HAL
open science

Identification in the bacterial symbiont Burkholderia insecticola of factors involved in antimicrobial peptide-resistance and colonization of the insect Riptortus pedestris

Joy Lachat

To cite this version:

Joy Lachat. Identification in the bacterial symbiont Burkholderia insecticola of factors involved in antimicrobial peptide-resistance and colonization of the insect Riptortus pedestris. Symbiosis. Université Paris Saclay (COmUE), 2019. English. NNT : 2019SACLS274 . tel-02947367

HAL Id: tel-02947367

https://theses.hal.science/tel-02947367v1

Submitted on 24 Sep 2020

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.

université PARIS-SACLAY

NNT : 2019SACLS274

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Identification des facteurs de résistance aux peptides antimicrobiens et de colonisation de l'insecte *Riptortus pedestris* chez la bactérie symbiotique *Burkholderia insecticola*.

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°567 : Sciences du végétal : du gène à l'écosystème (SDV)
Spécialité de doctorat: Biologie

Thèse présentée et soutenue à Gif-sur-Yvette, le 23 septembre 2019, par

Joy Lachat

Composition du Jury :

Stéphanie Bury-Moné Enseignant-Chercheur, CNRS (I2BC)	Présidente
Annette Vergunst Chargée de Recherches, INSERM (U 1047)	Rapporteur
Aurélien Carlier Directeur de Recherche, INRA (LIPM Toulouse)	Rapporteur
Anna Zaidman-Rémy Enseignant-Chercheur, CNRS (INSA Lyon)	Examinateur
Amandine Cornille Chargée de Recherches, INRA (GQE-Le Moulon)	Examinateur
Peter Mergaert Directeur de Recherche, CNRS (I2BC)	Directeur de thèse

Notre plus grande gloire
n'est point de tomber,
mais de savoir nous relever
chaque fois que nous tombons.
Confucius

Remerciements

Tout d'abord, je tiens à remercier mon directeur de thèse, Peter Mergaert, pour ses précieux conseils, son soutien, sa gentillesse, sa patience et ses enseignements durant ces trois années de thèse. Merci de m'avoir soutenue et encouragée pour mener ce travail à bien.

Je tiens aussi à remercier les membres de mon jury de thèse : Stéphanie Bury-Moné, Amandine Cornille, Anna Zaidman-Rémy, et en particulier Aurélien Carlier et Annette Vergunst pour avoir évaluer mon travail.

Je souhaite également remercier les membres de mon comité de thèse : Lionel Moulin, Erwan Gueguen et Akiko Sugio. Merci à vous pour votre aide et vos conseils avisés, par deux fois prodigués.

Je remercie aussi les collaborateurs du Japon qui m'ont accueilli chaleureusement pendant deux séjours dans leur laboratoire à Hokkaido, notamment le chef d'équipe Yoshitomo Kikuchi et sa famille, Hideomi Itoh et sa femme, ainsi que les étudiants Shiori Goto, Yuki Sano et Shunta qui m'ont fait découvrir la culture japonaise et avec qui j'ai passé de bons moments. Durant mon deuxième séjour avec le Summer Program JSPS, j'ai aussi pu rencontrer de nombreux autres étudiants en thèse venant des Etats-Unis qui étaient également à Sapporo avec moi, notamment Nikolaï Hensley et Lina Wang. Je tiens à les remercier énormément pour tous les bons souvenirs, les voyages, les fous rires qu'on a partagé ensembles au Japon.

Je tiens à remercier tous les membres permanents de mon équipe : notre co-chef d'équipe Denis Faure pour ses conseils, sa gentillesse, son humour et son sérieux qui m'ont beaucoup inspiré pour continuer dans la recherche, Yves Dessaux pour ses sages paroles, son humour et ses conseils, Benoit Alunni et Tania Timtchenko pour son aide durant le concours de thèse, Catherine GrandClément pour sa gentillesse, et Raynald Cossard avec qui j'ai pu travailler durant ma dernière année de thèse avec sa bonne humeur et son humour.

Je souhaite également remercier tous les autres membres de l'équipe : Jérémy Cigna, Marta Torres Béjar, Quentin Barrière, Nicolas Busset, Tsubasa Ohbayashi, Quentin Nicoud, mais aussi les « anciens » Pauline Blin, Almudena Gonzales-Mula et Florian Lamouche, qui m'ont beaucoup soutenu et avec lesquels j'ai partagé de bons moments, des fous rires, de chouettes

conversations et aussi quelques bières. Un grand merci aussi à tous les stagiaires qui sont venus au labo, notamment Christy Calif pour son implication dans le projet insectes avec qui j'ai pu travailler.

Parmi tous mes collègues, je réserve quand même une section spéciale pour Kévin Robic dit Kéké, mon acolyte thésard. Je te remercie pour tous les fous rires, les blagues, les discussions mangas, les blind tests de musiques, et bien sûr les discussions Tn-seq qui t'ont conduit à devenir le deuxième expert du labo. Maintenant, c'est toi qui devient « l'expert Tn-seq » de l'équipe. Merci pour cette bonne ambiance que tu as apporté au labo et à notre bureau, je suis contente d'avoir trouvé un ami avec qui je partage de bons souvenirs.

Enfin, je souhaite remercier toute ma famille, notamment mon frère William, mes sœurs Célia et Kathleen, mes beaux-frères Michel et Jules, mon père, et surtout mes amis proches Camille et Maxime pour m'avoir encouragé tout au long de ce travail et pour avoir cru en moi jusqu'au bout.

Liste des Abréviations

AMP: Antimicrobial peptide

Apo: Aposymbiotic

ARTIST: Analysis of high-Resolution Transposon-Insertion Sequences Technique

B. cenocepacia: *Burkholderia cenocepacia*

B. insecticola: *Burkholderia insecticola*

B. pseudomallei: *Burkholderia pseudomallei*

B. thailandensis: *Burkholderia thailandensis*

BCC&P: *Burkholderia cepacia* complex and *Burkholderia pseudomallei* group

BCR: Bacteriocyte-specific Cysteine Rich peptide

BGC: *Burkholderia glathei* clade

BHT: Bacteriohopanetetrol

BLAST: Basic Local Alignment Search Tool

Con-ARTIST: Conditionally essential loci ARTIST analysis

CCR: Crypt-specific Cysteine Rich peptide

cDNA: complementary DNA

CFU: Colony-Forming Unit

CI: Competitive Index

COG: Cluster of Orthologous Genes

CPP: Cell wall Penetrating Peptide

CR: Constricted Region

DNA: Deoxyribonucleic acid

Dpi: Days Post-Infection

E. coli: *Escherichia coli*

ED: Entner-Doudoroff pathway

EL-ARTIST: Essential-Loci ARTIST analysis

EMP: Embden-Meyerhof-Parnas pathway

ESR: Extracytoplasmic Stress Response

FISH: Fluorescence *In Situ* Hybridization

GFP: Green Fluorescent Protein

GNBP: Gram Negative Binding Protein

HITS: High-throughput Insertion Track by deep-Sequencing

HMM: Hidden Markov Model

IGV: Integrated Genome Viewer

InSEQ: Insertion Sequencing

IR: Inverted Repeats

Kb: Kilobase

Km: Kanamycin

LB: Luria Bertani

LPS: Lipopolysaccharide

M: Midgut

Mb: Megabase

MCP: Methyl-accepting Chemotaxis Protein

MIC: Minimum Inhibitory Concentration

MM: Minimal Medium

NCR: Nodule-specific Cysteine Rich peptide

ND: Not Determined

OD: Optical Density

OMP: Outer Membrane Protein

PBE: Plant-Beneficial and Environmental group

PBS: Phosphate-buffered saline

PCR: Polymerase Chain Reaction

PGA: Poly-β-1,6-N-acetyl-D-glucosamine

PGRP: Peptidoglycan Recognition Protein

PHA: Polyhydroxyalkanoate

PP: Pentose-phosphate pathway

PRR: Pattern Recognition Receptor

R. pedestris: *Riptortus pedestris*

RFP: Red Fluorescent Protein

Rif: Rifampicin

RNA: Ribonucleic acid

rRNA: ribosomic RNA

tRNA: transfert RNA

RNAi: RNA interference

RNA-seq: RNA-sequencing

RND: Resistance Nodulation and cell Division

ROS: Reactive Oxygen Species

Rpm: Rotation per minute

SAM: S-Adenosyl Methionine

SBE: Stinkbug-associated Beneficial and Environmental group

SDS: Sodium Dodecyl Sulfate

Sym: Symbiotic

TA: Thymine-Adenine

Tat: Twin-arginine translocation

TCA: Tricarboxylic acid cycle

Tn-seq: Transposon-sequencing

TPR: Tetratricopeptide repeat

TraDIS: Transposon-Directed Insertion site Sequencing

WT: Wild-Type

YG: Yeast extract and Glucose medium

Résumé de Thèse

Ce travail de thèse porte sur les interactions symbiotiques entre un insecte phytophage *Riptortus pedestris* et son partenaire bactérien *Burkholderia insecticola*, durant lequel je me suis focalisée sur l'identification des facteurs bactériens impliqués dans la résistance aux peptides antimicrobiens (AMPs) et dans la colonisation de l'insecte-hôte.

1. Contexte général

Les symbioses avec des microorganismes sont répandues dans le monde du vivant, en particulier chez les insectes (Douglas, 2011). En effet, le développement et la survie d'espèces d'insecte dépend beaucoup d'interactions symbiotiques, notamment avec des bactéries (Douglas, 2011). La plupart de ces associations permettent de compléter les apports en nutriments absents dans le régime alimentaire des insectes, tels que des acides aminés essentiels ou des vitamines (Engel et Moran, 2013; Pickard *et al.*, 2017). Ces nutriments sont synthétisés par leurs bactéries symbiotiques, comme par exemple la bactérie *Rhodococcus rhodnii* qui fournit des vitamines B à son hôte *Rhodnius prolixus* (Eichler et Schaub, 2002). En complément de ces avantages nutritionnels, les symbiotes bactériens peuvent également jouer un rôle dans la protection contre des agents pathogènes (Oliver *et al.*, 2010), la dégradation de composants complexes (Engel et Moran, 2013), et même la manipulation du système reproducteur (Werren *et al.*, 2008). Ces bactéries symbiotiques peuvent être transmises à l'hôte par deux manières, horizontale ou verticale (Moran, 2006). Une transmission de type horizontale s'effectue par l'acquisition du symbiote dans l'environnement à chaque nouvelle génération d'hôtes (Moran, 2006). A l'inverse, une transmission verticale signifie que les bactéries symbiotiques sont directement transmises de la génération parentale à la descendance (Moran, 2006; Salem *et al.*, 2015). Dans le cadre de symbioses verticales, il existe différents modèles symbiotiques étudiés chez les insectes dont l'association entre le puceron (*Acyrthosiphon pisum*) et la bactérie symbiotique *Buchnera aphidicola* (Shigenobu et Wilson, 2011). Cette bactérie appartient au groupe des γ-Proteobacteria, et est un symbiote obligatoire intracellulaire contenu à l'intérieur de cellule-hôtes spécifiques appelées bactériocytes (Wilson et Duncan, 2015). Ces bactériocytes sont eux-mêmes organisés en organes bilobés constituant ainsi les organes symbiotiques du puceron (Wilson et Duncan, 2015). Cet insecte est un phytophage et un nuisible agricole connu pour se nourrir du phloème des plantes telles que le pois, ayant donc un régime riche en

glucides mais pauvre en acides aminés essentiels (Hansen et Moran, 2011). Comme je l'ai mentionné précédemment, cette symbiose obligatoire est un exemple de symbiose nutritionnelle basée sur les échanges de nutriments essentiels pour la survie de l'hôte, mais également pour la survie de la bactérie (Hansen et Moran, 2011; Shigenobu et Wilson, 2011). Cependant, comme le puceron et son symbiote bactérien ne peuvent pas survivre indépendamment l'un de l'autre, il est donc difficile d'étudier les fonctions symbiotiques avec tous les outils de génomique fonctionnelle à disposition.

Dans le cadre de cette thèse, j'ai étudié un autre modèle d'études des interactions symbiotiques chez les insectes, impliquant la punaise *Riptortus pedestris* et son symbiote bactérien *Burkholderia insecticola* (Takeshita et Kikuchi, 2017). La punaise *R. pedestris* est un insecte phytophage qui sévit en Asie du sud-est, notamment au Japon et en Corée du Sud, se nourrissant préférentiellement des graines de légumineuses telles que le soja (Bae *et al.*, 2014; Kikuchi *et al.*, 2007). Il s'agit d'un insecte hémimétabole, c'est-à-dire à métamorphose incomplète, dont l'âge adulte est atteint après cinq stades larvaires au bout de 20 jours (Kikuchi *et al.*, 2011a). Cette punaise fait partie de la famille des Alydidae (sous-ordre des Hétéroptères), dont plusieurs membres sont en association avec des espèces bactériennes du genre *Burkholderia* (Kikuchi *et al.*, 2011b). Il a été montré que cet insecte possède un symbiote bactérien unique, appelé *B. insecticola*, résidant dans une portion spécifique de l'intestin dénommée la région M4 (Kikuchi *et al.*, 2007). Cette région est organisée en deux rangées de centaines de cryptes, au sein de laquelle la population bactérienne extracellulaire prolifère, et constitue ainsi l'organe symbiotique (Kikuchi *et al.*, 2007). L'insecte acquiert cette bactérie de manière horizontale dans son environnement à des stades larvaires précoces, notamment pendant le second stade larvaire (Kikuchi *et al.*, 2011a). Comme il s'agit d'une symbiose à transmission horizontale, chaque nouvelle génération d'insectes est dénuée de bactéries symbiotiques et donc qualifiée d'aposymbiotique (Takeshita et Kikuchi, 2017). Cette symbiose étant donc facultative, il est ainsi possible d'élever des insectes aposymbiotiques viables et de les maintenir au laboratoire, et également de cultiver *B. insecticola in vitro*, constituant ainsi un modèle idéal pour étudier l'hôte et le symbiote séparément l'un de l'autre (Takeshita et Kikuchi, 2017). Cependant, chaque insecte collecté dans la nature est en association symbiotique avec *B. insecticola*, ce qui suggère une forte affinité entre ces deux partenaires et une coexistence nécessaire et stable malgré ce caractère facultatif (Kikuchi *et al.*, 2007). En

comparant des insectes aposymbiotiques et des insectes en symbiose avec *B. insecticola*, il s'avère que la symbiose contribue à améliorer la morphologie et le développement de l'insecte, s'illustrant par un temps de développement plus rapide, une croissance plus importante et un taux de fécondité plus élevé chez les insectes symbiotiques (Kikuchi *et al.*, 2007; Takeshita et Kikuchi, 2017). En revanche, les bénéfices de cette symbiose pour la bactérie restent encore méconnus. Comme la population bactérienne semble piégée à l'intérieur de l'organe symbiotique et n'est jamais excrétée pendant toute la durée de vie de l'insecte, il pourrait être possible que cette population soit libérée dans l'environnement après la mort de l'insecte. Le symbiote *B. insecticola* se multipliant largement dans l'organe symbiotique, ce largage de bactéries lors de la mort de l'insecte pourrait augmenter la prévalence de la population de *B. insecticola* dans le sol. Comme mentionné précédemment, le caractère facultatif de cette symbiose permet d'étudier indépendamment l'insecte-hôte et la bactérie symbiotique. Il a donc été possible d'inactiver des gènes de l'hôte par ARN d'interférence (ARNi) (Futahashi *et al.*, 2011) et également de séquencer et manipuler le génome de *B. insecticola* (Takeshita *et al.*, 2018). L'application de ces outils de génomique, la facilité d'élevage des insectes, et le caractère horizontal et facultative de cette symbiose rendent ce système idéal pour l'étude des interactions symbiotiques dans un contexte d'association insecte-bactérie (Kim et Lee, 2015; Takeshita et Kikuchi, 2017).

Dans le cadre des symbioses, l'immunité des organismes hôtes doit prendre en compte le symbiote bactérien étranger et ne pas le considérer comme un pathogène menaçant l'intégrité et la survie de l'hôte. Dans de telles associations, les organismes hôtes produisent des AMPs spécifiques dits symbiotiques, participant au contrôle et au maintien de la population symbiotique résidente (Mergaert, 2018). Les AMPs forment une très large famille de peptides produites par un vaste nombre d'organismes et possédant des propriétés antimicrobiennes (Bechinger et Gorr, 2017; Brogden, 2005). La plupart de ces AMPs sont cationiques et ciblent la paroi des bactéries chargée négativement, ce qui entraîne une perturbation de l'intégrité membranaire et donc, à terme, provoque une lyse bactérienne (Kumar *et al.*, 2018). Chez le modèle *Rhizobium*-légumineuses, il existe une catégorie spécifique d'AMPs produits exclusivement au sein des structures renfermant les bactéries symbiotiques appelées nodosités (Kondorosi *et al.*, 2013; Mergaert *et al.*, 2003). Ces AMPs dénommés « Nodule Cysteine-rich Peptides » ou peptides NCR représentent une large famille

de peptides à activité antimicrobienne riches en cystéines comprenant près de 600 membres chez *Medicago truncatula* (Mergaert *et al.*, 2003). Les rhizobia internalisés dans les nodosités optent pour un changement de morphologie très allongée avec un contenu en ADN plus important, et sont renommées bactéroïdes (Kondorosi *et al*., 2013). Comme il a été montré que les peptides NCR sont internalisés par les bactéries (Guefrachi *et al.*, 2015), il est possible que ces peptides régulent des gènes impliqués dans la morphologie observée chez les bactéroïdes (Barrière *et al.*, 2017; Guefrachi *et al*., 2015; Kondorosi *et al.*, 2013). De manière similaire, les insectes-hôtes produisent également des AMPs symbiotiques, comme le puceron produisant certains types d'AMPs exclusivement au sein des bactériocytes appelés « Bacteriocyte Cysteine-rich Peptides » ou peptides BCR (Uchi *et al.*, 2019). Il s'avère que chez un autre insecte, le charençon (espèces *Sitophilus*), en symbiose avec une bactérie spécifique localisée également dans des bactériocytes, il y a une production d'un AMP symbiotique appelé coléoptéricine A qui jouerait un rôle dans le maintien de la population symbiotique à l'intérieur des bactériocytes (Anselme *et al.*, 2008; Login *et al.,* 2011). En ce qui concerne la punaise *R. pedestris*, il a été montré récemment que cet insecte produit aussi une catégorie spécifique d'AMPs à l'intérieur de l'organe symbiotique appelés « Crypt-specific Cysteine-rich Peptides » ou peptides CCR (Futahashi *et al.*, 2013). En plus de ces peptides CCR, cette punaise produit d'autres AMPs comme la riptocine, les trialysines, la rip-thanatine et la rip-défensine (Kim *et al.*, 2016; Lee *et al*., 2017; Park *et al.*, 2018). Pour que les symbiotes colonisent leurs hôtes, il leur faut être capable de résister à l'activité antimicrobienne de ces AMPs symbiotiques. En effet, les espèces du genre *Burkholderia* sont connues pour être résistantes à un large spectre d'antibiotiques tels que les β-lactames, les macrolides, les aminoglycosides et également les AMPs comme les polymyxines (Rhodes et Schweizer, 2016; Sfeir, 2018). Comme je l'ai mentionné précédemment, les AMPs comme les polymyxines ciblant préférentiellement la paroi bactérienne, les mécanismes de résistance caractérisés chez les espèces du genre *Burkholderia* incluent des modifications membranaires (Rhodes et Schweizer, 2016). Par exemple, la présence d'un groupement 4-amino-4-déoxy-arabinose (Ara4N) dans les lipopolysaccharides (LPS) présents à la surface de la membrane externe des bactéries du genre *Burkholderia* permet de réduire considérablement les charges négatives, et donc de diminuer les interactions avec les AMPs (Ortega *et al*., 2007; Sfeir, 2018). Récemment, il a été montré au laboratoire que le symbiote *B. insecticola* est également

résistant à l'activité antimicrobienne des polymyxines, ainsi qu'aux peptides CCR.
Sachant que *B. insecticola* est également résistant aux AMPs, existe-t-il une corrélation entre la faculté de résister aux AMPs et la capacité de coloniser efficacement l'organe symbiotique de *R. pedestris* ? Reformulée d'une autre manière, on peut se demander si les AMPs produits par *R. pedestris*, dont les peptides CCR, sont-ils impliqués dans la colonisation spécifique de l'organe symbiotique par le symbiote *B. insecticola* ?
Pour répondre à cette question, ce travail de thèse s'est basé sur l'utilisation de la technique de Transposon-sequencing ou Tn-seq et s'est réparti en trois axes. Dans un premier temps, la technique de Tn-seq a été mise au point chez *B. insecticola* et a permis d'identifier le génome essentiel de cette bactérie. De plus, j'ai évalué la robustesse de cette méthode en identifiant des gènes participant à l'exploitation de sources de carbones. Dans un second temps, j'ai déterminé les gènes impliqués dans la résistance aux AMPs via une approche gènes candidats et également avec l'approche Tn-seq. Et enfin dans une dernière partie, j'ai décrit les fonctions symbiotiques identifiées pour la colonisation de l'organe symbiotique de *R. pedestris* grâce à une expérience de Tn-seq *in vivo*.
Durant ce travail de thèse, j'ai eu l'occasion de me rendre chez des collaborateurs au Japon, à l'AIST (Advanced national Institute of Science and Technology) d'Hokkaido, qui m'ont enseigné les méthodes d'élevage d'insectes et de dissections nécessaires pour mener à bien ce projet à mon laboratoire.

2. Détermination du génome essentiel de *B. insecticola* par Tn-seq

Durant ce travail de thèse, j'ai mis en place la technique de Tn-seq pour l'appliquer sur les bactéries étudiées au laboratoire, dont *B. insecticola*. Cette approche repose sur l'utilisation d'une banque de bactéries mutées aléatoirement par l'insertion d'un transposon pour identifier les gènes requis pour une condition donnée par séquençage haut débit (Chao *et al.*, 2016; van Opijnen *et al.*, 2009). En effet, lors d'un traitement spécifique, les bactéries mutantes contenues dans cette banque ayant perdu leur capacité à croître sont donc mutées pour des gènes impliqués dans cette condition particulière (Chao *et al.*, 2016; van Opijnen *et al.*, 2009). De plus, en utilisant cette banque de mutants aléatoires et en localisant le site

d'insertion du transposon dans le génome de la population bactérienne mise en culture dans un milieu riche, il est possible d'identifier les gènes strictement essentiels pour la viabilité bactérienne (Chao *et al*., 2016; van Opijnen *et al.*, 2009). Ainsi, le Tn-seq a permis d'identifier le génome essentiel de plusieurs espèces bactériennes, comprenant notamment des pathogènes humains comme *Vibrio cholerae* (Chao *et al*., 2013; Kamp *et al.*, 2013) ou *Mycobacterium tuberculosis* (DeJesus *et al*., 2017).

Dans un premier temps, j'ai généré une banque de mutants de *B. insecticola* par insertion d'un transposon de type *mariner*, qui cible spécifiquement les dinucléotides TA dans le génome. Sachant qu'il y a 110735 sites TA dans le génome de *B. insecticola* (6,96 Mb, possédant trois chromosomes et deux plasmides) et qu'il n'y a que 1,7% des gènes qui ne possèdent pas de sites TA, il était donc possible de générer une banque suffisamment conséquente et représentative en mutants aléatoires via le recours à ce type de transposon. Ainsi, nous avons pu obtenir une banque de *B. insecticola* contenant approximativement $2x10^7$ mutants indépendants. Pour pouvoir utiliser cette banque de mutants pour les futurs objectifs de la thèse, il fallait tout d'abord vérifier la robustesse de cette banque ainsi que l'efficacité de la méthode Tn-seq pour identifier des gènes spécifiques pour une condition donnée. Le premier aspect de ce travail de thèse a donc consisté en la validation de la banque créée et de vérifier la cohérence entre la condition testée et les fonctions identifiées avec l'analyse bioinformatique choisie pour le Tn-seq. Pour analyser ces données Tn-seq, j'ai utilisé l'analyse ARTIST basée sur un modèle HMM (« Hidden Markov Model ») pour prédire les gènes essentiels d'une bactérie (El-ARTIST), et également utilisée pour prédire les gènes requis pour une condition spécifique (Con-ARTIST) (Pritchard *et al.*, 2014).

Tout d'abord, j'ai utilisé la banque de mutants préalablement construite pour identifier le génome essentiel de *B. insecticola* dans un milieu riche. Avec l'analyse El-ARTIST, j'ai pu trouver 1080 gènes essentiels (sur un nombre total de 6352 gènes) pour la survie de *B. insecticola* en milieu riche. Ces gènes étaient notamment localisés sur le chromosome 1 (NC_021287.1) et le plasmide 1 (NC_021289.1), et leurs fonctions étaient principalement associées aux mécanismes de transcription et de traduction, à la production d'énergie, aux activités métaboliques, et à la biosynthèse des membranes bactériennes. Les gènes identifiés comme essentiels étaient représentatifs des activités vitales pour la survie de la bactérie, tels que les gènes codant pour les protéines ribosomales 50S et 30S, pour la biosynthèse du

peptidoglycane, et pour les sous-unités de l'ATP synthase. En comparant ce génome essentiel identifié chez *B. insecticola* avec les génomes essentiels caractérisés chez d'autres espèces de *Burkholderia* via Tn-seq, comme *B. pseudomallei* (Moule *et al.*, 2014), *B. cenocepacia* (Wong *et al.*, 2016) et *B. thailandensis* (Baugh *et al.*, 2013), j'ai remarqué que ces mêmes fonctions essentielles étaient également partagées entre ces quatre espèces. Avant d'obtenir les résultats Tn-seq pour *B. insecticola*, la comparaison des gènes essentiels identifiés pour les trois autres espèces de *Burkholderia* a montré que seulement 164 gènes étaient communément essentiels (Wong *et al.*, 2016). Lorsque j'ai effectué cette même comparaison avec *B. insecticola*, j'ai trouvé que 151 gènes essentiels étaient partagés entre les quatre espèces, ce qui est très semblable au résultat décrit précédemment, et représente un pool de gènes essentiels pour le genre *Burkholderia*.

Une fois que le génome essentiel de *B. insecticola* a été identifié, je l'ai gardé comme référentiel pour déterminer les gènes requis pour la fitness bactérienne sous une condition spécifique, telles que la présence d'AMPs ou encore la colonisation *in vivo* que nous souhaitions réaliser. Pour pouvoir valider la robustesse de la méthode Tn-seq, j'ai choisi d'identifier les gènes impliqués dans l'exploitation de deux sources de carbones, le glucose et le succinate, dans un milieu minimum. Le choix de ces deux conditions était motivé par le fait que nous nous attendions à identifier des gènes impliqués dans la glycolyse, la gluconéogenèse et le transport de ces molécules, et qui confirmerait la validité de l'analyse Tn-seq mise en place. En présence du glucose, j'ai trouvé plusieurs gènes impliqués dans la glycolyse de type Entner-Doudoroff ainsi que de gènes codant un système ABC pour le transport de glucose. Pour le succinate, j'ai également trouvé un transporteur de dicarboxylates, ainsi que plusieurs gènes indiquant que le succinate peut être directement être intégré au cycle de Krebs, ou alors servir d'intermédiaire pour créer du glucose via la voie de gluconéogenèse. Grâce à ces résultats, j'ai pu conclure que le Tn-seq était suffisamment robuste pour détecter des gènes d'intérêts impliqués dans les conditions testées.

Comme je l'ai mentionné précédemment, l'utilisation du Tn-seq s'est généralisée au laboratoire et a pu ainsi être appliquée à d'autres bactéries étudiées au sein de l'équipe, notamment sur *Agrobacterium tumefaciens*. Le génome essentiel d'*A. tumefaciens* a été déterminé, ainsi que les gènes impliqués dans l'exploitation de sources de carbones telles que

le sucrose et le GHB, et ces résultats obtenus par Tn-seq ont fait l'objet d'une publication dont je suis deuxième auteur (Gonzalez-Mula *et al.*, 2018, voir **Publications**).

3. Identification des facteurs bactériens impliqués dans la résistance aux peptides antimicrobiens chez *B. insecticola*

Ayant confirmé la robustesse de la méthode Tn-seq chez *B. insecticola* durant les expériences précédentes, il est devenu possible d'utiliser cette méthode pour identifier les gènes bactériens impliqués dans la résistance aux AMPs. En complément de l'approche Tn-seq, j'ai également étudié d'autres cibles potentielles ayant été décrites chez d'autres espèces du genre *Burkholderia* pour leur participation à la résistance aux AMPs. Basé sur des recherches bibliographiques, je me suis focalisée sur trois composants membranaires ayant été décrits comme des facteurs de résistance aux AMPs chez *B. cenocepacia* et *B. multivorans* : le LPS (Loutet *et al*., 2006), les hopanoïdes (Malott *et al.*, 2012; Schmerk *et al.*, 2011), et le facteur RpoE de réponse au stress extracellulaire ou ESR (Flannagan and Valvano, 2008). Le LPS est un composant majeur des parois membranaires des bactéries à Gram négatif, composé de trois parties : le lipide A, le domaine « core oligosaccharide » et l'antigène O. Il a été montré que le « core oligosaccharide », en particulier la partie interne, est nécessaire pour maintenir la capacité de résistance aux AMPs chez *B. cenocepacia* (Loutet *et al.*, 2006). Les hopanoïdes sont des molécules lipidiques de type triterpènes, faisant partie des membranes de plusieurs espèces bactériennes, en particulier celles du genre *Burkholderia* (Kannenberg and Poralla, 1999; Pearson *et al*., 2007). Semblables au cholestérol chez les organismes eucaryotes, les hopanoïdes participent à la rigidité membranaire et également à la résistance aux AMPs chez *B. cenocepacia* et *B. multivorans* (Malott *et al.*, 2012; Schmerk *et al*., 2011). Le dernier composant membranaire étudié correspond au facteur RpoE faisant partie des mécanismes d'ESR, également appelé facteur σ^E (Flores-Kim and Darwin, 2014; Guest and Raivio, 2016). Suite à un dommage membranaire causé lors d'un stress tel que la présence d'AMPs, une cascade d'enzymes protéolytiques s'enclenche pour libérer le facteur de transcription RpoE maintenu dans un état inactif à la face cytoplasmique de la membrane interne, et va ainsi activer la transcription des gènes impliqués dans la réparation de la paroi bactérienne (Flores-

Kim et Darwin, 2014; Guest et Raivio, 2016). En outre, le facteur RpoE a été décrit chez *B. cenocepacia* pour son rôle dans la résistance aux AMPs ainsi que pour le maintien de la paroi bactérienne face à de fortes températures (Flannagan et Valvano, 2008). Nous avons ainsi sélectionné trois gènes impliqués dans la biosynthèse du « core oligosaccharide » du LPS (*waaC*, *waaF*, *wbiF*), cinq gènes participant à la voie de biosynthèse des hopanoïdes (*shc*, *hpnA*, *hpnH*, *hpnJ*, *hpnN*), et deux gènes impliqués dans la réponse aux stress extracellulaires (ESR) de type RpoE (*rpoE*, *mucD*). Comme les AMPs représentent une famille très large de peptides dotés de propriétés physicochimiques différentes et produits par divers organismes, nous avons tenu à tester plusieurs AMPs pour évaluer les capacités de résistance de *B. insecticola*. Pour cela, nous avons choisi cinq AMPs dont : la polymyxine B, couramment utilisée pour traiter des cas cliniques de maladies bactériennes (Cai *et al*., 2015) ; le LL-37, également appelé cathélicidine, produite par les polynucléaires neutrophiles chez l'homme (Fabisiak *et al.*, 2016) ; ainsi que trois AMPs produits par l'insecte-hôte *R. pedestris* comprenant la riptocine (Kim *et al.*, 2016a), et deux peptides CCR (CCR179 et CCR480) (Futahashi *et al.*, 2013). En testant l'effet de ces AMPs sur les mutants de *B. insecticola* pour les gènes cités précédemment, il s'avère que seules les souches mutées pour la biosynthèse du core oligosaccharide du LPS étaient hypersensibles à tous les AMPs, ainsi que le mutant *rpoE* qui était sensible uniquement à la riptocine.

En réalisant l'approche Tn-seq en présence de ces cinq AMPs, j'ai identifié 42, 42, 15, 21 et 39 gènes requis respectivement pour la fitness bactérienne en présence de la polymyxine B, du LL-37, de la riptocine, du peptide CCR179 et du peptide CCR480. Parmi tous ces facteurs requis pour la résistance aux AMPs, seulement trois gènes étaient communément retrouvés pour ces cinq AMPs codant pour les sous-unités du transporteur Tat ou « twin-arginine transporting system » connu pour participer à la stabilité de la membrane externe (Robinson *et al*., 2011a). Concernant les autres gènes identifiés, la plupart d'entre eux codent pour des composants de la paroi bactérienne incluant les voies de biosynthèse du « core oligosaccharide » et de l'antigène O du LPS, de protéines associées à la membrane externe, d'un transporteur de l'antigène O à travers la membrane interne, ainsi que du système de transport Tol-Pal. Sachant que les AMPs ciblent majoritairement les membranes bactériennes (Kumar *et al.*, 2018), il n'était pas surprenant de trouver que la biosynthèse de la paroi était la catégorie fonctionnelle la plus représentée parmi tous les gènes requis pour la résistance aux AMPs. Pour confirmer

ces résultats Tn-seq, j'ai choisi cinq gènes identifiés pour plusieurs AMPs en ciblant le transporteur Wzm/Wzt de l'antigène O (*wzm*) (Ortega *et al.*, 2005), la biosynthèse de dTDP-L-rhamnose qui est un des composants de l'antigène O (*rfbA*, *rfbC*) (Tsukioka *et al.*, 1997), et le transporteur Tol-Pal (*tolB*, *tolQ*) (Lloubès *et al*., 2001). En générant des souches de *B. insecticola* mutées pour ces cinq gènes, j'ai constaté que ces cinq mutants étaient tous hypersensibles aux AMPs, validant ainsi les analyses Tn-seq précédemment obtenues.

Afin d'évaluer la participation de ces facteurs bactériens étudiés au cours de ces deux approches dans la colonisation de l'insecte-hôte, j'ai réalisé des mono-infections avec toutes ces souches sur de jeunes insectes au second stade larvaire. Sachant que toutes les souches mutantes construites exprimaient la GFP, j'ai pu observer leur devenir dans l'organe symbiotique et j'ai également dénombrer la population bactérienne présente au sein des cryptes. J'ai découvert que les mutants présentant une sensibilité aux AMPs présentaient tous des défauts de colonisation plus ou moins sévères avec : des souches présentant une incapacité totale à coloniser l'insecte-hôte (*waaC*, *waaF*, *rfbA*), des souches pouvant coloniser partiellement l'organe symbiotique (*tolB*, *tolQ*) et des souches capables d'infecter de manière efficace seulement une partie de la population d'insectes (*wbiF*, *rfbC*, *wzm*). En effectuant des expériences de compétitions avec la souche sauvage *in vivo*, je me suis rendue compte que toutes ces souches mutantes présentant des défauts de colonisation étaient moins compétitives que la souche sauvage, et donc devenaient incapables de coloniser l'organe symbiotique en présence de la souche sauvage. Au cours de ce travail, j'ai pu encadrer une étudiante de Master 2, Christy Calif, qui a continué de travailler sur ces données Tn-seq en obtenant trois autres mutants de *B. insecticola* pour les gènes *tatB*, *dsbA* et *mlaD*. Le gène *tatB*, identifié communément pour les cinq AMPs testés, code pour une des trois sous-unités du transporteur Tat (Robinson *et al.*, 2011a). Le gène *dsbA* code pour une protéase périplasmique impliquée dans le contrôle qualité des protéines (Manta *et al.*, 2019), et a été identifié spécifiquement en présence de la polymyxine B. Enfin, le gène *mlaD* code une des sous-unités d'un transporteur ABC impliqué dans l'export de phospholipides, jouant un rôle dans le maintien de la membrane externe (Bernier *et al.*, 2018). En effectuant des mono-infections et des co-infections d'insectes avec ces trois mutants, il s'est avéré qu'aucun d'entre eux n'était capable d'infecter efficacement l'organe symbiotique de *R. pedestris*. Bien que ces mutants soient toujours en cours de validation, notamment en ce qui concerne leurs

phénotypes vis-à-vis des AMPs, il a été montré que le mutant *dsbA* était hypersensible uniquement en présence de la polymyxine B, validant ainsi l'identification de ce gène par Tn-seq. Ainsi l'ensemble de ces résultats suggèrent qu'il existe bien un lien entre les facteurs de résistance aux AMPs chez *B. insecticola* et la capacité de coloniser efficacement l'insecte-hôte. Comme cette symbiose apporte des effets bénéfiques d'ordre morphologique et développemental à l'insecte-hôte (Kikuchi *et al.*, 2007), je me suis également intéressée à l'impact que pouvaient avoir ces souches mutantes sur ces différents paramètres chez *R. pedestris*. Un des résultats les plus intéressants que j'ai trouvé est que les insectes infectés avec le mutant *mucD*, capable de coloniser pleinement l'organe symbiotique, présentent tous des caractéristiques morphologiques d'insectes aposymbiotiques. Cependant, leur temps de développement rapide est semblable à celui des insectes symbiotiques, ce qui indique que le temps de développement jusqu'à l'âge adulte et la croissance des insectes sont deux processus indépendants. Pour les insectes infectés par le mutant *mucD*, ces observations suggèrent que ces insectes ont un défaut de croissance probablement dû à des déficits métaboliques chez ce mutant qui ne fournit plus de nutriments essentiels pour assurer le bon développement de son hôte. Concernant les mutants ne pouvant pas coloniser l'organe symbiotique (*waaC*, *waaF* et *rfbA*), les insectes infectés par ces souches présentent bien des caractéristiques aposymbiotiques. Cependant, le temps de développement et la morphologie des insectes infectés par des souches présentant des phénotypes intermédiaires de colonisation de l'organe symbiotique étaient plutôt semblables à ceux d'insectes infectés par la souche sauvage.

4. Les fonctions symbiotiques impliquées dans la colonisation de *R. pedestris* par *B. insecticola* ont été identifiées par Tn-seq *in vivo*

Dans ce dernier volet de mon travail de thèse, je me suis focalisée sur la mise en place d'une expérience de Tn-seq *in vivo* chez *R. pedestris* afin d'identifier directement les gènes bactériens requis pour la colonisation de l'organe symbiotique. L'anatomie de l'intestin de *R. pedestris* révèle qu'il existe un passage très étroit pour atteindre l'organe symbiotique (région M4) depuis l'organe précédent (région M3), s'appelant la région resserrée ou « constricted

region » (CR) (Ohbayashi *et al.*, 2015). Cette zone étroite constitue une barrière anatomique et physicochimique, ou un goulot d'étranglement à franchir par la population symbiotique (Ohbayashi *et al.*, 2015), qui peut entraîner une réduction du nombre de bactéries pouvant réussir à coloniser initialement l'organe symbiotique avec succès. Ainsi avant de procéder au Tn-seq *in vivo*, j'ai évalué l'ampleur du goulot d'étranglement imposé par l'anatomie de l'insecte chez *B. insecticola* en utilisant l'approche Tn-seq. En infectant indépendamment soixante insectes avec la banque Tn-seq de *B. insecticola* à hauteur de 10^6 bactéries par insecte, j'ai découvert qu'environ 10000 bactéries en moyenne par insecte étaient capables d'infecter l'organe symbiotique. Ces résultats indiquent que l'ampleur du goulot d'étranglement est bien à prendre en compte pour mettre en place des expériences de Tn-seq *in vivo*. A partir de ces résultats, j'ai calculé qu'il faudrait utiliser une centaine d'insectes par réplicat biologique pour s'assurer d'une bonne représentativité de la banque Tn-seq dans l'organe symbiotique (environ 10^6 mutants), et pour avoir également suffisamment de matériel génomique pour le séquençage.

Une fois que l'effet du goulot d'étranglement sur la population symbiotique a été déterminé, j'ai pu réaliser une expérience Tn-seq *in vivo* en étudiant plusieurs organes intestinaux à différents stades larvaires. En effet, j'ai prélevé la région M4 au second et au troisième stades larvaires, afin d'évaluer l'effet de la première mue de l'hôte sur la dynamique de cette population symbiotique. J'ai également prélevé deux autres organes intestinaux M1 et M3 au deuxième stade larvaire, organes qui précèdent la région M4 et constituent des zones transitoires du passage de la population symbiotique lors de la colonisation. En comparant ces conditions *in vivo* avec la condition en milieu riche avec l'analyse Con-ARTIST, j'ai trouvé 37, 18, 129 et 329 gènes requis respectivement pour la colonisation des organes M1, M3, M4 au second stade larvaire et M4 au troisième stade larvaire. En me focalisant sur les fonctions biologiques codées par tous ces gènes identifiés pour la colonisation de l'organe symbiotique, il s'avère qu'elles peuvent être classées dans cinq grandes catégories fonctionnelles : les mécanismes de réparation de l'ADN, le métabolisme, les réponses au stress, la biosynthèse de la paroi bactérienne, et la motilité. Concernant la réparation de l'ADN, j'ai identifié des mécanismes de recombinaison homologue comme la réparation de jonctions Holliday et la réquisition de l'exodéoxyribonucléase de type V impliquée dans la réparation des cassures de l'ADN double brin (Lohman et Fazio, 2018; Wardrope *et al.*, 2009). Diverses activités

métaboliques étaient aussi identifiées comme la biosynthèse des acides nucléiques, faisant le lien avec la réparation de l'ADN, la biosynthèse de différents acides aminés essentiels comme la méthionine ou l'arginine, les voies de biosynthèse des vitamines B6 et B12, et également la voie des pentoses-phosphates pour la glycolyse. De plus, j'ai identifié des transporteurs d'éléments inorganiques tels que le magnésium (Maloney et Valvano, 2006) et le zinc (Gabbianelli *et al.*, 2011) faisant également partie des activités métaboliques requises pour *B. insecticola* au sein de la région M4. Comme je l'ai mentionné, il y a aussi différents mécanismes de réponse au stress qui s'avèrent nécessaires pour le symbiote dans la région M4 comme les processus de contrôle qualité des protéines avec différentes protéases pour dégrader les protéines mal repliées (Aertsen *et al.*, 2004; Manta *et al.*, 2019; Seol *et al.*, 1991), ou l'accumulation de tréhalose pour lutter contre la pression osmotique (Joseph *et al*., 2010). En plus de ces réponses au stress, j'ai également détecté de nombreux composants de la paroi bactérienne tels que les voies de biosynthèse du LPS, notamment pour le « core oligosaccharide » et l'antigène O (Loutet *et al.*, 2006), ainsi que des systèmes de transport comme le complexe Tol-Pal (Lazzaroni *et al.*, 1999) et le système Tat (Robinson *et al*., 2011a). Enfin, il s'avère que la fonction biologique la plus prédominante identifiée pour la région M4 est celle de la motilité bactérienne, notamment impliquant toutes les sous-unités du flagelle (Rajagopala *et al.*, 2007) et aussi de nombreuses protéines impliquées dans le chimiotactisme (Baker *et al.*, 2006).

En partant des résultats de l'analyse Tn-seq, j'ai remarqué que le nombre de gènes identifiés pour la colonisation de la région M4 au troisième stade larvaire était bien plus conséquent qu'au second stade larvaire. Une grande majorité de ces gènes additionnels étaient localisés sur le plasmide 2 (NC_021295.1), ce qui indiquait que le plasmide 2 devenait important pour *B. insecticola* durant le troisième stade larvaire de l'hôte. Cependant, d'autres résultats au laboratoire ont montré qu'une partie de la population symbiotique perdait le plasmide 2 à partir du troisième stade larvaire de l'insecte (données non publiées). La perte de ce plasmide signifie donc qu'il ne sera pas séquencé lors du Tn-seq, et donc sa perte est interprétée comme un critère d'essentialité avec l'analyse Tn-seq. Ainsi les résultats Tn-seq corroborent les précédentes observations réalisées au laboratoire, et suggèrent que les bactéries symbiotiques perdent leur plasmide 2 durant le passage au troisième stade larvaire.

Comme notre hypothèse s'appuie sur le fait que la résistance aux AMPs est une caractéristique

permettant de coloniser spécifiquement l'organe symbiotique de *R. pedestris* par *B. insecticola*, j'ai donc comparé les facteurs bactériens identifiés précédemment pour la résistance aux AMPs avec la liste des gènes requis pour la colonisation de la région M4, tous identifiés par Tn-seq. J'ai trouvé que près de 28% des gènes identifiés pour la colonisation de l'organe symbiotique participaient également à la résistance aux AMPs. Donc la capacité de résister aux AMPs est une condition requise et très importante pour *B. insecticola* pour pouvoir coloniser efficacement l'organe symbiotique de *R. pedestris.*

5. Conclusions et perspectives

Durant ce travail de thèse, j'ai mis au point l'approche Tn-seq chez *B. insecticola* grâce à laquelle j'ai pu identifier le génome essentiel de ce symbiote, qui représente 17% du génome total. En comparant ce génome essentiel avec ceux de trois autres espèces de *Burkholderia*, j'ai ainsi pu identifier un pool de gènes essentiels attribué au genre *Burkholderia*. En utilisant deux sources de carbones différentes, j'ai pu identifier des gènes impliqués dans l'exploitation de ces molécules, ce qui m'a permis de valider l'approche Tn-seq mise en place au laboratoire pour *B. insecticola*.

Grâce à l'approche Tn-seq, j'ai pu déterminer les facteurs bactériens impliqués dans la résistance à cinq AMPs différents. Les facteurs de résistance aux AMPs identifiés chez *B. insecticola* sont majoritairement des composants membranaires tels que le LPS, les transporteurs Tat et Tol-Pal, ainsi que de nombreuses autres protéines de la membrane externe. J'ai ainsi pu valider plusieurs de ces gènes identifiés par Tn-seq pour leur rôle dans la résistance aux AMPs. D'autres gènes sont également en cours de validation au laboratoire, en ciblant notamment les trois gènes en commun entre les cinq AMPs qui représentent les sous-unités du transporteur Tat. J'ai pu montrer que les mutants de *B. insecticola* hypersensibles aux AMPs présentaient également un défaut de colonisation de l'organe symbiotique de *R. pedestris*. Les expériences de compétitions *in vivo* se sont révélées être pertinentes pour décrire les phénotypes de colonisation dans un contexte de co-infections, reflétant ce qui pourrait se produire dans le cadre de l'acquisition de plusieurs bactéries dans l'environnement de l'insecte. En m'intéressant aux caractéristiques physiques des insectes infectés par ces souches, j'ai pu conclure que le temps de développement et la croissance des différentes parties anatomiques de l'insecte étaient deux phénomènes découplés au cours du

développement de l'insecte.

En déterminant l'effet du goulot d'étranglement appliqué par l'insecte sur la population symbiotique, j'ai pu réaliser une expérience de Tn-seq *in vivo* me permettant d'identifier les facteurs bactériens requis pour la colonisation de l'organe symbiotique. Cinq grandes fonctions biologiques s'avèrent jouer un rôle important au cours de la symbiose, que sont la motilité, la paroi bactérienne, les réponses aux stress, la réparation de l'ADN et différentes voies métaboliques. Il apparaît que la voie de glycolyse Entner-Doudoroff est la voie de dégradation prioritaire utilisée par *B. insecticola* en culture *in vitro*, alors que la voie des pentoses-phosphates devient la voie de dégradation préférentiellement choisie par *B. insecticola* durant son maintien dans l'organe symbiotique de *R. pedestris*. Toutes ces fonctions identifiées suggèrent que l'organe symbiotique constitue un environnement stressant pour le symbiote, notamment via des stress de type oxydatif, osmotique, et également par la présence d'AMPs. Il s'avère qu'il existe une corrélation entre les facteurs impliqués dans la résistance aux AMPs et les facteurs requis pour la colonisation de l'organe symbiotique de l'insecte-hôte chez *B. insecticola*. En effet, près de 28% des facteurs symbiotiques sont aussi dédiés à la résistance aux AMPs. Cependant, les autres fonctions symbiotiques identifiées suggèrent que la résistance aux AMPs n'est pas la seule caractéristique nécessaire pour expliquer cette spécificité de sélection pour *B. insecticola*.

Ce travail de thèse amène de nombreuses perspectives, notamment dans la confirmation des résultats Tn-seq obtenus pour la colonisation de l'organe symbiotique. Il sera indispensable de créer des mutants de *B. insecticola* pour plusieurs gènes prédits comme participant à la symbiose afin de valider leur implication dans la colonisation de l'insecte-hôte. L'étude du rôle du plasmide 2 est également en cours au laboratoire. Parmi les fonctions symbiotiques identifiées, il s'avère que le chimiotactisme semble jouer un rôle primordial dans la colonisation de l'hôte. Il serait donc envisageable et intéressant d'identifier les molécules chimio-attractives produites par *R. pedestris* pour guider la population symbiotique de *B. insecticola* dans la région M4. De plus, maintenant que le Tn-seq *in vivo* est réalisable chez *R. pedestris*, il serait possible d'identifier les facteurs symbiotiques requis pour la colonisation de l'organe symbiotique durant les autres stades larvaires, y compris l'âge adulte. Les données Tn-seq générées pourraient compléter la liste des facteurs symbiotiques déjà obtenus durant cette étude, et ainsi obtenir une vision globale et dynamique des fonctions bactériennes

nécessaires pour le maintien de la population symbiotique sur le long terme. De plus, en ce qui concerne les caractéristiques aposymbiotiques des insectes infectés par le mutant *mucD*, il serait possible d'étudier le métabolome de *R. pedestris* ainsi que de *B. insecticola* afin de découvrir quelles sont les molécules échangées par ces deux partenaires au cours de la symbiose. Sachant que de telles expériences ont déjà été réalisées chez l'abeille et son microbiote intestinal (Kešnerová *et al.*, 2017; Zheng *et al*., 2017), il serait envisageable de réaliser de telles expériences sur le modèle *Riptortus-Burkholderia*.

Table des matières

Remerciements.....

Liste des abréviations.....

Résumé de la Thèse.....

Chapter I: General Introduction 1

1. Symbiosis.....2

1.1 Concept 2

1.2 Symbiosis models.....3

1.3 Insect symbioses 5

2. The *Riptortus pedestris-Burkholderia insecticola* symbiosis 9

2.1 The stinkbug *Riptortus pedestris*.....9

2.2 The symbiosis of Pentatomomorpha insects..... 12

2.3 The symbiosis of *Riptortus pedestris* 17

2.4 *Burkholderia insecticola* symbiotic functions 20

3. *Burkholderia* species 24

3.1 Classification and phylogeny 24

3.2 Pathogenic *Burkholderia* species 26

3.3 Environmental and beneficial *Burkholderia* species 27

3.4 Antimicrobial resistance in *Burkholderia*..... 29

4. Antimicrobial peptides (AMPs) 34

4.1 Classes of AMPs 34

4.2 Modes of action of AMPs 37

4.3 Antimicrobial peptides and symbiosis 38

4.4 Immunity and antimicrobial peptides in *R. pedestris* 42

Thesis objectives 48

Chapter II: Inferring the Essential Genome of *Burkholderia insecticola* by Transposon Sequencing 49

1. Introduction..... 50

2. Contributions..... 52

3. Materials and methods 53

4. Results 57

4.1 Transposon mutagenesis of the *Burkholderia* symbiont..... 57

4.2 Genome-wide screening of essential functions in the *B. insecticola* symbiont 60

4.3 Identification of genes for growth on succinate and glucose as carbon sources 65

4.4 Comparative transposon mutagenesis on *Burkholderia* species 73

5. Discussion 75

Chapter III: Identification of Factors Involved in Antimicrobial Peptide Resistance in the *Burkholderia insecticola* Symbiont 81

1. Introduction 82

2. Contributions 84

3. Materials and methods 85

4. Results 91

4.1 Candidate gene approach 91

4.1.1 LPS biosynthesis genes 91

4.1.2 Hopanoids 102

4.1.3 Extracellular Stress Response 110

4.2 Tn-seq approach 115

4.2.1 Bacterial functions involved in global AMP resistance 121

4.2.2 Resistance factors for specific AMPs 131

4.2.3 Validation of bacterial genes involved in AMP resistance 135

4.3 Host colonization competitions experiments 147

4.4 Host fitness parameters in infections with AMP-sensitive mutants 158

5. Discussion 166

Chapter IV: Host Colonization Functions of *Burkholderia insecticola* Identified by *in vivo* Tn-seq 177

1. Introduction 178

2. Contributions 181

3. Materials and methods 182

4. Results 185

4.1 Bottleneck size estimation 185

4.2 Dynamics of the *B. insecticola* population in the *R. pedestris* midgut 191

4.3 Bacterial symbiotic functions 193

4.3.1 DNA repair, transcription and translation modulations 201

4.3.2 Stress response elements 204

4.3.3 Metabolism 208

4.3.4 Envelope biogenesis functions 216

4.3.5 Motility and chemotaxis 222

4.3.6 Plasmid 2 genes 225

4.4 Correlation between symbiosis factors and AMP resistance 234

5. Discussion 244

Chapter V: General Discussion and Perspectives 251
References 263
Publications 301
Annexes 319
Annexe 1 320
Annexe 2 366
Annexe 3 371
Annexe 4 376
Annexe 5 382
Annexe 6 411
Annexe 7 415
Annexe 8 418
Annexe 9 422
Annexe 10 426
Annexe 11 430
Annexe 12 432
Annexe 13 433
Annexe 14 436
Annexe 15 439
Annexe 16 442
Annexe 17 450

Chapter I

General Introduction

1. Symbiosis

1.1. Concept

Most species of life's diversity, if not all, are colonized by microbial communities, which closely interact with their host. Such intimate and long-term interactions are referred to as symbiosis, a term which was defined for the first time by de Bary in 1879 as "the living together of unlike organisms" (de Bary, 1879; Oulhen *et al.*, 2016). As this definition suggests, symbiosis involves two different organisms and engulfs the three main kinds of relationships known as mutualism, commensalism and parasitism. These states depend on the interaction context, whether the fitness of one organism is positively or negatively affected by the other: in mutualistic symbiosis, both organisms exchange mutual benefits in order to survive; for commensalism, only one organism is positively affected without harming the second partner; and finally, parasitic relationships are illustrated by pathogens which take advantage of their host by causing severe damages (López-García *et al.*, 2017). However, symbiosis is often confounded with mutualism in the literature.

Most symbioses are interactions of microbial symbionts, such as bacteria, archaea and fungi, with their eukaryotic host. These symbionts are categorized either as facultative or obligate partners to sustain the host's life, and can be involved in ecto- or endosymbiotic interactions depending on their physical localization on the host (Zilber-Rosenberg and Rosenberg, 2008). Two modes of transmission are possible to maintain symbiosis over generations: horizontal and vertical transmissions (Moran, 2006). In the case of horizontal transmission, the host is exposed to a large spectrum of environmental microbes and needs to acquire its symbiont through a selective mechanism to promote a high specificity of colonization in each successive generation (Moran, 2006). In contrast, vertical transmission or transmission from mother to offspring, ensures maintenance of the coevolved symbiont in every generation of the host with a strong stability (Moran, 2006).

During symbiosis, the host and the microbial symbionts communicate with each other and exchange several services. The microbial communities contribute to many host metabolic processes: they provide essential nutrients, degrade recalcitrant food components or recycle waste molecules produced by the host; they can participate to the protection against pathogenic invasion due to the niche occupation or by inter-specific competitions; they

stimulate the host immune system; and they promote the host's development and fitness (Engel and Moran, 2013; Hooper and Gordon, 2001; Pickard *et al.*, 2017). On the symbiont side, microbial symbionts living inside the host are protected against competitive environmental microorganisms to get resources including nutrients, and can thus grow efficiently without competitors (Garcia and Gerardo, 2014; Wier *et al.*, 2010).

1.2. Symbiosis models

As symbioses are widespread among living organisms, different biological models were investigated to seek how symbiosis is maintained and regulated by both partners, or in other terms, to identify the mechanisms drive the interplay between the host and its symbiotic population.

In plants, well studied models are the *Rhizobium*-legume and the *Frankia*-actinorhizal plant symbioses. The *Rhizobium*-legume symbiosis plays a critical role in land ecosystems with nitrogen-limited soils where the legume partner, belonging to the *Fabaceae* family, interacts with Gram-negative α- and β-proteobacteria called α- and β-rhizobia, respectively (Kondorosi *et al.*, 2013). In this nutritionally limiting condition, the symbiont provides ammonia through nitrogen fixation, thus promoting the host plant growth. As these two organisms can live independently from each other, the microbial symbiont needs to be acquired from the soil environment when the plant needs additional nitrogen. This horizontal transmission relies on a specific molecular dialog between the two partners in order to enable the plant to select specifically the compatible *Rhizobium* bacteria. The legume plant produces flavonoids detected by the symbiont, which triggers the production of diffusible host-specific signals called the Nod factors (Kondorosi *et al.*, 2013). These bacterial signals are lipochitooligosaccharides that initiate the bacterial infection of the plant roots (Kondorosi *et al.*, 2013). In parallel with the infection and at the sites of infection, the Nod factors also trigger the formation of a specific symbiotic organ called the root nodule. In these organs, large numbers of the symbiotic bacteria are trapped within intracellular compartments termed symbiosomes (**Figure 1**). These intracellular rhizobia differentiate into a nitrogen-fixing form called bacteroids. In some host plants, this endosymbiotic lifestyle results in a terminal bacterial differentiation, an irreversible process where the symbiotic bacteria are unable to return to their free-living state (Mergaert *et al.*, 2006). The possibility of genetic manipulations

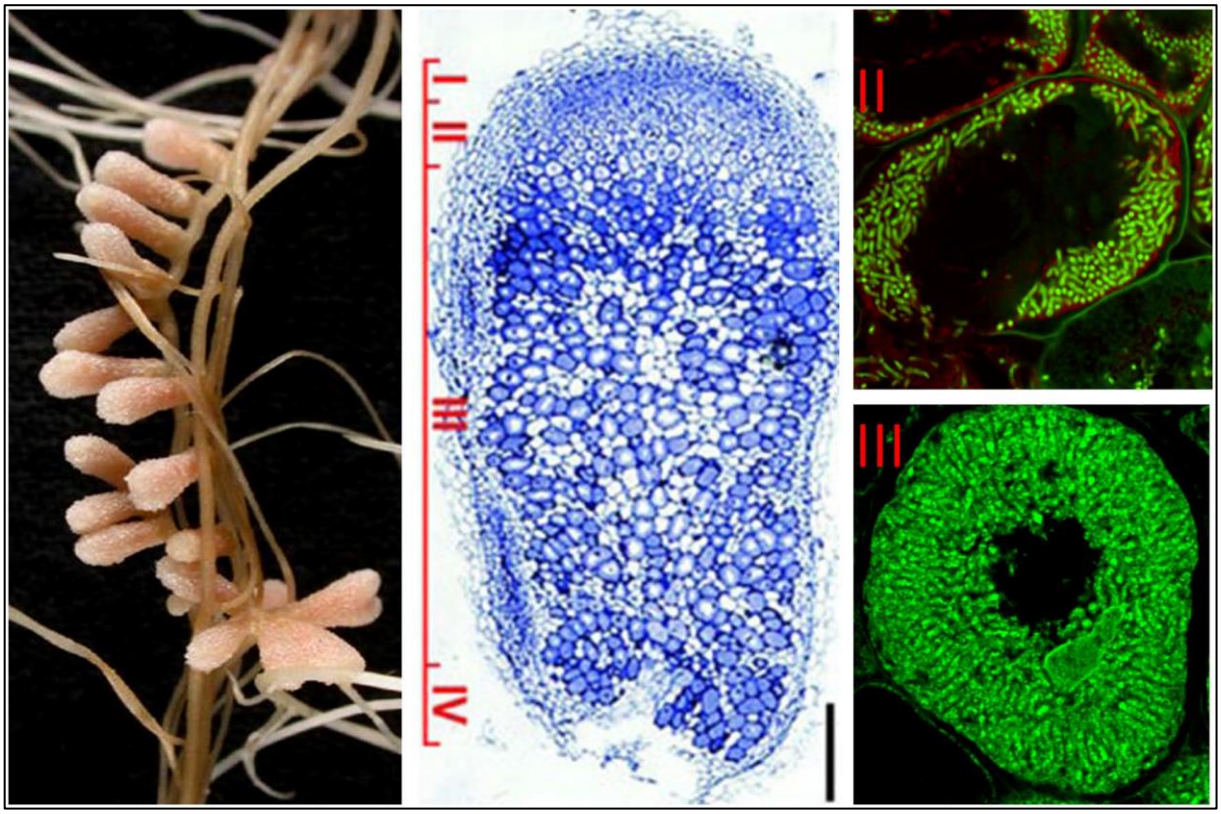


Figure 1: *Rhizobium*-legume symbiosis.
Nodules observed on *Medicago truncatula* roots, with different nodule zones (I to IV) indicated in a longitudinal section. Symbiotic *Sinorhizobium meliloti* cells are visible with green fluorescence inside the nodule structure (Syto9 staining).
Taken from *Maróti and Kondorosi, 2014*.

of both the *Rhizobium* species and the host plants, both with available genome sequences (Krishnakumar *et al.*, 2015; Reeve *et al.*, 2010), has made this interaction one of the best understood symbioses at the molecular level.

A similar symbiotic signalling, infection and organogenesis mechanism is used by the Gram-positive *Frankia* species to colonize their hosts known as actinorhizal plants, such as *Alnus* sp. (Froussart *et al.*, 2016). Even if the generation of *Frankia* defective mutants is not yet possible, silencing host plant genes with RNAi (Clavijo *et al.*, 2015) and host transcriptomic analyses are useful tools available to decipher the molecular relationship between these two partners.

In the animal kingdom, one of the most interesting symbiosis models is the association between the Hawaiian bobtail squid *Euprymna scolopes* and *Vibrio fischeri* (Mandel and Dunn, 2016). In this mono-interaction, the bioluminescent bacterium *V. fischeri* is retrieved from the seawater by its host quickly after birth, and stored extracellularly inside its light organ (McFall-Ngai, 2014). The bioluminescence produced by the symbiont provides a specific camouflage

for the host known as counterillumination, which has an antipredatory function. Thus in this particular case, the symbiosis confers a behavioural advantage to the host, while most of the known symbiotic associations lead to nutritional gains like in the *Rhizobium*-legume interaction. Similar to the *Rhizobium*-legume symbiosis, the horizontally-transmitted *Vibrio* symbiont needs to be specifically selected by its host from the diverse microbial species present in the seawater. The ability to manipulate the host through all stages of development, coupled with recent genomic analyses of several *V. fischeri* strains (Bongrand *et al.*, 2016; Gyllborg *et al.*, 2012; Ruby *et al.*, 2005) are useful tools to study extensively this priming molecular dialogue (Pankey *et al.*, 2014). Other model interactions of animals with specific bacteria are found in insects. These symbioses are discussed in detail in the following section (see **section 1.3**).

Unlike the monospecific associations in these biological models, animals and plants are generally colonized by complex microbial communities, constituting specific microbiota. Hence, multiple models arose that are under intensive investigation, notably in mammals and in social insects. In humans, the gut intestinal tract represents one of the largest interfaces between the host and its microbiota. There are ten times more microbial cells in the intestinal tract than human cells in the whole body (Thursby and Juge, 2017). This microbiota is constituted of up to 10^{14} bacterial cells, of which more than 90% of the bacterial species belong to the Proteobacteria, Firmicutes, Bacteroidetes and Actinomycetes, with a strong prevalence of Bacteroidetes and Firmicutes species (Chow *et al.*, 2010; Thursby and Juge, 2017).

