattenden T, Puccini AJ (2018) Evolutionary stability of the lysis-lysogeny decision: why be virulent? *Evolution* 73(1):evo.13648. <https://doi.org/10.1111/evo.13648>
- Waters CM, Bassler BL (2005) QUORUM SENSING: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* 21(1):319–346. <https://doi.org/10.1146/annurev.cellbio.21.012704.131001>
- Watve S, Barrasso K, Jung SA, Davis KJ, Hawver LA, Khataokar A, Palaganas RG, Neiditch MB, Perez LJ, Ng W-L (2019) Ethanolamine regulates CqsR quorum-sensing signaling in *Vibrio cholerae*. *BioRxiv* March:589390. <https://doi.org/10.1101/589390>
- Yang P, Yu S, Lin C, Ning K (2019) Meta-network: optimized species-species network analysis for microbial communities. *BMC Genomics* 20(S2):187. <https://doi.org/10.1186/s12864-019-5471-1>
- Zhang G, Zhang F, Ding G, Li J, Guo X, Zhu J, Zhou L et al (2012) Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon. *ISME J* 6(7):1336–1344. <https://doi.org/10.1038/ismej.2011.203>

7. Travaux en cours

7.1. Détection de variants divergents de gènes de QS par des recherches itératives d'homologues et critères de conservation de synténie

Les identifications de QSSs candidats par recherche d'homologie avec des QSSs déjà validées expérimentalement décrites précédemment (**sections 3 et 4**) avaient pour point commun d'être assez stringentes dans les seuils définis pour considérer d'éventuels hits. Or, des seuils plus permissifs pourraient permettre d'identifier des homologues plus divergents de protéine de QS, ce qui serait intéressant à la fois d'un point de vue évolutif comme appliqué. Par exemple, ces homologues divergents seraient sans doute plus susceptibles de produire/reconnaître des molécules de communication à la structure inédite. L'identification de telles molécules pourrait alors déboucher sur de possibles applications dans l'inhibition des QSSs de pathogènes résistants aux antibiotiques (voir **section 1.3**). En guise d'exemple, un travail en cours se propose d'identifier des synthèses d'AHLs divergentes de la famille LuxI, dans le but d'identifier par expression hétérologue chez *E. coli* des AHLs à la structure moléculaire inédite.

Les synthèses d'AHLs de la famille LuxI déjà caractérisées peuvent être classées en deux sous familles : celles qui utilisent préférentiellement l'acyl-CoA et celles qui utilisent préférentiellement l'acyl-ACP comme premier substrat (voir figure supplémentaire 1 de la **section 3**). L'utilisation du second substrat, la S-adenosyl-L-méthionine (SAM), est conservée pour toutes les protéines de la famille LuxI. Le domaine protéique d'incorporation du premier précurseur moléculaire (acyl-CoA ou acyl-ACP) contribue de façon majeure à la structure finale de l'AHL produite, si bien que les AHLs synthèses ayant un tropisme pour l'acylCoA produisent des molécules (e.g. Isovaleryl-AHL, pCoumaroyl-AHL, pCinnamoyl-AHL, **Figure 4**) aux structures vraiment différentes de celles produites par les AHLs utilisant préférentiellement l'acyl-ACP. Ainsi, si des synthèses d'AHLs divergentes présentent un domaine divergent ou alternatif d'incorporation du premier substrat, il est probable que ces synthèses concourent à la production de nouvelles molécules, même par expression hétérologue chez *E. coli*, pour peu que leurs substrats soient disponibles dans cet organisme modèle. Des programmes de prédiction de structure 3D tels qu'Alphafold pourraient notamment être mis à contribution pour identifier les synthèses divergentes dont la structure 3D est peu similaire à celles de synthèses déjà connues, et donc, les synthèses les plus susceptibles de produire de nouveaux signaux (Jumper *et al.*, 2021). Par ailleurs, les résultats de l'étude de (Wellington Miranda *et al.*, 2021) pourraient également être exploités pour définir des candidats intéressants. Dans cette étude, les acides aminés des séquences LuxI covariant avec les acides aminés de leur LuxR associé respectif

furent identifiés. Les auteurs ont ensuite caractérisés expérimentalement que dans la famille LuxI ces résidus sont déterminants dans la sélectivité des synthèses envers les AHLs produites. De même, dans la famille LuxR, ces résidus sont important dans la sélectivité des récepteurs envers les AHLs reconnues. Ainsi, des LuxI divergents avec des acides aminés inédits aux positions identifiées par Miranda et al. comme impliquées dans la sélectivité des AHLs produites pourraient représenter des candidats intéressants au titre de synthèses de molécules de communication encore inconnues.

Toutefois, la détection de protéines de QS divergentes s'accompagne de nombreux risques de faux positifs, et les synthèses d'AHLs n'y font pas exception, en particulier car cette famille est affiliée à une superfamille appelée GNAT-acyltransférase qui catalyse l'acétylation d'une très large diversité de substrats, dont la majorité ne sont pas impliqués dans la biosynthèse de molécules de QS. L'analyse suivante se propose d'appréhender ce problème en détectant des QSSs divergents de type *luxI-luxR*.

Une façon de minimiser les risques de faux positifs dans la détection de synthèses d'AHL divergentes de la famille LuxI est de tirer profit de leur tendance à la colocalisation génomique avec leur récepteur associé de la famille LuxR. Ainsi, la présence d'un gène *luxR* dans le voisinage génomique d'un gène codant une protéine LuxI divergente rassurera sur la qualité de synthase d'AHLs de cette dernière.

D'un point de vue purement méthodologique, la recherche de profils HMM offre un très bon compromis entre sensibilité et spécificité. Puisque les HMMs sont construits avec des protéines de référence dont on sait qu'elles sont homologues, leur recherche s'avère plus fiable pour détecter des homologues divergents que PSI-BLAST (Altschul *et al.*, 1997), un programme de recherches itératives d'homologues se proposant d'actualiser à chaque itération la matrice de position d'acides aminés utilisée en requête avec les hits significatifs identifiés à l'itération précédente. En effet l'inclusion de protéines non homologues à cette matrice, c'est à dire n'ayant pas évolué à partir d'un ancêtre commun aux protéines de départ, peut en effet biaiser la recherche d'homologues vers des faux positifs, résultant alors en un phénomène de dévoiement. L'idée de recherche itérative étant toutefois très intéressante, mon laboratoire d'accueil a récemment développé un outil appelé « ISHF » pour Iterative-Safe-Homologs-Finder, se proposant de lancer des BLASTs itératifs (en utilisant les homologues trouvés à l'itération précédente comme requêtes pour trouver des homologues distants dans l'itération en cours) en tentant de minimiser le phénomène de dérive décrit précédemment. À la différence de PSI-BLAST, qui met à jour une matrice de position sur l'intersection des régions alignées, et aboutit de ce fait à détecter des petits domaines conservés sur des séquences très divergentes, ISHF définit un segment aligné comme une partie de séquence pour laquelle il existe une chaîne d'alignements depuis une séquence de référence (requête initiale) telle que l'intersection mutuelle de toutes les régions alignées couvre une pourcentage minimale de la longueur de chaque séquence (e.g. 80%). À chaque fin d'itération, ISHF met à jour les segments alignés des hits identifiés et ne reporte que ceux qui respectent cette condition de chaînage. De la sorte, ISHF ne conserve que

les séquences qui s'alignent sur par exemple plus de 80% de leur longueur commune, ce qui aboutit à la détection d'une similarité faible mais globale, indice d'une homologie véritable.

La combinaison de ces méthodes, ISHF, recherche de profils HMMs, et contrôle de synténie pourrait alors permettre d'identifier des homologues divergents de synthases d'AHLS en minimisant les risques de faux positifs. Pour l'illustrer, l'analyse lancée au cours de cette thèse se propose d'utiliser la base de données de protéines prédites à partir de métagénomiques de la base de données MGnify comme jeu de données cible (Mitchell *et al.*, 2020). Cette base de données comprend 1,106,951,200 séquences protéiques partitionnées en 304,820,129 clusters à 90% d'identité, issues de 326,788 échantillons de prélèvement au sein 491 biomes différents (types d'environnements).

En premier lieu, ISHF fut mis à contribution selon le protocole suivant. Les 75 séquences représentatives de la diversité des protéines de la famille LuxI validées expérimentalement (clustering à 90% d'identité par CD-HIT (Huang *et al.*, 2010)) furent utilisées comme requêtes initiales pour une recherche itérative d'homologues par ISHF contre les séquences représentatives des 304,820,129 clusters de MGnify selon les paramètres suivants : DIAMOND comme aligneur, identité de séquences minimale $\geq 25\%$, couverture mutuelle minimale prédite $\geq 80\%$ avec toutes les séquences. 2111 séquences ont été trouvées en 7 itérations (run1: 1074, run2: 253, run3: 52, run4: 66, run 5: 551, run 6: 97, run 7: 7). Le lancement d'ISHF avec les 72 séquences représentatives des clusters à 90 % d'identité des LuxR validées expérimentalement a permis l'identification de 35,520 séquences en 7 itérations (run1: 2200, run2: 31,506, run3: 1952, run4: 2, run 5: 1, run 6: 1, run 7: 1).

En second lieu, le lancement d'hmmsearch avec les motifs HMMs spécifiques, d'une part aux synthases d'AHLS de la famille LuxI, et d'autre part au domaine N-terminal de liaison à l'AHLS de la famille LuxR (voir table supplémentaire 1 de l'article précédent) a abouti à l'identification de 2086 hits pour LuxI et 4391 pour LuxR. Ensemble, ces deux méthodes ont permis d'identifier 2774 homologues putatifs de LuxI et 37,351 de LuxR. Après cette recherche d'homologues divergents, les positions génomiques sur chacun des contigs des séquences codantes de ces hits (et de l'ensemble des protéines des clusters qu'ils représentent) furent extraites des métadonnées du MGnify. Cela a permis d'identifier que 864 clusters de LuxI putatifs comprennent au moins une protéine codée par un gène adjacent à un gène codant LuxR putatif, soit 31% des hits LuxI identifiés (**Figure 24**). 144 hits de LuxI ont été trouvés par ailleurs sur un même contig qu'un hit de LuxR sans y être adjacent.

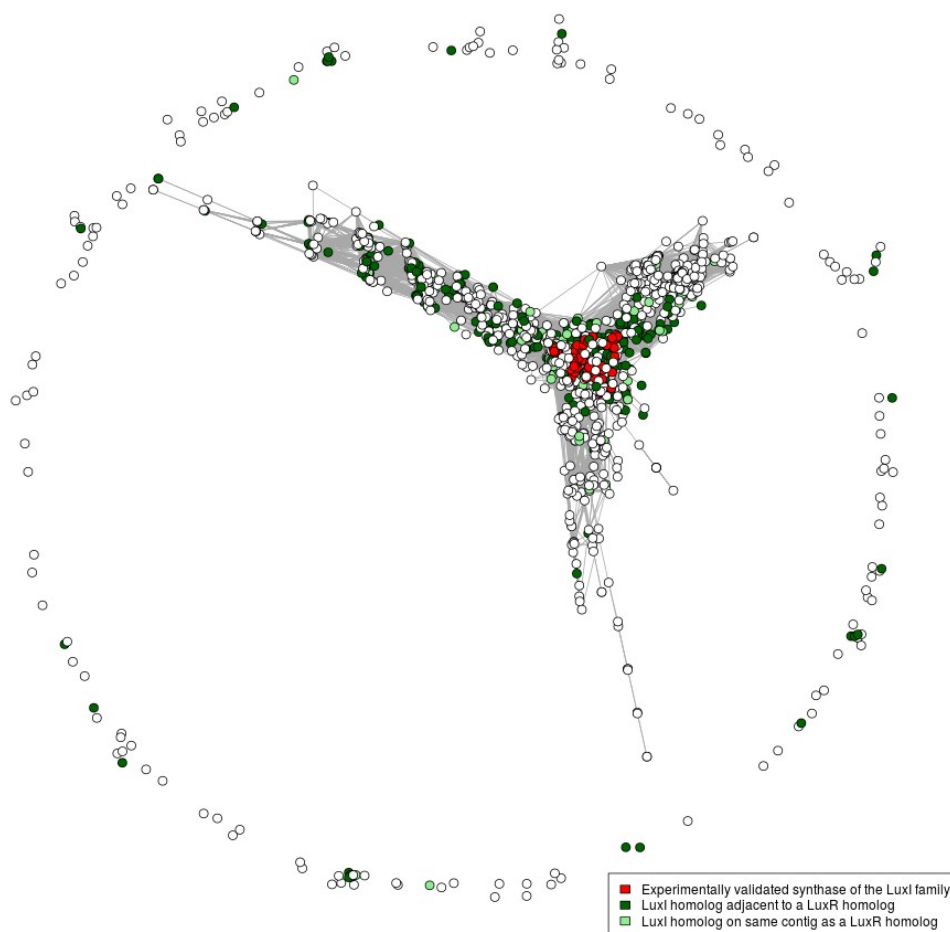


Figure 24 : Réseau de similarité de séquences de la famille LuxI. Les séquences protéiques de LuxI validées expérimentalement sont représentées en rouge, celles des hits adjacents à un *luxR* putatif dans MGnify en vert foncé, celles des hits présents sur un même contig que *luxR* en vert pâle et celles des autres hits en blanc. Une arête représente un pourcentage d'identité $\geq 25\%$ sur au moins 80% de la longueur des deux protéines.

Les composantes connexes isolées ont été identifiées par HMM et sont annotées par MGnify comme étant tronquées en N-terminal et/ou en C-terminal, ce qui explique pourquoi elles ne passent pas le seuil de couverture mutuelle nécessaire à leur incorporation à la composante connexe principale. Comme on peut le voir, des séquences très éloignées du noyau de séquences validées expérimentalement ont été identifiées, ce qui en fait de bons candidats pour la synthèse d'AHLs à la structure inédite. La prédiction de structure 3D de ces candidats avec Alphafold et leur alignement avec les structures 3D des protéines validées expérimentalement pourrait notamment permettre d'identifier ceux dont la structure laisse présager une susceptibilité importante de produire de nouvelles molécules de communication (Jumper *et al.*, 2021). Pour pouvoir tester expérimentalement des candidats représentatifs de la diversité de séquences du réseau ci-dessus, des communautés de nœuds (ayant tendance à être plus connectés entre eux qu'avec le reste du réseau) ont été identifiées avec l'algorithme de Leiden (Traag, Waltman and van Eck, 2019) (**Figure 25**). Les séquences

codantes des candidats les plus intéressantes situées à côté d'un *luxR* dans ces différentes communautés de séquences seront commandées et exprimées de façon hétérologue chez *E. coli*.

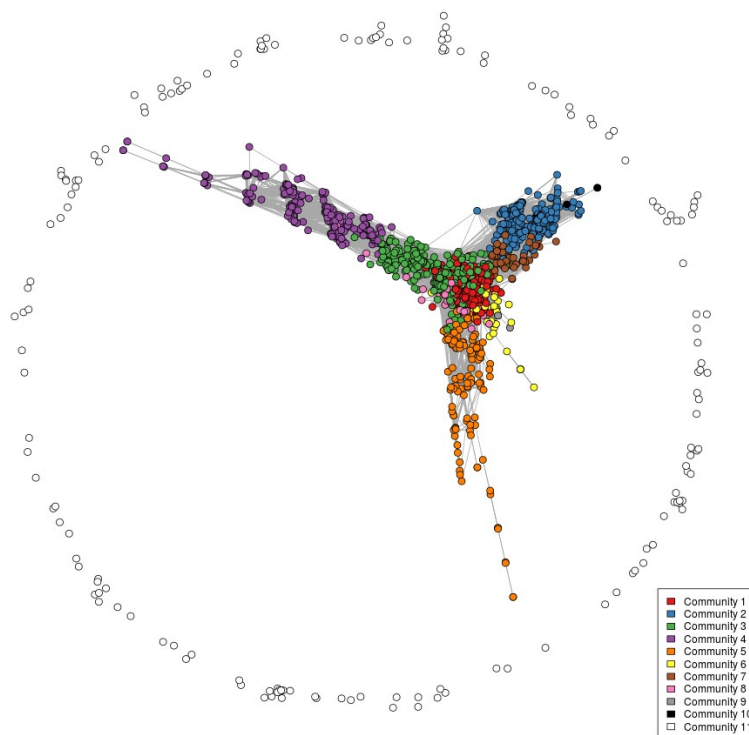


Figure 25 : Les différentes communautés de Leiden de LuxI putatifs dans le réseau de similarité de séquences

Bien sûr, cette analyse préliminaire est incomplète, notamment parce qu'elle est surtout centrée autour des synthases, et non des récepteurs. Ainsi, la sélection de candidats pourrait par exemple être affinée vers des LuxI dont le LuxR associé présente une structure primaire ou secondaire inhabituelle, ainsi que des résidus alternatifs dans la sélectivité des AHLs reconnus. Des analyses d'identité de séquences des protéines identifiées dans MGnify avec des protéines présentes dans la base de données publique NR pourrait aussi permettre de minimiser les faux positifs en éliminant les séquences moins proches des LuxI validées expérimentalement que des séquences présentes dans NR annotées pour des fonctions différentes de celle de QS. Enfin, l'assignation taxonomique des différents nœud du réseau de similarité de séquences (e.g. assignation taxonomique du contig codant le hit LuxI) pourrait également permettre d'ajouter une composante évolutive à cette analyse. Au regard des résultats prometteurs d'ores et déjà révélés par cette analyse préliminaire, des efforts de recherche seront consacrés dans cette direction après ma thèse.

7.2. Vers une inférence à large échelle des fonctions régulées par QSSs dont le récepteur est à une composante

Si l'inférence des fonctions biologiques régulées par des QSSs présentée dans la **section 5** pouvait être réalisée de façon systématique et à grande échelle parmi de nombreux génomes / MGEs, cela permettrait sans doute de révéler de nouvelles fonctions densité-dépendantes ainsi que de quantifier la prévalence relative des différentes réponses au QS (e.g. biofilm formation, competence, bioluminescence etc.) chez les bactéries et leurs MGEs lorsque ces réponses sont initiées par des récepteurs à une composante. Toutefois, la tendance à la colocalisation des gènes cibles et des QSSs ne représente pas une vérité absolue, d'autant plus si ces QSSs sont chromosomiques. En conséquence, des inférences erronées peuvent émerger de cette règle d'association. Néanmoins, la génomique comparative peut être mise à contribution pour tenter de limiter ces faux positifs. En effet, puisque les réarrangements génomiques résultant en une colocalisation des gènes participant à un même processus biologique semblent favorisés par la sélection naturelle, les contextes génomiques où les gènes régulés par un QSS lui sont adjacents devraient théoriquement être davantage conservés évolutivement que les contextes où les gènes adjacents à un QSS ne sont pas régulés par ce dernier. Selon cette logique, une famille de gènes représentée plusieurs fois en adjacence avec un QSS dans différentes espèces aura une probabilité plus importante d'être régulée par QS qu'une famille de gènes retrouvée seulement une fois en adjacence à un QSS.

Une tentative d'analyse fut initiée pour explorer cette intuition lors de cette thèse mais fut malheureusement abandonnée en cours de route, faute de temps. Premièrement, 2949 couples LuxI-LuxR adjacents furent identifiés par recherche de profils HMMs au sein des génomes complets représentatifs de la diversité des Protéobactéries du NCBI. Puis, l'opéron au sein duquel chaque *luxI* détecté appartient fut inféré comme une succession de gènes séparés par une distance intergénique inférieure à 200 nucléotides et localisé sur le même brin d'ADN (gènes do-directionnels). Les 22,985 séquences des protéines codées par les différents gènes de ces opérons putatifs furent groupées en 1590 familles dans un BLAST all vs all (Altschul *et al.*, 1990). Chaque famille fut définie comme une composante connexe dans un graphe de similarité de séquence où chaque arête entre deux nœuds/protéines correspond à une identité de séquence $\geq 25\%$ sur au moins 70 % de la longueur des deux protéines. Une matrice de présence/absence « famille par opéron » fut créée et donnée en entrée à la fonction `cooccur` du package R `cooccur` pour comptabiliser les co-occurrences des différentes familles au sein des différents opérons *luxI* et estimer celles qui sont statistiquement significatives (Griffith, Veech and Marsh, 2016). En fonction du nombre de gènes présents dans chacun des différents opérons et de la taille de chaque famille, `cooccur` assigne en effet un nombre de co-occurrences attendu entre deux familles par le modèle probabiliste de Veech (Veech, 2013) pour

chaque paire de famille. Enfin, cooccur assigne à partir des co-occurrences observées et attendues de de chaque couple de familles, une probabilité que ces deux familles soit positivement ou négativement associées. Ici, seuls les couples dont les familles avaient une probabilité d'être positivement associées supérieure à 0.05 furent conservés. Les résultats furent ensuite visualisés dans un graphe de co-occurrence où la taille d'un nœud représente la prévalence d'une famille dans les opérons *luxI* et les arêtes entre deux nœuds furent pondérées par le ratio « nombre de co-occurrences observées / nombre de co-occurrences attendues » entre deux familles. Les familles LuxI et LuxR qui étaient logiquement les plus prévalentes dans ces opérons ne furent pas représentées dans le graphe pour mieux visualiser les co-occurrences des autres nœuds. Des communautés de nœuds significativement associés au QS furent ensuite identifiées dans ce réseau par l'algorithme de Leiden pour identifier d'éventuels modules fonctionnels, annotés par un entier naturel dans la **figure 26** ci-après (Traag, Waltman and van Eck, 2019).

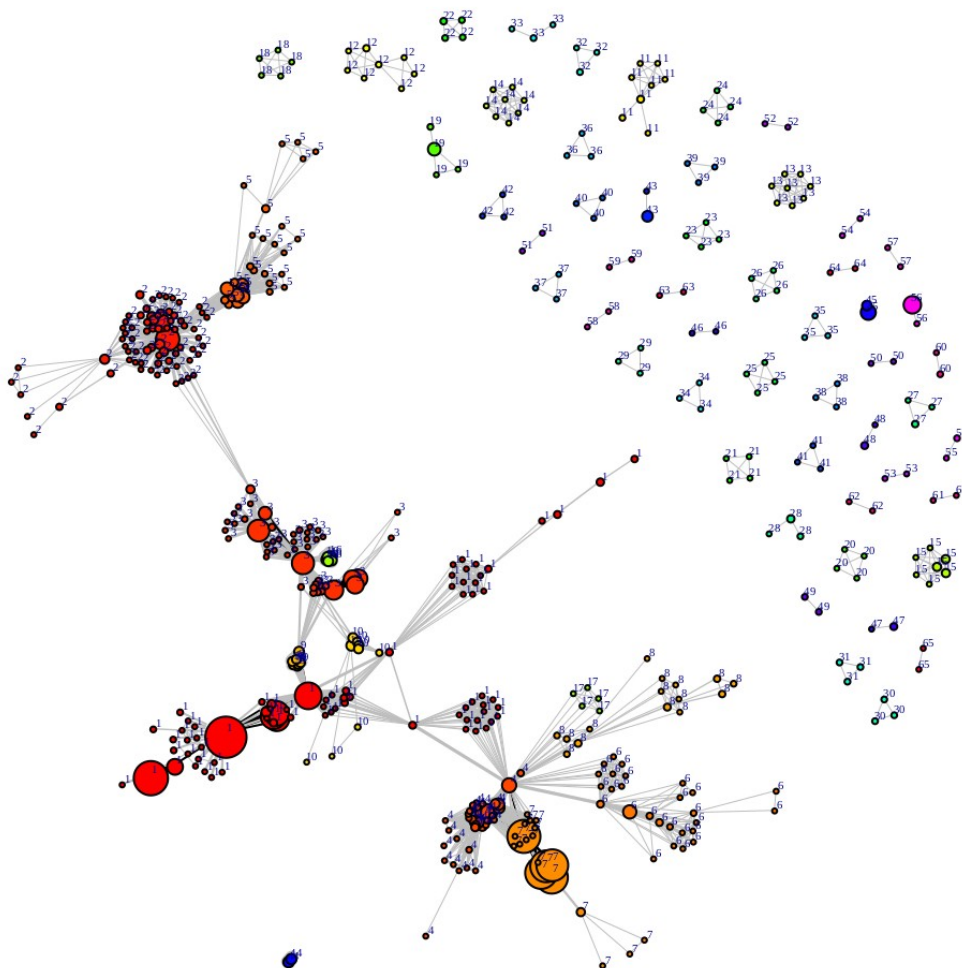


Figure 26 : Réseau de co-occurrences de différentes familles de gènes au sein d'opérons *luxI*, ainsi inférées comme étant régulées par QS

Malheureusement, cette analyse fut abandonnée à cette étape, et la cartographie de labels fonctionnels sur chaque nœud du réseau n'a pas encore été réalisée (e.g. KEGG), ce qui prive cette analyse de résultats interprétables sous le prisme de la biologie. Il serait intéressant d'identifier si parmi les modules identifiées ci-avant, certains pouvaient être associés à des fonctions encore non décrites dans le champ de recherche du QS. Il serait également intéressant de représenter chaque nœud du réseau comme un camembert dont les différentes portions correspondraient à différents niveaux taxonomiques (e.g. genre ou famille) pour avoir une idée de la spécificité taxonomique des différents modules. Enfin, on aperçoit bien sur le réseaux, que les modules sont composés de nœuds de tailles plus ou moins importantes et correspondent ainsi vraisemblablement à des réponses associées au QS de type *luxI-luxR* plus au moins prévalentes chez les Protéobactéries. Une analyse quantitative de ces prévalences relatives serait également intéressant pour donner un aperçu global des fonctions régulées par les QSSs de type *luxI-luxR* dans ce phylum bactérien.

8. Conclusions et perspectives

Le travail effectué au cours de cette thèse sur le QS a permis à la fois des avancées méthodologiques importantes dans la détection de systèmes génétiques de QS (QSSs) et des découvertes inédites dans le champ du QS tels que l'identification des premiers phages « bilingues » ou des premiers systèmes de QSSs de phages prédits pour manipuler la biologie de leur hôte bactérien. À l'heure de tirer un bilan de ce travail de thèse, je résumerai d'abord succinctement les points clés de mes développements méthodologiques, les opportunités qu'ils ont déverrouillées ainsi que leurs axes d'amélioration. Je résumerai ensuite les découvertes biologiques révélées par l'application de ces méthodes à des jeux de données génomiques de large échelle. Enfin, j'aborderai les limitations de mon travail de thèse et présenterai quelques perspectives que j'aurais souhaité explorer si le temps de thèse avait pu être extensible à souhait.

Développements méthodologiques

Détection de QSSs dont les composantes sont homologues à des gènes de QS déjà caractérisés

Une première contribution méthodologique de mon travail de thèse a été de réunir les ressources nécessaires à l'identification d'homologues de QSSs déjà connus au sein de génomes d'intérêt. Ces ressources sont le fruit d'un travail minutieux de recensement des séquences protéiques validées expérimentalement pour chaque famille de synthase, propeptide, récepteur membranaire ou récepteur intracellulaire d'un signal de QS connu et des HMMs de motifs clés spécifiques à ces familles. Les identifiants de ces protéines et de ces HMMs sont disponibles dans les informations supplémentaires de l'article présenté en **section 3**. Les fichiers fastas et HMMs sont disponibles sur demande en attendant leur mise en ligne prochaine sur un dépôt GitLab. Ces ressources déverrouillent l'identification d'une large diversité de QSSs par des méthodes basées sur l'homologie de séquence (e.g. BLASTP (Altschul *et al.*, 1990) et/ou HMMSearch (Eddy, 2011b)) (e.g. **sections 3 et 4.1**). Avec l'avènement de la métagénomique, ces ressources pourraient s'avérer précieuses dans le cadre d'investigations écologiques puisqu'elles permettent notamment l'identification de systèmes de communication au sein de communautés microbiennes environnementales. Pour en témoigner, plusieurs requêtes d'accès à ces fastas et ces HMMs de protéines de QS de référence nous ont déjà été formulées à des fins d'identification de gènes de QS au sein de bactéries isolées de microbiotes intestinaux humains, de microbiomes d'éponges de l'Antarctique ou encore de souches de *Streptomyces* étudiées pour leur riche réservoir de métabolites secondaires. Une limite à ces ressources est qu'elles ne sont pas encore intégrées dans un programme clé en main de détection d'homologues de QSSs de référence. Néanmoins, je prévois de palier partiellement ce problème en mettant à profit quelques unes de ces données dans un programme qui sera appelé QSS_detector et

qui regroupera RRNPP_detector avec des algorithmes de détection de systèmes synténiques *luxI-luxR* basés sur les AHLs (Protéobactéries) et de systèmes synténiques *afsA-arpA* basés sur les GBLs (Actinobactéries). QSS_detector permettra ainsi la détection automatisée de systèmes de communication spécifiques de 3 phyla bactériens différents.