1.3. Insect symbioses

The development and survival of many insects strongly depend on beneficial microorganisms, especially symbiotic bacteria, which are able to colonize their gut, tissues and cells. For example, *Wolbachia* endosymbionts can infect more than half of all insect species (Sazama *et al.*, 2019). The study of resident gut microbiota of insects is of a major interest, notably because equivalent interactions can be found in mammals, such as the human gut microbiota. Even so, these two animal groups differ in their microbial diversity, which tends to be much more complex in mammals than in insects. Hence, the insects' gut microbial community generally consists of few taxa (less than 20-30) compared to the mammal gut microbial

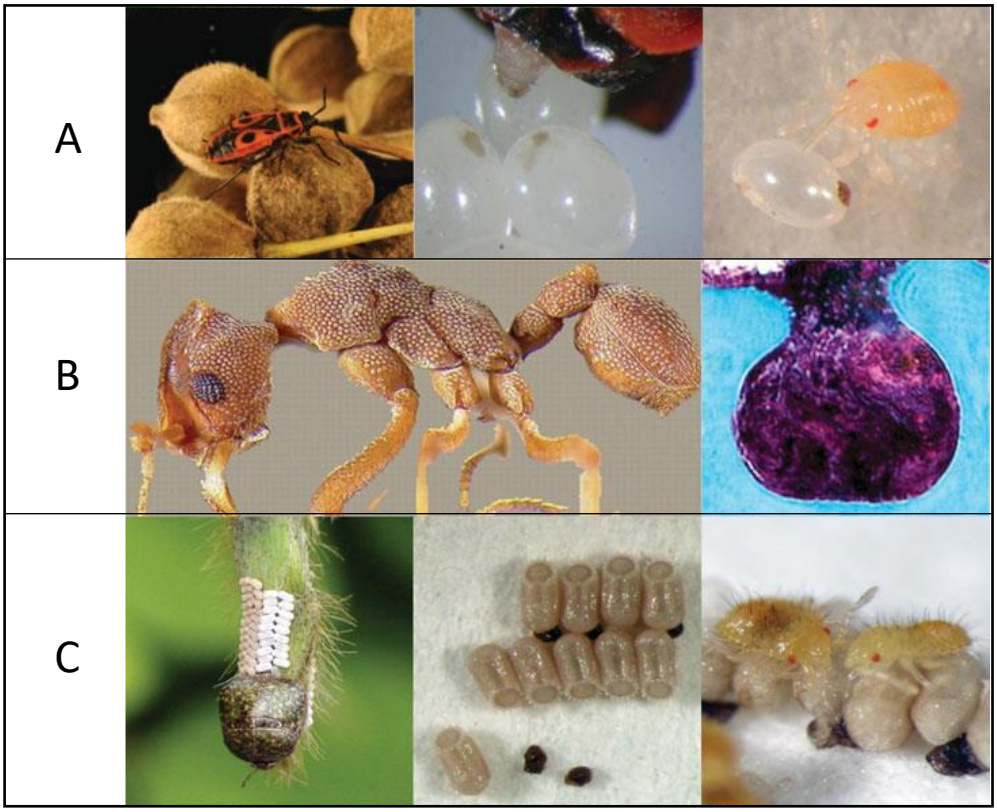


Figure 2: Examples of vertical transmission mechanisms in insects.
A) Egg smearing in the European firebug. B) Social behaviour in fungus-growing ants with symbiotic bacteria present in specific regions of their cuticle. C) Capsules deposition ingested by hatched insects in plataspid stinkbugs.
Adapted from *Salem et al., 2015*.

community containing typically 500-1000 taxa. However, as the phylogenetic diversity of insects is greater than in mammals, more bacterial taxa are able to colonize the gut of insects (Douglas, 2011). Besides the gut, another common habitat for microbial symbionts in insects is cells. About 10-20% of insects have intracellular symbionts localized in specialized cells, called bacteriocytes, whose only function is to house and maintain these symbionts. In addition, microorganisms have been described in cells of various organs, including the fat body, gut epithelium, and gonads. Some of these bacteria (*e.g., Wolbachia*, *Hamiltonella*) can occupy multiple compartments, within and between the cells of insect organs and in the hemolymph. Most of these bacterial associations contribute to the insect diet and help them to degrade complex food components or alternatively they produce metabolites that are lacking in the diet. For example, the wood-feeding termites harbour a specific symbiotic community which participates to lignocellulose digestion, and provides nitrogen and carbon sources to complement its host nutritional diet (Engel and Moran, 2013). Phytophagous

insects or blood-sucking insects have a specific diet that often lack essential nutrients. These nutrients are synthesized by the bacterial symbionts of the insects, like *Rhodococcus rhodnii* which provides B vitamins to its blood-feeding host *Rhodnius prolixus* (Eichler and Schaub, 2002). Additionally, symbiotic interactions can play a protective role in parasitic infections, such as facultative symbionts of aphids which protect their host against parasitoid wasps (Oliver *et al.*, 2010). Symbiotic bacteria have also the ability to manipulate the host's reproductive system, like *Wolbachia* and *Spiroplasma* species which induce male-killing in the offspring of their *Drosophila* host (Werren *et al.*, 2008).

As mentioned earlier, there are two ways to ensure symbiont transmission across generations, either via horizontal or vertical transmission. In insect species, vertical transmission is very common and can be accomplished through several mechanisms such as coprophagy, where the offspring acquire their symbionts through probing on their mother's faeces (termites, cockroaches); egg smearing, when the mother spreads a superficial layer of symbiotic bacteria directly on the eggs (firebugs, shield bugs); capsule formation, with deposition of capsules next to the eggs containing symbiotic bacteria which are eaten by the offspring (plataspid bugs); and transovarial transfer in which the ovaries are infected with symbionts inside the female body (aphids) (Salem *et al.*, 2015). Additionally, social behaviours strongly contribute to the transfer of mutualistic microbial communities, known as trophallaxis, and has been described in ants, bees and termites (**Figure 2**) (Onchuru *et al.*, 2018; Salem *et al.*, 2015).

Among these vertically-transmitted interactions, multiple bacterial symbiotic systems are well studied such as *Buchnera* and aphids, and *Wigglesworthia* associated with tse-tse flies. One of the best-studied model for insect symbiosis is the pea aphid (*Acyrthosiphon pisum*) with its intracellular endosymbiont *Buchnera aphidicola* (Shigenobu and Wilson, 2011). Similar to the *Rhizobium*-legume symbiosis, these obligate γ-proteobacteria are stored inside the specialized bacteriocytes, which can house tens of thousands bacterial cells. These bacteriocytes are grouped in bilobed organ-like structures named bacteriomes, thus representing the symbiotic organs (**Figure 3**) (Wilson and Duncan, 2015). In addition, aphids can acquire facultative symbionts, like *Hamiltonella defensa*, *Regiella insecticola* and *Serratia symbiotica*, but these symbionts colonize different insect tissues or distinct cells within the bacteriomes (Koga *et al.*, 2012). Aphids are major plant-sucking crop pests, hence they feed on plant phloem sap which lacks essential nutrients. The *Buchnera* symbiont completes the nutritional requirements of

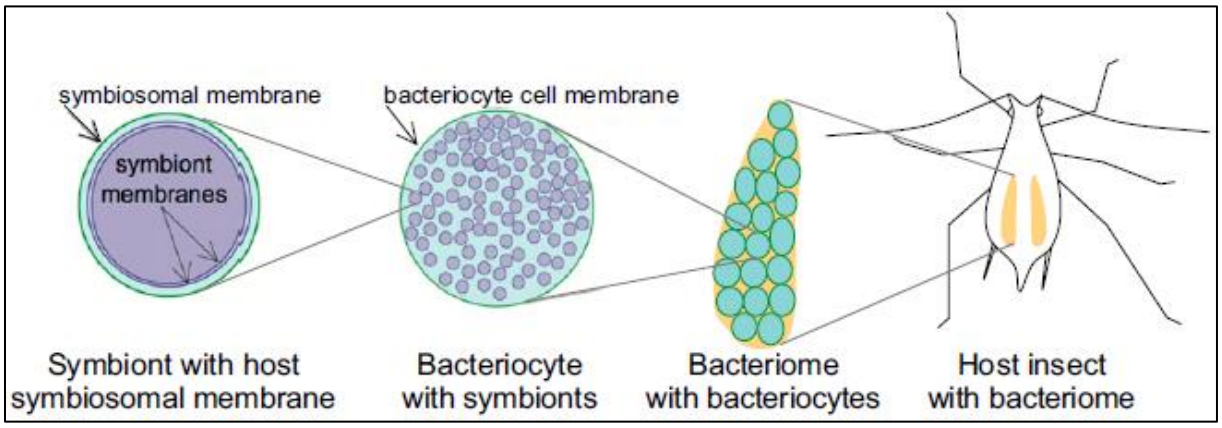


Figure 3: *Buchnera aphidicola*-aphids symbiosis organization.
Taken from *Wilson and Duncan, 2015*.

its host by providing essential amino acids, which cannot be synthesized *de novo* by aphids (Hansen and Moran, 2011). On the bacterial side, *B. aphidicola* possesses a tiny genome of 0.64 Mb which is devoid of several non-essential amino acid biosynthesis genes (Hansen and Moran, 2011). These deficient biosynthesis pathways are complemented by the host, which produces these nutrients lacking in the bacteria. Thus, this obligate symbiosis relies essentially on an intimate metabolic cooperation (Hansen and Moran, 2011; Shigenobu and Wilson, 2011).

Another vertical symbiosis system well described in insects concerns the association between tse-tse flies (*Glossina* species) and *Wigglesworthia* bacteria (Wang *et al.*, 2013). Tse-tse flies are vectors of trypanosome parasites (*Trypanosoma brucei*), causing the Human African Trypanosomiasis also known as sleeping sickness (Wang *et al.*, 2013). These insects are colonized by an obligate *Wigglesworthia* symbiont and can possess two additional facultative symbionts, *Wolbachia* and *Sodalis* (Kim and Lee, 2015). Similar to *B. aphidicola*, the *Wigglesworthia* bacteria live inside bacteriocytes, housed in a unique bacteriome (Wang *et al.*, 2013). However, a small fraction of the *Wigglesworthia* population is also present extracellularly in the milk gland secretions, which will be transmitted to the larvae *in utero* (Attardo *et al.*, 2008). These two partners also depend on their mutualistic nutritional exchanges, such as vitamins and amino acids (Wang *et al.*, 2013).

Interestingly, *Buchnera* and *Wigglesworthia* species share a same striking characteristic which is their extremely small genome size of less than 1 Mb (Shigenobu *et al.*, 2000; Wang *et al.*, 2013). Compared to horizontally-transmitted symbionts, these maternally-inherited bacteria are strictly associated to their host in a long-term obligate interaction for their survival, which

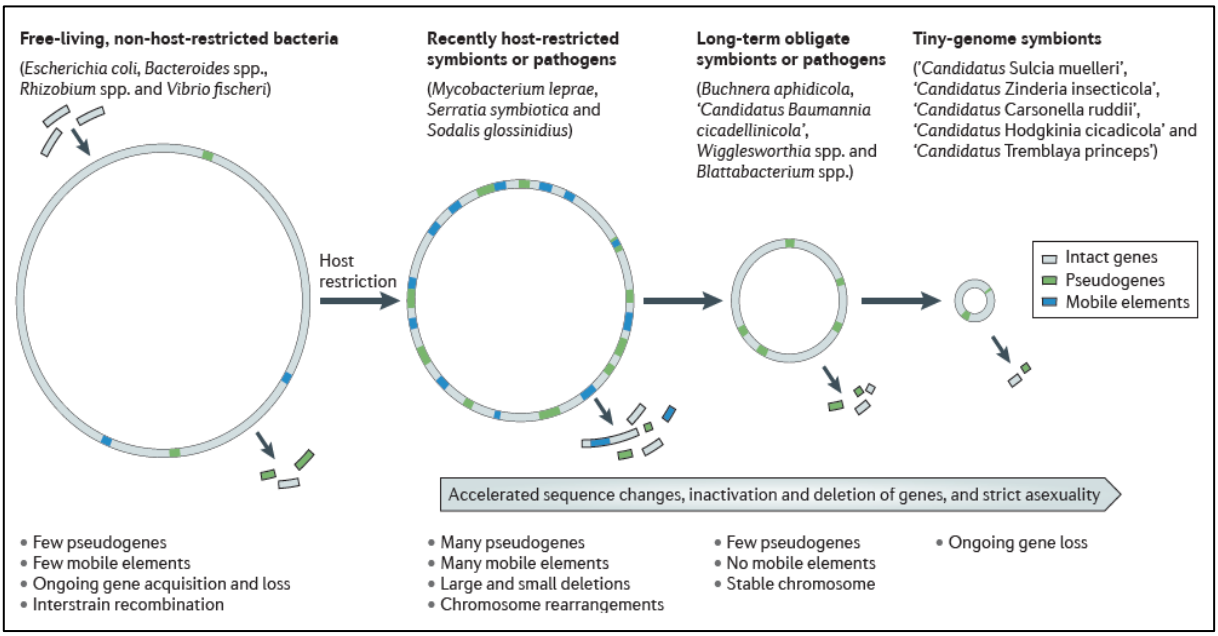


Figure 4: Genome reduction effect on long-term obligate symbiotic bacteria.
Taken from *McCutcheon and Moran, 2011*.

underlies a strong coevolutionary process. Hence, their intracellular lifestyles had a dramatic consequence on their genome that underwent significant gene losses (**Figure 4**) (McCutcheon and Moran, 2011).

As a result, *Buchnera* and *Wigglesworthia* symbionts are not able to live independently from their insect host, hence cannot be cultured *in vitro*. And similarly, their hosts can not be reared without the symbionts. These features imply strong limitations on the available possibilities for experimentation and in particular for genetic analysis. Studies on mechanisms in these symbiotic interactions are therefore restricted mostly to genomic analyses (genome analyses, transcriptomics), reverse genetics by RNAi and histological experiments (Chaudhary *et al.*, 2014; Shigenobu and Stern, 2013). Thus, other insect-bacteria interactions, in which both partners can be studied independently, would be attractive models to apply up-to-date functional genomic tools.

2. The *Riptortus pedestris-Burkholderia insecticola* symbiosis

2.1. The stinkbug *Riptortus pedestris*

Recently, *Riptortus pedestris* in association with *Burkholderia insecticola* was recognized as an ideal model system to study insect-bacteria interactions (Kikuchi *et al.*, 2007; Kim and Lee, 2015; Takeshita and Kikuchi, 2017). The stinkbug *R. pedestris* (Order: Hemiptera, Suborder:

Heteroptera, Infraorder: Pentatomomorpha, Superfamily: Coreoidea, Family: Alydidae (broad-headed bugs)), known with the common name “bean bug”, is a phytophagous insect.

Figure 5: *Riptortus pedestris* adult and its dramatic impact on soybean seeds.
Taken from *Bae et al., 2014*.

This insect is a notorious crop pest in South-Eastern Asia, notably in Japan and in South Korea, but also in India. These economically significant pest is feeding preferentially on soybean seeds

Figure 6: Insect pests from the Pentatomomorpha infraorder.
A) *Cletus punctiger*. B) *Halyomorpha halys*. C) *Nezara viridula*. D) *Pyrrhocoris apterus*.
Pictures taken from https://www.inaturalist.org/

and other crop legumes such as pigeon pea, cowpea and chickpea thanks to their piercing-sucking mouthparts, called rostrum or proboscis (**Figure 5**) (Bae *et al.*, 2014).

Besides the bean bug, many other species of the Pentatomomorpha infraorder are pests or nuisances (**Figure 6**) (Henry, 1997; Schaefer and Panizzi, 2000).

For instance, *Cletus punctiger* (**Figure 6A**), a species of the Coreidae family, closely related to the Alydidae, is a serious pest of rice cultures (Paik *et al.*, 2007). Another example is the brown marmorated stinkbug *Halyomorpha halys* (**Figure 6B**), an insect of the family Pentatomidae which is a native species in China, Japan and the Korean peninsula, but is currently an invasive species in America and Europe (Bergmann *et al.*, 2016). This stinkbug is a nuisance, invading homes, and more importantly, it is a polyphagous pest feeding on a wide array of plants, such as tree fruits, legumes, field crops and ornamentals (Lu *et al.*, 2017). Other well-known examples are the southern green stinkbug *Nezara viridula* (**Figure 6C**), another polyphagous insect pest of mainly legumes but also of tomatoes in America (Gordon *et al.*, 2017), and the red firebug *Pyrrhocoris apterus* (**Figure 6D**), a common species in Europe which feeds on linden tree seeds (Sudakaran *et al.*, 2012).

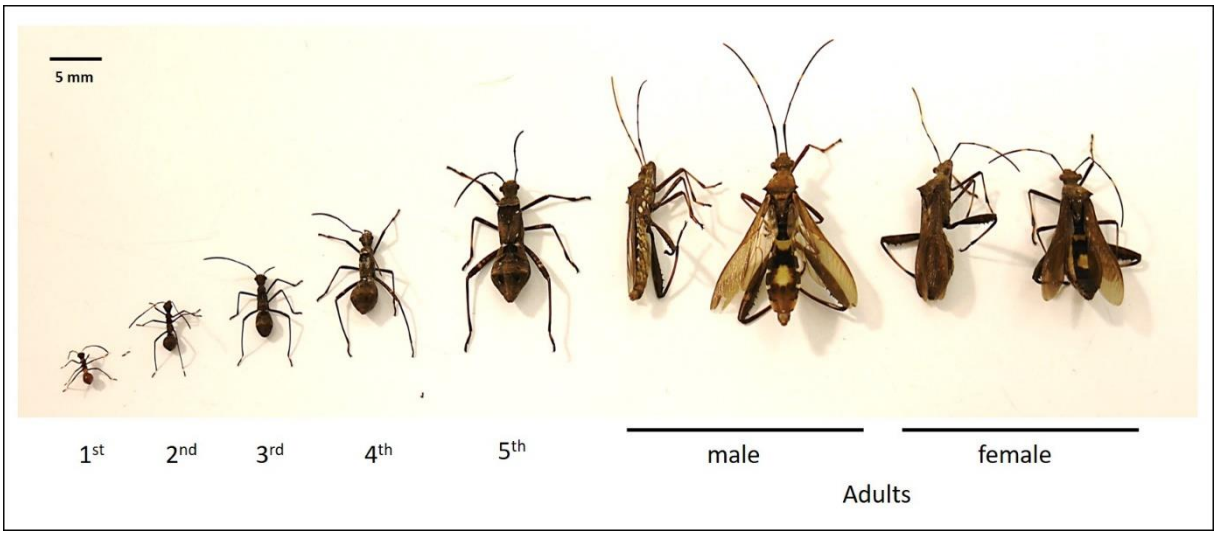


Figure 7: Development and morphologies of the different larval stages of *Riptortus pedestris*.

Picture is showing the dorsal view, with additional lateral view for the adult insects.

R. pedestris undergoes five molting stages or instars (**Figure 7**), and reach their adult form in less than 20 days after hatching (Kikuchi *et al.*, 2011a). The adults are characterized by a thin body of approximately 1.6 cm long with a brown coloration (**Figure 7**). Small patches of a yellowish color on the upperside of the abdomen can be visible, but are generally hidden by two pairs of wings (forewings and hindwings), and they bear two long and thin hindlegs (**Figure 7**). Physical differences between males and females can be noticed only at the adult stage, by

checking their abdominal genital morphology. Additionally, males are thinner than females and can be punctuated by lateral whitish dots localized on their thorax (**Figure 7**).

2.2. The symbiosis of Pentatomomorpha insects

The Pentatomomorpha infraorder to which *R. pedestris* belongs, is one of the six infraorders of Heteroptera, comprising over 12,500 insect species known as stinkbugs (Henry, 1997). Except for some predacious and mycophagous species, the majority of the pentatomomorphan stinkbugs are phytophagous (Henry, 1997; Schaefer and Panizzi, 2000).

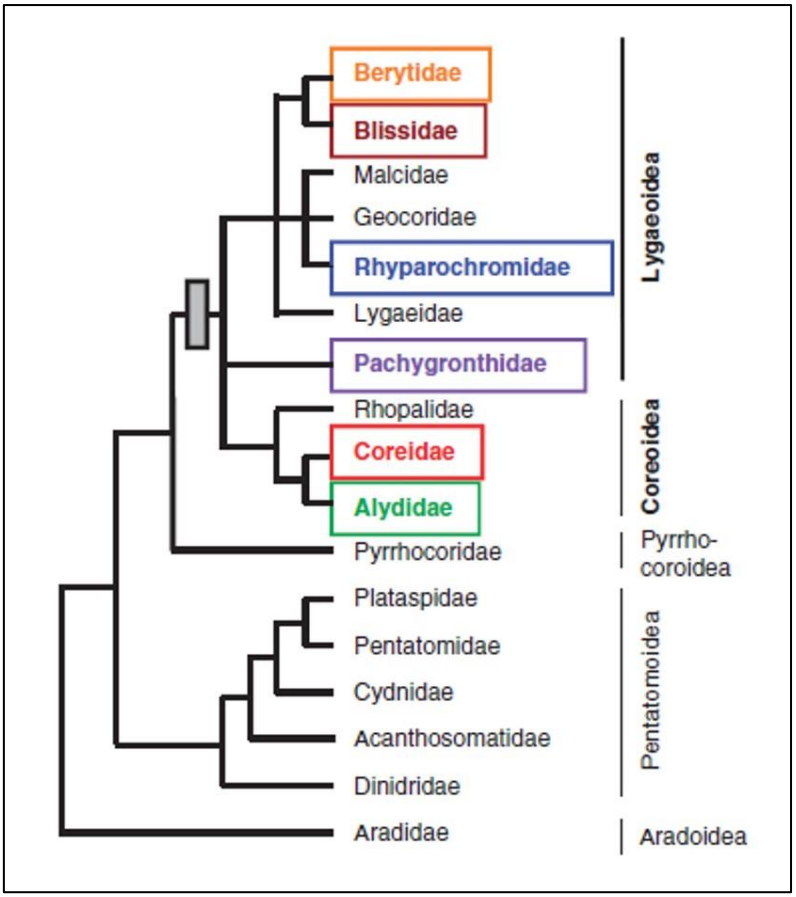


Figure 8: Phylogenetic tree of the different stinkbug families from the Pentatomomorpha infraorder.

The tree was built based on nuclear 18S rRNA gene sequences and mitochondrial whole-genome phylogeny. Superfamilies are indicated on the right.

Taken from *Kikuchi et al., 2011*.

Within the Heteroptera, the Pentatomomorpha infraorder forms a monophyletic group and consists of five superfamilies: the Lygaeoidea, Coreoidea, Pyrrhocoroidea, Pentatomoidea and Aradoidea (Hua *et al.*, 2008; Kikuchi *et al.*, 2011b; Li *et al.*, 2005; Sudakaran *et al.*, 2012) (**Figure 8**). Many heteropteran insects possess extracellular symbiotic bacteria which are harboured

in specific sac-like or tube-like outgrowths named crypts or caeca, constituting specialized organs of the insect's midgut posterior region (Buchner, 1965; Dasch, 1984; Engel and Moran, 2013; Fukatsu and Hosokawa, 2002; Glasgow, 1913). These sac-like organs vary considerably in their number, morphology and arrangement in the different families of the Heteroptera

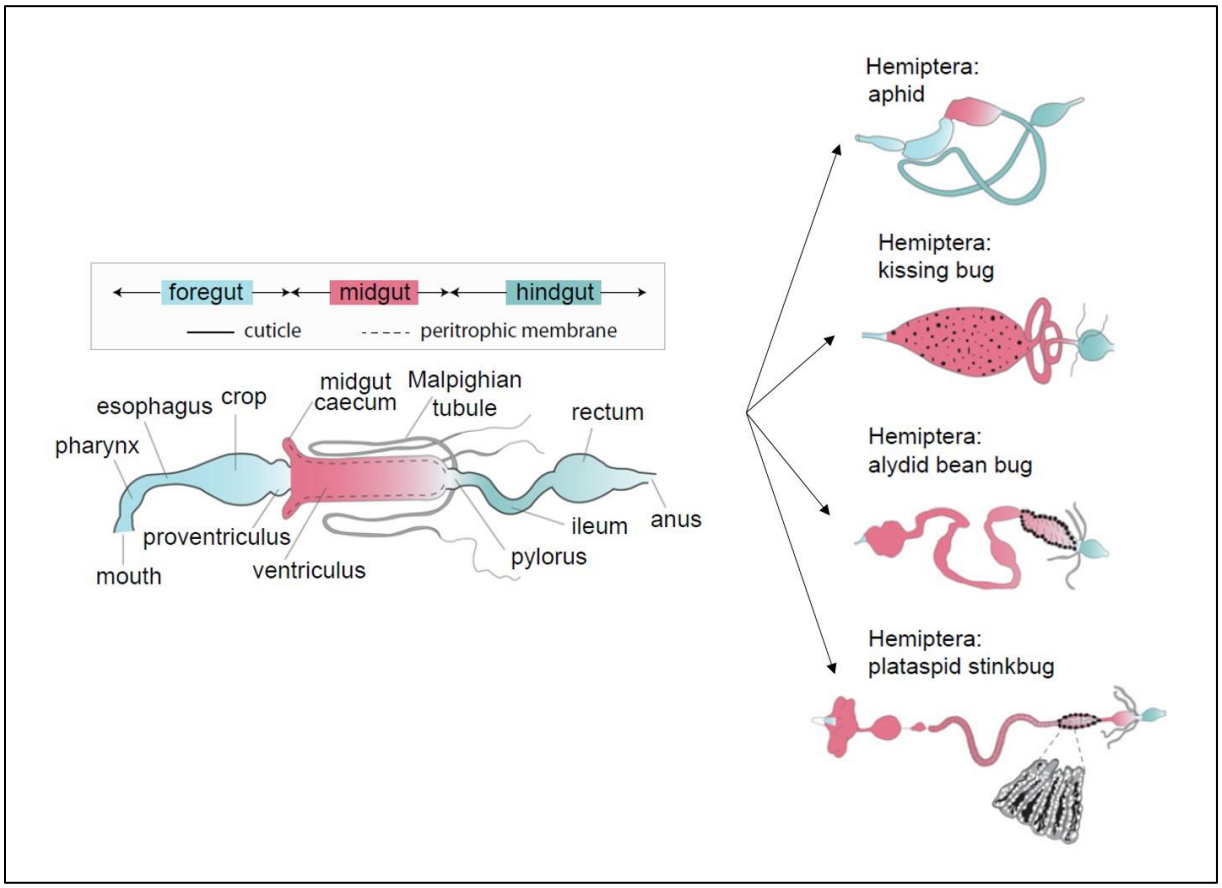


Figure 9: Insect intestinal tract organisation and morphologies in Heteroptera.
For the alydid bean bug and the plataspid stinkbug, crypts or caeca are outlined in black.
Taken from *Engel and Moran, 2013.*

(Engel and Moran, 2013; Fukatsu and Hosokawa, 2002; Kikuchi *et al.*, 2011b) (**Figure 9**).

However, a few stinkbug species from the Lygaeoidea superfamily, mostly from the Blissidae and the Lygaeidae families, lack the specialized midgut crypt region and have instead intracellular endosymbionts hosted within specialized bacteriomes similar to the *Buchnera*-aphid symbiosis (Kuechler *et al.*, 2011, 2012; Matsuura *et al.*, 2012). In the Lygaeidae family, the birch catkin bug *Kleidocerys resedae* harbours a γ-proteobacterial endosymbiont in a unique red-colored bacteriome also called mycetome, located close to the midgut (Küchler *et al.*, 2010). In the Blissidae family, the stinkbug *Ischnodemus sabuleti* possesses a specific endosymbiont closely related to *Baumannia cicadellinicola* and localized in a pair of whitish bacteriomes (Kuechler *et al.*, 2012).

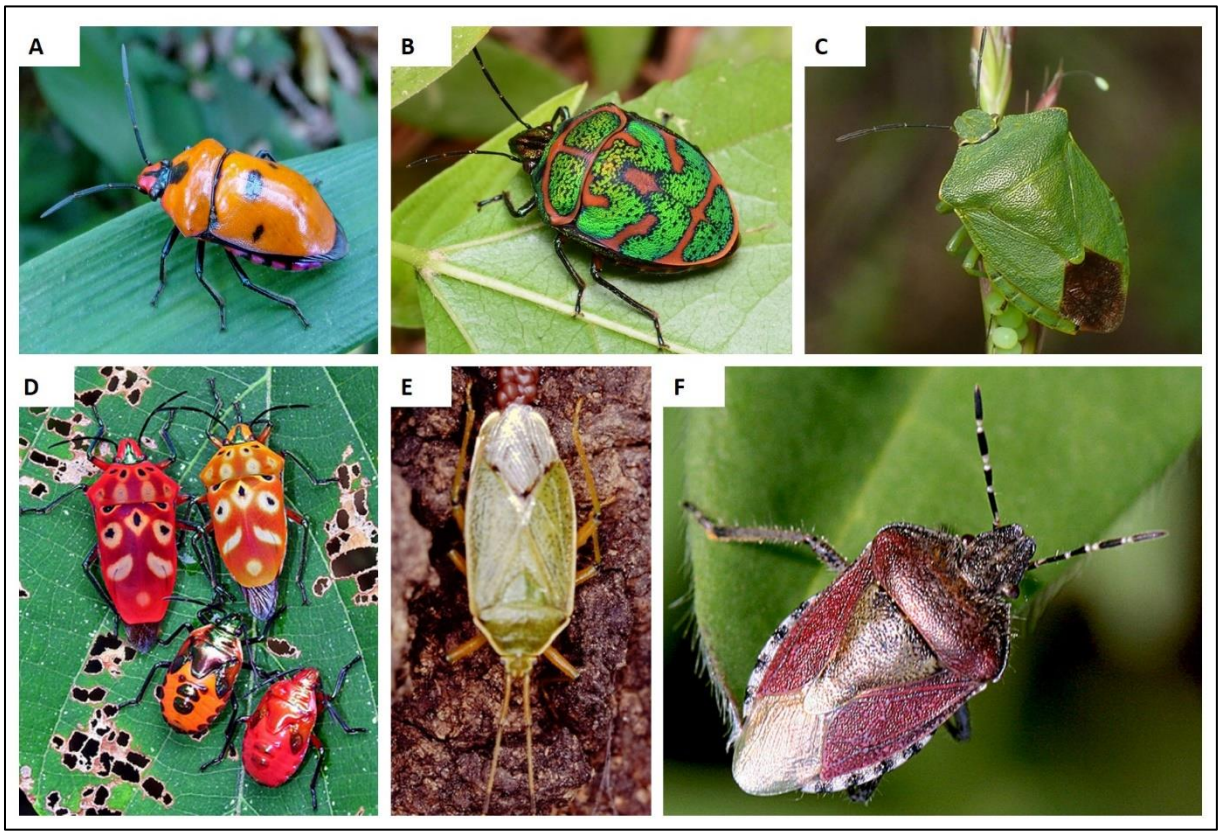


Figure 10: Pentatomomorphan insect species associated with γ-proteobacteria extracellular symbionts.
A) *Eucorysses grandis*. B) *Poecilocoris lewisi*. C) *Palomena angulosa*. D) *Cantao ocellatus*. E) *Urostylis westwoodii*. F) *Dolycoris baccarum*.
Pictures taken from https://www.inaturalist.org/

Species of the family Pyrrhocoridae of the superfamily Pyrrhocoroidea also lack crypts and the corresponding midgut region is underdeveloped and does not contain any symbiotic microbes. However, these species harbour conserved specific microbiota in another region of the midgut, the so-called M3 region (Salem *et al.*, 2013; Sudakaran *et al.*, 2012, 2015). The M3 microbiota mainly consists of Actinobacteria, Firmicutes and Proteobacteria and is transmitted vertically. It is believed that the acquisition of this specific microbiota in this family of stinkbugs has facilitated the adaptation of these insects to their host plants belonging to the angiosperm order Malvales (Salem *et al.*, 2013; Sudakaran *et al.*, 2015).

As mentioned before, stinkbugs from the Pentatomomorpha infraorder are mostly associated with extracellular symbionts located in the crypts or the caeca of the midgut. In stinkbug species of the families Plataspidae, Pentatomidae, Acanthosomatidae and Cydnidae (**Figure 8**), they are associated with distinct lineages of γ-proteobacteria which are all vertically-transmitted to the next generation (Fukatsu and Hosokawa, 2002; Hosokawa *et al.*, 2006;

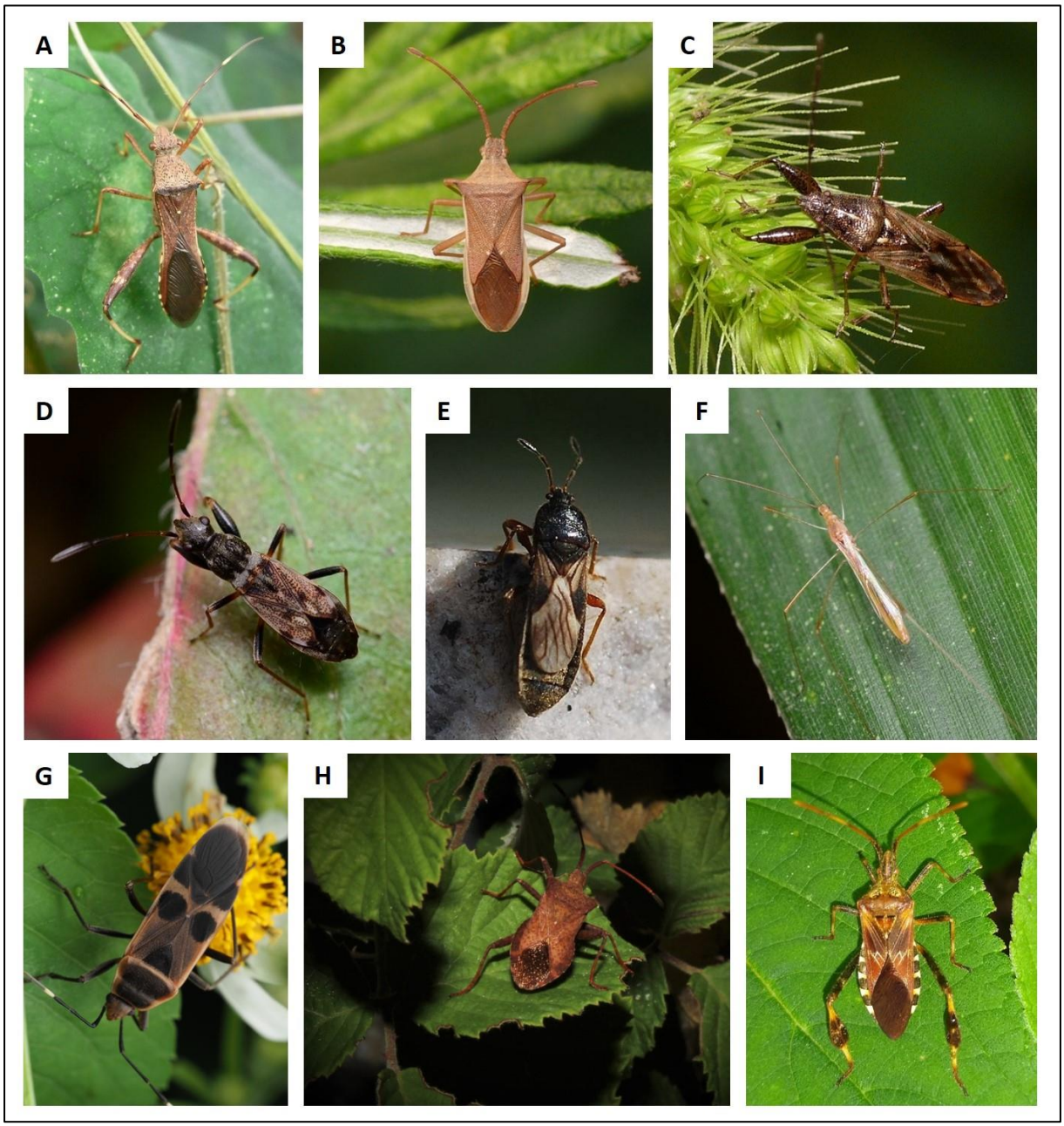


Figure 11: Pentatomomorphan insect species associated with *Burkholderia* β-proteobacteria extracellular symbionts.

A) *Riptortus pedestris*. B) *Cletus punctiger*. C) *Pachygrontha antennata*. D) *Togo hemipterus*. E) *Dimorphopterus pallipes*. F) *Yemma exilis*. G) *Physopelta gutta*. H) *Coreus marginatus*. I) *Leptoglossus occidentalis*.

Pictures taken from https://www.inaturalist.org/

Kikuchi *et al.*, 2009). For example, multiple stinkbug species are associated with *Sodalis* γ-proteobacterial symbionts in their midgut, such as *Eucorysses grandis* (Kaiwa *et al.*, 2011) (**Figure 10A**), *Poecilocoris lewisi* (Hosokawa *et al.*, 2015) (**Figure 10B**), *Palomena angulosa* (Hosokawa *et al.*, 2015) (**Figure 10C**), *Cantao ocellatus* (Hosokawa *et al.*, 2015) (**Figure 10D**),

Urostylis westwoodii (Kaiwa *et al.*, 2014) (**Figure 10E**) and *Dolycoris baccarum* (Hosokawa *et al.*, 2015) (**Figure 10F**). Another well-known example is the plataspid stinkbug *Megacopta punctatissima*, which harbours a specific γ-proteobacterial symbiont called *Candidatus Ishikawaella capsulata* (Fukatsu and Hosokawa, 2002; Hosokawa *et al.*, 2006, 2007).

In contrast, within the Pentatomomorpha infraorder, most members of the superfamilies Lygaeoidea, and Coreoidea, including all analyzed species in the Alydidae family to which *R. pedestris* belongs, as well as members of the Largidae family of the above-mentioned superfamily Pyrrhocoroidea, are associated with β-proteobacterial symbionts from the *Burkholderia* genus (**Figure 8**). Seven representative families, the Alydidae, Coreidae, Pachygronthidae, Rhyparochromidae, Blissidae, Berytidae and Largidae were shown to be associated with *Burkholderia* symbionts (**Figure 8**). These symbiotic interactions were demonstrated for various insect species, including the bean bug *R. pedestris* (Alydidae) (**Figure 11A**), the rice bug *C. punctiger* (Coreidae) (**Figure 11B**), *Pachygrontha antennata* (Pachygronthidae) **(Figure 11C**), *Togo hemipterus* (Rhyparochromidae) (**Figure 11D**), *Dimorphopterus pallipes* (Blissidae) (**Figure 11E**), *Yemma exilis* (Berytidae) (**Figure 11F**), *Physopelta gutta* (Largidae) (**Figure 11G**), *Coreus marginatus* (Coreidae) (**Figure 11H**) and *Leptoglossus occidentalis* (Coreidae) (**Figure 11I**) (Kikuchi *et al.*, 2011b; Ohbayashi *et al.*, 2019a; Takeshita *et al.*, 2015). We characterized the symbiosis of the two latter species after having them collected respectively from rumex plants in the prairie fields and from pine trees surrounding our laboratory in Gif-sur-Yvette (Ohbayashi *et al.*, 2019a, see **Publications**; and unpublished data).

Contrary to the stinkbugs that possess γ-proteobacteria in crypts or in bacteriomes or specific M3 microbiota, the *Burkholderia*-infected stinkbugs acquire their symbionts always from the environment, either from the soil or possibly from the host plant (*i.e.* horizontal symbiont transmission). Despite this, the *Burkholderia* species that are associated with the stinkbugs belong to diverse but specific clades of the *Burkholderia* (see **section 3**). The particular resident bacteria that are found in a given insect species are determined in the first place by the host, indicating the existence of stringent selection mechanisms (Takeshita *et al.*, 2015). However, also the geographic origin of insect specimens can influence the phylogenetic placement of its *Burkholderia* symbiont, meaning that even if the selection mechanism is stringent, it also displays some level of flexibility. This adaptability probably allows the insects

to acquire symbionts in nature independently of the local prevalence of *Burkholderia* species (Ohbayashi *et al.*, 2019a).

2.3. The symbiosis of *Riptortus pedestris*

R. pedestris is colonized by a specific *Burkholderia* species, named *B. insecticola*. This symbiont is a Gram-negative β-proteobacteria, rod-shaped, aerobic and motile, which was first isolated from *R. pedestris* midgut in 2007 (Takeshita *et al.*, 2018). Its genome of 6.96 Mb was completely sequenced and bears five replicons, comprising three chromosomes and two plasmids (Shibata *et al.*, 2013). The insect orally acquires its unique *Burkholderia* symbiont by horizontal transmission from the rhizosphere environment at early stages of development, mostly during the 2nd larval stage (Kikuchi *et al.*, 2011a). Due to this horizontal acquisition, each generation is born symbiont-free or aposymbiotic (Kikuchi *et al.*, 2007). In nature, all adult insects or late instars are symbiotic, but in the laboratory, it is possible to maintain them symbiont-free (Kikuchi *et al.*, 2007). The midgut (M) of *R. pedestris* is divided in five compartments: a large organ (M1 region), a second long tubular organ (M2 region), a swollen part (M3 region), a small bulk organ (M4B region) which is separated by a constricted region (CR) from the M3 section, and the crypt-bearing organ (M4 region) (**Figure 12**) (Takeshita and Kikuchi, 2017). Once ingested via drinking or feeding, the symbiotic bacteria move along the different midgut compartments and colonize the most posterior M4 midgut region. The establishment of a symbiont population in the M4 crypts is a very fast process (Kikuchi and

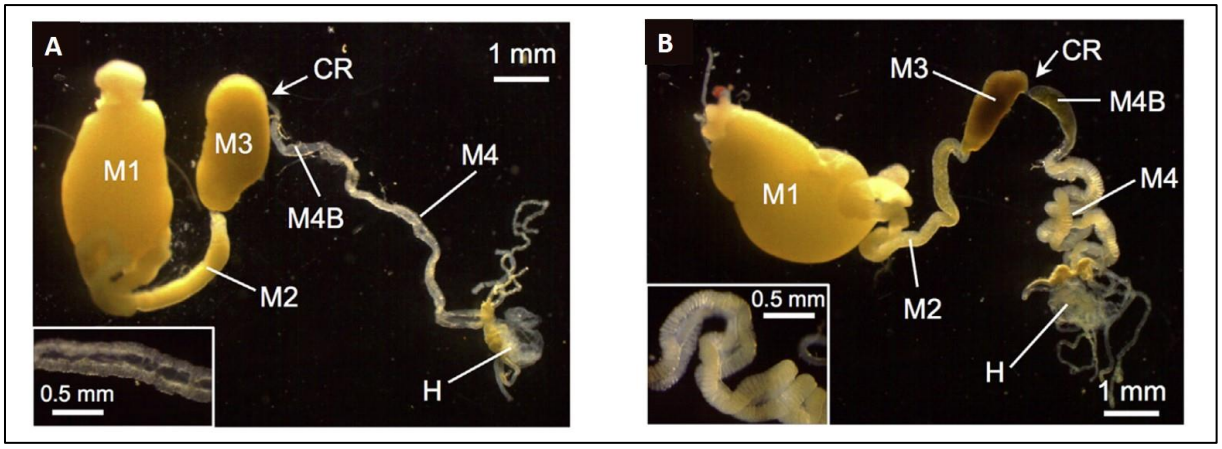


Figure 12: Midgut sections of dissected *R. pedestris* adults.
A) Aposymbiotic insect. B) Symbiotic insect.
M: midgut, CR: constricted region, H: hindgut.
Taken from *Takeshita and Kikuchi, 2017.*

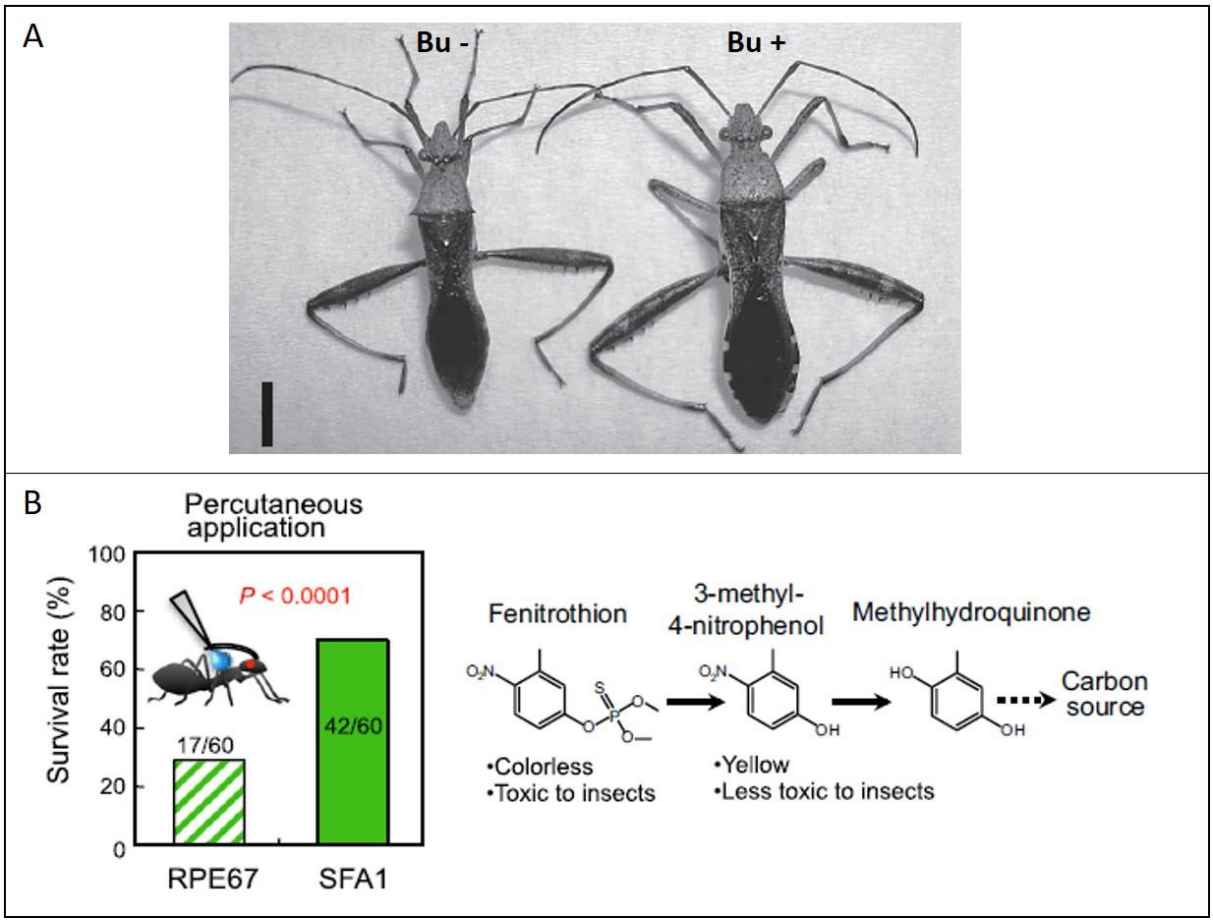


Figure 13: Host fitness benefits upon symbiotic association with *Burkholderia insecticola*. A) Morphometric differences between aposymbiotic (Bu-) and symbiotic (Bu+) *Riptortus pedestris* adults. The scale bar represents 5 mm. B) Effect of fenitrothion insecticide on the survival rate of 3rd instar *Riptortus pedestris*. On the left side, the survival rate was measured on insects infected with fenitrothion-degrading *Burkholderia* strain (SFA1) and with non-degrading *Burkholderia* strain (RPE67). On the right side, fenitrothion degradation is performed by the *Burkholderia* symbiont to exploit this waste compound as a carbon source. Taken from *Kikuchi 2009 and Kikuchi et al., 2012*.

Fukatsu, 2014). Six hours after the initial uptake of the bacteria, they appear at the junction between the M3 and the M4B midgut regions and start passing through the CR and entering in the symbiotic organ (Kikuchi and Fukatsu, 2014). At 24 hours, the central duct of the M4 and some of the crypts are filled and at 48 to 72 hours, all crypts are entirely filled with the bacterial symbiont (Kikuchi and Fukatsu, 2014). In symbiotic insects, the M4 region is morphologically differentiated with large, open and whitish crypts compared with the crypts in aposymbiotic insects which are collapsed and have a transparent appearance (**Figure 12**) (Takeshita and Kikuchi, 2017).

The M4 region houses as many as 10^8 *B. insecticola* cells extracellularly in their lumen, thus

constituting the symbiotic organ (Kikuchi and Fukatsu, 2014). The M4B region, closely associated to the M4 section, seems to be involved in the symbiont digestion, thus suggesting a role of this organ in the control of the symbiont population or in the extraction of useful nutrients from the digested symbionts (Byeon *et al.*, 2015; Kim *et al.*, 2013a; Ohbayashi *et al.*, 2019b). Unlike obligate symbiosis systems, aposymbiotic bean bugs can survive in laboratory conditions, meaning that the *Burkholderia* symbiont is not absolutely essential for its host's survival in optimal laboratory rearing conditions. Although this interaction remains facultative, the symbiont reaches nearly 100% prevalence in wild *R. pedestris* populations, highlighting a strong and stable selection for this bacterial species and suggesting the essential nature of the symbiosis in natural living conditions (Kikuchi *et al.*, 2007). In agreement, for several other stinkbug species carrying horizontally acquired *Burkholderia* symbionts, it was found that aposymbiotic insects do not survive, even in laboratory conditions (Ohbayashi *et al.*, 2019a). The contribution of the symbiotic bacteria is illustrated by morphological and developmental benefits for the host fitness, such as the improvement of the insect growth (**Figure 13A**), the shorter developmental time (*i.e.* time to reach adulthood) and the higher fecundity of females compared to aposymbiotic insects (Kikuchi, 2009; Kikuchi *et al.*, 2007, 2011a). Additionally, some *Burkholderia* strains confer resistance against fenitrothion to its host by degrading this insecticide to a non-toxic derivative (**Figure 13B**) (Itoh *et al.*, 2014; Kikuchi *et al.*, 2012). In contrast, the benefits of the interaction for the bacterial symbiont are still unclear. The ingestion of a small number of bacteria from the environment is sufficient for a full occupation of the symbiotic organ in a few days (Kikuchi and Yumoto, 2013). Thus, thanks to the interaction, the symbiotic bacteria can multiply enormously inside the midgut M4 region. But as the host insect does not secrete these symbiotic bacteria during its lifetime, they are trapped in the body of the insect. However, it is possible that a fraction of the bacterial population can colonize the soil after the host's death.

Due to the facultative nature of this symbiosis, it is possible to study the host and the symbiont independently from each other, by *in vitro* culture of the symbiont and by the generation of aposymbiotic insect lineages (Takeshita and Kikuchi, 2017). Moreover, genetic manipulations of the *Burkholderia* symbiont by mutagenesis, as well as RNAi experiments on the host (Futahashi *et al.*, 2011), are highly efficient and can be applied as useful tools to study this

symbiosis system. Hence, this insect model displays various advantages to study host-symbiont interactions at the molecular level.

2.4. *Burkholderia insecticola* symbiotic functions

In recent years, classical bacterial genetics and genomic approaches, including proteomic and transcriptomic analyses, have described a number of functions in *B. insecticola* that are required for the colonization of the *R. pedestris* symbiotic organ.

To investigate the effect of cell motility and bacterial morphology on the symbiotic association, Lee *et al*., have studied the role of the *amiC* gene based on a previous study on *E. coli* which reveals that mutants in the *ami* gene cluster led to abnormal cell morphology (Heidrich *et al.*, 2002; Lee *et al.*, 2015). This *amiC* gene encodes an N-acetylmuramyl-L-alanine amidase which degrades the peptidoglycan and is involved in daughter-cell separation during bacterial cell proliferation (Lee *et al.*, 2015). In *B. insecticola*, the *amiC* mutant exerted abnormal cell morphology with elongated filamentous shape, but did not affect the bacterial growth rate *in vitro* (Lee *et al.*, 2015). In addition, this mutant showed a defect in cell motility compared to the wild-type strain (Lee *et al.*, 2015). During *in vivo* experiments, when the *amiC* mutant was given with an initial inoculum concentrated at 10^7 bacteria per mL, the mutant was not able to colonize the symbiotic organ of *R. pedestris* nymphs in contrast to the wild-type strain (Lee *et al.*, 2015). However, when the *amiC* mutant was administered with a 1000-fold higher initial inoculum, the infection rate was similar to that of the wild-type strain (Lee *et al.*, 2015). In another study based on random transposon Tn*5* mutagenesis of *B. insecticola*, Ohbayashi *et al*., have identified several motility-deficient bacterial mutants (Ohbayashi *et al.*, 2015). These mutants contained the transposon insertion in genes encoding bacterial flagella subunits (*fliC*, *fliF*, *fliM*, *fliR*) and genes involved in the regulation of flagella formation (*flhA*, *fliK* and *flhF*) (Ohbayashi *et al.*, 2015). It was shown that these mutants were able to infect the midgut until the M3 region, but were not able to colonize the M4 region (Ohbayashi *et al.*, 2015). As these two midgut regions are separated by the thin CR, these results suggested that the bacterial motility is important to pass through the CR and to reach the M4 region (Ohbayashi *et al.*, 2015).

In order to identify symbiosis-related molecules, Kim *et al*., have compared the global protein profiles of *in vivo* and *in vitro B. insecticola* cells by SDS-PAGE and identified one protein,

named phasin (PhaP), which was more abundant in *in vivo* cells than *in vitro* cells (Kim *et al.*, 2013b). This PhaP protein is present at the surface of polyhydroxyalkanoate (PHA) intracellular granules of bacteria (York *et al.*, 2001). PHAs are linear polyesters produced by many bacterial species, usually during nutritional stress conditions, which accumulate as granules and serve as carbon and energy storage or as sinks for excess reducing power generated when the available carbon and nitrogen are not in balance (Anderson and Dawes, 1990; Poblete-Castro *et al.*, 2012). When *R. pedestris* nymphs were infected with the *phaP B. insecticola* mutant, it was able to colonize the symbiotic organ with the same efficiency as the wild-type strain (Kim et al., 2013b). However, insects infected with bacterial mutants in the PHA biosynthesis genes *phaB* and *phaC* exhibited a low colonization efficiency of the M4 region compared with the wild-type strain (Kim *et al.*, 2013b). In addition, the *B. insecticola* mutant of the *phaR* gene, which encodes a negative regulator of *phaP* expression, was also colonizing less efficiently the symbiotic organ during *in vivo* experiments (Jang *et al.*, 2017). Thus, the bacterial mutants *phaB*, *phaC* and *phaR*, which were not able to produce PHA granules, also demonstrated a weak colonization of the symbiotic organ of *R. pedestris* indicating that buffering the reducing power and balancing carbon and nitrogen is essential in the nutritional conditions of the crypt environment (Jang *et al.*, 2017; Kim *et al.*, 2013b).

Also based on a random transposon Tn*5* mutagenized library of *B. insecticola*, Kim *et al.*, have screened bacterial mutants for their colonization capabilities of the host midgut (Kim *et al.*, 2014a). In this study, Kim *et al*., identified one symbiotic-deficient mutant in which the *purL* gene was interrupted by the transposon (Kim *et al.*, 2014a). This *purL* gene encodes the N-formylglycinamidine ribonucleotide synthetase, involved in *de novo* purine biosynthesis in bacteria (Zhang *et al.*, 2008). When the *purL* mutant of *B. insecticola* was inoculated to the host, only 30% of the insect population tested was infected, but the bacterial load in the M4 region was 100-fold lower than that of insects infected with the wild-type strain (Kim *et al.*, 2014a).

As mentioned before, the bacterial morphology was previously investigated with the *amiC* mutant (Lee *et al.*, 2015), but other functions related to the bacterial cell wall were studied. As the cell envelope components are in direct contact with the surrounding host environment, Kim *et al*., targeted a gene, *uppP*, encoding the undecaprenyl-diphosphatase involved in the biosynthesis of a lipid carrier precursor for LPS and peptidoglycan biosynthesis (Kim *et al.*,

2013c). This *uppP* mutant was more susceptible to different environmental stress conditions, such as osmotic shock or lysozyme treatment, thus indicating an impaired cell wall integrity (Kim *et al.*, 2013c). Concerning the colonization efficiency of the symbiotic organ, the *uppP* mutant was able to reach and enter the M4 region but could not proliferate, suggesting that the mutant failed to establish symbiosis during the insect development (Kim *et al.*, 2013c). In another study, it was observed that symbiotic *Burkholderia* cells (the bacteria present inside the M4 crypts) lack the O-antigen subunit of LPS molecules in the cell envelope while free-living cells produce LPS with a long O-antigen chain (Kim *et al.*, 2015a). As the O-antigen was reported to be essential to establish symbiosis in the *Rhizobium*-legumes symbiosis (Ormeño-Orrillo *et al.*, 2008) and in the *Vibrio*-squid symbiosis (Post *et al.*, 2012), Kim *et al*., studied the role of the *B. insecticola* O-antigen in *R. pedestris* symbiosis (Kim *et al.*, 2016a). Several candidate genes involved in O-antigen biosynthesis were targeted, with three glycosyltransferase genes (*wbxA*, *wbxB* and *wbiF*) and one epimerase gene (*wbiG*). It was demonstrated that the bacterial mutants which exerted a reduced O-antigen expression (*wbxA*, *wbiF* and *wbiG*) showed a lower infection rate (55 to 67.5%) and a reduced bacterial population (30 to 100-fold) in the symbiotic organ than the wild-type strain (Kim *et al.*, 2016a). Hence, even if the O-antigen is lost during symbiosis, these results suggest that the O-antigen of LPS surface molecules are required for the initial colonization of the host's midgut (Kim *et al.*, 2016a). As the O-antigen is not present at the surface of symbiotic cells, the LPS molecules display the core oligosaccharide at the bacterial surface. The role of the core oligosaccharide biosynthesis genes in the colonization of the symbiotic organ was assessed by generating bacterial mutants with different core oligosaccharide lengths (Kim *et al.*, 2017). Bacterial mutants in the heptosyltranserase I and II genes (*waaC* and *waaF*, respectively) led to severe truncated forms of the core oligosaccharide, whereas mutants in two glycosyltransferase genes (*wabS* and *wabO*) exhibited a small reduction of the core oligosaccharide chain length (Kim *et al.*, 2017). Except for the *wabS* mutant, which was not affected for symbiosis, these glycosyltransferase mutants were not able to colonize efficiently the symbiotic organ compared to the wild-type strain (Kim *et al.*, 2017).

Thus, bacterial functions involved in cell motility, PHA biosynthesis, *de novo* purine biosynthesis, and the synthesis of the LPS and peptidoglycan cell wall components are important to establish the symbiotic interaction with *R. pedestris*.

More recently, Ohbayashi *et al*., have performed a comparative transcriptomic analysis of cultured and symbiotic (isolated from the M4 crypts) *B. insecticola* cells and have revealed which bacterial functions are regulated during the host colonization (Ohbayashi *et al.*, 2019b). It was found that genes involved in cell division, DNA replication, protein biosynthesis, cellular respiration process, LPS and peptidoglycan biosynthesis were highly expressed during *in vivo* condition, at a similar level to an exponential growth *in vitro* condition, thus indicating that bacterial cells are actively proliferating in the host midgut and that they required intact cell wall structures (Ohbayashi *et al.*, 2019b). By comparing these *in vitro* and *in vivo* conditions, 527 *in vivo* upregulated genes and 638 *in vivo* downregulated genes were found (Ohbayashi *et al.*, 2019b). Among the upregulated gene functions, the transcriptome revealed that *in vivo* cells actively use metabolic pathways involved in the uptake and degradation of carbohydrates (rhamnose, ribose, myo-inositol), fatty acids, diverse nitrogen sources (allantoin, urea) and sulfur sources (taurine, alkanesulphonates), strongly suggesting that the host insect provides these nutrient sources to the symbiont (Ohbayashi *et al.*, 2019b). In addition, biosynthetic pathways of B vitamins, methionine and tryptophan were highly expressed in the *in vivo* condition, suggesting that the symbiont might provide these vitamins and amino acids to its host (Ohbayashi *et al.*, 2019b). In contrast, the downregulated genes were involved in cell motility, chemotaxis, glucose transport and fatty acid biosynthesis (Ohbayashi *et al.*, 2019b).
In the same study, Ohbayashi *et al*., found that the bacterial morphology of the *in vivo* cells is altered. They exhibited a cocci-like shape with some membrane blebs (**Figure 14B**) compared

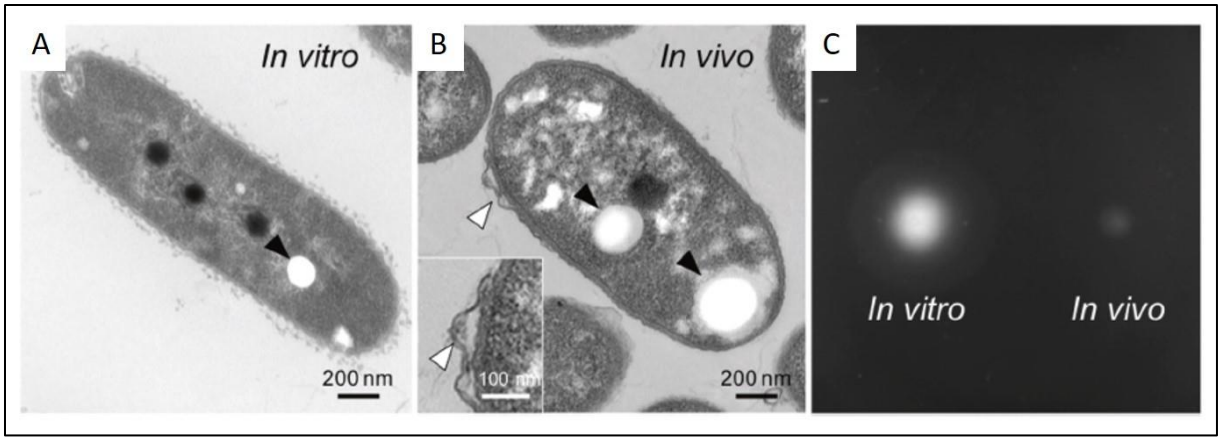


Figure 14: Bacterial morphology and motility of *in vitro* and *in vivo B. insecticola* cells.
A and B) Pictures were obtained by transmission electron microscopy. Black arrows indicate PHA granules. White arrows indicate membrane blebs. A) *In vitro* bacterial cell. B) *In vivo* bacterial cell. C) Motility test of *in vitro* and *in vivo* cells.
Taken from *Ohbayashi et al., 2019b.*

to the intact rod-shape of *in vitro* growing cells (**Figure 14A**) (Ohbayashi *et al.*, 2019b). Additionally, *in vivo* bacterial cells have a smaller size (**Figure 14B**), exert a low DNA content, accumulate PHA granules (**Figure 14B**), and lose their flagellar motility (in accordance with the downregulation of cell motility genes during the *in vivo* condition) (**Figure 14C**) (Ohbayashi *et al.*, 2019b). *In vivo* symbiotic cells were also more susceptible to different environmental stresses, such as antimicrobial peptides, osmotic shock and detergents (Ohbayashi *et al.*, 2019b). Thus, these observations demonstrated that *B. insecticola* undergoes severe morphological alterations inside the symbiotic organ, which suggest that the host environment is stressful for the symbiotic bacteria (Ohbayashi *et al.*, 2019b).

3. *Burkholderia* species

3.1. Classification and phylogeny

The genus *Burkholderia* was first introduced by Yabuuchi *et al.* in 1992 based on rRNA homology group II of pseudomonads, and included only seven species at that time (Compant *et al*., 2008). These former seven species were *Burkholderia pseudomallei*, *Burkholderia mallei*, *Burkholderia caryophilli*, *Burkholderia gladioli*, *Burkholderia cepacia*, and the remaining two species *Burkholderia picketii* and *Burkholderia solanacearum* were later transferred to the *Ralstonia* genus (Compant *et al.*, 2008). Presently, this *Burkholderia* genus represents more than 100 species (http://www.bacterio.net/burkholderia.html). Species of this genus inhabit a large variety of ecological niches, with a majority being soil bacteria (Coenye and Vandamme, 2003). Belonging to the class of β-proteobacteria, these bacteria are rod-shaped Gram-negative species (Coenye and Vandamme, 2003).

Based on phylogenetic analyses, the genus *Burkholderia* can be divided into three different clades: the BCC&P clade (*Burkholderia cepacia* complex and *Burkholderia pseudomallei* group), the PBE clade (plant-associated beneficial and environmental group) and the SBE clade (stinkbug-associated beneficial and environmental group) also known as BGC (*Burkholderia glathei* clade) (**Figure 15**) (Takeshita and Kikuchi, 2017).

The BCC&P clade (**Figure 15**) comprises pathogen species of animals and plants, including the human pathogens *B. mallei* and *B. pseudomallei*, and many opportunistic human pathogens like *Burkholderia cenocepacia* (Coenye and Vandamme, 2003; Compant *et al.*, 2008).

Interestingly, some members of this clade can be used as biocontrol agents in agriculture, such as *Burkholderia vietnamensis* and *Burkholderia ambifaria*, but their application is highly restricted by the US Environmental Protection Agency due to their close relationship with opportunistic pathogens (Eberl and Vandamme, 2016). In the PBE clade (**Figure 15**), the plant-associated species are non-pathogenic and can establish epiphytic and endophytic relationships. These bacteria are able to fix nitrogen, to promote plant growth and resistance against plant pathogens, and to form nitrogen-fixing nodules on legume roots (Coenye and Vandamme, 2003; Compant *et al*., 2008; Takeshita and Kikuchi, 2017). The last and third clade, the SBE clade (**Figure 15**), contains various environmental species and most of symbiotic species isolated from the gut of stinkbugs, including *B. insecticola* (previously named *Burkholderia sp.* RPE64) (Takeshita *et al.*, 2018). Only the symbionts of the stinkbug species of the Largidae family belong to the PBE clade (Takeshita *et al.*, 2015).

However, the *Burkholderia* classification is in a constant remodelling due to improvements in phylogenetic clustering methods. Recently, multiple species from the genus *Burkholderia* were transferred to two others, newly created genera named *Paraburkholderia* (Sawana *et al.*, 2014) regrouping the PBE species and *Caballeronia* (Dobritsa and Samadpour, 2016), which harbours the SBE clade. Thus, these two new genera contain only environmental and beneficial plant-associated species. The pathogen-containing BCC&P clade is maintained in the genus *Burkholderia*. Moreover, three additional genera, each containing only one species were proposed (Estrada-de Los Santos *et al.*, 2018; Lopes-Santos *et al.*, 2017), and thus the group *Burkholderia* sensu lato is currently divided into the genera *Burkholderia* sensu stricto, *Caballeronia*, *Paraburkholderia*, *Robbsia*, *Mycetohabitans* and *Trinickia*.