Détection de nouveaux QSSs candidats, non homologues à des QSSs déjà caractérisés

La contribution méthodologique majeure de cette thèse correspond au développement de RRNPP_detector, un outil de détection de QSSs de type RRNPP non basé sur l'homologie de séquence et permettant ainsi l'identification de nouveaux gènes de QS (**section 4.2**). Pour rappel, les systèmes RRNPP sont spécifiques des chromosomes, plasmides et phages de Firmicutes et représentent à ce jour le seul type de QSSs identifiés comme complets et fonctionnels au sein de génomes de phages. Le développement de RRNPP_detector fut motivé par l'observation qu'en dépit d'une diversité génétique importante, les systèmes RRNPP caractérisés fonctionnent, à de rares exceptions près, selon un même mécanisme canonique, sous-tendant alors une signature commune regroupant plusieurs critères génomiques et structuraux. Ces critères sont au nombre de 5 : i) le propeptide est une petite protéine (10-70aa), ii) le propeptide arbore une séquence signal d'adressage au SEC translocon en N-terminal, iii) le récepteur est une protéine de taille moyenne (250-500aa), iv) le récepteur présente des répétitions de tétratricopeptides (TPRs) impliquées dans la liaison de petits peptides, v) les gènes codant le propeptide et le récepteur sont adjacents, à faible distance inter-génique. Puisqu'un outil nommé SignalP avait déjà été développé dans le but de prédire la présence d'une séquence d'adressage au SEC translocon au sein de protéines (Almagro Armenteros *et al.*, 2019) et puisqu'une large gamme d'HMMs de TPRs publiques était d'ores et déjà disponible pour identifier des protéines arborant ces motifs structuraux, j'ai pu déterminer que la signature spécifique aux systèmes RRNPP était détectable *in silico* et j'ai alors développé RRNPP_detector pour automatiser son identification. Puisque les systèmes RRNPP constituent une des 4 grandes classes de QSSs (**Figure 2**), cet outil constitue une avancée importante dans le champ de recherche du QS. Comme RRNPP_detector ne repose pas sur de l'homologie de séquence, ce programme permet notamment d'étendre la diversité de QSSs candidats de type RRNPP. Conséquence de cette expansion, RRNPP_detector représente un jalon important dans l'identification de nouveaux systèmes de QSSs codés par des phages, un champ de recherche émergent qui ne comprenait jusqu'alors qu'un seul objet d'étude avec le système arbitrium (Erez *et al.*, 2017; Stokar-Avihail *et al.*, 2019). Cette diversité nouvelle de QSSs révélée par RRNPP_detector déverrouille aussi l'identification de nouveaux processus biologiques régulés en fonction de la densité de population codant le QSS candidat, parmi lesquels pourraient notamment se trouver des mécanismes jusqu'alors inconnus d'activation de virulence chez certains pathogènes humains et qui pourraient alors être ciblés à des fins thérapeutiques. Enfin, puisque les systèmes RRNPP sont spécifiques aux Firmicutes et leurs

MGEs et puisque ce phylum est le plus prévalent au sein des microbiotes intestinaux humains, RRNPP_detector pourrait s'avérer utile dans le cadre de recherches sur la biocommunication et les processus régulés de façon densité dépendante au sein de microbiomes, plasmidomes et viromes humains. Un axe d'amélioration important pour cet outil est lié au fait que les séquences codantes de bon nombre de propeptides et de peptides leaderless (de moins d'une dizaine d'acides aminés) ne sont pas présents dans les annotations de génomes déposées sur les bases de données publiques et passent sous les radars de Prodigal, qui constitue la dépendance actuelle de RRNPP_detector pour la détection d'ORFs au sein de séquences nucléotidiques (Hyatt *et al.*, 2010). En effet, Prodigal n'est pas optimisé pour la détection de micropeptides, ni pour la détection de gènes au sein de génomes de phages, un des objets d'étude pourtant majeur de RRNPP_detector. Une conséquence importante de cette limite technique est que RRNPP_detector n'est pas encore armé pour gérer les exceptions des leaderless peptides et des propeptides présentant une séquence signal pour le système alternatif PptAB (e.g. Shp2 et PrgX) puisqu'ils sont beaucoup plus petits que les propeptides excrétés par le SEC translocon sur lesquels RRNPP_detector se concentre actuellement. Cependant, tout cela pourrait changer dans un futur très proche. Une version 2 de RRNPP_detector est en effet prévue et sera incluse dans le programme QSS_detector évoqué plus haut. Cette version 2 proposera alors Phanotate comme alternative à Prodigal dans le cas où l'utilisateur serait intéressé par des QSSs viraux (McNair *et al.*, 2019). Par ailleurs, cette version 2 bénéficiera de l'expérience gagnée dans un projet collaboratif en cours pour quantifier la prévalence des QSSs basés sur des peptides leaderless chez les bactéries. En effet, suite à la publication de RRNPP_detector, le professeur Muthiah Kumaraswami (Houston Methodist), le scientifique à l'origine de la découverte des peptides de QS « leaderless » m'a contacté pour mener cette analyse (Do *et al.*, 2017, 2019). L'automatisation de la détection de ces peptides de QS sur laquelle j'ai commencé à travailler (non présentée dans cette thèse) a d'ores et déjà abouti au design du pipeline suivant. Premièrement, les régions nucléotidiques flanquant la séquence codante d'un récepteur candidat présentant des TPRs non accouplés à un propeptide excrété par le SEC translocon seront extraites (-400 nucléotides en amont et 400 nucléotides en aval). Deuxièmement, l'outil MiPepid, entraîné à détecter des petites séquences transcrites (codantes ou non) au sein de séquences nucléotidiques, sera appliqué sur ces régions voisines (Zhu and Gribskov, 2019). Troisièmement, les 21 nucléotides en amont du codon start des séquences codantes de micropeptides préalablement identifiées seront soumis au crible d'une hiérarchie de 27 expressions régulières pour détecter un potentiel motif Shine-Dalgarno de RBS (Ribosome Binding Site) correspondant à un site d'initiation de la traduction (voir table 2 de (Hyatt *et al.*, 2010)). Le score de probabilité que le micropeptide soit bien fonctionnel (donné par MiPepid) et le score assigné au RBS détecté en amont du peptide seront ensuite moyennés ; un filtre sera alors appliqué pour ne garder que les micropeptides à forte probabilité de transcription et de traduction. Enfin, si la séquence codante la plus directement adjacente à celle du récepteur correspond à un tel

micropeptide, ce couple de gènes sera considéré comme formant un QSS candidat de type RRNPP et sera classé dans une catégorie spéciale regroupant les peptides sécrétés de façon non canonique. En gérant ces cas particuliers, la version 2 de RRNPP_detector permettra de lever encore davantage le voile dissimulant la diversité des QSSs de type RRNPP.

Inférence fonctionnelle de QSSs dont le récepteur est intracellulaire (récepteur à une composante)

La veille bibliographique effectuée au cours de cette thèse a permis d'identifier une propriété de certaines classes de QSSs pouvant être exploitée à des fins d'inférence fonctionnelle. Lorsque le récepteur est à une composante, les gènes régulés par le QSS ont tendance à se trouver dans le voisinage génomique du QSS (Engbrecht, Nealson and Silverman, 1983; Hoover *et al.*, 2015; Erez *et al.*, 2017; Brotherton, Medema and Greenberg, 2018; Stokar-Avihail *et al.*, 2019; Rued *et al.*, 2021). Au cours de cette thèse, j'ai exploité cette propriété à des fins d'identification de clusters biosynthétiques de gènes (BGCs) probablement sous contrôle de QS, car ces BGCs sont des cibles de choix dans le champ disciplinaire de la découverte de produits naturels (**section 4.2**). En effet, puisqu'un des défis majeurs de cette discipline est que beaucoup de BGCs ne sont pas exprimés dans les conditions de laboratoire, le lien entre expression et densité de population existant pour les BGCs régulés par QS représente un recours prometteur pour déclencher la biosynthèse du métabolite réalisée par chaque BGC. En outre, un autre projet de ma thèse s'est concentré sur l'analyse de la présence de gènes de régulation de sporulation au sein du voisinage génomique de certains QSSs (**section 5**). Dans ces deux cas, l'analyse fonctionnelle des voisinages de QSSs n'était pas systématique car seules certaines fonctions d'intérêt étaient ciblées. Une tentative d'inférer de façon plus systématique les fonctions régulées de façon densité dépendante par des QSSs de type *luxI-luxR* fut néanmoins initiée au cours de cette thèse (**section 7.2**). Un des problèmes à surmonter dans ce genre d'analyses à grande échelle est que le lien fonctionnel entre gènes régulés par un QSS et proximité génomique au QSS ne représente qu'une tendance et pas une vérité absolue, ce qui peut donner lieu à de nombreuses inférences fonctionnelles erronées. Pour tenter de limiter ces faux positifs, la génomique comparative fut mise à profit. En effet, puisque les réarrangements génomiques résultant en une colocalisation des gènes participant à un même processus biologique semblent favorisés par la sélection naturelle, j'ai fait l'hypothèse que les contextes génomiques où les gènes régulés par un QSS sont adjacents à ce dernier devaient théoriquement être davantage conservés évolutivement que les contextes où les gènes adjacents à un QSS ne sont pas régulés par ce dernier. Selon cette logique, j'ai assemblé un réseau de co-occurrence de familles de gènes (chacune définie comme une composante connexe d'un réseau de similarité de séquence) au sein d'opérons *luxI-luxR* identifiés dans les génomes de Protéobactéries. Dans ces réseaux de co-occurrence, la taille d'un nœud (famille de gènes) représentant la prévalence d'une famille au sein d'opérons *luxI-luxR* et le poids d'une arête entre deux nœuds représentant le nombre de co-occurrence de deux familles

distinctes au sein d'opérons *luxI-luxR*. Un tel réseau offre en effet l'avantage de pouvoir être filtré sur la taille des nœuds et le poids des arêtes pour ne garder que les familles de gènes à forte probabilité d'être régulées par QS. Faute de temps, ni la cartographie de labels fonctionnels (e.g. KEGG) sur chacun des nœuds du réseau, ni la vérification que certains nœuds correspondent bel et bien à des fonctions définies comme régulées par QS dans la littérature n'ont pu être entrepris. Cette tentative n'a donc pas pu aboutir sur de réels interprétations biologiques. Bien que ces idées n'aient pas été implémentées dans une méthode concrète d'inférence fonctionnelle de QS à large échelle, leur formulation dans cette thèse pourrait néanmoins constituer un premier pas vers de futurs développements en cette direction. Des outils de détection d'orthologues et d'inférence de synténie ancestrale par des méthodes de phylogénomique pourraient notamment permettre des prédictions plus rigoureuses. Ces approches pourraient alors révéler de nouveaux processus régulés par QS et permettre de quantifier la prévalence des différentes réponses au QS prédites chez les bactéries.

Inférence de liens de communication entre génomes et reconstruction de réseaux de communication

Un schéma de méthode d'inférence de réseaux de communication fut esquissé au cours de cette thèse (**section 6**). Dans ces réseaux, les nœuds seraient des génomes et les arêtes seraient des liens de communication orientés, pondérés et colorés par type de signal de QS. L'idée à la base de tels réseaux fut que si deux génomes distincts codent des composantes de QSSs très semblables, il est probable que ces deux génomes puissent émettre et/ou reconnaître au moins une molécule de QS en commun, justifiant alors la prédiction d'un lien potentiel de communication entre eux. La veille bibliographique effectuée au cours de cette thèse (**section 2.5**) a permis d'identifier une limite majeure à dépasser pour implémenter cette idée d'apparence triviale : les seuils d'identité de séquence et de couverture mutuelle (sur toute la longueur ou sur certaines régions clés) entre protéines de QS homologues sur lesquels se reposer pour définir ces liens potentiels de communication sont variables d'une famille de protéines à l'autre. Il est donc primordial d'obtenir des données liant une séquence à la gamme de molécules produites ou reconnues par cette séquence pour chaque famille de protéine de QS afin d'identifier les résidus déterminants dans la sélectivité des protéines de QS envers leurs signaux associés et de justifier des seuils d'identité entre régions de protéines de QS, ou entre structures 3D prédites par AlphaFold (Jumper *et al.*, 2021) sur des bases statistiques solides. Ces données pourraient par ailleurs permettre d'entraîner des modèles de machine learning pour calculer la probabilité de production ou de reconnaissance d'une molécule en commun pour deux synthases ou deux récepteurs d'une même famille protéique. Or, pour l'instant, ces données ne sont disponibles que pour très peu de familles de protéines de QS (LuxI, LuxR (Rajput, Kaur and Kumar, 2016), Rap et Phr (Even-Tov *et al.*, 2016)). Sans posséder ces données, une manière d'inférer les résidus importants pour la sélectivité des synthases et des récepteurs pour leur molécules de QS hypothétiquement produites/reconnues pourrait toutefois consister à effectuer

des analyses à grande échelle de covariation de résidus pour entre chaque composante d'un couple famille de synthases / famille de récepteurs, selon les intuitions révélées par l'étude de (Wellington Miranda *et al.*, 2021). Au regard de tous ces prérequis méthodologiques, un travail de longue haleine serait nécessaire à l'identification de règles propres à chaque couple de familles d'émetteur et de récepteur pour définir la probabilité d'un lien de communication entre deux génomes. Pour ces raisons, ma contribution à cet objectif de reconstruction hypothétique de réseaux de communication fut limitée au seul cadre théorique mais aura néanmoins eu le mérite d'identifier les limites à dépasser pour développer cette méthode d'inférence. Par ailleurs, mon travail de recensement des molécules produites et reconnues par les différents membres des familles LuxI et LuxR aura fourni des données quantitatives sur lesquelles baser les règles évoquées ci-dessus pour deux des familles de protéines de QS les plus étudiées dans le champ du QS (**Figure 18**). Bien que l'inférence de réseaux de communication est à ce jour difficile à appréhender, un objet intermédiaire pourrait néanmoins permettre quelques investigations intéressantes sur la distribution des gènes de QS dans les communautés microbiennes. Cet objet est un réseau de partage de gènes de QS tripartite (famille de récepteurs, génome/MAG et famille de synthases), tel que généré dans l'article de la **section 3**. En attribuant des labels d'assignation taxonomique et d'environnement de séquençage aux MAGs de ce réseau, certaines analyses intéressantes pourraient être réalisées, comme par exemple la quantification de la propension des MAGs d'une même espèce à coder des QSSs associées à des langues moléculaires (groupes de signaux à la structure moléculaire semblable) différentes en fonction de leur environnement d'origine.

Découvertes biologiques

Détection d'homologues de protéines de QS chez des lignées sous-explorées et des métagénomes environnementaux

Après la constitution de ma base de données de séquences et d'HMMs de protéines de QS, mon laboratoire d'accueil fut intéressé par la propagation de ces connaissances aux lignées procaryotes nouvellement découvertes que sont les CPRs et les DPANNs (**Section 3**). Puisque ces lignées sont très différentes des lignées connues auparavant, leur découverte a permis une expansion massive de la diversité décrite des organismes habitant actuellement sur Terre (Hug *et al.*, 2016). Néanmoins, peu de représentants de CPRs et de DPANNs sont cultivables en laboratoire et la biologie des membres de ces différentes lignées reste encore très peu connue à ce jour. Puisque la réduction de leur répertoire de gènes fut une force importante de l'évolution des génomes des CPRs et des DPANNs, les membres de ces lignées sont souvent décrits comme dépendants d'autres espèces pour survivre, ce qui pose la question de savoir comment ils interagissent avec d'autres espèces. Mon travail a alors révélé un riche répertoire d'homologues de protéines de QS codés par ces génomes qui pourrait être

impliqué dans leurs interactions avec d'autres espèces. La majorité de ces homologues étaient orphelins, c'est à dire non adjacents à leur composante de QS complémentaire (synthase ou récepteur). Ces observations suggèrent une possible utilisation de ces protéines à des fins de manipulation dans le cas d'homologues de synthases ou d'espionnage dans le cas d'homologues récepteurs. Il est par ailleurs fortement probable que des membres de ces lignées codent des QSSs qui leur sont propres et que mon analyse basée sur l'homologie de séquence avec des systèmes connus n'a pas pu capturer. Quoiqu'il en soit, au regard de l'absence de gènes clés impliqués dans la néosynthèse de nucléotides ou d'acides aminés dans ces lignées (Castelle *et al.*, 2018), la présence de gènes codant des homologues de protéines de QS dans certains génomes de CPRs et de DPANNs souligne l'importance capitale que pourrait revêtir certains traits sociaux dans la survie de certains micro-organismes. À des fins de vérification expérimentale, certains variants de synthases de QS identifiés au cours de cette analyse ont été commandés et optimisés pour l'utilisation de codons d'*Escherichia coli*. Ces variants seront exprimés de façon hétérologue chez *E. coli* pour caractériser s'ils réalisent effectivement la biosynthèse de molécules de QS. Ce travail est toujours en cours à ce jour dans mon autre laboratoire d'accueil dirigé par Yanyan Li. Dans cette même veine, mes études ont identifié de nombreux variants, y compris très divergents de synthases et de récepteurs des familles LuxI et LuxR ont été identifiés au cours de cette thèse dans des génomes environnementaux recensés dans la base de données MGnify de l'EBI (Mitchell *et al.*, 2020) (**section 7.1**). De façon très intéressante, bon nombre de ces variants ont été identifiés dans un contexte d'adjacence *luxI-luxR*, ce qui renforce la prédiction que ces variants forment des QSSs fonctionnels. L'expression hétérologue d'une sélection de ces QSSs extrêmement divergents est prévue à plus long terme et pourra peut être révéler des structures moléculaires de signaux de QS encore inédites. Une application pourrait émerger de ces travaux si parmi ces nouvelles molécules de QS, certaines pouvaient être identifiées comme capable d'antagoniser les récepteurs de QS de pathogènes humains.

Détection de nouveaux QSSs au sein de génomes, plasmides et phages de Firmicutes

L'application de RRNPP_detector aux génomes complets de Firmicutes, aux prophages détectés dans ces génomes, aux génomes de virus séquencés ainsi qu'à la Gut Phage Database (Camarillo-Guerrero *et al.*, 2021) aura permis l'identification d'une cinquantaine de nouvelles familles de QSSs candidats de type RRNPP, distribuées parmi une grande diversité de taxa (21 différentes familles taxonomiques de Firmicutes) et d'éléments génétiques (chromosomes bactériens, plasmides, prophages et génomes de phages). Ces QSSs candidats constituent une importante librairie dans laquelle les expérimentalistes pourront piocher afin de mener des études fonctionnelles sur des processus bactériens, plasmidiques et viraux contrôlés par QS encore inconnus. Ces études pourraient notamment révéler de nouvelles stratégies évolutives régulées de façon densité-dépendante chez ces différentes entités microbiennes. Il est par ailleurs probable que cette librairie

déverrouille l'accès à certaines applications, puisque certains QSSs candidats pourraient très bien réguler des mécanismes à combattre d'un point de vue thérapeutique (e.g. virulence) où à exploiter à des fins médicales ou agronomiques (e.g. production de composés antimicrobiens). À ce dernier égard, ma détection de 46 clusters de gènes biosynthétiques (BGCs) présents dans le voisinage génomique de certains de ces QSSs candidats offre un catalogue de choix pour faciliter la découverte de ces composés. En effet, la prédiction que ces BGCs sont exprimés à haute densité de population suggère des rôles écologiques importants pour les molécules qu'ils concourent à synthétiser, parmi lesquelles pourraient se trouver de nouveaux métabolites de défense d'intérêt médical (Hoover *et al.*, 2015; Rued *et al.*, 2021). Une des limitations de mon travail est qu'il ne fut pas accompagné de validations expérimentales. Concevant toutefois ces recherches comme une pierre dans un édifice construit à plusieurs, j'ai bon espoir que d'autres chercheurs s'intéressant aux peptides de communication puissent déployer leurs ressources pour faire certains de ces tests fonctionnels. Mon travail fut en effet avant tout inscrit dans une démarche exploratoire de partage afin de faciliter les découvertes sur la biocommunication des Firmicutes et de leurs MGEs par le concours du plus grand nombre.

Contributions à la recherche sur le quorum-sensing des bactériophages

Une contribution importante de mon travail fut la découverte de nouvelles familles de QSSs candidats codés par des phages, alors que seule la famille des systèmes arbitrium était connue avant ma thèse (Erez *et al.*, 2017; Stokar-Avihail *et al.*, 2019). Ce travail aura notamment permis d'identifier les premiers génomes de phages 'bilingues', c'est à dire codant deux QSSs distincts, de familles génétiques différentes. La validation expérimentale de tous les candidats de QSSs viraux présentés au cours de cette thèse pourrait notamment révéler de nouveaux processus viraux régulés de façon densité-dépendante. À cet égard, une découverte importante réalisée au cours de cette thèse concerne l'identification de multiples occurrences d'une probable signature génomique de modulation densité-dépendante de la sporulation bactérienne au sein de génomes de phages de Firmicutes. D'une part, j'ai identifié plus d'une centaine de systèmes Rap-Phr codés par des phages dont les homologues bactériens sont connus pour réguler la compétence et/ou la sporulation (**Sections 4.1 et 5**). D'autre part, j'ai identifié de nombreux contextes génomiques où un gène viral homologue d'un régulateur bactérien de sporulation était adjacent à un QSS candidat viral de type RRNPP, et donc probablement sous son contrôle (**Section 5**). Avant mon travail de thèse, la communication entre phages était seulement caractérisée comme régulant la transition lyse-lysogénie, soit des processus strictement viraux. Or, une implication des prédictions formulées au cours de cette thèse pourrait être que la communication entre phages permet également de guider la manipulation densité-dépendante de processus d'hôtes bactériens, au bénéfice évolutif du phage ou du collectif hôte-prophage durant le cycle lysogénique.

Perspectives

Mon travail de thèse aura révélé un nombre important de QSSs candidats au sein de métagénomiques environnementaux, de génomes complets de Firmicutes, de plasmides et de génomes de phages. Certaines fonctions biologiques remarquables ont même été inférées pour quelques uns de ces QSSs, notamment celle, jusqu'alors inconnue, d'une manipulation densité-dépendante des voies de sporulation de l'hôte bactérien par certains bactériophages tempérés. Une des perspectives majeure de cette thèse concerne bien évidemment la validation expérimentale de toutes ces prédictions. Des études fonctionnelles sur le rôle biologique des autres QSSs candidats révélés par mon travail pourraient par ailleurs aboutir à la découverte de nouveaux processus régulés par QS qui pourraient peut-être transformer notre vision des interactions, de l'adaptation et de l'évolution des micro-organismes et des virus.

Si ce travail de thèse s'est focalisé avant tout sur les QSSs de type RRNPP (formant l'une des quatre grandes classes de QSSs), les connaissances glanées sur le QS au cours de ces trois années de thèse pourraient également être mises à profit pour découvrir de nouveaux QSSs d'autres classes. Par exemple, la combinaison de critères de synténie avec des librairies d'HMMs de superfamilles de motifs - qui s'est avérée efficace dans l'identification de QSSs candidats de type RRNPP - pourrait aussi être exploitée à des fins de détection de QSSs inconnus basés sur des peptides modifiés dont le module récepteur est à deux composantes. En effet, les QSSs validés expérimentalement appartenant à cette classe sont souvent organisés en opéron et partagent des caractéristiques communes. Par exemple, ces opérons comprennent un récepteur membranaire adjacent à un régulateur de réponse, ainsi qu'un propeptide de petite taille et une ou plusieurs enzymes de modifications de peptide (Sturme *et al.*, 2002). L'identification de peptides de petite taille (propeptide) pourrait alors servir d'ancre pour constituer un catalogue d'opérons candidats, opérons dont les protéines codées seraient soumises au crible de modèles HMMs de récepteurs membranaires, de régulateur de réponse compagnon et d'enzymes modificatrices de peptides afin de définir d'éventuels QSSs candidats. Ces HMMs pourraient être identifiés grâce à l'application d'Interproscan sur chacune des protéines codées au sein des opérons de QS déjà validés expérimentalement (Jones *et al.*, 2014). À la lueur des analyses préliminaires effectuées sur ces systèmes, j'ai confiance dans le fait qu'un programme analogue à RRNPP_detector mais dédié aux systèmes de transduction de peptides de QS à deux composantes puisse être développé. Si le temps me le permet dans le futur, j'essayerais de capitaliser sur le code de RRNPP_detector pour écrire ce programme.

Mes efforts de recherche sur la détection de nouveaux QSSs furent orientés principalement sur les QSSs dont le signal associé est de nature peptidique. Il est en effet beaucoup plus difficile d'appréhender par la bioinformatique les QSSs associés à de petites molécules biosynthétisées par

des synthèses. Cette difficulté est principalement due au fait que la séquence protéique d'une synthase ne code pas directement la structure finale du signal, contrairement aux propeptides qui codent la séquence non modifiée du futur peptide de QS dans leur partie C-terminal. Il est par ailleurs difficile d'assigner avec certitude une fonction de QS à une enzyme, car un doute peut subsister sur sa possible implication dans le métabolisme de molécules non affiliées au QS. Néanmoins, j'ai eu vent dans ma dernière année de thèse d'une technique qui pourrait peut-être être exploitée à des fins de détection de nouveaux QSSs basés sur des petites molécules. Cette technique s'appelle le profilage phylogénétique et se propose d'inférer des ensembles de gènes concourant au même processus biologique en traquant les différents groupes de gènes orthologues qui ont été perdus ou gagnés conjointement dans différentes lignées au cours de l'évolution (Pellegrini *et al.*, 1999; Škunca and Dessimoz, 2015). L'avantage de cette technique est qu'elle permet de s'affranchir de la synténie pour inférer de possible liens fonctionnels entre familles d'orthologues. Des enzymes de synthèse et des récepteurs aux histoires évolutives similaires pourraient alors peut-être constituer des candidats au titre de QSSs. Cependant, une limitation à cette méthode est que la phylogénie des procaryotes est assez mal résolue. De plus, les événements introgressifs tels que le transfert latéral de gènes contribuent significativement à l'évolution des génomes procaryotes et peuvent brouiller le signal potentiel de gains/pertes partagés. Une façon de valider la pertinence de ces techniques dans l'identification de nouveaux QSSs seraient de l'appliquer à la problématique des LuxR orphelins, c'est à dire non adjacents à une synthase de QS. Comme évoqué précédemment, un récepteur de type LuxR a deux domaines : un de détection d'un signal moléculaire et l'autre de liaison à l'ADN. Or, une variabilité importante peut exister dans le domaine de détection d'un signal moléculaire chez les LuxR orphelins. Ainsi, l'arbre phylogénétique de la superfamille des LuxR orphelins révèle différents clades d'orthologues, chacun d'entre eux pouvant potentiellement être dédié à la reconnaissance d'un type de signal moléculaire (Tobias *et al.*, 2019). Si ce signal peut être exogène, il peut aussi être endogène dans le cas où la synthase de QS n'est pas colocalisée avec le gène *luxR*. Par exemple, les membres du clade des LuxR orphelins chez le genre *Photobacterium* (*e.g. pluR* et *pauR*) fonctionnent en couple avec une kétosynthase non adjacente à *luxR* dans les génomes de ces bactéries (*e.g. ppyS* et *darB*) (Brameyer and Heermann, 2017). Si le profilage phylogénétique pouvait de façon non supervisée identifier ces kétosynthases de QS comme des candidates au titre de partenaires fonctionnels des LuxR de ce clade, cela fournirait une preuve de concept pour légitimer l'utilisation de cette méthode d'inférence pour tenter d'identifier pour chaque clade de LuxR orphelins, des synthèses de QS candidates, révélant par la même de nouveaux QSSs candidats.

Sans être aussi ambitieux avec la détection de nouvelles familles de QSSs, la détection d'homologues de QSSs basés sur des petites molécules au sein de plasmides et de prophages de Protéobactéries et d'Actinobactéries n'a pas été abordée au cours de cette thèse malgré des résultats qui pourraient peut-être ouvrir de nouvelles perspectives. Jusqu'à maintenant, seuls des récepteurs

de petites molécules de QS ont été trouvés dans des génomes de bactériophages, mais jamais des QSSs complets (Silpe and Bassler, 2019a, 2019b). L'identification du premier QSS basé non pas sur un peptide, mais sur une petite molécule de QS au sein de bactériophages aurait sans doute un retentissement certain dans la communauté de chercheurs s'intéressant au QS. Comme cette thèse m'a permis de développer des scripts pour la détection de QSSs peptidiques codés par des éléments génétiques mobiles, j'envisage de réutiliser ces scripts pour appréhender à moindre effort cette nouvelle perspective de recherche, ou faute de temps, de libérer ces scripts pour que d'autres chercheurs intéressés par cette thématique puissent les appliquer.

Outre la détection de QSS, ma thèse aura aussi été marquée par des efforts d'une part dans l'inférence des fonctions régulées par les QSSs dont le récepteur est à une composante et d'autre part dans l'inférence de réseaux de communication. Avec des méthodes rigoureuses et robustes pour effectuer ces deux types d'inférence, il serait alors possible d'inférer des réseaux de communication environnementaux et de cartographier certaines fonctions prédites pour être régulées par ces liens de communication sur les différents nœuds de ces réseaux. De telles annotations permettraient alors d'appréhender les dynamiques sociales et fonctionnelles au sein de communautés microbiennes à partir de seules séquences nucléotidiques de MAGs. Malgré de nombreuses difficultés à surmonter pour implémenter ces méthodes, les enseignements que pourraient révéler leur application à grande échelle, notamment au sein de microbiotes humains, pourraient s'avérer importants aussi bien d'un point de vue écologique qu'appliqué, par exemple vis à vis de collectifs de taxa co-occurents spécifiquement chez les patients atteints de certaines pathologies.

9. Références

- Adiliaghdam, F. *et al.* (2019) 'Targeting bacterial quorum sensing shows promise in improving intestinal barrier function following burn-site infection', *Molecular Medicine Reports*. Spandidos Publications, 49(5), pp. 4057–4066. doi: 10.3892/mmr.2019.10071.
- Almagro Armenteros, J. J. *et al.* (2019) 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nature Biotechnology*, 37(4), pp. 420–423. doi: 10.1038/s41587-019-0036-z.
- Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.
- Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: A new generation of protein database search programs', *Nucleic Acids Research*. Oxford University Press, pp. 3389–3402. doi: 10.1093/nar/25.17.3389.
- Anderson, J. K. *et al.* (2015) 'Chemorepulsion from the quorum signal autoinducer-2 promotes *Helicobacter pylori* biofilm dispersal', *mBio*. American Society for Microbiology, 6(4). doi: 10.1128/mBio.00379-15.
- Bassler, B. L., Greenberg, E. P. and Stevens, A. M. (1997) *Cross-Species Induction of Luminescence in the Quorum-Sensing Bacterium Vibrio harveyi*, *JOURNAL OF BACTERIOLOGY*.
- Bedree, J. K. *et al.* (2018) 'Quorum Sensing Modulates the Epibiotic-Parasitic Relationship Between *Actinomyces odontolyticus* and Its Saccharibacteria epibiont, a *Nanosynbacter lyticus* Strain, TM7x.', *Frontiers in microbiology*. Frontiers Media SA, 9, p. 2049. doi: 10.3389/fmicb.2018.02049.
- Bertrand, J. C. *et al.* (2015) 'Biogeochemical cycles', in *Environmental Microbiology: Fundamentals and Applications*. Springer Netherlands, pp. 511–617. doi: 10.1007/978-94-017-9118-2_14.
- Bhatt, V. S. (2019) 'Quorum sensing mechanisms in gram positive bacteria', in *Implication of Quorum Sensing System in Biofilm Formation and Virulence*. Springer Singapore, pp. 297–311. doi: 10.1007/978-981-13-2429-1_20.
- Bischofs, I. B. *et al.* (2009) 'Complexity in bacterial cell-cell communication: quorum signal integration and subpopulation signaling in the *Bacillus subtilis* phosphorelay.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(16), pp. 6459–64. doi: 10.1073/pnas.0810878106.
- Boettcher, K. J. and Ruby, E. G. (1990) 'Depressed light emission by symbiotic *Vibrio fischeri* of the sepiolid squid *Euprymna scolopes*', *Journal of Bacteriology*. American Society for Microbiology (ASM), 172(7), pp. 3701–3706. doi: 10.1128/jb.172.7.3701-3706.1990.
- Brameyer, S. and Heermann, R. (2017) 'Quorum sensing and LuxR solos in *Photobacterium*', in *Current Topics in Microbiology and Immunology*. Springer Verlag, pp. 103–119. doi: 10.1007/82_2016_28.
- Bridges, A. A. and Bassler, B. L. (2019) 'The intragenus and interspecies quorum-sensing autoinducers exert distinct control over *Vibrio cholerae* biofilm formation and dispersal', *PLOS Biology*. Edited by V. Sourjik. Public Library of Science, 17(11), p. e3000429. doi: 10.1371/journal.pbio.3000429.

- Brotherton, C. A., Medema, M. H. and Greenberg, E. P. (2018) 'luxR Homolog-Linked Biosynthetic Gene Clusters in Proteobacteria', *mSystems*. American Society for Microbiology, 3(3). doi: 10.1128/msystems.00208-17.
- Buchfink, B., Xie, C. and Huson, D. H. (2015) 'Fast and sensitive protein alignment using DIAMOND', *Nature Methods*. Nature Publishing Group, 12(1), pp. 59–60. doi: 10.1038/nmeth.3176.
- Camarillo-Guerrero, L. F. *et al.* (2021) 'Massive expansion of human gut bacteriophage diversity', *Cell*. Elsevier, 184(4), pp. 1098-1109.e9. doi: 10.1016/j.cell.2021.01.029.
- Capella-Gutierrez, S., Silla-Martinez, J. M. and Gabaldon, T. (2009) 'trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses', *Bioinformatics*, 25(15), pp. 1972–1973. doi: 10.1093/bioinformatics/btp348.
- Castelle, C. J. *et al.* (2018) 'Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations', *Nature Reviews Microbiology*, 16(10), pp. 629–645. doi: 10.1038/s41579-018-0076-2.
- Castelle, C. J. and Banfield, J. F. (2018) 'Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life.', *Cell*. Elsevier, 172(6), pp. 1181–1197. doi: 10.1016/j.cell.2018.02.016.
- Chu, E. K., Groisman, A. and Levchenko, A. (2019) 'Environmental sensing in dynamic quorum responses', *bioRxiv*. Cold Spring Harbor Laboratory, p. 745091. doi: 10.1101/745091.
- Churchill, M. E. A. and Chen, L. (2011) 'Structural basis of acyl-homoserine lactone-dependent signaling', *Chemical Reviews*. NIH Public Access, 111(1), pp. 68–85. doi: 10.1021/cr1000817.
- Clewell, D. B. and Weaver, K. E. (1989) 'Sex pheromones and plasmid transfer in *Enterococcus faecalis*', *Plasmid*. Plasmid, pp. 175–184. doi: 10.1016/0147-619X(89)90041-3.
- Co, A. D. *et al.* (2020) 'Short-range interactions govern the dynamics and functions of microbial communities', *Nature Ecology and Evolution*. Nature Research, 4(3), pp. 366–375. doi: 10.1038/s41559-019-1080-2.
- Cornforth, D. M. *et al.* (2014) 'Combinatorial quorum sensing allows bacteria to resolve their social and physical environment.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(11), pp. 4280–4. doi: 10.1073/pnas.1319175111.
- Coutinho, B. G. *et al.* (2018) 'A plant-responsive bacterial-signaling system senses an ethanolamine derivative', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(39), pp. 9785–9790. doi: 10.1073/pnas.1809611115.
- Davis, R. M., Muller, R. Y. and Haynes, K. A. (2015) 'Can the natural diversity of quorum-sensing advance synthetic biology?', *Frontiers in Bioengineering and Biotechnology*. Frontiers Media S.A., p. 30. doi: 10.3389/fbioe.2015.00030.
- Diggle, S. P., Gardner, A., *et al.* (2007) 'Evolutionary theory of bacterial quorum sensing: When is a signal not a signal?', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society, pp. 1241–1249. doi: 10.1098/rstb.2007.2049.

- Diggle, S. P., Matthijs, S., *et al.* (2007) 'The *Pseudomonas aeruginosa* 4-Quinolone Signal Molecules HHQ and PQS Play Multifunctional Roles in Quorum Sensing and Iron Entrapment', *Chemistry and Biology*. Cell Press, 14(1), pp. 87–96. doi: 10.1016/j.chembiol.2006.11.014.
- Do, H. *et al.* (2017) 'Leaderless secreted peptide signaling molecule alters global gene expression and increases virulence of a human bacterial pathogen', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(40), pp. E8498–E8507. doi: 10.1073/pnas.1705972114.
- Do, H. *et al.* (2019) 'Environmental pH and peptide signaling control virulence of *Streptococcus pyogenes* via a quorum-sensing pathway', *Nature Communications*. Nature Publishing Group, 10(1), pp. 1–14. doi: 10.1038/s41467-019-10556-8.
- Dombrowski, N. *et al.* (2019) 'Genomic diversity, lifestyles and evolutionary origins of DPANN archaea', *FEMS Microbiology Letters*. Narnia, 366(2). doi: 10.1093/femsle/fnz008.
- Donato, V. *et al.* (2017) 'Bacillus subtilis biofilm extends *Caenorhabditis elegans* longevity through downregulation of the insulin-like signalling pathway', *Nature Communications*. Nature Publishing Group, 8(1), pp. 1–15. doi: 10.1038/ncomms14332.
- Dufour, D. and Lévesque, C. M. (2013) 'Cell death of *Streptococcus mutans* induced by a quorum-sensing peptide occurs via a conserved streptococcal autolysin', *Journal of Bacteriology*. J Bacteriol, 195(1), pp. 105–114. doi: 10.1128/JB.00926-12.
- Eberhard, A. (1972) 'Inhibition and activation of bacterial luciferase synthesis.', *Journal of Bacteriology*. American Society for Microbiology (ASM), 109(3), pp. 1101–1105. doi: 10.1128/jb.109.3.1101-1105.1972.
- Eberhard, A. *et al.* (1981) 'Structural Identification of Autoinducer of *Photobacterium fischeri* Luciferase', *Biochemistry*. Biochemistry, 20(9), pp. 2444–2449. doi: 10.1021/bi00512a013.
- Eddy, S. R. (2011a) 'Accelerated Profile HMM Searches', *PLoS Computational Biology*. Edited by W. R. Pearson. Public Library of Science, 7(10), p. e1002195. doi: 10.1371/journal.pcbi.1002195.
- Eddy, S. R. (2011b) 'Accelerated Profile HMM Searches', *PLoS Computational Biology*. Edited by W. R. Pearson. Public Library of Science, 7(10), p. e1002195. doi: 10.1371/journal.pcbi.1002195.
- Edgar, R. C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput.', *Nucleic acids research*. Oxford University Press, 32(5), pp. 1792–7. doi: 10.1093/nar/gkh340.
- Engebrecht, J., Nealson, K. and Silverman, M. (1983) 'Bacterial bioluminescence: Isolation and genetic analysis of functions from *Vibrio fischeri*', *Cell*. Cell, 32(3), pp. 773–781. doi: 10.1016/0092-8674(83)90063-6.
- Erez, Z. *et al.* (2017) 'Communication between viruses guides lysis-lysogeny decisions.', *Nature*. Europe PMC Funders, 541(7638), pp. 488–493. doi: 10.1038/nature21049.
- Even-Tov, E. *et al.* (2016) 'Transient Duplication-Dependent Divergence and Horizontal Transfer Underlie the Evolutionary Dynamics of Bacterial Cell–Cell Signaling', *PLOS Biology*. Edited by J. Gore. Public Library of Science, 14(12), p. e2000330. doi: 10.1371/journal.pbio.2000330.

- Federle, M. J. and Bassler, B. L. (2003) 'Interspecies communication in bacteria', *Journal of Clinical Investigation*. The American Society for Clinical Investigation, pp. 1291–1299. doi: 10.1172/JCI20195.
- Fuqua, W. C., Winans, S. C. and Greenberg, E. P. (1994) 'Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.', *Journal of bacteriology*. American Society for Microbiology (ASM), 176(2), pp. 269–75. doi: 10.1128/jb.176.2.269-275.1994.
- van Gestel, J. *et al.* (2021) 'Short-range quorum sensing controls horizontal gene transfer at micron scale in bacterial communities', *Nature Communications*. Nature Research, 12(1), pp. 1–11. doi: 10.1038/s41467-021-22649-4.
- Gilbert, J. A. *et al.* (2018) 'Current understanding of the human microbiome', *Nature Medicine*. Nature Publishing Group, 24(4), pp. 392–400. doi: 10.1038/nm.4517.
- Gondry, M. *et al.* (2018) 'A Comprehensive Overview of the Cyclodipeptide Synthase Family Enriched with the Characterization of 32 New Enzymes', *Frontiers in Microbiology*. Frontiers Media S.A., 9(FEB), p. 46. doi: 10.3389/fmicb.2018.00046.
- González-Pastor, J. E. (2011) 'Cannibalism: A social behavior in sporulating *Bacillus subtilis*', *FEMS Microbiology Reviews*, 35(3), pp. 415–424. doi: 10.1111/j.1574-6976.2010.00253.x.
- González, J. F. and Venturi, V. (2013) 'A novel widespread interkingdom signaling circuit', *Trends in Plant Science*, 18(3), pp. 167–174. doi: 10.1016/j.tplants.2012.09.007.
- González, O. *et al.* (2017) 'Non-ribosomal Peptide Synthases from *Pseudomonas aeruginosa* Play a Role in Cyclodipeptide Biosynthesis, Quorum-Sensing Regulation, and Root Development in a Plant Host', *Microbial Ecology*. Springer New York LLC, 73(3), pp. 616–629. doi: 10.1007/s00248-016-0896-4.
- Grandclément, C. *et al.* (2015) 'Quorum quenching: Role in nature and applied developments', *FEMS Microbiology Reviews*. Oxford University Press, pp. 86–116. doi: 10.1093/femsre/fuv038.
- Griffith, D. M., Veech, J. A. and Marsh, C. J. (2016) 'Cooccur: Probabilistic species co-occurrence analysis in R', *Journal of Statistical Software*. American Statistical Association, 69(1), pp. 1–17. doi: 10.18637/jss.v069.c02.
- Haft, D. H. *et al.* (2013) 'TIGRFAMs and genome properties in 2013', *Nucleic Acids Research*. Nucleic Acids Res, 41(D1). doi: 10.1093/nar/gks1234.
- Hastings, J. W. (1971) 'Light to hide by: Ventral luminescence to camouflage the silhouette', *Science*. Science, 173(4001), pp. 1016–1017. doi: 10.1126/science.173.4001.1016.
- Hawver, L. A., Jung, S. A. and Ng, W. L. (2016) 'Specificity and complexity in bacterial quorum-sensing systemsa', *FEMS Microbiology Reviews*. Oxford University Press, pp. 738–752. doi: 10.1093/femsre/fuw014.
- Heilmann, S., Krishna, S. and Kerr, B. (2015) 'Why do bacteria regulate public goods by quorum sensing?-How the shapes of cost and benefit functions determine the form of optimal regulation', *Frontiers in Microbiology*. Frontiers Media S.A., 6(AUG), p. 767. doi: 10.3389/fmicb.2015.00767.

- Hochstrasser, R. and Hilbi, H. (2017) 'Intra-Species and Inter-Kingdom Signaling of Legionella pneumophila', *Frontiers in Microbiology*. Frontiers Media SA, 8, p. 79. doi: 10.3389/FMICB.2017.00079.
- Hoover, S. E. *et al.* (2015) 'A new quorum-sensing system (TprA/PhrA) for Streptococcus pneumoniae D39 that regulates a lantibiotic biosynthesis gene cluster', *Molecular Microbiology*. Blackwell Publishing Ltd, 97(2), pp. 229–243. doi: 10.1111/mmi.13029.
- Horinouchi, S. and Beppu, T. (1992) 'Autoregulatory Factors and Communication in Actinomycetes', *Annual Review of Microbiology*. Annual Reviews, 46(1), pp. 377–392. doi: 10.1146/annurev.mi.46.100192.002113.
- Hornby, J. M. *et al.* (2001) 'Quorum Sensing in the Dimorphic Fungus *Candida albicans* Is Mediated by Farnesol', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 67(7), pp. 2982–2992. doi: 10.1128/AEM.67.7.2982-2992.2001.
- Horswill, A. R. *et al.* (2007) 'The effect of the chemical, biological, and physical environment on quorum sensing in structured microbial communities', *Analytical and Bioanalytical Chemistry*. Springer, 387(2), pp. 371–380. doi: 10.1007/s00216-006-0720-y.
- Hsiao, A. *et al.* (2014) 'Members of the human gut microbiota involved in recovery from *Vibrio cholerae* infection', *Nature*. Nature Publishing Group, 515(7527), pp. 423–426. doi: 10.1038/nature13738.
- Huang, Y. *et al.* (2010) 'CD-HIT Suite: a web server for clustering and comparing biological sequences', *Bioinformatics*, 26(5), pp. 680–682. doi: 10.1093/bioinformatics/btq003.
- Hubbard, T. D., Murray, I. A. and Perdew, G. H. (2015) 'Special section on drug metabolism and the microbiome - Minireview indole and tryptophan metabolism: Endogenous and dietary routes to ah receptor activation', *Drug Metabolism and Disposition*. American Society for Pharmacology and Experimental Therapy, pp. 1522–1535. doi: 10.1124/dmd.115.064246.
- Hudaiberdiev, S. *et al.* (2015) 'Census of solo LuxR genes in prokaryotic genomes', *Frontiers in Cellular and Infection Microbiology*, 5, p. 20. doi: 10.3389/fcimb.2015.00020.
- Hug, L. A. *et al.* (2016) 'A new view of the tree of life', *Nature Microbiology*. Nature Publishing Group, 1(5), p. 16048. doi: 10.1038/nmicrobiol.2016.48.
- Hyatt, D. *et al.* (2010) 'Prodigal: Prokaryotic gene recognition and translation initiation site identification', *BMC Bioinformatics*. BMC Bioinformatics, 11. doi: 10.1186/1471-2105-11-119.
- Ismail, A. S., Valastyan, J. S. and Bassler, B. L. (2016) 'A Host-Produced Autoinducer-2 Mimic Activates Bacterial Quorum Sensing', *Cell Host and Microbe*. Cell Press, 19(4), pp. 470–480. doi: 10.1016/j.chom.2016.02.020.
- Jang, Y. J. *et al.* (2013) 'Autoinducer 2 of *Fusobacterium nucleatum* as a target molecule to inhibit biofilm formation of periodontopathogens', *Archives of Oral Biology*, 58(1), pp. 17–27. doi: 10.1016/j.archoralbio.2012.04.016.
- Jiang, Q. *et al.* (2019) 'Quorum Sensing: A Prospective Therapeutic Target for Bacterial Diseases', *BioMed Research International*. Hindawi Limited. doi: 10.1155/2019/2015978.

- Jimenez, A. G. and Sperandio, V. (2019) 'Quorum Sensing and the Gut Microbiome', in *Quorum Sensing: Molecular Mechanism and Biotechnological Application*. Elsevier, pp. 151–169. doi: 10.1016/B978-0-12-814905-8.00006-X.
- Jones, P. *et al.* (2014) 'InterProScan 5: Genome-scale protein function classification', *Bioinformatics*. Oxford University Press, 30(9), pp. 1236–1240. doi: 10.1093/bioinformatics/btu031.
- Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*. Nature Research, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.
- Kalamara, M. *et al.* (2018) 'Social behaviours by *Bacillus subtilis*: quorum sensing, kin discrimination and beyond', *Molecular Microbiology*. Wiley-Blackwell, 110(6), p. 863. doi: 10.1111/MMI.14127.
- Kaufmann, G. F. *et al.* (2005) 'Revisiting quorum sensing: Discovery of additional chemical and biological functions for 3-oxo-N-acylhomoserine lactones', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 102(2), pp. 309–314. doi: 10.1073/pnas.0408639102.
- Kempner, E. S. and Hanson, F. E. (1968) 'Aspects of light production by *Photobacterium fischeri*.', *Journal of bacteriology*. American Society for Microbiology (ASM), 95(3), pp. 975–979. doi: 10.1128/jb.95.3.975-979.1968.
- Kim, C. S. *et al.* (2020) 'Characterization of Autoinducer-3 Structure and Biosynthesis in *E. coli* ', *ACS Central Science*. American Chemical Society (ACS). doi: 10.1021/acscentsci.9b01076.
- Kleerebezem, M. (2004) 'Quorum sensing control of lantibiotic production; nisin and subtilin autoregulate their own biosynthesis', *Peptides*, pp. 1405–1414. doi: 10.1016/j.peptides.2003.10.021.
- Kumar, S. and Engelberg-Kulka, H. (2014) 'Quorum sensing peptides mediating interspecies bacterial cell death as a novel class of antimicrobial agents', *Current Opinion in Microbiology*. Elsevier Ltd, pp. 22–27. doi: 10.1016/j.mib.2014.09.001.
- Kuraku, S. *et al.* (2013) 'aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity', *Nucleic Acids Research*. Narnia, 41(W1), pp. W22–W28. doi: 10.1093/nar/gkt389.
- Landman, C. *et al.* (2018) 'Inter-kingdom effect on epithelial cells of the N-Acyl homoserine lactone 3-oxo-C12:2, a major quorum-sensing molecule from gut microbiota', *PLoS ONE*. Public Library of Science, 13(8). doi: 10.1371/journal.pone.0202587.
- Lee, J. H. and Lee, J. (2010) 'Indole as an intercellular signal in microbial communities', *FEMS Microbiology Reviews*. Blackwell Publishing Ltd, pp. 426–444. doi: 10.1111/j.1574-6976.2009.00204.x.
- Lee, W. *et al.* (2018) 'Vibrio vulnificus quorum-sensing molecule cyclo(Phe-Pro) inhibits RIG-I-mediated antiviral innate immunity', *Nature Communications*. Nature Publishing Group, 9(1), p. 1606. doi: 10.1038/s41467-018-04075-1.
- Letunic, I., Khedkar, S. and Bork, P. (2021) 'SMART: Recent updates, new developments and status in 2020', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D458–D460. doi: 10.1093/nar/gkaa937.

- Li, Q. *et al.* (2019) 'Autoinducer-2 of gut microbiota, a potential novel marker for human colorectal cancer, is associated with the activation of TNFSF9 signaling in macrophages', *Oncolmmunology*. Taylor and Francis Inc., 8(10). doi: 10.1080/2162402X.2019.1626192.
- Maire, T. and Youk, H. (2015) 'Molecular-Level Tuning of Cellular Autonomy Controls the Collective Behaviors of Cell Populations', *Cell Systems*. Cell Press, 1(5), pp. 349–360. doi: 10.1016/j.cels.2015.10.012.
- Martha Antunes, L. C. *et al.* (2005) 'Bacteroides species produce Vibrio harveyi autoinducer 2-related molecules', *Anaerobe*. Anaerobe, 11(5), pp. 295–301. doi: 10.1016/j.anaerobe.2005.03.003.
- Mcnair, K. *et al.* (2019) 'PHANOTATE: A novel approach to gene identification in phage genomes', *Bioinformatics*. Oxford University Press, 35(22), pp. 4537–4542. doi: 10.1093/bioinformatics/btz265.
- Mehta, P. *et al.* (2009) 'Information processing and signal integration in bacterial quorum sensing', *Molecular Systems Biology*, 5(1), p. 325. doi: 10.1038/msb.2009.79.
- Miller, M. B. and Bassler, B. L. (2001) 'Quorum Sensing in Bacteria', *Annual Review of Microbiology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 55(1), pp. 165–199. doi: 10.1146/annurev.micro.55.1.165.
- Millikan, D. S. and Ruby, E. G. (2004) 'Vibrio fischeri flagellin A is essential for normal motility and for symbiotic competence during initial squid light organ colonization', *Journal of Bacteriology*. American Society for Microbiology (ASM), 186(13), pp. 4315–4325. doi: 10.1128/JB.186.13.4315-4325.2004.
- Mistry, J. *et al.* (2021) 'Pfam: The protein families database in 2021', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D412–D419. doi: 10.1093/nar/gkaa913.
- Mitchell, A. L. *et al.* (2020) 'MGnify: The microbiome analysis resource in 2020', *Nucleic Acids Research*. Oxford University Press, 48(D1), pp. D570–D578. doi: 10.1093/nar/gkz1035.
- Moreno-Gómez, S. *et al.* (2017) 'Quorum sensing integrates environmental cues, cell density and cell history to control bacterial competence', *Nature Communications*. Nature Publishing Group, 8(1). doi: 10.1038/s41467-017-00903-y.
- NCBI Resource Coordinators (2016) 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Research*, 44(D1), pp. D7–D19. doi: 10.1093/nar/gkv1290.
- Nealson, K. H. (1977) 'Autoinduction of bacterial luciferase - Occurrence, mechanism and significance', *Archives of Microbiology*. Springer-Verlag, 112(1), pp. 73–79. doi: 10.1007/BF00446657.
- Nealson, K. H., Platt, T. and Hastings, J. W. (1970) 'Cellular control of the synthesis and activity of the bacterial luminescent system.', *Journal of bacteriology*. American Society for Microbiology (ASM), 104(1), pp. 313–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/5473898> (Accessed: 25 September 2019).
- Neiditch, M. B. *et al.* (2017) 'Genetic and Structural Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria', *Annual Review of Genetics*. Annual Reviews Inc., pp. 311–333. doi: 10.1146/annurev-genet-120116-023507.
- Oh, K. B. *et al.* (2001) 'Purification and characterization of an autoregulatory substance capable of regulating the morphological transition in *Candida albicans*', *Proceedings of the National Academy of*

Sciences of the United States of America. National Academy of Sciences, 98(8), pp. 4664–4668. doi: 10.1073/pnas.071404698.

Papenfort, K. *et al.* (2017a) 'A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation', *Nature Chemical Biology*. Nature Publishing Group, 13(5), pp. 551–557. doi: 10.1038/nchembio.2336.

Papenfort, K. *et al.* (2017b) 'A *Vibrio cholerae* autoinducer-receptor pair that controls biofilm formation', *Nature Chemical Biology*. Nature Publishing Group, 13(5), pp. 551–557. doi: 10.1038/nchembio.2336.

Papenfort, K. and Bassler, B. L. (2016) 'Quorum sensing signal-response systems in Gram-negative bacteria.', *Nature reviews. Microbiology*. NIH Public Access, 14(9), pp. 576–88. doi: 10.1038/nrmicro.2016.89.

Passador, L. *et al.* (1993) 'Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication', *Science*. Science, 260(5111), pp. 1127–1130. doi: 10.1126/science.8493556.

Pellegrini, M. *et al.* (1999) 'Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 96(8), pp. 4285–4288. doi: 10.1073/pnas.96.8.4285.

Pereira, C. S. *et al.* (2009) 'Identification of functional LsrB-like autoinducer-2 receptors', *Journal of Bacteriology*. J Bacteriol, 191(22), pp. 6975–6987. doi: 10.1128/JB.00976-09.

Pietschke, C. *et al.* (2017) 'Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in bacterial symbionts', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(40), pp. E8488–E8497. doi: 10.1073/pnas.1706879114.

Pirhonen, M. *et al.* (1993) 'A small diffusible signal molecule is responsible for the global control of virulence and exoenzyme production in the plant pathogen *Erwinia carotovora*', *EMBO Journal*. European Molecular Biology Organization, 12(6), pp. 2467–2476. doi: 10.1002/j.1460-2075.1993.tb05901.x.

Polkade, A. V. *et al.* (2016) 'Quorum sensing: An under-explored phenomenon in the phylum Actinobacteria', *Frontiers in Microbiology*. Frontiers Media S.A. doi: 10.3389/fmicb.2016.00131.

Pollak, S. *et al.* (2016) 'Facultative cheating supports the coexistence of diverse quorum-sensing alleles', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(8), pp. 2152–2157. doi: 10.1073/pnas.1520615113.

Pottathil, M. and Lazazzera, B. A. (2003) 'The extracellular PHR peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*', *Frontiers in Bioscience*, 8(4), p. 913. doi: 10.2741/913.

Rajput, A., Kaur, K. and Kumar, M. (2016) 'SigMol: repertoire of quorum sensing signaling molecules in prokaryotes', *Nucleic Acids Research*. Narnia, 44(D1), pp. D634–D639. doi: 10.1093/nar/gkv1076.

Rao, R. M., Pasha, S. N. and Sowdhamini, R. (2016) 'Genome-wide survey and phylogeny of S-Ribosylhomocysteinase (LuxS) enzyme in bacterial genomes', *BMC Genomics*. BioMed Central Ltd., 17(1), p. 742. doi: 10.1186/s12864-016-3002-x.