Strictly speaking, *B. insecticola* belongs thus to the genus *Caballeronia,* but the genus name *Burkholderia* is kept because the symbiosis of stinkbugs is known since its original description as the *Riptortus-Burkholderia* symbiosis (Takeshita and Kikuchi, 2017). Moreover, the division of *Burkholderia* sensu lato in distinct genera is not generally accepted, based on the arguments that the groups are not distinguished by sufficiently definable and clear phenotypes, and by consistent phylogenetic and phylogenomic support (Takeshita *et al.*, 2018; Vandamme *et al.*, 2017).

3.2. Pathogenic *Burkholderia* species

The two most prevalent pathogenic species for humans and animals in the genus *Burkholderia* are *B. pseudomallei* and *B. mallei*. *B. pseudomallei* is the causative agent of melioidosis, a zoonosis, which is predominant in South-Eastern Asia and Northern Australia (Hemarajata *et al.*, 2016). This infectious disease can have multiple forms, ranging from skin lesions to a chronic infection that can evolve into septicaemia (Hemarajata *et al.*, 2016; Titball *et al.*, 2017). The second pathogen, *B. mallei*, is causing glanders, an infectious zoonosis which can be contracted by donkeys, horses and humans (Saikh and Mott, 2017). This bacterial species is an intracellular pathogen which leads to chronic lung infection, followed also by septicaemia (Saikh and Mott, 2017). Regarding their symptoms, these diseases can be easily confused with tuberculosis (Titball *et al.*, 2017). As these two pathogens can be acquired through inhalation, they are classified as Tier 1 select agents by the US Centers for Disease Control and Prevention due to their potential use as bioweapons (Hemarajata *et al.*, 2016; Titball *et al.*, 2017).

Additionally, other members of the BCC&P clade (**Figure 15**) such as *B. cenocepacia*, are environmental species and are frequently identified as opportunistic pathogens in cystic fibrosis patients (Scoffone *et al.*, 2017). These soil bacteria are ubiquitous in the environment and induce pulmonary function decline in these patients as opportunistic infections, and can even lead to necrotizing pneumonia syndrome (Scoffone *et al.*, 2017).

In plants, *Burkholderia* species from the BCC&P clade (**Figure 15**) can also be found as phytopathogens, like *B. gladioli* which is responsible for soft rot disease in onions (Compant *et al.*, 2008). Another phytopathogen, *Burkholderia glumae*, is causing grain rot in rice and wilting symptoms in more than 20 plant species (Coenye and Vandamme, 2003; Compant *et al.*, 2008). Interestingly, some phytopathogenic *B. gladioli* strains have evolved into mutualists of herbivorous Lagriinae beetles. These beetle symbionts, which are vertically transmitted and harboured extracellularly in glands connected to the female reproductive system, are also present on the surface of eggs and protect them against fungal infections via the production of a cocktail of antifungal compounds (Flórez *et al.*, 2017). Remarkably, these insect symbionts can be transmitted from the insects to the plants, systemically infect the latter and reduce their fitness. Moreover, it was proposed that one of the antimicrobials that protect the eggs is at the same time involved in plant pathogenicity (Flórez *et al.*, 2017).

3.3. Environmental and beneficial *Burkholderia* species

The *Paraburkholderia* and *Caballeronia* genera comprise environmental species, many of which are known to interact with eukaryotic hosts, conferring beneficial effects to them (Compant *et al.*, 2008; Eberl and Vandamme, 2016). Concerning plant hosts, beneficial *Burkholderia* species stimulate plant growth development by production of phytohormones, siderophores or ammonium (Divan Baldani *et al.*, 2000; Esmaeel *et al.*, 2018). Some species are also known to protect the plant tissues against phytopathogens (Coenye and Vandamme, 2003; Compant *et al.*, 2008). For example, the endophytic *Burkholderia phytofirmans* strain PsJN (reclassified as *Paraburkholderia phytofirmans*) protects its host plants (*e.g.* potatoes, tomatoes, grapevine and other crops) by inhibiting the growth of various phytopathogenic fungi and bacteria (Esmaeel *et al.*, 2018; Sessitsch *et al.*, 2005). Still others, like *Paraburkholderia phymatum* and *Paraburkholderia tuberum*, are even nitrogen-fixing rhizobia, capable of inducing and infecting root nodules on legumes (Moulin *et al.*, 2001). These *Burkholderia* are known as β-rhizobia, referring to their membership to the β-proteobacteria, as opposed to the large majority of described rhizobium species which are α-proteobacteria (Gyaneshwar *et al.*, 2011; Lemaire *et al.*, 2016).

The *Caballeronia* genus contains the earlier mentioned symbiotic bacteria isolated from the gut of different stinkbug families from the Pentatomomorpha infraorder (Kikuchi and Yumoto, 2013; Kikuchi *et al.*, 2011b). Additionally, symbiotic *Burkholderia* species were detected in the gut of the ant *Tetraponera binghami* (van Borm *et al.*, 2002), which suggests that the presence of *Burkholderia* genus might be underestimated for its associations with insect species (Flórez and Kaltenpoth, 2017; Flórez *et al.*, 2017). Other remarkable species of the *Caballeronia* genus are plant symbionts which form leaf nodules or galls at the surface of the leaves, such as *Candidatus Burkholderia kirkii* (Carlier and Eberl, 2012; Carlier *et al.*, 2013). These leaf nodule symbioses have been described in *Psychotria*, *Pavetta* and *Ardisia* plant species, located in tropical and sub-tropical Africa (Lemaire *et al.*, 2011; Pinto-Carbó *et al.*, 2016, 2018). Interestingly, the symbiotic *Burkholderia* in these plants are transmitted vertically to the plant progeny by colonizing the developing seeds, a rare phenomenon in plant symbioses (Lemaire *et al.*, 2012).

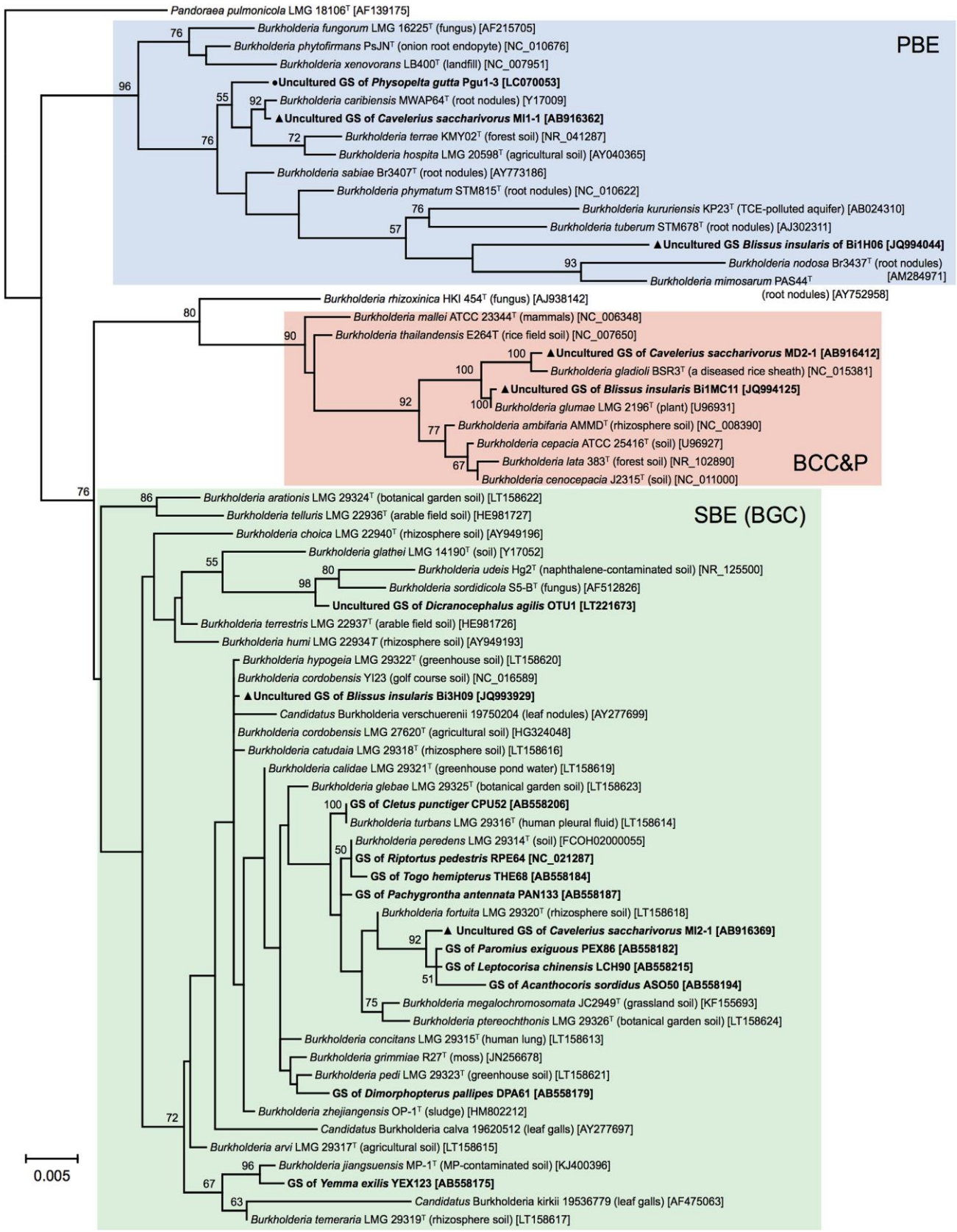


Figure 15: Phylogenetic tree of *Burkholderia* species.
This tree was built based on 16S rRNA gene sequences.
Taken from *Takeshita and Kikuchi, 2017*.

Similarly as the vertically-transmitted symbionts of insects, these leaf symbionts show a high degree of genome erosion and have by consequence lost their capacity for free-living growth (Pinto-Carbó *et al.*, 2018). Genomic and proteomic analyses revealed that *Candidatus Burkholderia kirkii* possesses a unique cluster of genes involved in the biosynthesis of C_7N aminocyclitol derivatives, which has no homologs in the *Burkholderia* genus (Carlier and Eberl, 2012; Carlier *et al.*, 2013). One C_7N aminocyclitol molecule named kirkamide was isolated from leaf nodules of *Psychotria kirkii* plants and was shown to exhibit cytotoxic and insecticidal activities, hence suggesting that the bacterial symbiont may have a protective beneficial role for the fitness of the host plant (Carlier and Eberl, 2012; Carlier *et al.*, 2013; Sieber *et al.*, 2015). Also other *Burkholderia* leaf symbionts produce secondary metabolites, which might have a protective role against herbivorous insects, suggesting that this is a common function of leaf nodule symbioses (Crüsemann *et al.,* 2018). However, this is certainly not the sole function of the symbiosis because in the tested cases, aposymbiotic plants develop poorly or not at all, suggesting that the symbiosis also affects plant development (Lemaire *et al*., 2012 ; Sinnesael *et al.*, 2019).

3.4. Antimicrobial resistance in *Burkholderia*

Although little is known about antibiotic resistance of environmental *Burkholderia* species, the pathogenic *Burkholderia* species are notoriously highly resistant to various antibiotics and it is believed that their pathogenicity depends on their resistance mechanisms (Sfeir, 2018). The human pathogens *B. mallei* and *B. pseudomallei* are highly resistant against a broad spectrum of antibiotics, such as β-lactams, aminoglycosides, polymyxins and macrolides, thus decreasing treatment efficiency in patients affected by these infectious diseases (Hemarajata *et al.*, 2016; Saikh and Mott, 2017; Sfeir, 2018). The same resistance pattern is observed in clinical cases infected by the opportunistic pathogens from the BCC&P clade, with natural resistance against cephalosporins, polymyxins and carboxypenicillins (El-Halfawy and Valvano, 2013; Sfeir, 2018). For the moment, the most effective treatment against BCC&P infections remains the association of four antibiotics: trimethoprim-sulfamethoxazole, ceftazidim, meropenem and doripenem (Sfeir, 2018).

Different molecular mechanisms were described to explain this large antibiotic resistance. In *B. mallei*, *B. pseudomallei* and *B. thailandensis*, the gene *penA* encodes a β-lactamase which

is secreted extracellularly and inactivates β-lactams antibiotics (**Figure 16**) (Rhodes and Schweizer, 2016). Moreover, various *Burkholderia* pathogens express efflux pumps from the RND (Resistance Nodulation cell Division) family, especially the AmrAB-OprA and the BpeEF-OprC systems, to export diverse antibiotic classes such as chloramphenicol, fluoroquinolones, aminoglycosides, tetracyclines and macrolides (**Figure 16**) (Rhodes and Schweizer, 2016). Compared to enterobacterial species, members of the *Burkholderia* genus also demonstrate changes in outer membrane permeability (Rhodes and Schweizer, 2016). One particular example is the modification of the lipid A component of LPS (**Figure 16**) with 4-amino-4-deoxy-arabinose (Ara4N) moiety, which decreases the net negative charge of the outer membrane, thus reducing the potent interaction with cationic antimicrobial peptides, including polymyxins (Ortega *et al.*, 2009; Rhodes and Schweizer, 2016). This weak membrane permeability also contributes to resistance against aminoglycosides and β-lactams (Rhodes and Schweizer, 2016). Interestingly, this Ara4N modification is essential for *Burkholderia* viability (Ortega *et al.*, 2007), while in the Enterobacteriaceae, Ara4N is introduced on the LPS only upon sensing of AMPs by the PhoPQ two-component system (Dalebroux and Miller,

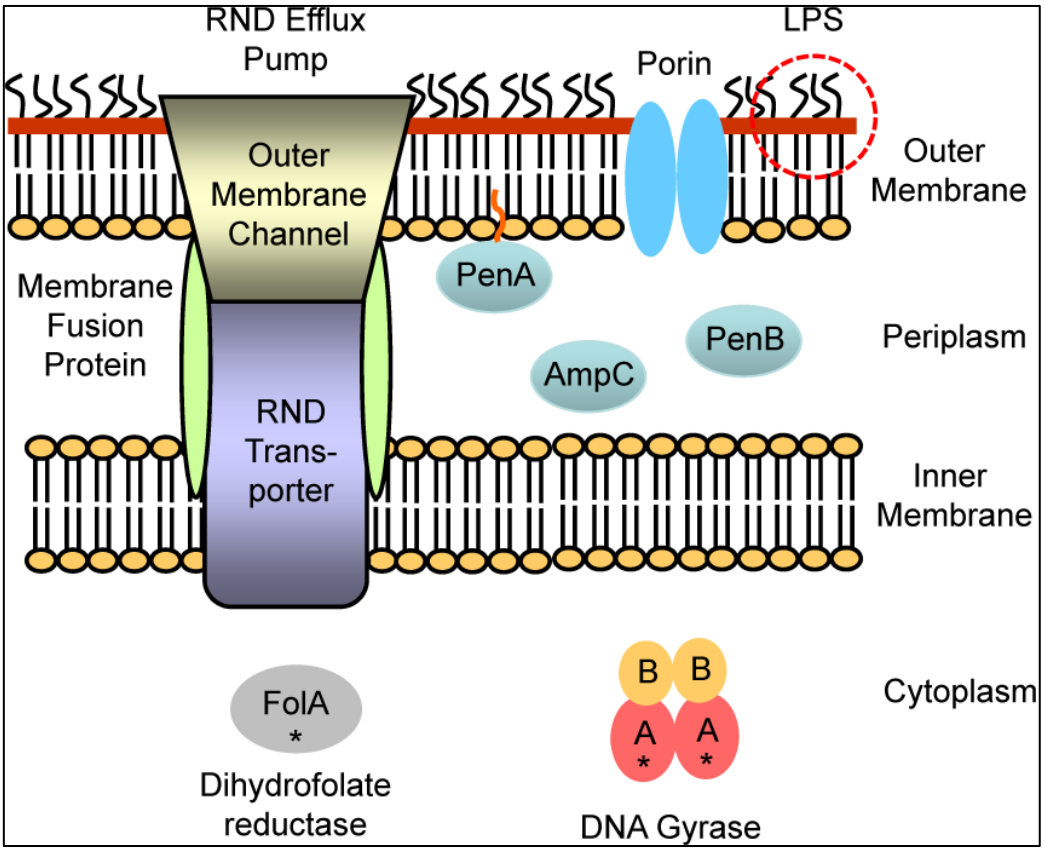


Figure 16: Resistance mechanisms against antibiotics in *Burkholderia* species.
Taken from *Rhodes and Schweizer, 2016*.

2014). Another resistance mechanism of *Burkholderia* is the modification of drug targets, known for the fluoroquinolones target (GyrA subunit of topoisomerase IV) and the trimethoprim target (dihydrofolate reductase, involved in tetrahydrofolate biosynthesis) (**Figure 16**) (Rhodes and Schweizer, 2016).

It is generally recognized that *Burkholderia* species share mechanisms to resist to polymyxins, which belong to the category of antimicrobial peptides. In *Burkholderia* species, these resistance mechanisms involve specific cell wall structures with the LPS and the hopanoids as well as the extracytoplasmic stress response, also known as envelope stress response (ESR) (Loutet and Valvano, 2011).

As mentioned before, modifications of the LPS molecules with the presence of the Ara4N moiety strongly contribute to the resistance towards polymyxins in *B. cenocepacia* (Ortega *et al.*, 2009; Rhodes and Schweizer, 2016). Another region of the LPS molecules, the core oligosaccharide, was also demonstrated to be involved in polymyxin B resistance in *Burkholderia* species (Burtnick and Woods, 1999; Loutet and Valvano, 2011; Loutet *et al.*, 2006; Ortega *et al.*, 2009). In *B. cenocepacia*, progressive truncations of the core oligosaccharide led to increasing sensitivity to polymyxin B (Ortega *et al.*, 2009). This sensitivity is particularly high when the first sugar moieties of the core oligosaccharide are lost in *B. cenocepacia* mutants targeting the first steps of the LPS core oligosaccharide biosynthesis (Loutet *et al.*, 2006). In another study, it was reported that a *B. pseudomallei* mutant in the *waaF* gene, encoding the heptosyltransferase II involved in the LPS core oligosaccharide biosynthesis (see section **2.4**), was more sensitive to polymyxin B compared to the wild-type strain (Burtnick and Woods, 1999). It is thus striking that the LPS modifications of *B. insecticola* needed for proper colonization of the M4 crypts are the same as those in other *Burkholderia* species to resist polymyxin B.

In some *Burkholderia* species, another mechanism enables resistance to polymyxins that consists of the presence of hopanoids in the bacterial membranes. Hopanoids are pentacyclic triterpenoid bacterial lipids, analogous to sterols of eukaryotic membranes (Belin *et al.*, 2018; Kannenberg and Poralla, 1999). Interestingly, sedimentary hopanoids are massively abundant in rocks and are used as fossil molecules that testimony the presence of ancient life (Ourisson and Albrecht, 1992). These lipids are synthesized from the cyclization of squalene molecules and the following enzymatic steps lead to different hopanoid molecules, whose final

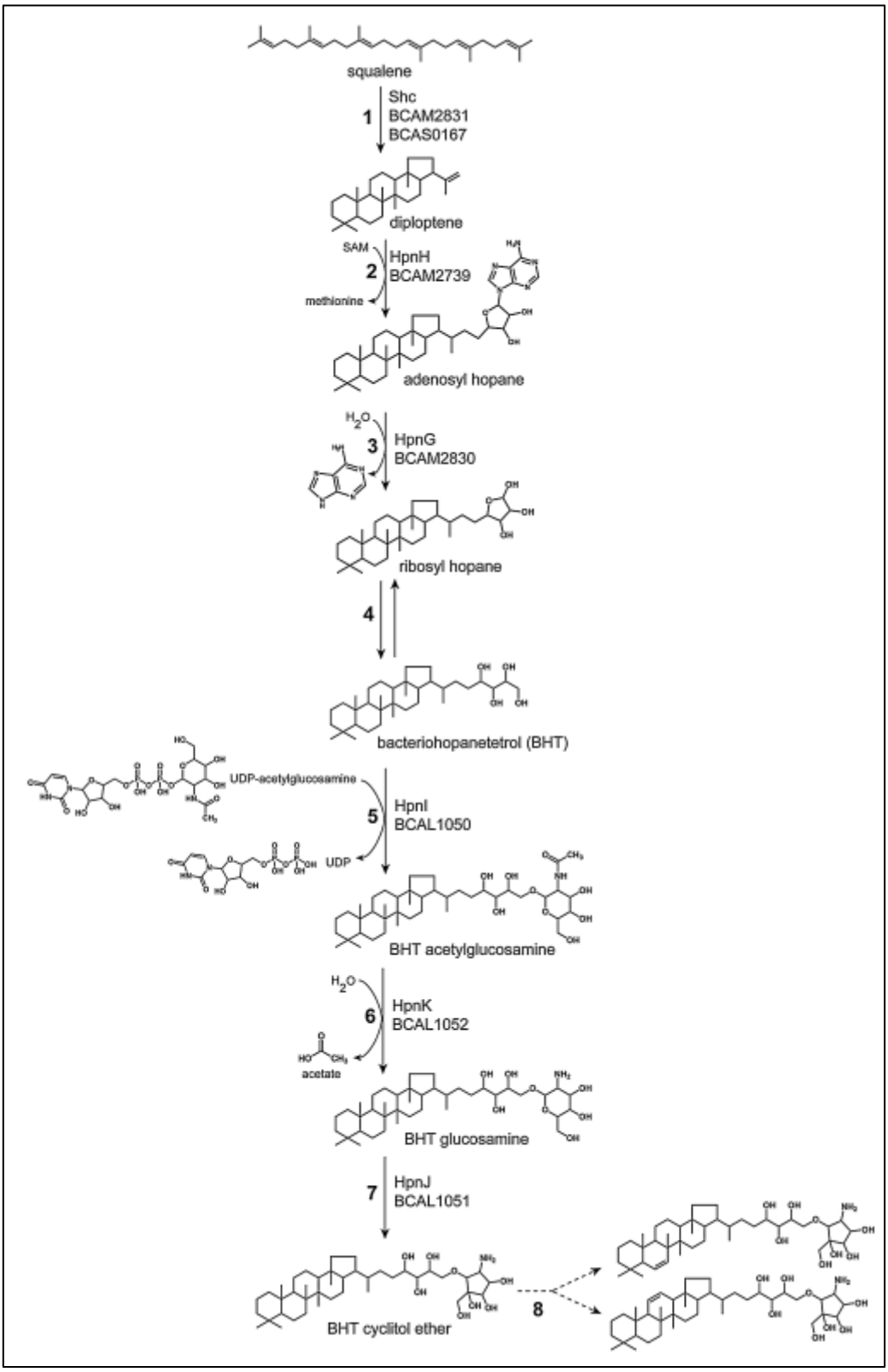


Figure 17: Proposed biosynthesis pathway of hopanoids in *B. cenocepacia*.
Taken from *Schmerk et al., 2015*.

structures vary, depending on the bacterial species (Belin *et al.*, 2018; Kannenberg and Poralla, 1999; Sahm *et al.*, 1993). These sterol-like molecules are found in both Gram-negative and

Gram-positive bacterial species (Kannenberg and Poralla, 1999; Poralla *et al., 2000; Sahm et al.*, 1993), but were mostly reported for numerous Gram-negative bacteria (Pearson *et al.*, 2007) such as *Desulfovibrio bastinii* (Blumenberg *et al.*, 2009), *Geobacter* species (Härtner *et al.*, 2005), *Bradyrhizobium diazoefficiens* (Kulkarni *et al.*, 2015) and several *Burkholderia* species (Cvejic *et al.*, 2000). In *Burkholderia multivorans*, hopanoids were found to contribute to the outer membrane permeability thus promoting resistance to polymyxins (Malott *et al.*, 2012, 2014). Similarly, in *B. cenocepacia*, Schmerk *et al*., demonstrated that hopanoids are required for polymyxin B resistance, bacterial motility and tolerance to low pH environments (Schmerk *et al*., 2011). More recently, Schmerk *et al*., have studied the hopanoid biosynthesis pathway in *B. cenocepacia* and unraveled the possible enzymatic steps involved in this pathway, from squalene to different hopanoids like diploptene and bacteriohopanetetrol (BHT) (**Figure 17**) (Schmerk *et al.*, 2015).

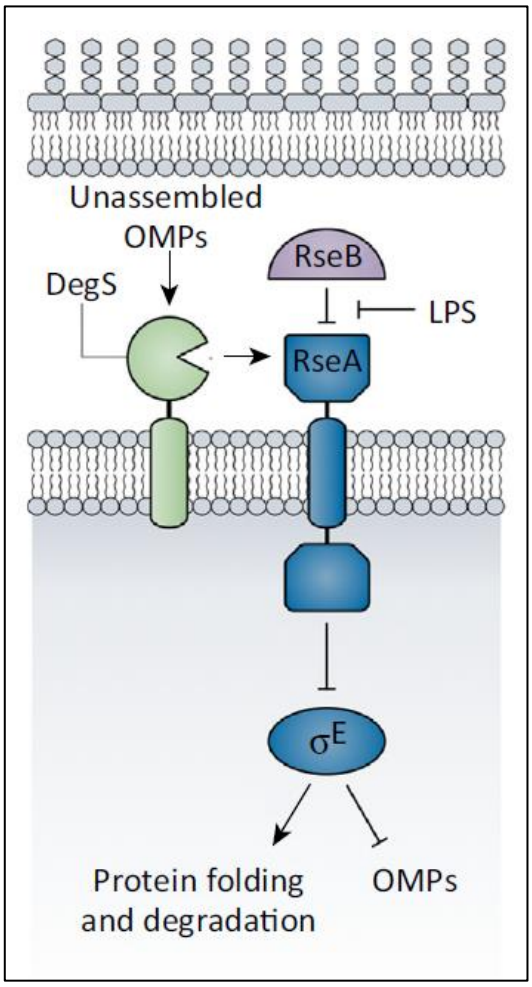


Figure 18: The ESR pathway of the σ^E response.
Taken from *Guest and Raivio, 2016*.

A third resistance mechanism to polymyxins in *Burkholderia* species involves the ESR. The ESR constitutes a signalling pathway that is activated when the outer membrane integrity is compromised by different environmental stresses such as temperature, pH, osmotic and oxidative variations or by the presence of misfolded and aggregated proteins in the periplasm (Flores-Kim and Darwin, 2014; Guest and Raivio, 2016; Raivio, 2005). To date, five ESR signalling systems are described in *E. coli*, the Bae, Cpx, Psp, Rcs and σ^E signalling pathways and these pathways are widely conserved in Gram-negative bacteria (Bury-Moné *et al.*, 2009; Guest and Raivio, 2016). These ESR pathways are regulating genes involved in biogenesis and repair of bacterial membranes, but can also modulate the cell motility and regulate the biofilm formation (Bury-Moné *et al.*, 2009; Guest and Raivio, 2016). One of these signalling pathways,

the σ^E (RpoE) response pathway was shown to be involved in polymyxin B resistance in *B. cenocepacia* (Flannagan and Valvano, 2008; Loutet *et al.*, 2011). In *E. coli*, this ESR pathway detects two kinds of perturbations in the outer membrane: misfolded outer membrane proteins (OMPs) which interact with the DegS periplasmic protease; or damaged LPS molecules that interact with the RseB protein, an anti-anti-σ factor (**Figure 18**) (Bury-Moné *et al.*, 2009; Flores-Kim and Darwin, 2014; Guest and Raivio, 2016). These two interactions trigger the signalling cascade by releasing the RpoE factor from the internal membrane, which then activates the transcription of different genes encoding chaperones, membrane biogenesis proteins, proteases and a small set of small RNAs that will downregulate the OMP production (**Figure 18**) (Flores-Kim and Darwin, 2014; Guest and Raivio, 2016). In addition to polymyxin B resistance, the RpoE factor in *Burkholderia* is also involved in heat stress response (Vanaporn *et al.*, 2008), oxidative stress and biofilm formation in *B. pseudomallei* (Korbsrisate *et al.*, 2005).

4. Antimicrobial peptides (AMPs)

4.1. Classes of AMPs

Antimicrobial peptides or AMPs are small ubiquitous molecules exerting antimicrobial activity that are secreted by all living organisms as part of their innate immune system. AMPs of eukaryotes are gene-encoded and synthesized by ribosomes while in prokaryotes, besides ribosomally-synthesized AMPs, also AMPs produced by non-ribosomal peptide synthetases exist (Hamidi, 2013). They have a broad range of activity against Gram-negative and Gram-positive bacteria, fungi and even parasites, and can modulate the immune system of their producers (Brogden, 2005; Lee *et al*., 2016). Due to their enormous diversity and large spectrum of targeted organisms, they generate a very broad interest for pharmaceutical applications. Therefore, most studies on AMPs are currently focused on the search for new molecules and on the design of modified molecules for drug development (Fjell *et al.*, 2011; Wang *et al.*, 2015).

AMPs can be classified based on their size, net charge, structure, level of hydrophobicity and amino acid composition, knowing that all of these physicochemical properties affect the AMPs' activity and target specificity (Bahar and Ren, 2013; Lee *et al.*, 2016). Regarding the net

charge, anionic AMPs are small peptides active against Gram-negative and Gram-positive bacteria, and are usually present in mucosal secretions, such as dermcidin in human sweat (Wang, 2014). Cationic AMPs exert antimicrobial activity against Gram-positive and Gram-negative bacterial species, but display much more variation in size and amino acid composition (Brogden, 2005). The majority of these cationic AMPs can be enriched in specific amino acids in their sequence like proline, arginine, lysine, glycine and cysteine (Bahar and Ren, 2013; Brogden, 2005).

As mentioned earlier, AMPs are produced by all organisms ranging from bacteria, archaea, fungi, plants, invertebrates and vertebrates (Bahar and Ren, 2013; Brogden, 2005). In bacteria and archaea, AMPs known as bacteriocins and archaeocins respectively, contribute to interspecies competition and shaping microbial communities in specific ecological niches (Besse *et al.*, 2015; Chikindas *et al.*, 2018; Nishie *et al.*, 2012). In invertebrates, AMPs can be detected as potent toxins isolated from venoms of bees, wasps, spiders, scorpions and snakes such as the king cobra cathelicidin (Zhao *et al.*, 2018), and these molecules are specifically interesting for their pharmaceutical potential (Primon-Barros and José Macedo, 2017).

Also in the invertebrates, one of the first described AMPs was the insect AMP cecropin isolated from *Hyalophora cecropia* (Order: Lepidoptera) (Steiner *et al.*, 1981). The majority of insect AMPs are small cationic and amphiphilic molecules which can protect their host against a large set of pathogenic microorganisms (Bulet *et al.*, 1999; Wu *et al.*, 2018). Insect AMPs were mostly studied in *Drosophila melanogaster*, which comprise at least seven categories of AMPs, including cecropins, attacins, defensins, drosomycins, diptericins, drosocin and metchnikowins (Yi *et al.*, 2014). Cecropins, a general term derived from cecropin isolated from *H. cecropia* (Steiner *et al.*, 1981), are small peptides (approximately 35 amino acids) produced by various dipteran and lepidopteran species (Wu *et al.*, 2018). They exert antibacterial activity against both Gram-negative and Gram-positive bacterial species, and do not possess cysteines in their sequence (Wu *et al.*, 2018). Attacins, also discovered in *H. cecropia* (Hultmark *et al.*, 1983), are glycine-rich antimicrobial peptides which are active against Gram-negative bacteria (Wu *et al.*, 2018). Insect defensins are effective against both Gram-positive and Gram-negative bacteria, and display a conserved cysteine motif with six to eight residues which usually form three disulphide bridges and stabilize the molecule (Zhu and Gao, 2013). Excepting lepidopteran insects, these insect defensins were reported in hemipterans,

coleopterans, dipterans and hymenopterans (Bulet *et al.*, 1999; Hoffmann and Hetru, 1992; Wu *et al.*, 2018). Concerning drosomycins (Fehlbaum *et al.*, 1994) and metchnikowins (Levashina *et al.*, 1995), they were isolated in *D. melanogaster* and exhibit antifungal properties (Lemaitre and Hoffmann, 2007; Wu *et al.*, 2018). Another category of insect AMPs, the diptericins, represents glycine-rich AMPs isolated from the hemolymph of dipteran insects (Cudic *et al.*, 1999), and are active against a limited number of Gram-negative bacterial species (Wu *et al.*, 2018). The last category of insect AMPs is represented by its unique member, drosocin, which is produced by *D. melanogaster* (Wu *et al.*, 2018). Drosocin is a proline-rich AMP which is active against a broad range of microorganisms, such as Gram-negative bacteria, Gram-positive bacteria and fungi (Bulet *et al.*, 1996). Although drosocin constitutes a unique class of insect AMPs, it was reported that apidaecin IB from honeybees shared significant sequence homology with drosocin (Gobbo *et al.*, 2002).

In vertebrates, like humans, the immune system is much more developed than in other organisms, which suggests more complex and diverse set of secreted AMPs to protect their host against infections (Zhang and Gallo, 2016). In humans, AMPs are constitutively secreted by a large variety of tissues in mucosal surfaces, such as skin, eyes, saliva, lung airways, intestinal and urinary tracts (Wang, 2014). However, the expression of certain AMPs can vary depending on the host status like its age and the frequency of microbial infections (Wang, 2014). For example, the human β-defensin 2 (hBD-2) is overexpressed in the gingival epithelium of older individuals (Matsuzaka *et al.*, 2006). Furthermore, human AMPs such as defensins and cathelicidin (LL-37), can play a role in immune modulation of adaptive immune cells (Lai and Gallo, 2009). Hence, according to the host immune context, different sets of AMPs can be recruited to deal with various encountered infections during the host's life (Wang, 2014).

To understand the biological roles of AMPs, genetic manipulations on different hosts by deleting or silencing genes encoding AMPs were performed (Maróti *et al.*, 2011; Mergaert, 2018). For example, the inactivation of the CRAMP gene in mice (cathelin-related antimicrobial peptide), the analogue of the human cathelicidin, led to higher susceptibility to necrotic skin infections caused by *Streptococcus* group A bacteria (Maróti *et al.*, 2011; Nizet *et al.*, 2001). In *D. melanogaster*, multiple deletions were performed and enabled the generation of fly lines which lack different AMP combinations (Hanson *et al.*, 2019). It was shown that

multiple AMPs-deficient flies which lack drosocin, attacin and diptericin were more susceptible to systemic infection by the Gram-negative bacterium *Providencia burhodogranariea* than the single-AMP deficient flies (Hanson *et al.*, 2019). Hence, this study brought evidence of the synergistic immune activity of these different AMPs when facing a systemic bacterial infection, regardless to their individual protective role (Hanson *et al.*, 2019). During pathogenic infections, as the immune system recruits several immune cells and AMPs which may act together to eradicate the pathogen, it is of interest to study combinatorial AMPs mutations to decipher their contribution to the host defence (Hanson *et al.*, 2019; Maróti *et al.*, 2011; Mergaert, 2018).

4.2. Modes of action of AMPs

Considering bacterial species, it is generally assumed that the primary targets of AMPs are the bacterial membranes (Bechinger and Gorr, 2017; Brogden, 2005). Cationic AMPs, such as polymyxins, interact with negatively charged molecules at the bacterial surface through electrostatic binding, mostly provided by LPS and teichoic acids in Gram-negative and Gram-positive bacteria, respectively (Brogden, 2005). Following these electrostatic interactions, AMPs accumulate at the cell surface and different models are proposed to explain what happens after this attraction step. These models suggest two modes of action: pore formation and non-pore structures (Bechinger and Gorr, 2017).

Considering the pore formation models, AMPs can either interact together and form “barrel-stave” pores in the membranes (**Figure 19**), or directly insert themselves in the lipid bilayer without AMP-AMP interactions and form a “toroidal” pore (**Figure 19**) (Kumar *et al.*, 2018; Yu *et al.*, 2015). AMPs can also interact with the membrane without forming any pores, known as the “carpet” model (Brogden, 2005; Kumar *et al.*, 2018). This model proposes that AMPs cover the whole cell surface, hence creating a “carpet”, which can evolve into detergent-like model or toroidal pore model through membrane disruption (**Figure 19**) (Kumar *et al.*, 2018). All of these models share the same final pattern with disruption of the membrane integrity, which mediates killing by abolishing the membrane potential and even by cell lysis (**Figure 19**) (Brogden, 2005; Kumar *et al.*, 2018; Wu *et al.*, 2018; Yu *et al.*, 2015).

Even if membrane rupture is the major mode of action of most characterized AMPs, some eukaryotic AMPs, and most bacteriocins and archeocins can interact with intracellular targets

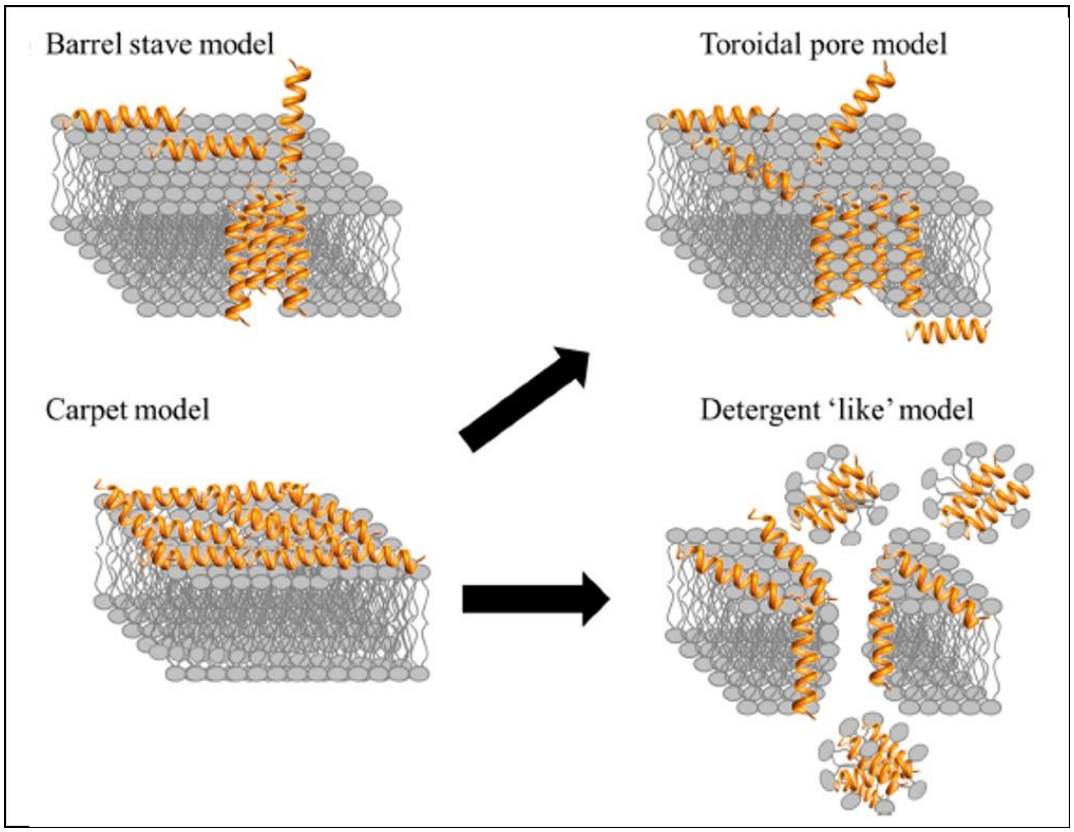


Figure 19: Models of interactions between cationic AMPs and bacterial cell surface.
AMPs are depicted in orange and bacterial surface in grey.
Taken from *Kumar et al., 2018*.

by penetrating through membranes in a non-damaging way (Falanga *et al.*, 2015). Once in the cytoplasm, these peptides can inhibit various vital intracellular pathways such as the cell wall, protein, RNA and DNA biosynthesis (Falanga *et al.*, 2015). Some peptides with intracellular targets can self-translocate across bacterial membranes (Kauffman *et al.*, 2015; Scocchi *et al.*, 2016). Hence, these AMPs are categorized as cell penetrating peptides or CPPs, such as dermaseptin from frog species which inhibit both protein and nucleic acid synthesis (Kumar *et al.*, 2018). Other peptides highjack bacterial transporters to cross the bacterial membranes, such as the proline-rich antimicrobial peptides (Graf *et al.*, 2017) or the bacteriocins (Jakes and Cramer, 2012; Nishie *et al.*, 2012). Due to the large diversity of AMPs, providing a precise mode of action of these peptides remains challenging and is still under investigation.

4.3. Antimicrobial peptides and symbiosis

While host AMPs are well-known since a long time for their primordial role in innate immunity and in the elimination of infecting microbes, more recently it became clear that eukaryotic

hosts also produce AMPs in response to the symbiotic bacterial populations they carry inside or at the surface of their bodies (Mergaert, 2018). The AMPs that are recruited in symbiotic associations can be the same ones that are involved in innate immunity, but some AMPs are specifically and only produced during long-term symbiotic associations, referred to as symbiotic AMPs (Mergaert, 2018). Since in symbiosis, hosts deliberately maintain symbiotic bacteria while in innate immunity they (try to) eliminate infecting bacteria with AMPs, what then can be the role of AMPs in symbiosis?

In the well-characterized *Rhizobium*-legumes symbiosis, some host plants like *Medicago truncatula* produce a specific family of AMPs named nodule-specific cysteine-rich (NCR) peptides, which are exclusively expressed inside the nodules (Mergaert *et al.*, 2003; Van de Velde *et al.*, 2010). This large family of peptides (up to 600 NCR peptides in *M. truncatula*) are short peptides of 30 to 50 amino acids long, many of them are cationic, and they possess a typical cysteine motif with four to six conserved cysteine residues that can form disulphide bridges (Kondorosi *et al.*, 2013; Mergaert *et al.*, 2003). The NCR peptides have antimicrobial activity against the *Rhizobium* symbionts but also against many other bacteria and even fungi and yeasts (Farkas *et al.*, 2017, 2018; Van de Velde *et al.*, 2010). However, in the legume nodules, the NCR peptides do not kill the bacteroids (term designating the intracellular rhizobia in nodules), but their function is to induce them into a specific differentiated and irreversible but metabolically active state, characterized by a blockage of the bacterial division, an amplification of the bacterial genome, a very strong cell enlargement, a high nitrogen fixing activity and a partial permeabilization of the bacterial membrane (Mergaert *et al.*, 2006; Van de Velde *et al.*, 2010). At concentrations that are below the killing activity, NCR peptides can induce these features *in vitro* (Kondorosi *et al.*, 2013). Thus, although individual NCR peptides showed antimicrobial activity against rhizobia species *in vitro*, the differentiated bacteria or bacteroids inside the nodules remain viable (Van de Velde *et al.*, 2010). This can be explained by the large expressed cocktail of NCR peptides which could act synergistically at smaller concentrations than *in vitro* conditions to maintain the symbiotic bacterial population (Kondorosi *et al.*, 2013). Moreover, the membrane structure of the bacteroids may undergo dramatic conformational and physicochemical changes which can contribute to NCR peptides resistance (Kondorosi *et al.*, 2013; Van de Velde *et al.*, 2010). In addition, it was shown that the ABC transporter BacA or BclA in rhizobia species is able to mediate the uptake

of NCR peptides inside the bacterial cell, thus limiting their potent antimicrobial activity to the cell surface (Barrière *et al.*, 2017; Guefrachi *et al.*, 2015; Haag *et al.*, 2011). As these bacteroids demonstrate an exaggerated elongated form with a high DNA content, these NCR peptides, which are internally transported in the bacterial cells, may regulate some bacterial genes involved in the cell morphology of the symbiotic bacteria inside the symbiosome (Barrière *et al.*, 2017; Kondorosi *et al.*, 2013; Van de Velde *et al.*, 2010).

Also in animal symbioses, AMPs are known to have key functions in the interaction (reviewed in Mergaert, 2018). In the *Vibrio*-squid model, the squid host produces a specific antimicrobial peptide, known as galaxin or EsGal1, in the light organ where the symbiont *V. fischeri* is maintained (Heath-Heckman *et al.*, 2014). Similar to NCR peptides, galaxin is a cysteine-rich peptide that was localized in the mucus layer of the ciliated epithelia and in the extracellular space of the light organ (also organized in crypts), where the symbiont proliferates (Heath-Heckman *et al.*, 2014). This peptide was shown to inhibit the growth of Gram-positive bacteria, but not Gram-negative marine species including *V. fischeri* (Heath-Heckman *et al.*, 2014). Such selective elimination may participate to reduce the number of environmental species present in the seawater and contribute to the selection of the desired symbiont (Heath-Heckman *et al.*, 2014).

Similarly, symbiotic AMPs were reported in different insect symbioses (Mergaert, 2018). The pea aphid *A. pisum* expresses AMPs exclusively in the bacteriocytes where the endosymbiotic obligate symbiont *Buchnera aphidicola* is stored (Shigenobu and Stern, 2013). These AMPs, known as bacteriocytes-specific cysteine-rich peptides or BCR peptides, comprise only seven peptide members which contain six to eight conserved cysteine residues (Shigenobu and Stern, 2013; Uchi *et al.*, 2019). Among these seven BCR peptides, only four of them had antimicrobial activity against *E. coli*, suggesting that each BCR peptide might play a different role during symbiosis (Shigenobu and Stern, 2013; Uchi *et al.*, 2019).

Cereal weevils (*Sitophilus* genus) also possess a specific endosymbiont named *Sitophilus* primary endosymbiont (SPE) which is sequestered in bacteriocytes that are grouped together to form a bacteriome (Charles *et al.*, 2001; Heddi *et al.*, 1998). Transcriptomic analyses revealed that one gene was particularly up-regulated inside the bacteriocytes, which is encoding a specific AMP called coleoptericin A or ColA (Anselme *et al.*, 2008). When the *colA* gene is silenced by RNAi, the endosymbiont is able to exit the bacteriocytes and invade the

surrounding tissues, which suggests that ColA play a role as a border controlling agent (Login *et al.*, 2011). Additionally, this endosymbiont exhibit the same elongated morphology as the *Rhizobium* symbiont of legumes (Login *et al.*, 2011). This phenotype is attributed to the ColA action which targets bacterial cytokinesis without inhibiting DNA replication, thus resulting in a form of "symbiont domestication" process (Login *et al.*, 2011).

Other examples of symbiotic AMPs are peptides secreted by cnidarians (*Hydra*), hematophagous annelids (leeches), and also by amoeba (Mergaert, 2018). In the medicinal leech *Hirudo verbana*, the gut is colonized by two γ-proteobacterial species, *Aeromonas veronii* and *Mucinivorans hirudinis* (Kikuchi and Graf, 2007; Nelson *et al.*, 2015; Worthen *et al*., 2006). It was demonstrated that both the host leech and the gut symbiont *Aeromonas veronii* secrete AMPs which enable a reciprocal protection against bacterial invaders and provide a suitable niche for the two gut symbionts (Tasiemski *et al.*, 2015). Belonging to the *Cnidaria* phylum, *Hydra* species are known to possess species-specific bacterial communities in their endodermal and ectodermal surfaces, which is directly in contact with the surrounding environment (Augustin *et al.*, 2010). During embryogenesis, *Hydra* expresses a specific category of AMPs named periculins which promote the establishment of the bacterial microbiota in their surfaces (Fraune *et al.*, 2010). Another category of AMPs produced by *Hydra* species, the arminins, and more specifically the arminin 1a showed a strong antibacterial activity against methicillin-resistant *Staphylococcus aureus* (Augustin *et al.*, 2009). It was observed that arminin-silenced *Hydra* species by RNAi have a decreased ability to select their native microbiota compared to the wild-type species (Franzenburg *et al.*, 2013). Thus, this result suggests that arminins participate in the selection and the composition of the surface microbiota in *Hydra* species (Franzenburg *et al.*, 2013; Mergaert, 2018). Similarly, the microbiota composition and containment in the gut of mammals strongly depends on the production of AMPs by the gut epithelial cells (Salzman *et al.,* 2010; Vaishnava *et al.*, 2011).

The amoeba *Paulinella chromatophora* which harbours a photosynthetic organelle derived from a cyanobacterial endosymbiont, constitutes a suitable model to study organellogenesis (C.M. Nowack, 2014). Proteomic analyses of this organelle named chromatophore showed that a specific group of short nuclear-encoded peptides, and thus produced by the amoeba host, are abundantly found in the chromatophore (Singer *et al.*, 2017). These imported peptides are AMP-like peptides, and might be involved in the control of the chromatophore

growth, division or metabolites exchange (Mergaert *et al*., 2017; Singer *et al.*, 2017).

All of the above-described symbiotic AMPs share similarities, being mostly cationic peptides, exhibiting antimicrobial activity and targeting their host specific symbiotic bacteria. Hence, symbiotic AMPs have key functions in selecting, maintaining and controlling symbiotic communities to promote specific host-symbiont relationships (Mergaert, 2018; Mergaert *et al.*, 2017).

4.4. Immunity and antimicrobial peptides in *Riptortus pedestris*

In insects, the best characterized immune system is from the holometabolous (superorder of insects that display complete metamorphosis) insect *Drosophila melanogaster* (Lemaitre and Hoffmann, 2007). As the stinkbug *R. pedestris* is a hemimetabolous (superorder of insects that display incomplete metamorphosis) insect, the understanding of its immunity is limited due to incomplete functional and genomic analyses. However, transcriptomic analysis in *R. pedestris* with the generation and annotation of cDNA libraries (Futahashi *et al.*, 2013) allowed to make connections with known immune factors from *Drosophila* studies (Lemaitre and Hoffmann, 2007). In addition, immune pathways of the brown-winged green stinkbug *Plautia stali* were recently identified by transcriptomic analysis, hence providing some knowledge on immunity of hemimetabolous insects (Nishide *et al.*, 2019).

Upon microbial infections, insects exhibit an immune system less complex than in mammals, which consists mainly of innate cellular and humoral immune mechanisms (**Figure 20**) (Lemaitre and Hoffmann, 2007). Concerning the humoral immunity, the main immune-response organ is the fat body, the equivalent of the mammalian liver, which produces AMPs and releases them in the hemolymph, a circulating fluid in the interior of the insect body that is in direct contact with the animal's tissues and that is analogous to the blood in vertebrates (Arrese and Soulages, 2010; Kanost, 2009; Lemaitre and Hoffmann, 2007). AMPs are also produced locally, in infected tissues (Lemaitre and Hoffmann, 2007). For example, the epithelial cells of the midgut produce AMPs and secrete them in the gut lumen when pathogenic bacteria are ingested by the insect (Lemaitre and Hoffmann, 2007). In addition to AMPs, pathogenic microbes can also induce the production of reactive oxygen species or ROS

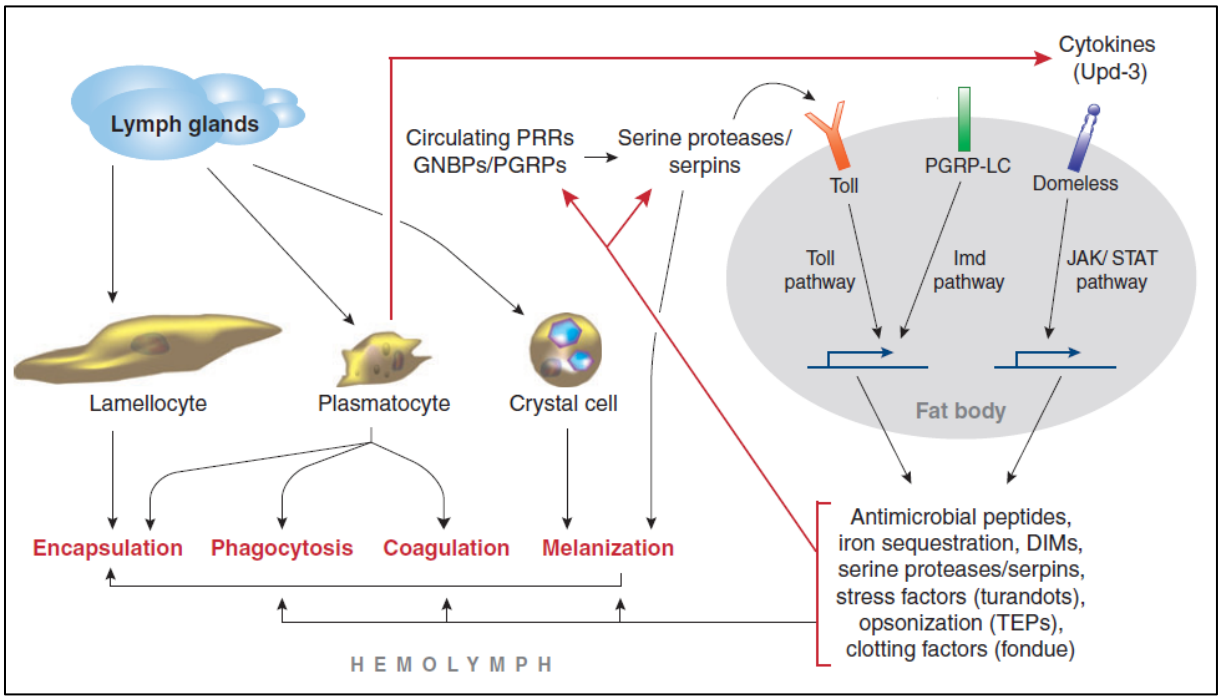


Figure 20: Overview of *Drosophila melanogaster* immune system.
Taken from *Lemaitre and Hoffmann, 2007*.

by cells nearby the infection site (Lemaitre and Hoffmann, 2007). In *Drosophila*, the cellular immunity is represented by hemocytes constantly circulating in the hemolymph (Kanost, 2009; Lemaitre and Hoffmann, 2007). These insect cells are able to eliminate invading microbes by phagocytosis and encapsulation (accumulation of immune cells at the microbial surface), but they also have additional roles in wound healing through coagulation and melanisation (**Figure 20**) (Kanost, 2009; Lemaitre and Hoffmann, 2007). Both the fat body and hemocytes are present in *R. pedestris*, though the different subgroups of immune cells are not defined (**Figure 20**) (Kim *et al.*, 2015b).

In the fat body of *Drosophila*, the regulation of AMP production is governed by two main signalling pathways called the Toll and Imd pathways (**Figure 20**) (De Gregorio *et al.*, 2002; Lemaitre and Hoffmann, 2007). These signalling cascades are activated upon sensing microbial patterns at the host cell surface through specific receptors known as pattern recognition receptors or PRRs, such as PGRPs (peptidoglycan recognition proteins) and GNBPs (Gram-negative binding proteins) (Lemaitre and Hoffmann, 2007). The Toll pathway controls the NF-κB transcription factors Dorsal and Dif while the Imd pathway controls the NF-κB transcription factor Relish. These transcription factors regulate the expression of distinct sets of AMP genes. The Toll pathway is induced in the presence of yeasts, filamentous fungi and Gram-positive

bacteria (Rutschmann *et al.*, 2002), whereas the Imd pathway is activated by sensing Gram-negative bacterial species (Gottar *et al.*, 2002). These two pathways and PGRPs are present in the stinkbugs *P. stali* (Nishide *et al.*, 2019) and *R. pedestris* and are currently under investigation.

Different AMPs are produced by *R. pedestris* which include riptocin, rip-thanatin, rip-defensin, lysozyme, two rip-trialysins and crypt-specific cysteine-rich peptides or CCR peptides (Futahashi *et al.*, 2013; Kim *et al.*, 2016b; Lee *et al.*, 2017; Park *et al.*, 2018; unpublished data). These molecules, except the CCRs which are produced in the midgut, are secreted by the fat body and can be localized in different organs and body fluids, such as the salivary glands, the hemolymph, the fat body and the midgut (Lee *et al.*, 2017).

The saliva constitutes one of the first physicochemical barrier for entomopathogens (Lee *et al.*, 2017). Two AMPs were isolated from the salivary glands: rip-trialysin-1 and rip-trialysin-2 (Lee *et al.*, 2017). These two peptides are strictly localized in the salivary fluid and exert antimicrobial activity against *E. coli* (Lee *et al.*, 2017). However, they are not effective against the *B. insecticola* symbiont, neither against the entomopathogenic *Serratia marcescens* (Lee *et al.*, 2017). Interestingly, *Serratia* is able to escape saliva defence mechanisms by cleaving trialysins, hence inducing strong hemolymph bacteremia followed by insect killing (Lee *et al.*, 2017).

Three AMPs were isolated from the hemolymph: riptocin, rip-defensin and rip-thanatin (Kim *et al.*, 2015a). All of these peptides are expressed in both the hemolymph and the fat body (Kim *et al.*, 2015a; Park *et al.*, 2018), but also in the symbiotic organ (M4 region) (unpublished data), and are up-regulated during a septic shock (Park *et al.*, 2018). It has been demonstrated that the *in vivo Burkholderia* symbiont is more sensitive to riptocin and rip-defensin than the cultured *in vitro Burkholderia* symbiont (Kim *et al.*, 2015a). Interestingly, rip-thanatin is active against a broad spectrum of microbes such as Gram-negative bacteria, Gram-positive bacteria and filamentous fungi, except on the cultured *Burkholderia* symbiont which is specifically resistant to rip-thanatin *in vitro* (Park *et al.*, 2018). Closely related to thanatin from *Podisus maculiventris*, this AMP possesses one arginine residue on the C-terminal region and two conserved cysteine residues which contribute to its antimicrobial activity (Park *et al.*, 2018). Silencing *rip-thanatin* gene in *R. pedestris* by RNAi during a septic shock showed that the titer of *Burkholderia* symbiont was dramatically increased compared to insects not infected by

pathogenic bacteria (Park *et al.*, 2018). In addition, *rip-thanatin* is highly expressed when the *Burkholderia* symbiont is present in the symbiotic organ (Park *et al.*, 2018). Thus, these results suggest that rip-thanatin may have a function in controlling the symbiont population upon pathogenic infection (Park *et al.*, 2018).

Transcriptomic analyses of different midgut regions of *R. pedestris* revealed that a specific category of AMPs is strictly expressed in the symbiotic crypts-containing organ (the M4 region) and also in the spatially-closed M4B region (**Figure 21**) (Futahashi *et al.*, 2013; unpublished data). These peptides were named CCR peptides (Futahashi *et al.*, 2013) in analogy to the symbiotic AMPs of legumes, the NCR peptides (Mergaert *et al.,* 2003), or of the pea aphid, the BCR peptides (Uchi *et al.*, 2019). Similar to these peptides, the CCR peptides are cationic cysteine-enriched peptides with six to eight conserved cysteine residues that can form disulphide bridges (**Figure 22**) (Futahashi *et al.*, 2013). Some of these peptides were tested and shown to exert antimicrobial activity (unpublished data). In total, there are 97 CCR peptides identified with various sizes, ranging from smaller peptides of 70-90 amino acids to larger molecules of 100-180 amino acids (Futahashi *et al.*, 2013). Interestingly, the CCR peptides are not activated in the fat body during an immune response (**Figure 21**). On the other hand, the typical innate immune peptides riptocin, rip-thanatin, and rip-defensin are strongly activated in the fat body during a septic shock response of the insect (**Figure 21**). This indicates that the CCR peptides are specific symbiotic peptides that are not involved in an immune response against pathogens. Moreover, this transcriptional pattern of the different *R. pedestris* AMP genes indicates that the M4 region does not trigger a typical immune response despite the massive presence of bacteria within this organ (**Figure 21**).

Nevertheless, it was recently demonstrated that pre- and post-molting stages of *R. pedestris* have a strong impact on the expression of riptocin, as well as c-type lysozyme, with a dramatic increase of antimicrobial activity against the *Burkholderia* symbiont in the M4 (Kim *et al.*, 2014b, 2016b). As mentioned before, the *Burkholderia* symbiont also exerts drastic envelope changes with a cocci-shape morphology (Ohbayashi *et al.*, 2019b) (see section **2.3**), but it was also demonstrated that these morphological changes were observed when *in vitro Burkholderia* cells were treated with M4 lysates (Kim *et al.*, 2015a). In addition, the LPS of *in vivo Burkholderia* cells are deprived of the O-antigen part (Kim *et al.*, 2015a) (see section **2.3**).

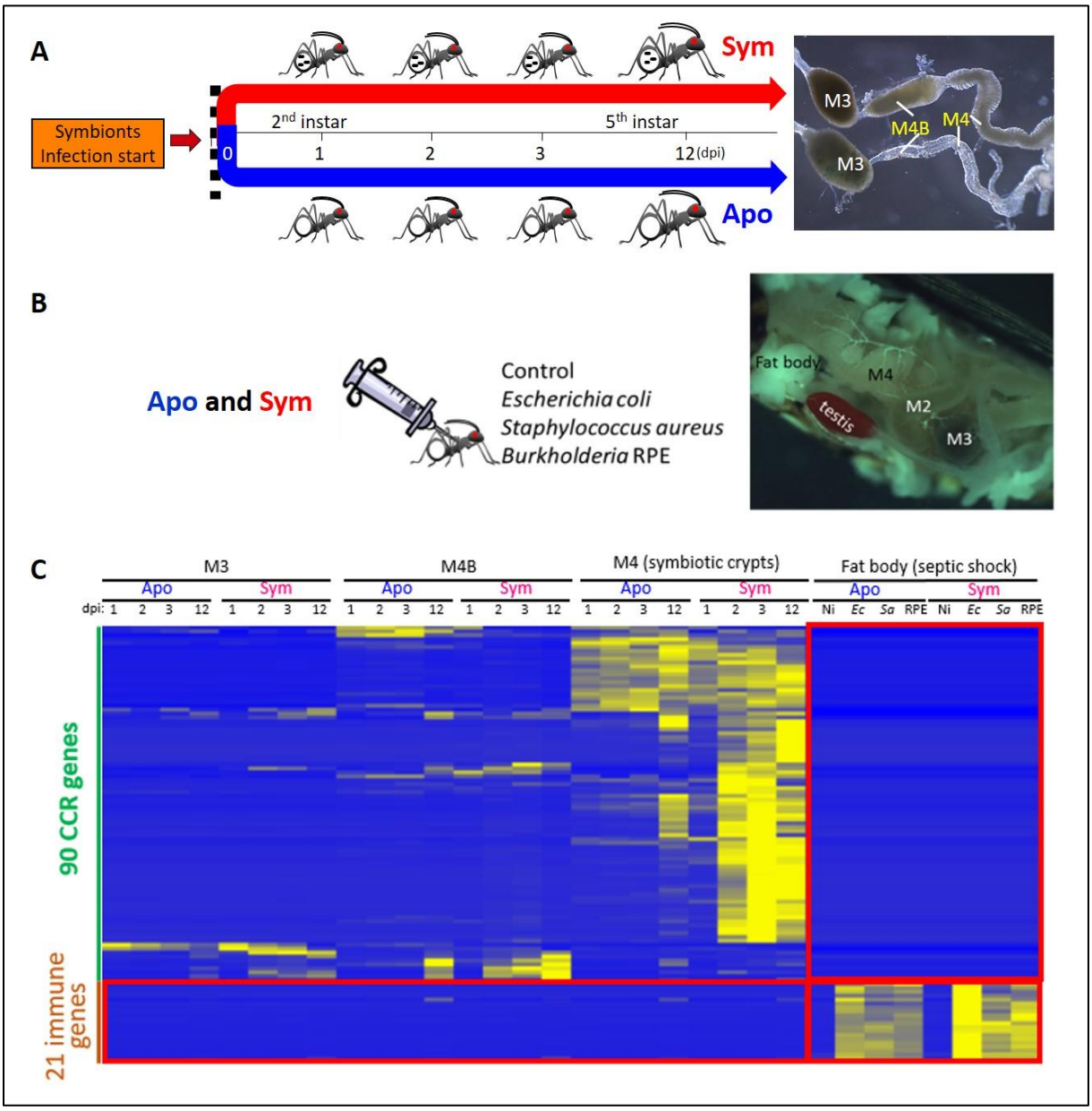


Figure 21: Transcriptomic analysis of different organs of *Riptortus pedestris* with a focus on immune-related genes.

A) Experimental setup of the different conditions studied for the host transcriptomic analysis. Host total RNA were extracted from three midgut compartments (M3, M4B and M4) in Apo and Sym insects at 1, 2, 3 and 12 dpi. B) Experimental setup of transcriptomic analysis on the fat body of Apo and Sym insects after a septic shock with different bacterial species (Control or non inoculated, *E. coli*, *Staphylococcus aureus* and *Burkholderia insecticola* or RPE) injected in the hemolymph of young second instar nymphs. C) Heatmap of the gene expression level of 90 CCR genes and 21 immune-related genes (including riptocin, rip-thanatin, rip-defensin). Blue: no gene expression, yellow: the gene is expressed. Abbreviations: dpi: days post-infection, Sym: symbiotic insects, Apo: aposymbiotic insects, Ni: non inoculated, Ec: *E. coli*, Sa: *Staphylococcus aureus*, RPE: *Riptortus pedestris* endosymbiont.