- Redfield, R. J. (2002) 'Is quorum sensing a side effect of diffusion sensing?', *Trends in Microbiology*. Trends Microbiol, pp. 365–370. doi: 10.1016/S0966-842X(02)02400-9.
- Reyes, A. *et al.* (2010) 'Viruses in the faecal microbiota of monozygotic twins and their mothers', *Nature*. NIH Public Access, 466(7304), pp. 334–338. doi: 10.1038/nature09199.
- Rezzonico, F., Smits, T. H. M. and Duffy, B. (2012) 'Detection of AI-2 Receptors in Genomes of Enterobacteriaceae Suggests a Role of Type-2 Quorum Sensing in Closed Ecosystems', *Sensors*. Molecular Diversity Preservation International, 12(5), pp. 6645–6665. doi: 10.3390/s120506645.
- Rudner, D. Z. *et al.* (1991) 'The spo0K locus of *Bacillus subtilis* is homologous to the oligopeptide permease locus and is required for sporulation and competence', *Journal of Bacteriology*. J Bacteriol, 173(4), pp. 1388–1398. doi: 10.1128/jb.173.4.1388-1398.1991.
- Rued, B. E. *et al.* (2021) 'Quorum sensing in *Streptococcus mutans* regulates production of tryglysin, a novel ras-ripp antimicrobial compound', *mBio*. American Society for Microbiology, 12(2), pp. 1–25. doi: 10.1128/mBio.02688-20.
- Schaefer, A. L. *et al.* (1996) 'Quorum sensing in *Vibrio fischeri*: Probing autoinducer-LuxR interactions with autoinducer analogs', *Journal of Bacteriology*. American Society for Microbiology, 178(10), pp. 2897–2901. doi: 10.1128/jb.178.10.2897-2901.1996.
- Sender, R., Fuchs, S. and Milo, R. (2016) 'Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans', *Cell*. Cell Press, pp. 337–340. doi: 10.1016/j.cell.2016.01.013.
- Shi, Y. M. and Bode, H. B. (2017) 'Microbiology: A new language for small talk', *Nature Chemical Biology*. Nature Publishing Group, pp. 453–454. doi: 10.1038/nchembio.2362.
- Shimada, Y. *et al.* (2013) 'Commensal bacteria-dependent indole production enhances epithelial barrier function in the colon', *PLoS ONE*. PLoS One, 8(11). doi: 10.1371/journal.pone.0080604.
- Sillitoe, I. *et al.* (2021) 'CATH: Increased structural coverage of functional space', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D266–D273. doi: 10.1093/nar/gkaa1079.
- Silpe, J. E. and Bassler, B. L. (2019a) 'A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision', *Cell*, 176(1–2), pp. 268-280.e13. doi: 10.1016/j.cell.2018.10.059.
- Silpe, J. E. and Bassler, B. L. (2019b) 'Phage-encoded LuxR-type receptors responsive to host-produced bacterial quorum-sensing autoinducers', *mBio*. American Society for Microbiology, 10(2). doi: 10.1128/mBio.00638-19.
- Škunca, N. and Dessimoz, C. (2015) 'Phylogenetic Profiling: How Much Input Data Is Enough?', *PLOS ONE*. Edited by H. Escriva. Public Library of Science, 10(2), p. e0114701. doi: 10.1371/journal.pone.0114701.
- Stokar-Avihail, A. *et al.* (2019) 'Widespread Utilization of Peptide Communication in Phages Infecting Soil and Pathogenic Bacteria', *Cell Host & Microbe*, 25(5), pp. 746-755.e5. doi: 10.1016/j.chom.2019.03.017.
- Sturme, M. H. J. *et al.* (2002) 'Cell to cell communication by autoinducing peptides in gram-positive bacteria', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 81(1–4), pp. 233–243. doi: 10.1023/A:1020522919555.

- Sturme, M. H. J. (2005) 'Analysis of quorum sensing regulatory systems in the human isolate *Lactobacillus plantarum*'. Available at: <https://agris.fao.org/agris-search/search.do?recordID=NL2012024093> (Accessed: 16 October 2021).
- Sun, J. *et al.* (2004) 'Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways', *BMC Evolutionary Biology*. BioMed Central, 4(1), p. 36. doi: 10.1186/1471-2148-4-36.
- Surette, M. G., Miller, M. B. and Bassler, B. L. (1999) 'Quorum sensing in *Escherichia coli*, *Salmonella typhimurium*, and *Vibrio harveyi*: A new family of genes responsible for autoinducer production', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 96(4), pp. 1639–1644. doi: 10.1073/pnas.96.4.1639.
- Tang, M. *et al.* (2020) 'Modulation of quorum sensing as an adaptation to nodule cell infection during experimental evolution of legume symbionts', *mBio*. American Society for Microbiology, 11(1). doi: 10.1128/mBio.03129-19.
- Thompson, J. A. *et al.* (2015) 'Manipulation of the quorum sensing signal AI-2 affects the antibiotic-treated gut microbiota', *Cell Reports*. Elsevier, 10(11), pp. 1861–1871. doi: 10.1016/j.celrep.2015.02.049.
- Tian, X. *et al.* (2018) '*Cryptococcus neoformans* sexual reproduction is controlled by a quorum sensing peptide', *Nature Microbiology*. Nature Publishing Group, 3(6), pp. 698–707. doi: 10.1038/s41564-018-0160-4.
- Tobias, N. J. *et al.* (2019) 'New Vocabulary for Bacterial Communication', *ChemBioChem*. Wiley-VCH Verlag, p. cbic.201900580. doi: 10.1002/cbic.201900580.
- Torcato, I. M. *et al.* (2019) 'Identification of novel autoinducer-2 receptors in *Clostridia* reveals plasticity in the binding site of the LsrB receptor family', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 294(12), pp. 4450–4463. doi: 10.1074/jbc.RA118.006938.
- Traag, V. A., Waltman, L. and van Eck, N. J. (2019) 'From Louvain to Leiden: guaranteeing well-connected communities', *Scientific Reports*. Nature Publishing Group, 9(1), pp. 1–12. doi: 10.1038/s41598-019-41695-z.
- Veech, J. A. (2013) 'A probabilistic model for analysing species co-occurrence', *Global Ecology and Biogeography*. Edited by P. Peres-Neto. John Wiley & Sons, Ltd, 22(2), pp. 252–260. doi: 10.1111/j.1466-8238.2012.00789.x.
- Vendeville, A. *et al.* (2005) 'Making “sense” of metabolism: Autoinducer-2, LuxS and pathogenic bacteria', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 383–396. doi: 10.1038/nrmicro1146.
- Venkatesh, M. *et al.* (2014) 'Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and toll-like receptor 4', *Immunity*. Cell Press, 41(2), pp. 296–310. doi: 10.1016/j.immuni.2014.06.014.
- Verbeke, F. *et al.* (2017) 'Peptides as quorum sensing molecules: Measurement techniques and obtained levels in vitro and in vivo', *Frontiers in Neuroscience*. Frontiers Research Foundation, p. 183. doi: 10.3389/fnins.2017.00183.

- Visick, K. L. *et al.* (2000) 'Vibrio fischeri lux genes play an important role in colonization and development of the host light organ', *Journal of Bacteriology*. American Society for Microbiology (ASM), 182(16), pp. 4578–4586. doi: 10.1128/JB.182.16.4578-4586.2000.
- Voichek, M. *et al.* (2020) 'Peptide-based quorum sensing systems in Paenibacillus polymyxa', *Life Science Alliance*. Rockefeller University Press, 3(10). doi: 10.26508/LSA.202000847.
- Waters, C. M. and Bassler, B. L. (2005) 'QUORUM SENSING: Cell-to-Cell Communication in Bacteria', *Annual Review of Cell and Developmental Biology*, 21(1), pp. 319–346. doi: 10.1146/annurev.cellbio.21.012704.131001.
- Wellington Miranda, S. *et al.* (2021) 'A covariation analysis reveals elements of selectivity in quorum sensing systems', *eLife*. eLife Sciences Publications Ltd, 10. doi: 10.7554/eLife.69169.
- Wellington, S. and Peter Greenberg, E. (2019) 'Quorum sensing signal selectivity and the potential for interspecies cross talk', *mBio*. American Society for Microbiology, 10(2). doi: 10.1128/MBIO.00146-19.
- Werner, K. M. *et al.* (2014) 'Caenorhabditis elegans recognizes a bacterial quorum-sensing signal molecule through theawcon neuron', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 289(38), pp. 26566–26573. doi: 10.1074/jbc.M114.573832.
- West, S. A. *et al.* (2012) 'Quorum sensing and the confusion about diffusion', *Trends in Microbiology*. Elsevier Current Trends, pp. 586–594. doi: 10.1016/j.tim.2012.09.004.
- Wilson, D. *et al.* (2009) 'SUPERFAMILY - Sophisticated comparative genomics, data mining, visualization and phylogeny', *Nucleic Acids Research*. Oxford Academic, 37(SUPPL. 1), pp. D380–D386. doi: 10.1093/nar/gkn762.
- Von Wintersdorff, C. J. H. *et al.* (2016) 'Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer', *Frontiers in Microbiology*. Frontiers Media S.A., p. 173. doi: 10.3389/fmicb.2016.00173.
- Witzany, G. (2011) 'Introduction: Key Levels of Biocommunication of Bacteria', in: Springer, Berlin, Heidelberg, pp. 1–34. doi: 10.1007/978-3-642-14512-4_1.
- Wu, L. and Luo, Y. (2021) 'Bacterial Quorum-Sensing Systems and Their Role in Intestinal Bacteria-Host Crosstalk', *Frontiers in Microbiology*. Frontiers Media S.A., p. 611413. doi: 10.3389/fmicb.2021.611413.
- Wynendaele, E. *et al.* (2013) 'Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides', *Nucleic Acids Research*. Oxford Academic, 41(D1), pp. D655–D659. doi: 10.1093/nar/gks1137.
- Youk, H. and Lim, W. A. (2014) 'Secreting and sensing the same molecule allows cells to achieve versatile social behaviors', *Science*. American Association for the Advancement of Science, 343(6171), p. 1242782. doi: 10.1126/science.1242782.
- Zargar, A. *et al.* (2015) 'Bacterial secretions of nonpathogenic Escherichia coli elicit inflammatory pathways: A closer investigation of interkingdom signaling', *mBio*. American Society for Microbiology, 6(2). doi: 10.1128/mBio.00025-15.

Zhang, B. *et al.* (2019) 'The AI-2/luxS Quorum Sensing System Affects the Growth Characteristics, Biofilm Formation, and Virulence of *Haemophilus parasuis*', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., 9(MAR), p. 62. doi: 10.3389/fcimb.2019.00062.

Zhang, Guishan *et al.* (2012) 'Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon.', *The ISME journal*. Nature Publishing Group, 6(7), pp. 1336–44. doi: 10.1038/ismej.2011.203.

Zhang, L. *et al.* (2020) 'Sensing of autoinducer-2 by functionally distinct receptors in prokaryotes', *Nature Communications*. Nature Research, 11(1), pp. 1–13. doi: 10.1038/s41467-020-19243-5.

Zhu, M. and Gribskov, M. (2019) 'MiPepid: MicroPeptide identification tool using machine learning', *BMC Bioinformatics*. BioMed Central Ltd., 20(1), p. 559. doi: 10.1186/s12859-019-3033-9.

10. Annexes



Review

Interspecific interactions that affect ageing: Age-distorters manipulate host ageing to their own evolutionary benefits

Jérôme Teulière, Charles Bernard, Eric Bapteste *

Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Muséum National d'Histoire Naturelle, EPHE, Université des Antilles, Paris, France

ARTICLE INFO

Keywords:

Antagonistic pleiotropy
Disposable soma theory
Mutation accumulation theory
Virus
Evolutionary conflict
Symbiont
Parasite
Coevolution

ABSTRACT

Genetic causes for ageing are traditionally investigated within a species. Yet, the lifecycles of many organisms intersect. Additional evolutionary and genetic causes of ageing, external to a focal species/organism, may thus be overlooked. Here, we introduce the phrase and concept of age-distorters and its evidence. Age-distorters carry ageing interfering genes, used to manipulate the biological age of other entities upon which the reproduction of age-distorters relies, e.g. age-distorters bias the reproduction/maintenance trade-offs of cells/organisms for their own evolutionary interests. Candidate age-distorters include viruses, parasites and symbionts, operating through specific, genetically encoded interferences resulting from co-evolution and arms race between manipulative non-kins and manipulable species. This interference results in organismal ageing when age-distorters prompt manipulated organisms to favor their reproduction at the expense of their maintenance, turning these hosts into expanded disposable soma. By relying on reproduction/maintenance trade-offs affecting disposable entities, which are left ageing to the reproductive benefit of other physically connected lineages with conflicting evolutionary interests, the concept of age-distorters expands the logic of the Disposable Soma theory beyond species with fixed germline/soma distinctions. Moreover, acknowledging age-distorters as external sources of mutation accumulation and antagonistic pleiotropic genes expands the scope of the mutation accumulation and of the antagonistic pleiotropy theories.

1. Introduction

A broad range of cellular organisms, from microbes to animals and plants, have been shown to age physiologically and even replicatively, prompting the search for the proximal and ultimate causes of ageing across increasingly larger portions of the Tree of Life. Accordingly, diverse kinds of biochemical, ecological and evolutionary processes and events have been proposed to contribute to ageing, but these explanations remain incomplete. Importantly, mainstream explanations of ageing that adopt an evolutionary perspective (Kirkwood and Holliday, 1979; Medawar, 1952; Williams, 1957) have always considered conspecific organisms as the relevant units of selection, i.e. the entities in which longevity genes, antagonistic pleiotropic genes, genes involved in trade-offs that indirectly lead to ageing, or late-expressed deleterious mutations may evolve.

In other words, the genetic causes for ageing have been and still are commonly searched for within a given species, and even within a kin (Bourke, 2007; Hamilton, 1966; Kirkwood and Holliday, 1979; Medawar, 1952; Williams, 1957), but seldom outside the focal ageing

lineage (Johnson et al., 2019). Indeed, the three mutually non-exclusive major evolutionary theories of ageing (Johnson et al., 2019) rest on a common argument: all organisms inevitably die from extrinsic sources, be it by accident, predation, disease outbreak or due to exceptionally harsh conditions (Reichard, 2017). Since no individual is immortal, irrespectively of the existence of ageing, as progressively fewer individuals within populations escape these extrinsic sources of mortality and are thus able to experience the effects of natural selection, the later periods of any organism's lives are predicted to be less important from an evolutionary viewpoint than their early periods (Reichard, 2017). Thus, the Mutation Accumulation theory of ageing (MA) (Medawar, 1952) predicts that late-expressed deleterious mutations will accumulate in organismal genomes, when the strength of natural selection has diminished so much that it is insufficient to purge the older members of the populations from their "rotten" late-expressed genes and to further postpone these genes expressions later in the course of the organisms' lifetime. Hence, senescence, the gradual decrease in biological functions, occurs when such detrimental genes contribute to the decline of organisms late in their lives. The Antagonistic Pleiotropy theory of

* Corresponding author.

E-mail address: eric.bapteste@mnhn.fr (E. Bapteste).

<https://doi.org/10.1016/j.arr.2021.101375>

Received 27 January 2021; Received in revised form 22 May 2021; Accepted 26 May 2021

Available online 31 May 2021

1568-1637/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ageing (AP) (Williams, 1957), proposed by Williams shortly after MA, also fundamentally explains senescence by the decreasing strength of natural selection acting on chronologically ageing individuals. As external death is inevitable, the selection for genes that benefit organisms should be more active early in their lives, even when these genes with early positive fitness effects impair the survival of their bearers later in their lives. Such genes with positive fitness effects early in life, but negative fitness effects later, are therefore called antagonistic pleiotropic genes. Consequently, organismal ageing results from trade-offs favoring the reproduction of individuals over their long-term survival. Finally, the Disposable Soma theory (DS) developed by Kirkwood (Kirkwood and Holliday, 1979; Kirkwood, 1977) stresses that, because resources are limited, most organisms will be evolutionarily more successful if they invest their finite energy into their germline (i.e., the sperm, eggs and their precursor cells) and into their reproduction rather than into the maintenance of their soma (i.e., their other cells, doomed to form temporary bodies that contain, protect and help to propagate the genetic information they share with the germ cells) (Reichard, 2017). Thus, DS principally differs from AP by explicitly assuming optimal dynamical allocation of resources within the life history trade-offs (reproduction, survival) to maximize lifetime fitness (Reichard, 2017). Consequently, both AP and DS theories expect a trade-off between ageing and fecundity (Johnson et al., 2019).

While these approaches and their conclusions are sound and important, it also strikes us that searching for the causes of ageing solely within kin or within a species may obfuscate some additional genetic causes of organismal ageing, which would be external to the focal species, yet directly related to organismal ageing. We agree with Johnson et al. (Johnson et al., 2019), who recently stressed that “*the trend in all three theories has been to designate the organism as the dynamic locus of aging while treating the broader environment as the undifferentiated cumulative source of selective pressures and resources*”. This trend has likely practical and theoretical consequences: “*the locus of aging may need theoretical expansion to embrace more of the pathways that link environment, ecology, population, and the organism to each other*” (Johnson et al., 2019). Convinced by the argument that “*aging, on this proposed view, does not happen to a discrete organism affected by its undifferentiated environment, but rather to an organism variably implicated in a multi-dimensional environment involving ecology and population* (Johnson et al., 2019).”, in this paper we specifically focused on how interspecific interactions on ageing could enhance current explanations of ageing.

Importantly, no cellular organism and no species lives alone. All organisms, including microbial taxa (i.e., archaea, bacteria, and the unicellular eukaryotes, called protists), are affected by genetic interactions with non-kin entities, fundamentally altering the basic biology of these hosts through interspecific interactions. Discoveries about viruses offer a first intuitive path to introduce the general notion of age-distorter (AD) as well as its possible significant impact on ageing, including on human ageing, because viral infections are extremely common and can be widely distributed, easily transmitted and challenging to control. Accordingly, when all biodiversity is considered, specialists estimate that more than 10^{25} infections of cells by viruses occur on Earth each second (Frost et al., 2005), with deep evolutionary consequences (Cobián Güemes et al., 2016; Coutinho et al., 2020; Koskella and Brockhurst, 2014; Pal et al., 2007; Villarreal and Witzany, 2010). Indeed, viruses commonly manipulate cells and gene expression within cells to their own advantage through co-evolution and evolutionary arms races between viruses and cells (Alperovitch-Lavy et al., 2011; De Smet et al., 2017; Feschotte and Gilbert, 2012; Hoover et al., 2011; Iranzo et al., 2014; Labrie et al., 2010; Roossinck, 2011). From a mechanistic perspective, some of these interspecific interferences rely on viral acquisition of a range of host genes, including protein coding genes called virokinases, which are suspected to mimic or to antagonize their host homologs in order to favor viral infection (Mlih et al., 2018; Sodhi et al., 2004). Host-cell communication, host immune response (Aiello et al., 2017; Frasca and Blomberg, 2016) and apoptosis figure

amongst the targets of these genes that are pirated by viruses (Lamiabe et al., 2016; Mlih et al., 2018; Schönrich et al., 2017). Typically, Cytomegaloviruses (CMVs) have long been suspected of harboring genes whose products can modify the cell cycle, interfere with apoptosis, induce site-specific breakage of chromosomes, dysregulate cellular proliferation (Schleiss, 2011), and to use homologs of cellular immune effectors, which may contribute to chronically impact the host immune response (Aiello et al., 2017; Frasca and Blomberg, 2016), seemingly promoting host inflammation, which in turn could facilitate the widespread dissemination of these viruses in their infected hosts.

Fundamentally, the evolutionary fate of viruses and cellular organisms appears nowadays to be uncontroversially intertwined because organismal and viral lifecycles necessarily intersect. In order to complete its own lifecycle and to multiply, a virus must always exploit a host cell, and, at some point, this virus will turn its host cell into a viral factory, also called a virocell in some evolutionary models (Forterre, 2013). From this infected cell, hundreds to thousands of viral particles will eventually emerge, further spreading the virus under the form of viral particles, which confers a role analogous to that of a germline to the viral particles, whereas the virocells would act as the viral soma (Forterre, 2013): a host cell turned into a viral body. Thus, an hijacked host cell constitutes a *bona fide* part of the viral cycle (Forterre, 2016).

These examples show that the gene content and gene expression of host cells truly lie at the crossroads between multiple selection regimes: on the one hand, a selection, well-taken into account by the traditional models of ageing (Kirkwood and Holliday, 1979; Medawar, 1952; Williams, 1957), that acts at the level of the host and that eventually declines with chronological age; on the other hand, a selection that acts at the level of interfering entities from another species, such as viruses taking control over host cells, a phenomenon not currently modeled by the mainstream theories of ageing.

Importantly, viruses are not the only entities whose lifecycles intersect with that of a host: a broad diversity of non-viral parasites and symbionts also affects the lifecycle of their hosts. Therefore, more generally, a host organism or, at least some of its infected cells can lie at the crossroads between (at least) two lifecycles: that of the manipulative taxa (typically viruses, non-viral parasites or non-viral symbionts) and that of the manipulable taxa (the host lineage). As such, a host or at least some of its cells can be and probably should be seen as a common body, temporarily shared by (at least) two distinct entities with possibly contrasted evolutionary interests.

Below, we argue that this sharing of a common body by two entities, be it limited to some cells or corresponding to an entire ageing host organism, matters for ageing studies. To establish our case, we first introduce the phrase and general concept of AD (Fig. 1), and we propose that on the one hand cellular ageing, both in unicellular and in multicellular organisms, but also on the other hand organismal ageing, could figure among the essential cellular/hosts traits that ADs may manipulate for evolutionary reasons. Second, to support our age-distorter hypothesis, we successively present evidence for the existence of viral ADs and of non-viral ADs. Third, although any host species could be impacted by ADs, we briefly discuss how theoretically immortal or very long-lived organisms may resist ADs, whereas prokaryotes are likely to be affected. Fourth, we discuss how the age-distorter hypothesis and experiments to test it should update current knowledge and major evolutionary theories of ageing. We conclude, that for fundamental as well as for practical reasons, the age-distorter hypothesis should now encourage ageing studies to identify what species may be ‘aged’ and used as ‘disposable soma’ by what other biological entities, and what genes contribute to externalizing ageing outside a given cellular lineage in natural communities.

2. Introducing the age-distorter hypothesis

ADs are viruses, microorganisms or multicellular organisms that interfere with host ageing from another species for their own

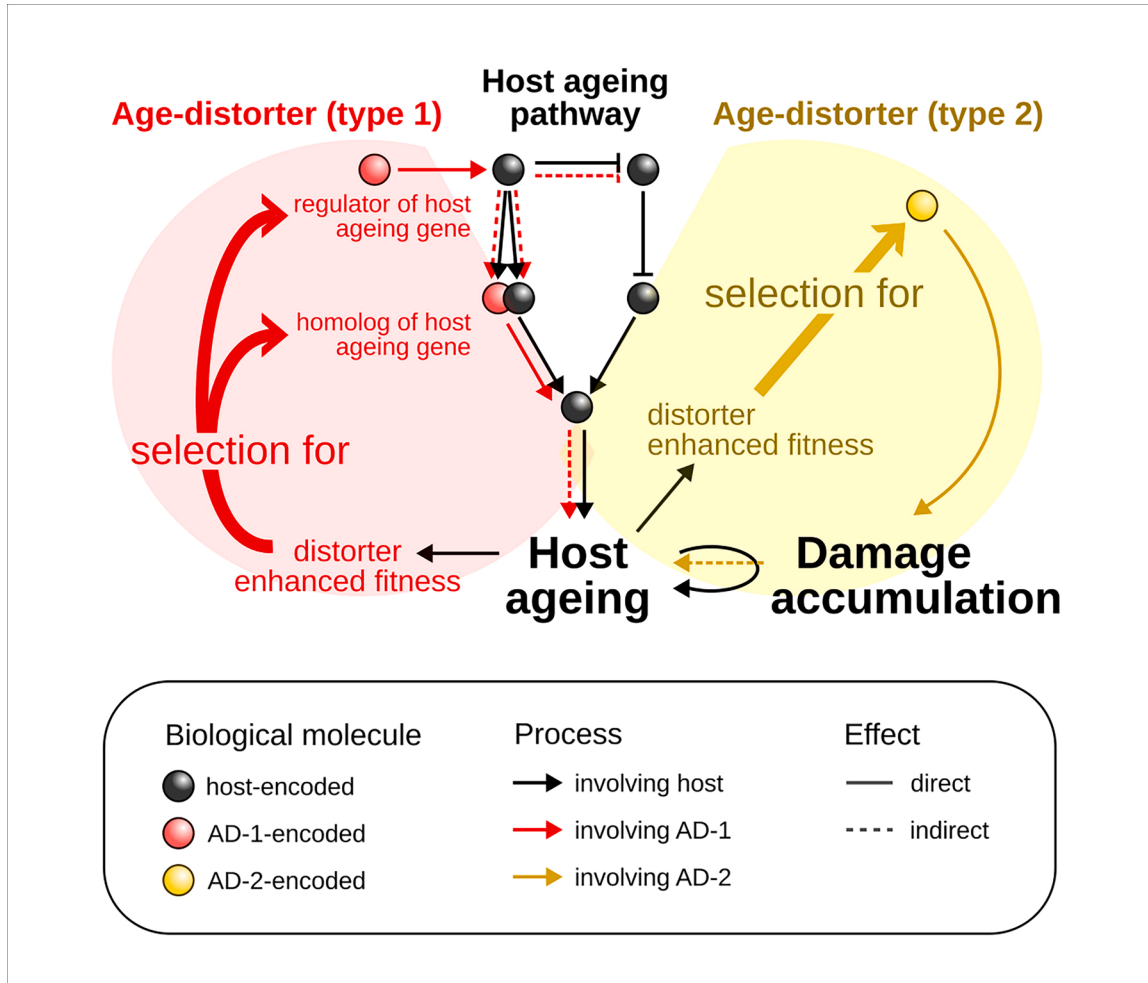


Fig. 1. Introducing the concept of age-distorters.

Age-distorters are viruses, microorganisms or multicellular organisms proposed to hinder or enhance cellular or general host ageing with a evolutionary benefit. We distinguish between age-distorters of type 1 (in red) that hijack the host ageing pathway (in black) and age-distorters of type 2 (in yellow) that indirectly influence host ageing through the modulation of damage accumulation within the host. Age-distorters of type 1 are proposed to act through a regulator (transcription factor, synthase of a signalling molecule, non-coding RNA).

evolutionary benefit. In particular, when the AD lacks an evolutionary interest for maintaining its host or infected host cells in good shape, the AD might take advantage of its own genes, to bias the trade-off between maintenance and reproduction for its own evolutionary interest, towards minimal energetic investment in the maintenance of host infected cells or of the whole host organism. Typically, when infected by ADs able to interfere with their maintenance/reproduction trade-offs, host cells are turned into an expanded disposable soma of the AD. In principle, based on their distinct mode of action, two kinds of ADs could be theoretically distinguished. Age-distorters of type 1 (AD-1) use their own genes to directly hijack and manipulate host ageing pathways, whereas age-distorters of type 2 (AD-2) use their own genes to interfere with the balance of damage accumulation within their hosts and thus indirectly influence host ageing by tipping this balance (Fig. 1). In what follows, unless otherwise specified, we will focus exclusively on AD-1, when discussing ADs.

Many species, including the bacteria *E. coli* K12 (Nyström, 2002), encode trade-offs resulting in ageing (Baudisch and Vaupel, 2012; Johnson et al., 2019; Kirkwood and Holliday, 1979; Williams, 1957), and all of them entertain interspecific interactions with at least some viral, parasitic or symbiotic entities, with which their lifecycles intersect. Therefore, demonstrating the existence of AD requires only two necessary conditions:

- (1) When their two lifecycles intersect, the gene content of an AD

must be able to directly influence the ageing of its host or at least of some of its host cells by interfering with a host maintenance/reproduction trade-off, for example by introducing additional copies of homologous genes or proteins involved in that trade-off in the host, thus in host ageing pathways, or its regulation;

- (2) It must be evolutionarily advantageous for an AD to directly manipulate the ageing of its host or of some of its infected cells, even if this manipulation is detrimental for its hosts (i.e. results in accelerated host ageing with respect to naive hosts).