Taken from unpublished data.

```
                                                   C1          C2    C3    C4    C5  C6
Riptortus pedestris Rped-0033[AK416895]      (36)GWCDQWVT------QHDGMCNNHCRAHHYKSGYCDKDGVCHCKR
Bombyx mori defensin[NP_001037370]           (56)IWCEFEEA------TETAICQEHCLPKGYSYGICVS-NTCSCI-
Drosophila melamogaster defensin[NP_523672] (53)ATCDLLS----KWNWNHTACAGHCIAKGFKGGYCNDKAVCVCRN
Protophormia terraenovae defensinA[CAA39152](55)ATCDLLS----GTGINHSACAAHCLLRGNRGGYCNGKGVCVCRN
Aedes aegypti defensinA[AAL11703]            (59)ATCDLLS----GFGVGDSACAAHCIARGNRGGYCNSKKVCVCRN
Eristalis tenax defensin[CAM92111]           (55)ATCDLLS----FLNVKDAACAAHCLAKGYRGGYCDGRKVCNCRR
Sarcophaga peregrina sapecin[AAA29984]       (55)ATCDLLS----GTGINHSACAAHCLLRGNRGGYCNGKAVCVCRN
Allomyrina dichotoma defenisn[AAB36306]       (1)VTCDLLSFEAKGFAANHSLCAAHCLAIGRRGGSCER-GVCICRE
Apis mellifera defensin-1[NP_001011616]      (44)VTCDLLSFKGQ---VNDSACAANCLSLGKAGGHCEK-VGCICRK
Tribolium castaneum defensin1[XP_967194]     (40)VTCDLLSAEAKGVKVNHAACAAHCLLKRKRGGYCNKRRICVCRN
Zophobas atratus defensin[P80033]             (1)FTCDVLGFEIAGTKLNSAACGAHCLALGRRGGYCNSKSVCVCR-
Pyrrhocoris apterus defensin[P37364]          (1)ATCDILSFQSQWVTPNHAGCALHCVIKGYKGGQCKI-TVCHCRR
Tenebrio molitor Tenecin1[BAA04552]          (42)VTCDILSVEAKGVKLNDAACAAHCLFRGRSGGYCNGKRVCVCR-
                                                  *                 *   *         *     * *
```

Figure 22: Conservation of specific cysteine motifs in CCR peptides of *Riptortus pedestris*. Alignments of one CCR peptide (Rped-0033) with other insect defensins amino acid sequences. Conserved cysteine residues are shown in red. Possible disulphide bridges are illustrated above. Each number in parentheses corresponds to the amino acid position in the peptide sequence.
Taken from *Futahashi et al., 2013*.

Thus, these results suggest that unidentified host factor(s) in the M4 region may act on the alteration of the membrane integrity of symbiotic bacteria (Kim *et al.*, 2015a).

Thus taken together, a large diversity of AMPs are produced by *R. pedestris* which exhibit antimicrobial activities towards bacterial species, including the *B. insecticola* symbiont. This likely implies that resistance towards antimicrobial peptides is a crucial feature of the symbiont for infecting the symbiotic organ and chronically establishing within its host.

Thesis objectives

The importance of the diverse AMPs from *R. pedestris* in the establishment and the maintaining of the symbiotic association with *B. insecticola* seems to be obvious from the expression pattern of these peptides. Therefore, we hypothesized that these AMPs, including the CCR peptides, may participate to the specific colonization of the symbiont *B. insecticola* and that resistance of these bacteria to AMPs is a key feature that allows them to be competent for the symbiosis.

In order to validate or to reject this hypothesis, I had two main axis of work during my thesis:

- Determine the bacterial factors involved in antimicrobial peptides resistance for *B. insecticola* ;
- Identify the symbiotic functions required for the host colonization, and find if a correlation subsists between the ability of the symbiont to resist antimicrobial peptides and its host colonization efficiency.

In this work, I used a Tn-seq approach on the *R. pedestris* symbiont. Prior to start this work, I had to check the efficiency of this new method and to implement in the laboratory the bioinformatic tools required to analyse the high-throughput sequencing data generated by Tn-seq.

In the first chapter of the results, I introduce the Tn-seq methodology and its usage for the description of the essential genome of *B. insecticola*.

In the second chapter, I describe different bacterial factors involved in antimicrobial peptides resistance of *B. insecticola,* based on a candidate-gene approach and on the Tn-seq method.

Finally, in the third chapter, I present the study of the bottleneck on the symbiotic population during the infection of the symbiotic organ and the identification of bacterial genes involved in *R. pedestris* colonization by Tn-seq.

Chapter II

Inferring the Essential Genome of *Burkholderia insecticola* by Transposon Sequencing

1. Introduction

Knowing that bacterial genomes are submissive to genomic changes to cope with variable environmental conditions and adapt themselves to multiple niches, this raises the question about which genes become dispensable and which genes are required for the bacterial viability (Gil *et al.*, 2004). For obligate symbiotic interactions, it was previously reported that some bacterial symbionts were shown to possess eroded genomes, such as *Buchnera aphidicola* (Carlier and Eberl, 2012; Manzano-Marín *et al.*; Moran and Mira, 2001). As the genomes of these symbiotic bacteria are strongly reduced compared to other bacterial species, notably because of a strong coevolutionary process with their respective host (Moran and Mira, 2001), it suggests that the minimal essential genome to sustain the bacterial viability for symbiotic bacteria is probably different from other bacterial species. Hence, the study of these essential bacterial genomes, especially in a symbiotic context, can help to understand what are the minimum fundamental cellular functions required for bacterial species to promote their survival and their growth.

Recently, it has become possible to screen essential genes on a genome-wide scale with high-throughput sequencing methods known as Tn-seq or transposon-sequencing (van Opijnen *et al.*, 2009), INSeq or insertion-sequencing (Goodman *et al.*, 2009), RB-TnSeq or random bar code transposon-site sequencing (Wetmore *et al.*, 2015), HITS or high-throughput insertion track by deep-sequencing (Gawronski *et al.*, 2009) and TraDIS or transposon-directed insertion site sequencing (Langridge *et al.*, 2009). These approaches rely on the creation of a (saturated) transposon mutant library which will be grown as a pool in a defined condition, so that the output bacterial population recovered will be compared to the initial library by high-throughput sequencing (Chao *et al.*, 2016). In the initial library, these approaches enable to identify the essential genes of the bacterium studied, which represent the genes that are strictly required to sustain the bacterial viability to promote an optimal growth condition. In other terms, mutations in these essential genes would be lethal for the bacteria, and thus, do not exist. On the other hand, by using these sequencing methods to compare the initial library to an output bacterial population grown in a defined condition, this leads to the identification of bacterial fitness genes required for the growth on this defined condition. Here, mutations in these genes would still lead to viable bacteria, present in the input population, but they

wouldn't be able to grow on the condition studied, and thus, would be absent in the output population.

These methods were mainly used to determine the essential gene sets of human pathogens because the characterization of essential functions can identify new putative drug targets, for example in *Pseudomonas aeruginosa* (Gallagher *et al.,* 2011; Lee *et al.*, 2015), *Mycobacterium tuberculosis* (Carey *et al.*, 2018; DeJesus *et al.*, 2017), *Staphylococcus aureus* (Valentino *et al.*, 2014; Wilde *et al.*, 2015), *Campylobacter jejuni* (Gao *et al.,* 2017; de Vries *et al.*, 2017), *B. cenocepacia* (Higgins *et al.*, 2017; Wong *et al.*, 2016) and *Vibrio cholerae* (Chao *et al.,* 2013; Fu *et al.*, 2013; Kamp *et al.*, 2013). In addition to *in vitro* growth conditions, it is also possible to identify with these methods fitness genes involved in virulence and pathogenesis during *in vivo* experiments (Gutierrez *et al.*, 2015; Skurnik *et al.*, 2013). Nonetheless, only a few symbiotic bacteria were studied through transposon sequencing methods, like the *Vibrio* symbiont of squid (Brooks *et al.*, 2014; Lyell *et al.*, 2017), *Snodgrassella alvi* of the honey bee gut microbiota (Powell *et al.*, 2016), the plant growth-promoting rhizobacterium *Pseudomonas simiae* (Cole *et al.*, 2017) or *Borrelia burgorferi* (the causative agent of Lyme disease in humans) in its tick host *Ixodes scapularis* (Phelan *et al.*, 2019).

Here, I report on the construction of a *B. insecticola* transposon mutant library and the identification of essential genes using Tn-seq. I obtained a large-scale transposon bacterial population of the *R. pedestris* symbiont, which enabled to define its essential genome in rich medium and also to obtain additional fitness gene lists involved in glucose or succinate exploitation. The essential genes from *B. insecticola* were compared to previously published essential gene repertoires in diverse *Burkholderia* species, and this comparison provided an overview of fundamental cellular functions shared between *Burkholderia* species. This work is the first step towards a better understanding of *B. insecticola* adaptive mechanisms for its survival in a context of the interaction with its host.

2. Contributions

Quality control and sequencing of the Tn-seq samples were performed by the I2BC sequencing platform (CNRS Gif-sur-Yvette, France). The *Escherichia coli* MFD*pir* strain carrying the plasmid pSAM_Ec, which used as a donor strain for transposon mutagenesis, was kindly provided by Erwan Gueguen (Université Claude Bernard Lyon 1, CNRS INSA). The *B. insecticola* Tn-seq library was constructed with the help of Quentin Nicoud (Master 1 student).

3. Materials and methods

Bacterial strains and growth conditions

The *B. insecticola* strain RPE64 and derivatives were routinely cultured in YG medium (5 g.L^{-1} yeast extract, 1 g.L^{-1} NaCl, 4 g.L^{-1} glucose) at 28°C. The modified strain *B. insecticola* RPE75 carrying a resistance to rifampicin (Rif) was used for transposon mutagenesis and cultured in YG medium supplemented with 30 µg.mL^{-1} of Rif at 28°C. The *Escherichia coli* MFD*pir* strain (Δ*dapA* derivative, auxotroph for diaminopimelic acid (DAP) synthesis) carrying the plasmid pSAM_Ec (Wiles *et al.*, 2013) was used as a donor strain for transposon mutagenesis, and cultured in LB broth (5 g.L^{-1} yeast extract, 10 g.L^{-1} tryptone, 5 g.L^{-1} NaCl) supplemented with 300 µg.mL^{-1} of DAP and 50 µg.mL^{-1} of kanamycin at 37°C. For cultures on solid medium, the media were supplemented with 1.5% agar. All strains were stored at -80°C in 20% glycerol for long-term conservation.

Transposon library generation

For transposon mutagenesis, we used the plasmid pSAM_Ec containing a modified *Himar1 mariner* transposon carrying the kanamycin (Km) resistance gene. The donor strain *E. coli* MFD*pir* pSAM_Ec and the recipient strain *B. insecticola* RPE75 were grown until exponential growth phase with a final OD_{600nm} of 1 in 50 mL cultures at 180 rpm. The cultures were washed twice by centrifugation at 4000 rpm for 10 minutes at 4°C and resuspensed in fresh medium. The pellets were resuspended in fresh medium to obtain a final OD_{600nm} of 50. For conjugation, the donor strain and the recipient strain were mixed at a ratio 1:1. The bacterial mix was spotted on YG agar plates supplemented with 300 µg.mL^{-1} of DAP and incubated at 28°C. After 1 hour of incubation (allowing conjugation of the pSAM_Ec plasmid from the donor *E. coli* strain to the RPE75 recipient strain and transposition of the transposon in the genome of the target strain), the spots were resuspended in YG medium, a dilution series was plated on a selective medium carrying Rif and Km and subjected to CFU counting to assess the number of individual bacterial mutants obtained by the mutagenesis. In parallel, the totality of the remaining bacterial suspension was spread on YG agar plates supplemented with 30 µg.mL^{-1} of Rif and 50 µg.mL^{-1} of Km to obtain the *B. insecticola* transposon mutant population. After 2 days of incubation at 28°C, the transposon library was resuspended from the agar plates in

fresh liquid YG medium. The suspension was adjusted to 20% glycerol, aliquoted and stored at -80°C.

Before further use, a quality control was performed on the library. The presence of the *mariner* transposon and the absence of the transposon donor plasmid pSAM_Ec was verified by PCR on 20 randomly selected transposon library clones. The transposon borders of these clones were amplified as described below and cloned in the pGEM-T Easy plasmid (Promega). For each library clone, 10 pGEM-T easy plasmid clones were sequenced, revealing that each clone had a single transposon, that the two borders of the transposon were obtained, that each clone had a distinct insertion and that the insertions were spread over the genome.

Tn-seq library screening for *in vitro* growth conditions

One aliquot of the Tn-seq library was diluted to obtain an initial OD_{600nm} of 0.01 into 20 mL of different nutrient-growth media. Three growth conditions were tested : YG rich medium corresponding to the input pool, minimal medium (1 $g.L^{-1}$ KH_2PO_4; 2 $g.L^{-1}$ K_2HPO_4; 1 $g.L^{-1}$ $(NH_4)_2SO_4$; 0.2 $g.L^{-1}$ NaCl; 0.1 $g.L^{-1}$ $MgSO_4$, $7H_2O$; 2.46 $mg.L^{-1}$ $FeSO_4$, $7H_2O$; 3.31 $mg.L^{-1}$ EDTA, 2Na; 50 $mg.L^{-1}$ $CaCl_2$, $2H_2O$) supplemented with 0.2% of glucose, and minimal medium supplemented with 0.2% of succinate as carbon sources, representing the output pools. These three growth cultures were incubated at 28°C, with shaking at 180 rpm. When the cultures reached an OD_{600nm} of 1, corresponding to approximately 7 generations of multiplication, bacteria were collected by centrifugation at 4000 rpm for 20 minutes at 4°C and the pellets were kept for DNA extraction. Each growth culture was performed in triplicates.

DNA extraction and preparation of the high-throughput sequencing libraries

Genomic DNA was extracted from the bacterial pellets using the MasterPure™ Complete DNA and RNA purification kit (Epicentre). Samples of 10 µg DNA were digested for one hour at 37°C with 1 µL of MmeI enzyme (2000 $U.mL^{-1}$, New England BioLabs, reference R0637L), 25 µL of 10X CutSmart buffer (New England BioLabs, reference B7204S) and 10 µL of S-adenosine-methionine (1.5 mM, New England BioLabs, reference B9003S) in a total volume of 250 µL. Subsequently, 1 µL of FastAP Thermosensitive Alkaline Phosphatase (1 $U.\mu L^{-1}$, ThermoScientific, reference EF0651) was added to the digestion mixes and samples were incubated for one additional hour at 37°C. The enzymes were then heat-inactivated at 75°C for 5 minutes. Digested DNA samples were

purified using the QIAquick PCR purification kit (QIAGEN). 700 ng of each digested DNA was ligated to specific barcoded adaptors (5 µM) (**Table 1**) using T4 DNA ligase (1 U.µL^{-1}, ThermoScientific, reference EL0016) in a final volume of 20 µL and incubated overnight at 16°C. The double stranded adaptors were prepared beforehand by mixing 25 µL of each corresponding single stranded primer at 200 µM (**Table 1**) and 1 µL of TrisHCl (100 µM, pH 8.3), denaturing the primers in the mixture at 92°C for 1 min and promoting the annealing of the complementary primers by gradual cooling of the samples (2°C per min) in a PCR thermocycler. Transposon borders were subsequently amplified by PCR from the adapter-ligated DNA samples using 1 µL of them as template. The PCR was performed for 22 cycles using the EuroBio Taq polymerase (5 U.µL^{-1}, reference GAETAQ00-4W) in a final volume of 20 µL according to the manufacturer's instructions, with 0.5 µM of the forward P7 Illumina primer (5'-CAAGCAGAAGACGGCATACGAGATAGACCGGGGACTTATCATCCAACCTGT-3') and 0.5 µM of the reverse P5 Illumina primer (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3').

The amplified products were purified by gel extraction on a 2.5% agarose gel using the QIAquick gel extraction kit (QIAGEN) and concentrated in a final volume of 30 µL. The concentration and the quality control of these Tn-seq samples were assessed by the I2BC sequencing platform (CNRS Gif-sur-Yvette, France) using Qubit fluorometric quantification (ThermoFisher) and a Bioanalyzer instrument (Agilent), respectively.

Sequencing and bioinformatics

The Tn-seq samples were mixed in equimolar amounts and sequenced by an Illumina NextSeq 500 instrument with 2 x 75 paired-end run at the I2BC sequencing platform (CNRS Gif-sur-Yvette, France). The generated data were demultiplexed using bcl2fastq2 software (bcl2fastq v2.15.0; Illumina, San Diego, USA) and FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Only read 1 from each sequenced fragment was used for further analysis. The 3' transposon sequence was trimmed using Trimmomatic (Bolger *et al.*, 2014), and reads with a length of 75 nucleotides were removed (reads without the transposon insertion). After the trimming step, reads with a length between 19 and 23 bp were reverse-complemented and only the reads starting with TA dinucleotides were mapped

using Bowtie (bowtie-1.1.2) (Langmead *et al.*, 2009; Li and Durbin, 2010) to the reference genome of *B. insecticola* (accession n° NC_021287.1 (chromosome 1), NC_021294.1 (chromosome 2), NC_021288.1 (chromosome 3), NC_021289.1 (plasmid 1), NC_021295.1 (plasmid 2)). BAM output files were sorted with Samtools (http://www.htslib.org/). FeatureCounts (Liao *et al.*, 2014) was used to evaluate the number of reads by gene. BAM output files were converted with Samtools on the Galaxy server (https://usegalaxy.org/) into non-binary SAM files, the appropriate format to use for further analysis.

The SAM files were analyzed with the ARTIST pipeline (Pritchard *et al.*, 2014) for *mariner* transposon data, working on MatLab software (The MathWorks, Natick, MA). The ARTIST pipeline provides two different types of analyses, El-ARTIST (Essential loci analysis) and Con-ARTIST (Conditionally-essential loci analysis).

El-ARTIST performs a within-sample analysis and identifies loci required for growth in the considered condition. First, the initial raw reads dataset is normalized with a sliding-window approach. Second, the El-ARTIST analysis compares the transposon insertion distribution with a theoretical model proposed by running a Hidden Markov Model (HMM), coupled with a Mann and Whitney U-test conducted with a p-value of 0.03 to identify essential genes. The *B. insecticola* mutant library grown in YG medium was analysed by El-ARTIST to define the essential gene set in this bacterium and to define the input pool used in Con-ARTIST.

Depending on the growth condition of the strain, *B. insecticola* can maintain multiple copies of its genome (Ohbayashi *et al.*, 2019b). Therefore, mutants could in principle carry a transposon insertion in a gene on one copy of a replicon and still maintain a wild-type allele

Replicate	Barcode sequence	Forward phospho-adaptor sequence	Reverse adaptor sequence
YG replicate 1	GGTAT	p-GGTATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTATACCNN
YG replicate 2	CTAGA	p-CTAGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTTCTAGNN
YG replicate 3	GCTGC	p-GCTGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTGCAGCNN
MM Glucose replicate 1	CTCGG	p-CTCGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTCCGAGNN
MM Glucose replicate 2	ATCTT	p-ATCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTAAGATNN
MM Glucose replicate 3	TACTA	p-TACTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTTAGTANN
MM Succinate replicate 1	CAGAC	p-CAGACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTGTCTGNN
MM Succinate replicate 2	TCACG	p-TCACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTCGTGANN
MM Succinate replicate 3	GAGCT	p-GAGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-p	TTCCCTACACGACGCTCTTCCGATCTAGCTCNN

Table 1: List of adaptor sequences and barcodes for the different Tn-seq conditions.

on another copy. The identification of essential genes on all replicons of the genome (except plasmid 2) indicates however that eventual heterozygosity in the transposon mutant library had been resolved before our Tn-seq analyses and that it should have no impact on the phenotypic analyses reported here.

Con-ARTIST was used to compare the reads distribution across the *B. insecticola* genome between the input pool and each of the two minimal medium conditions. First, simulation-based resampling is used to normalize the input pool dataset (YG medium) to account for random loss of mutants in the output pool condition (minimal medium or MM). Second, the Con-ARTIST analysis compares this normalized input dataset with the output pool dataset by training the previously generated HMM during the El-ARTIST analysis, coupled with a Mann and Whitney U-test performed with a p-value of 0.01. This comparison allows to associate genes and metabolic pathways required for the bacterial fitness under two different carbon sources.

The distribution of transposon insertions across the *B. insecticola* genome for all conditions was visualized using Integrative Genomics Viewer version 2.3.83 (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013). The circular genome representations were realised using the interactive web-based service ClicO FS (Cheong *et al.*, 2015). For comparative Tn-seq analysis, transposon mutagenesis data from *Burkholderia pseudomallei* strain K96243 (Moule *et al.*, 2014), *Burkholderia cenocepacia* strain J2315 (Wong *et al.*, 2016) and *Burkholderia thailandensis* strain E264 (Baugh *et al.*, 2013) were downloaded from supplementary materials of the respective publications. Orthologs of *B. insecticola* proteins in *B. pseudomallei*, *B. cenocepacia* and *B. thailandensis* were obtained with the Comparative Genomics tools in the MicroScope platform (Médigue *et al.*, 2017) by BlastP Bidirectional Best Hit with at least 35% identity on 80% of the query sequence.

4. Results

4.1. Transposon mutagenesis of the *Burkholderia* symbiont

For the construction of *B. insecticola* Tn-seq library, a *Himar1 mariner* transposon was used, which targets TA dinucleotides for insertion that is accomplished with an excision of the transposon from the donor site and a duplication of the TA dinucleotides in the acceptor site.

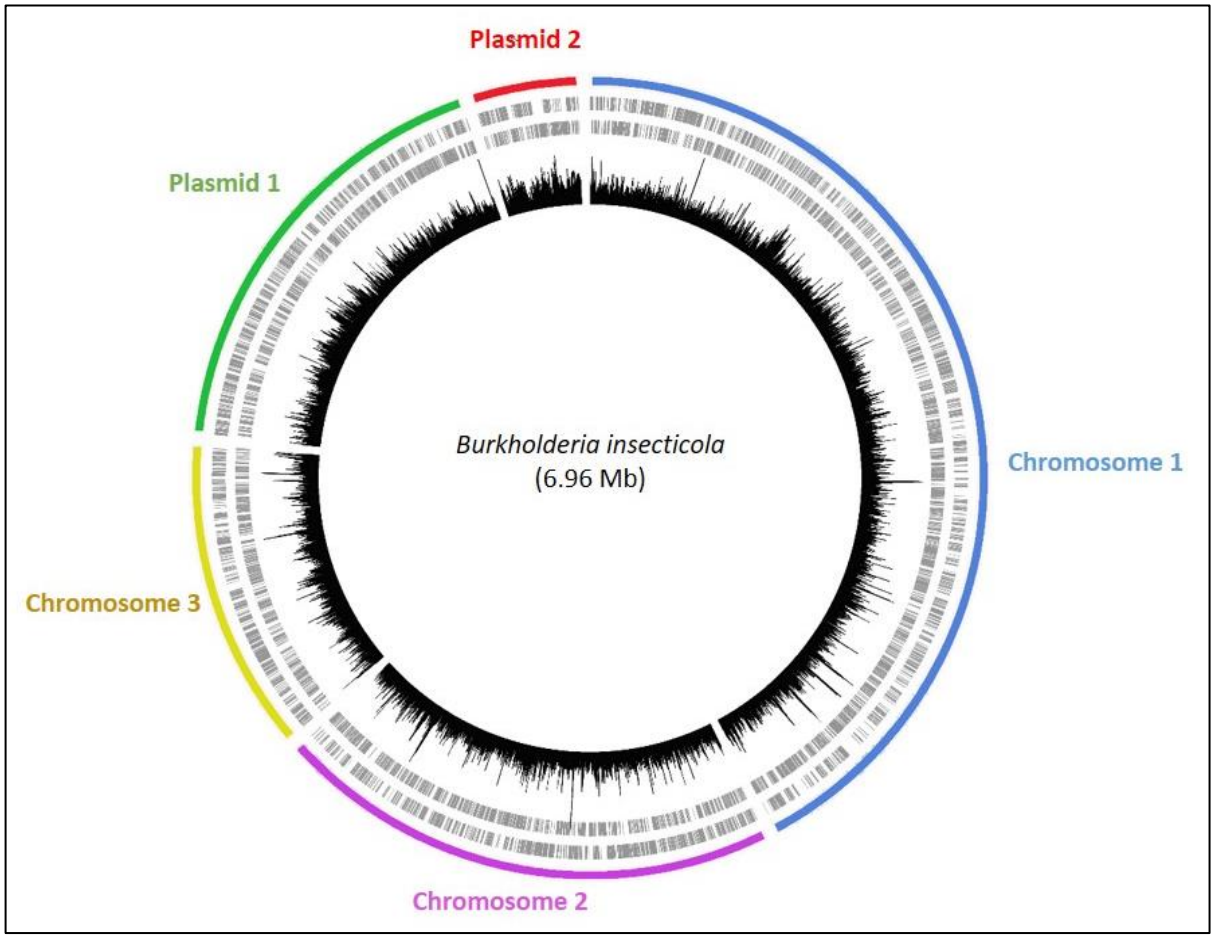


Figure 23: TA sites distribution on the *B. insecticola* genome.
Circular representation of the *B. insecticola* genome consisting of chromosome 1, 2 and 3 and plasmids 1 and 2. From outer to inner rings: forward CDS (grey bars), reverse CDS (grey bars), number of TA sites per kb (black histograms).

The genome of *B. insecticola* (6.96 Mb) contains 110,735 TA sites that are potential targets for *mariner* transposon mutagenesis, with 84,898 TA sites located in coding regions. Among the 6,352 genes, located on three chromosomes (chromosomes 1, 2 and 3) and two plasmids (plasmids 1 and 2), 6,244 genes contain TA sites, hence covering approximately the whole genome (**Figure 23**). The genes lacking TA sites are short to very short open reading frames varying in length from 33 to 513 nucleotides and encoding mostly peptides of unknown function. On the other hand, genes with TA sites have as a mean of 13.6 TA sites in their sequence. Thus as the proportion of genes without TA sites in the genome is small (1.7%) and the large majority of genes have a high number of TA sites, it was feasible to produce a genome-wide mutagenized transposon library with a good coverage of the *Burkholderia* symbiont using the *mariner* transposon.

Here, we used for mutagenesis the conjugative plasmid pSAM_Ec which contains a modified

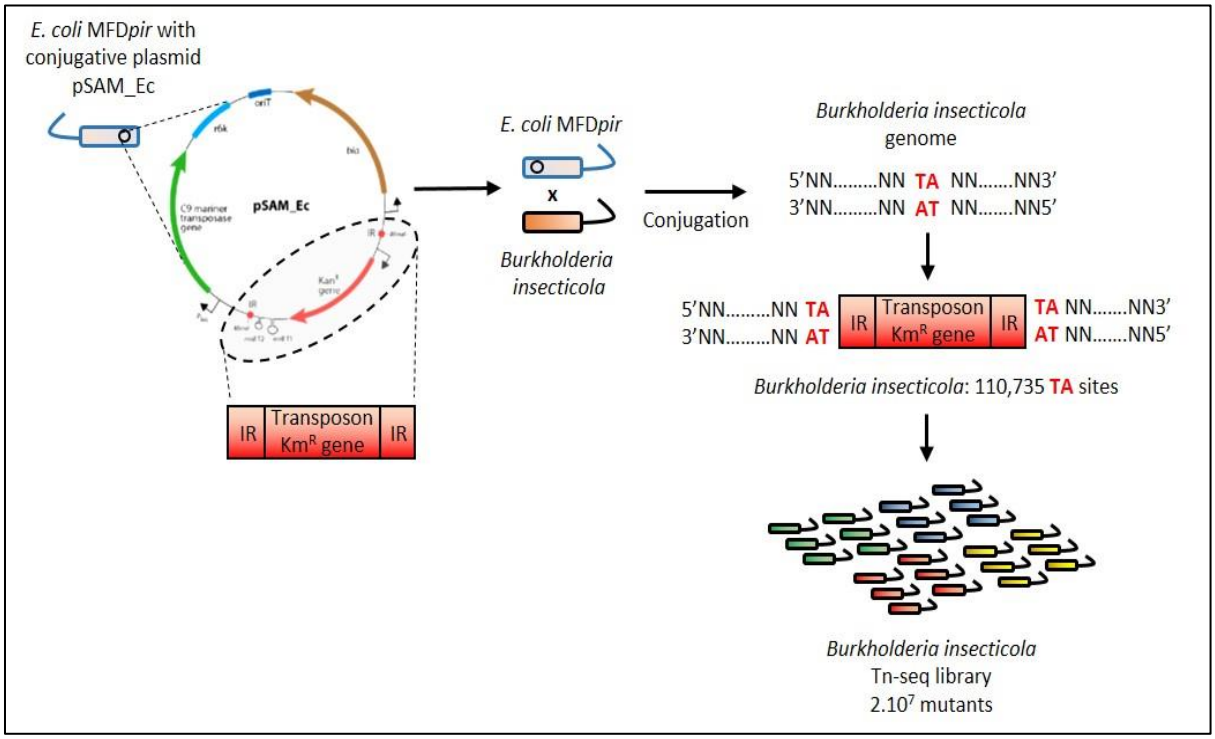


Figure 24: Schematic representation of the Tn-seq library construction.
For the pSAM_Ec plasmid characteristics: the Km resistance gene is indicated in red bordered by two IR indicated by red dots which contained the MmeI restriction sites, the C9 mariner transposase gene is indicated in green, the origin of replication r6K is indicated in light blue, the origin of transfer oriT is indicated in dark blue, the *bla* gene encoding a β-lactamase responsible for ampicillin resistance is indicated in orange.

transposon bearing a kanamycin resistance gene cassette, bordered by the *mariner* inverted repeats (IR) (**Figure 24**). After conjugation with the *B. insecticola* recipient strain, the transposon can insert in TA sites in the recipient genome (**Figure 24**). As the presence of the *pir* gene is required for the replication of the vector pSAM_Ec (which is present in the *E. coli* donor strain), the plasmid is not maintained in the donor strain and selection for kanamycin resistance allows to obtain mutants in which the transposon has been effectively inserted in the genome. Since the transposase gene, required for transposition, is located on the plasmid but not within the transposon itself (**Figure 24**), the transposons remain stable in the genome once the plasmid is lost.

After 50 independent conjugations performed between the donor strain and the *B. insecticola* recipient strain, we harvested 2.10^7 individual clones on YG medium, which corresponds to an 180-fold coverage of the total TA sites number in the genome. This transposon library was homogenized and finally concentrated to 2.10^{10} mutants per aliquot. Randomly selected

clones of the mutant library were verified for the presence of a single transposon inserted into the genome.

4.2. Genome-wide screening of essential functions in the *B. insecticola* symbiont

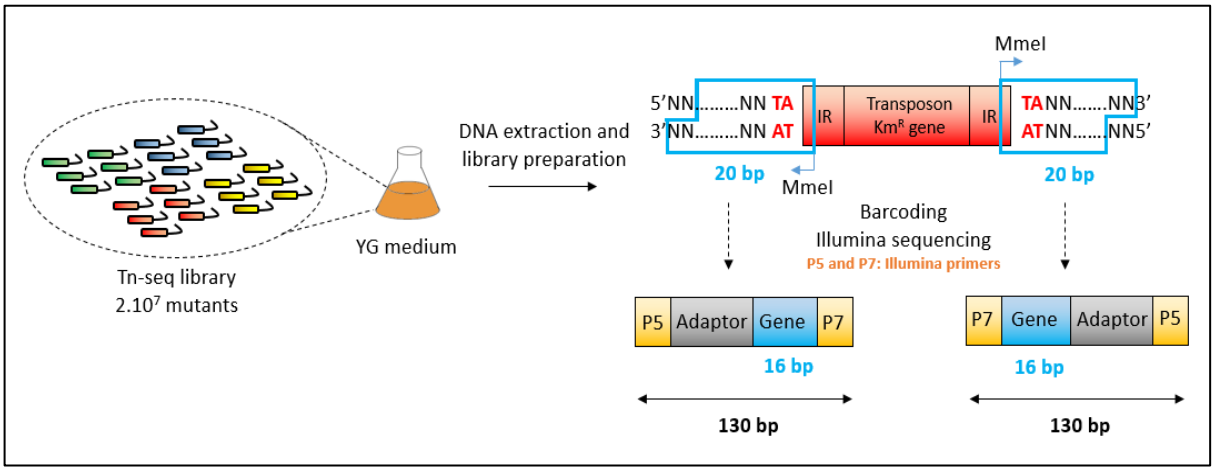


Figure 25: Schematic representation of Illumina sequencing library preparation.

As the transposon library was built in YG medium, I assessed the essential gene set required for the symbiont's proliferation on this rich medium by Tn-seq. I prepared liquid YG cultures of the transposon mutant population in triplicates, and prepared the bacterial DNA to sequence the transposon borders. The genomic DNA was digested with MmeI which is a type II restriction enzyme that cuts 20 bp upstream of its restriction site. Besides the recognition sites in the *B. insecticola* genome, two engineered MmeI sites are present in the *mariner* transposon located at the 4 bp from the TA site within the IR sequence, hence generating transposon fragments extended with 16 bp of genome tags and with two bases 3' overhangs (**Figure 25**). The obtained restriction fragments were ligated to an adaptor containing an experiment-specific barcode sequence, which is used to identify the associated experimental condition after sequencing. Finally, this product is PCR-amplified with adapter- and IR-specific primers that are extended with the P5 and P7 Illumina sequences generating a final fragment of 130 bp (**Figure 25**). After next-generation sequencing and trimming of the IR and adapter sequences, small sequence fragments of 16 bp, corresponding to the transposon insertion borders, are obtained and mapped to the *B. insecticola* genome. The number of read counts per TA site were determined and this count is taken as a relative estimate of the abundance

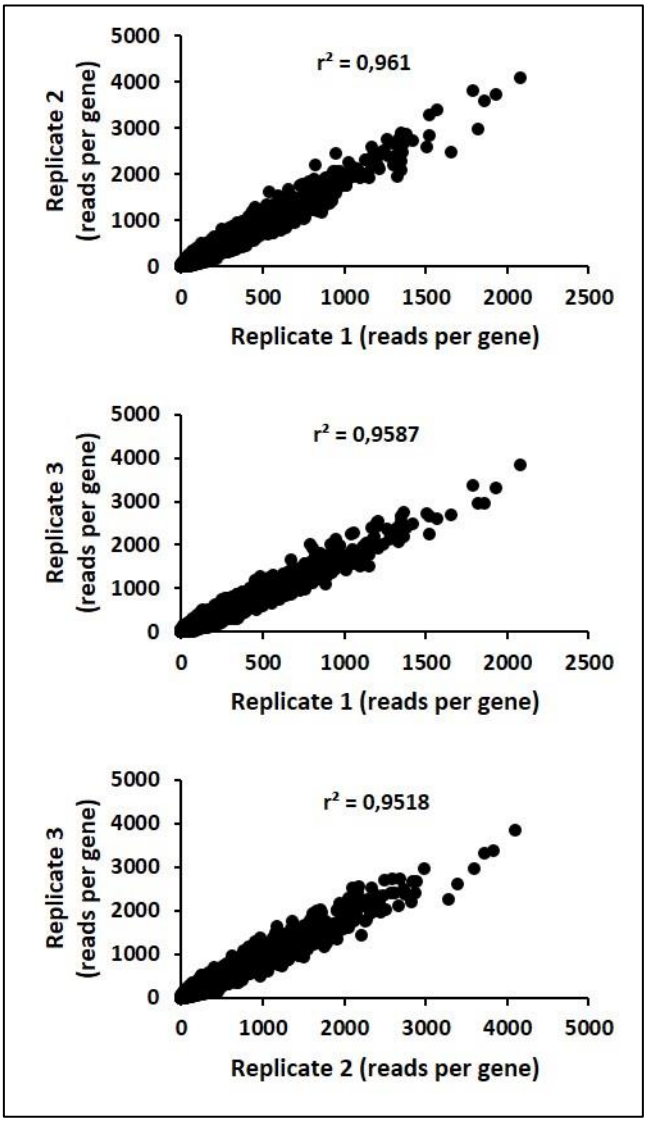


Figure 26: Correlations between read counts distribution in the three replicates of YG rich medium condition.

Dot plot representations of the comparison of each transposon insertions distribution between the three Tn-seq replicates of YG medium condition. The number of reads per gene is displayed for each replicate. The Pearson correlation coefficient r^2 was calculated for each comparison and indicated on each graph.

of the corresponding transposon mutants in the culture.

For each of the three replicates, approximately two million filtered reads were obtained. The correlation coefficient was high between these triplicates ($r^2 > 0.95$), allowing me to pool the sequencing data of the three replicates together for the El-ARTIST analysis (**Figure 26**). The pooled data contained 6,791,796 filtered reads with 88% of them aligning on the *B. insecticola* genome. Additionally, among the genes targeted by transposon insertions, almost 90% of their total TA sites were mutated (**Figure 27**). Hence, this Tn-seq library grown in rich medium

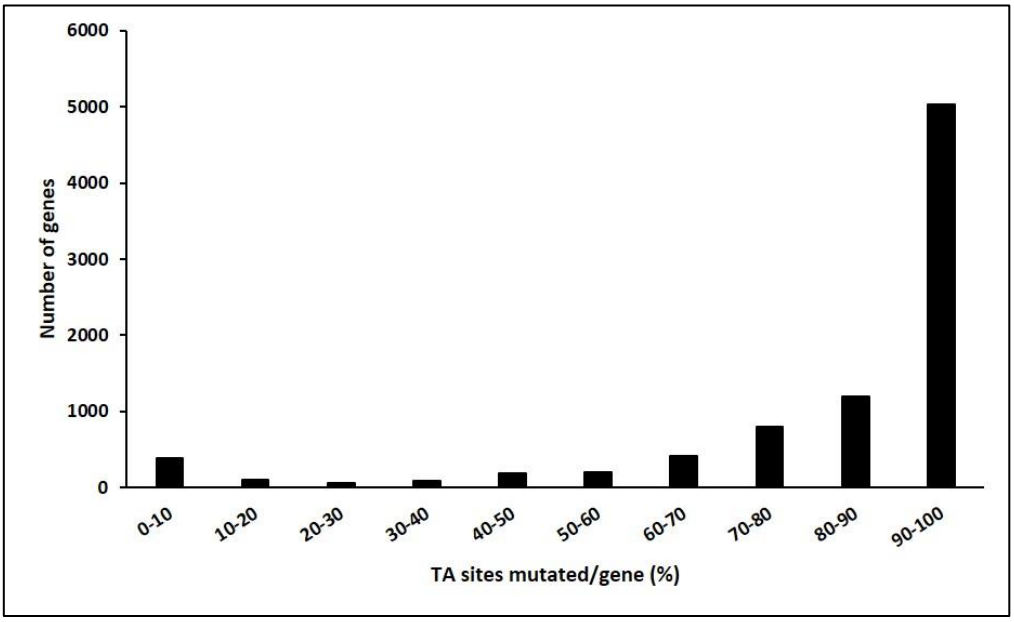


Figure 27: TA sites proportion targeted by transposon insertions in YG rich medium. The percentage of TA sites mutated for each gene was obtained by the El-ARTIST analysis and the total number of genes for each range of mutation percentage was calculated. The majority of genes (5,036 genes) has their entire TA sites targeted by transposon insertions.

displayed a true random transposon distribution according to van Opijnen *et al.* (2009).

Mutations, which provoke a loss in bacterial viability or a reduced fitness in this medium, will not be represented or under-represented in the sequenced population. Thus, the corresponding genes harbour less or no transposon insertions in the sequencing profile compared to the rest of the genome. Such genes are considered as essential for the bacterial survival.

The majority of the genome (4,966 genes) was categorized as “non-essential” by the El-ARTIST analysis (essentiality El-ARTIST score = 1 for “non-essential genes”) for the bacterial fitness on YG medium. 198 genes harboured a domain with a small number of transposon insertions which have a negative impact on the bacterial fitness (essentiality El-ARTIST score = 3 for “domain-essential genes”). A total of 1,080 genes were characterized as “essential” for the fitness on YG rich medium (essentiality El-ARTIST score = 2 for “essential genes”) (see **Annexe 1** for the list of essential genes in YG medium) (**Figure 28**). Moreover, I noticed that the distribution of reads was relatively weak in the chromosome 3 compared to the other replicons (**Figure 28**), which could mean that the chromosome 3 was less prone to transposon insertions, although the reason for that is unclear at present.

The 1,080 essential genes identified were mostly located on the chromosome 1 (44.35%) and

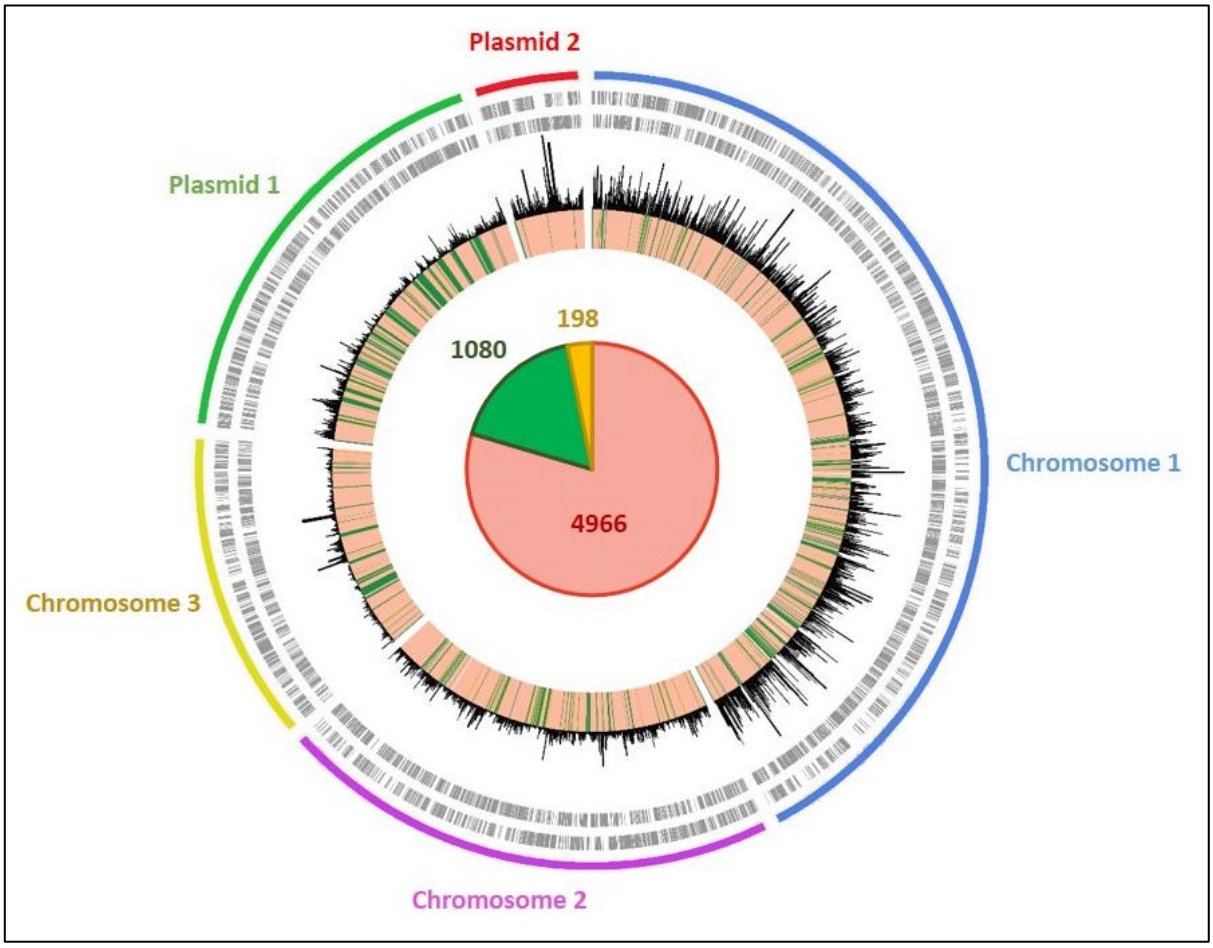


Figure 28: Essential genes required for *B. insecticola* growth on YG rich medium identified by El-ARTIST analysis.

Circular representation of the *B. insecticola* genome consisting of chromosome 1, 2 and 3 and plasmids 1 and 2. From outer to inner rings: forward CDS (grey bars), reverse CDS (grey bars), read counts per TA site (black histograms), El-ARTIST essentiality scores (heatmap, red : non-essential genes, yellow : domain-essential genes, green : essential genes), distribution of El-ARTIST essentiality scores for the whole genome (pie chart).

the plasmid 1 (29.91%). This essential gene set represented 17% of the genome, a number which is the same range as previous reports for different bacterial transposon mutagenesis studies (Christen *et al.*, 2014; DeJesus and Ioerger, 2013; Griffin *et al.*, 2011; Hooven *et al.*, 2016).

According to the COG (Clusters of Orthologous Genes) classification (Tatusov *et al.*, 2000), the most representative essential category was related to translation, ribosomal structure and biogenesis (J category) (**Figure 29**). Some members of this functional class are illustrated by genes encoding for 30S and 50S ribosomal proteins (*BRPE64_RS12355-BRPE64_RS12450*) (see the summary illustration below **Figure 33A**). The energy conversion and production category (C category) also contains many essential functions with genes involved in the respiration

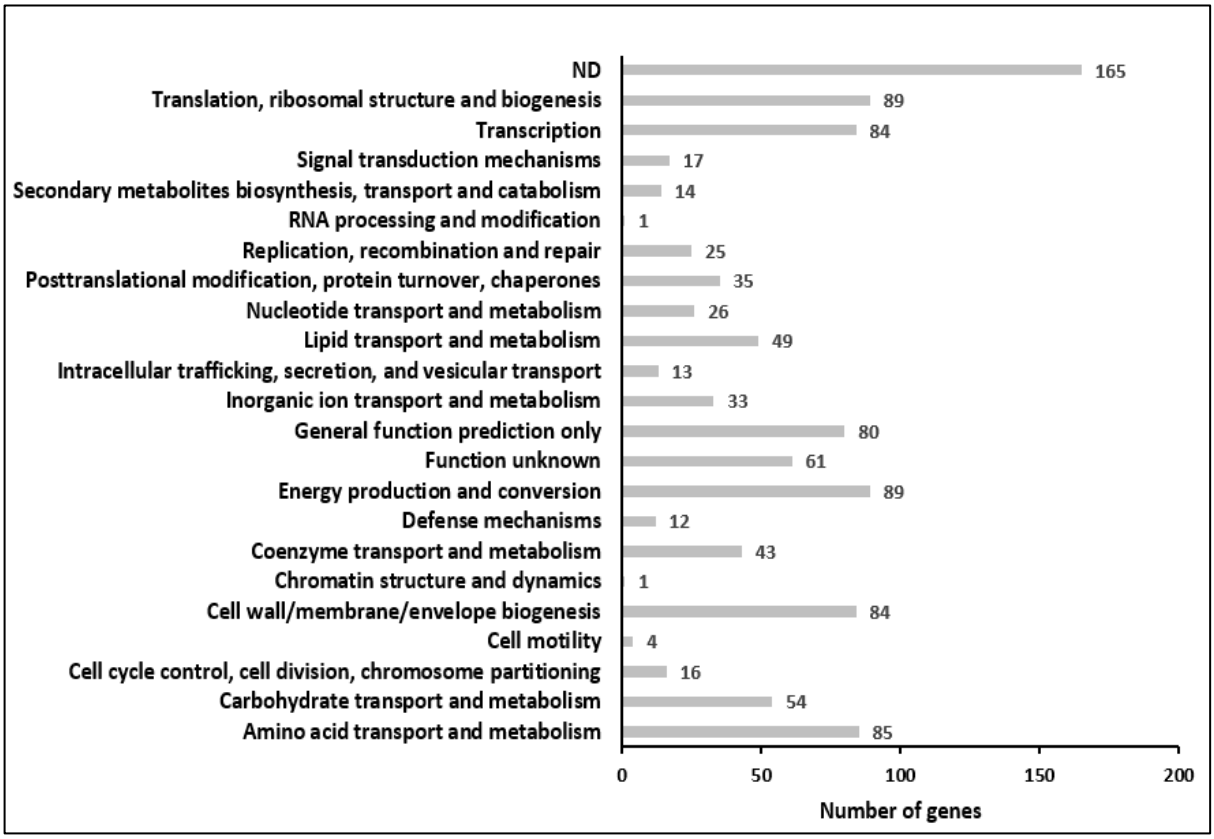


Figure 29: COG categories of essential genes identified in YG rich medium.
The total number of essential genes is indicated for each COG category.
ND: Not determined

process, like the ATP synthase subunits (*BRPE64_RS13345-BRPE64_RS13385*) (see the summary illustration below **Figure 33B**). Other highly represented categories are amino acid metabolism (E category) with multiple transporters (*e.g. BRPE64_RS05615*, *BRPE64_RS16125*, *BRPE64_RS22260*, *BRPE64_RS24435*, *BRPE64_RS25555*, *BRPE64_RS26255*), the transcription machinery (K category) with the RNA polymerase subunits (*BRPE64_RS03690, BRPE64_RS12320, BRPE64_RS12485, BRPE64_RS12490*), and cell wall biogenesis process (M category) including the genes involved in lipid A biosynthesis (*BRPE64_RS05785, BRPE64_RS05790, BRPE64_RS10785, BRPE64_RS11575*). It is striking that 306 genes (the combined categories ND, General function prediction only and Function unknown in **Figure 29)** (out of the 1,080 fitness genes) are encoding hypothetical proteins with unknown functions, mostly located on the chromosome 1 and plasmid 1.

The determination of the essential functions in the B. *insecticola* genome further provide some insights on the genome organization of this bacterium. The genome has five replicons, designated as chromosome 1, 2 and 3 and plasmids 1 and 2 (Shibata *et al.*, 2013).

Chromosome 1 has a typical chromosomal organization similar as the principal chromosome of *B. cenocepacia,* carrying on the replication origin locus with the genes *rpmH* (*BRPE64_RS14035*), *rnpA* (*BRPE64_RS14030*), *dnaA* (*BRPE64_RS00005*), *dnaN* (*BRPE64_RS00010*), and *gyrB* (*BRPE64_RS00015*), and on a nearby locus the chromosome partitioning genes *parA* (*BRPE64_RS13400*) and *parB* (*BRPE64_RS13395*) (Dubarry *et al.*, 2006). All these genes were found to be essential in my Tn-seq analysis. On the other hand, chromosomes 2 and 3 have, similar to plasmids 1 and 2, a plasmid-like replication origin locus, each carrying its own distinct *parABS* system and a plasmid-like replication protein (chromosome 2: *BRPE64_RS14050*, *BRPE64_RS14055*, *BRPE64_RS14060*; chromosome 3: *BRPE64_RS20740*, *BRPE64_RS20745*, *BRPE64_RS20750*; plasmid 1: *BRPE64_RS24690*, *BRPE64_RS24695*, *BRPE64_RS24700*; plasmid 2: *BRPE64_RS30485*, *BRPE64_RS30490*, *BRPE64_RS30495*). These *parABS* and replication protein encoding genes were found to be essential in this analysis for the chromosomes 2 and 3 and the plasmids 1 and 2, indicating that each of these replicons are specifically replicated and partitioned by their cognate machinery. It has to be noted that the essentiality of these genes in this case does not mean essential for the cell growth, but essential for the maintenance of the replicon. Nevertheless, since the chromosomes 2 and 3 and the plasmid 1 contain a large number of essential genes, these replicons are essential themselves. Thus, chromosomes 2 and 3 have plasmid-like features but carry essential genes. According to a new classification of bacterial replicons, the chromosomes 2 and 3 of the *Burkholderia* strain RPE64 should be named “chromid” (Harrison *et al.*, 2010; diCenzo *et al.*, 2017). On the other hand, the plasmid 2 is not essential and can be removed from the bacterium without affecting its fitness (see **Chapter IV**).

4.3. Identification of genes for growth on succinate and glucose as carbon sources

To further verify the *B. insecticola* library and check the robustness of the Tn-seq methodology, I tested different carbon sources to decipher which genes are involved in their exploitation (**Figure 30**). For that purpose, I incubated the *B. insecticola* transposon library in a minimal medium (MM) with either glucose or succinate as the sole carbon source. Since these metabolites are predicted to be imported by the cell via different transporters and they are integrated in different nodes of the central metabolism of the bacterium, I expected that

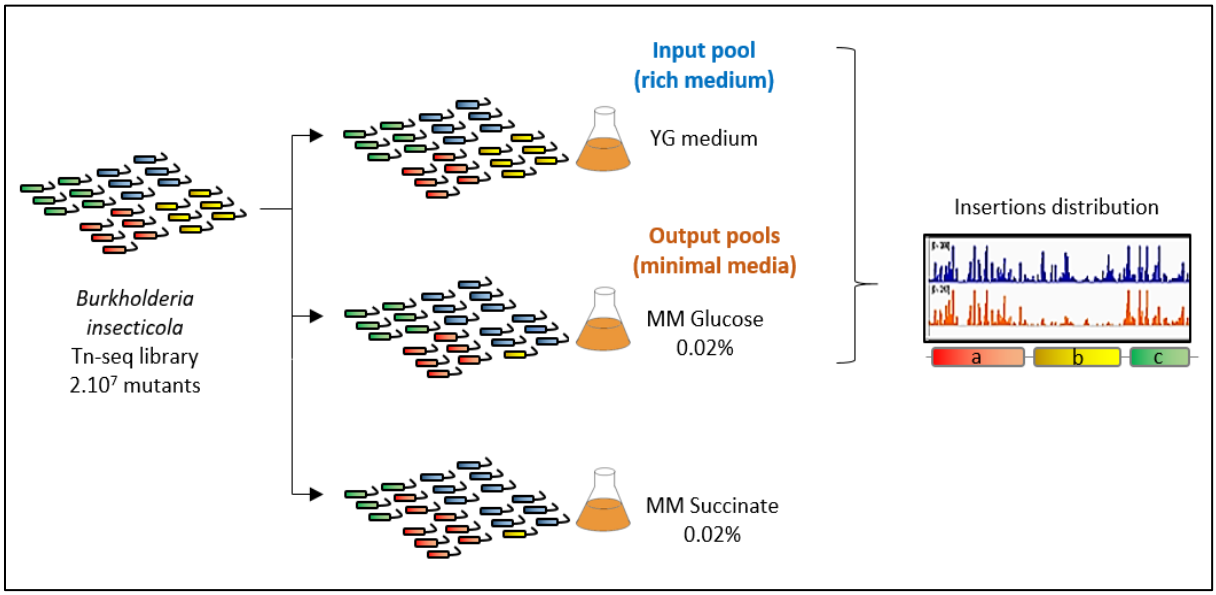


Figure 30: Schematic representation of the Tn-seq experiment conducted with different carbon sources.

The bacterial mutants colored in red, yellow and green correspond to mutations in the genes a, b and c, respectively. The sequencing results of the input pool (blue) and output pools (orange) are depicted in the right figure, which can be obtained through IGV.
MM: minimal medium.

different genes sets are required for growth in these two defined media. In addition, large differences in essential genes in both defined media were expected compared with growth in the YG rich medium, particularly in anabolic pathways. Hence, comparing the sequencing profiles of transposon borders in the transposon mutant population grown in YG (input population) and grown in MM (output population) will identify those fitness genes (**Figure 30**). After the sequencing, I obtained between one and four million filtered reads per replicate of the MM conditions, except for one succinate replicate that showed an insufficient number of reads and was discarded for the ARTIST analysis. The insertions distribution was highly correlated between replicates of a same condition ($r^2 > 0.95$ for glucose and $r^2 > 0.92$ for succinate) (**Figure 31**). Therefore, I pooled the sequencing data for the glucose triplicates and the succinate duplicates for the Con-ARTIST analysis. The pooled glucose and succinate conditions contained 9,365,715 and 3,398,153 reads respectively, with 92% of alignment against the symbiont genome.

By applying the Con-ARTIST analysis comparing the insertions distribution between the standard growth condition in YG rich medium and each of the two MM conditions, I found that 53 and 54 genes (essentiality Con-ARTIST score = 2) were required for the fitness in the

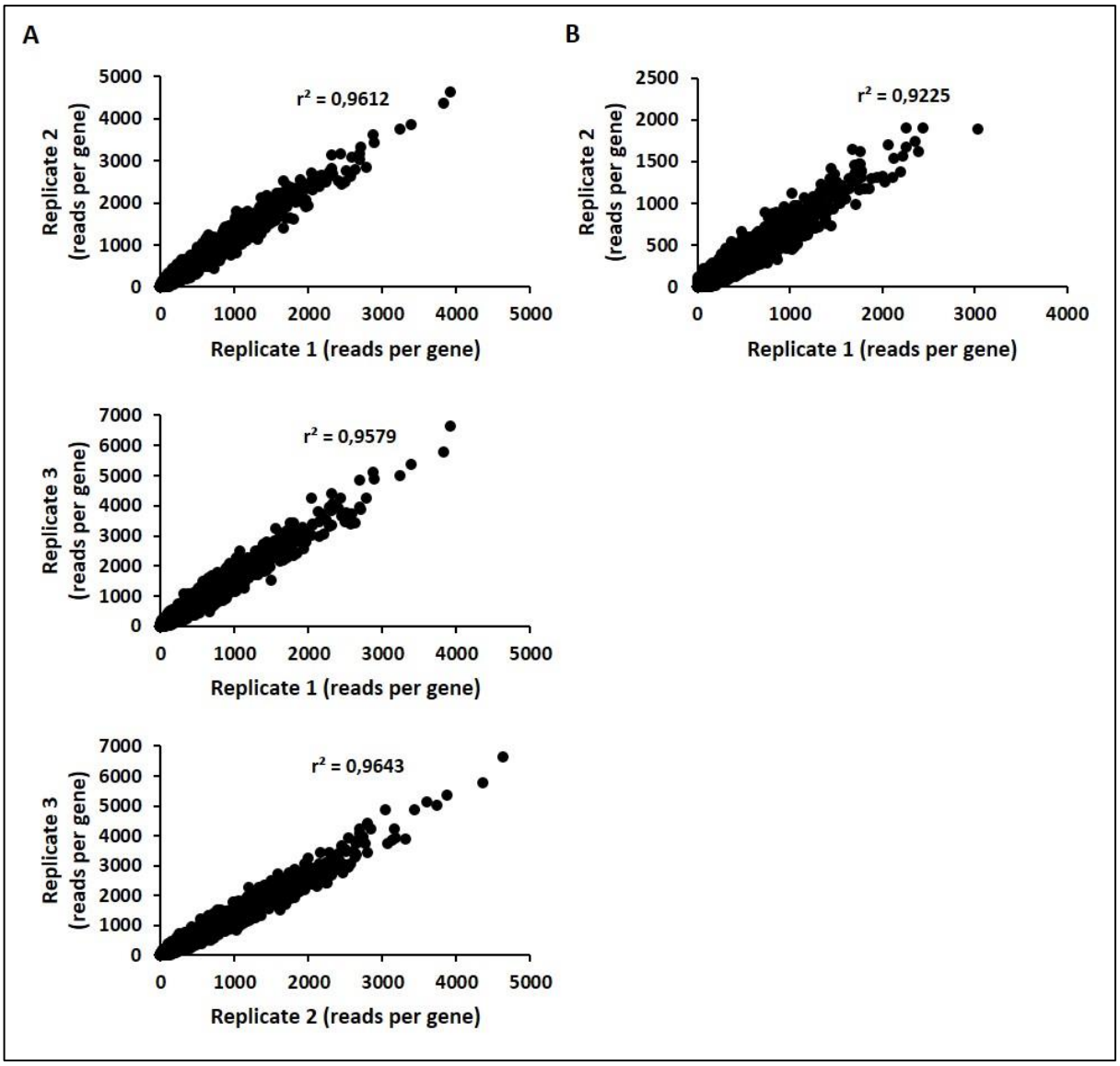


Figure 31: Correlations between read counts distribution in the replicates of minimal media conditions.

Dot plot representations of the comparison of each transposon insertions distribution between the Tn-seq replicates of minimal media conditions. The number of reads per gene is displayed for each replicate. The Pearson correlation coefficient r^2 was calculated for each comparison and indicated on each graph. A) Glucose condition (three replicates). B) Succinate condition (only two replicates).

presence of glucose and succinate, respectively (see **Annexes 2 and 3** for the lists of fitness genes in MM with glucose and succinate, respectively) (**Figure 32**). In addition, 18 and 23 genes were considered as "domain-essential genes" for glucose and succinate, respectively (essentiality Con-ARTIST score = 1).

Among the 34 fitness genes shared between the MM conditions, most of them belong to the amino acid biosynthesis pathways like the tryptophan biosynthesis gene cluster

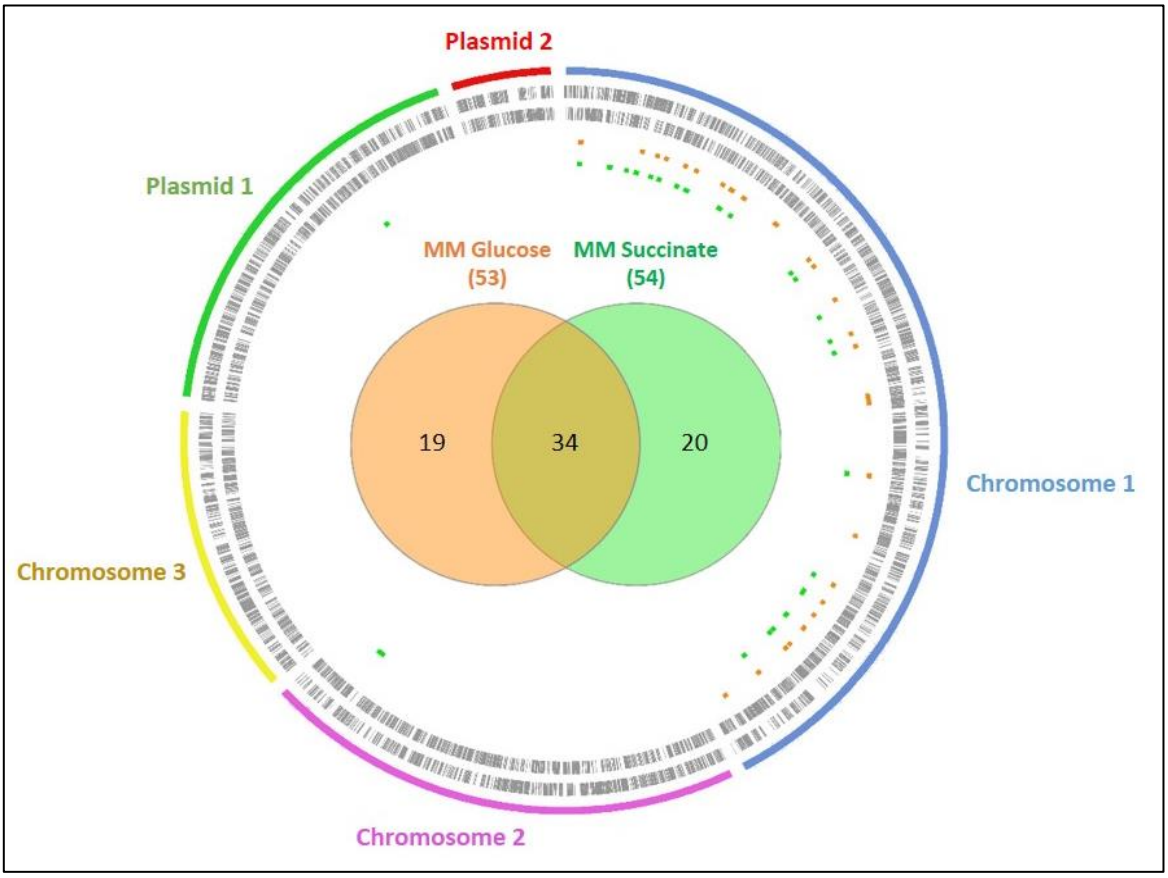
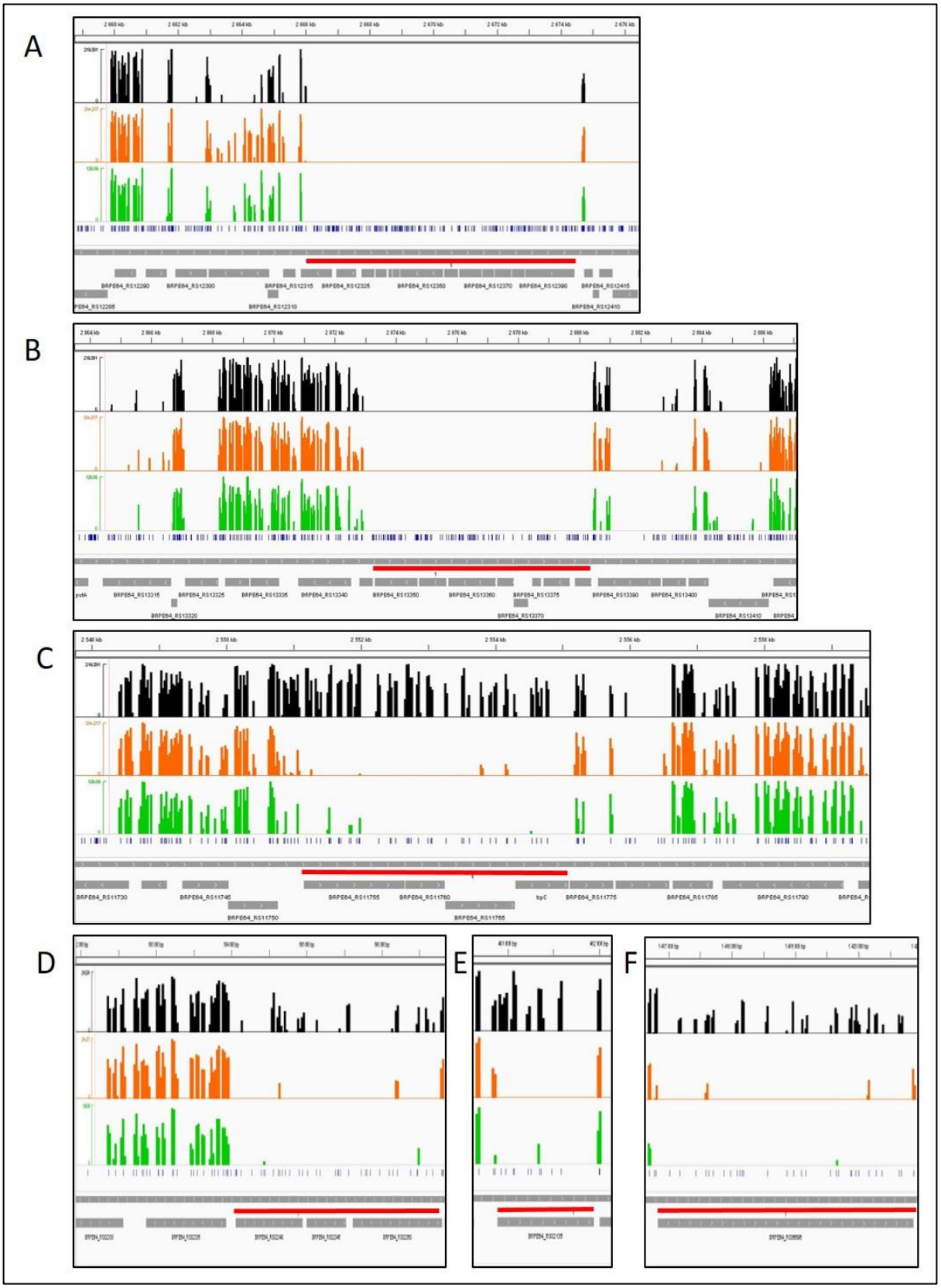


Figure 32: Fitness genes identified by Con-ARTIST analysis for glucose and succinate exploitation in *B. insecticola*.