The ageing of whole host organisms can indeed be impacted either as a result of cellular senescence in the case of unicellular organisms or due to bystander effects of cellular senescence in the case of multicellular organisms. Of note, ADs should not be confused with sex-distorters (Bennett and Moran, 2015). The latter bias the gender of the progeny produced by a manipulated host, whereas the former could bias the timing of the reproduction, and interfere with the ageing of a manipulated entity. Importantly, from an evolutionary perspective, there is in fact a constant selective pressure for age-distorting genes to evolve, because the evolutionary interests of symbionts/parasites or of viruses that replicate faster than their hosts are generally not aligned with the evolutionary interest of their hosts, due to selfish tendencies of the former entities, even when these entities are obligately vertically inherited (Bennett and Moran, 2015).

3. Some evidence for the age-distorter hypothesis

As scientific inquiries are increasingly questioning the causal connections between (human) ageing, ageing-associated diseases and viral infections (Nikolich-Zugich et al., 2020; Nikolich-Zugich and van Lier, 2017; Redeker et al., 2018; Sansoni et al., 2014), recent ageing studies contain some evidence that viruses can act as AD in multicellular host species, ours included. Therefore, and owing to the prevalence of viral infections in nature, we will first review the evidence for viral ADs, before turning onto the cases of non-viral ADs.

3.1. Evidence for viral age-distorters

Several arguments support the existence of viral ADs. First, viruses interact with host defense mechanisms for their own benefits. This is notably the case for inflammation and apoptosis, as illustrated by the classic work by (Gillet and Brun, 1996), who twenty-five years ago presented a diversity of viruses recognized for their ability to counteract inflammation and apoptosis in the aim to keep their host cells alive and to maximize their own reproduction. When apoptosis is used by their hosts as a defense mechanism (Connolly and Fearnhead, 2017), latent viruses generally prevent the death of infected host cells, often through expression of anti-apoptotic factors (Connolly and Fearnhead, 2017), unraveling interconnections between the viral life cycle and host cell stress and programmed cell death pathways (Iranzo et al., 2014; Veyer et al., 2017; Yatim and Albert, 2011). In short, many viruses (including Papovaviricetes, such as the Human Papillomavirus, the *Macaca mulatta* polyomavirus 1 SV40 (Connolly and Fearnhead, 2017); Herpesvirales, such as the Epstein-Barr virus, the Kaposi's sarcoma-associated herpesvirus (Connolly and Fearnhead, 2017; Kalvakolanu, 1999); Riboviria, such as the Human T-lymphotropic retrovirus (HTLV-1) or the influenza virus (Yatim and Albert, 2011; Zamaraev et al., 2020); Bamfordvirae, such as the African swine fever virus, the Human adenoviruses 12 and 18, or Poxviruses (Kalvakolanu, 1999; Kaminsky and Zhivotovsky, 2010); and Baculoviruses (Kaminsky and Zhivotovsky, 2010) have developed mechanisms to suppress or to modify their host cell apoptotic responses for their own benefit (Zamaraev et al., 2020). Importantly, such viral interactions with apoptosis are recognized to affect host homeostasis. By affecting the systemic environment of an organism, dysregulation of apoptosis appears as a pleiotropic driving force of tissue ageing. Too little apoptosis can promote cancer or age-related and detrimental cell senescence, while too much can promote tissue atrophy or hyperplastic overgrowth through compensatory proliferation (Mlih et al., 2018). Logically, viruses that interfere with apoptosis could manipulate or tip this subtle balance and thus, in metazoans at least, interfere with host ageing in the long run (Gillet and Brun, 1996).

Importantly, inflammation and apoptosis belong to a broader range of host defense mechanisms, in which the implication of some longevity-promoting genes, such as sirtuin 1 (Kim et al., 2016), of some ageing promoting genes, such as the mechanical Target Of Rapamycin mTOR kinase (Choi et al., 2018; Kaminsky and Zhivotovsky, 2010), or of some ageing regulating genes, such the onco-suppressive transcription factor p53, which notably contributes to control cell cycle arrest, cell ageing, DNA repair, and regulated cell death (Zamaraev et al., 2020), have been reported. Therefore, host defense and host ageing can be mechanistically connected, as notably recognized for autophagy (Choi et al., 2018) and for cellular senescence (Mlih et al., 2018). Since it is reasonable to expect that viruses will evolve to overcome some of these host defenses to successfully replicate and spread (Zamaraev et al., 2020), it is logically expected, and in some cases even demonstrated, that viruses have evolved some genetic potential to interfere with host ageing at the cellular and/or at the general levels.

For example, RNA viruses have acquired a powerful ability to hijack and subvert autophagy for their own benefit. Thus, the Zika virus uses its NS42 and NS4B proteins to inhibit mTOR activation in fetal neural stem cells to induce autophagy; and a highly pathogenic avian H5N1 strain of

IAV can also induce autophagy by inhibiting mTOR (Choi et al., 2018). Likewise, poliovirus benefits from the autophagy-inducer and mTOR inhibitor rapamycin to increase its replication (Choi et al., 2018). Additionally, many viruses (such as adenovirus, SV40, smallpox virus, HIV-1) use their own proteins to manipulate p53 (Zamaraev et al., 2020); and the Env-encoded protein of HIV-1 triggers apoptosis through the activation of mTOR and p38 MAP kinase and transactivation of the p53-dependent genes PUMA and Bax (Kaminsky and Zhivotovsky, 2010).

Moreover, when cellular senescence is considered as an innate defense mechanism to limit microbial replication (Humphreys et al., 2020), cellular senescence increasingly appears as a target of choice for viral manipulation. During ageing, increasing the number of senescent cells contributes to the chronic inflammation and deregulation of the immune function, which in turn increases susceptibility to infectious diseases in ageing organisms (Humphreys et al., 2020). Therefore, cellular senescence is, from that perspective, a weakness of ageing hosts, which some viruses (and some bacterial pathogens, as we will mention later) have learned to exploit during their evolution, in particular through the use of senescence-inducing toxins (Humphreys et al., 2020). As senescent cells do not replicate but stay viable and metabolically active, a broad range of viruses (see Table 1 in (Seoane et al., 2020)) are able to induce cellular senescence pathways to exploit senescent cells to improve their own replication (Humphreys et al., 2020; Seoane et al., 2020). By contrast, other viruses encode proteins that specifically inhibit cellular senescence (see table 2 in (Seoane et al., 2020)). For example, the Kaposi's sarcoma-associated herpesvirus uses vCyclin that contributes to suppress replicative senescence in primary human lymphatic endothelial cells (Humphreys et al., 2020). Likewise, HPV viral proteins E6 and E7 inhibit cellular senescence by destabilizing pRb and activating telomerases, and an oncogenic virus, such as the Epstein-Barr Virus, uses its LMP1 protein to induce activation of NF-kB and promote proliferation and survival of host infected cells (Zamaraev et al., 2020).

Consequently, there is no doubt that some viruses interfere with host ageing at the cellular level, and in many different ways (Kelley et al., 2020). However, the evolutionary reasons for viral interference with cellular senescence are debated. On the one hand, acute induction of cellular senescence by viruses could be part of the host's antiviral immune response, when this defense mechanism allows virus-infected cells to secrete SASP factors that limit viral replication in neighbouring cells (Kelley et al., 2020). Under this interpretation, viruses would be manipulated by their hosts accrued senescence, rather than the opposite. Conversely, in some instances at least, viruses may have instead evolved the ability to manipulate host cellular senescence for their own benefits, supporting the age-distorter hypothesis. This second interpretation is likely, since some viruses, including human papillomavirus or the varicella zoster virus, replicate more efficiently in senescent human bronchial epithelial cells or in human dermal fibroblasts than in non-senescent cells, and have evolved machinery to specifically overcome cellular senescence (Kelley et al., 2020). Indeed, these viruses have been proposed to hijack the senescent machinery within the infected cell to downregulate the interferon (IFN) pathway, to promote viral packaging and ultimately to enhance viral replication within the cell (Kelley et al., 2020; Kim et al., 2016).

Importantly, virally-induced manipulation of cellular senescence is expected to affect global host ageing beyond the fate of host infected cells. Virally-induced senescence is expected to contribute to age-associated diseases (Kelley et al., 2020), due to its resulting dysregulation of the host homeostasis, e.g. when senescent cells accumulate faster than they are removed from an ageing host due to a viral AD, and/or due to an aggravation of immunosenescence, caused by the known bystander effect exerted by senescent cells on their neighbours (Karin et al., 2019).

In addition to this evidence that some viruses would have evolutionary interest to manipulate host cell ageing, sometimes with systemic consequences on general host ageing, both the scientific literature

Table 1
Some possible instances of age-distorters.

| Candidate AD | Host | Age-distorting component | Proposed mechanism of action | References |
|---|---|---|---|---|
| Viral AD | | | | |
| Human immunodeficiency virus 1 (HIV-1) | Multicellular (Human) | <i>nef, tat</i> genes | Stress-induced premature senescence of non-immune cell and immunosenescence, inducing osteoporosis and osteopenia | Cohen and Torres, 2017 |
| Dengue virus (DENV) | Multicellular (Human) | viral receptor DC-SIGN | Exploitation of host senescence program | Seoane et al., 2020 |
| Insect viruses (<i>Entomopoxvirinae</i> , <i>Ascoviridae</i> and <i>Baculoviridae</i>) | Multicellular (Drosophila) | <i>diadel</i> | Exploitation of a host endogenous pathway: the immune deficiency pathway, to modulate NF- κ B signalling and promote fly survival during the antiviral response | Lamiabile et al., 2016 |
| Hepatitis B (HBV) and C (HCV), Epstein Barr virus (EBV) | Multicellular (Human) | HBV-encoded HBx, EBV-encoded LMP1 | Virally-induced oxidative stress, triggering metabolic changes and the activation of the NF- κ B signalling pathway, which promotes oncogenesis in liver tissues and blood | Zamaraev et al., 2020 |
| Non-viral AD | | | | |
| <i>Listeria monocytogenes</i> | Multicellular (Human) | Debated | Dose-dependent depletion of host dNTPs pools, interference with the host DNA replication, culminating in fork arrest and DNA strand breaks, possible hijacking of host signalling pathways, triggering DNA damage/repair signaling, all of which introduces an S-phase delay in host infected cells | Leitão et al., 2014 |
| <i>Salmonella typhi</i> | Multicellular (Human) | Typhoid toxin (<i>ctdB, pltA</i> and <i>pltB</i> genes) | Production of typhoid toxin driving transmissible senescence within human through persistent DNA damages | Ibler et al., 2019; Humphreys et al., 2020 |
| <i>Pseudomonas aeruginosa, Helicobacter pylori, Neisseria gonorrhoea</i> and <i>Shigella flexneri</i> | Multicellular (Human) | Cytolethal distending toxins | Production of ROS that promote oxidative DNA damage, thereby impairing host DNA repair mechanisms | Leitão et al., 2014; Benedetti et al., 2021 |
| Gram-negative bacteria | Multicellular (Human) | PAMPs, such as lipopolysaccharides (LPS) | Induction of dental stem cells senescence via persistent inflammatory response | Feng et al., 2018 |
| <i>Escherichia coli</i> | Multicellular (<i>C. elegans</i>) | 29 genes, including CA regulators <i>lon, hns, uidC, hyfR</i> and <i>sadC</i> | Interference with the insulin/IGF-1 and the TOR signalling pathways; downregulation of colonic acid production, an important co-factor of host mitochondrial repair mechanisms, resulting in accelerated host ageing | Han et al., 2017 |
| <i>Bacillus subtilis</i> | Multicellular (<i>C. elegans</i>) | <i>nosA, phrC</i> genes | Nitric oxide production, quorum sensing CSF pentapeptide release and biofilm formation, all independently inducing an upregulation of host genes involved in thermoresistance and longevity | Gusarov et al., 2013; Donato et al., 2017; Smolentseva et al., 2017 |
| <i>Comamonas</i> DA1877 | Multicellular (<i>C. elegans</i>) | Yet uncharacterized signal that is interpreted by <i>C.elegans</i> intestinal cells | Alteration, through NHR-23, of the cyclic expression of host-encoded genes associated with molting, resulting in accelerated development of the host, coupled with a shorter host lifespan | MacNeil et al., 2013 |
| <i>Acetobacter pomorum</i> | Multicellular (Drosophila) | Pyrrroloquinoline quinone-dependent alcohol dehydrogenase | <i>A. pomorum</i> -produced acetic acid is an important co-factor of the insulin/insulin-like growth factor signalling (IIS) in Drosophila, which regulates host homeostatic programs controlling developmental rate, body size, energy metabolism, and intestinal stem cell activity | Shin et al., 2011 |
| <i>Proteobacteria (Klebsiella variicola, Acetobacter fabarum)</i> | Multicellular (Drosophila) | <i>cbs</i> and <i>cgl</i> genes | Driving of metabolic flux through transsulfuration, extending <i>D. melanogaster</i> lifespan | Matthews et al., 2020 |
| <i>Hymenolepis diminuta</i> (tapeworm) | Multicellular (<i>Tenebrio molitor</i>) | Uncharacterized | Reduction of host ageing, through the induction of a host-derived signalling molecule that blocks resource allocation to reproduction in <i>T. molitor</i> females; and a diminution of sex pheromone production and of the response to female copulatory pheromone in <i>T. molitor</i> males. | Hurd et al., 2001 |
| Trematodes (e.g. <i>Renicola rescovita, Microphallus pseudopygmaeus, Parorchis acanthus</i>) | Multicellular (e.g. <i>Littorina littorea, Onoba aculeus, Cerithidea californis</i>) | Uncharacterized | Host castration, which leads to host gigantism and prolonged lifespan | Sorensen and Minchella, 2001 |
| <i>Acanthocheilonema viteae</i> (nematode) | Multicellular (Mouse) | Excretory-secretory molecule ES-62 | ES-62 preferentially targets male gut-metabolic tissue axis in male HCD-mice, acting to normalize gut microbiome changes and loss of intestinal barrier integrity to protect against mortality | Crowe et al., 2020 |

(Kaminsky and Zhivotovsky, 2010; Schleiss, 2011) and a survey of empirical data (Fig. 2) provide compelling molecular evidence further supporting the case that viral ADs could operate by hijacking host ageing pathways. Namely, beyond known virokines, homologs of genes involved in host ageing pathways in model ageing organisms are present in numerous viruses, which, owing to the strong selective pressures to limit viral genome sizes, indicates that these viral-encoded genes are likely of some use to viruses. Current state-of-the-art knowledge

regarding the genetic potential of viruses to directly impact host ageing pathways using viral homologs of these pathways can be deduced by conjugating three types of data: i) curated host protein-protein interaction networks, defining longevity networks (Bell et al., 2009; Budovsky et al., 2007; Fernandes et al., 2016), ii) annotations from the CellAge (Avelar et al., 2020) and the GenAge (Tacutu et al., 2018) databases, which summarize the association of host genes with ageing in various eukaryotic hosts, and iii) complete viral genomes. Results of homology

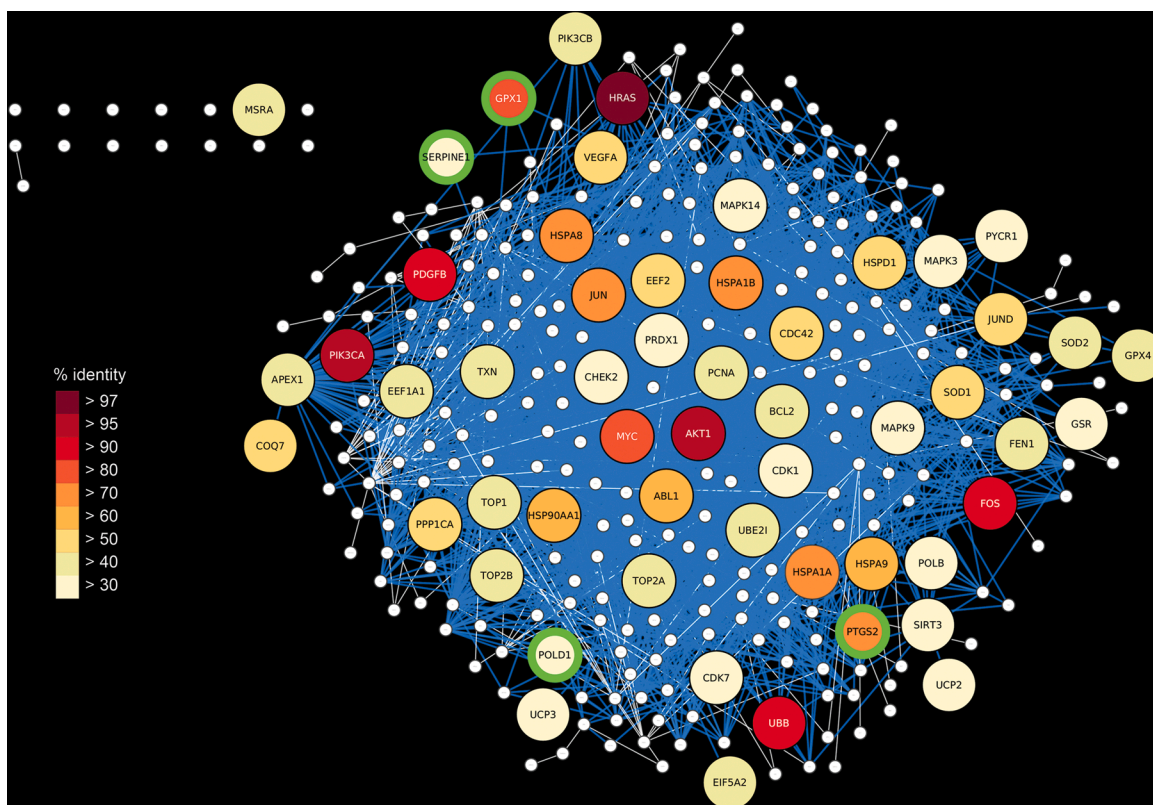


Fig. 2. Mapping of all known viral homologs of ageing-associated genes on the human protein protein interaction network. The human longevity-associated protein-protein interaction (PPI) network is derived from the human PPI network from Biogrid (Release 4.3.195), restricted to the list of human proteins involved in regulating longevity, ageing and cell senescence present in the GenAge (Build 20) and CellAge (Build 2) databases. Enlarged colored nodes (17,6% of the network) represent proteins with homologs among all viral proteins available at the NCBI as of January 2021, identified by BLAST homology searches with a minimal coverage of 70 % for increasingly high percentage of identity thresholds (E-value <1e-5). Node colors reflect the maximum percentage of identity (see color legend) for which at least one viral hit was found for the corresponding human protein. Nodes outlined in green indicate proteins with homologs in viruses able to infect *Homo sapiens*. Blue edges (representing 94,7% of the interactions of the network) indicate direct interactions involving nodes with viral homologs, stressing their centrality. White nodes correspond to proteins involved in regulating longevity, ageing and cell senescence without known viral homologs. White edges represent direct interactions between proteins involved in regulating longevity, ageing and cell senescence without known viral homologs.

searches between viral and host proteins can then be represented. Reporting these homologies on curated protein-protein longevity network (Fig. 2) yields an overview of the proteins associated with host ageing, including those also encoded by currently sequenced viruses (hence likely pirated by viruses (Schönrich et al., 2017)). Such a network also shows the numerous direct interactions of proteins encoded by such ageing-associated genes, highlighting the potential reach and centrality of proteins with viral homologs within the longevity network. In the human longevity network, these viral homologs include, for example, viral homologs with extremely high similarity (E-value < E-5; mutual cover ≥ 90 %, %ID ≥ 90 %) to human age-associated genes, such as the oncogenes Src, Ras, PI3K, and p21/CIP, experimentally demonstrated to interfere with the progression of host cellular cell-cycles by activating cell division, and also reported for their contribution to ageing, via their connection with the insulin, MAPkinases and Cyclin/CDK pathways. Therefore, both theoretical and current empirical knowledge support the hypothesis that viruses may carry genes that can directly influence cellular ageing and host ageing by interfering with host ageing pathways.

Finally, published analyses of the impact of specific viral infections on animal and/or human health can be reinterpreted as published examples of virally induced age-distortion. Thus, the remarkable work by Lamiable et al. (Lamiable et al., 2016) provides documented evidence of a viral AD enhancing its host lifespan. These authors showed that insect DNA viruses encode a homolog of an anti-apoptotic cytokine, the *Drosophila* gene *diedel* (*die*), demonstrated to be involved in *Drosophila* longevity. When secreted from muscle or adipose tissues, *Diedel* can

attenuate age-related intestinal tissue degeneration by promoting proliferative homeostasis of stem cells (Mlih et al., 2018). This circulating protein is both necessary and sufficient to limit tissue degeneration and to regulate lifespan. By blocking cell death in organs during ageing, *Diedel* limits deterioration of organ function and extends the organism's lifespan (Mlih et al., 2018). Consistently, *die* mutant flies have shortened lifespans and succumb more rapidly than controls when infected by viruses that do not harbor a viral homolog of *die*, as a result of a deregulated activation of the immune deficiency (IMD) pathway of host defense. Precisely, this latter endogenous pathway is manipulated by insect viruses harboring viral homologs of *die*. Such viruses modulate Nuclear Factor κB (NF-κB) signalling, prevent IMD pathway activation and promote fly survival during the antiviral response, while allowing viruses to spread (Mlih et al., 2018). Significantly, viral homologs of *die* are found in the genomes of Entomopoxvirinae, Ascoviridae and Baculoviridae, three large classes of unrelated insect viruses, testifying of the evolutionary success of this particular strategy of hijacking.

Likewise, studies of HIV infection provide other well-documented instances of host age-distortion induced by a virus. Experiments demonstrate that several virally-encoded proteins, including Env, Vpr, Tat, Nef and Vou, contribute to HIV-induced cell apoptosis (Schank et al., 2021). Each of these proteins can use bystander-induced apoptosis, during which infected cells expressing virally-encoded proteins can interact with neighboring uninfected cells to induce apoptosis. For example, Tat is able to trigger apoptosis and take part in the bystander effect and infection of neighboring cells through CCR5 and CXCR4 receptors. In CD4 T cells infected with HIV, Tat exposure alone is

also able to induce telomerase reduction. Thus, Tat damages the replicative potential of both infected and noninfected CD4 T cells, potentially contributing to the depletion of functional CD4 T cells in people living with HIV (Schank et al., 2021), contributing to their severe acquired immunodeficiency, characterized by susceptibility to infections and certain types of infection-associated cancer (Kaposi sarcoma and B cell lymphoma) (Humphreys et al., 2020). By diminishing the numbers of CD4 + T cells, HIV infection induces premature senescence of non-immune cells and immunosenescence (Cohen and Torres, 2017; Schank et al., 2021). Consistently, expression of HIV Tat in endothelial cells or in transgenic mice upregulates the expression of the microRNA miR-34a, a molecule that targets sirtuin 1, leading to the induction of senescence (Seoane et al., 2020). Furthermore, Tat and Nef are known to interfere with different host mechanisms, the NF- κ B signalling pathway for Tat and the inhibition of autophagy for Nef, inducing premature senescence in bone marrow mesenchymal stem cells, which impairs their osteoblastic potential (Cohen and Torres, 2017). Consistently, it has been proposed that HIV-infected patients suffer from accelerated ageing (Cohen and Torres, 2017), including HIV1-related osteoporosis and osteopenia (Seoane et al., 2020).

Other viruses (Table 1) with characterized age-distorting effects, but often less well characterized effector genes, include the human respiratory syncytial virus, which can cause DNA damage-induced senescence in lung epithelial and mononuclear cells in mice, ultimately leading to airway tissue remodelling and permanent tissue damage (Kelley et al., 2020); the Dengue virus (DENV), which has been demonstrated to lead to cellular senescence in infected human umbilical vein endothelial cells, possibly because senescent cells expose an increased number of viral receptors, such as DC-SIGN, to which DENV can attach, which facilitates its infection (Seoane et al., 2020); or the SARS-CoV-2 virus, recently proposed to promote senescence in coronary artery endothelial cells, possibly by inducing mitochondrial dysfunction indirectly via its ACE-2 entry receptor (Humphreys et al., 2020; Singh et al., 2020). Moreover, the relationship between human cytomegalovirus (HCMV) infections and accelerated immune senescence has also been debated for a long time, and appears supported by some recent studies demonstrating that a high infectious dose is a prerequisite for CMV-associated immune senescence (Redeker et al., 2018), and that HCMV infection results in telomere erosion and replicative senescence, faster and more frequently in CD8+ cytotoxic T lymphocytes compared to CD4 + T helper cells (Humphreys et al., 2020), through continuous antigen exposure. The latest consensus is that CMV does not have a clear and measurable negative immune senescence impact (Nikolich-Zugich et al., 2020).

In addition to viral ADs, non-viral ADs also affect ageing in multicellular organisms.

3.2. Evidence for non-viral age-distorters

Beyond viruses, any entity able to manipulate the lifecycle of another, using its own genetic information, in ways that bias the reproduction/maintenance trade-off of the cells/organisms with which its lifecycle intersects, can act as a potential AD. Supporting this view, evidence for non-viral ADs affecting multicellular hosts is also available throughout the literature. We present it next, organized based on the taxonomy of ADs (parasitic bacteria, then endosymbiotic bacteria, then bacteria from the host microbiota, and multicellular eukaryotic parasites).

The human-adapted intracellular pathogen *Salmonella typhi* offers a clear illustration of pro-ageing bacterial AD (of type 2). *S. typhi* invades human cells within *Salmonella*-containing vacuole and may induce mortality either via typhoid fever or chronic infections. Ibler et al. characterized that one mechanism underlying *Salmonella typhi* infections involves senescence-promotion in the host via a secreted virulence factor, called the typhoid toxin (Ibler et al., 2019). Specifically, this toxin enters bystander human cells and causes DNA damage through

nuclease activity. Toxin-induced DNA damage leads to the accumulation of single-stranded DNA (ssDNA) substrates, which progressively overloads the host Replication Protein A (RPA) pathway, a process that protects ssDNA from breakage in preparation for subsequent repair. RPA exhaustion within host cells results in a Senescence-Associated Secretory Phenotype (SASP), a hallmark of cellular senescence characterized by the excretion of proteins into the extracellular medium that induce secondary senescence in bystander cells. The production of the typhoid toxin thus confers an evolutionary advantage upon *Salmonella typhi* as it drives transmissible senescence within human cells, a phenotype shown to be significantly associated with enhanced *Salmonella* invasion. Importantly, the described deleterious effect of the typhoid toxin on host ageing is just a representative example. Many other cases highlight that toxin-induced senescence via DNA-breaks is a common evolutionary strategy employed by pathogens to facilitate their proliferation within the host. These additional examples of induced senescence involve the toxin B of *Clostridioides difficile* (Fettucciari et al., 2018), the colibactin toxin of certain *Escherichia coli* strains (Smolentseva et al., 2017), the pyocyanin toxin of *Pseudomonas aeruginosa* (Muller, 2006), the cytotoxic distending toxin of *Haemophilus ducrei* (Blazkova et al., 2010) and the pneumolysin toxin of *Streptococcus pneumoniae* (Kwon et al., 2017).

In addition to toxins, pathogens-associated molecular patterns (PAMPs) might induce senescence of host bystander cells, strengthening the case for pro-ageing bacterial ADs. Indeed, Feng et al. observed that human dental stem cells undergo senescence when treated with lipopolysaccharide (LPS), a major component of the cell wall of gram-negative bacteria (Feng et al., 2014). Later, Feng et al. identified that in LPS-induced senescent stem cells, LPS overstimulates the pattern recognition Toll-like receptor 4 (TLR4), which, in turn, activates NF- κ B that induces the expression of pro-inflammatory cytokines (Feng et al., 2018). In other words, PAMPs, like toxins, can also induce host senescence for the benefit of bacterial invasion, albeit rather through the promotion of a persistent inflammatory micro-environment.

Moreover, many bacteria are known to interfere with their host's cell cycle. Thus, *Shigella* and pathogenic *E. coli* can block their host cells in the G2/M phase transition, whereas *Neisseria gonorrhoea* and *Porphyromonas gingivalis* can inhibit cell proliferation via G1 arrest (Leitão et al., 2014). However, the case of the human intracellular pathogen *Listeria monocytogenes* arguably provides the most compelling evidence for bacterial ADs interfering with their host cell reproduction. These bacteria multiply in the host cytoplasm, rerouting resources available for the host metabolism, such as dNTPs, for their own benefit. This interference depletes the pools of dNTPs, interferes with the host DNA replication, culminates in fork arrest and DNA strand breaks within the infected host cells. *L. monocytogenes* is also proposed to hijack host signalling pathways, triggering DNA damage/repair signalling that possibly leads to S-phase delay, until these breaks are repaired by the host. Delaying the progression of its host cell cycle in a dose dependent way would benefit *L. monocytogenes* for two reasons: first, it favors the bacterial infection, because *L. monocytogenes* infected cells divide giving rise to infected daughter cells; second it optimizes the control of the intracellular numbers of *L. monocytogenes* to avoid further cell damage (Leitão et al., 2014).