Circular representation of the *B. insecticola* genome consisting of chromosome 1, 2 and 3 and plasmids 1 and 2. From outer to inner rings: forward CDS (grey bars), reverse CDS (grey bars), conditionally-essential genes (Con-ARTIST essentiality score = 2) for glucose (orange dots), conditionally-essential genes (Con-ARTIST essentiality score = 2) for succinate (green dots), distribution of conditionally-essential genes between the two minimal media (MM) conditions (Venn diagram).

(*BRPE64_RS11755/trpE, BRPE64_RS11760, BRPE64_RS11765/trpD, BRPE64_RS11770/trpC)* (**Figure 33C**), but also the cysteine, proline, lysine, isoleucine, arginine and glutamate biosynthesis genes. Additionally, fitness genes involved in nucleotide metabolism were identified, with the *de novo* purine biosynthesis pathway (*BRPE64_RS02135/purM*, *BRPE64_RS02340/purC, BRPE64_RS02345/purE, BRPE64_RS02350/purK* and *BRPE64_RS06595/purL*) (**Figure 33 D-F**). As the MM condition is deprived of any amino acid and nucleotide sources, these bacteria have to synthesize these components *de novo*, so the genes involved in their biosynthesis are required for the bacterial survival in this type of environment.

I next focused on the fitness genes that were associated specifically to glucose or succinate



exploitation, respectively (**Figures 34 and 35**). For the glucose condition, it is known that glucose can be metabolized through different catabolic pathways in microorganisms, all referred to as glycolysis. These glucose metabolisms in bacteria include the Embden-

Figure 33: Identification of *B. insecticola* essential genes and fitness genes shared in the two minimal media conditions.

The insertion distributions (log^{10} scale) are displayed for the conditions YG (black bars), MM with glucose (orange bars) and MM with succinate (green bars). The different positions on the chromosome 1 is indicated above each figure. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Red lines are highlighting essential genes in panels A and B, and fitness genes for MM conditions in panels C, D, E and F. A) Essential genes encoding 30S and 50S ribosomal proteins from *BRPE64_RS12355* to *BRPE64_RS12450.* B) Essential genes encoding ATP synthase subunits from *BRPE64_RS13345* to *BRPE64_RS13385.* C) Fitness genes involved in tryptophan biosynthesis from *BRPE64_RS11755* to *BRPE64_RS11770.* D, E and F) Fitness genes involved in *de novo* purine biosynthesis with *purC*, *purE* and *purK* (D), *purM* (E) and *purL* (F).

Meyerhof-Parnas (EMP) pathway which is the most common type of glycolysis, the pentose-phosphate (PP) pathway and the Entner-Doudoroff (ED) pathway (**Figure 35**). Most steps of the EMP pathway are reversible and shared with gluconeogenesis except the conversion of fructose-6-phosphate to fructose 1,6-bisphosphate which is catalysed by 6-phosphofructokinase in the EMP pathway and by fructose bisphosphatase in gluconeogenesis. Because the 6-phosphofructokinase gene is absent in the *B. insecticola* genome, I can conclude that the EMP pathway is not used for glucose utilisation in this bacterium. However, the gluconeogenesis-specific fructose bisphosphatase gene is present and is essential in MM with succinate, demonstrating that gluconeogenesis is essential for growth on this carbon source (**Figure 35**). The other steps of the EMP and gluconeogenesis pathways are either essential in all conditions or essential for growth in MM with both carbon sources, confirming a key role of the anabolic gluconeogenesis for growth on succinate and suggesting an essential catabolic function downstream of the PP and ED pathways for growth on glucose and/or in YG rich medium (**Figures 34 and 35**). In the MM condition with glucose, I found genes involved in glucose uptake through an ABC sugar-transporting system (*BRPE64_RS03960-BRPE64_RS03975*) and the genes of the ED pathway (*BRPE64_RS03980*, *BRPE64_RS03985 and BRPE64_RS11130*/*edd*) (**Figures 34 and 35**). Thus, the ED pathway seems to be the principal pathway mobilized for glucose catabolism. I furthermore noticed that the PP pathway was required for growth in all conditions. The PP pathway can have a catabolic function parallel to the ED and EMP pathways but also an important anabolic role, in the generation of the pentose ribose-5-phosphate and erythrose-4-phosphate for nucleotide and aromatic amino acid biosynthesis, respectively.

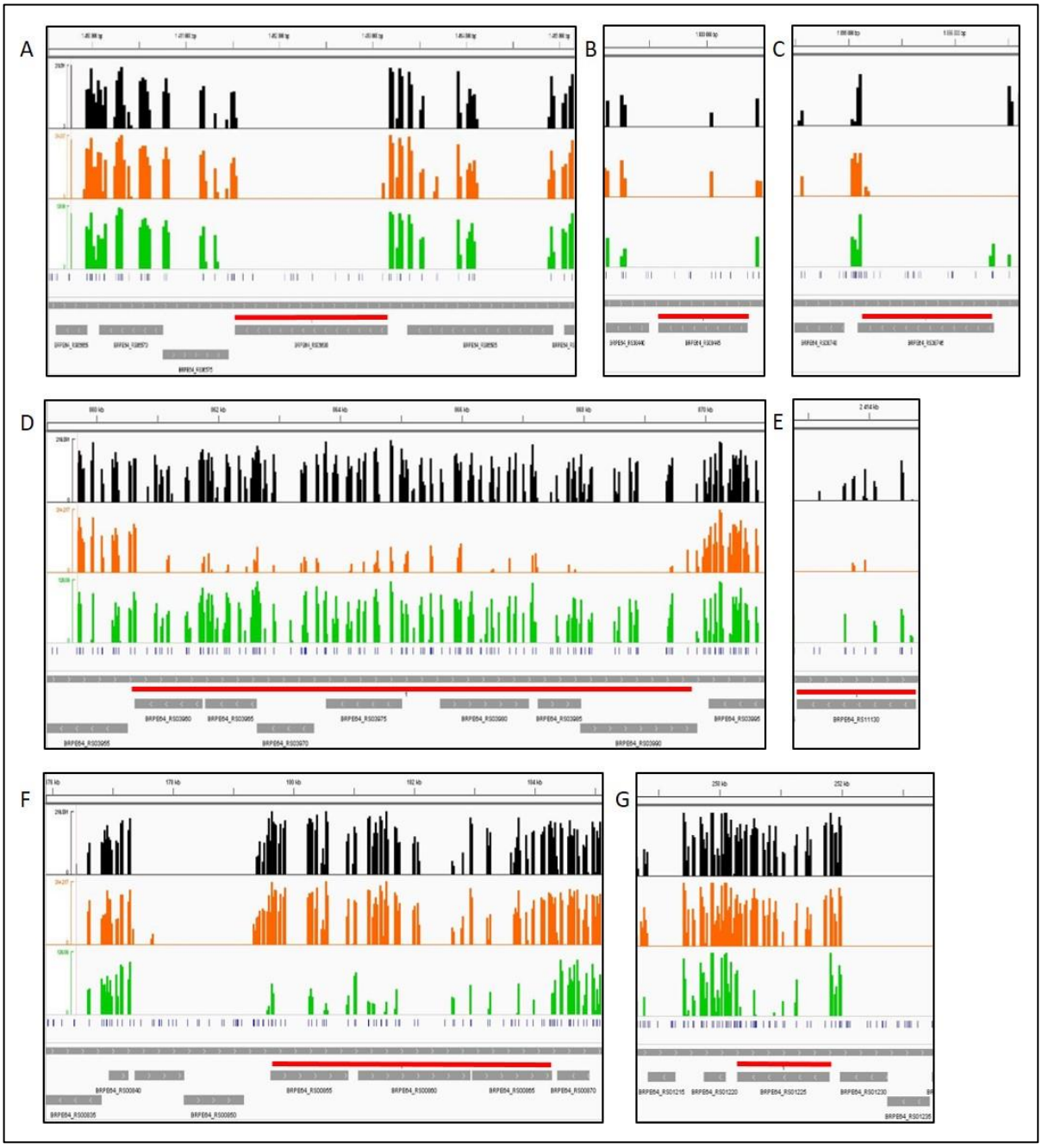
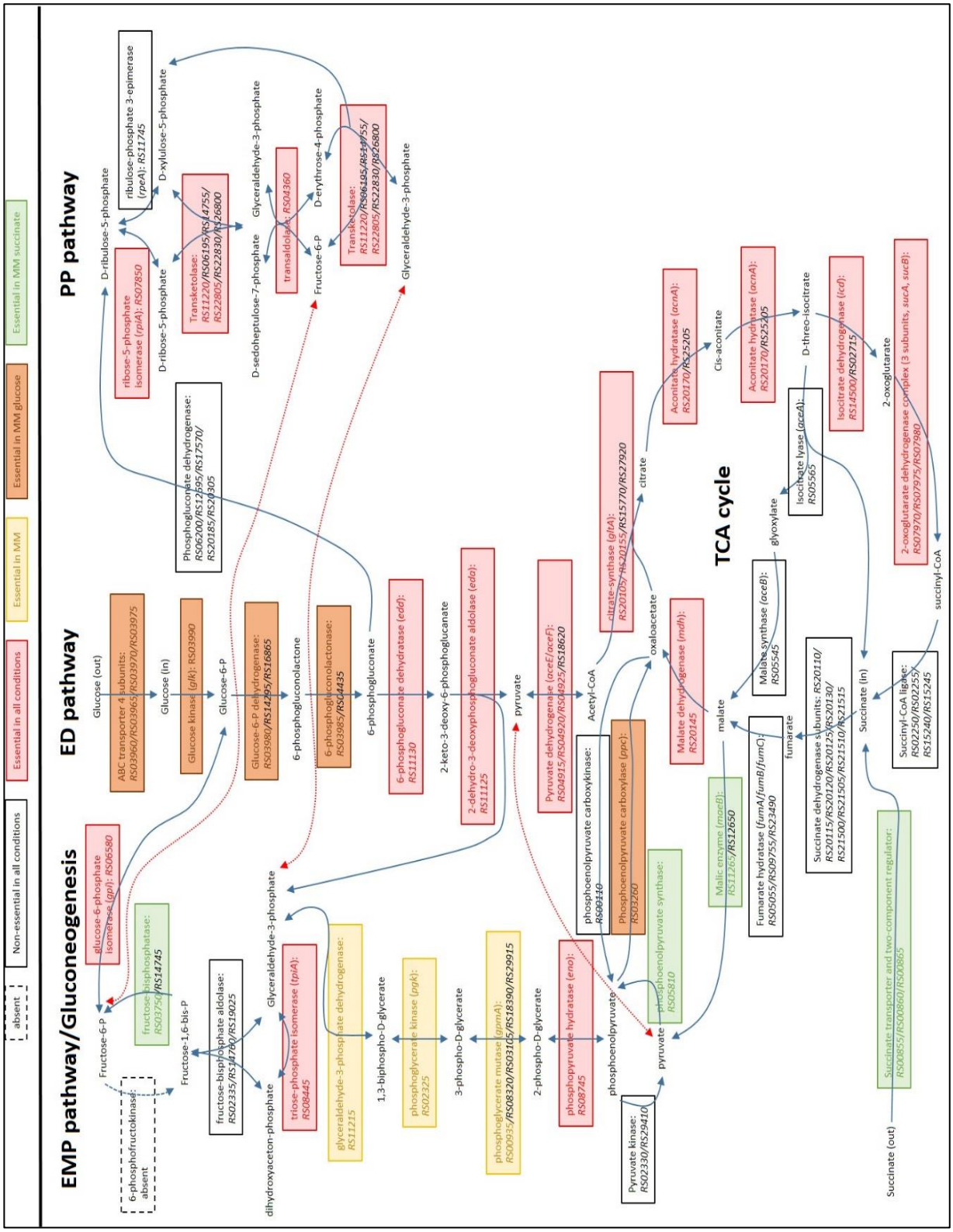


Figure 34: Identification of *B. insecticola* fitness genes required for glucose and succinate exploitation.

The insertion distributions (log^{10} scale) are displayed for the conditions YG (black bars), MM with glucose (orange bars) and MM with succinate (green bars). The different positions on the chromosome 1 is indicated above each figure. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Red lines are highlighting essential genes or fitness genes. A, B and C) Essential genes involved in the EMP/gluconeogenesis pathways with *pgi* (A), *tpiA* (B) and *eno* (C) genes. D and E) Fitness genes required for glucose catabolism with a sugar ABC transporter (D) and genes involved in the ED pathway (D and E). F and G) Fitness genes involved in succinate exploitation with genes encoding a C4-dicarboxylate transporter, a two-component system (F) and the RNA-polymerase sigma-54 factor (G).



In the growth on succinate, a noticeable group of fitness genes was *BRPE64_RS01225*, encoding an RNA polymerase sigma-54 factor, and *BRPE64_RS00860-BRPE64_RS00865*, encoding a two-component system which is dependent on the sigma-54 transcription factor

Figure 35: Identification of fitness genes involved in the different glycolysis pathways. The different steps are summarized for the three glycolysis pathways with the EMP/gluconeogenesis, ED and PP pathways, with also the TCA cycle. For each enzymatic step, the corresponding gene names and locus tags are indicated (the " BRPE64_" was removed for each locus tag to fit the figure). The legend is indicated on the figure with non-essential genes (black), essential genes in the rich medium and all conditions (red), specific genes for the MM conditions (yellow), specific genes for the glucose condition (orange) and for the succinate condition (green). The red dotted arrows indicate the locations of similar molecules in the different pathways. The blue arrows indicate the direction of the enzymatic reactions.

family (**Figure 34**). This class of transcription factors was reported to be activated in the presence of signalling molecules, such as succinic semi-aldehyde or gamma-aminobutyric acid, which suggests that succinate might act as a signalling molecule (Peng *et al.*, 2014, 2015; Söhling and Gottschalk, 1996). In the same locus as the two-component system genes, I found a fitness gene encoding a specific transporter called C4-dicarboxylate transporter, probably involved in the succinate transport (*BRPE64_RS00855*) (Valentini *et al.*, 2011). Another fitness gene for growth in succinate, *BRPE64_RS11265*, is encoding a malate dehydrogenase (malic enzyme, *maeB*) which branches from the tricarboxylic acid (TCA) cycle and generates the acetyl-CoA that is essential for feeding the cycle, from the excess of carbon skeleton resulting from the direct succinate input in the TCA cycle (**Figure 35**). The other enzymes involved in the TCA cycle are encoded by essential or duplicated genes. An additional gene required for growth on succinate encodes the phosphoenolpyruvate synthase that can shuttle carbon from succinate into the gluconeogenesis pathway (**Figure 35**).

4.4. Comparative transposon mutagenesis on *Burkholderia* species

In previous studies, transposon mutagenesis techniques were used to identify essential genes in rich medium for two pathogenic *Burkholderia* species, *B. pseudomallei* strain K96243 (Moule *et al.*, 2014) and *B. cenocepacia* strain J2315 (Wong *et al.*, 2016), and additionally for one environmental species, *B. thailandensis* strain E264 (Baugh *et al.*, 2013). These studies revealed that 505, 470 and 406 essential genes were found for *B. pseudomallei* K96243, *B. cenocepacia* J2315 and *B. thailandensis* E264, respectively.

By comparing these essential gene sets with the *B. insecticola* essential genes on YG rich

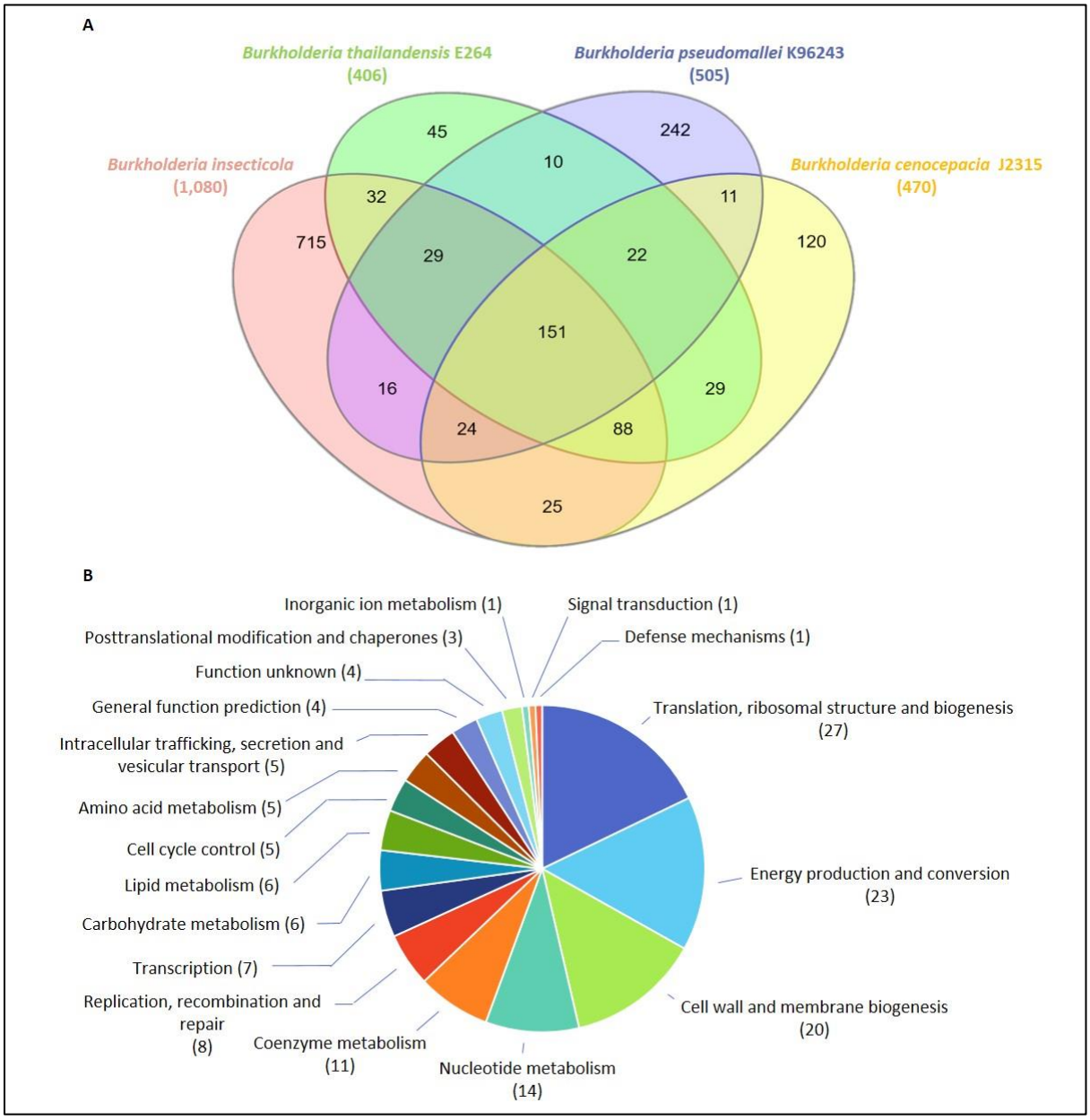


Figure 36: Identification of essential functions in rich medium for *Burkholderia* species by comparative Tn-seq analysis.

A) Comparison of essential gene sets in rich medium identified in *B. insecticola* (1,080 genes, YG medium pink), *B. thailandensis* E264 (406 genes, LB medium, green), *B. pseudomallei* K96243 (505 genes, LB medium, blue) and *B. cenocepacia* J2315 (470 genes, LB medium, yellow). 151 genes were shared between these four *Burkholderia* species. B) COG classes distribution of the 151 genes shared between the four *Burkholderia* species. The number of genes is indicated for each cellular function.

medium, only 151 genes were shared between these four species (see **Annexe 4** for the list of shared essential genes, **Figure 36A**). An approximatively similar number (164) was previously reported when the three *Burkholderia* species datasets were compared together (Wong *et al.*, 2016). The majority of these 151 fitness genes belongs to the translation process (17.9%), the

energy production (15.2%) and the cell wall biosynthesis COG categories (13.2%) (**Figure 36 B**). Examples of conserved essential genes are those encoding the ribosomal subunits, initiation factors of translation, the ATP synthase subunits and enzymes involved in peptidoglycan biosynthesis. One striking example is the 4-amino-4-deoxy-L-arabinose (Ara4N) lipid A modification gene cluster (*BRPE64_RS06345-BRPE64_RS06365* in *B. insecticola*), which was experimentally demonstrated to be essential for *B. cenocepacia* J2315 viability (Ortega *et al.*, 2007).

Nonetheless, I noticed a strong difference in the essential genome proportion between *B. insecticola* (17%) and *B. pseudomallei* (8.5%), *B. cenocepacia* (6.1%) and *B. thailandensis* (7.1%). Hence, I observed that 715 fitness genes are specific to the *B. insecticola* symbiont (see **Annexe 5** for the list of *B. insecticola*-specific essential genes). Among these genes, I found previous essential functions like the amino acid metabolism (9%), the energy production (8%), the cell wall formation (6%), but also another COG category which is the transcription machinery (10%). However, most of these specific genes are encoding hypothetical proteins with unknown functions (38%).

5. Discussion

Deciphering the essential genome of bacteria is a key step towards the understanding of the main biological functions that are strictly required to sustain life (Gil *et al.*, 2004; Qiu, 2012). For that purpose, numerous studies were performed on different pathogenic and environmental bacteria using transposon mutagenesis approaches (Armbruster *et al.*, 2017; Bishop *et al.*, 2014; Hooven *et al.*, 2016; Klein *et al.*, 2012; Pechter *et al.*, 2016; Weerdenburg *et al.*, 2015). In the case of the *R. pedestris-Burkholderia* interaction, the creation of a Tn-seq-compatible transposon mutant library of *B. insecticola*, and the genome-wide identification of essential genes constitutes a first step towards the characterization of the fitness landscape of this bacterium during its different lifestyles, which include colonization of soil, plants and of the symbiotic organ of *R. pedestris*. The *mariner* Tn-seq library of *B. insecticola* was highly concentrated with 2.10^7 independent clones counted, compared to other Tn-seq libraries previously generated (van Opijnen *et al.*, 2009; Wiles *et al.*, 2013). This might be due to differences in transformation efficiency between these different bacterial species. Additionally, the majority of the genes showed approximately a 90% coverage of their total

TA sites disrupted by the transposon, which is higher than the 75% coverage previously reported (Chao *et al.*, 2013), thus demonstrating that this library is saturated in transposon insertions.

By using this transposon mutant library, I identified 1,080 essential genes for the bacterial fitness during growth on a rich medium, representing 17% of the *B. insecticola* genome. This essential genome proportion is quite closed to those found for *M. tuberculosis* (DeJesus and Ioerger, 2013; Griffin *et al.*, 2011), *Caulobacter crescentus* (Christen *et al.*, 2014) and *E. coli* (Gerdes *et al.*, 2003). The majority of these essential genes are encoding for ribosomal subunits, transcription factors, RNA polymerase subunits, DNA replication components, the electron transport chain, ATP synthase subunits, peptidoglycan and LPS biosynthesis proteins and different ABC transporters linked to amino acid transporting systems (**Figure 37**). All of these pathways highlight cellular functions like translation, transcription, energy production and cell wall biosynthesis, known to represent vital processes for the bacterial survival (Glass *et al.*, 2006; Grazziotin *et al.*, 2015). It was striking to notice that approximately half of the genes of the plasmid 1 were considered as essential, which could suggest that the plasmid 1 is lost in a fraction of the bacterial population during *in vitro* growth cultures, as it was previously reported for the plasmid pC3 in *B. cenocepacia* H111 (Agnoli *et al.*, 2014). However, the loss of the plasmid 1 was not observed in axenic culture of *B. insecticola*, so the number of essential genes identified by Tn-seq may truly reflect the essentiality of these genes.

Nonetheless, a consistent part of essential genes (28.3%) represents hypothetical proteins with unknown functions. Our attempts to identify these proteins by BLASTp analysis revealed that they are closely related to hypothetical proteins of *Burkholderia sp.* YI23 (Lim *et al.*, 2012). This observation is quite relevant due to the short phylogenetic distance between *Burkholderia sp.* YI23 and *B. insecticola* (Sawana *et al.*, 2014). Furthermore, these two *Burkholderia* species are soil microorganisms, hence potentially sharing the same ecological niches in the environment (Kim *et al.*, 2009). Among these hypothetical proteins identified, there might be some functions involved in soil habitat adaptation which became essential for the bacterial survival to cope with different environmental selective pressures.

To assess the robustness of this transposon mutagenesis approach, I tested the procedure in two slightly different growth conditions – growth in MM with glucose or succinate as carbon source – that have predictable requirements on genes needed for growth in these conditions.

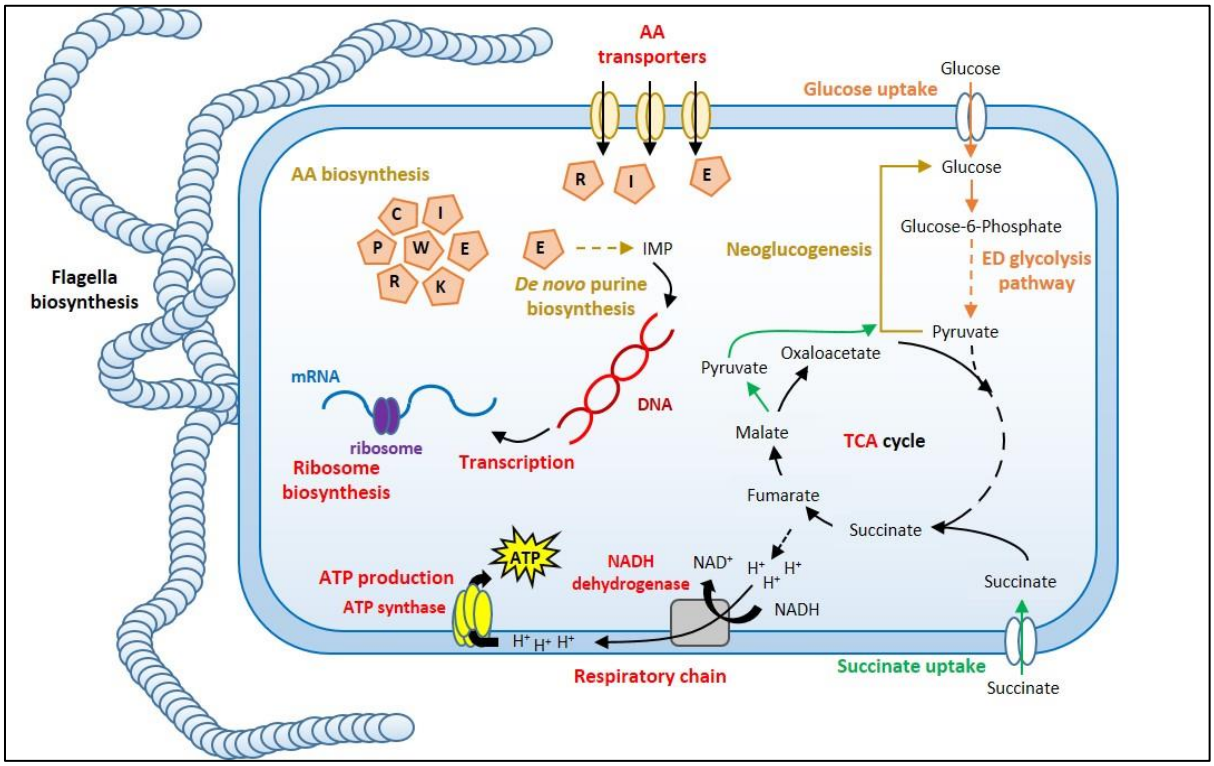


Figure 37: Overview of the main essential cellular pathways and specific pathways for glucose or succinate exploitation identified in *B. insecticola.*

Biological functions required for YG rich medium are highlighted in red, required for MM conditions are highlighted in yellow, and required for glucose and succinate conditions are highlighted in orange and green, respectively. The biological functions highlighted in black represent non-essential functions. The enzymatic steps required for MM conditions are indicated by yellow arrows, required for glucose condition are indicated by orange arrows, and required for succinate condition are indicated by green arrows. Abbreviations: AA: amino acid, C: cysteine, E: glutamic acid, I: isoleucine, IMP: inositol mono-phosphate, L: lysine P: proline, R: arginine, W: tryptophan.

I have found 53 and 54 essential genes required for glucose and succinate exploitation in a minimal medium, respectively. As expected, for growth in a MM containing besides the carbon sources no other organic compounds, 34 genes involved in amino acid and nucleotide metabolisms were commonly found to be essential between these two conditions (**Figure 37**). In the presence of glucose, I identified different steps of the ED glycolysis pathway, as well as a glucose transporter, which is encoded in the same gene cluster, as essential for the bacterial fitness (**Figure 37**). The EMP glycolysis is one of the main cellular pathways to degrade glucose, but because the first specific step of the pathway, the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate, is missing, the EMP pathway can not be used to catabolize glucose in *B. insecticola*. The ED pathway is an alternative route for glucose degradation known to be used by different bacterial species, including *Pseudomonas* and

Zymomonas species (Godbey, 2014). When succinate was used as a carbon source, I could identify fitness genes that are involved in its transport, encoding a C4-dicarboxylate transporter (Valentini *et al.*, 2011) as well as a sigma-54 transcription factor that is potentially involved in the regulation of this transporter (Peng *et al.*, 2014; Söhling and Gottschalk, 1996). The two additional key genes specifically required for growth on succinate allow the supply of acetyl-CoA into the TCA cycle and to feed the gluconeogenesis anabolic pathway for the synthesis of the carbohydrate metabolites of the cell (**Figure 37**). Taken together, the identification of amino acid and nucleotide biosynthesis genes, as well as carbon source-specific genes showed the robustness of transposon mutagenesis approaches to detect critical cellular functions required under specific conditions.

In previous works, this robust transposon sequencing technique facilitated the discovery of 505 essential genes for *B. pseudomallei* (Moule *et al.*, 2014), which is the causative agent of melioidosis and a potential bioterrorism agent, 470 essential genes for *B. cenocepacia* (Wong *et al.*, 2016), mostly associated to opportunistic infections in cystic fibrosis patients, and 406 essential genes for *B. thailandensis* (Baugh *et al.*, 2013), an environmental microorganism closely related to *B. pseudomallei*. By comparing these essential gene sets together with the essential genes of *B. insecticola*, I found 151 essential genes shared between these four *Burkholderia* species on rich medium conditions. Such comparison was already performed in a previous study where a total of 164 genes were predicted to be essential in the other three *Burkholderia* species (Wong *et al.*, 2016). This common set of essential genes is mostly encoding proteins involved in cellular functions like translation, energy production and cell wall biosynthesis (**Figure 37**). Notable among these genes, there is the *arn* gene cluster which encodes the enzymes for the production of the Ara4N, a lipopolysaccharide modification. This gene cluster was in an independent approach found to be essential for *B. cenocepacia* viability (Ortega *et al.*, 2007). This Ara4N modification of lipopolysaccharide is a mechanism known to mediate resistance against cationic AMPs in many Gram-negative bacteria (Ernst *et al.*, 2007; Shafer *et al.*, 1984), including *Burkholderia* species (Loutet and Valvano, 2011). This kind of genes may be an attractive target to develop new inhibitory drugs against *Burkholderia* pathogens. Hence, by using these transposon mutagenesis approaches, this essential gene repertoire found may outline the core essential genome of *Burkholderia* species.

Nonetheless, I could notice that there is a significant gap between the essential genes

proportion between *B. insecticola* (17%) and the other three *Burkholderia* species (6.1 to 8.5%). That difference might be explained by the different transposon mutagenesis techniques employed, Tn-seq (Baugh *et al.*, 2013; this work) and TraDIS (Moule *et al.*, 2014; Wong *et al.*, 2016), which are moreover linked to specific bioinformatic tools to determine the essential genes proportion for each species. The El-ARTIST analysis that I used for the *B. insecticola* data is based on a Hidden-Markov Model (Pritchard *et al.*, 2014), in which essentiality implies both (nearly) absolute requirement for growth as well as a reduced fitness (translated by less sequencing reads as expected in the gene). Thus, the higher number of genes labelled as essential in *B. insecticola* might be due to a difference in the stringency of the criteria determining essentiality although the associated cellular functions of these additional genes showed a clear biological relevance (see below). Additionally, the mutant libraries for these *Burkholderia* species used distinct transposon types with *mariner* (*B. insecticola*), Tn*5* (*B. pseudomallei* K96243 and *B. cenocepacia* J2315) and T23 (*B. thailandensis* E264) transposons, which target distinct genomic sites, hence producing mutant populations with different mutation coverages. Due to these differences, we found 715 essential genes which were strictly specific to *B. insecticola*. These *B. insecticola*-specific essential genes belonged to the same vital functional categories like transcription, translation and energy production as the conserved essential genes, and they completed entirely specific pathways as essential, suggesting the biological relevance of the *B. insecticola* essential gene set. Additionally, the majority of the *B. insecticola*-specific essential genes are encoding hypothetical proteins with unknown functions, closely related to *Burkholderia sp.* YI23 hypothetical proteins, and could be involved in the ecological niche adaptation of soil-related *Burkholderia* species. Hence, it would be of interest to study the essential genome of other environmental *Burkholderia* species closely related to *B. insecticola,* which could provide more information on their mechanisms to adapt themselves to their environment.

Chapter III

Identification of Factors Involved in Antimicrobial Peptide Resistance in the *Burkholderia insecticola* Symbiont

1. Introduction

Antibiotic resistance is a common feature in *Burkholderia* species, and more specifically towards antimicrobial peptides (AMPs) or polymyxins (Loutet and Valvano, 2011; Rhodes and Schweizer, 2016). Different resistance mechanisms against polymyxins were previously characterized in the *Burkholderia* genus, especially in the BCC&P clade and it is believed that this characteristic is one of the important adaptations of these bacteria that render them successful infectious agents (Loutet and Valvano, 2011; Rhodes and Schweizer, 2016; Scoffone *et al*., 2017).

One of the well-known mechanisms is the structure of the lipopolysaccharides (LPS) (Loutet and Valvano, 2011; Ortega *et al.*, 2009, 2007). As mentioned before, LPS molecules are made of three parts: the lipid A, the core oligosaccharide and the O-antigen (see **Chapter I**). In the *Burkholderia* genus, 4-amino-4-deoxy-L-arabinose (Ara4N) modifications of the lipid A are able to reduce the net negative charge of the LPS molecule, thus reducing the interactions between cationic polymyxins and anionic bacterial membranes (Loutet and Valvano, 2011; Ortega *et al.*, 2007). It was recently shown that the gene cluster involved in the biosynthesis of Ara4N moiety is essential for the bacterial viability in *B. cenocepacia* (Ortega *et al.*, 2007) and other *Burkholderia* species, including *B. insecticola* (see **Chapter II**). This Ara4N modification seems to be constitutively present on LPS in species belonging to the BCC&P clade, which highlights its conserved requirement (Ortega *et al.*, 2007) but in other bacteria like *Salmonella*, this LPS modification is not essential and is introduced only upon sensing of AMPs by the PhoPQ two-component system (Dalebroux and Miller, 2014). Additionally, the core oligosaccharide also contributes to the negative charges of the bacterial membrane, and can be subdivided into two parts, an inner core and an outer core (Ortega *et al.*, 2009). Hence, it was demonstrated that core oligosaccharide deficient *B. cenocepacia* mutants, which led to various truncated forms of the core oligosaccharide, are more sensitive towards polymyxin B (Ortega *et al.*, 2009). Thus, the composition of the LPS molecule is important for polymyxin resistance in *Burkholderia* species (Loutet and Valvano, 2011; Rhodes and Schweizer, 2016).

Another polymyxin resistance factor was described in *B. cenocepacia* which is the alternative sigma factor σ^E or RpoE, the key regulator of the extracytoplasmic stress response or ESR pathway (see **Chapter I**) (Flannagan and Valvano, 2008). More specifically, the RpoE factor

plays a role in envelope integrity, and regulating stress response mechanisms during membrane damages and environmental changes, such as elevated temperature and osmotic shock (Flores-Kim and Darwin, 2014; Guest and Raivio, 2016). This RpoE factor was mostly studied in *E. coli* and its precise function and impact on gene regulation is poorly understood in the *Burkholderia* genus (Guest and Raivio, 2016; Loutet and Valvano, 2011; Loutet *et al.*, 2011).

Finally, a specific category of membrane components, known as hopanoids, was shown to participate in polymyxin B resistance in *B. multivorans* (Malott *et al.*, 2012) and in *B. cenocepacia* (see **Chapter I**) (Schmerk *et al.*, 2011, 2015). Hopanoids are triterpenoids or sterol-like molecules, which are the bacterial analogues of cholesterol molecules in eukaryotic membranes (Kannenberg and Poralla, 1999; Sahm *et al.*, 1993). Although they are not widespread among bacterial species, their biosynthesis gene cluster seems to be conserved in the *Burkholderia* genus, and they are thought to be involved in membrane fluidity and permeability (Malott *et al.*, 2012; Ourisson and Albrecht, 1992; Pearson *et al.*, 2007; Schmerk *et al.*, 2011, 2015).

Based on these studies in other *Burkholderia* species, which identified the core oligosaccharide of LPS, the sigma factor σ^E and hopanoids as polymyxin resistance factors in some *Burkholderia* species (Flannagan and Valvano, 2008; Loutet and Valvano, 2011; Malott *et al.*, 2012; Ortega *et al.*, 2009; Schmerk *et al.*, 2011, 2015), I created mutants and investigated the role of these different membrane components in AMP resistance in *B. insecticola*. In addition to this candidate gene approach, I have performed a Tn-seq experiment with different AMPs, including *R. pedestris* AMPs, on *B. insecticola* and I have identified its fitness genes involved in AMPs resistance. Some genes of this fitness gene list were mutagenized in order to validate the Tn-seq analysis. Each of these *Burkholderia* deficient mutants was assessed for its AMPs sensitivity and was also tested for its colonization efficiency of *R. pedestris*. Thus, this work highlighted new bacterial factors involved in AMPs resistance in *B. insecticola* and suggested a connection between AMPs resistance and host colonization capability.

2. Contributions

Quality control and sequencing of the Tn-seq samples were performed by the I2BC sequencing platform (CNRS Gif-sur-Yvette, France). Flow cytometry experiments were conducted by the the I2BC ImaGif platform (CNRS Gif-sur-Yvette, France). The lipid analysis (hopanoid determination) was performed by Quentin Nicoud during his Master 1 internship, with the assistance of Frédéric Gressent (IRD, Montpellier) and Philippe Schaeffer (CNRS, Université de Strasbourg). Additional results presented in the discussion part were obtained by a Master 2 student, Christy Calif, that I participated in her supervision.

3. Materials and methods

Bacterial strains and growth conditions

The *B. insecticola* strain RPE64 was routinely cultured in YG medium (5 g.L^{-1} yeast extract, 1 g.L^{-1} NaCl, 4 g.L^{-1} glucose) at 28°C. The modified strain *B. insecticola* RPE75 carrying a resistance to rifampicin (Rif) was used for transposon mutagenesis and cultured in YG medium supplemented with 30 µg.mL^{-1} of Rif at 28°C. Each *B. insecticola* mutant was cultured in YG

Strains	Characteristics	Reference or source
B. insecticola RPE75	Wild-type strain with Rif^R	Kikuchi et al., 2011
B. insecticola RPE75 GFP	Wild-type strain with Rif^R, containing a Tn7 insertion with a GFP cassette a Km^R cassette	Kikuchi and Fukatsu, 2014
B. insecticola RPE75 RFP	Wild-type strain with Rif^R, containing a Tn7 insertion with a RFP cassette a Km^R cassette	Kindly provided by Yoshitomo Kikuchi, AIST Hokkaido (Japan)
B. insecticola Δ*waaC*	Deletion of the *waaC* gene (*BRPE64_RS10300*) in a *B. insecticola* RPE75 GFP background	Kim et al., 2017
B. insecticola Δ*waaF*	Deletion of the *waaF* gene (*BRPE64_RS02300*) in a *B. insecticola* RPE75 GFP background	Kim et al., 2017
B. insecticola Δ*wbiF*	Deletion of the *wbiF* gene (*BRPE64_RS10490*) in a *B. insecticola* RPE75 GFP background	Kim et al., 2016
B. insecticola shc::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *shc* gene (*BRPE64_RS14420*) in a *B. insecticola* RPE75 background	This study
B. insecticola hpnA::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *hpnA* gene (*BRPE64_RS14505*) in a *B. insecticola* RPE75 background	This study
B. insecticola hpnH::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *hpnH* gene (*BRPE64_RS14480*) in a *B. insecticola* RPE75 background	This study
B. insecticola hpnJ::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *hpnJ* gene (*BRPE64_RS14180*) in a *B. insecticola* RPE75 background	This study
B. insecticola hpnN::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *hpnN* gene (*BRPE64_RS14440*) in a *B. insecticola* RPE75 background	This study
B. insecticola rpoE::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *rpoE* gene (*BRPE64_RS09525*) in a *B. insecticola* RPE75 background	This study
B. insecticola mucD::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *mucD* gene (*BRPE64_RS09510*) in a *B. insecticola* RPE75 background	This study
B. insecticola wzm::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *wzm* gene (*BRPE64_RS10560*) in a *B. insecticola* RPE75 background	This study
B. insecticola rfbA::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *rfbA* gene (*BRPE64_RS10590*) in a *B. insecticola* RPE75 background	This study
B. insecticola rfbC::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *rfbC* gene (*BRPE64_RS10585*) in a *B. insecticola* RPE75 background	This study
B. insecticola tolB::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *tolB* gene (*BRPE64_RS11040*) in a *B. insecticola* RPE75 background	This study
B. insecticola tolQ::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *tolQ* gene (*BRPE64_RS11025*) in a *B. insecticola* RPE75 background	This study
B. insecticola tatB::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *tatB* gene (*BRPE64_RS12015*) in a *B. insecticola* RPE75 background	This study
B. insecticola dsbA::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *dsbA* gene (*BRPE64_RS00670*) in a *B. insecticola* RPE75 background	This study
B. insecticola mlaD::pVO155	Insertion of the plasmid pVO155-nptII-GFP (with Km^R) in the *mlaD* gene (*BRPE64_RS12120*) in a *B. insecticola* RPE75 background	This study

Table 2: Bacterial strains used in this study.

medium supplemented with 30 µg.mL^{-1} of Rif and 50 µg.mL^{-1} of kanamycin (Km) (see **Table 2**). For cultures on solid medium, the media were supplemented with 1.5% agar. All strains were stored at -80°C into 20% glycerol for long-term conservation.

Insect rearing

Adult *R. pedestris* insects were reared in plastic boxes with soybean seeds and sterile water supplemented with 0.05% ascorbic acid in a 25°C room with a light and dark cycle of 16 hours and 8 hours, respectively. Newly hatched insects were collected every day and reared in Petri dishes with the same conditions as adult insects.

Bacterial mutant construction

B. insecticola RPE75 was mutagenized by insertion mutagenesis. For that purpose, internal fragments (300-600 bp) of the target gene were amplified by PCR and cloned into the pVO155-nptII-GFP vector with a couple of specific restriction enzymes (SalI-XbaI or XhoI-XbaI). *E. coli* DH5α was transformed with the resulting construct by heat shock, and transformed bacteria were spread onto LB agar plates supplemented with 50 µg.mL^{-1} of Km. Colonies bearing the correct construct were confirmed by colony-PCR and Sanger sequencing (Eurofins Genomics). The plasmid construct was transferred to the recipient *B. insecticola* RPE75 strain by triparental conjugation, with the *E. coli* DH5α donor strain and the *E. coli* HB pRK600 helper strain. Each conjugation was first incubated on YG agar plates for 24 hours and was subsequently transferred on YG agar plates supplemented with 30 µg.mL^{-1} of Rif and 50 µg.mL^{-1} of Km. Candidate *B. insecticola* mutants were verified by colony-PCR and by checking the GFP fluorescence. All mutant strains were stored at -80°C in 20% glycerol.

Tn-seq library screening for *in vitro* growth with AMPs

One aliquot of the Tn-seq library was diluted to obtain an initial OD_{600nm} of 0.01 into a final volume of 1.2 mL of minimal medium (1 g.L^{-1} KH_2PO_4; 2 g.L^{-1} K_2HPO_4; 1 g.L^{-1} $(NH_4)_2SO_4$; 0.2 g.L^{-1} NaCl; 0.1 g.L^{-1} $MgSO_4$, $7H_2O$; 2.46 mg.L^{-1} $FeSO_4$, $7H_2O$; 3.31 mg.L^{-1} EDTA, 2Na; 50 mg.L^{-1} $CaCl_2$, $2H_2O$) supplemented with 0.2% of glucose. In the control growth condition, corresponding to the input pool, no AMP was added. For the output pools, the minimal medium was supplemented with five different AMPs, each tested with two concentrations: polymyxin B (1.5 µg.mL^{-1} and 12.5 µg.mL^{-1}), LL-37 (1.5 µg.mL^{-1} and 12.5 µg.mL^{-1}), riptocin (100 µg.mL^{-1} and

200 µg.mL^{-1}), CCR179 peptide (100 µg.mL^{-1} and 200 µg.mL^{-1}) and CCR480 peptide (25 µg.mL^{-1} and 100 µg.mL^{-1}). These growth cultures were all supplemented with 30 µg.mL^{-1} of Rif and 50 µg.mL^{-1} of Km, and were incubated at 28°C, with shaking at 180 rpm. Once the cultures reached an OD_{600nm} of 1, corresponding to approximately 7 generations, the bacteria were collected by centrifugation at 4000 g for 20 minutes at 4°C and the pellets were kept for DNA extraction. Each condition was performed in triplicates. DNA extractions, preparation of the Illumina sequencing libraries, sequencing and bioinformatics analysis to identify fitness genes were performed essentially as described in **Chapter II section 3**.

In vitro susceptibility tests

B. insecticola mutants were tested for their susceptibility to varying concentrations of a range of stress molecules in microdilution assays. Microdilution assays were performed in ninety-six flat well plates in MM supplemented with 0.2% of glucose. Exponential phase cultures of each *Burkholderia* strain were prepared and inoculated to an initial OD_{600nm} of 0.05 in the microtiter plates. Different chemicals were tested in a two-fold serial dilutions manner with specific ranges of concentrations as follows (minimal-maximal concentrations): polymyxin B (0.39-200 µg.mL^{-1}); colistin (0.39-200 µg.mL^{-1}); sodium-dodecyl sulphate (SDS) (0.002-1%); NaCl (0.977-500 mM); H_2O_2 (0.005-2.5 mM); gentamicin (0.05-25 µg.mL^{-1}); ampicillin (0.05-25.6 mg.mL^{-1}); tetracycline (0.006-3.125 µg.mL^{-1}); chloramphenicol (0.2-100 µg.mL^{-1}); trimethoprim (0.78-400 µg.mL^{-1}); CCR008 (1-100 µg.mL^{-1}); CCR179 peptide (0.2-100 µg.mL^{-1}); CCR480 peptide (0.2-100 µg.mL^{-1}); LL-37 (0.2-100 µg.mL^{-1}) and riptocin (0.39-200 µg.mL^{-1}). The riptocin, LL-37 and CCR peptides were obtained by chemical synthesis by a commercial peptide synthesis service (Proteogenix). The plates were incubated at 100 rpm at 28°C. The OD_{600nm} was measured with an automated microtiter plate reader after 15 hours of incubation. Each tested condition was performed in triplicates.

Additional conditions that were tested were temperature variations (20°C, 28°C, 37°C and 40°C) and pH variations in MM (6, 7 and 8) and YG medium (5, 6.3 and 8), in the same conditions as described above.

Swimming motility tests

Each *Burkholderia* strain was cultured until reaching an exponential phase ($0.4 < OD_{600nm} < 0.8$) and washed twice with 0.8% NaCl solution. Each bacterial pellet was resuspended with

0.8% NaCl solution at a final OD_{600nm} of 0.3. 1 µL of this bacterial suspension was inoculated on soft agar YG plates (0.3% agar) and incubated at room temperature for 24 hours. The swimming radius for each strain was measured and plates were photographed after 24 hours of incubation. Statistical analyses on the radius measurements were performed using a parametric one-way ANOVA test followed by a Tukey post-hoc test with a p-value < 0.05 on R software version 3.4.2.

Insect mono-infections

B. insecticola mutants were tested for their capacity to establish symbiosis in *R. pedestris*. Insects were collected two days after birth, at the second instar larval stage, and water was removed to make them thirsty before subsequent infection the following day. At three days after birth, the second instar nymphs were infected with a *Burkholderia* inoculum solution diluted at 10^7 CFU.mL^{-1} in sterile water, and kept in Petri dishes with soybean seeds in a 25°C room. The infection test was repeated three times for each tested *Burkholderia* strain.

Measurement of symbiont titers in M4 organs

At three days and five days post-infection, corresponding to second and third instar larval stages respectively, ten *R. pedestris* insects were dissected under a binocular microscope for direct observation of GFP signal, and pictures were taken with Leica LAS EZ software version 3.4. The M4 organ was collected in 250 µL of sterilized PBS solution (170 mM NaCl, 3.3 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4). Each M4 organ was homogenized with a plastic pestle and after homogenization, the pestle was washed with 250 µL of sterilized PBS solution. The resulting pooled suspension of the M4 material (final volume of 500 µL) was serially diluted in sterile water and each dilution was spread onto YG agar plates containing 30 µg.mL^{-1} Rif, and YG agar plates supplemented with 30 µg.mL^{-1} Rif and 50 µg.mL^{-1} Km, and incubated at 28°C for two days. After incubation, colonies were counted and the total number of CFU per insect was assessed. The same procedure was performed for the M3 organs with the LPS mutants. Statistical analyses were performed using a non-parametric Kruskal-Wallis test followed by a Dunn post-hoc test with a p-value < 0.05 on R software version 3.4.2.

Competition assays and flow cytometry analysis

For *in vitro* competitions between the wild-type strain and the mutant strains, exponential

phase cultures of the wild-type *B. insecticola* RPE525 strain, which carries a RFP-tag and the *Burkholderia* mutant strain tagged with GFP were prepared and mixed together at a 1:1 ratio to an initial concentration of 10^7 CFU.mL^{-1} in YG medium. This initial mix was subjected to CFU counting on YG agar plates supplemented with 30 µg.mL^{-1} of Rif, to check the initial proportions contained in the inoculum. Each mix was spotted in triplicates. The mixed bacterial suspension was incubated at 28°C, with shaking 180 rpm. After 20 hours of incubation, the bacterial mix was subjected to flow cytometry measurements and CFU counting on YG agar plates supplemented with 30 µg.mL^{-1} of Rif. Each mix was prepared in triplicates.

For *in vivo* competitions between the wild-type strain and the mutant strains, insects were collected two days after birth, at second instar larval stage, and water was removed one day before subsequent infection. At three days after birth, second instar nymphs were co-infected with the wild-type RFP-tagged *B. insecticola* RPE525 strain and the GFP-tagged *Burkholderia* mutant strain, mixed together at a 1:1 ratio to an initial concentration of 10^7 CFU.mL^{-1} in sterile water. Infected insects were kept in Petri dishes with soybean seeds in a 25°C room. At three days post-infection, corresponding to second instar larval stage, ten *R. pedestris* insects were dissected under a binocular microscope for direct observation of RFP and GFP signals, and pictures were taken with Leica LAS EZ software version 3.4. The merged fluorescence pictures were obtained with ImageJ version 1.8.0. The M4 organ was collected in 250 µL of sterilized PBS solution. Each M4 organ was homogenized with a plastic pestle and after homogenization, the pestle was washed with 250 µL of sterilized PBS solution. The resulting suspension of the M4 bacteria was used for flow cytometry measurements and a small fraction was serially diluted in sterile water for CFU counting.

For CFU counting, each dilution was spread onto YG agar plates containing 30 µg.mL^{-1} of Rif and incubated at 28°C for two days. After incubation, RFP and GFP colonies were counted and the total number of CFU per insect was assessed.

For flow cytometry analysis, bacteria were filtered through a 50-µm nylon filter and analysed by a CytoFlex S flow cytometer (Beckman Coulter) driven by Cytexpert software. A first gating was made on the forward-scatter (FSC)-side scatter (SSC) dot plot to focus on bacteria. Doublets were discarded using the SSC_Area-SSC_Height dot plot. GFP fluorescence was excited by a 488-nm laser and collected through a 525/40 nm band pass filter; RFP

fluorescence was excited by a 561-nm laser and collected through a 610/20 nm band pass filter. Data acquisition for a total of 15,000-20,000 bacteria was performed for each condition. Thresholds for considering positive events for GFP and RFP were determined using non-fluorescent control bacteria.

Competitive index (CI) values were calculated for each competition mix based on the following formula (Macho *et al.*, 2016):

$$\text{CI} = \frac{(\text{proportion of mutant strain} \div \text{proportion of wild-type srain})\ \textit{in vivo}\ \text{or}\ \textit{in vitro}\ \text{competitions}}{(\text{proportion of mutant strain} \div \text{proportion of wild-type srain})\ \text{inoculum}}.$$

For *in vivo* competitions, as I have generated triplicates of inoculum and I have tested ten insects for *in vivo* competition tests, I have calculated thirty CI values for competition mix. For *in vitro* competitions, as I have generated triplicates of inoculum and triplicates of *in vitro* competition tests, I have calculated nine CI values for each competition mix. As the CI values obtained did not follow a normal distribution, statistical analyses were performed using a non-parametric Kruskal-Wallis test followed by a Dunn post-hoc test with a p-value < 0.05 on R software version 3.4.2.

Determination of host fitness parameters

The adult emergence rate was monitored with daily inspections by counting the number of newly molted adult insects from fifth instar nymphs. For body weight and measurements, young adult insects were immersed in 100% acetone for one month by changing the acetone bath every two weeks, and were dried for one day at room temperature. The total body size, abdomen size, thorax size, abdomen width, thorax width and the dry weight were measured for each young adult insect and their gender was characterized by checking their abdomen genital region. For each *Burkholderia* strain, three batches of approximately fifty insects were infected and submitted to these measurements.

For each morphometric parameter, statistical analyses were performed using a parametric one-way ANOVA test followed by a Tukey post-hoc test with a p-value < 0.05 on R software version 3.4.2. For the adult emergence rate, the area under the curve (AUC value) was calculated for each developmental rate curve and for each of the triplicate insect batches. Statistical analyses were performed using these AUC values by running a non-parametric Kruskal-Wallis test followed by a Dunn post-hoc test with a p-value < 0.05 on R software version 3.4.2.

Hopanoid analysis

Hopanoids were determined in the RPE75 wild-type strain and its *shc*, *hpnA*, *hpnH*, *hpnJ* and *hpnN* mutant derivatives. Pre-cultures were grown overnight, diluted 100-fold and grown for 4 more hours. A volume of cultures was taken to obtain about $1.5x10^9$ CFU. This volume was centrifuged at 4000 g for 10 min, the supernatant was discarded and bacteria were kept at -80°C until the extraction. 1 mL of sterile water was used to resuspend the pellet, bacteria were transferred in glass tubes and 2.5 mL of MeOH was added to the mixture. Samples were sonicated for 15 min to lyse the bacteria. As sonication boiled samples, part of the MeOH evaporated. Samples were re-equilibrated by adding the lost volume of MeOH. 1.25 mL of dichloromethane (DCM) was added to the mixtures. Samples were incubated at room temperature during 30 min and then centrifuged at 4000 g for 10 min. The supernatant was transferred into a clean glass tube, 2.3 mL of sterile water and 2.3 mL of DCM were added, and the mixture was centrifuged again at 4000 g for 10 min. The organic, lower phase was harvested with a Pasteur pipette, taking care not to disturb the interphase. The remaining aqueous phase was washed with 2.3 mL of DCM followed by another centrifugation at 4000 g for 10 min. The organic phase was harvested as above and added to the first one. The DCM was evaporated using a N_2 flow at 40°C. Samples were subsequently analysed by LC-MS and GC-MS at Strasbourg University.

4. Results

4.1. Candidate gene approach

4.1.1. LPS biosynthesis genes

In *B. insecticola*, numerous genes from the LPS biosynthesis pathway were shown to be involved in the host colonization efficiency, such as *waaC* (Kim *et al*., 2017), *waaF* (Kim *et al.*, 2017) and *wbiF* genes (Kim *et al.*, 2016). The two genes *waaC* (*BRPE64_RS10300*) and *waaF* (*BRPE64_RS02300*) encode heptosyltransferases that are involved in the first steps of the inner core oligosaccharide LPS biosynthesis (Kim *et al.*, 2017). The *wbiF* gene (*BRPE64_RS10490*) encodes a glycosyl transferase that is associated to the outer core oligosaccharide LPS biosynthesis (Kim *et al.*, 2016). As *B. insecticola in vivo* cells were shown to lose the O-antigen part of LPS molecules, these previous studies focused on the ability of

these mutant strains to colonize the host midgut (Kim *et al.*, 2016, 2017).

Here, I have tested the sensitivity of these three mutant strains towards different membrane

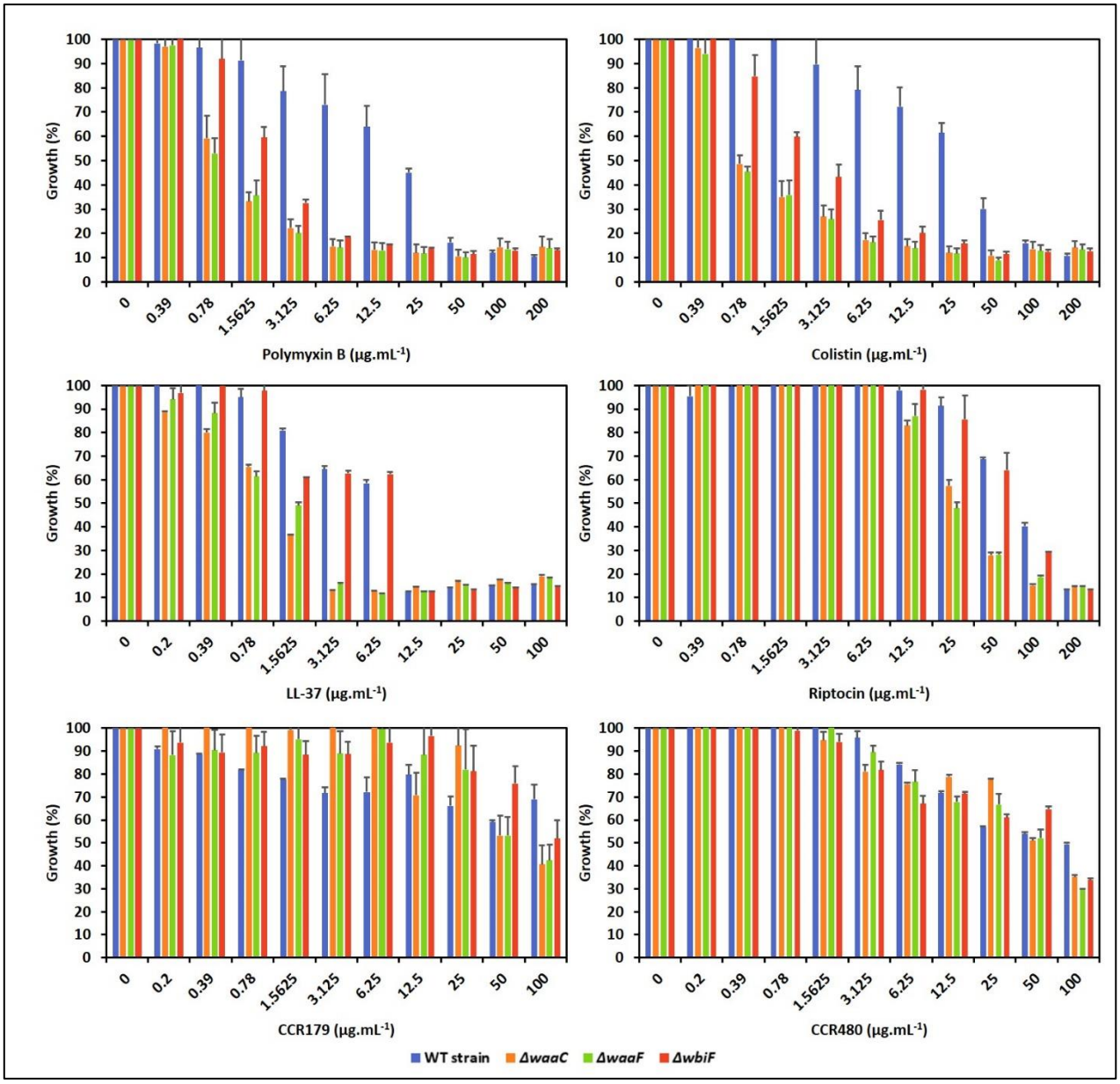


Figure 38: AMP sensitivity of *B. insecticola* LPS mutant strains.
Growth of wild-type and LPS mutant strains of *B. insecticola* in MM supplemented with increasing concentrations of AMPs (polymyxin B, colistin, LL-37, riptocin, CCR179 peptide, CCR480 peptide). Growth is expressed as a percentage of growth observed in the MM without AMPs, based on the measured OD_{600nm}. Abbreviations: WT: wild-type.

stressors, such as AMPs and other membrane-damaging agents (**Figures 38, 39 and 40**). For AMP sensitivity, I have tested six different AMPs: polymyxin B and colistin or polymyxin E (pentacationic polypeptides containing 10 AA with a fatty acid tail) which are produced by *Paenibacillus polymyxa* and are routinely administered to treat bacterial infections in clinical

cases (Cai *et al.*, 2015); LL-37 or the human cathelicidin (37 AA, arginine- and lysine-rich peptide) which is mostly produced by neutrophils (Fabisiak *et al.*, 2016); riptocin (19 AA, proline-rich peptide), an immunity-related AMP produced by *R. pedestris* (see **Chapter I**)

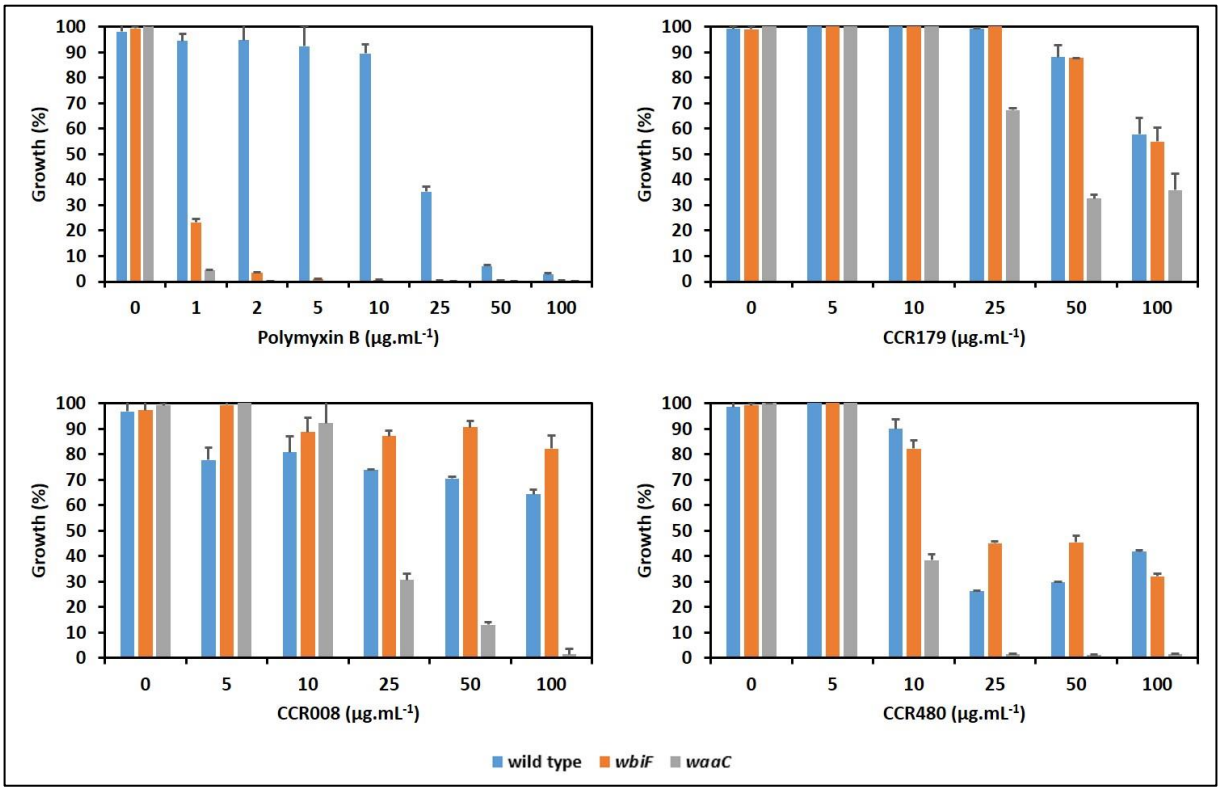


Figure 39: CCR sensitivity of *B. insecticola* LPS mutant strains.
Growth of wild-type and LPS mutant strains of *B. insecticola* in MM supplemented with increasing concentrations of AMPs (polymyxin B, CCR008 peptide, CCR179 peptide, CCR480 peptide). Growth is expressed as a percentage of growth observed in the MM without AMPs, based on the measured OD_{600nm}.

(Kim et al., 2015); and two CCR peptides that are specifically expressed in the M4 organ of *R. pedestris* (see **Chapter I**) (Futahashi *et al.*, 2013), CCR179 (53 AA, cysteine-rich peptide, 3 predicted disulphide bridges) and CCR480 (83 AA, cysteine-rich peptide, two predicted disulphide bridges). In addition to these AMPs, I have tested other membrane stressors, such as oxidative stress (H_2O_2), osmotic shock (NaCl), detergents (sodium dodecyl sulfate or SDS) and variations of growth temperatures and pH of the medium. Concerning the sensitivity towards AMPs, I noticed that the minimum inhibitory concentrations (MIC) that could inhibit the growth of 90% of the bacterial population (MIC_{90}) for the three LPS mutant strains towards polymyxin B and colistin were lower (6.25 µg.mL^{-1}) compared to the MIC_{90} of the wild-type strain (50 µg.mL^{-1}) (**Figure 38**). The MIC_{90} values of LL-37 and riptocin for Δ*waaC* and Δ*waaF*

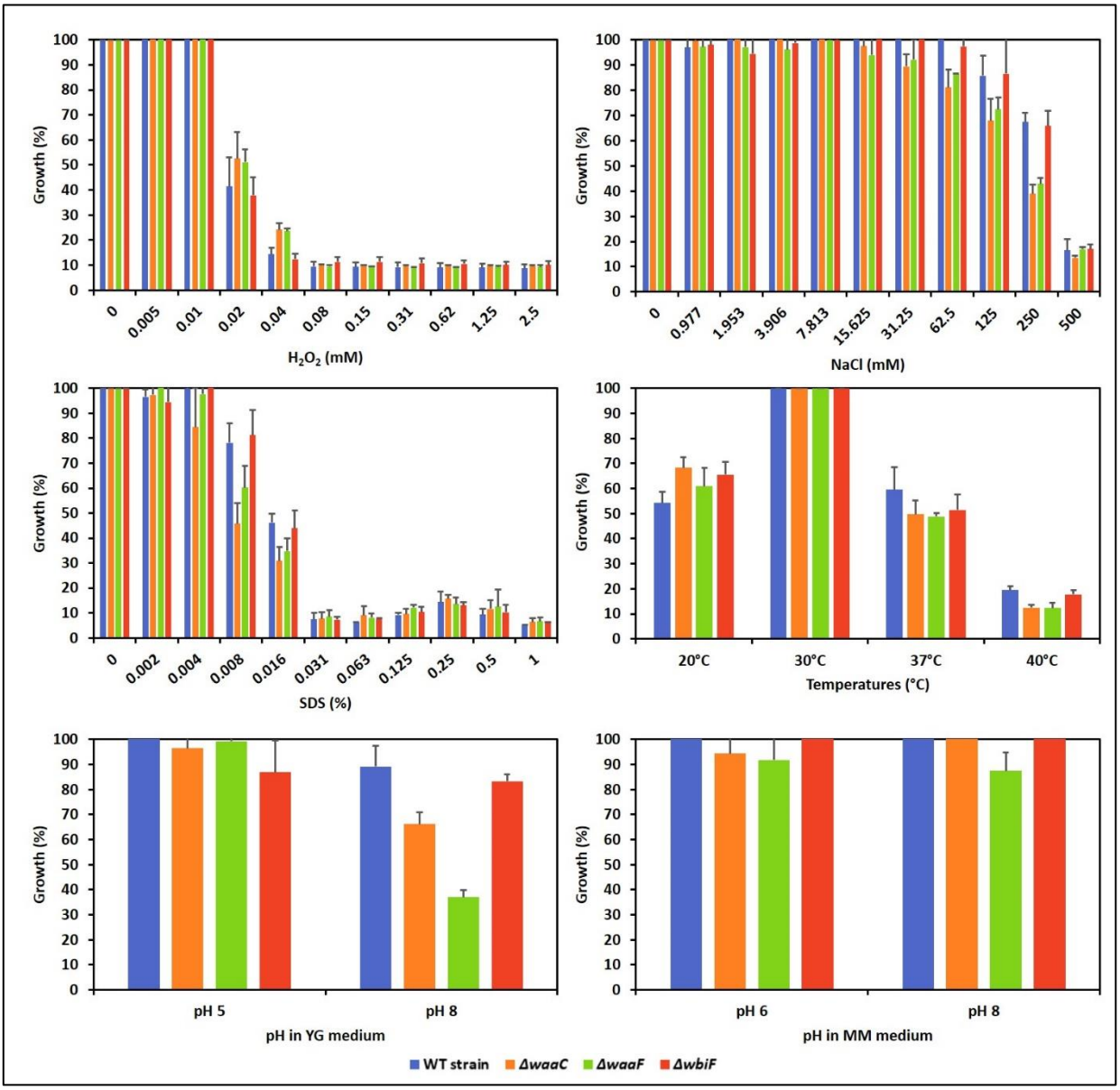


Figure 40: Sensitivity to various membrane stressors of *B. insecticola* LPS mutant strains. Growth of wild-type and LPS mutant strains of *B. insecticola* in MM supplemented with increasing concentrations of membrane damaging agents (H_2O_2, NaCl, SDS) and in various temperatures and pH growth conditions. Growth is expressed as a percentage of growth observed in the MM without the membrane stressor, based on the measured OD_{600nm}. For temperature variations, the growth is expressed as a percentage of growth in the MM at optimal growth temperature at 28°C. For pH variations in YG and MM media, the growth is expressed as a percentage of growth at optimal pH conditions measured in YG and MM media (pH of 6.3 and pH of 7, respectively). Abbreviations: WT: wild-type.

mutants (3.125 and 100 µg.mL^{-1}, respectively) were also lower than the MIC_{90} of the wild-type strain (12.5 and 200 µg.mL^{-1}, respectively) (**Figure 38**). However, the Δ*wbiF* mutant had the same sensitivity than the wild-type strain towards LL-37 and riptocin (**Figure 38**). For the two CCR peptides, I wasn't able to determine the MIC_{90} values for each strain, including the wild-

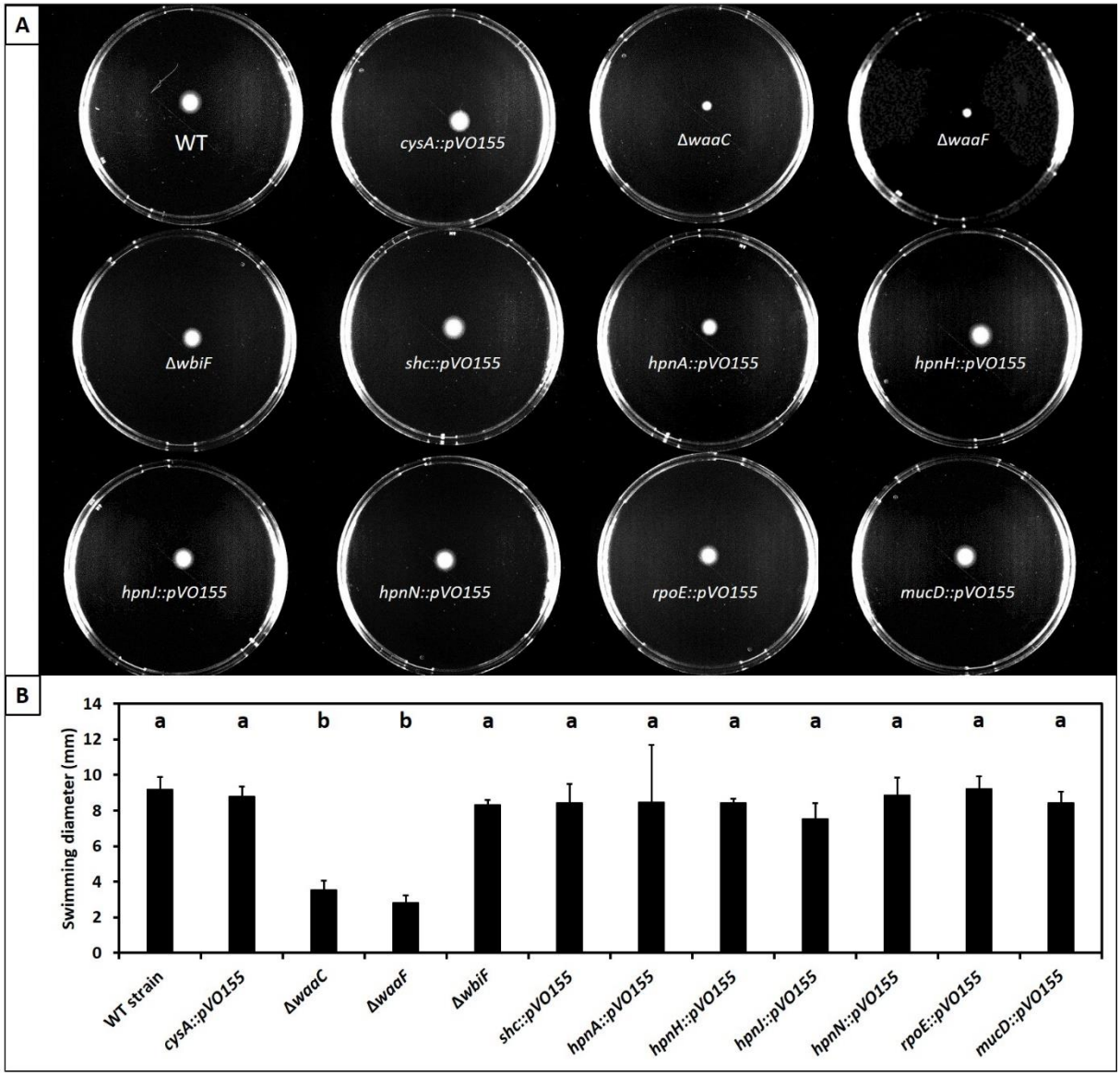


Figure 41: Swimming motility of *B. insecticola* mutants of candidate genes.
A) Pictures of swimming motility assays for each strain in YG soft agar plates after 24h. B) Swimming diameter measurements for each strain in YG soft agar plates. The above letters indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey post-hoc correction). Abbreviations: WT: wild-type.

type strain, for the concentration range that I have used (**Figure 38**). However, I noticed that the MIC_{50} values (the concentration for which half of the bacterial growth is inhibited) for CCR480 were similar between the LPS mutants and the wild-type *B. insecticola* (**Figure 38**). For CCR179, I could not determine the MIC_{50} value for the wild-type strain, but I observed that the LPS mutant strains were more susceptible to CCR179 at 100 $\mu g.mL^{-1}$ compared to the wild-type strain (**Figure 38**).

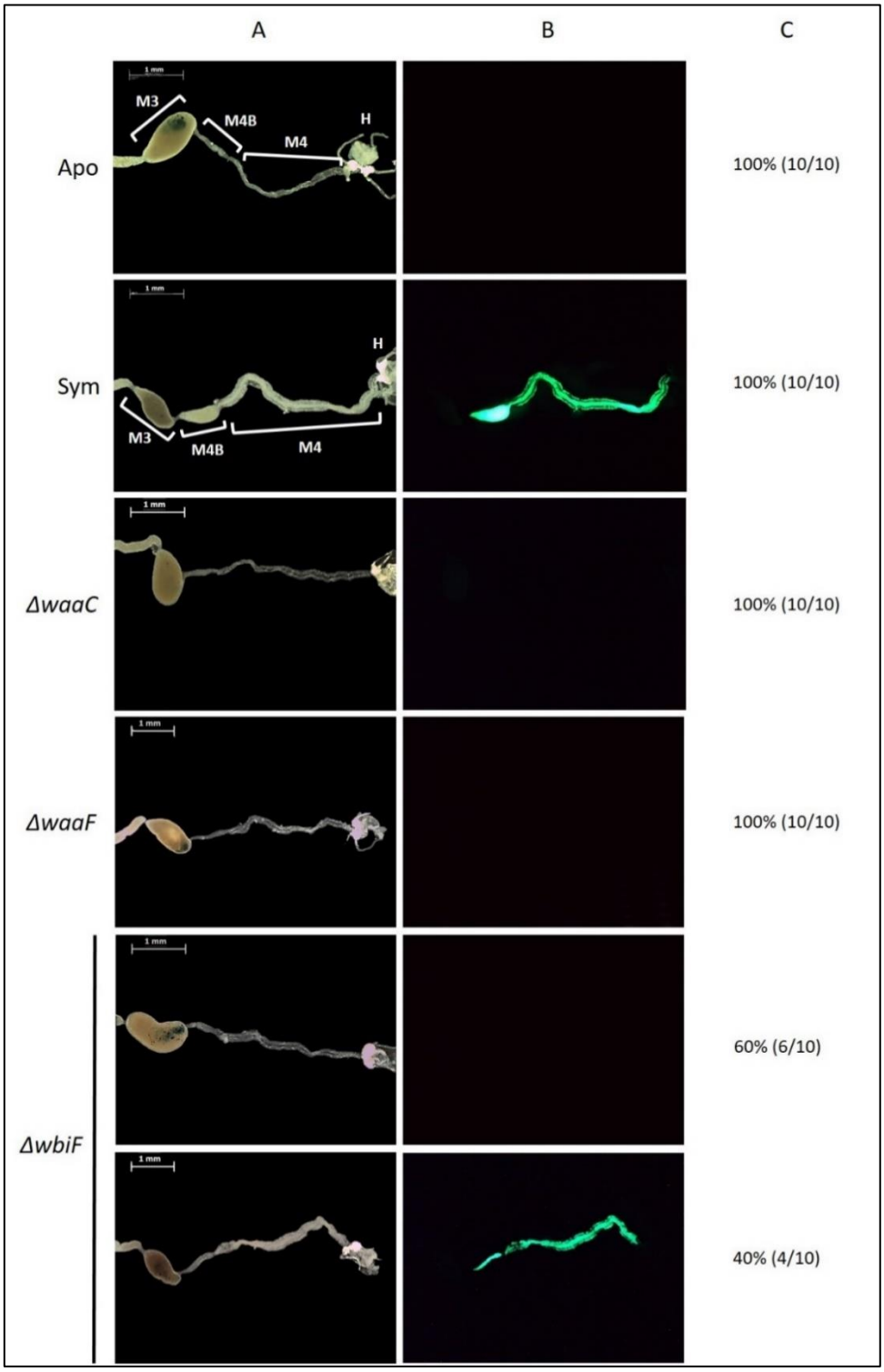


Figure 42: Midgut morphologies of second instar *R. pedestris* insects infected with LPS mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

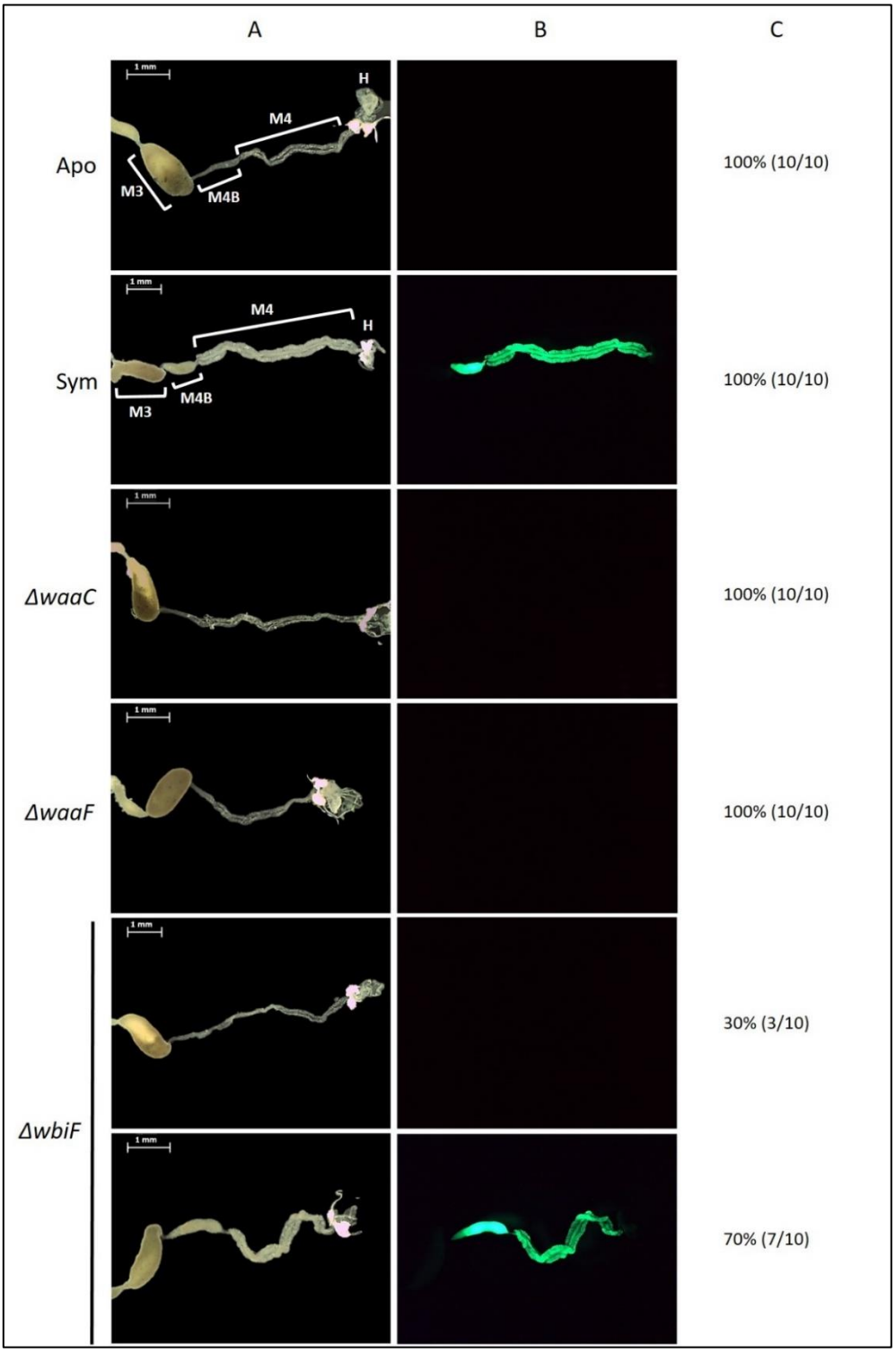


Figure 43: Midgut morphologies of third instar *R. pedestris* insects infected with LPS mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

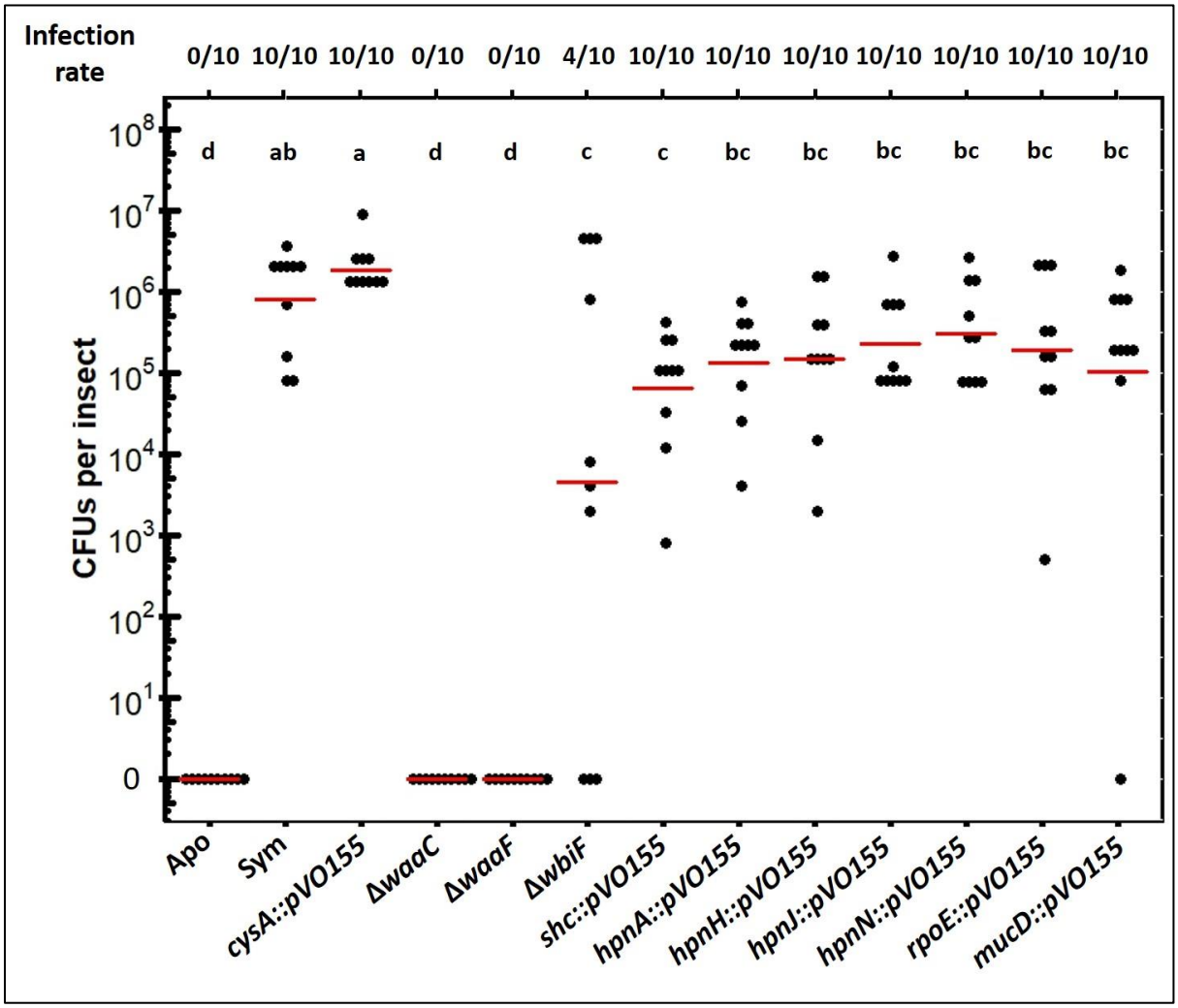


Figure 44: Symbiont titers of *B. insecticola* strains of candidate genes at the second instar stage in the M4 organ.

The total number of CFUs per insect was calculated for each condition. The related infection rate corresponds to the number of infected insects out of ten insects and is indicated above each condition. Means are indicated by red bars for each condition (n = 10 insects). The above letters indicate statistically significant differences (p-value < 0.05, Kruskal-Wallis). Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

Nevertheless, using an independent batch of peptides, a clear hypersensitivity to three different CCR peptides, CCR008, CCR179, and CCR480, was observed for the Δ*waaC* and Δ*waaF* mutants but not for the Δ*wbiF* mutant (**Figure 39**). The variability between batches (and even between experiments with the same batch) illustrates the difficulty to work with these peptides. Similar difficulties are encountered with other peptides like the legume NCRs in the laboratory and by other researchers. These problems can be related to the instability of the peptides during storage or in the assays, as well as the redox-state of the peptides with the presence or absence of disulphide bridges between cysteine residues.

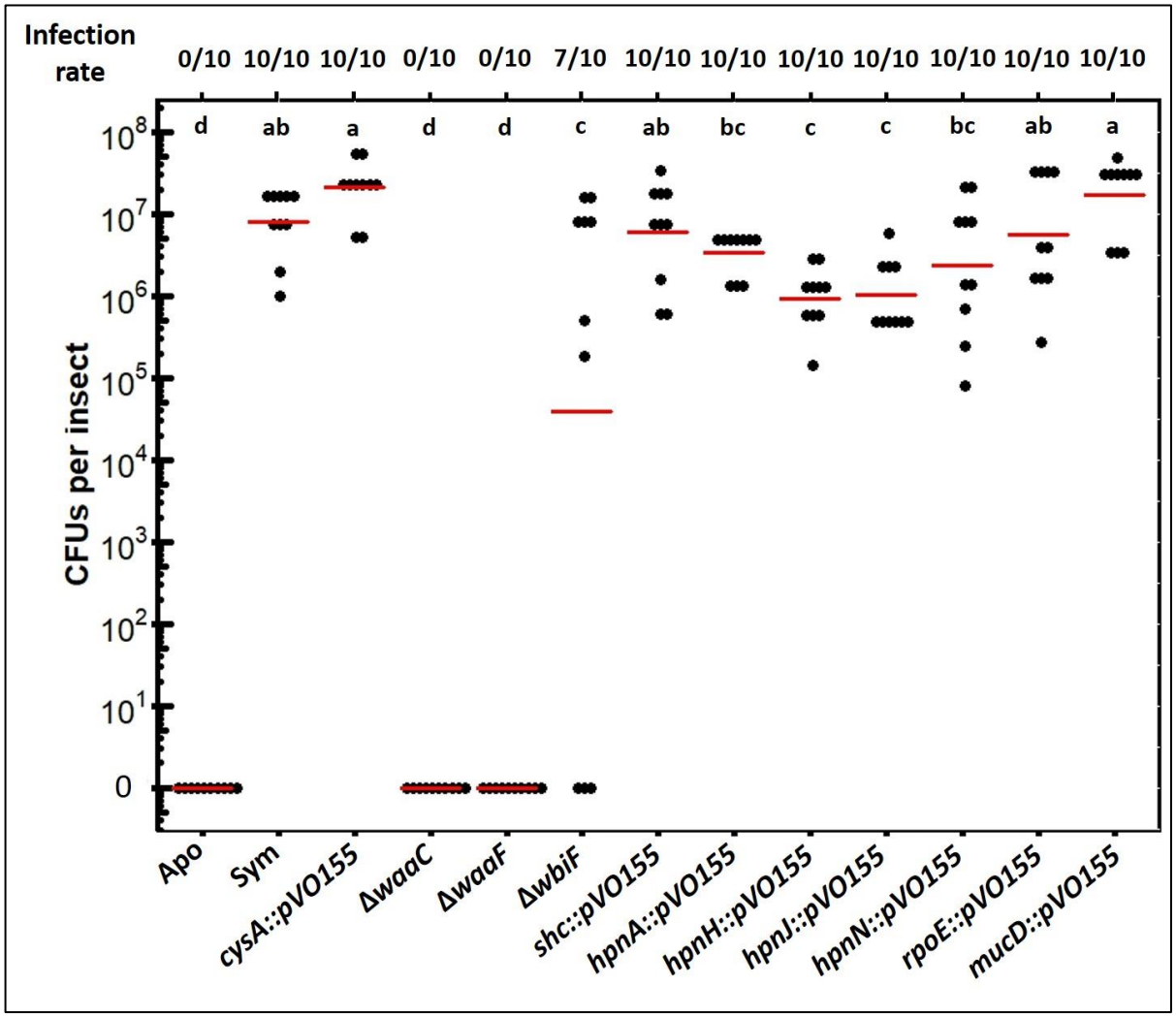


Figure 45: Symbiont titers of *B. insecticola* strains of candidate genes at the third instar stage in the M4 organ.

The total number of CFUs per insect was calculated for each condition. The related infection rate corresponds to the number of infected insects out of ten insects and is indicated above each condition. Means are indicated by red bars for each condition (n = 10 insects). The above letters indicate statistically significant differences (p-value < 0.05, Kruskal-Wallis). Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

Regarding the other membrane stressors, the LPS mutants did not show any significant difference of sensitivity towards these conditions compared to the wild-type strain, except for the Δ*waaC* and Δ*waaF* mutants that grew less efficiently at pH 8 in YG medium compared to the growth of the wild-type strain (**Figure 40**). In addition to these *in vitro* sensitivity tests, I also checked the swimming motility of these mutants in YG soft agar plates (**Figure 41A**). The Δ*waaC* and Δ*waaF* mutants were less motile than the wild-type strain, whereas the Δ*wbiF* mutant had the same swimming diameter than the wild-type strain (**Figure 41B**). These results demonstrate that LPS mutants, especially the inner core oligosaccharide biosynthesis mutants

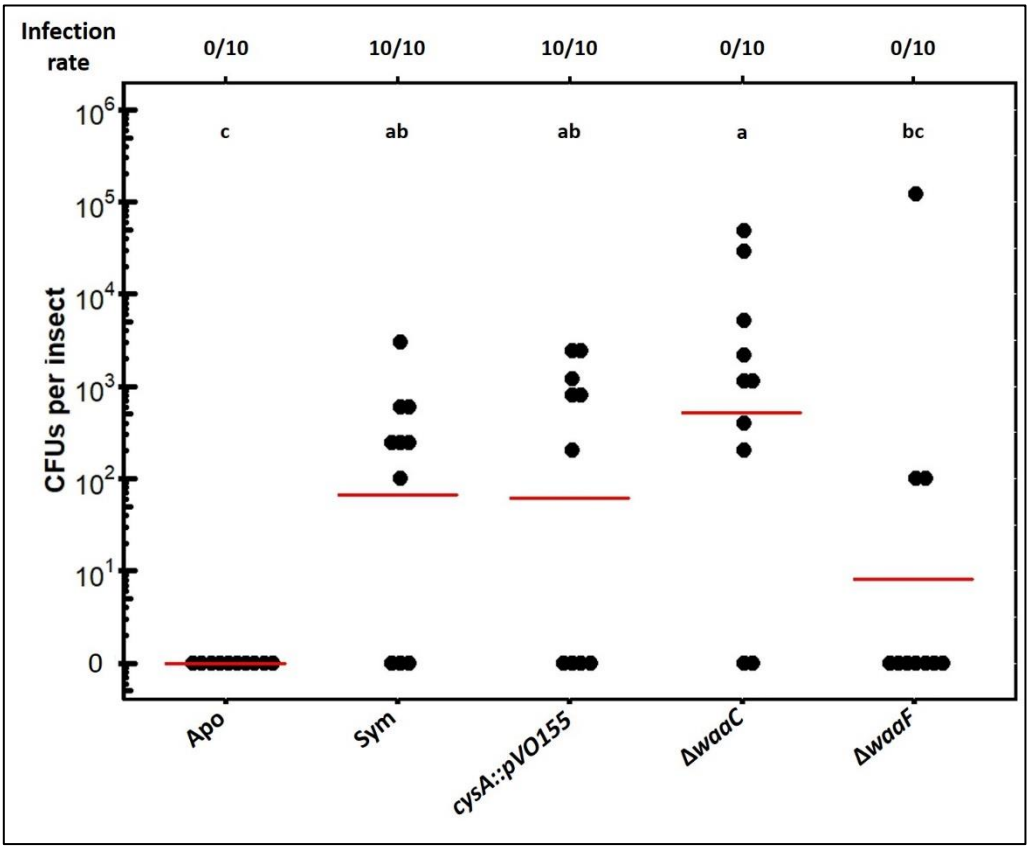


Figure 46: Symbiont titers of *waaC* and *waaF B. insecticola* mutants at the second instar stage in the M3 organ.

The total number of CFUs per insect was calculated for each condition. The related infection rate corresponds to the number of infected insects out of ten insects observed in the M4 organ, and is indicated above each condition. Means are indicated by red bars for each condition (n = 10 insects). The above letters indicate statistically significant differences (p-value < 0.05, Kruskal-Wallis). Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

Δ*waaC* and Δ*waaF*, are more susceptible to AMPs and are less motile than the wild-type strain. I conducted mono-infections of *R. pedestris* insects with these mutants in order to verify their capacity to colonize the M4 crypts. For that purpose, I infected young insects at early second instar stage and I have checked the presence of symbiotic bacteria in the M4 organ at the second and third instar stages (**Figures 42, 43, 44 and 45**). As these mutant strains were GFP-labelled, as well as the wild-type strain, the success of the mutants to establish in the crypts was observed by the presence of fluorescent bacteria inside the M4 organ (**Figures 42 and 43**). Insects infected with the wild-type strain showed a morphological differentiation of the M4 organ compared to the aposymbiotic insects (see **Chapter I**), with GFP-labelled symbionts visible in the M4 and the M4B regions (**Figures 42 and 43**). At the second and third instar stages, I observed that the Δ*waaC* and Δ*waaF* mutants were not able to colonize the M4

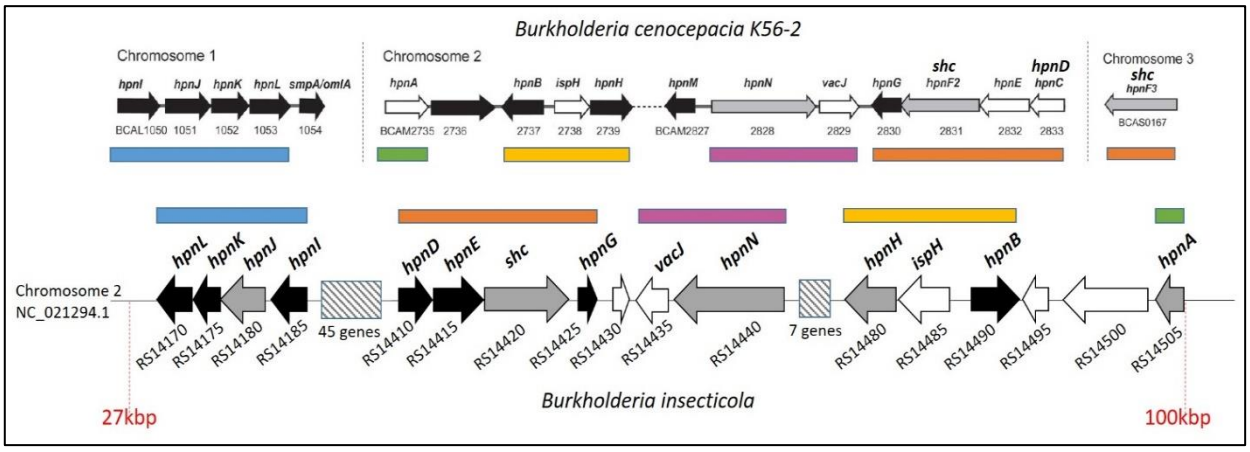


Figure 47: Identification of hopanoid biosynthesis genes.
Hopanoid biosynthetic gene clusters are indicated for *B. cenocepacia* on chromosomes 1, 2 and 3 with the above picture taken from *Schmerk et al.,* 2015, and for *B. insecticola* on chromosome 2. For *B. insecticola*, the black arrows indicate hopanoid genes that were not studied, grey arrows indicate hopanoid genes that have been targeted for this study, and white arrows indicate genes that are not involved in hopanoid biosynthesis. For *B. insecticola*, each gene name was attributed based on synteny results by BLAST analysis with *B. cenocepacia* homologues. The different colors indicate the synteny conservation of gene cluster organization between *B. cenocepacia* and *B. insecticola*. For each gene, the *B. insecticola* identifier was indicated below, for which "BRPE64_" was not included on the figure.

organs (**Figures 42 and 43**). The Δ*wbiF* mutant was able to colonize 40% of the insect population at the second instar stage (**Figure 42**), and this infection rate reached 70% at the third instar stage (**Figure 43**). By quantification of the symbiont titers contained in the M4 region, I clearly noticed that there were no symbiotic bacteria recovered from the M4 organs of insects infected with the Δ*waaC* and Δ*waaF* mutants at both instar stages (**Figures 44 and 45**). For the other LPS mutant, there were significant differences between the symbiont titers of insects infected with the Δ*wbiF* mutant and insects infected with the wild-type strain, with a mean bacterial load per insect of 10^4 and 10^6 at the second instar stage, respectively, and of 10^5 and 10^7 at the third instar stage, respectively (**Figures 44 and 45**). As the Δ*waaC* and Δ*waaF* mutants were not detected in the M4 organs of second and third instar insects, I have assessed the presence of these bacterial strains inside the M3 organ, which is the last midgut section before entering the symbiotic organ. I observed that both these LPS mutant strains were present in the M3 organ, with similar bacterial loads than the wild-type strain close to 10^2 CFUs per insect (**Figure 46**). Thus, the Δ*waaC* and Δ*waaF* mutants did not colonize the symbiotic organ but were able to reach the M3 organ (**Figure 46**). Taken together, these results showed that the three LPS mutant strains, more particularly the Δ*waaC* and Δ*waaF*

mutants, are not able to colonize efficiently the symbiotic organ, and confirmed previous published observations (Kim *et al.*, 2016, 2017). Interestingly, the two strains that were hypersensitive towards AMPs were the ones that were not able to colonize the symbiotic organ, which suggests a possible link between AMP resistance and colonization efficiency.

4.1.2. Hopanoids

One of the targets chosen for the candidate gene approach was the hopanoid biosynthesis pathway. As mentioned before, hopanoids are triterpenoid lipids that are present in the membranes of some Gram-negative bacteria, including *Burkholderia* species (see **Chapter I**) (Malott *et al.*, 2012; Schmerk *et al.*, 2011, 2015). It was shown that deletion mutants of *B. cenocepacia* and *B. multivorans* for hopanoid biosynthesis genes were more sensitive towards AMPs such as polymyxin B compared to the wild-type strains (Malott *et al.*, 2012; Schmerk *et al.*, 2011). The overall genes involved in the hopanoid biosynthesis pathway were mostly

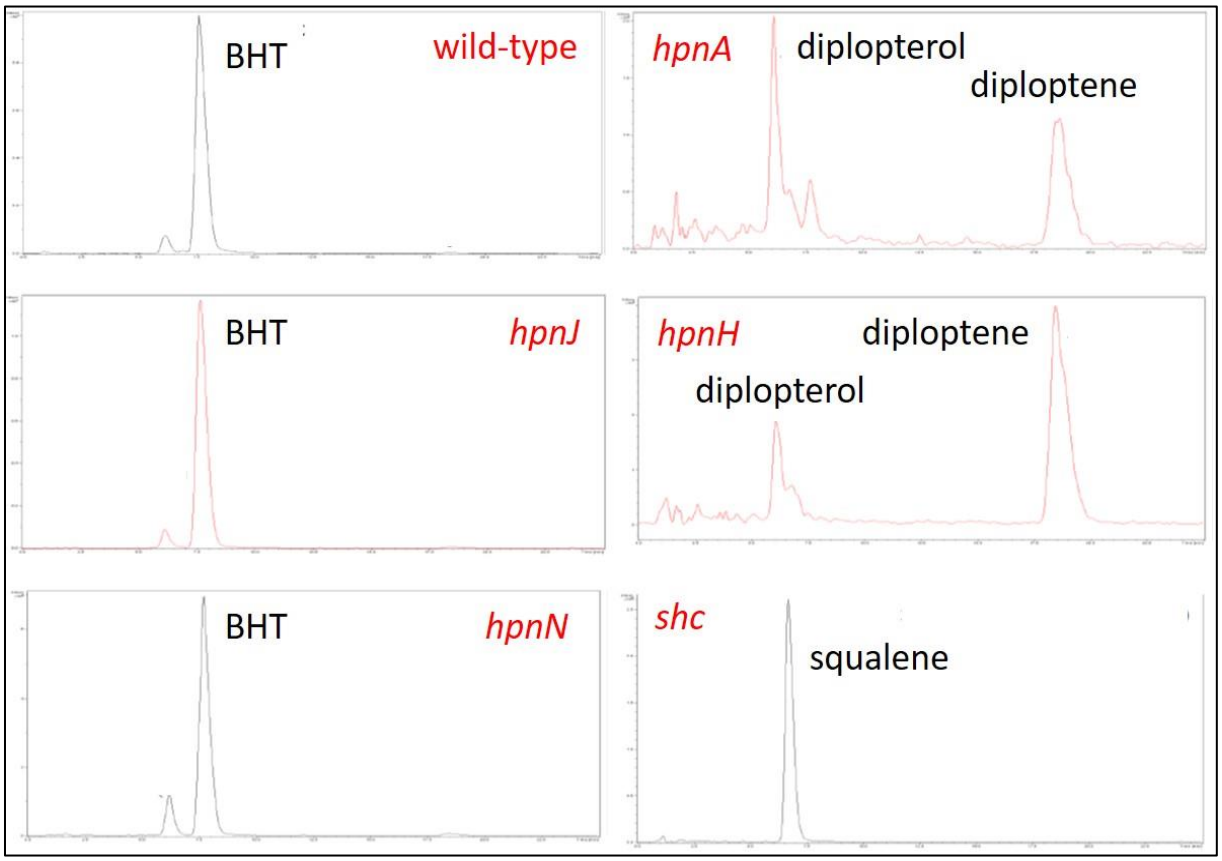


Figure 48: Hopanoid analysis in *B. insecticola* mutants.
HPLC chromatograms of lipid extracts of the indicated strains are shown. The identity of the peaks was confirmed by standard molecules and MS analysis. Samples were also analyzed by GC-MS (not shown), confirming the LC-MS analyses.

identified in *B. cenocepacia* (**Figure 47**) (see **Chapter I**) (Schmerk *et al.*, 2015). I identified the homologous genes in *B. insecticola* and targeted five of them for mutagenesis: *shc* (*BRPE64_RS14420*), *hpnA* (*BRPE64_RS14505*), *hpnH* (*BRPE64_RS14480*), *hpnJ* (*BRPE64_RS14180*), and *hpnN* (*BRPE64_RS14440*) (**Figure 47**). The *shc* gene encodes the squalene-hopene cyclase, which performs the first step of hopanoid synthesis from two molecules of squalene and forms diploptene, a C30 hopanoid (**Figure 17**) (Schmerk *et al.*, 2015). The *hpnA* gene encodes a sugar epimerase but its specific function in the pathway is not defined yet. The *hpnH* and *hpnJ* genes, corresponding to radical SAM proteins, are involved in the modification and processing of hopanoid molecules, producing extended forms of hopanoids (the C35 bacteriohopanetetrol (BHT) and the modified C35 BHT cyclitol ether) (**Figure 17**) (Schmerk *et al.*, 2015). The last target gene, *hpnN*, encodes an RND (resistance of nodulation and cell division) transporter which was shown to be involved in the translocation of hopanoids from the inner to the outer membrane of *Rhodopseudomonas palustris* (Doughty *et al.*, 2011).

Biochemical analysis of the hopanoids by HPLC-MS (**Figure 48**) and GC-MS (data not shown) demonstrated that the wild-type strain produced mainly BHT but surprisingly, no BHT glucosamine or BHT cyclitol ether were detected in the lipid extracts despite the presence of the genes *hpnIKJ* (**Figure 47**) encoding the enzymes involved in the production of these modified hopanoids (**Figure 17**). Different reasons could explain the absence of these molecules in our extracts. First, it is possible that the extraction method did not allow the extraction of the glycosylated hopanoids. However, applying essentially the same extraction method as here, glycosylated hopanoids could be extracted from another *Burkholderia* strain (Schmerk *et al*., 2015) making this option not very likely. Another explanation could be that the *hpnIKJ* genes are conditionally expressed and not highly enough in the standard growth conditions used for this experiment. Whatever the reason, in agreement with the absence of modified hopanoids, the *hpnJ* mutant produced exactly the same hopanoids as the wild-type strain (**Figure 48**). In addition, the *hnpN* mutant also produced the same molecules as the wild-type strain, as expected (**Figure 48**). In the *hpnH* and *hpnA* mutants, only the C30 intermediate diploptene was detected (**Figure 48**). This finding is coherent with the function of HpnH and provides some new indications on the role of the *hpnA* gene, which should be, as *hpnH*, involved in the modification of diploptene. Finally, the *shc* mutant did, as expected, not

produce any hopanoids but accumulated squalene (**Figure 48**). Taken together, the lipid characterization of the strains showed that the principal hopanoid in *B. insecticola* is BHT and confirmed that the mutants are affected in the hopanoid biosynthesis as expected from their predicted function. Together, the strains could represent four different classes: the wild-type strain and the *hpnJ* mutant produce normal BHT hopanoids, the *hpnN* mutant is expected to have reduced hopanoids in the outer membrane, the *hpnA* and *hpnH* mutants produce mainly diploptene and finally, the *shc* mutant accumulates squalene.

Next, I have tested the sensitivity of the hopanoid biosynthesis mutants towards AMPs, but disappointingly, none of them showed a difference in their resistance profiles towards polymyxin B, LL-37 and the two tested CCR peptides (**Figure 49**). Only the MIC_{90} of riptocin for these hopanoid mutants (100 $\mu g.mL^{-1}$) was lower than the MIC_{90} of the wild-type strain and the *hpnJ* mutant (200 $\mu g.mL^{-1}$) (**Figure 49**). In addition, I could notice that the growth rate of these hopanoid mutants was similar to the one of the wild-type strain. When I tested other membrane stressors, I observed that the hopanoid mutants' growth was not affected by them (**Figure 50**). Concerning the motility of these mutant strains, the swimming diameter was similar to the one of the wild-type strain in YG soft agar plates (**Figure 41**). Thus, contrary to previous studies showing that hopanoids in *Burkholderia* are required for AMP resistance, low pH tolerance and motility (Doughty *et al.*, 2011; Malott *et al*., 2012; Schmerk *et al.*, 2011), the analysis of these mutants clearly showed that hopanoids produced by *B. insecticola* are not involved in these processes.

Despite the absence of any differences in the AMP resistance profiles of these hopanoid mutants compared to the wild-type strain, I evaluated their capacity to colonize the *R. pedestris* host. As these hopanoid genes were interrupted by a plasmid (pVO155) which contains a GFP cassette that is under the control of a different promoter than the wild-type GFP-labelled strain, resulting in a clearly detectable but relatively lower fluorescence signal, I have used a pVO155 insertion mutant in the *cysA* gene as a fluorescence control for pVO155 insertion mutants (**Figures 51 and 52**). It was previously shown that this *cysA* mutant strain was able to colonize the symbiotic organ as efficiently as the wild-type strain (**Figures 44, 45, 51 and 52**).

In the same conditions as previously described (see **section 4.1.1**), I observed that the five hopanoid mutants were able to colonize the M4 region, both at the second and the third instar

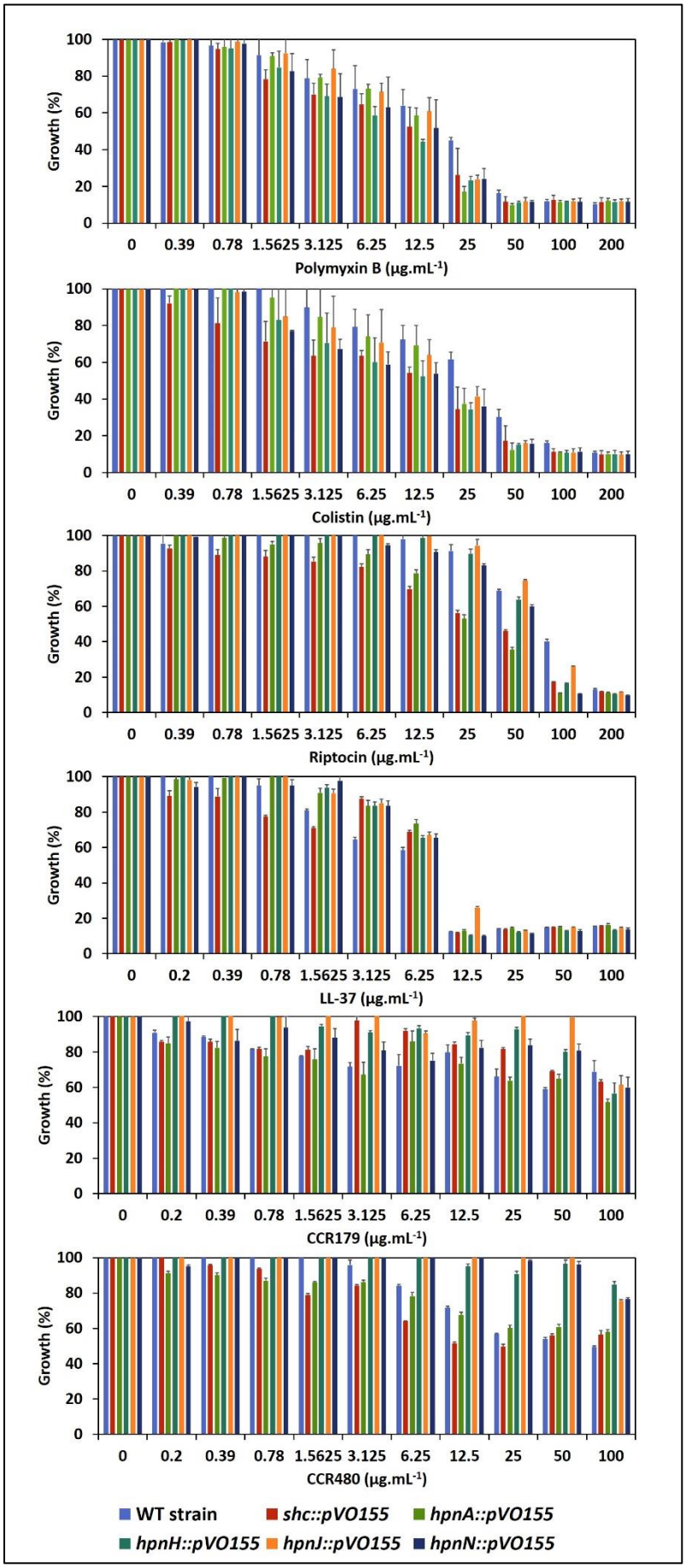


Figure 49: AMP sensitivity of *B. insecticola* hopanoid mutant strains.
Growth of wild-type and hopanoid mutant strains of *B. insecticola* in MM supplemented with increasing concentrations of AMPs (polymyxin B, colistin, LL-37, riptocin, CCR179 peptide, CCR480 peptide). Growth is expressed as a percentage of growth observed in the MM without AMPs, based on the measured OD_{600nm}. Abbreviations: WT: wild-type.

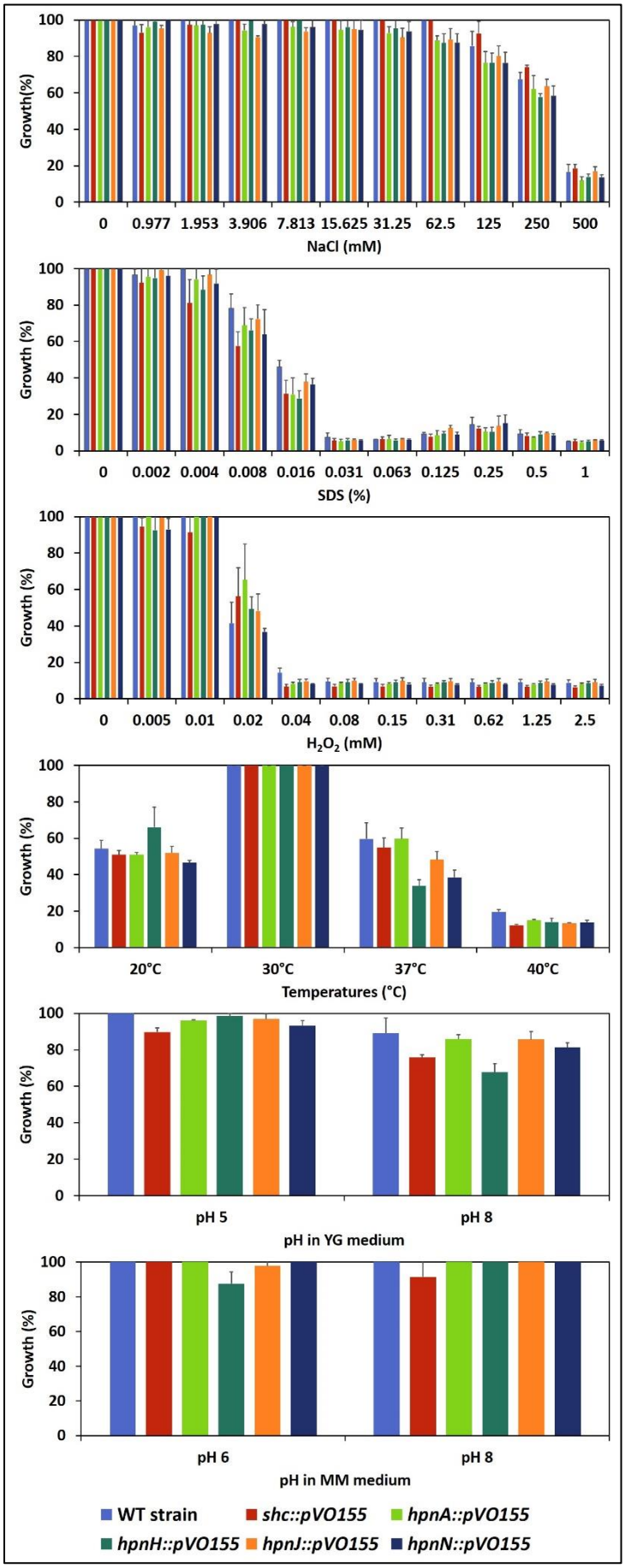


Figure 50: Sensitivity to various membrane stressors of *B. insecticola* hopanoid mutant strains.

Growth of wild-type and hopanoid mutant strains of *B. insecticola* in MM supplemented with increasing concentrations of membrane damaging agents (H_2O_2, NaCl, SDS) and in various temperatures and pH growth conditions. Growth is expressed as a percentage of growth observed in the MM without the membrane stressor, based on the measured OD_{600nm}. For temperature variations, the growth is expressed as a percentage of growth in the MM at optimal growth temperature at 28°C. For pH variations in YG and MM media, the growth is expressed as a percentage of growth at optimal pH conditions measured in YG and MM media (pH of 6.3 and pH of 7, respectively). Abbreviations: WT: wild-type.

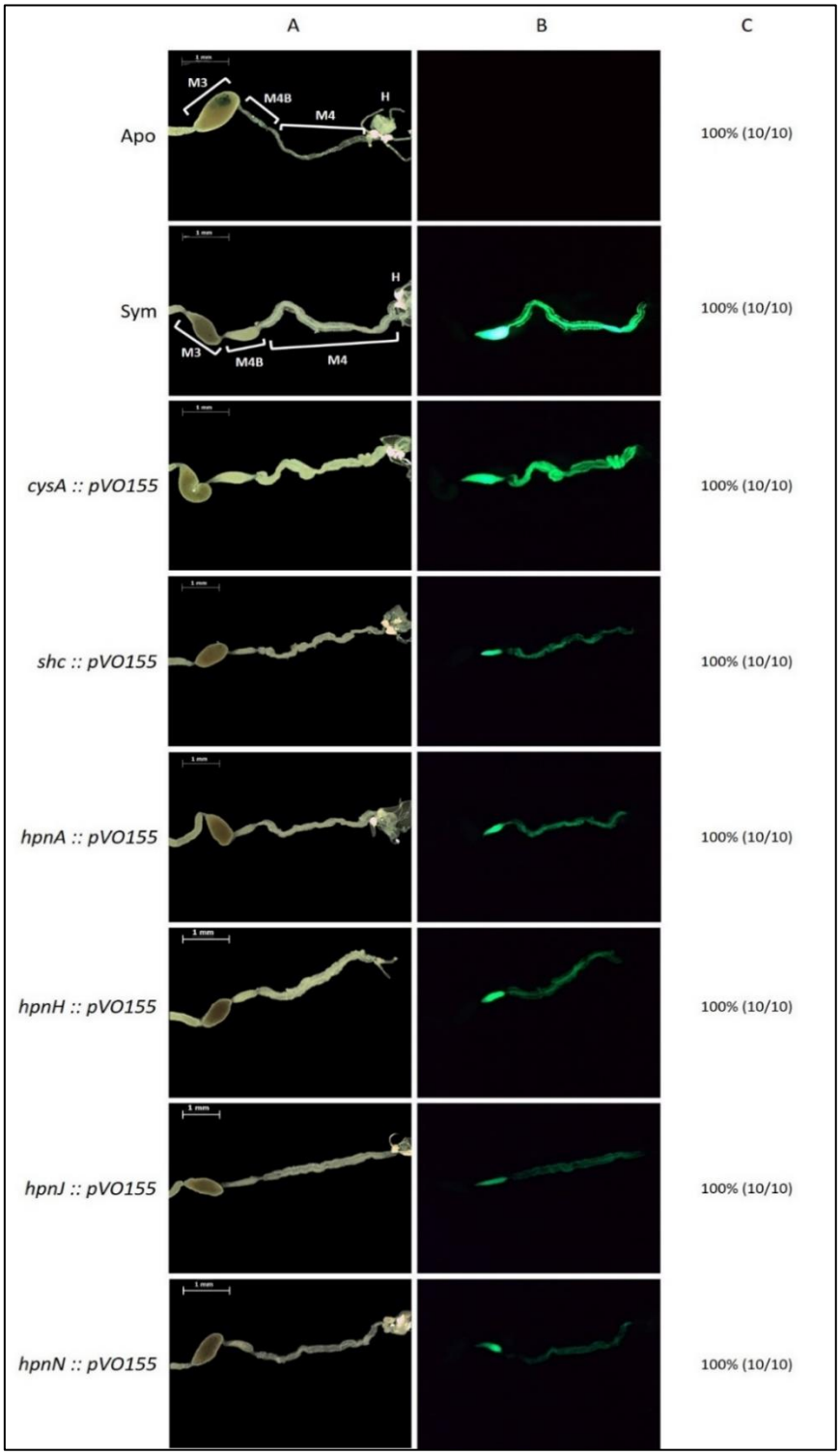


Figure 51: Midgut morphologies of second instar *R. pedestris* insects infected with hopanoid mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

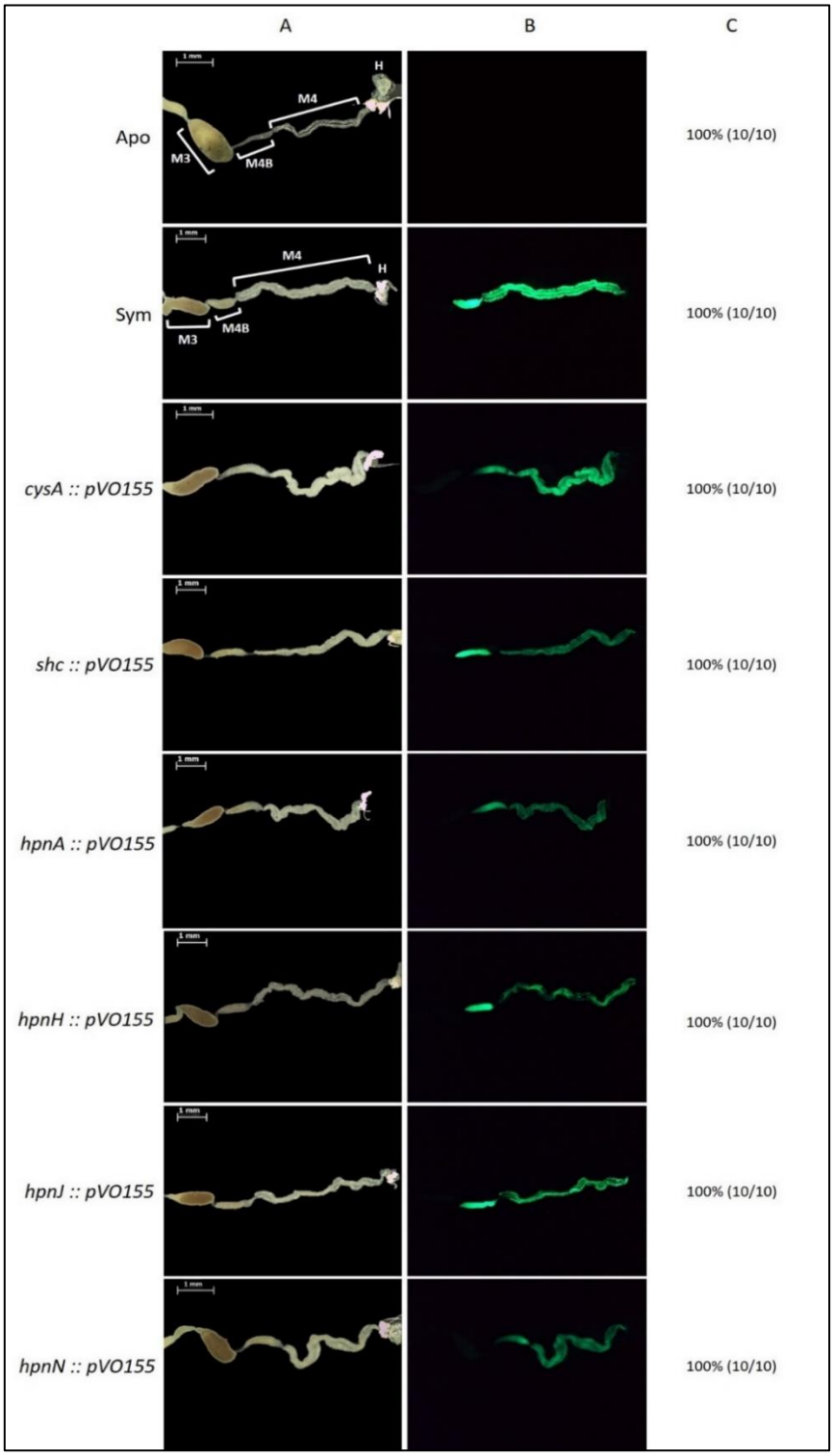


Figure 52: Midgut morphologies of third instar *R. pedestris* insects infected with hopanoid mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

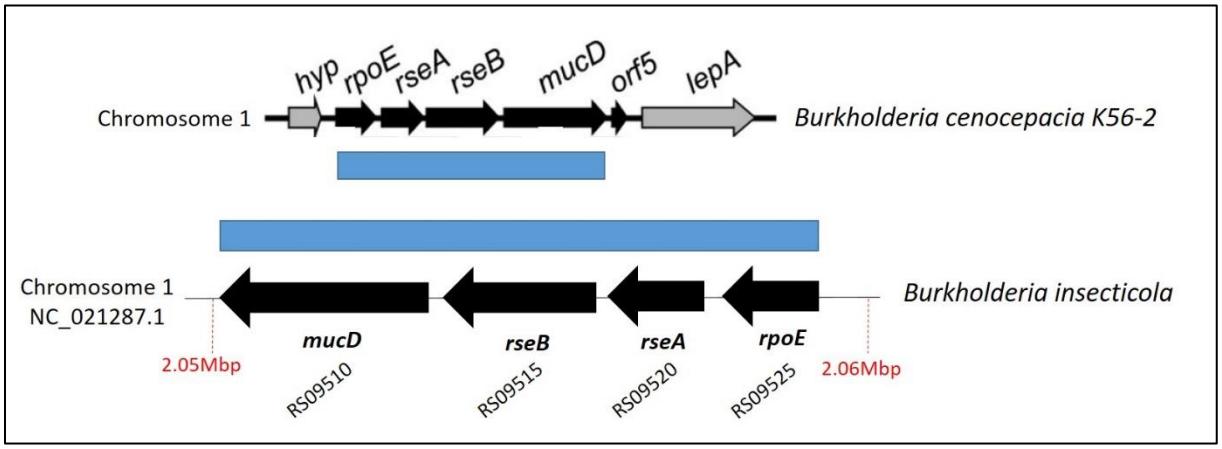


Figure 53: Identification of the *rpoE* operon.
The *rpoE* operon is indicated for *B. cenocepacia* on chromosome 1 with the above picture taken from Flannagan and Valvano, 2008, and for *B. insecticola* on chromosome 1. The black arrows indicate genes that are involved in the RpoE response, and grey arrows indicate genes that are not part of the rpoE operon. For *B. insecticola*, each gene name was attributed based on synteny results by BLAST analysis with *B. cenocepacia* homologues. The blue rectangles indicate the synteny conservation of this gene cluster organization between *B. cenocepacia* and *B. insecticola*. For each gene, the *B. insecticola* identifier was indicated below, for which "BRPE64_" was not included on the figure.

stages (**Figures 51 and 52**). However, the fluorescence intensity of these symbiotic mutants was weaker than the ones of the wild-type strain and the *cysA* mutant (**Figures 51 and 52**). By checking the proportion of viable bacteria inside the M4 organ at the second instar stage, I noticed that the mean bacterial load of insects infected with the *shc* mutant is significantly reduced (≈ 10^5 CFUs per insect) compared to insects infected with the wild-type strain and the other mutants (≈ 10^6 CFUs per insect) (**Figure 44**). At the third instar stage, the bacterial population contained in the symbiotic organ for *shc, hpnA and hpnN* mutants reached the same proportion contained in symbiotic insects (≈ 10^7 CFUs per insect) (**Figure 45**). However, the symbiotic population of *hpnH* and *hpnJ* mutant strains, close to 10^6 CFUs per insect, was significantly lower (**Figure 45**). Thus, even if I observed that the hopanoid mutant strains were able to colonize the *R. pedestris* host, these mutants were impacted in their efficiency for colonizing the symbiotic organ with a symbiotic population that cannot reach the same colonization level of the wild-type strain, especially for the *hpnH* and *hpnJ* mutants. However, this colonization defect is possibly not linked to a decreased AMP resistance in these mutants, as I originally hypothesized.