Illustrating yet another mechanism of bacterial age-distortion, *Chlamydia trachomatis* can use its genes to enhance its host cell's longevity. This bacterium was reported to inhibit, in cultured human fibroblasts, telomeric DNA damage via transient upregulation of the human telomerase catalytic subunit (Padberg et al., 2013). In cells cleared of *C. trachomatis* infection, average telomere length was increased and the activity of the γ H2A.X marker of DNA damage was reduced. *C. trachomatis*-induced alteration of the host ageing pathways might result in host cells with increased longevity, more prone to genomic instabilities and could therefore underlie the observed association between *C. trachomatis* infection and cervical cancer (Anttila et al., 2001). Hence, according to Humphreys et al., this age-distorting mechanism could perhaps help the bacterium to reprogram host cells

in order to create a suitable niche for proliferation (Humphreys et al., 2020).

Some endosymbiotic bacteria have also been demonstrated to interfere with host longevity. For instance, non-virulent strains of *Wolbachia* have been shown to extend the longevity in certain *D. melanogaster* genotypes (Fry and Rand, 2002). In hybrid crosses between two strains, one that lives longer with *Wolbachia* and one that does not, statistical analyses revealed that host survival in *D. melanogaster* depends on interactions between *Wolbachia* infection, host genotype, host sex, and whether reproduction occurs. Endosymbionts are also likely able to interfere with the rate of sexual development and with reproduction, or with the development of maintenance systems (Chow et al., 2010), as they engage in co-evolution and arms races with their hosts (Bennett and Moran, 2015). For example, *Wolbachia* infection can lead to an increase in the mitotic activity of germline stem cells (GSC) (Fast et al., 2011). This was first hinted by the observation that the total number of eggs laid per *Wolbachia*-infected female (W+) was 3.5 times higher than that observed in non-infected *D. melanogaster* (W-). Then, Fast et al. found that relative to W- flies, W+ flies were characterized by 2 times more GSC division and by 50 % less programmed cell death in the germline. The increased egg production explained by these two *Wolbachia*-induced manipulations was proposed to maximize *Wolbachia* spreading through maternal transmission.

In addition to parasitic or endosymbiotic bacteria, an increasingly supported source of ADs is attributed to host-associated microbiomes (Garagnani et al., 2014; Heintz and Mair, 2014). Notably, host-associated bacteria have been shown to interfere with the survival and reproduction of their hosts or of their host's cells. For instance, by comparing the lifespan of *D. melanogaster* when fed with killed bacteria as opposed to living bacteria, Brummel et al. observed that the mean lifespan of axenic flies was 30 % less than that of untreated flies (Brummel et al., 2004). Consistently, when fed with living bacteria and further treated with antibiotics, the lifespan of *D. melanogaster* was reduced by 35 % as compared to flies maintained with the same food but without antibiotics. Next, Brummel et al. studied the effect of adding living bacteria to adult flies that had been raised axenically to adulthood. They found that the introduction of living bacteria during the first week of adult life can enhance lifespan whereas later in life, this bacterial presence can reduce lifespan. More specifically, some bacterial strains have been shown to be associated with either increased or decreased host longevity (Matthews et al., 2020). Indeed, in a systematic experiment assessing the individual contribution of 39 intestinal bacterial strains to the lifespan of *D. melanogaster*, Matthews et al. observed that *Acetobacteraceae* species tended to result in a shorter lifespan of the fly, as opposed to *Lactobacillus* species which were associated with enhanced host longevity, while other species did not follow this trend. On a smaller scale, Schwarzer et al. identified that host-associated microbes can interfere with the survival and reproduction of specific host cells. Indeed, they observed that DNA replication and proliferation of intestinal enterocytes and stem cells is increased by microbiota, both in *Drosophila* larvae and adults (Schwarzer et al., 2018).

In line with these observations, microbiota are expected to have systemic effects on host ageing. This is because, in metazoans, interactions between the host immune system and the gut microbiome maintain the homeostasis of the gut. Therefore, dysbiosis leads to immune dysregulation, promoting chronic inflammation and development of tumors. As gut microorganisms and their toxins migrate through the leaky guts of their ageing hosts via circulatory systems, gut microbes disequilibrate host physiology and contribute to inflammation and tumorigenesis in specific organs via the secretion of various neuroactive molecules through the gut-brain axis, the gut-hepatic axis, and gut-lung axis (Ge et al., 2021b). Typically, the immune-modulating effects of the human microbiota extend well beyond the gut, mostly through the small molecules they produce (Uribe-Herranz et al., 2021), and microbial communities associated with ageing host have therefore been proposed to contribute to several age-associated phenomena, such as

immunosenescence and inflamm-ageing (Biagi et al., 2011).

Giving credit to the existence of systemic effects of ADs involving a manipulation of the trade-offs between reproduction and maintenance, it has for instance been discovered that microbiota composition can influence the rate of development and time of sexual maturation of their hosts (McFall-Ngai, 2002; Schwarzer et al., 2018; Werren et al., 2008). For example, bacteria such as *Acidobacter pomorum* produce acetic acid, an important co-factor of the insulin/insulin-like growth factor signalling in its host, *D. melanogaster* (Shin et al., 2011). This molecular interaction contributes to the healthy homeostasis, in the host, of regulatory programs controlling developmental rate, body size, energy metabolism, and intestinal stem cell activity. Accordingly, Shin et al. proposed that the contribution of *A. pomorum* metabolism to the host homeostatic programs maintains the mutualism between the gut of *C. elegans* and *A. pomorum*. Another example is *Comamonas* DA1877 in *C. elegans* (MacNeil et al., 2013), which contributes to accelerated development of its hosts (and in the case of the worm to a shorter lifespan (MacNeil et al., 2013)). Indeed, MacNeil et al. observed that *C. elegans* fed with *Comamonas* DA1877 had a shorter lifespan than *C. elegans* fed with *E. coli* OP50. They showed that the negative effect of *Comamonas* diet on the host lifespan cannot be attributed to a weak nutritional value nor to a certain pathogenicity of the bacterium. Rather, the authors observed a differential expression of host-encoded molting-associated genes when the worm is fed with *Comamonas* as compared to a diet with *E. coli*. MacNeil et al. proposed that *Comamonas* generates a signal that is interpreted by gene regulatory networks of nematode cells, most likely intestinal cells. Thereupon, this signal would affect *C. elegans* cyclic gene expression during molting, likely through the host-encoded NHR-23 receptor, leading to the accelerated development and shortened lifespan observed in the host. A final example illustrates that microbiota can stimulate an earlier production of the ecdysone hormone, interfering downstream of cell-autonomous TOR signalling in the endocrine glands of their hosts (Schwarzer et al., 2018; Storelli et al., 2011), accelerating the development of *D. melanogaster*.

In addition to interfering with host developmental processes that affect host lifespan, genes of the microbiota can be even more directly intertwined with host ageing pathways. For instance, analyzing the gene content of bacterial strains influencing *D. melanogaster* lifespan, Matthews et al. identified an enrichment of genes associated with ageing pathways, namely glucagon signalling, and cysteine and methionine metabolism (Matthews et al., 2020). Pursuing with functional studies, they identified that *D. melanogaster* had a shortened lifespan in a mono-association with *Acetobacter fabarum* wild-type, as opposed to a mono-association with a *A. fabarum* strain engineered to overexpress the cystathionine beta synthase (*cbs*) and gamma ligase (*cgl*), two central enzymes of the multi-step transsulfuration pathway that converts methionine to cysteine. Matthews et al. proposed that bacterial transsulfuration enzymes can extend *D. melanogaster* lifespan by driving metabolic flux. Another example comes from studies of *Bacillus subtilis*. This commensal bacterium of *C. elegans*, on which the worm feeds during plant degradation, appears responsible of extending its host lifespan. *Bacillus*-induced longevity in *C. elegans* acts through nitric oxide (NO) synthase (Donato et al., 2017; Gusarov et al., 2013), quorum sensing CSF pentapeptide release (Donato et al., 2017), and biofilm formation (Donato et al., 2017; Smolentseva et al., 2017). All three factors independently induce an upregulation of host genes involved in thermoresistance and longevity and may thus impart a fitness benefit to *B. subtilis* by contributing to the evolutionary stability of its association with *C. elegans*, as the latter offers protection and a continuous flow of nutrients to the bacterium. Last but not least, bacterial genes affecting the longevity of *C. elegans* have been identified in a systematic experiment. Specifically, Han et al. discovered 29 *Escherichia coli*-encoded genes that increase *C. elegans* longevity upon deletion (Han et al., 2017), some of which notably interact with the insulin/IGF-1 and/or the TOR signalling pathways. Other genes like *lon*, *hns*, *uidC*, *hyfr* and *sadC* contribute to *C. elegans* ageing by independently downregulating, in

E. coli, colanic acid (CA) biosynthesis (Han et al., 2017). CA polymers, alone, are sufficient to extend the lifespan of the worm, which indicates that CA-induced longevity acts directly in the host and not through CA-dependent bacterial mechanisms. Specifically, upon CA-uptake by intestinal cells, Han et al. demonstrated that CA promotes longevity in *C. elegans* via the host-encoded ATFS-1 protein, a transcription factor that contributes to mitochondrial homeostasis and induces transcription of mitochondrial chaperones under stress conditions. In other words, *E. coli* accelerates host ageing via genes that inhibit the biosynthesis of CA, an important co-factor of host mitochondrial repair mechanisms. These deleterious effects on host longevity are consistent with the overall pathogenic tendency of *E. coli* strains relative to *C. elegans* (Merckx-Jacques et al., 2013).

Finally, interspecific interactions between hosts and non-microbial species can also be re-interpreted as evidence for the existence of multicellular ADs. Thus, parasitic helminths are known to release excretory-secretory molecules (ES) that act as immunomodulators within their host (Harnett, 2014). For instance, the ES-62 product of the nematode *Acanthocheilonema viteae* has been shown to resolve chronic inflammation in mice, by normalizing aberrant MyD88 signalling to homeostatically restore immunoregulation (Crowe et al., 2020). Because adoption of a high calorie diet (HCD) typically results in chronic MyD88 signalling, Crowe et al. investigated the potential therapeutic effect of ES-62 in HCD-fed mice. They found that a small weekly dose of ES-62 not only promoted healthy late-life in HCD-fed mice of both sexes but unexpectedly also extended by over 70 days the lifespan of male obese mice. According to the authors, the lifespan promoting effects of ES-62 on males could be explained by the fact that ES-62 appears to preferentially target the gut-metabolic tissue axis in male HCD-mice, acting to normalize gut microbiome changes and loss of intestinal barrier integrity.

Another example is provided by the parasitic rat tapeworm, *Hymenolepis diminuta*, reported by Hurd et al. to significantly prolong the longevity of its intermediate host, the beetle *Tenebrio molitor* in spite of fitness costs imposed by the parasite on this host (Hurd et al., 2001). Within the haemocoel of *T. molitor*, *H. diminuta* matures but does not multiply: its life cycle is completed only when the beetle eventually gets predated by suitable mammalian hosts. Accordingly, *H. diminuta*-induced prolonged longevity in *T. molitor* may maximize the probability that the parasite is transmitted to a mammalian host and might therefore impart a fitness benefit to *H. diminuta*. This parasite-induced reduction of host ageing was proposed by the authors to act, in *T. molitor* females, through the induction of a host-derived signalling molecule that blocks resource allocation to reproduction. Within males, the mechanism remains elusive but an association was observed between infection by *H. diminuta* and a diminution of both sex pheromone production and response to female copulatory pheromone. Based on the same rationale, trematodes are known to prolong the lifespan of snails through castration (Sorensen and Minchella, 2001), which eventually leads to gigantism in these intermediate hosts and may enhance the probability of a successful transmission to birds, the definitive hosts of trematodes. A field study on the population dynamics of trematodes, snails and wading birds even revealed a correlation between bird abundance and trematode prevalence in snails (Smith, 2001). In light of these observations, parasite-induced manipulation of the optimum trade-off between reproduction and longevity in intermediate hosts may actually be a widespread evolutionary strategy, as the recruitment rates to definitive hosts likely creates a selective pressure for it.

To conclude, the class of ADs that our work introduces is demonstrably broad and diverse, from viruses to bacteria and even parasitic animals. It could also encompass: i) virophages, which replicate by hijacking viral factories from giant viruses and lead to the production of defective giant virus particles, which nonetheless effectively propagates the virophages (La Scola et al., 2008); ii) plasmids, which, in prokaryotes, are another very common kind of mobile genetic element, that

take advantage of their host cells for their own replication; iii) parasites, genetically equipped to manipulate the reproductive agenda of their preys, such as some *Wolbachia*; and iv) symbionts, from ancestral mitochondria (Dalle Pezze et al., 2014) to present day gut microbes, when hosts and symbionts have not fully aligned their evolutionary interests. The generality of the age-distorter hypothesis, and its broad phylogenetic scope regarding both the variety of manipulated hosts and that of ADs makes it timely to expand current models of cellular-/organismal ageing to take the impact of manipulative interspecific interactions into account.

4. Age-distorters effects on theoretically immortal or extremely long-lived organisms and on prokaryotes

Above, we argued that ageing in any organism with endogenous ageing pathways, typically grounded in trade-offs between reproduction and maintenance, be it a multicellular or an unicellular entity, could in principle be affected by interspecific interactions with some ADs. However, because theoretically immortal or extremely long-lived organisms have developed adaptations that counteract their own ageing, and have heavily invested in their soma, one may wonder whether these remarkable organisms are also possibly immunized against viral, parasitic or symbiotic age-distortion.

We can only speculate about this question, but there are some reasons, briefly summarized below, to assume that age-distortion induced by interspecific interactions may interfere less with the ageing of theoretically immortal or extremely long-lived organisms than with the ageing of other organisms.

First, some extremely long-lived taxa harbor remarkably effective immune systems (Johnson et al., 2019), which means that they are adapted to counteract the deleterious effects of pathogenic or of viral infections, including that conducted by potential ADs. For example, very old bowhead whales (*Balaena mysticetus*) do not exhibit obvious signs of pathology, and benefit from especially robust anti-cancer machinery (Caulin and Maley, 2011). The genome of this species presents evidence of genetic adaptations in genes related with ageing, cancer, cell-cycle, and DNA repair (Johnson et al., 2019). Similarly, the naked mole rat (*Heterocephalus glaber*), is known for its remarkable healthspan, and negligible senescence, characterized by a strong cancer-resistance, a high fecundity until death, and no age-related increase in mortality. These animals have unique adaptations in genes involved in cancer resistance, DNA repair, DNA replication, and telomerase function, with evidence of positive selection for genes associated with autophagy, inflammation and cellular respiration, notably (Johnson et al., 2019). Likewise, the Greenland shark (*Somniosus microcephalus*), the longest-lived vertebrate, harbors a remarkably robust immune system (Johnson et al., 2019). Similarly, yet making an original optimal use of their immune system, long-lived bats, which are highly exposed to viruses due to their lifestyle, have notably evolved multiple ways to co-exist with viruses rather than to fight them (Laing et al., 2019), including enhanced autophagic activity with advancing age (Gorbunova et al., 2020; Huang et al., 2019).

Second, and non-exclusively, some extremely long-lived taxa, such as the slowly ageing giant tortoise species *Chelonoidis abingdonii* and *Aldabrachelys gigantea* (Quesada et al., 2018), the bowhead whale (Keane et al., 2015) or the naked mole rat, display unique investments in repair and maintenance mechanisms, which should reduce the impact of ADs. Similarly, planarian flatworms, such as *Schmidtea mediterranea*, are well known for their considerable regenerative capacity, supported by a pluripotent neoblast stem cell population, which make up approximately 25–30 % of all the cells in their body (Sahu et al., 2017). Damaging effects of ADs on such hosts may simply be cancelled as new tissues replace infected tissues, or even as new individuals emerge from some healthy parts of infected hosts. The same conclusion would hold for the regenerative acoele worm *Hofstenia miamia* or for *Hydra vulgaris*, which constantly renews tissues in its body (Boehm et al., 2012; Johnson

et al., 2019) and may experience asymmetrical damage transmission during division (Reichard, 2017).

In addition, some extremely long-lived taxa own very efficient ways to limit the numbers of their damaged cells or of their senescent cells. For instance, the quahog clam *Arctica islandica* has evolved increased resistance to oxidative stress-induced cell death and to produce less mitochondrial H₂O₂ (Johnson et al., 2019), which may explain its longevity (although environmental factors appear nonetheless to play a critical role in this species survival (Reichard, 2017)). Likewise, *Myotis myotis* bats evolved multiple strategies to fight ROS-induced damages and chronic inflammation, including that induced by viruses (Huang et al., 2019). Systemic effects of potential ADs are less likely to impair the global ageing of such host organisms, in particular because the general balance of their damaged and senescent cells with respect to healthy cells may be under a stronger control, and possibly further away from a dramatic tipping points, than in most other host organisms.

Overall, extremely long-lived and theoretically immortal organisms appear especially good at fixing their own bodies, and/or at renewing their own damaged cells, and/or at protecting themselves against diverse damages, including infections, implementing strategies that should reduce the impact of ADs.

Conversely, we suspect that viral ADs can affect some unicellular organisms (bacteria, archae and protists). Indeed, replicative ageing has been demonstrated to affect at least some unicellular bacterial and eukaryotic species (such as *E. coli*, *C. crescentus*, *S. cerevisiae*, etc.) (Stephens, 2005), and based on various lines of reasoning outside the scope of the present paper, replicative ageing and physiological ageing (the decline of biological functions with chronological time) have also been proposed to affect many more (possibly all) unicellular organisms, especially those with asymmetric cellular division (Ackermann, 2003; Kirkwood, 2005; Proenca et al., 2018; Teulière et al., 2020). Therefore, nothing opposes the fact that viruses may manipulate, when they exist, ageing of unicellular hosts for their own benefit. This proposition finds some indirect support in two different kinds of observations. First, viral manipulation of the cell cycles of unicellular taxa hijacking their host prokaryotic bodies has very recently been demonstrated to lead to a viral induction of asymmetrical cell division in an archaeon (Liu et al., 2021), and analogously, viral induction of cellular enlargement by interferences with cell division has also been demonstrated in *E. coli* (Ballesteros-Plaza et al., 2013). Thus, to enhance their own replication, some viruses interfere with cellular division and/or host DNA replication, two processes suspected to be associated with prokaryotic reproduction and rejuvenation (Teulière et al., 2020), with the consequence of reducing their enlarged host cells viability, whose energetic resources are exploited by viruses (Ballesteros-Plaza et al., 2013; Liu et al., 2021). Second, bacterial genes encoding transcription factors involved in a form of molecular trade-off between cellular reproduction and maintenance in *E. coli* K12 (Nyström, 2002) have homologs in viruses from closely related strains/species, which is consistent with a possible interference with this trade-off. Namely, sigma 70 gears the fate of the *E. coli* K12 cell towards reproduction, whereas sigma s gears the fate of *E. coli* K12 cell towards maintenance. When sigma 70 prevails over sigma s, *E. coli* K12 cells demonstrably age physiologically (Nyström, 2002). Homology search by tBLASTn against the NCBI database shows that highly conserved homologs of sigma 70/tpod, the pro-ageing transcription factor of *E. coli* K12, are notably present in two prophages of *Klebsiella* (*Klebsiella* phage ST147-VIM1 phi7.1 and *Klebsiella* phage ST13-OXA48phi12.1, with 100 % coverage, 92 % ID, E-value = 0 to rpod). This *Enterobacteriaceae* host is phylogenetically closely related to *E. coli*, so if *Klebsiella* uses the same cellular reproduction/maintenance trade-off as *E. coli* K12, its phages-encoded rpod may make *Klebsiella* age faster. Future research appears needed to test whether, as we surmise, viral AD may affect unicellular organisms in addition to multicellular organisms.

5. Research perspectives encouraged by the concept of age-distorters

Based on the existence of genetically-encoded trade-offs between organismal or cellular reproduction and maintenance and ageing pathways in host organisms, the age-distorter hypothesis predicts:

- (1) that novel ADs can be discovered, harboring genes that allow them to interfere with the host ageing pathways, in particular those involved in maintenance/reproduction trade-offs;
- (2) that it must be evolutionarily advantageous for an AD to directly manipulate the ageing of its host or of some of its infected cells, even if this manipulation is detrimental for its hosts (i.e. results in accelerated host ageing with respect to naive hosts).

Testing these two predictions of the age-distorter hypothesis would require experimental support that could be accrued by repurposing already established experimental systems, notably *in vitro*, in order to screen for new ADs, i.e. biological entities interfering with the ageing process of a host. For instance, as reviewed by Seoane et al., cultured human cells can be infected by several viruses that lead to, or inhibit, cellular senescence (Seoane et al., 2020). This phenotype is revealed by the analysis of specific cellular senescence biomarkers, including a persistent DNA damage response, a stable cell cycle arrest with induction of specific molecular regulators, characteristic morphological changes, a SASP and increased levels of senescence-associated beta-galactosidase (SA-βgal) activity at pH 6.0. To test prediction (1), namely that it is possible to identify new ADs, whether cellular senescence is activated or delayed could be systematically assayed for any virus able to infect human cultured cells, in comparison to non-infected cells, using a straightforward SA-βgal assay as screening readout, then cellular senescence phenotypes induced by viral candidates could be confirmed using the complete array of biomarkers.

Beyond *in vitro* screening, existing animal models could be used to discover new ADs and examine the effect of age-distortion on both host and AD fitness. As mentioned above, *Drosophila melanogaster*, one of the best studied models of ageing, has been shown to be manipulated by many insect-specific viruses through the expression of viral homologs of the secreted cytokine Dieldel (Lamiabile et al., 2016). Extending this model by monitoring stress response, lifespan, age at reproduction and fecundity in the fly model infected with candidate viral and microbial ADs, in comparison with different controls (such as non-infected/axenic controls and animals carrying ADs with inactivated candidate hijacking genes, homologous to host genes) could reveal possible AD-dependent lifespan changes in hosts. Such an assay could be used to infer AD-mediated interference on a reproduction/maintenance trade-off; for instance, hosts infected by lifespan-shortening ADs would also display weakened maintenance mechanisms (i.e. impaired DNA damage response activation or altered immune response, premature activation of apoptosis leading to organ failure) coupled to accelerated development and increased early fecundity, as described for gut microbiota-induced changes on *D. melanogaster*'s life history traits (Gould et al., 2018). A promising vertebrate experimental model for ageing studies, the African turquoise killifish (*Nothobranchius furzeri*) has also been shown to be susceptible to age-manipulation by its gut microbiota (Smith et al., 2017) and future experiments using this model could similarly uncover ADs acting on reproduction/maintenance life history trade-offs. Although more expensive, this approach may be used on mammals as well. High calorie diet mice, an experimental model of accelerated ageing, have allowed the discovery of a lifespan-extending, anti-inflammation helminthic peptide (Crowe et al., 2020). Using this model, one could screen for the ability of either infectious agents or other commensal organisms from the microbiota to alter the host's ageing rate. As showed in Crowe et al., this can be achieved by monitoring and comparing pathophysiological (including inflammatory), metabolic and microbiome parameters of ageing, or even using an epigenetic clock (Crowe et al., 2020; Thompson et al., 2018) between candidate AD-associated animals and AD-free controls.

Reproduction/maintenance trade-offs have been observed in mice subjected to lifespan-altering experimentation, including macronutrient intake manipulation (Solon-Biet et al., 2015), growth hormone signaling impairment (Bartke et al., 2013) or mating in early life independently of fertilization (Garratt et al., 2020). With the caveats that an alteration of these trade-offs by AD infection would depend on the mouse strain genetic background, interference would in principle be detectable by monitoring parameters such as age at sexual maturity, fertility, body weight, or susceptibility to cancer, in control and infected animals.

Once a new AD is identified, its classification into type 1 or 2 would have to be determined by probing the mechanisms leading to age-distortion in these systems. Type 1 AD could first be suspected on the basis of the presence of homologs of longevity-associated genes (LAGs) in their genomes, for instance the genes listed in the GenAge database (Tacutu et al., 2018). Type 1 AD could also be associated with a differential expression of LAGs, conditional on AD infection, and detected using comparative transcriptomic analyses of infected and control host cells. The functions of candidate age-distorting genes could also be tested by genetically engineering hosts, in which specific AD genes expression could be downregulated by targeted CRISPR or RNAi transgenes, and by comparing the ageing rate associated with AD infection in modified and non-modified hosts. Genetic manipulation of pairs of host and associated symbionts/pathogens, in particular using genome editing tools such as CRISPR/Cas9 technology, has already been performed on mouse models of bacterial and viral infection, including human-specific viruses in humanized transgenic mice (Masemann et al., 2020; Sarkar and Heise, 2019; Takaki et al., 2017) as well as on model animals and their microbiota (Hsu et al., 2020; Lee et al., 2019; Ramachandran and Bikard, 2019; Sun et al., 2020). More generally, hijacking of host pathways, typical of a type 1 AD, could be investigated not only by tracking the localization and the interaction partners of proteins encoded by homologs of LAG genes in the candidate AD genome upon expression in the host cells, using tagging, immunodetection or biochemical methods such as immunoprecipitation, but also by transgenic expression of these candidate hijacker genes in uninfected cells or animals, in an attempt to mimic the corresponding AD-induced pro- or anti-senescence phenotype.

To test prediction (2), it should be demonstrated that age-distortion increases the fitness of the AD, defined as its replicative success relative to competing organisms in a particular environment, *in vivo* or *in vitro*. In the case of viruses interfering with cellular senescence, the effect of age distortion on competitive viral fitness could be assayed using already developed relevant methods for the considered virus (reviewed in (Domingo et al., 2019; Wargo and Kurath, 2012)), in comparison to control viruses genetically inactivated for the candidate age-distorting genes (using RNAi or genome editing techniques), in control cells or in cells where cellular senescence can be induced (e.g. by oxidative stress, (Chen et al., 2007)) or eliminated (e.g. using senolytics that induce apoptosis of senescent cells, (von Kobbe, 2019)). In animal models, AD within-host relative fitness comparison could be performed similarly (methods reviewed in (Hughes and Andersson, 2015)). Besides these experimental approaches to uncover novel evidence for ADs, computational and mathematical modelling (Johnson et al., 2019) could be used to predict and quantify the effects of ADs on host ageing. Typically, ADs could be explicitly integrated into models that explore what drivers contribute to the accumulation of senescent cells with chronological age (Kowald et al., 2020).

Finally, the AD hypothesis brings about a third prediction, with potential for improving human health during ageing. Namely, both AD interfering products and their interference host targets could illuminate the underlying physiological mechanisms that regulate ageing, as proposed for viruses by (Mliĥ et al., 2018).

This prediction can be tested using comparative transcriptomics and gene set module analysis of human tissues pre- and post-infection, which could reveal AD-induced host gene expression changes of known ageing

genes. This new knowledge could then lead to novel therapeutic avenues to lengthen human healthspan, offering new strategies to counter inflamm-ageing and immunosenescence consequences in ageing humans, including cancer. Uncovering known viruses or bacteria with pro-ageing properties in humans would also promote the conversion of already existing vaccines or antibiotics into potential anti-ageing therapeutic tools. In addition, the knowledge of the molecular mechanisms at play may allow clinicians repurposing existing drugs as pharmacological inhibitors of these ADs. Finally, ADs themselves could be used as therapeutic tools with a potentially tissue-specific effect on ageing. Indeed, as pointed out by (Seoane et al., 2020), the ability of some viruses to exploit pathways involved in the induction of cellular senescence to improve their replication raises the opportunity to use them as pro- or anti-senescence agents (Seoane et al., 2020). These tools could improve human healthspan by, for instance, reducing cellular senescence-induced inflamm-ageing in the elderly.

In therapies aimed at cancers however, targeting cellular senescence is considered a double edged-sword (Schosserer et al., 2017): although senescence induction in tumor cells is beneficial to inhibit their replication, induction of senescence by current chemotherapeutic treatments in other cells creates an inflammatory, pro-tumorigenic environment (Wang et al., 2020a). In this context, senescence-manipulating viruses could be used as modulators of cellular senescence in combination with chemotherapy, to help tip the balance towards a better cancer prognosis. Uncovering ADs with similar properties in the human intestinal microbiota, in particular in the intestine, would allow relatively straightforward intervention by Fecal Microbiota Transplant (FMT) in order to manipulate cellular senescence in this tissue. Further, by screening for cellular senescence modifiers as proposed above, ADs that would trigger programmed cell death of senescent cells, i.e. behaving like natural senolytics (Ge et al., 2021a), may also be recovered and such properties could be evaluated for use in anti-ageing and anti-cancer therapies.