4.1.3. Extracellular Stress Response

The last target I have studied for this candidate gene approach was the ESR pathway, and more specifically the sigma factor o^E, also called RpoE. As mentioned before, the o^E type of ESR is activated during membrane damages by releasing the RpoE factor from the inner membrane, which will further activate the transcription of genes involved in membrane repair

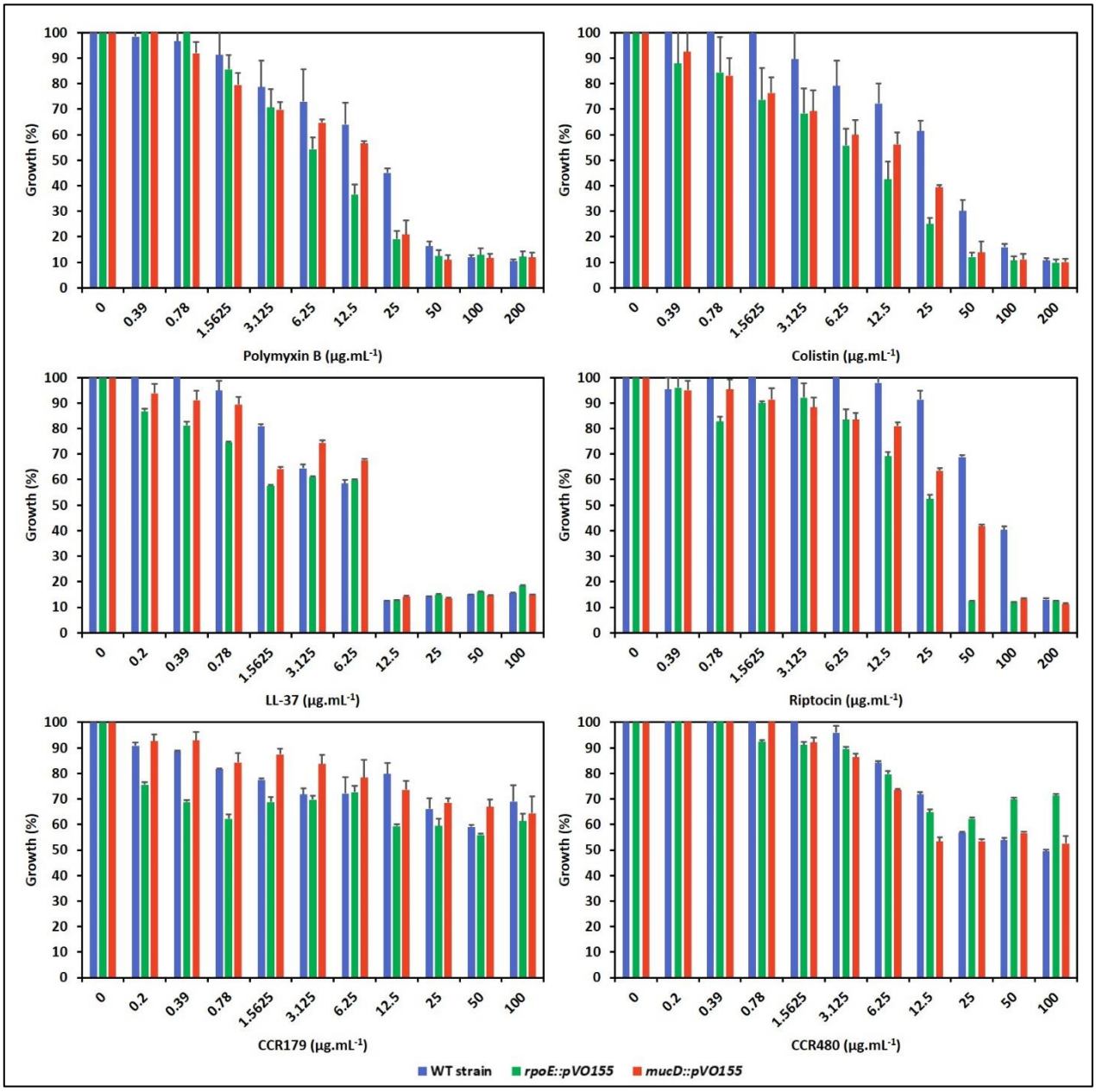


Figure 54: AMP sensitivity of *B. insecticola* ESR mutant strains. Growth of wild-type and ESR mutant strains of *B. insecticola* in MM supplemented with increasing concentrations of AMPs (polymyxin B, colistin, LL-37, riptocin, CCR179 peptide, CCR480 peptide). Growth is expressed as a percentage of growth observed in the MM without AMPs, based on the measured OD_{600nm}. Abbreviations: WT: wild-type.

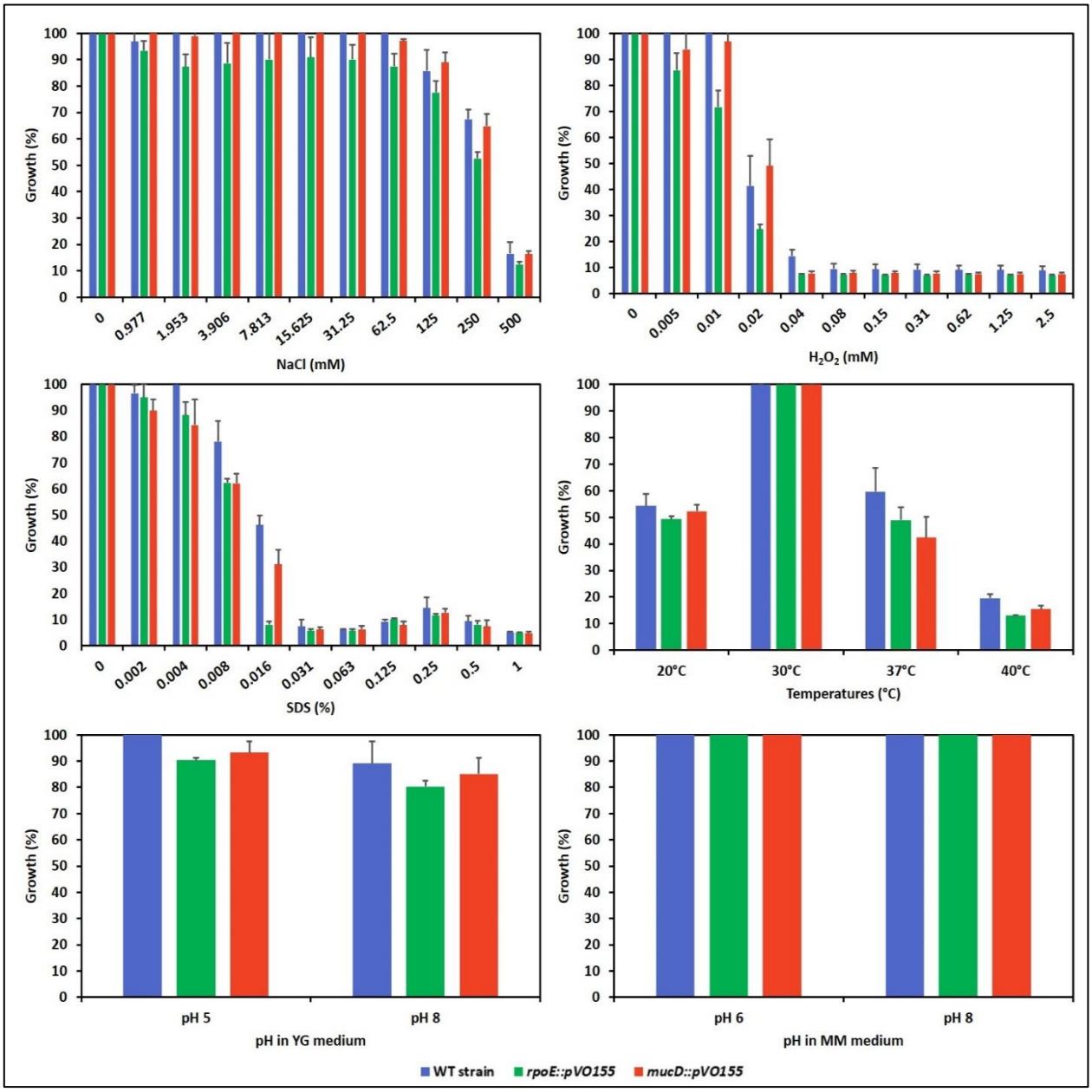


Figure 55: Sensitivity to various membrane stressors of *B. insecticola* ESR mutant strains. Growth of wild-type and ESR mutant strains of *B. insecticola* in MM supplemented with increasing concentrations of membrane damaging agents (H2O2, NaCl, SDS) and in various temperatures and pH growth conditions. Growth is expressed as a percentage of growth observed in the MM without the membrane stressor, based on the measured OD_{600nm}. For temperature variations, the growth is expressed as a percentage of growth in the MM at optimal growth temperature at 28°C. For pH variations in YG and MM media, the growth is expressed as a percentage of growth at optimal pH conditions measured in YG and MM media (pH of 6.3 and pH of 7, respectively). Abbreviations: WT: wild-type.

(see **Chapter I**) (Flores-Kim and Darwin, 2014; Guest and Raivio, 2016). Genes from the RpoE pathway were already identified in *B. cenocepacia* (Flannagan and Valvano, 2008), I identified their homologues in *B. insecticola* (**Figure 53**).

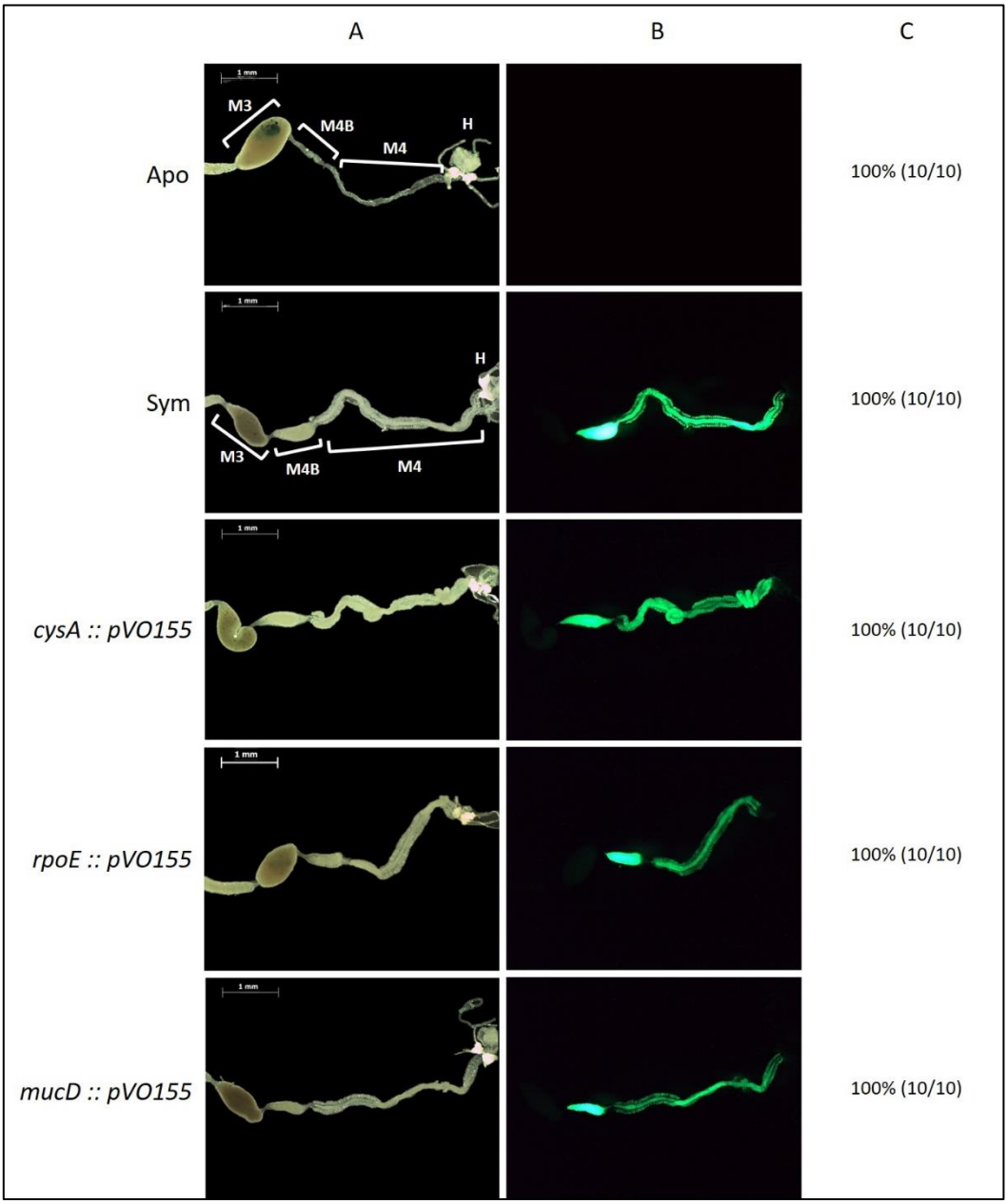


Figure 56: Midgut morphologies of second instar *R. pedestris* insects infected with ESR mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

The *rpoE* operon consists of four genes with *rpoE*, *rseA*, *rseB* and *mucD* (**Figure 53**) that are respectively coding for the sigma factor RpoE, the anti-sigma factors RseA and RseB which retain the RpoE factor to the membrane, and a protease Do (MucD) which could be potentially involved in the proteolytic cleavage to release the RpoE factor from the membrane (Flannagan

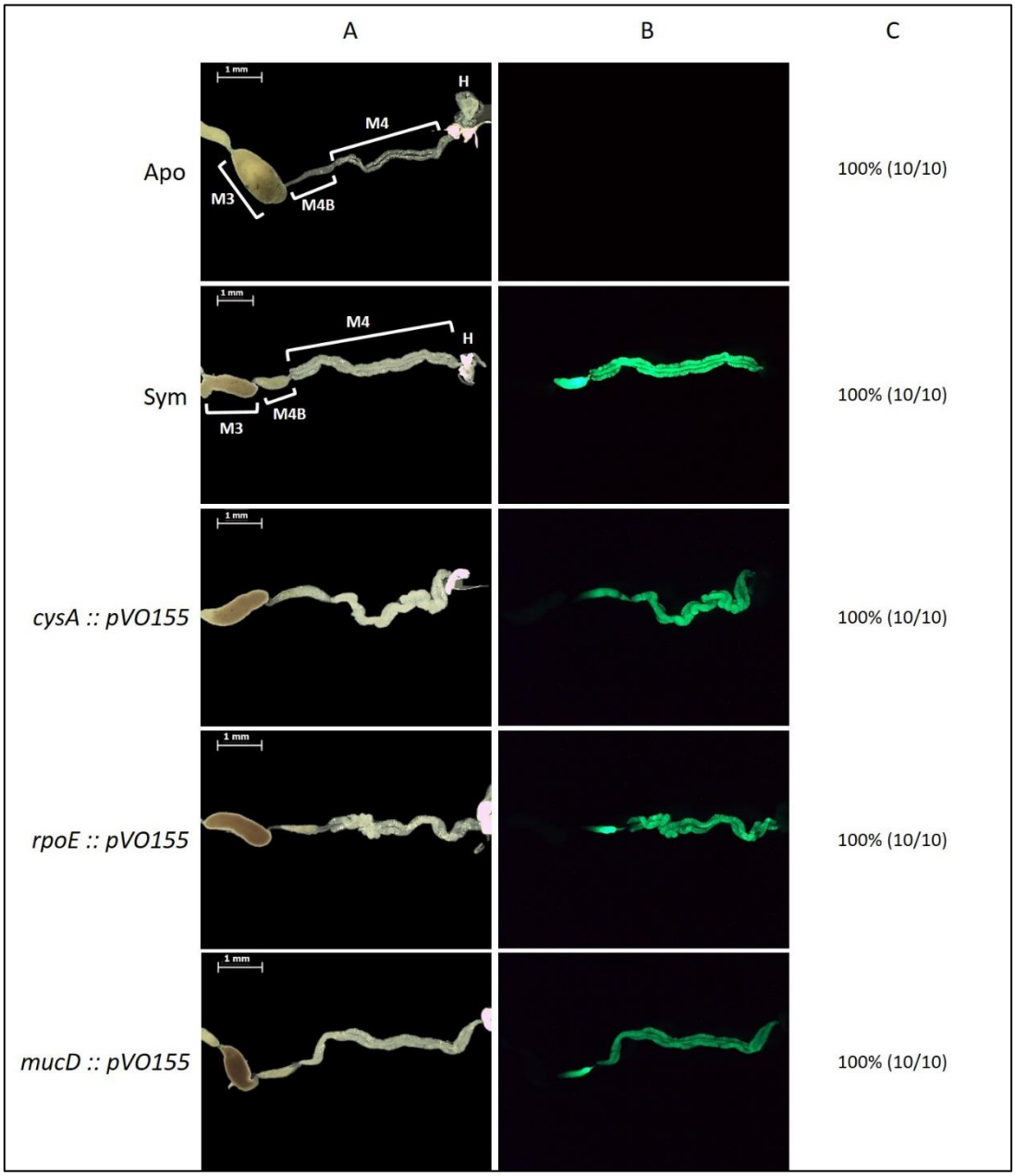


Figure 57: Midgut morphologies of third instar *R. pedestris* insects infected with ESR mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

and Valvano, 2008; Flores-Kim and Darwin, 2014). As the RpoE factor and MucD were reported to be involved in polymyxin B resistance in *B. cenocepacia* (Flannagan and Valvano, 2008), I have targeted the corresponding genes in *B. insecticola* for mutagenesis using pVO155 plasmid insertion. Similar to the previous mutant strains, I have tested the sensitivity of these

ESR mutants towards AMPs (**Figure 54**), membrane stressors (**Figure 55**) and also assessed their motility in YG soft agar plates (**Figure 41**). Concerning the AMP sensitivity, I observed that the *mucD* mutant had the same MIC_{90} as the wild-type strain for the six AMPs tested (**Figure 54**). The *rpoE* mutant had also similar sensitivity profiles compared to the wild-type strain for polymyxin B, LL-37, CCR179 and CCR480, but it was more sensitive to riptocin (MIC_{90} of 50 $\mu g.mL^{-1}$) than the wild-type strain (MIC_{90} of 200 $\mu g.mL^{-1}$) (**Figure 54**). Similar to the hopanoid mutants, I have also noticed that the *rpoE* and the *mucD* mutants were growing at the same rate as the wild-type strain. For the other membrane stressing conditions, the growth of the *mucD* mutant was similar to the growth of the wild-type strain (**Figure 55**). The *rpoE* mutant was more sensitive to SDS (MIC_{90} of 0.016%) compared to the wild-type strain (MIC_{90} of 0.031%), however the *rpoE* mutant grew at the same rate as the wild-type strain in the other conditions (**Figure 55**). Additionally, the swimming motility was not significantly different between these two ESR mutants and the wild-type strain (**Figure 41**). Thus, these results showed that the RpoE stress response is probably involved only in the riptocin resistance (with the hypersensitivity of the *rpoE* mutant), but is not required for the motility and the resistance towards the other AMPs tested.

I have subsequently checked the ability of these ESR mutant strains to colonize the host midgut, as previously mentioned (see **section 4.1.1**). Both at the second and the third instar stages, I noticed that the M4 region of the insects was successfully infected by the two ESR mutants (**Figures 56 and 57**). In addition, I have found that the *rpoE* and *mucD* mutants were able to proliferate in the M4 region to similar levels as the wild-type strain, both at the second and the third instar stages (**Figures 44 and 45**). These *in vivo* colonization parameters showed that the ESR mutant strains had equivalent colonization abilities than the wild-type strain to colonize and to maintain their symbiotic population inside the host midgut. So even if the RpoE stress response was shown to be involved in polymyxin B resistance in *B. cenocepacia* (Flannagan and Valvano, 2008), my results suggest that the RpoE factor in *B. insecticola* has a lesser importance in AMP resistance and also in host colonization during mono-infections (however, see **section 4.3** for co-infections with the wild-type strain).

In conclusion, the candidate gene approach confirmed that the LPS in *B. insecticola* is an important factor for AMP resistance and host colonization (see **section 4.1.1**), whereas the hopanoids and the RpoE stress response do not seem to participate in these processes or to a

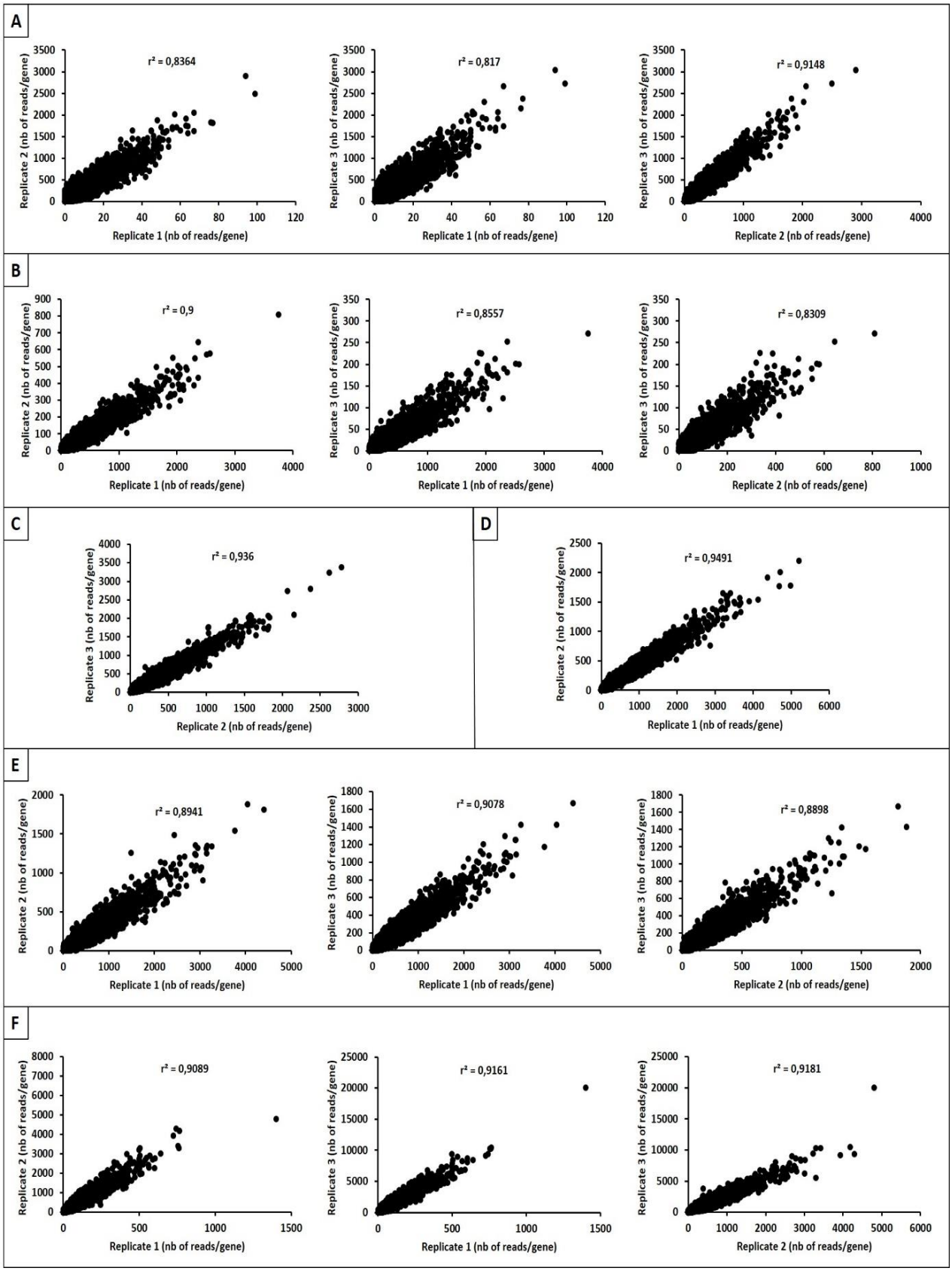
Sample	Nb of Post-Trim reads	% of Alignment against *Burkholderia insecticola* genome	Nb of Mapped reads
MM Glucose	4,909,911	68.9	3,381,676
Polymyxin B 1.5 µg.mL^{-1}	4,888,405	72.9	3,565,736
Polymyxin B 12.5 µg.mL^{-1}	3,884,958	68.4	2,656,595
LL-37 1.5 µg.mL^{-1}	5,409,490	67	3,623,348
LL-37 12.5 µg.mL^{-1}	6,265,036	79.5	4,981,739
Riptocin 100 µg.mL^{-1}	6,617,322	76	5,027,122
Riptocin 200 µg.mL^{-1}	13,305,508	83.6	11,127,900
CCR179 100 µg.mL^{-1}	6,772,599	76.7	5,192,840
CCR179 200 µg.mL^{-1}	8,572,555	73.6	6,306,647
CCR480 25 µg.mL^{-1}	7,014,452	76.5	5,367,970
CCR480 100 µg.mL^{-1}	9,347,705	73.6	6,884,647

Table 3: Sequencing results for AMPs Tn-seq conditions after pooling each replicate. The number of post-trim reads corresponds to the number of filtered reads after the trimming step. Abbreviations: Nb: number.

lesser extend (see **sections 4.1.2 and 4.1.3**). As hopanoids and the RpoE factor were shown to be important for *B. cenocepacia* and *B. multivorans* under stress conditions, these results suggest that these two membrane components may not be generally required for *Burkholderia* species or that their impact is only apparent under particular conditions.

4.2. Tn-seq approach

Based on previous studies on *Burkholderia* species, I could validate some bacterial factors involved in AMP resistance for *B. insecticola*, such as LPS (see **section 4.1**). In order to obtain a more global overview at the whole genome level of the bacterial functions involved in AMP resistance in *B. insecticola*, I have conducted a Tn-seq approach with the five different AMPs previously used for AMP sensitivity tests (see **section 4.1.1**): polymyxin B, LL-37, riptocin, and two CCR peptides, CCR179 and CCR480. As these AMPs are produced by different organisms, including *R. pedestris*, and have different structures with different activities on bacterial species, it would be possible to identify common and specific bacterial genes for AMP resistance. I tested two concentrations for each AMP, a low concentration that corresponds to the MIC_{90} of the AMP hypersensitive strains like the *waaC* mutant, and a higher concentration that is able to inhibit half of the growth of the wild-type strain. These two concentrations would discriminate fitness genes required for different selective pressure



intensities. As these AMPs are cationic peptides, I have used a defined nutrient medium for this Tn-seq experiment so that the activity of these peptides is not affected by the cationic

Figure 58: Correlations between read counts distribution in replicates of polymyxin B, LL-37 and riptocin Tn-seq conditions at both concentrations.
Dot plot representations of the comparison of each transposon insertions distribution between the Tn-seq replicates are shown for each AMP condition. The number of reads per gene is displayed for each replicate. The Pearson correlation coefficient r^2 was calculated for each comparison and indicated on each graph. A) Polymyxin B 1.5 µg.mL^{-1}. B) Polymyxin B 12.5 µg.mL^{-1}. C) LL-37 1.5 µg.mL^{-1}. D) LL-37 12.5 µg.mL^{-1}. E) Riptocin 100 µg.mL^{-1}. F) Riptocin 200 µg.mL^{-1}.

compounds present in the rich medium. Hence, I have performed the Tn-seq experiments in minimal medium (MM) with five AMPs at two concentrations in triplicates.

After the sequencing, each replicate of the ten AMP conditions contained at least one million filtered reads, with around 70 to 80% of these reads that aligned to the *B. insecticola* genome. However, one replicate of the LL-37 1.5 µg.mL^{-1} condition and one replicate of the LL-37 12.5 µg.mL^{-1} condition presented a low number of post-trim reads (< 400,000 reads). I discarded these replicates for further analysis. The calculated correlation coefficients between each replicate for each condition were high (r^2 > 0.81 for polymyxin B 1.5 µg.mL^{-1}, r^2 > 0.83 for polymyxin B 12.5 µg.mL^{-1}, r^2 > 0.93 for LL-37 1.5 µg.mL^{-1}, r^2 > 0.94 for LL-37 12.5 µg.mL^{-1}, r^2 > 0.88 for riptocin 100 µg.mL^{-1}, r^2 > 0.90 for riptocin 200 µg.mL^{-1}, r^2 > 0.93 for CCR179 100 µg.mL^{-1}, r^2 > 0.92 for CCR179 200 µg.mL^{-1}, r^2 > 0.86 for CCR480 25 µg.mL^{-1} and r^2 > 0.87 for CCR480 100 µg.mL^{-1}) (**Figures 58 and 59**). Therefore, I pooled the sequencing data together for each tested AMP condition. Thus, the pooled data obtained for each AMP condition contained at least two million reads mapped on the *B. insecticola* genome (**Table 3**), which constitutes a significant amount of reads for further analysis.

By Con-ARTIST analysis, I have found 19, 6, 12 and 17 conditionally-essential genes (Con-ARTIST essentiality score = 2, see **Chapter II**) for the lower concentrations of polymyxin B, riptocin, CCR179 and CCR480 peptides, respectively (see **Annexes 6, 7, 8, 9 and 10**). Unfortunately, I was not able to detect conditionally-essential genes for LL-37, probably due to the concentration of LL-37 that was not sufficient to observe a significant difference between the treated and the untreated bacteria. In addition, I found 5, 4, 6 and 1 domain-conditionally essential genes (Con-ARTIST essentiality score = 1, see **Chapter II**) for polymyxin B, riptocin, CCR179 and CCR480 peptides, respectively (see **Annexes 6, 7, 8, 9 and 10**). Concerning the high peptide concentration conditions, I identified 32, 31, 6, 16 and 32 conditionally-essential genes for the polymyxin B, LL-37, riptocin, CCR179 and CCR480

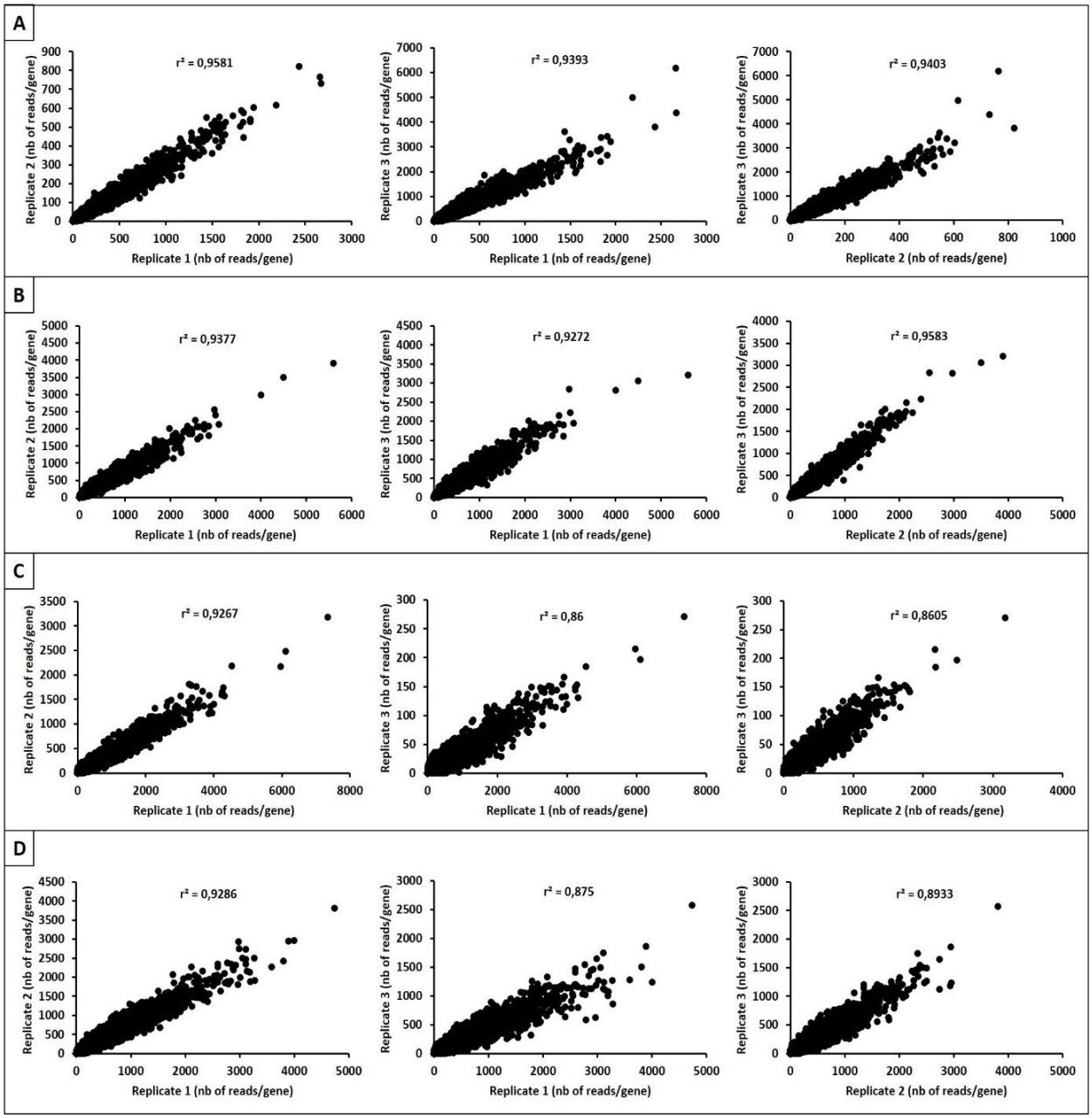


Figure 59: Correlations between read counts distribution in replicates of CCR179 and CCR480 Tn-seq conditions at both concentrations.

Dot plot representations of the comparison of each transposon insertions distribution between the Tn-seq replicates are shown for each AMP condition. The number of reads per gene is displayed for each replicate. The Pearson correlation coefficient r^2 was calculated for each comparison and indicated on each graph. A) CCR179 100 µg.mL^{-1}. B) CCR179 200 µg.mL^{-1}. C) CCR480 25 µg.mL^{-1}. D) CCR480 100 µg.mL^{-1}.

peptides, respectively (see **Annexes 6, 7, 8, 9 and 10**). Additionally, I also identified 10, 11, 9, 5 and 7 domain-conditionally essential genes for polymyxin B, LL-37, riptocin, CCR179 and CCR480 peptides, respectively (see **Annexes 6, 7, 8, 9 and 10**). These sets of fitness genes required for AMP resistance increased from lower to higher concentrations, thus confirming

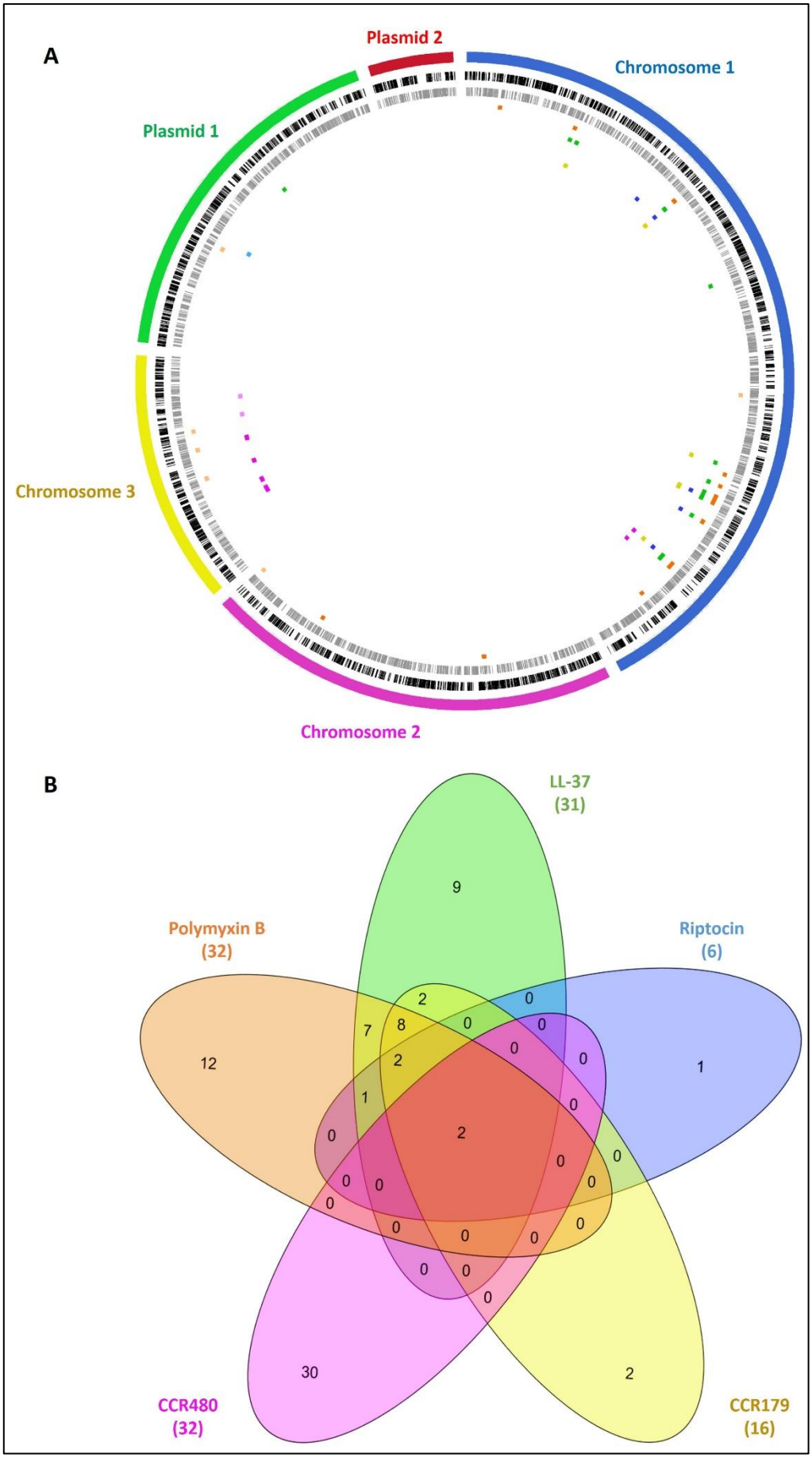

A
Plasmid 2
Chromosome 1
Plasmid 1
Chromosome 3
Chromosome 2
B
LL-37
(31)
Polymyxin B
(32)
Riptocin
(6)
CCR480
(32)
CCR179
(16)
9
12
1
30
2
7
8
2
2
1
2
0

Figure 60: Fitness genes identified by Con-ARTIST analysis in *B. insecticola* for AMP resistance.

A) Circular representation of the *B. insecticola* genome consisting of chromosomes 1, 2 and 3 and plasmids 1 and 2. From outer to inner rings: forward CDS (black bars), reverse CDS (grey bars), conditionally-essential genes (Con-ARTIST essentiality score = 2) for polymyxin B (orange dots), conditionally-essential genes for LL-37 (green dots), conditionally-essential genes for riptocin (blue dots), conditionally-essential genes for CCR179 peptide (yellow dots), conditionally-essential genes for CCR480 peptide (pink dots). The light coloured dots indicate fitness genes identified at low concentrations and the dark coloured dots indicate fitness genes identified at high concentrations. B) Distribution of conditionally-essential genes between the five AMP conditions (Venn diagram). The total number of conditionally-essential genes is indicated for each condition in parentheses.

that the increased selective pressure of the AMP treatments had an effect on the survival of the bacterial mutant population. Reassuringly, the majority of the AMP fitness genes identified at lower concentrations were also detected in the fitness gene sets identified for higher concentration conditions, thus indicating a linked conservation of these genes required for AMP resistance at different concentrations (see **Annexes 6, 7, 8, 9 and 10**).

On the conditionally-essential genes for each AMP condition, I observed that these genes are mostly located to the chromosome 1 of the *B. insecticola* genome (**Figure 60A**, see **Annexes 6, 7, 8, 9 and 10**). However, there was an exception for the CCR480 peptide-required genes that were mainly located on the chromosome 3 (**Figure 60A**, see **Annexe 10**). Moreover, some fitness genes required for polymyxin B resistance were located on the chromosomes 2 and 3, and also in the plasmid 1 (**Figure 60A**). Interestingly, none of these bacterial fitness genes required for AMP resistance was located on the plasmid 2 (**Figure 60A**).

As most of the fitness genes in the lower concentration conditions were also found in the higher concentration conditions, I compared the gene sets required for high AMP concentrations (**Figure 60B**). This comparison showed that multiple bacterial factors are shared between the polymyxin B, LL-37, riptocin and CCR179 peptide conditions (**Figure 60B**). Interestingly, only two genes were shared between the five AMP conditions (**Figure 60B**), the *tatA* (*BRPE64_RS12020*) and *tatB* (*BRPE64_RS12015*) genes that are encoding for two subunits of the Tat transporter. Except these two genes, the other fitness genes identified for CCR480 peptide condition are specifically required only for the CCR480 resistance (**Figure 60B**).

The COG categories (Tatusov *et al.*, 2000) of these fitness genes required for AMP resistance showed that the majority of these genes belong to the cell wall biogenesis category (M category) and the intracellular trafficking (U category) (**Figure 61**). Concerning the specific

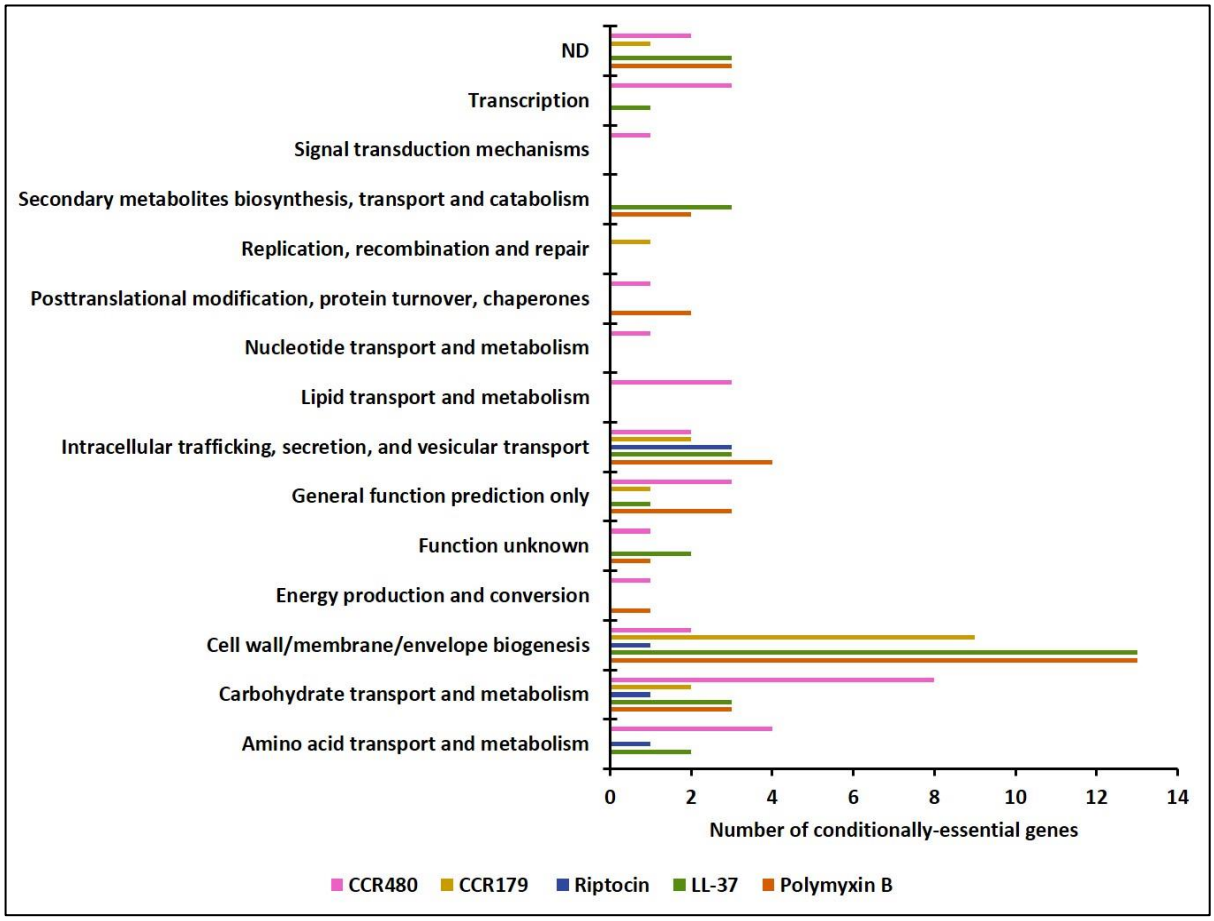


Figure 61: COG categories of bacterial fitness genes involved in AMP resistance.
The numbers of conditionally-essential genes identified for each AMP condition, polymyxin B (orange), LL-37 (green), riptocin (blue), CCR179 (yellow) and CCR480 (pink) are displayed for each COG category. ND: not determined.

genes required for CCR480 peptide, they were associated to diverse metabolic activities which mostly involved carbohydrate exploitation functions (G category) (**Figure 61**). Based on these results, I will focus first on the description of the fitness genes that were commonly found for different AMPs, and then, I also describe the specific bacterial factors identified for each AMP.

4.2.1. Bacterial functions involved in global AMP resistance

A striking result of the Tn-seq analysis was the identification of the *tatABC* gene cluster (*BRPE64_RS12010-BRPE64_RS12020*) as required for resistance to all tested AMPs (**Figure 62).** Interestingly, this gene cluster is also strictly essential for the bacterial growth in YG rich medium, but not in the MM medium (**Figure 62**). In previous transposon mutagenesis studies,

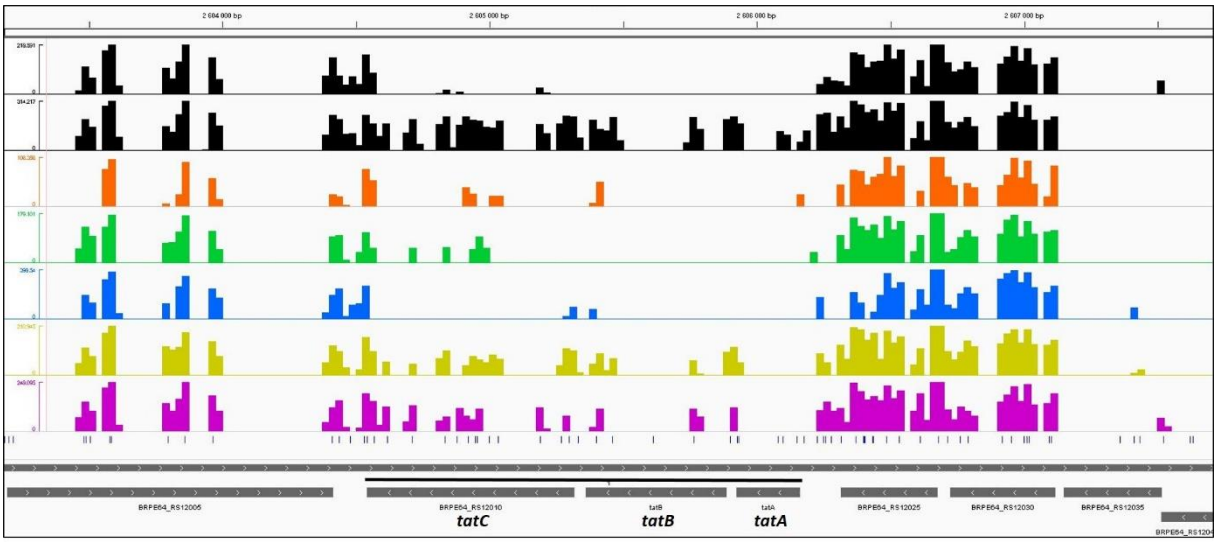


Figure 62: Fitness genes required for all AMPs resistance.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes are outlined by a black line which correspond to *tatA* (*BRPE64_RS12010*), *tatB* (*BRPE64_RS12015*) and *tatC* (*BRPE64_RS12020*).

it was reported that the *tatABC* genes of *Pseudomonas aeruginosa* and *Acinetobacter baylyi* were also essential genes for the bacterial viability in rich medium (de Berardinis *et al.*, 2008; Liberati *et al.*, 2006). Thus, the Tn-seq results strongly suggest that these three genes are not essential in the MM condition, but they are required for AMPs resistance in the same medium (**Figure 62**). The *tatABC* gene cluster encodes the twin-arginine translocation (Tat) transporting system, belonging to the intracellular trafficking functional COG category (U category). The Tat system is known to export across the cytoplasmic membrane large folded proteins that contain a specific consensus sequence S/T-R-R-x-F-L-K, also called the Tat signal motif (Robinson *et al.*, 2011a). In *E. coli*, it was previously described that *tat* mutants showed an impaired cell division and exerted an atypical chain-forming morphological shape (Ize *et al.*, 2003), whereas in *V. cholerae*, the Tat transporter was shown to participate in flagellar motility, biofilm formation and host colonization (Zhang *et al.*, 2009). Thus, the Tat system seems strongly required for the outer membrane stability of Gram-negative species. As the main target of AMPs are the bacterial membranes, the requirement of the Tat transporter seems consistent with these previous studies. To identify the putative proteins exported by the Tat transporter in *B. insecticola*, I conducted an *in silico* analysis with the dedicated prediction algorithm TATFIND (http://signalfind.org/tatfind.html), which searches the specific

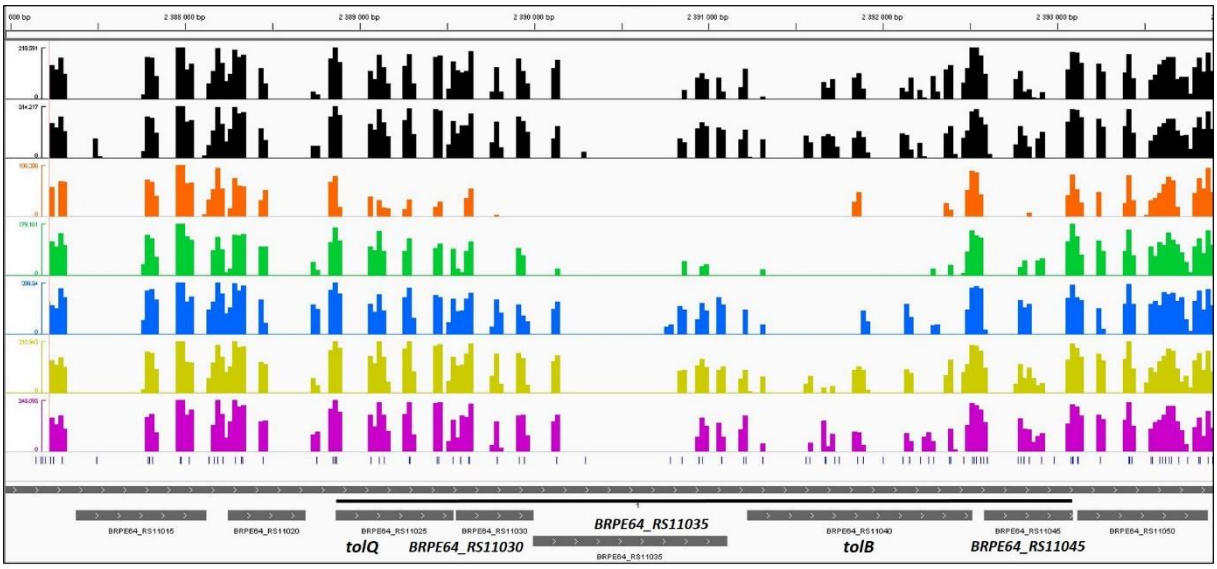


Figure 63: The Tol-Pal complex is required for AMPs resistance.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes are outlined by a black line which correspond to *tolQ* (*BRPE64_RS11025*), *BRPE64_RS11030*, *BRPE64_RS11035*, *tolB* (*BRPE64_RS11040*) and *BRPE64_RS11045*.

Tat signal motif in the first 35 amino acids of each protein (Dilks *et al.*, 2003; Rose *et al.*, 2002). In a previous study, 72 putative Tat substrates were identified in *P. aeruginosa* by this algorithm (Gimenez *et al.*, 2018). In *B. insecticola*, this analysis revealed that 64 proteins contained the Tat signal motif in their sequence and are therefore potentially exported by the Tat transporter (see **Annexe 11**). Among these genes, there is the *BRPE64_RS10880* gene which corresponds to the *amiC* gene that encodes the N-acetylmuramoyl-L-alanine amidase (see **Annexe 11**), and two genes *BRPE64_RS15055* and *BRPE64_RS23505* which encode β-lactamases (see **Annexe 11**), which were experimentally characterized as Tat substrates in *Mycobacterium smegmatis* (McDonough *et al.*, 2005) and *E. coli* (Ize *et al.*, 2003). In addition, there are genes encoding lipoproteins (*BRPE64_RS01050*, *BRPE64_RS06735*, *BRPE64_RS11155*), genes encoding extracellular solute-binding proteins (*BRPE64_RS23420*, *BRPE64_RS23700*, *BRPE64_RS23840*, *BRPE64_RS30900*), a gene that encodes a LPS-assembly protein (*BRPE64_RS11095*) and genes that encode ABC-type periplasmic proteins or substrate-binding proteins (*BRPE64_RS10210*, *BRPE64_RS20855*, *BRPE64_RS23725*, *BRPE64_RS24150*, *BRPE64_RS26230*, *BRPE64_RS27260*) (see **Annexe 11**). All these predicted

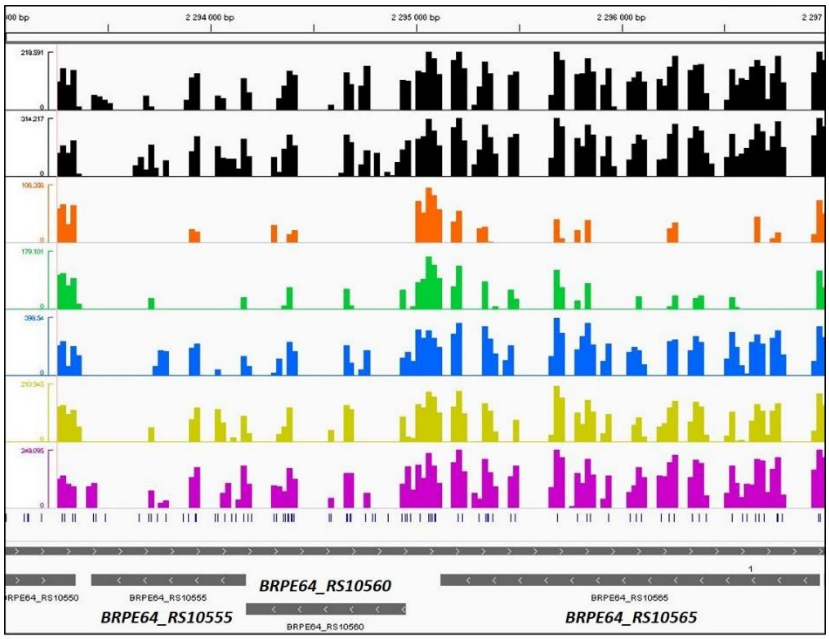


Figure 64: The Wzm/Wzt O-antigen transporter is required for AMPs resistance. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes are outlined by a black line which correspond to *BRPE64_RS10555*, *BRPE64_RS10560* for the Wzm/Wzt transporter and *BRPE64_RS11045* for a glycosyl transferase.

Tat substrates indicate that the Tat transporter is effectively contributing to the outer membrane stability. However, none of the genes encoding these Tat substrates was identified in the Tn-seq with AMPs, making it at present unclear how the transporter specifically contributes to AMPs resistance.

Another transporting system that was identified as a bacterial fitness determinant for riptocin, LL-37 and polymyxin B resistance was the Tol-Pal complex (**Figure 63**). This transporter consists of the five subunits TolQ, TolR, TolA, TolB and Pal that are encoded by the gene cluster *BRPE64_RS11025-BRPE64_RS11045* in *B. insecticola*. In this *tol-pal* gene cluster, only the *tolB* and the *pal* genes were required for riptocin resistance, whereas the complete gene cluster was important for the bacterial fitness in the presence of polymyxin B (**Figure 63**). For LL-37 resistance, almost all the genes of this cluster were also identified as conditionally-essential except for the *tolQ* gene (**Figure 63**). The Tol-Pal complex was reported to participate in the outer membrane permeability in Gram-negative species by transporting outer membrane proteins (Lazzaroni *et al.*, 1999; Lloubès *et al.*, 2001; Yeh *et al.*, 2010). In *E. coli*, it was also

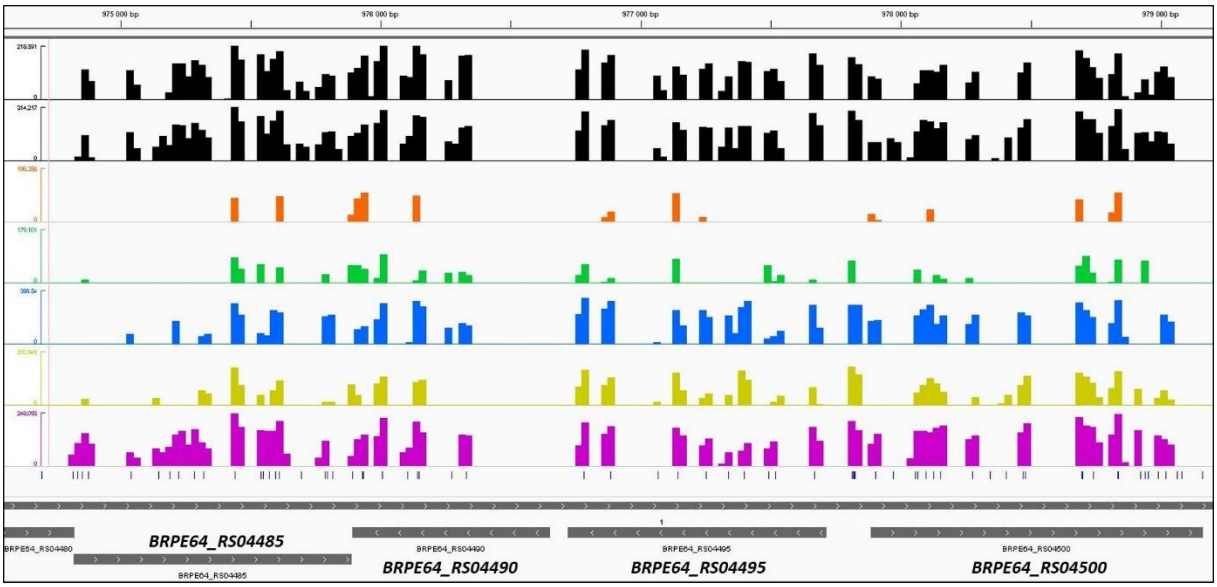


Figure 65: O-antigen biosynthesis genes are required for AMPs resistance.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes are indicated with *BRPE64_RS04485*, *BRPE64_RS04490, BRPE64_RS04495* and *BRPE64_RS04500*.

demonstrated that the Tol-Pal complex is involved in the cell division machinery (Gerding *et al.*, 2007), but also in the import of group A colicins and filamentous bacteriophages (Lazzaroni *et al.*, 2002; Lloubès *et al.*, 2001; Webster, 1991). As the Tol-Pal complex seems to play a role in the outer membrane permeability, this transporter may be strongly required to face membrane damages caused by AMPs.