6. Updating classic evolutionary theories of ageing to account for age-distorters

As trade-offs between (cellular) reproduction and maintenance have likely evolved in many taxa (Austad and Hoffman, 2018; Baudisch and Vaupel, 2012; Kirkwood and Holliday, 1979; Williams, 1957), direct selection biasing these trade-offs towards host (or host cell) reproduction, directly mediated by ADs, could be a new significant cause of ageing. Such effects that aggravate ageing may be very common, because viroplages (the virus of giant viruses), viruses, plasmids, parasites, and symbionts able to manipulate the reproductive agenda or the ageing pathways of their hosts are common in nature. Moreover, both the phylogenetic diversity of possible ADs and of their hosts taxa can be extremely large, spanning over the entire Tree of Life. Thus, the age-distorting hypothesis introduces a new, general theoretical mechanism for host ageing, with an uncommonly broad phylogenetic scope, and potential to update current knowledge and evolutionary theories of ageing.

The idea of viruses as ADs is quite novel. Although the impact of latent viruses on ageing have been empirically questioned for some years (Nikolich-Zugich and van Lier, 2017; Redeker et al., 2018; Sansoni et al., 2014), and although the causal role of some viruses (and some bacteria) in neurological disorders such as Alzheimer's disease is becoming increasingly discussed (Allnutt et al., 2020; Fülöp et al., 2018; Lathe et al., 2019; Wouk et al., 2021), and although the relationship between viruses and immunosenescence, and between viruses and host ageing, have recently become recognized as more complex than previously thought (Hurme, 2019; Mliĥ et al., 2018), viruses have nonetheless so far received very limited consideration in evolutionary theories of ageing. Firstly, viruses themselves are not suspected to age much. They are surely sensitive to the accumulation of material damage in their genetic information and in their proteic components with time, and to

the accumulation of mutations in their genomes, as they replicate. However, as viruses are negatively defined by their lack of binary division (Lwoff, 1957), because viral particles propagate from neo-synthesized genetic and proteic materials, viruses should not be affected by replicative ageing, which typically relies on asymmetric cellular division coupled with the differential sorting of aged components between clones to evolve (Ackermann, 2003; Barker and Walmsley, 1999; Florea, 2017; Kirkwood, 2005; Partridge and Barton, 1993; Pernice et al., 2017; Stewart et al., 2005; Teulière et al., 2020; Wang et al., 2020b). Secondly, viruses have occasionally been mentioned in ageing theories but almost exclusively as stressors, contributing to the external selective pressures exerted on cellular organisms (Williams, 1957). By raising the risks of organismal death within a hostile environment, viruses (and in general parasites), to which biologically older individuals appear to be more sensitive than younger organisms (Garsin et al., 2003; Kim et al., 2001; Medvedev, 1990), may indirectly reinforce the evolution of organismal ageing within lineages. It was also proposed that viral infections might enhance chaotic population dynamics, which may theoretically lead a hypothetical non-ageing cellular population to extinction. In these latter models, viral infections would encourage the evolution of ageing as they would buffer the risk of chaotic extinction (Mitteldorf and Pepper, 2009) of a population. Consequently, bringing forward the phrase and concept of viral AD should further support the timely emerging inquiries on the impact of viruses of host ageing, which constitute a new frontier for ageing research (Humphreys et al., 2020; Hurme, 2019; Mlih et al., 2018). In particular, the concept of viral ADs may help shedding a new light on: i) the role of viruses in carcinogenesis (as some oncogenic viruses (Benedetti et al., 2021; Zamaraev et al., 2020) may prove to be ADs), ii) on apoptosis and on cellular senescence (as ADs may more commonly than appreciated manipulate both processes to evade host immune response to their benefit (Kaminsky and Zhivotovsky, 2010)), iii) on immunosenescence (as diminution or abrogation of the host immune response may be a common strategy in latent viral ADs (Gillet and Brun, 1996)), and iv) on inflamm-ageing (as ADs may commonly modulate host inflammation to create system conditions conducive to their own replication (Weitzman and Weitzman, 2014)).

Moreover, virally-induced cellular age-distortion opens another under appreciated research avenue: the study of microbial ageing in a context of interspecific interactions. Discovering whether and which microbes have their age distorted would be important for multiple reasons. In nature, a significant proportion of microbes may be ageing at a rate that is different from that of their conspecifics due to viral infection by ADs, meaning that microbial populations may comprise a higher proportion of age-distorted cells than expected by any current theory. As it is estimated that in the oceans viral infections remove 20%–40% of all bacterial cells each day (Microbiology by numbers, 2011), up to 40% of the microbial cells may have their biological age distorted with respect to their conspecifics. Demonstrating virally-induced bacterial ageing would unravel a new pressure acting on present and on past microbiomes, and thus likely to affect and to have affected host health and ecosystem dynamics. For example, dysbiosis may be reached at a faster pace in a gut microbiome invaded by ADs, with negative consequences on its host. Second, if cellular ageing mediated by interspecific interactions with AD was successfully traced back to bacteria, archaea, or protists, this manipulable mode of host ageing would appear evolutionarily very ancient. Like with the admitted case of the co-option of ancestral endogenous retroviruses that provided novel regulatory elements for interferon response (Chuong et al., 2016), it would then become a relevant empirical question to determine whether our species inherited components of a manipulable cellular ageing processes from its microbial ancestors, and if yes, which ones. In that sense, ADs affecting microbes may hold keys to unravel hidden aspects of the origins of our own cellular replicative ageing.

Critically, ADs do not yet feature in any of the mainstream evolutionary theories of ageing (MA, AP and DS), which were developed in

the classical framework of population genetics and are thus all logically centered on organisms from a single, focal species or population (Johnson et al., 2019). However, the concept of ADs is consistent with these theories. ADs may be heuristically connected to each of these important theories, and thus expand their theoretical application scope to explain the causes of ageing beyond what happens within populations of conspecific individuals.

First, the MA theory can easily be enhanced to encompass ADs and their mutations as an additional source of late life damages affecting an organism. This is because, at its core, MA considers the declining strength of natural selection as the main driver of organismal ageing. Endorsing this hypothesis entails that, just like the expression and accumulation of genes with negative effects in the host genome can be neither effectively fought against nor postponed by an organism's adaptive mechanisms (because natural selection's effect is looser at later stages of an organism's life), the expression and accumulation of ADs in a host organism cannot be effectively fought against past a given chronological age by comparable adaptive mechanisms. For instance, as the selection exerted by the host on its symbiotic communities (Barroso-Batista et al., 2015; Bevins and Salzman, 2011) declines with age, not only more pathobionts (Rampelli et al., 2013; van der Lugt et al., 2018), but also more ADs can make their way inside the host. Yet, under the MA hypothesis, as the effects of selection on the host and therefore the efficacy of adaptive defense mechanisms in hosts, exerting their own selection on the microbiome recedes with time, genetic mutations in microbiomes, which are sometimes metaphorically qualified of 'second or third organismal genome' (Garagnani et al., 2014), or even accumulations of taxa with damaging effects late in life (Brummel et al., 2004; Garsin et al., 2003; Kim et al., 2001; Rampelli et al., 2013; van der Lugt et al., 2018), would aggravate host ageing, just like mutations in this host genome would do. The same is true for the genomes of endosymbionts, which accumulate mutations over time (Bennett and Moran, 2015; Komaki and Ishikawa, 2000). Thus, ADs could be considered an age-related threat, which provide additional extra-genomic sources of damaging late-expressed genes. Of note, in that framework, the ageing genes carried by ADs would not need to be restricted to genes involved in reproduction/maintenance trade-offs.

ADs can also be connected to the AP theory, by considering that some ADs can play the role of extra-genomic sources of antagonistic pleiotropic genes. For instance, a given symbiotic taxon or virus, by training or by priming a host's immune system early in its life may have positive fitness effects for this host early on, yet negative fitness effects later, when this host-microbe interaction becomes exceedingly costly for the host. This was suggested in *Drosophila*, where the presence of specific bacterial strains during the first week of adult life enhances longevity, whereas their presence in the last stage of life caused a slight decrease (Brummel et al., 2004). This was also reported in mice, where CMV infections in young mice seem to improve immunity against other pathogens but not in aged mice (Redeker et al., 2018). Similarly, virally-induced senescence was proposed to play a positive role by recruiting neutrophils and other immune cells in young hosts via the SASP, resulting in viral clearance and tissue repair, whereas in aged or vulnerable hosts, senescence induced by viral infection would aggravate the senescence already present and tip the balance towards pathology (Kelley et al., 2020). Thus, both MA and AP theories, once the locus of ageing is expanded to genes outside the genome of the focal ageing organism, offer a well-developed framework to analyze the effects of ADs on organismal ageing.

Furthermore, the age-distorter hypothesis also supports and expands, by analogy, the logic of the DS, by casting as its core the concept of a trade-off between reproduction and maintenance, either at the cellular level or at the general host level, and the idea that a cell or a host can become the disposable body of another entity with an intersecting life-cycle and with conflicting evolutionary interest. A fundamental difference with the case of the soma and germline lineages within multicellular organisms is that, here, the role of the soma is expanded beyond the

conventional fixed germline/soma distinction within multicellular organisms. By analogy at least, the disposable soma theory may thus be expanded in an original way: when the life cycle of a manipulative entity intersects with the lifecycle of a manipulable entity, the former may use the latter as its own 'disposable soma' to favor its own reproductive success. When this is the case, cellular or host ageing in a focal species is directly under the genetic control of an external species, meaning that, although natural selection is generally not expected to favor ageing within a given species (because this would lower organismal fitness), natural selection will nonetheless directly promote ageing in manipulated members of a focal species by acting on an AD. Entities that age because of ADs then appear as collateral victims, neither because of a neglect of evolution (Medawar, 1952) nor due to antagonistic pleiotropy (Kirkwood and Holliday, 1979; Williams, 1957), but because these age-distorted entities share (part of) their lifecycles and of their bodies with another entity with different evolutionary interests. Importantly, the age-distorter hypothesis does not necessarily require that a host and its AD would collectively form an overarching emerging selective unit, i. e. a larger evolutionary individual on which natural selection would act (Garagnani et al., 2014) (as suggested, for instance, in a theoretical model, considering that microbiomes could be selected to contribute to their host demise after their reproductive life (Blaser and Webb, 2014)). To be effective, the age-distorter hypothesis only, and more parsimoniously, requires that the lifecycles of two non-kin entities with conflicting evolutionary interests and with gene sets interfering with organismal ageing/reproduction trade-offs intersect.

7. Conclusion

Because no organism is an island, it makes sense to reconceptualize the locus at which ageing is thought to occur (Johnson et al., 2019). Consistently, we propose that cellular ageing and even organismal ageing could be directly selected for, but for the good of another, non kin entity. The age-distorter hypothesis proposes that when lifecycles of manipulative taxa intersect with lifecycles of manipulable taxa, for example during viral infection, manipulative taxa may contribute to the ageing of their hosts. Thus, unexpectedly, some instances of organismal ageing could be the direct result of natural selection, not because ageing would benefit to the ageing species (an adaptive explanation of ageing, which is rejected by all mainstream evolutionary theories on ageing (Kirkwood and Holliday, 1979; Medawar, 1952; Williams, 1957)), but rather because cellular or organismal ageing would benefit to non-kin: individuals from another species (e.g. viruses, parasites or symbionts) that intersect and manipulate the lifecycles of ageing host organisms and use these host organisms or at least some of their cells as a disposable soma for their own evolutionary interest. For AD-type 1, this contribution to host ageing would not be due to a stress/enhanced damage that is induced by the manipulative taxa, but to a genetic impetus to gear the trade-off between maintenance and reproduction towards reproduction at the expense of the preservation of the cell or of organismal physiological youth. We provided arguments and evidence that, when it is advantageous for viral or non-viral ADs to dispose of more cheaply-built hosts cells or of host organisms in which to replicate or to hide, ADs may benefit from manipulating host ageing (Humphreys et al., 2020). We argued that acknowledging ADs has a significant potential to strengthen the role attributed to viruses in ageing, from bacteria to humans, to prompt new experiments questioning the impact of AD on host health and ageing, and to expand the scope of the three major evolutionary theories of ageing. Owing to the impressive number of viruses, parasites and symbionts on Earth, future analyses are needed to determine whether cellular or organismal ageing driven from outside, as ADs use the cells or body of other species as their own disposable soma, is more than a hypothesis and a rare clade-specific event, but possibly a significant phenomenon at the scale of the planet.

Acknowledgements

The authors thank M. Habib, C. Vigliotti, F. Papale, P. Lopez and A. Watson and three anonymous reviewers for critical comments. 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. J. Teulière was supported by the over-heads of the EVOLUNET grant, European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement # 615274, category LS8) to E. Bapteste.

References

- Ackermann, M., 2003. Senescence in a bacterium with asymmetric division. *Science* 300. <https://doi.org/10.1126/science.1083532>, 1920–1920.
- Aiello, A.E., Chiu, Y.-L., Frasca, D., 2017. How does cytomegalovirus factor into diseases of aging and vaccine responses, and by what mechanisms? *GeroScience* 39, 261–271. <https://doi.org/10.1007/s11357-017-9983-9>.
- Allnutt, M.A., Johnson, K., Bennett, D.A., Connor, S.M., Troncoso, J.C., Pletnikova, O., Albert, M.S., Resnick, S.M., Scholz, S.W., De Jager, P.L., Jacobson, S., 2020. Human herpesvirus 6 detection in Alzheimer's disease cases and controls across multiple cohorts. *Neuron* 105, 1027–1035. <https://doi.org/10.1016/j.neuron.2019.12.031> e2.
- Alperovitch-Lavy, A., Sharon, I., Rohwer, F., Aro, E.-M., Glaser, F., Milo, R., Nelson, N., Beja, O., 2011. Reconstructing a puzzle: existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. *Environ. Microbiol.* 13, 24–32. <https://doi.org/10.1111/j.1462-2920.2010.02304.x>.
- Anttila, T., Saikku, P., Koskela, P., Bloigu, A., Dillner, J., Ikäheimo, I., Jellum, E., Lehtinen, M., Lenner, P., Hakulinen, T., Näränen, A., Pukkala, E., Thoresen, S., Youngman, L., Paavonen, J., 2001. Serotypes of *Chlamydia trachomatis* and risk for development of cervical squamous cell carcinoma. *JAMA* 285, 47–51. <https://doi.org/10.1001/jama.285.1.47>.
- Austad, S.N., Hoffman, J.M., 2018. Is antagonistic pleiotropy ubiquitous in aging biology? *Evol. Med. Public Health* 2018, 287–294. <https://doi.org/10.1093/emph/eoy033>.
- Avelar, R.A., Ortega, J.G., Tacutu, R., Tyler, E.J., Bennett, D., Binetti, P., Budovsky, A., Chatsirisupachai, K., Johnson, E., Murray, A., Shields, S., Tejada-Martinez, D., Thornton, D., Fraifeld, V.E., Bishop, C.L., de Magalhães, J.P., 2020. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol.* 21, 91. <https://doi.org/10.1186/s13059-020-01990-9>.
- Ballesteros-Plaza, D., Holguera, L., Scheffers, D.-J., Salas, M., Muñoz-Espín, D., 2013. Phage 29 phi protein p1 promotes replication by associating with the FtsZ ring of the divisome in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12313–12318. <https://doi.org/10.1073/pnas.1311524110>.
- Barker, M.G., Walmsley, R.M., 1999. Replicative ageing in the fission yeast *Schizosaccharomyces pombe*. *Yeast* Chichester Engl. 15, 1511–1518. [https://doi.org/10.1002/\(SICI\)1097-0061\(199910\)15:14<1511::AID-YEA482>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0061(199910)15:14<1511::AID-YEA482>3.0.CO;2-Y).
- Barroso-Batista, J., Demengeot, J., Gordo, I., 2015. Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. *Nat. Commun.* 6, 8945. <https://doi.org/10.1038/ncomms9945>.
- Bartke, A., Sun, L.Y., Longo, V., 2013. Somatotropic signaling: trade-offs between growth, reproductive development, and longevity. *Physiol. Rev.* 93, 571–598. <https://doi.org/10.1152/physrev.00006.2012>.
- Baudisch, A., Vaupel, J.W., 2012. Getting to the root of aging. *Science* 338, 618–619. <https://doi.org/10.1126/science.1226467>.
- Bell, R., Hubbard, A., Chettier, R., Chen, D., Miller, J.P., Kapahi, P., Tarnopolsky, M., Sahasrabudhe, S., Melov, S., Hughes, R.E., 2009. A human protein interaction network shows conservation of aging processes between human and invertebrate species. *PLoS Genet.* 5, e1000414. <https://doi.org/10.1371/journal.pgen.1000414>.
- Benedetti, F., Curreli, S., Gallo, R.C., Zella, D., 2021. Tampering of viruses and bacteria with host DNA repair: implications for cellular transformation. *Cancers* 13. <https://doi.org/10.3390/cancers13020241>.
- Bennett, G.M., Moran, N.A., 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci.* 112, 10169–10176. <https://doi.org/10.1073/pnas.1421388112>.
- Bevins, C.L., Salzman, N.H., 2011. The potter's wheel: the host's role in sculpting its microbiota. *Cell. Mol. Life Sci.* 68, 3675–3685. <https://doi.org/10.1007/s00018-011-0830-3>.
- Biagi, E., Candela, M., Franceschi, C., Brigidi, P., 2011. The aging gut microbiota: new perspectives. *Ageing Res. Rev.* 10, 428–429. <https://doi.org/10.1016/j.arr.2011.03.004>.
- Blaser, M.J., Webb, G.F., 2014. Host demise as a beneficial function of indigenous microbiota in human hosts. *mBio* 5, e02262–14. <https://doi.org/10.1128/mbio.02262-14>.
- Blazkova, H., Krejčíková, K., Moudry, P., Frisan, T., Hodný, Z., Bartek, J., 2010. Bacterial intoxication evokes cellular senescence with persistent DNA damage and cytokine signalling. *J. Cell. Mol. Med.* 14, 357–367. <https://doi.org/10.1111/j.1582-4934.2009.00862.x>.
- Boehm, A.-M., Khalturin, K., Anton-Erxleben, F., Hemmrich, G., Klostermeier, U.C., Lopez-Quintero, J.A., Oberg, H.-H., Puchert, M., Rosenstiel, P., Wittlieb, J., Bosch, T.

- C.G., 2012. FoxO is a critical regulator of stem cell maintenance in immortal Hydra. *Proc. Natl. Acad. Sci. U. S. A* 109, 19697–19702. <https://doi.org/10.1073/pnas.1209714109>.
- Bourke, A.F.G., 2007. Kin selection and the evolutionary theory of aging. *Annu. Rev. Ecol. Evol. Syst.* 38, 103–128. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095528>.
- Brummel, T., Ching, A., Seroude, L., Simon, A.F., Benzer, S., 2004. Drosophila lifespan enhancement by exogenous bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12974–12979. <https://doi.org/10.1073/pnas.0405207101>.
- Budovsky, A., Abramovich, A., Cohen, R., Chalifa-Caspi, V., Fraifeld, V., 2007. Longevity network: construction and implications. *Mech. Ageing Dev.* 128, 117–124. <https://doi.org/10.1016/j.mad.2006.11.018>.
- Caulin, A.F., Maley, C.C., 2011. Peto's Paradox: evolution's prescription for cancer prevention. *Trends Ecol. Evol.* 26, 175–182. <https://doi.org/10.1016/j.tree.2011.01.002>.
- Chen, J.-H., Ozanne, S.E., Hales, C.N., 2007. Methods of cellular senescence induction using oxidative stress. In: Tollefsbol, T.O. (Ed.), *Biological Aging, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 179–189. https://doi.org/10.1007/978-1-59745-361-5_14.
- Choi, Y., Bowman, J.W., Jung, J.U., 2018. Autophagy during viral infection - a double-edged sword. *Nat. Rev. Microbiol.* 16, 341–354. <https://doi.org/10.1038/s41579-018-0003-6>.
- Chow, J., Lee, S.M., Shen, Y., Khosravi, A., Mazmanian, S.K., 2010. Chapter 8 - host-bacterial symbiosis in health and disease. In: Fagarasan, S., Cerutti, A. (Eds.), *Advances in Immunology, Mucosal Immunity*. Academic Press, pp. 243–274. <https://doi.org/10.1016/B978-0-12-381300-8.00008-3>.
- Chuong, E.B., Elde, N.C., Feschotte, C., 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351, 1083–1087. <https://doi.org/10.1126/science.1254977>.
- Cobián Güemes, A.G., Youle, M., Cantú, V.A., Felts, B., Nulton, J., Rohwer, F., 2016. Viruses as winners in the game of life. *Annu. Rev. Virol.* 3, 197–214. <https://doi.org/10.1146/annurev-virology-100114-054952>.
- Cohen, J., Torres, C., 2017. HIV-associated cellular senescence: a contributor to accelerated aging. *Ageing Res. Rev.* 36, 117–124. <https://doi.org/10.1016/j.arr.2016.12.004>.
- Connolly, P.F., Fearnhead, H.O., 2017. Viral hijacking of host caspases: an emerging category of pathogen–host interactions. *Cell Death Differ.* 24, 1401–1410. <https://doi.org/10.1038/cdd.2017.59>.
- Coutinho, F.H., Cabello-Yeves, P.J., Gonzalez-Serrano, R., Rosselli, R., López-Pérez, M., Zemskaya, T.I., Zakharenko, A.S., Ivanov, V.G., Rodriguez-Valera, F., 2020. New viral biochemicochemical roles revealed through metagenomic analysis of Lake Baikal. *Microbiome* 8, 163. <https://doi.org/10.1186/s40168-020-00936-4>.
- Crowe, J., Lumb, F.E., Doonan, J., Broussard, M., Tarafdar, A., Pineda, M.A., Landabaso, C., Mulvey, L., Hoskisson, P.A., Babayan, S.A., Selman, C., Harnett, W., Harnett, M.M., 2020. The parasitic worm product ES-62 promotes health- and lifespan in a high calorie diet-accelerated mouse model of ageing. *PLoS Pathog.* 16, e1008391. <https://doi.org/10.1371/journal.ppat.1008391>.
- Dalle Pezze, P., Nelson, G., Otten, E.G., Korolchuk, V.I., Kirkwood, T.B.L., von Zglinicki, T., Shanley, D.P., 2014. Dynamic modelling of pathways to cellular senescence reveals strategies for targeted interventions. *PLoS Comput. Biol.* 10, e1003728. <https://doi.org/10.1371/journal.pcbi.1003728>.
- De Smet, J., Hendrix, H., Blasdel, B.G., Danis-Wlodarczyk, K., Lavigne, R., 2017. Pseudomonas predators: understanding and exploiting phage-host interactions. *Nat. Rev. Microbiol.* 15, 517–530. <https://doi.org/10.1038/nrmicro.2017.61>.
- Domingo, E., de Ávila, A.I., Gallego, I., Sheldon, J., Perales, C., 2019. Viral fitness: history and relevance for viral pathogenesis and antiviral interventions. *Pathog. Dis.* 77. <https://doi.org/10.1093/femspd/ftz021>.
- Donato, V., Ayala, F.R., Cogliati, S., Bauman, C., Costa, J.G., Leinini, C., Grau, R., 2017. Bacillus subtilis biofilm extends Caenorhabditis elegans longevity through downregulation of the insulin-like signalling pathway. *Nat. Commun.* 8. <https://doi.org/10.1038/ncomms14332>.
- Fast, E.M., Toomey, M.E., Panaram, K., Desjardins, D., Kolaczky, E.D., Frydman, H.M., 2011. Wolbachia enhance Drosophila stem cell proliferation and target the germline stem cell niche. *Science* 334, 990–992. <https://doi.org/10.1126/science.1209609>.
- Feng, X., Feng, G., Xing, J., Shen, B., Tan, W., Huang, D., Lu, X., Tao, T., Zhang, J., Li, L., Gu, Z., 2014. Repeated lipopolysaccharide stimulation promotes cellular senescence in human dental pulp stem cells (DPSCs). *Cell Tissue Res.* 356, 369–380. <https://doi.org/10.1007/s00441-014-1799-7>.
- Feng, G., Zheng, K., Cao, T., Zhang, J., Lian, M., Huang, D., Wei, C., Gu, Z., Feng, X., 2018. Repeated stimulation by LPS promotes the senescence of DPSCs via TLR4/MyD88-NF- κ B-p53/p21 signaling. *Cytotechnology* 70, 1023–1035. <https://doi.org/10.1007/s10616-017-0180-6>.
- Fernandes, M., Wan, C., Tacutu, R., Barardo, D., Rajput, A., Wang, J., Thoppil, H., Thornton, D., Yang, C., Freitas, A., de Magalhães, J.P., 2016. Systematic analysis of the gerontome reveals links between aging and age-related diseases. *Hum. Mol. Genet.* <https://doi.org/10.1093/hmg/ddw307>.
- Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* 13, 283–296. <https://doi.org/10.1038/nrg3199>.
- Fettucciari, K., Macchioni, L., Davidescu, M., Scarpelli, P., Palumbo, C., Corazzi, L., Marchegiani, A., Cerquetella, M., Spaterna, A., Marconi, P., Bassotti, G., 2018. Clostridium difficile toxin B induces senescence in enteric glial cells: a potential new mechanism of Clostridium difficile pathogenesis. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1865, 1945–1958. <https://doi.org/10.1016/j.bbamcr.2018.10.007>.
- Florea, M., 2017. Aging and immortality in unicellular species. *Mech. Ageing Dev.* 167, 5–15. <https://doi.org/10.1016/j.mad.2017.08.006>.
- Forterre, P., 2013. The virocell concept and environmental microbiology. *ISME J.* 7, 233–236. <https://doi.org/10.1038/ismej.2012.110>.
- Forterre, P., 2016. To be or not to be alive: how recent discoveries challenge the traditional definitions of viruses and life. *Stud. Hist. Philos. Biomed. Sci.* 59, 100–108. <https://doi.org/10.1016/j.shpsc.2016.02.013>.
- Frasca, D., Blomberg, B.B., 2016. Inflammaging decreases adaptive and innate immune responses in mice and humans. *Biogerontology* 17, 7–19. <https://doi.org/10.1007/s10522-015-9578-8>.
- Frost, L.S., Leplae, R., Summers, A.O., Toussaint, A., 2005. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732. <https://doi.org/10.1038/nrmicro1235>.
- Fry, A.J., Rand, D.M., 2002. Wolbachia interactions that determine Drosophila melanogaster survival. *Evolution* 56, 1976–1981. <https://doi.org/10.1111/j.0014-3820.2002.tb00123.x>.
- Fülöp, T., Itzhaki, R.F., Balin, B.J., Miklossy, J., Barron, A.E., 2018. Role of microbes in the development of Alzheimer's disease: state of the art – an international symposium presented at the 2017 IAGG congress in San Francisco. *Front. Genet.* 9. <https://doi.org/10.3389/fgene.2018.00362>.
- Garagnani, P., Pirazzini, C., Giuliani, C., Candela, M., Brigidi, P., Sevinci, F., Luiselli, D., Bacalini, M.G., Salvioli, S., Capri, M., Monti, D., Mari, D., Collino, S., Delledonne, M., Descombes, P., Franceschi, C., 2014. The three genetics (nuclear DNA, mitochondrial DNA, and gut microbiome) of longevity in humans considered as metaorganisms. *Biomed. Res. Int.* 2014, 560340. <https://doi.org/10.1155/2014/560340>.
- Garratt, M., Try, H., Smiley, K.O., Grattan, D.R., Brooks, R.C., 2020. Mating in the absence of fertilization promotes a growth-reproduction versus lifespan trade-off in female mice. *Proc. Natl. Acad. Sci.* 117, 15748–15754. <https://doi.org/10.1073/pnas.2003159117>.
- Garsin, D.A., Villanueva, J.M., Begun, J., Kim, D.H., Sifri, C.D., Calderwood, S.B., Ruvkun, G., Ausubel, F.M., 2003. Long-lived C. elegans daf-2 mutants are resistant to bacterial pathogens. *Science* 300, 1921. <https://doi.org/10.1126/science.1080147>.
- Ge, M., Hu, L., Ao, H., Zi, M., Kong, Q., He, Y., 2021a. Senolytic targets and new strategies for clearing senescent cells. *Mech. Ageing Dev.* 195, 111468. <https://doi.org/10.1016/j.mad.2021.111468>.
- Ge, Y., Wang, X., Guo, Y., Yan, Junting, Abuduwaili, A., Aximujiang, K., Yan, Jie, Wu, M., 2021b. Gut microbiota influence tumor development and alter interactions with the human immune system. *J. Exp. Clin. Cancer Res.* 40. <https://doi.org/10.1186/s13046-021-01845-6>.
- Gillet, G., Brun, G., 1996. Viral inhibition of apoptosis. *Trends Microbiol.* 4, 312–317. [https://doi.org/10.1016/0966-842x\(96\)10047-0](https://doi.org/10.1016/0966-842x(96)10047-0).
- Gorbunova, V., Seluanov, A., Kennedy, B.K., 2020. The world goes bats: living longer and tolerating viruses. *Cell Metab.* 32, 31–43. <https://doi.org/10.1016/j.cmet.2020.06.013>.
- Gould, A.L., Zhang, V., Lamberti, L., Jones, E.W., Obadia, B., Korasidis, N., Gavyrushkin, A., Carlson, J.M., Beerwinkel, N., Ludington, W.B., 2018. Microbiome interactions shape host fitness. *Proc. Natl. Acad. Sci.* 115, E11951–E11960. <https://doi.org/10.1073/pnas.1809349115>.
- Gusarov, I., Gautier, L., Smolentseva, O., Shamovskiy, I., Eremina, S., Mironov, A., Nudler, E., 2013. Bacterial nitric oxide extends the lifespan of C. elegans. *Cell* 152, 818–830. <https://doi.org/10.1016/j.cell.2012.12.043>.
- Hamilton, W.D., 1966. The moulding of senescence by natural selection. *J. Theor. Biol.* 12, 12–45. [https://doi.org/10.1016/0022-5193\(66\)90184-6](https://doi.org/10.1016/0022-5193(66)90184-6).
- Han, B., Sivaramakrishnan, P., Lin, C.-C.J., Neve, I.A.A., He, J., Tay, L.W.R., Sowa, J.N., Sizovs, A., Du, G., Wang, J., Herman, C., Wang, M.C., 2017. Microbial genetic composition tunes host longevity. *Cell* 169, 1249–1262. <https://doi.org/10.1016/j.cell.2017.05.036>.
- Harnett, W., 2014. Secretory products of helminth parasites as immunomodulators. *Mol. Biochem. Parasitol.* 195, 130–136. <https://doi.org/10.1016/j.molbiopara.2014.03.007>. Special Issue on the 35th Anniversary of Molecular and Biochemical Parasitology.
- Heintz, C., Mair, W., 2014. You are what you host: microbiome modulation of the aging process. *Cell* 156, 408–411. <https://doi.org/10.1016/j.cell.2014.01.025>.
- Hoover, K., Grove, M., Gardner, M., Hughes, D.P., McNeil, J., Slavicek, J., 2011. A gene for an extended phenotype. *Science* 1401. <https://doi.org/10.1126/science.1209199>.
- Hsu, B.B., Plant, I.N., Lyon, L., Anastassacos, F.M., Way, J.C., Silver, P.A., 2020. In situ reprogramming of gut bacteria by oral delivery. *Nat. Commun.* 11, 5030. <https://doi.org/10.1038/s41467-020-18614-2>.
- Huang, Z., Whelan, C.V., Foley, N.M., Jebb, D., Touzalin, F., Petit, E.J., Puechmaile, S.J., Teeling, E.C., 2019. Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats. *Nat. Ecol. Evol.* 3, 1110–1120. <https://doi.org/10.1038/s41559-019-0913-3>.
- Hughes, D., Andersson, D.I., 2015. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nat. Rev. Genet.* 16, 459–471. <https://doi.org/10.1038/nrg3922>.
- Humphreys, D., ElGhazaly, M., Frisan, T., 2020. Senescence and host–pathogen interactions. *Cells* 9, 1747. <https://doi.org/10.3390/cells9071747>.
- Hurd, H., Warr, E., Polwart, A., 2001. A parasite that increases host lifespan. *Proc. R. Soc. Lond. B Biol. Sci.* <https://doi.org/10.1098/rspb.2001.1729>.
- Hurme, M., 2019. Viruses and immunosenescence - more players in the game. *Immun. Ageing A* 16, 13. <https://doi.org/10.1186/s12979-019-0152-0>.
- Ibler, A.E.M., ElGhazaly, M., Naylor, K.L., Bulgakova, N.A., F. El-Khamisy, S., Humphreys, D., 2019. Typhoid toxin exhausts the RPA response to DNA replication stress driving senescence and Salmonella infection. *Nat. Commun.* 10, 4040. <https://doi.org/10.1038/s41467-019-12064-1>.

- Iranzo, J., Lobkovsky, A.E., Wolf, Y.I., Koonin, E.V., 2014. Virus-host arms race at the joint origin of multicellularity and programmed cell death. *Cell Cycle* 13, 3083–3088. <https://doi.org/10.4161/15384101.2014.949496>.
- Johnson, A.A., Shokhirev, M.N., Shoshitaishvili, B., 2019. Revamping the evolutionary theories of aging. *Ageing Res. Rev.* 55, 100947 <https://doi.org/10.1016/j.arr.2019.100947>.
- Kalvakolanu, D.V., 1999. Virus interception of cytokine-regulated pathways. *Trends Microbiol.* 7, 166–171. [https://doi.org/10.1016/s0966-842x\(99\)01476-6](https://doi.org/10.1016/s0966-842x(99)01476-6).
- Kaminsky, V., Zhivotovsky, B., 2010. To kill or be killed: how viruses interact with the cell death machinery. *J. Intern. Med.* 267, 473–482. <https://doi.org/10.1111/j.1365-2796.2010.02222.x>.
- Karin, O., Agrawal, A., Porat, Z., Krizhanovsky, V., Alon, U., 2019. Senescent cell turnover slows with age providing an explanation for the Gompertz law. *Nat. Commun.* 10, 5495. <https://doi.org/10.1038/s41467-019-13192-4>.
- Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I., Michalak, P., Kang, L., Bhak, J., Yim, H.-S., Grishin, N.V., Nielsen, N.H., Heide-Jørgensen, M.P., Oziolor, E.M., Matsun, C.W., Church, G.M., Stuart, G.W., Patton, J.C., George, J.C., Suydam, R., Larsen, K., López-Otín, C., O'Connell, M.J., Bickham, J.W., Thomsen, B., de Magalhães, J.P., 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep.* 10, 112–122. <https://doi.org/10.1016/j.celrep.2014.12.008>.
- Kelley, W.J., Zemans, R.L., Goldstein, D.R., 2020. Cellular senescence: friend or foe to respiratory viral infections? *Eur. Respir. J.* 56, 2002708 <https://doi.org/10.1183/13993003.02708-2020>.
- Kim, Y.S., Nam, H.J., Chung, H.Y., Kim, N.D., Ryu, J.H., Lee, W.J., Arking, R., Yoo, M.A., 2001. Role of xanthine dehydrogenase and aging on the innate immune response of *Drosophila*. *J. Am. Aging Assoc.* 24, 187–193. <https://doi.org/10.1007/s11357-001-0020-6>.
- Kim, J.-A., Seong, R.-K., Shin, O.S., 2016. Enhanced viral replication by cellular replicative senescence. *Immune Netw.* 16, 286–295. <https://doi.org/10.4110/in.2016.16.5.286>.
- Kirkwood, T.B., 1977. Evolution of ageing. *Nature* 270, 301–304. <https://doi.org/10.1038/270301a0>.
- Kirkwood, T.B.L., 2005. Asymmetry and the origins of ageing. *Mech. Ageing Dev.* 126, 533–534. <https://doi.org/10.1016/j.mad.2005.02.001>.
- Kirkwood, T., Holliday, R., 1979. The evolution of ageing and longevity. *Proc. R. Soc. Lond. B Biol. Sci.* 205, 531–546. <https://doi.org/10.1098/rspb.1979.0083>.
- Komaki, K., Ishikawa, H., 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem. Mol. Biol.* 30, 253–258. [https://doi.org/10.1016/s0965-1748\(99\)00125-3](https://doi.org/10.1016/s0965-1748(99)00125-3).
- Koskella, B., Brockhurst, M.A., 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* 38, 916–931. <https://doi.org/10.1111/1574-6976.12072>.
- Kowald, A., Passos, J.F., Kirkwood, T.B.L., 2020. On the evolution of cellular senescence. *Aging Cell* 19, e13270. <https://doi.org/10.1111/acel.13270>.
- Kwon, I.-S., Kim, J., Rhee, D.-K., Kim, B.-O., Pyo, S., 2017. Pneumolysin induces cellular senescence by increasing ROS production and activation of MAPK/NF- κ B signal pathway in glial cells. *Toxicol. Int.* 129, 100–112. <https://doi.org/10.1016/j.toxicol.2017.02.017>.
- La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., Raoult, D., 2008. The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104. <https://doi.org/10.1038/nature07218>.
- Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327. <https://doi.org/10.1038/nrmicro2315>.
- Laing, E.D., Sterling, S.L., Weir, D.L., Beauregard, C.R., Smith, I.L., Larsen, S.E., Wang, L.-F., Snow, A.L., Schaefer, B.C., Broder, C.C., 2019. Enhanced autophagy contributes to reduced viral infection in black flying fox cells. *Viruses* 11. <https://doi.org/10.3390/v11030260>.
- Lamielle, O., Kellenberger, C., Kemp, C., Troxler, L., Pelte, N., Boutros, M., Marques, J. T., Daefliger, L., Hoffmann, J.A., Roussel, A., Imler, J.-L., 2016. Cytokine Diederl and a viral homologue suppress the IMD pathway in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 113, 698–703. <https://doi.org/10.1073/pnas.1516122113>.
- Lathe, R., Tzeng, N.-S., Itzhaki, R., 2019. Herpes infections and dementia: rebutting alternative fact. *Neurotherapeutics* 16, 176–179. <https://doi.org/10.1007/s13311-018-00700-5>.
- Lee, H.-Y., Lee, S.-H., Lee, J.-H., Lee, W.-J., Min, K.-J., 2019. The role of commensal microbes in the lifespan of *Drosophila melanogaster*. *Aging*. <https://doi.org/10.18632/aging.102073>.
- Leitão, E., Costa, A.C., Brito, C., Costa, L., Pombinho, R., Cabanes, D., Sousa, S., 2014. *Listeria monocytogenes* induces host DNA damage and delays the host cell cycle to promote infection. *Cell Cycle Georget. Tex.* 13, 928–940. <https://doi.org/10.4161/cc.27780>.
- Liu, J., Cvirkaitė-Krupovic, V., Baquero, D.P., Yang, Y., Zhang, Q., Shen, Y., Krupovic, M., 2021. Virus-induced cell gigantism and asymmetric cell division in archaea. *Proc. Natl. Acad. Sci. U. S. A.* 118. <https://doi.org/10.1073/pnas.2022578118>.
- Lwoff, A., 1957. The concept of virus. *Microbiology* 17, 239–253. <https://doi.org/10.1099/00221287-17-2-239>.
- MacNeil, L.T., Watson, E., Arda, H.E., Zhu, L.J., Walhout, A.J.M., 2013. Diet-induced developmental acceleration independent of TOR and insulin in *C. Elegans*. *Cell* 153, 240–252. <https://doi.org/10.1016/j.cell.2013.02.049>.
- Masemann, D., Ludwig, S., Boergeling, Y., 2020. Advances in transgenic mouse models to study infections by human pathogenic viruses. *Int. J. Mol. Sci.* 21, 9289. <https://doi.org/10.3390/ijms21239289>.
- Matthews, M.K., Wilcox, H., Hughes, R., Veloz, M., Hammer, A., Banks, B., Walters, A., Schneider, K.J., Sexton, C.E., Chaston, J.M., 2020. Genetic influences of the microbiota on the life span of *Drosophila melanogaster*. *Appl. Environ. Microbiol.* 86. <https://doi.org/10.1128/AEM.00305-20>.
- McFall-Ngai, M.J., 2002. Unseen forces: the influence of bacteria on animal development. *Dev. Biol.* 242, 1–14. <https://doi.org/10.1006/dbio.2001.0522>.
- Medawar, P.B., 1952. *An unsolved problem of biology*. College.
- Medvedev, Z.A., 1990. An attempt at a rational classification of theories of ageing. *Biol. Rev.* 65, 375–398. <https://doi.org/10.1111/j.1469-185x.1990.tb01428.x>.
- Merkx-Jacques, A., Coors, A., Brousseau, R., Masson, L., Mazza, A., Tien, Y.-C., Topp, E., 2013. Evaluating the pathogenic potential of environmental *Escherichia coli* by using the *Caenorhabditis elegans* infection model. *Appl. Environ. Microbiol.* 79, 2435–2445. <https://doi.org/10.1128/AEM.03501-12>.
- Microbiology by numbers, 2011. *Nat. Rev. Microbiol.* 9. <https://doi.org/10.1038/nrmicro2644>, 628–628.
- Mitteldorf, J., Pepper, J., 2009. Senescence as an adaptation to limit the spread of disease. *J. Theor. Biol.* 260, 186–195. <https://doi.org/10.1016/j.jtbi.2009.05.013>.
- Mliih, M., Khericha, M., Birdwell, C., West, A.P., Karpac, J., 2018. A virus-acquired host cytokine controls systemic aging by antagonizing apoptosis. *PLoS Biol.* 16, e2005796. <https://doi.org/10.1371/journal.pbio.2005796>.
- Muller, M., 2006. Premature cellular senescence induced by pyocyanin, a redox-active *Pseudomonas aeruginosa* toxin. *Free Radic. Biol. Med.* 41, 1670–1677. <https://doi.org/10.1016/j.freeradbiomed.2006.09.004>.
- Nikolich-Zugich, J., van Lier, R.A.W., 2017. Cytomegalovirus (CMV) research in immune senescence comes of age: overview of the 6th International Workshop on CMV and Immunosenescence. *GeroScience* 39, 245–249. <https://doi.org/10.1007/s11357-017-9984-8>.
- Nikolich-Zugich, J., Čičin-Šain, L., Collins-McMillen, D., Jackson, S., Oxenius, A., Sinclair, J., Snyder, C., Wills, M., Lemmermann, N., 2020. Advances in cytomegalovirus (CMV) biology and its relationship to health, diseases, and aging. *GeroScience* 42, 495–504. <https://doi.org/10.1007/s11357-020-00170-8>.
- Nyström, T., 2002. Aging in bacteria. *Curr. Opin. Microbiol.* 5, 596–601. [https://doi.org/10.1016/s1369-5274\(02\)00367-3](https://doi.org/10.1016/s1369-5274(02)00367-3).
- Padberg, I., Janßen, S., Meyer, T.F., 2013. Chlamydia trachomatis inhibits telomeric DNA damage signaling via transient hTERT upregulation. *Int. J. Med. Microbiol.* 303, 463–474. <https://doi.org/10.1016/j.ijmm.2013.06.001>.
- Pal, C., Maciá, M.D., Oliver, A., Schachar, I., Buckling, A., 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450, 1079–1081. <https://doi.org/10.1038/nature06350>.
- Partridge, L., Barton, N.H., 1993. Optimally, mutation and the evolution of ageing. *Nature* 362, 305–311. <https://doi.org/10.1038/362305a0>.
- Pernice, W.M., Swaney, T.C., Boldogh, I.R., Pon, L.A., 2017. Mitochondrial tethers and their impact on lifespan in budding yeast. *Front. Cell Dev. Biol.* 5, 120. <https://doi.org/10.3389/fcell.2017.00120>.
- Proença, A.M., Rang, C.U., Buetz, C., Shi, C., Chao, L., 2018. Age structure landscapes emerge from the equilibrium between aging and rejuvenation in bacterial populations. *Nat. Commun.* 9, 3722. <https://doi.org/10.1038/s41467-018-06154-9>.
- Quesada, V., Freitas-Rodríguez, S., Miller, J., Pérez-Silva, J.G., Jiang, Z.-F., Tapia, W., Santiago-Fernández, O., Campos-Iglesias, D., Kuderna, L.F.K., Quinzin, M., Álvarez, M.G., Carrero, D., Beheregaray, L.B., Gibbs, J.P., Chiari, Y., Glaberman, S., Ciofi, C., Araujo-Voces, M., Mayoral, P., Arango, J.R., Tamargo-Gómez, I., Roiz-Valle, D., Prascual-Torner, M., Evans, B.R., Edwards, D.L., Garrick, R.C., Russello, M.A., Poulakakis, N., Gaughran, S.J., Rueda, D.O., Bretones, G., Marqués-Bonet, T., White, K.P., Caccione, A., López-Otín, C., 2018. Giant tortoise genomes provide insights into longevity and age-related disease. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-018-0733-x>.
- Ramachandran, G., Bikard, D., 2019. Editing the microbiome the CRISPR way. *Philos. Trans. R. Soc. B Biol. Sci.* 374, 20180103. <https://doi.org/10.1098/rstb.2018.0103>.
- Rampelli, S., Candela, M., Turroni, S., Biagi, E., Collino, S., Franceschi, C., O'Toole, P.W., Brigidi, P., 2013. Functional metagenomic profiling of intestinal microbiome in extreme ageing. *Aging* 5, 902–912. <https://doi.org/10.18632/aging.100623>.
- Redeker, A., Remmerswaal, E.B.M., van der Gracht, E.T.I., Welten, S.P.M., Höllt, T., Koning, F., Čičin-Šain, L., Nikolich-Zugich, J., ten Berge, I.J.M., van Lier, R.A.W., van Unen, V., Arens, R., 2018. The contribution of cytomegalovirus infection to immune senescence is set by the infectious dose. *Front. Immunol.* 8. <https://doi.org/10.3389/fimmu.2017.01953>.
- Reichard, M., 2017. Evolutionary perspectives on ageing. *Semin. Cell Dev. Biol.* 70, 99–107. <https://doi.org/10.1016/j.semcdb.2017.05.013>.
- Roosinck, M.J., 2011. The good viruses: viral mutualistic symbioses. *Nat. Rev. Microbiol.* 9, 99–108. <https://doi.org/10.1038/nrmicro2491>.
- Sahu, S., Dattani, A., Aboobaker, A.A., 2017. Secrets from immortal worms: What can we learn about biological ageing from the planarian model system? *Semin. Cell Dev. Biol.* 70, 108–121. <https://doi.org/10.1016/j.semcdb.2017.08.028>.
- Sanson, P., Vecovini, R., Fagnoni, F.F., Akbar, A., Arens, R., Chiu, Y.-L., Čičin-Šain, L., Dechanet-Merville, J., Derhovanessian, E., Ferrando-Martínez, S., Franceschi, C., Frasca, D., Fulöp, T., Furman, D., Gkrania-Klotsas, E., Goodrum, F., Grubeck-Loebenstein, B., Hurme, M., Kern, F., Lillier, D., López-Botet, M., Maier, A.B., Marand, T., Marchant, A., Matheï, C., Moss, P., Muntasell, A., Remmerswaal, E.B.M., Riddell, N.E., Rothe, K., Sauce, D., Shin, E.-C., Simanek, A.M., Smithey, M.J., Söderberg-Nauclér, C., Solana, R., Thomas, P.G., van Lier, R., Pawelec, G., Nikolich-Zugich, J., 2014. New advances in CMV and immunosenescence. *Exp. Gerontol.* 55, 54–62. <https://doi.org/10.1016/j.exger.2014.03.020>.
- Sarkar, S., Heise, M.T., 2019. Mouse models as resources for studying infectious diseases. *Clin. Ther.* 41, 1912–1922. <https://doi.org/10.1016/j.clinthera.2019.08.010>.

- Schank, M., Zhao, J., Moorman, J.P., Yao, Z.Q., 2021. The impact of HIV- and ART-induced mitochondrial dysfunction in cellular senescence and aging. *Cells* 10. <https://doi.org/10.3390/cells10010174>.
- Schleiss, M.R., 2011. Congenital cytomegalovirus infection: molecular mechanisms mediating viral pathogenesis. *Infect. Disord. Drug Targets* 11, 449–465. <https://doi.org/10.2174/187152611797636721>.
- Schönrich, G., Abdelaziz, M.O., Raftery, M.J., 2017. Herpesviral capture of immunomodulatory host genes. *Virus Genes* 53, 762–773. <https://doi.org/10.1007/s11262-017-1460-0>.
- Schossere, M., Grillari, J., Breitenbach, M., 2017. The dual role of cellular senescence in developing tumors and their response to cancer therapy. *Front. Oncol.* 7, 278. <https://doi.org/10.3389/fonc.2017.00278>.
- Schwarzer, M., Strigini, M., Leulier, F., 2018. Gut microbiota and host juvenile growth. *Calcif. Tissue Int.* 102. <https://doi.org/10.1007/s00223-017-0368-y>.
- Seoane, R., Vidal, S., Bouzaher, Y.H., El Motiam, A., Rivas, C., 2020. The interaction of viruses with the cellular senescence response. *Biology* 9. <https://doi.org/10.3390/biology9120455>.
- Shin, S.C., Kim, S.-H., You, H., Kim, B., Kim, A.C., Lee, K.-A., Yoon, J.-H., Ryu, J.-H., Lee, W.-J., 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 334, 670–674. <https://doi.org/10.1126/science.1212782>.
- Singh, K.K., Chaubey, G., Chen, J.Y., Suravajhala, P., 2020. Decoding SARS-CoV-2 hijacking of host mitochondria in COVID-19 pathogenesis. *Am. J. Physiol. Cell Physiol.* 319, C258–C267. <https://doi.org/10.1152/ajpcell.00224.2020>.
- Smith, N.F., 2001. Spatial heterogeneity in recruitment of larval trematodes to snail intermediate hosts. *Oecologia* 127, 115–122. <https://doi.org/10.1007/s004420000560>.
- Smith, P., Willemsen, D., Popkes, M., Metge, F., Gandiwa, E., Reichard, M., Valenzano, D. R., 2017. Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. *eLife* 6, e27014. <https://doi.org/10.7554/eLife.27014>.
- Smolentseva, O., Gusarov, I., Gautier, L., Shamovsky, I., DeFrancesco, A.S., Losick, R., Nudler, E., 2017. Mechanism of biofilm-mediated stress resistance and lifespan extension in *C. elegans*. *Sci. Rep.* 7, 7137. <https://doi.org/10.1038/s41598-017-07222-8>.
- Sodhi, A., Montaner, S., Gutkind, J.S., 2004. Viral hijacking of G-protein-coupled-receptor signalling networks. *Nat. Rev. Mol. Cell Biol.* 5, 998–1012. <https://doi.org/10.1038/nrm1529>.
- Solon-Biet, S.M., Walters, K.A., Simanainen, U.K., McMahon, A.C., Ruohonen, K., Ballard, J.W.O., Raubenheimer, D., Handelsman, D.J., Le Couteur, D.G., Simpson, S. J., 2015. Macronutrient balance, reproductive function, and lifespan in aging mice. *Proc. Natl. Acad. Sci.* 112, 3481–3486. <https://doi.org/10.1073/pnas.1422041112>.
- Sorensen, R.E., Minchella, D.J., 2001. Snail-trematode life history interactions: past trends and future directions. *Parasitology* 123 (Suppl), S3–18. <https://doi.org/10.1017/s0031182001007843>.
- Stephens, C., 2005. Senescence: even bacteria get old. *Curr. Biol.* 15, R308–310. <https://doi.org/10.1016/j.cub.2005.04.006>.
- Stewart, E.J., Madden, R., Paul, G., Taddei, F., 2005. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* 3, e45. <https://doi.org/10.1371/journal.pbio.0030045>.
- Storelli, G., Defaye, A., Erkosar, B., Hols, P., Royet, J., Leulier, F., 2011. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metab.* 14, 403–414. <https://doi.org/10.1016/j.cmet.2011.07.012>.
- Sun, Q., Vega, N.M., Cervantes, B., Mancuso, C.P., Mao, N., Taylor, M., Collins, J.J., Khalil, A.S., Gore, J., Lu, T.K., 2020. Colonization with heterologous bacteria reprograms a *Caenorhabditis elegans* nutritional phenotype (preprint). *Microbiology*. <https://doi.org/10.1101/2020.03.01.972349>.
- Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V.E., de Magalhães, J.P., 2018. Human ageing genomic resources: new and updated databases. *Nucleic Acids Res.* 46, D1083–D1090. <https://doi.org/10.1093/nar/gkx1042>.
- Takaki, H., Oshiumi, H., Shingai, M., Matsumoto, M., Seya, T., 2017. Development of mouse models for analysis of human virus infections: mouse models for viral infection. *Microbiol. Immunol.* 61, 107–113. <https://doi.org/10.1111/1348-0421.12477>.
- Teulière, J., Bhattacharya, D., Bapteste, E., 2020. Ancestral germs/soma distinction in microbes: expanding the disposable soma theory of aging to all unicellular lineages. *Ageing Res. Rev.* 101064. <https://doi.org/10.1016/j.arr.2020.101064>.
- Thompson, M.J., Chwiałkowska, K., Rubbi, L., Lusi, A.J., Davis, R.C., Srivastava, A., Korstanje, R., Churchill, G.A., Horvath, S., Pellegrini, M., 2018. A multi-tissue full lifespan epigenetic clock for mice. *Aging* 10, 2832–2854. <https://doi.org/10/gfvm68>.
- Uribe-Herranz, M., Klein-González, N., Rodríguez-Lobato, L.G., Juan, M., Fernández de Larrea, C., 2021. Gut microbiota influence in hematological malignancies: from genesis to cure. *Int. J. Mol. Sci.* 22. <https://doi.org/10.3390/ijms22031026>.
- van der Lugt, B., Rusli, F., Lute, C., Lamprakis, A., Salazar, E., Boekschoten, M.V., Hooiveld, G.J., Müller, M., Vervoort, J., Kersten, S., Belzer, C., Kok, D.E.G., Steegenga, W.T., 2018. Integrative analysis of gut microbiota composition, host colonic gene expression and intraluminal metabolites in aging C57BL/6J mice. *Aging* 10, 930–950. <https://doi.org/10.18632/aging.101439>.
- Veyer, D.L., Carrara, G., Maluquer de Motes, C., Smith, G.L., 2017. Vaccinia virus evasion of regulated cell death. *Immunol. Lett.* 186, 68–80. <https://doi.org/10.1016/j.imlet.2017.03.015>.
- Villarreal, L.P., Witzany, G., 2010. Viruses are essential agents within the roots and stem of the tree of life. *J. Theor. Biol.* 262, 698–710. <https://doi.org/10.1016/j.jtbi.2009.10.014>.
- von Kobbe, C., 2019. Targeting senescent cells: approaches, opportunities, challenges. *Aging* 11, 12844–12861. <https://doi.org/10.18632/aging.102557>.
- Wang, B., Kohli, J., Demaria, M., 2020a. Senescent cells in cancer therapy: friends or foes? *Trends Cancer* 6, 838–857. <https://doi.org/10.1016/j.trecan.2020.05.004>.
- Wang, Y., Lo, W.-C., Chou, C.-S., 2020b. Modelling stem cell ageing: a multi-compartment continuum approach. *R. Soc. Open Sci.* 7, 191848. <https://doi.org/10.1098/rsos.191848>.
- Wargo, A.R., Kurath, G., 2012. Viral fitness: definitions, measurement, and current insights. *Curr. Opin. Virol.* 2, 538–545. <https://doi.org/10.1016/j.coviro.2012.07.007>.
- Weitzman, M.D., Weitzman, J.B., 2014. What's the damage? The impact of pathogens on pathways that maintain host genome integrity. *Cell Host Microbe* 15, 283–294. <https://doi.org/10.1016/j.chom.2014.02.010>.
- Werren, J.H., Baldo, L., Clark, M.E., 2008. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6, 741–751. <https://doi.org/10.1038/nrmicro1969>.
- Williams, G.C., 1957. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 11, 398–411. <https://doi.org/10.1111/j.1558-5646.1957.tb02911.x>.
- Wouk, J., Rechenchoski, D.Z., Rodrigues, B.C.D., Ribelato, E.V., Faccin-Galhardi, L.C., 2021. Viral infections and their relationship to neurological disorders. *Arch. Virol.* 1–21. <https://doi.org/10.1007/s00705-021-04959-6>.
- Yatim, N., Albert, M.L., 2011. Dying to replicate: the orchestration of the viral life cycle, cell death pathways, and immunity. *Immunity* 35, 478–490. <https://doi.org/10.1016/j.immuni.2011.10.010>.
- Zamaraev, A.V., Zhivotovsky, B., Kopeina, G.S., 2020. Viral infections: negative regulators of apoptosis and oncogenic factors. *Biochem. Biokhimiia* 85, 1191–1201. <https://doi.org/10.1134/s0006297920100077>.