In addition to these two transporting systems, I found two fitness genes for polymyxin B, LL-37, riptocin and CCR179 peptide conditions, *BRPE64_RS10555* and *BRPE64_RS10560*, that were homologous to the *wzt* and *wzm* genes from *B. cenocepacia*, respectively (with 42.5% and 27.03% of identity, respectively) (**Figure 64**) (Ortega *et al.*, 2005). These two genes encode the Wzm/Wzt O-antigen transporting system which is an ABC transporter located in the inner membrane that exports the O-antigen chain of LPS from the cytoplasm to the periplasm (Greenfield and Whitfield, 2012; Raetz and Whitfield, 2002). This O-antigen transporter was shown to be present in *Klebsiella pneumonia* (Raetz and Whitfield, 2002), *P. aeruginosa* (Rocchetta and Lam, 1997), *E. coli* (Greenfield and Whitfield, 2012) and also in *Burkholderia* species (Ortega *et al.*, 2005; Yuen *et al.*, 2012). In addition to the O-antigen transport, it was shown that this transporter is involved in biofilm formation and participates to the bacterial

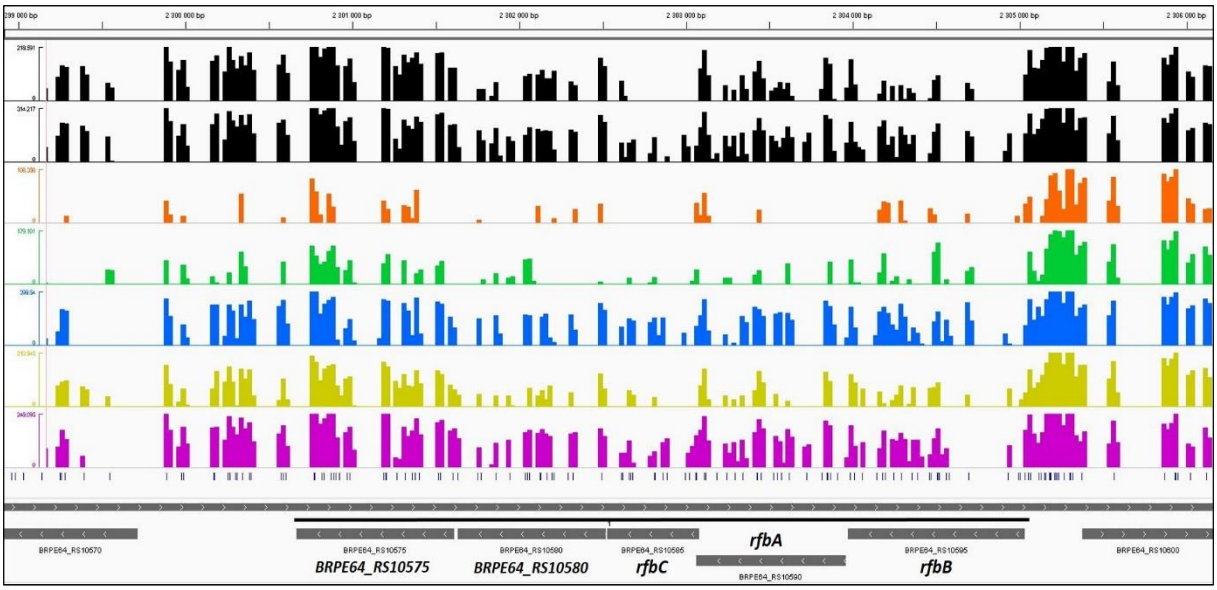


Figure 66: dTDP-L-rhamnose biosynthesis genes are required for AMPs resistance. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes are outlined by a black line which correspond to *BRPE64_RS10575*, *BRPE64_RS10580, rfbC* (*BRPE64_RS10585*), *rfbA* (*BRPE64_RS10590*) and *rfbB* (*BRPE64_RS10595*).

membrane integrity in *B. pseudomallei* (Ortega *et al.*, 2005). Moreover, I identified in the polymyxin B, LL-37, riptocin and CCR179 peptide conditions a gene cluster (*BRPE64_RS04485-BRPE64_RS04500*) that is involved in O-antigen biosynthesis (**Figure 65**). From this cluster, the *BRPE64_RS04485*, *BRPE64_RS04490* and *BRPE64_RS04495* genes encode glycosyl transferases, and the last gene *BRPE64_RS04500* encodes the O-antigen polymerase. In addition, there was another glycosyl transferase (encoded by *BRPE64_RS10565*) also involved in the O-antigen biosynthesis that was found as a fitness determinant for polymyxin B, LL-37 and CC179 peptide resistance (**Figure 64**). Also linked to O-antigen biosynthesis, I identified the gene cluster (*BRPE64_RS10575-BRPE64_RS10595*) responsible for the biosynthesis of dTDP-L-rhamnose in the polymyxin B, LL-37 and CCR179 peptide conditions (**Figure 66**). In some Gram-negative bacteria, dTDP-L-rhamnose is the precursor of rhamnose moieties in the O-antigen of LPS molecules (Tsukioka *et al*., 1997; Vinion-Dubiel and Goldberg, 2003). In this gene cluster, there are four genes (*BRPE64_RS10580*, *rfbC* or *BRPE64_RS10585*, *rfbA* or *BRPE64_RS10590*, and *rfbB* or *BRPE64_RS10595*) that encode the four enzymes required to synthesize dTDP-L-rhamnose from glucose-1-phosphate (Tsukioka *et al*., 1997), and there is

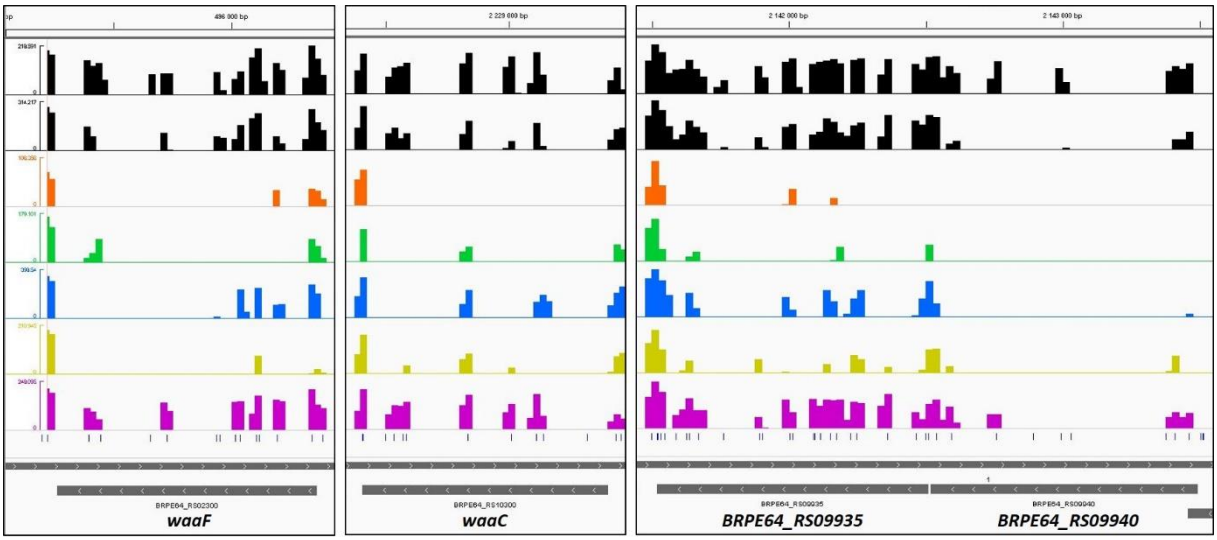


Figure 67: Core oligosaccharide biosynthesis genes are required for AMPs resistance. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. From left to the right: fitness genes are indicated with *waaF (BRPE64_RS02300)*, *waaC* (*BRPE64_RS10300*), *BRPE64_RS09935* and *BRPE64_RS09940*.

an additional gene, *BRPE64_RS10575*, which encodes a rhamnosyltransferase that transfers L-rhamnose from the dTDP-L-rhamnose donor to the O-antigen (Steiner *et al.*, 2010). As mentioned before, these genes are present in Gram-negative bacteria and are required for O-antigen biosynthesis, but are also conserved in Gram-positive bacteria for the L-rhamnosylation of cell wall teichoic acids (Carvalho *et al.*, 2015). Interestingly, it was shown that the deletion of the complete *rfb* gene cluster (also named *rml* gene cluster for some bacterial species) in *Listeria monocytogenes* led to an increased sensitivity towards AMPs such as LL-37 (Carvalho *et al.*, 2015).

In addition to the O-antigen, the biosynthetic pathway of the core oligosaccharide in LPS was also identified as required for AMPs resistance. In this Tn-seq analysis, I found four fitness genes involved in the inner core oligosaccharide biosynthesis including the *waaF* (*BRPE64_RS02300*) and *BRPE64_RS09935* (homologue of the *rfaD* gene from *Paraburkholderia xenovorans*, with 89.09% of identity) genes that were identified in the polymyxin B, LL-37 and CCR179 peptide conditions (**Figure 67**) (de Kievit and Lam, 1997; Kneidinger *et al.*, 2002); and the *waaC* (*BRPE64_RS10300*) and *BRPE64_RS09940* (homologue of the *rfaE* gene from *Paraburkholderia fungorum*, with 82.7% of identity) genes which were

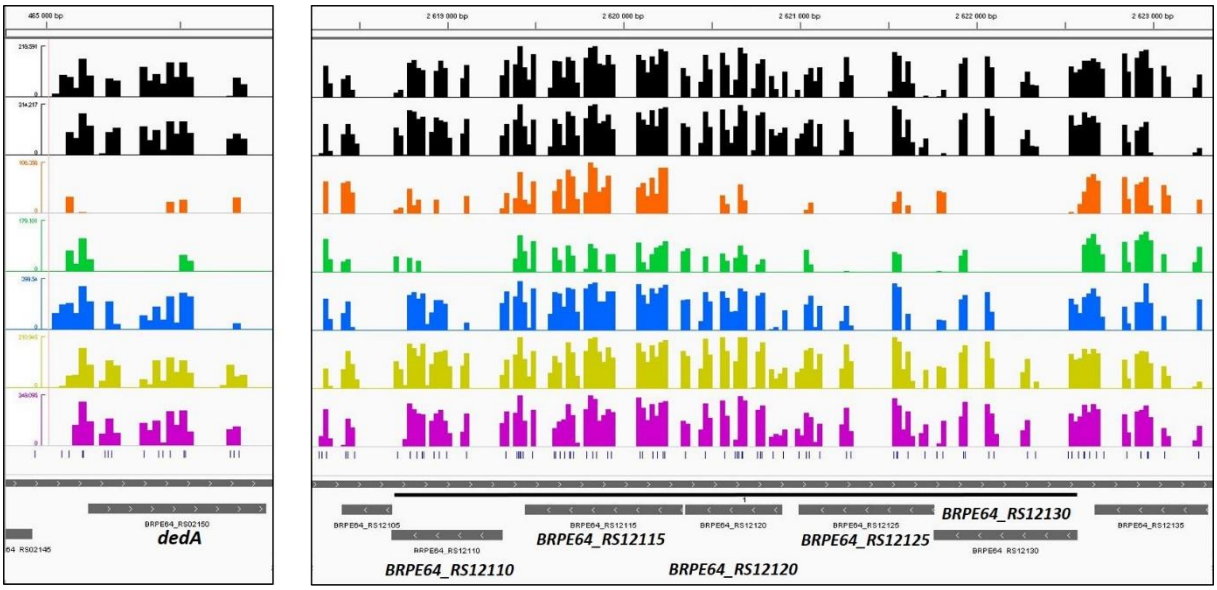


Figure 68: The DedA protein and Mla proteins are required for polymyxin B and LL-37 resistance.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Left picture: *dedA (BRPE64_RS02150)*. Right picture: fitness genes are outlined by a black line with *BRPE64_RS12110, BRPE64_RS12115, BRPE64_RS12120, BRPE64_RS12125* and *BRPE64_RS12130*.

required for polymyxin B and LL-37 resistance (**Figure 67**) (Izquierdo *et al.*, 2002; de Kievit and Lam, 1997). Interestingly, I found both the *waaC* and *waaF* genes for which I have previously studied their role in AMPs resistance in the candidate gene approach (see **section 4.1.1**). Moreover, I found four other fitness genes organized in a cluster in polymyxin B and LL-37 conditions, from *BRPE64_RS10475* to *BRPE64_RS10490*, that are responsible for the outer core oligosaccharide biosynthesis in *Burkholderia* species (Ortega *et al*., 2009; Vinion-Dubiel and Goldberg, 2003). Among these bacterial fitness genes, I identified the *wbiF* or *BRPE64_RS10490* gene that was previously shown to participate in polymyxin B resistance in the candidate gene approach (see **section 4.1.1**).

Concerning other cell wall components, I found an interesting gene (*BRPE64_RS05760*) for riptocin, LL-37 and polymyxin B treatments that corresponds to the *rseP* gene from *B. multivorans* (72.39% of identity). This gene encodes a membrane-associated zinc metalloprotease that is known to participate in the proteolytic cleavage of RseA, which is part of the σ^E ESR pathway (see **section 4.1.3**) (Li *et al.*, 2009). Despite the identification of the *rseP*

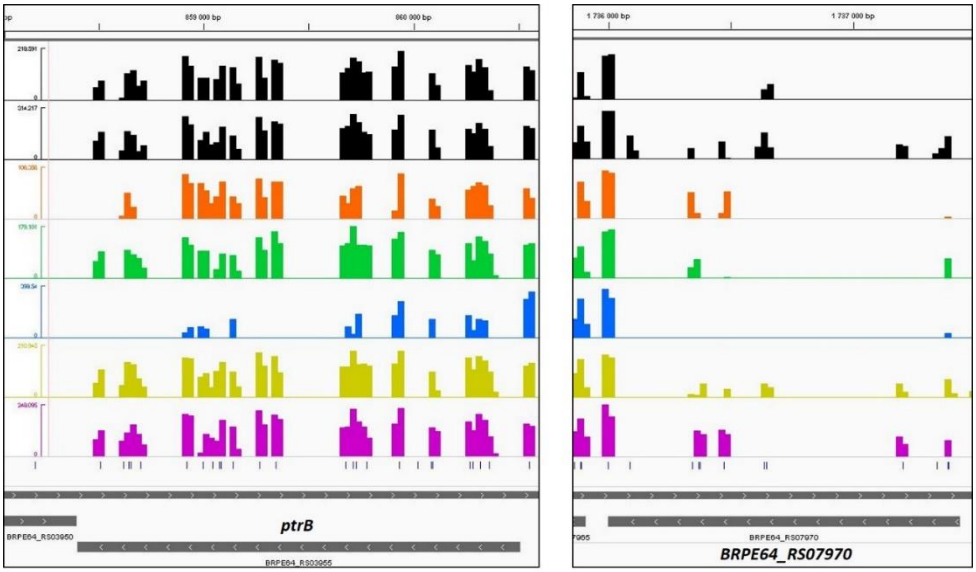


Figure 69: Specific fitness genes required for riptocin resistance.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Left picture: *ptrB* gene (*BRPE64_RS03955*). Right picture: *BRPE64_RS07970* gene.

gene, I did not find any of the other genes involved in the RpoE ESR pathway in the AMPs conditions. Even if the RseP protease seems to be specifically recruited for the RpoE stress response, it was shown that this protease had the ability to cleave a broad range of membrane proteins which might explain the non-essentiality of the σ^E gene cluster for AMPs resistance (Akiyama *et al.*, 2004).

Specifically involved in polymyxin B and LL-37 resistance, I identified four fitness genes including *dedA* (*BRPE64_RS02150*) that encodes a specific membrane protein, and three other genes (*BRPE64_RS12120*, *BRPE64_RS12125* and *BRPE64_RS12130*) that encode the Mla phospholipid transport system (**Figure 68**). For the *dedA* gene, the function of its encoded membrane protein is poorly understood, but members of the DedA membrane protein family were reported to be involved in temperature sensitivity, the cell division process and the regulation of the membrane composition in some Gram-negative species (Doerrler *et al.*, 2013). Interestingly, DedA proteins were shown to be involved in cationic AMPs resistance in *Salmonella enterica* and *Neisseria meningitidis* (Shi *et al.*, 2004; Tzeng and Stephens, 2015), and may act as proton-dependent transporters as these proteins are closely related to the LeuT superfamily of amino acid transporters (Doerrler *et al.*, 2013; Kumar and Doerrler, 2014).

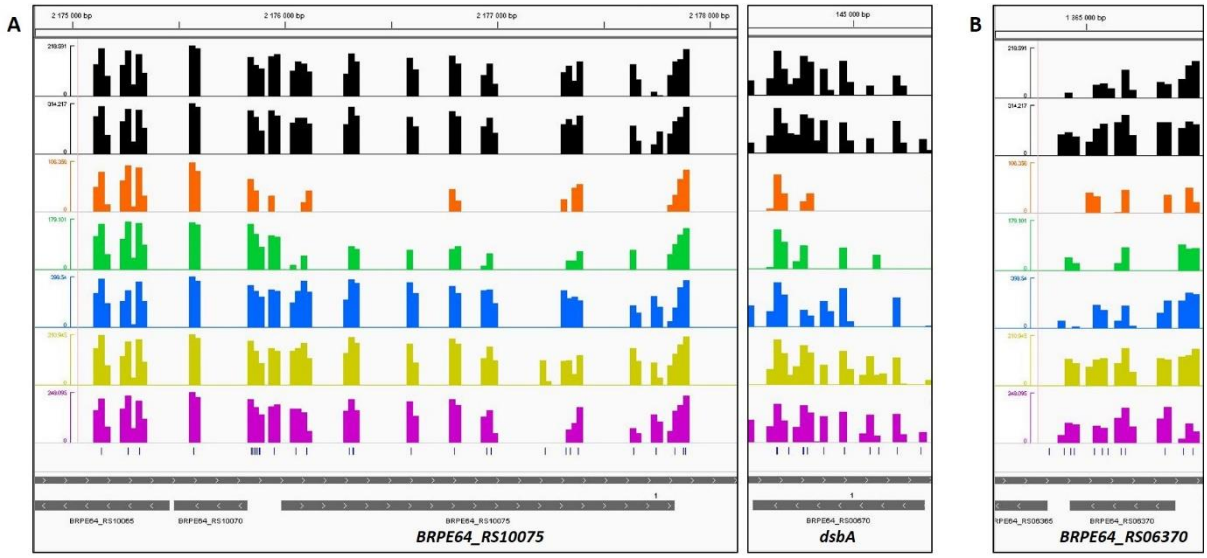


Figure 70: Specific fitness genes required for polymyxin B and LL-37 resistance.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. A) Polymyxin B specific genes with *BRPE64_RS10075* (left) and *dsbA (BRPE64_RS00670)* (right) genes. B) LL-37 specific gene with *BRPE64_RS06370*.

As mentioned above, I also identified three genes in a small cluster, from *BRPE64_RS12120* to *BRPE64_RS12130*, which are annotated as toluene tolerance Ttg2C-like proteins. In *Pseudomonas putida*, the *ttg2* gene cluster encodes an ABC transporter that might be involved in toluene export in toluene-tolerant bacteria (Kim *et al.*, 1998). Additionally, I found that *BRPE64_RS12120*, *BRPE64_RS12125* and *BRPE64_RS12130* genes are respectively homologous to the *yrbD*, *yrbE* and *yrbF* genes from *E. coli*, which encode an ABC transporter known as MlaDEF, regulating the outer membrane lipid asymmetry known (Thong *et al.*, 2016). Interestingly in *B. cenocepacia*, mutants in the Mla pathway showed an increased sensitivity towards antimicrobial agents, a reduced motility and an impaired colony morphology compared to the wild-type strain (Bernier *et al*., 2018).

Taken together, it appears that the common targets of AMPs in *B. insecticola* identified here are membrane components (see the recapitulative functions illustrated in **Figure 72**). On the one hand, the genes that are involved in the biosynthesis of the O-antigen and the core oligosaccharide of the LPS molecules are key components of the genetic repertoire of the strain allowing it to resist to AMPs. On the other hand, several transporters including the Tol-

Pal, TatABC, DedA and Mla are also key contributors and their common feature is that they have an impact on the composition and the stability of the bacterial membranes.

4.2.2. Resistance factors for specific AMPs

For riptocin resistance, the only specific gene identified by Tn-seq was *ptrB* (*BRPE64_RS03955*) that encodes a serine-type prolyl endopeptidase, also known as oligopeptidase B (**Figure 69**). This peptidase is able to hydrolyse peptide bonds after lysine and arginine residues in short protein sequences, and has been described to inactivate proline-rich AMPs (Mattiuzzo *et al.*, 2014; Morty *et al.*, 2002). As riptocin was the only proline-rich AMP tested (16.67% of the protein sequence), this might explain why I did not find the *ptrB* gene in the fitness gene sets for the other AMPs. Additionally, I found two domain-essential genes (*BRPE64_RS07970*, *BRPE64_RS07975*) involved in the first steps of the tricarboxylic acid (TCA) cycle that are responsible for the hydrolysis of pyruvate into acetyl-CoA (**Figure 69**) (de Kok *et al.*, 1998). However, as these genes were identified as domain-essential genes, they might play a minor

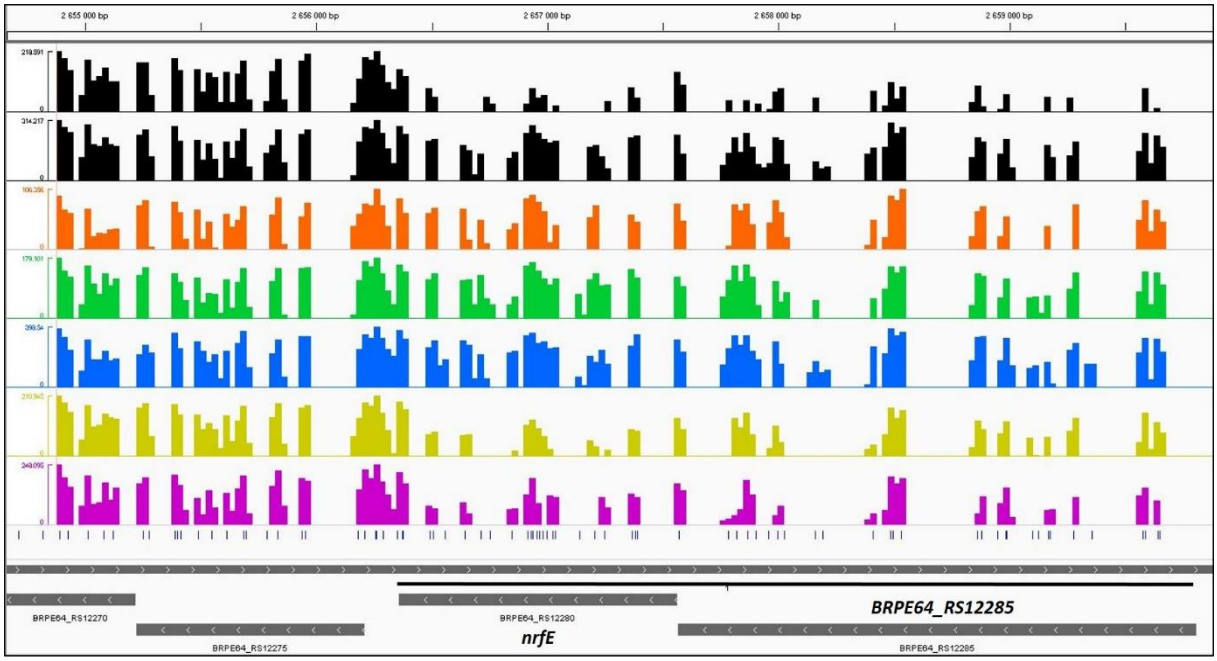


Figure 71: Specific domain-essential genes required for CCR peptides resistance. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), polymyxin B (orange bars), LL-37 (green bars), riptocin (blue bars), CCR179 (yellow bars) and CCR480 (pink bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes that are specific for CCR peptides are outlined by a black line which correspond to *nrfE* (*BRPE64_RS12280*) and *BRPE64_RS12285*.

role in riptocin resistance or the riptocin treatment may have a little impact on the energy production of the symbiont (see the recapitulative functions illustrated in **Figure 72**).

BRPE64_RS08480 was specifically required for polymyxin B resistance. This gene codes for a NLP/P60 protein, a papain-like protease involved in cell wall remodelling (Xu *et al.*, 2015). Also specific to polymyxin B resistance, there were the *dsbA* gene (*BRPE64_RS00670*) that is involved in protein stabilization (Manta *et al.*, 2019; Meehan *et al.*, 2017), and two genes (*BRPE64_RS16130*, *BRPE64_RS16135*) encoding heat shock proteins from the Hsp20 family that are chaperones, also involved in protein quality control (**Figure 70A**) (Mercer *et al.*, 2017). Moreover, I identified the *BRPE64_RS10075* gene that encodes a protein containing a TPR (tetratricopeptide repeat) motif, known to mediate protein-protein interactions (Blatch and Lässle, 1999; D'Andrea and Regan, 2003). This specific TPR motif is found in several proteins that participate in diverse cellular processes, such as cell cycle control, protein quality control, protein export and transcription (Blatch and Lässle, 1999). In addition to these polymyxin B resistance genes, there was the *BRPE64_RS19345* gene that encodes an outer membrane protein OmpC type which is participating in the cell wall maintenance (Wang *et al.*, 2007). Hence, these specific fitness genes suggest that polymyxin B targets the cell wall of *B. insecticola*, but also affects the protein quality control.

Concerning LL-37 resistance genes, I identified one gene (*BRPE64_RS12115*) that encodes the lipoprotein VacJ, which is localized in the Mla-encoding cluster (*BRPE64_RS12110-BRPE64_RS12130*) and was reported to play a role in the maintenance of lipid asymmetry in Gram-negative bacteria (**Figure 68**) (Malinverni and Silhavy, 2009; Suzuki *et al*., 1994). In addition, I also found the *BRPE64_RS06370* gene that encodes an exporter which belongs to the drug/metabolite transporter (DMT) superfamily, which could be involved in the export of LL-37 out of the bacterial cell (**Figure 70B**) (Jack *et al*., 2001) (see the recapitulative functions illustrated in **Figure 72**).

For CCR peptide resistance, I have found two domain-essential genes, *nrfE* (*BRPE64_RS12280*) and *BRPE64_RS12285* that were required for both the CCR179 and the CCR480 peptides. These two genes encode proteins involved in cytochrome c biogenesis, which are part of the aerobic respiratory chain to produce energy for the bacterial cell (**Figure 71**) (Ahuja *et al.*, 2009; Le Brun *et al.*, 2000). Regarding CCR179, I found one fitness gene (*BRPE64_RS10515*) that encodes a protein containing a MBOAT (membrane-bound O-acyltransferase) domain

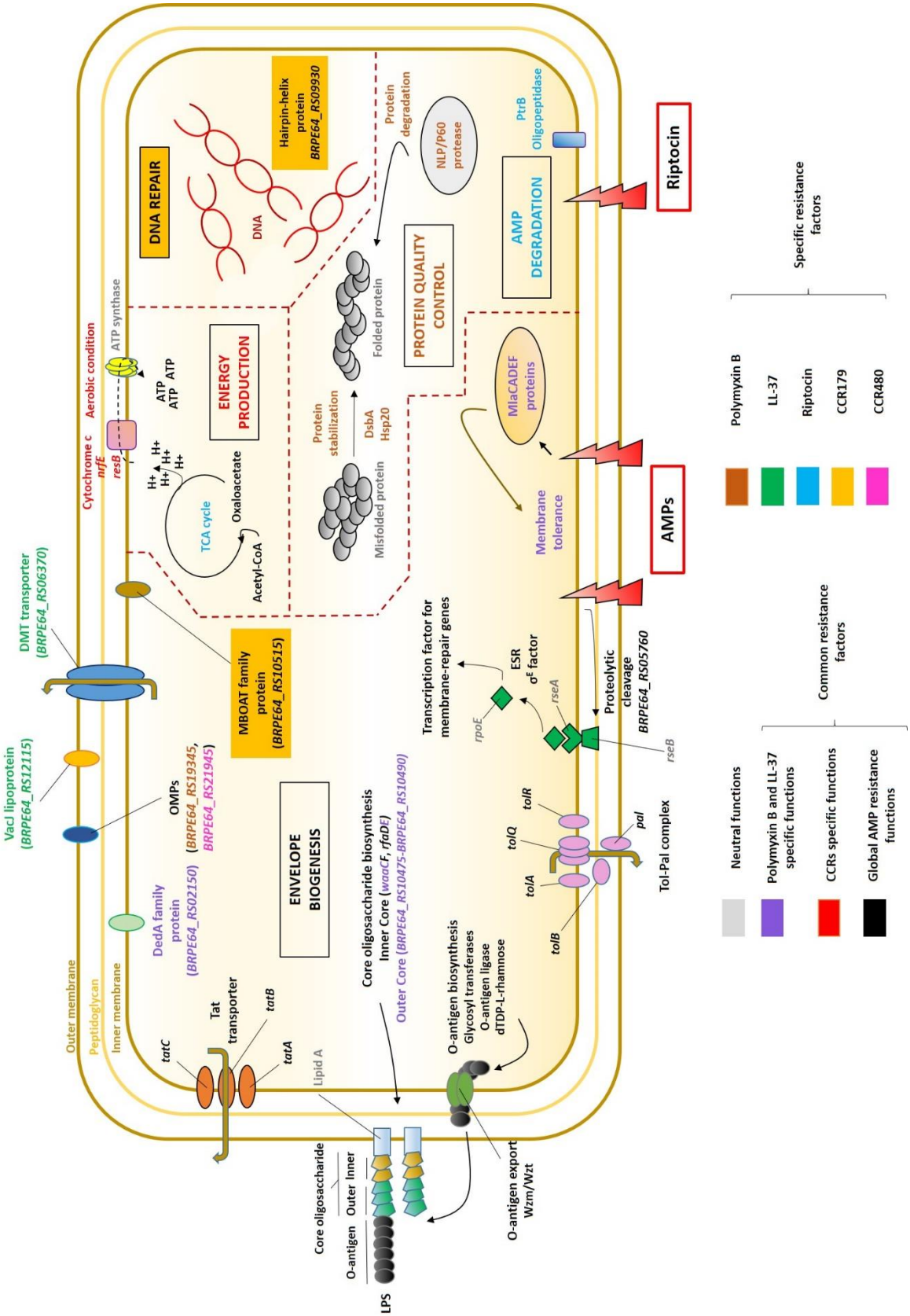

Outer membrane
Peptidoglycan
Inner membrane
Cytochrome c
nrfE
resB
Aerobic condition
ATP synthase
ATP
ATP
ATP
H+
H+
H+
H+
TCA cycle
Oxaloacetate
Acetyl-CoA
ENERGY PRODUCTION
DNA REPAIR
DNA
Hairpin-helix protein BRPE64_RS09930
Protein degradation
NLP/P60 protease
PtrB Oligopeptidase
AMP DEGRADATION
Riptocin
PROTEIN QUALITY CONTROL
Folded protein
Protein stabilization
DsbA
Hsp20
Misfolded protein
MiaCADEF proteins
Membrane tolerance
AMPs
DMT transporter (BRPE64_RS06370)
VacJ lipoprotein (BRPE64_RS12115)
MBOAT family protein (BRPE64_RS10515)
OMPs (BRPE64_RS19345, BRPE64_RS21945)
DedA family protein (BRPE64_RS02150)
ENVELOPE BIOGENESIS
Tat transporter
tatC
tatB
tatA
Lipid A
Core oligosaccharide biosynthesis
Inner Core (waaCF, rfaDE)
Outer Core (BRPE64_RS10475-BRPE64_RS10490)
O-antigen biosynthesis
Glycosyl transferases
O-antigen ligase
dTDP-L-rhamnose
Transcription factor for membrane-repair genes
ESR
σE factor
rpoE
rseA
rseB
Proteolytic cleavage
BRPE64_RS05760
tolR
tolQ
tolA
tolB
pal
Tol-Pal complex
Core oligosaccharide
O-antigen
Outer
Inner
LPS
O-antigen export
Wzm/Wzt
Specific resistance factors
Polymyxin B
LL-37
Riptocin
CCR179
CCR480
Common resistance factors
Neutral functions
Polymyxin B and LL-37 specific functions
CCRs specific functions
Global AMP resistance functions

Figure 72: Overview of AMP resistance factors in *B. insecticola.*
Each gene name and function is coloured according to its requirement for certain categories of AMPs. Genes and functions required for resistance towards multiple AMPs are indicated in black (all AMPs), in red (the two CCR peptides) and in purple (polymyxin B and LL-37). Genes and functions required for resistance towards specific AMPs are indicated in orange (polymyxin B), in green (LL-37), in blue (riptocin), in yellow (CCR179) and in pink (CCR480). Neutral genes and functions are indicated in grey. Abbreviations: OMP: outer membrane proteins.

that is often associated to phospholipid remodelling in the bacterial membranes (Hofmann, 2000). Additionally, I identified the *BRPE64_RS09930* gene as a resistance factor towards CCR179, which is encoding a hairpin-helix-hairpin motif protein that is often associated to DNA repair (Aravind *et al.*, 1999). As previously mentioned, the majority of the fitness genes that I have identified during CCR480 peptide treatment were specifically associated to CCR480 peptide resistance, and were not detected with the other tested AMPs (see **section 4.2**). These genes were mostly present on the chromosome 3 of the *B. insecticola* genome, and their functions were mostly associated to carbohydrate and amino acid exploitation with annotated sugar ABC (ATP-binding cassette) transporters (*BRPE64_RS21855*, *BRPE64_RS21860*, *BRPE64_RS21870*, *BRPE64_RS21900*, *BRPE64_RS21905*, *BRPE64_RS21910*). As the chromosome 3 is the least mutated replicon in the *B. insecticola* genome (see **Chapter II**), the reduction of transposon insertions inside these genes was not tremendously different from the MM control condition, but the CCR480 peptide treatment may have an impact on the metabolic activities of the symbiont. Concerning cell wall biogenesis functions encoded by some of these genes located on the chromosome 3, I found a glycosyl transferase (*BRPE64_RS22650*), an outer membrane protein OmpC type (*BRPE64_RS21945*) and an RND efflux transporter (*BRPE64_RS23525*, *BRPE64_RS23530*). Hence, these fitness genes suggest that each CCR peptide has its own specific cell target, with the CCR179 peptide that is mostly affecting the bacterial membranes (see **section 4.2.1**) and the CCR480 peptide that targets the bacterial metabolism (see the recapitulative functions illustrated in **Figure 72**).

4.2.3. Validation of bacterial genes involved in AMP resistance

I targeted several fitness genes that were shared between different AMPs from the lists of conditionally-essential genes in order to confirm the Tn-seq results. I was interested in three main gene clusters which were the *rfb* gene cluster that is involved in dTDP-L-rhamnose biosynthesis, the *tol* genes encoding the Tol-Pal complex and the *BRPE64_RS10555-BRPE64_RS10560* genes which are encoding the Wzm/Wzt O-antigen transport system (see **section 4.2.1**). I have generated bacterial mutants in these targeted genes with the same mutagenesis approach previously used to create hopanoids and ESR mutant strains (see **sections 4.1.2 and 4.1.3**). Specifically, I produced mutants in the *rfbA* (*BRPE64_RS10590*), *rfbC* (*BRPE64_RS110585*), *tolB* (*BRPE64_RS11040*), *tolQ* (*BRPE64_RS11025*) and *wzm* (*BRPE64_RS10560*) genes and investigated their *in vitro* and *in vivo* phenotypes in detail.

Similar to the previous mutants studied in the candidate gene approach, I have checked the AMP sensitivity of these mutant strains, as well as their sensitivity towards other cell wall stressors and their swimming motility. For AMP sensitivity, it appeared that all the mutants were hypersensitive towards polymyxin B and colistin, with a MIC_{90} of 3.125 µg.mL^{-1} for *wzm*, *rfbA* and *tolB* mutants, 12.5 µg.mL^{-1} for *rfbC* mutant and 25 µg.mL^{-1} for the *tolQ* mutant, compared to 50-100 µg.mL^{-1} for the wild-type strain (**Figure 73**). For LL-37, only the *tolB* mutant showed a lower MIC_{90} value (6.25 µg.mL^{-1}) compared to the other strains and the wild-type strain (12.5 µg.mL^{-1}) (**Figure 73**). Similar to the polymyxin B treatment, all the mutants were also susceptible to riptocin with MIC_{90} values of 25 µg.mL^{-1} for the *tolB* mutant, 50 µg.mL^{-1} for the *wzm* mutant, and 100 µg.mL^{-1} for *rfbA*, *rfbC* and *tolQ* mutants (**Figure 73**). As I mentioned before, the MIC_{90} values of the wild-type strain for the two CCR peptides could not be determined here, so I could check only the MIC_{50} values when it was possible (**Figure 73**). Here, I found that the *wzm* and *rfbA* mutants were more sensitive than the wild-type strain for the CCR179 peptide, with a MIC_{50} value of 100 µg.mL^{-1}, whereas the other mutants had the same resistance profile than the wild-type strain (**Figure 73**). Curiously, the *tolB* and *tolQ* mutants showed a higher resistance towards the CCR480 peptide compared to the *wzm*, *rfbA*, *rfbC* mutants and the wild-type strain (**Figure 73**). Thus, the five mutant strains were more

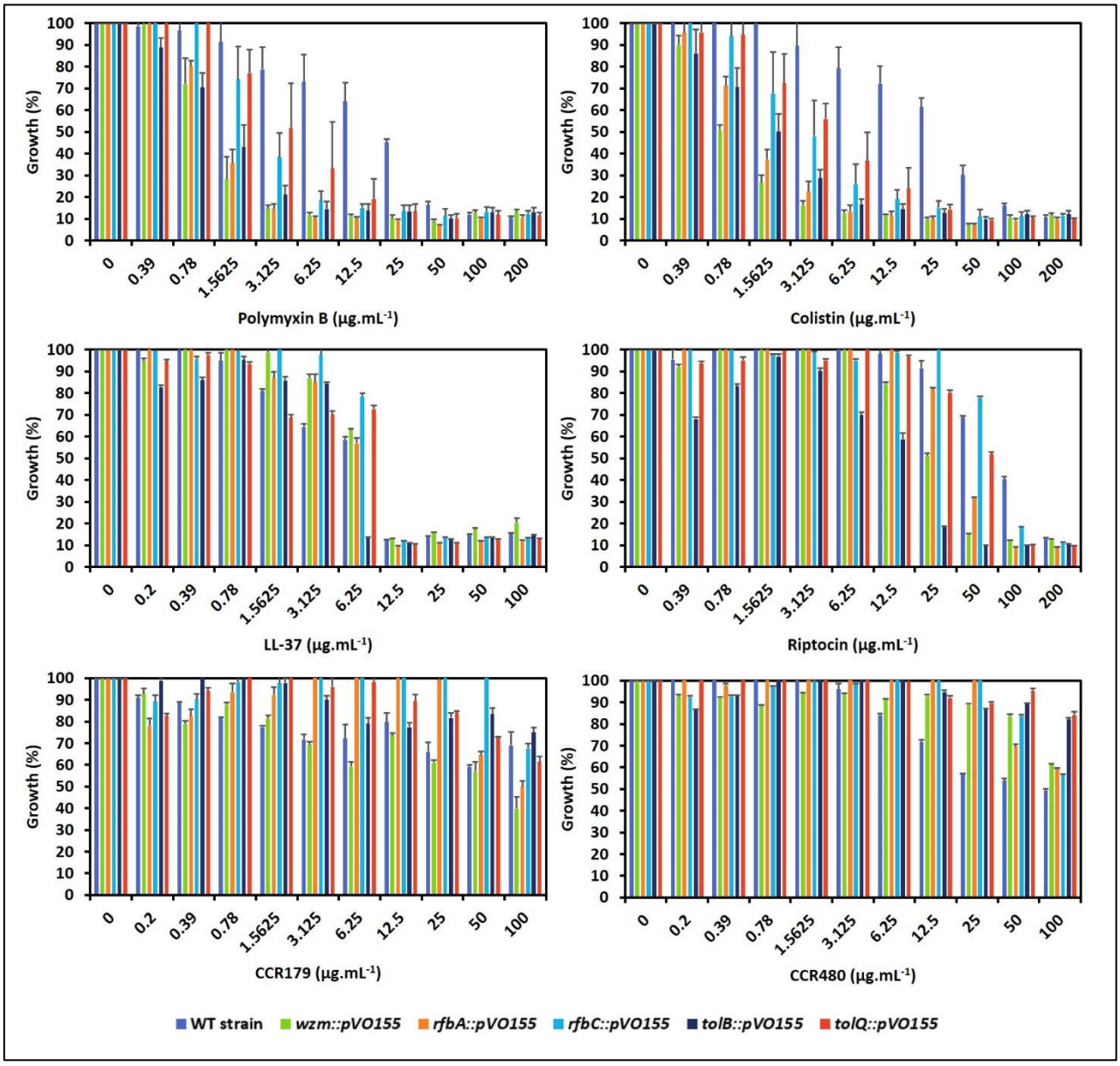


Figure 73: AMP sensitivity of *B. insecticola* mutant strains of genes identified by Tn-seq. Growth of wild-type and mutant strains of genes identified by Tn-seq of *B. insecticola* in MM supplemented with increasing concentrations of AMPs (polymyxin B, colistin, LL-37, riptocin, CCR179 peptide, CCR480 peptide). Growth is expressed as a percentage of growth observed in the MM without AMPs, based on the measured OD_{600nm}. Abbreviations: WT: wild-type.

susceptible to AMPs, especially towards polymyxin B, colistin and riptocin (**Figure 73**). Thus, these results demonstrated that mutants in the fitness genes previously identified by Tn-seq showed a stronger sensitivity towards AMPs, except for the CCR480 peptide. Concerning the other membrane stressors, it appears that the *wzm*, *rfbA*, *tolB* and *tolQ* mutants were more sensitive to an osmotic shock triggered by NaCl (MIC_{90} of 250 mM) compared to the *rfbC* mutant and the wild-type strain (MIC_{90} of 500 mM) (**Figure 74**). I also checked the effect of an

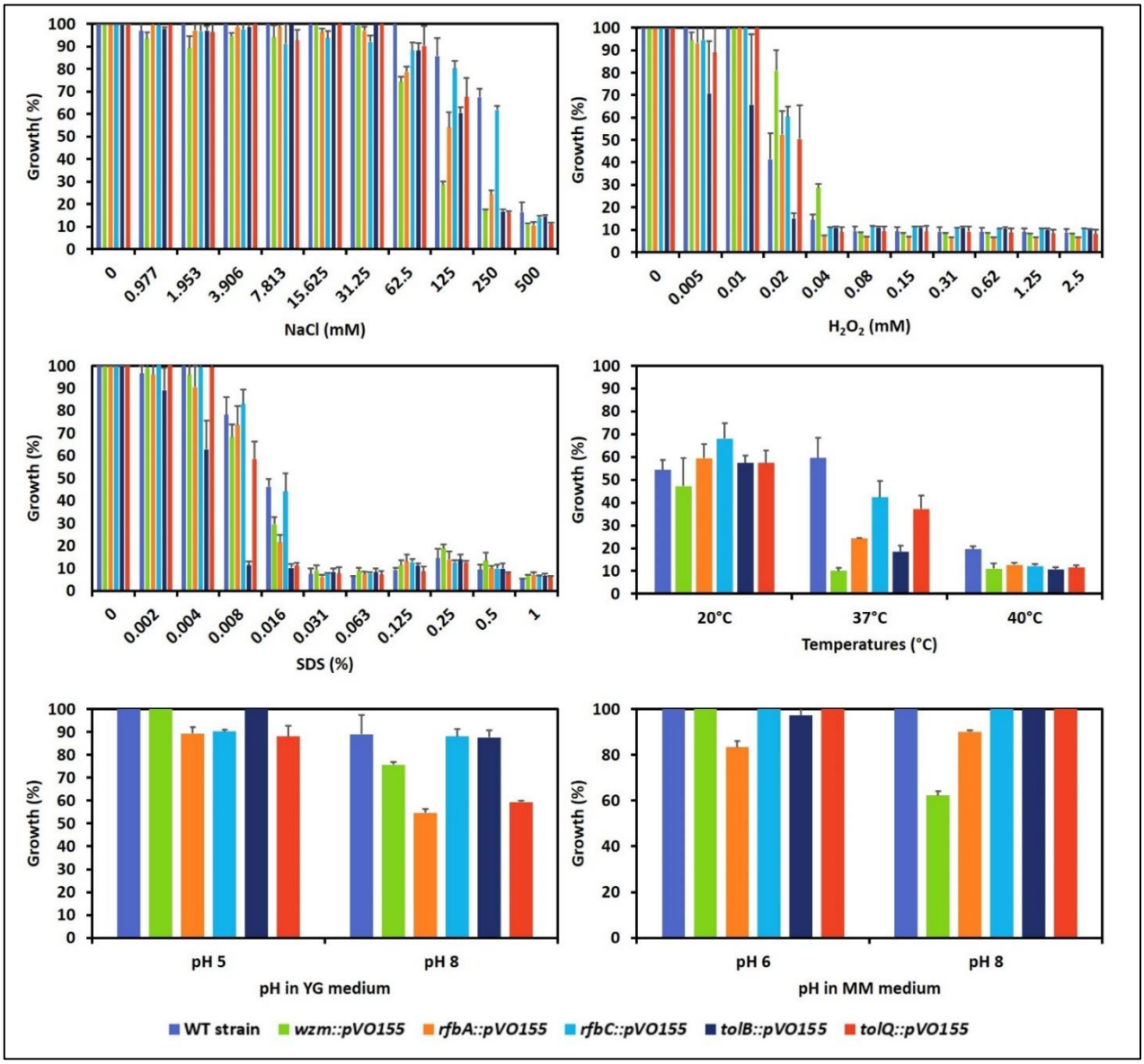


Figure 74: Sensitivity to various membrane stressors of *B. insecticola* mutant strains identified by Tn-seq.

Growth of wild-type and mutant strains of genes identified by Tn-seq of *B. insecticola* in MM supplemented with increasing concentrations of membrane damaging agents (H_2O_2, NaCl, SDS) and in various temperatures and pH growth conditions. Growth is expressed as a percentage of growth observed in the MM without the membrane stressor, based on the measured OD_{600nm}. For temperature variations, the growth is expressed as a percentage of growth in the MM at optimal growth temperature at 28°C. For pH variations in YG and MM media, the growth is expressed as a percentage of growth at optimal pH conditions measured in YG and MM media (pH of 6.3 and pH of 7, respectively). Abbreviations: WT: wild-type.

oxidative stress with H_2O_2 treatment, and I observed that the *tolB* mutant was slightly more sensitive to H_2O_2 (MIC_{90} of 0.02 mM) compared to the other strains (MIC_{90} of 0.04 mM) (**Figure 74**). In the presence of detergents such as SDS, I found that the *tolB* and *tolQ* mutants had an

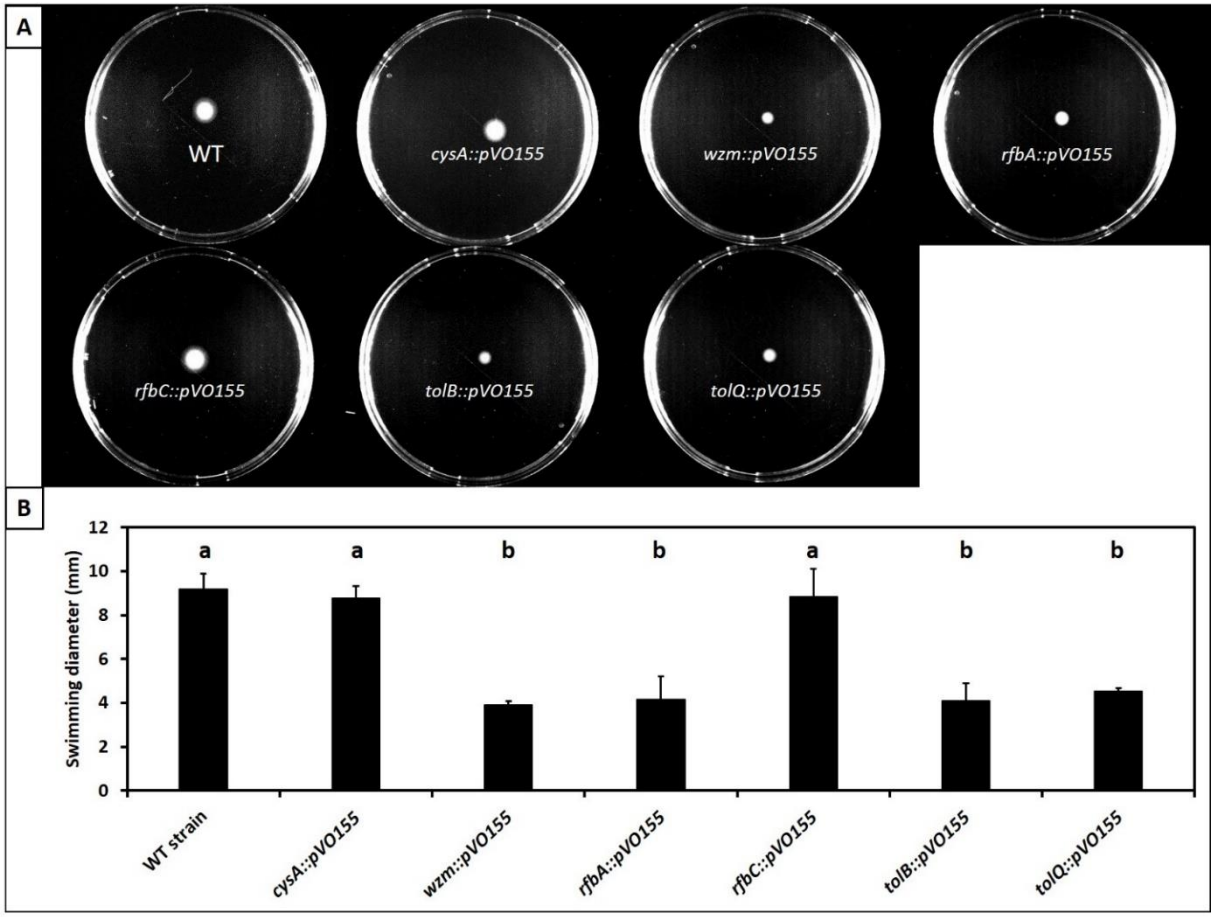


Figure 75: Swimming motility of *B. insecticola* mutants of genes identified by Tn-seq. A) Pictures of swimming motility assays for each strain in YG soft agar plates after 24h. B) Swimming diameter measurements for each strain in YG soft agar plates. The above letters indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey post-hoc correction). Abbreviations: WT: wild-type.

increased sensitivity towards SDS with MIC_{90} values of 0.008% and 0.016%, respectively, compared to the *wzm*, *rfbA*, *rfbC* mutants and the wild-type strain with a MIC_{90} value of 0.031% (**Figure 74**). Moreover, by assessing the effect of the temperature, I noticed that the growth of the *wzm*, *rfbA* and *tolB* mutant strains was reduced at 37°C compared to the growth of the wild-type strain (**Figure 74**). Interestingly, none of these strains could grow at 40°C, including the wild-type strain, which showed that the *Burkholderia* symbiont is a heat-sensitive species (**Figure 74**). In addition, I observed that the *wzm*, *rfbA* and *tolQ* mutants were more sensitive at an alkaline pH in a rich medium, whereas only the *wzm* mutant strain was more susceptible at an alkaline pH in a poor nutrient medium compared to the wild-type strain (**Figure 74**). Thus, these results showed that the *tol* mutants, especially the *tolB* mutant, were more sensitive to detergents, an oxidative stress, an increased temperature and an osmotic

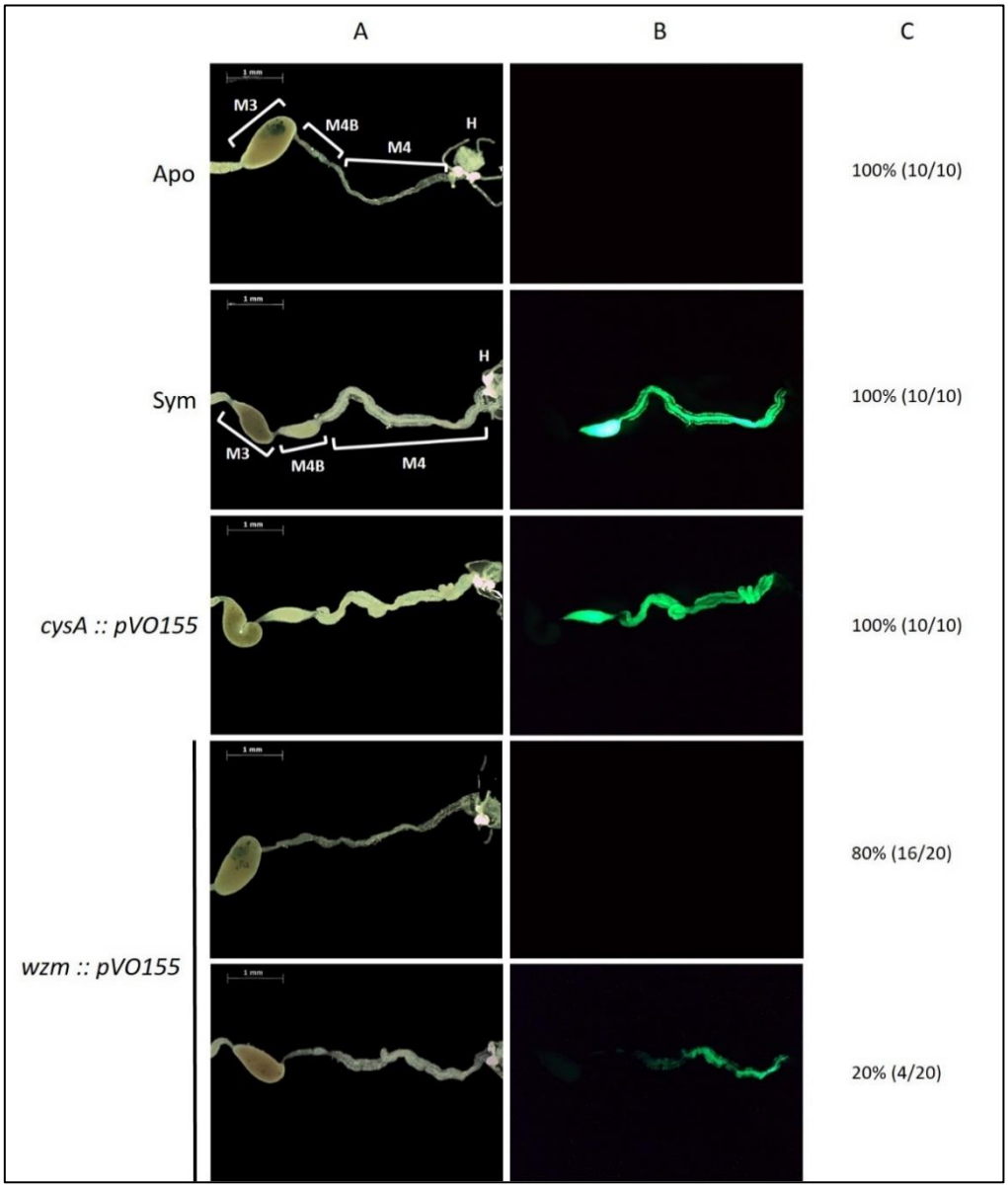


Figure 76: Midgut morphologies of second instar *R. pedestris* insects infected with *wzm* mutant strain of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

shock. The O-antigen mutants in the *wzm* and *rfbA* genes were highly susceptible to an osmotic shock, an increased temperature and an increased pH, whereas the *rfbC* mutant had similar sensitivities than the wild-type strain. In addition, I checked the swimming motility of these mutant strains and I observed that the *wzm*, *rfbA*, *tolB* and *tolQ* mutants exerted a significantly reduced swimming diameter compared to the *cysA*, *rfbC* mutants and the wild-

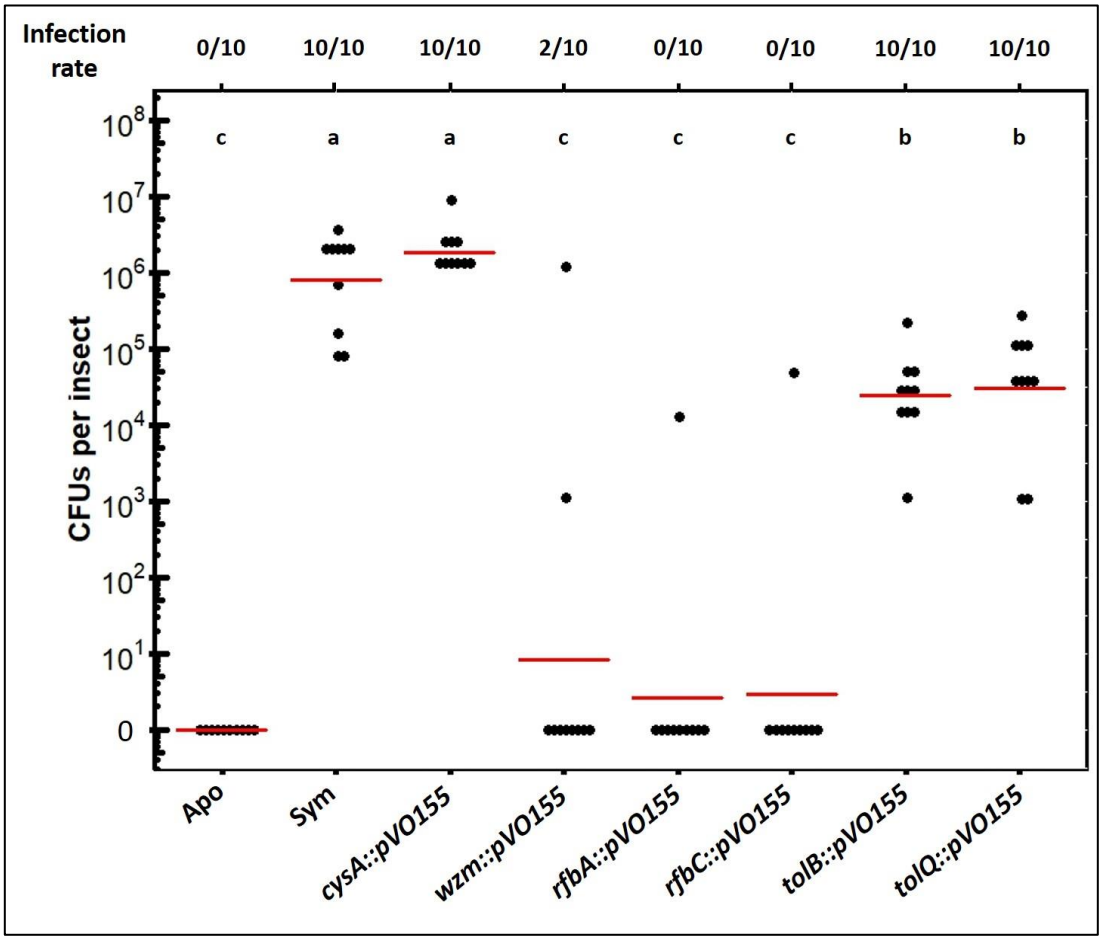


Figure 77: Symbiont titers of *B. insecticola* strains of genes identified by Tn-seq at the second instar stage in the M4 organ.

The total number of CFUs per insect was calculated for each condition. The related infection rate corresponds to the number of infected insects out of ten insects and is indicated above each condition. Means are indicated by red bars for each condition (n = 10 insects). The above letters indicate statistically significant differences (p-value < 0.05, Kruskal-Wallis). Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

type strain (**Figure 75**).

Similar to the previous mutant strains, I evaluated the colonization efficiency of these five mutants by analyzing the crypt morphology and quantifying the symbiont population at the second and the third instar stages of the host insect (see **section 4.1.1**). With the *wzm* mutant strain, I observed that 80% of the dissected insects appeared aposymbiotic and only 20% of the insects were partially colonized by the *wzm* mutant in the M4 organ at the second instar stage (**Figure 76**). In agreement, the *wzm* mutant was detected only inside the M4 organ of partially colonized insects with moreover a decreased population level compared to symbiotic insects colonized by the wild-type strain at the second instar stage (**Figure 77**). At the third instar stage, I found that the proportion of insects infected with the *wzm* mutant showed a

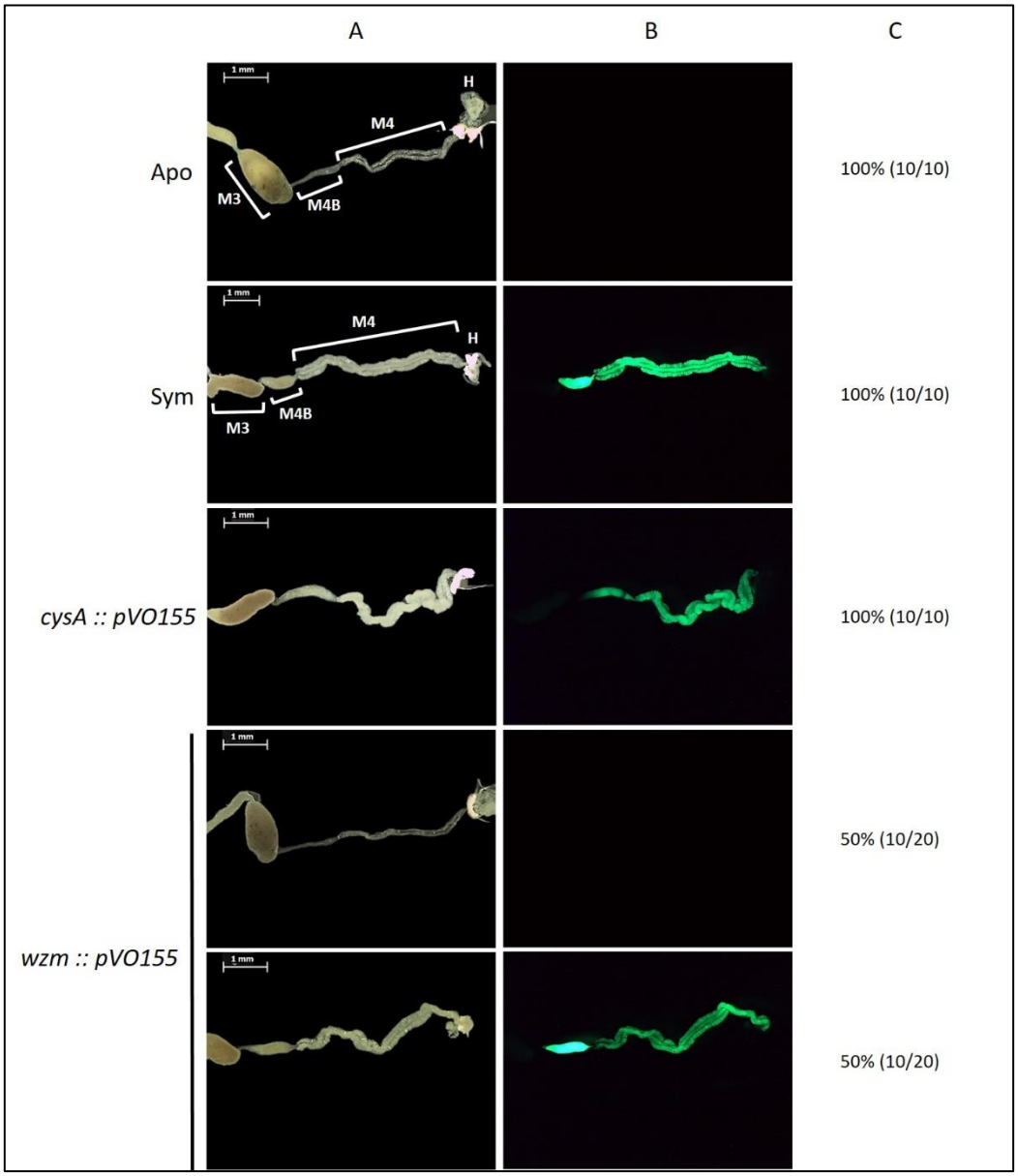


Figure 78: Midgut morphologies of third instar *R. pedestris* insects infected with *wzm* mutant strain of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

similar profile of colonization than insects infected with the wild-type strain in 50% of the insects while the other insects remained uninfected (**Figure 78**). This pattern remained constant thereafter in the fourth and fifth instar stages. Interestingly, the *wzm* mutant population present in colonized insects showed similar CFU counts in the M4 organ than the

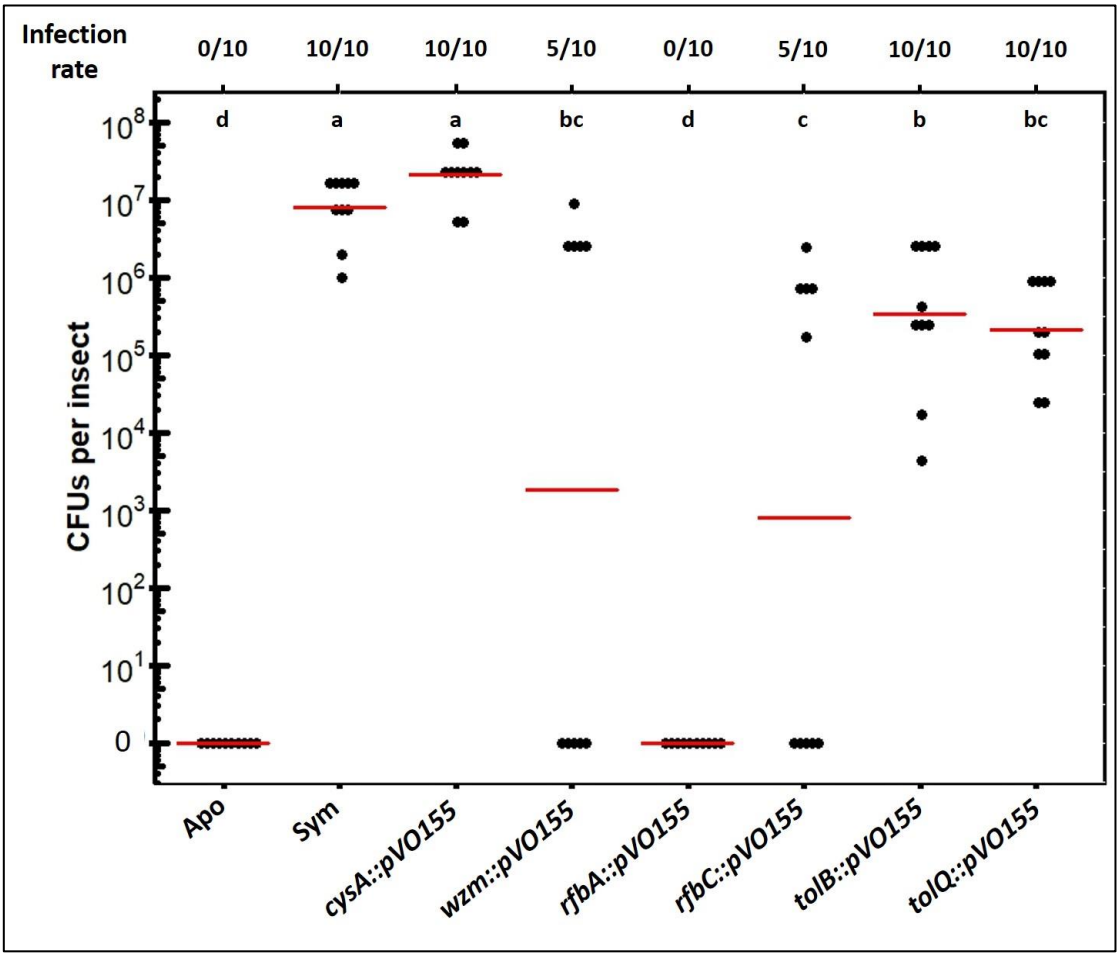


Figure 79: Symbiont titers of *B. insecticola* strains of genes identified by Tn-seq at the third instar stage in the M4 organ.

The total number of CFUs per insect was calculated for each condition. The related infection rate corresponds to the number of infected insects out of ten insects and is indicated above each condition. Means are indicated by red bars for each condition (n = 10 insects). The above letters indicate statistically significant differences (p-value < 0.05, Kruskal-Wallis). Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

wild-type strain population in symbiotic insects (**Figure 79**). However, as 50% of the insect population was not colonized by the *wzm* mutant strain, the mean proportion of this mutant population inside the M4 organ (≈ 10^3 CFUs per insect) was significantly different from the symbiont population counted in symbiotic insects at the third instar stage (≈ 10^7 CFUs per insect) (**Figure 79**). The *rfbA* and *rfbC* mutants were unable to colonize the symbiotic organ at the second instar stage (**Figures 77 and 80**)**.** Similar to the second instar stage, the *rfbA* mutant was not able to colonize the M4 region at the third instar stage (**Figures 79 and 81**). However, third instar insects infected by the *rfbC* mutant showed the same profile than the *wzm* mutant with 50% of the insect population being successfully colonized by the mutant strain and 50%

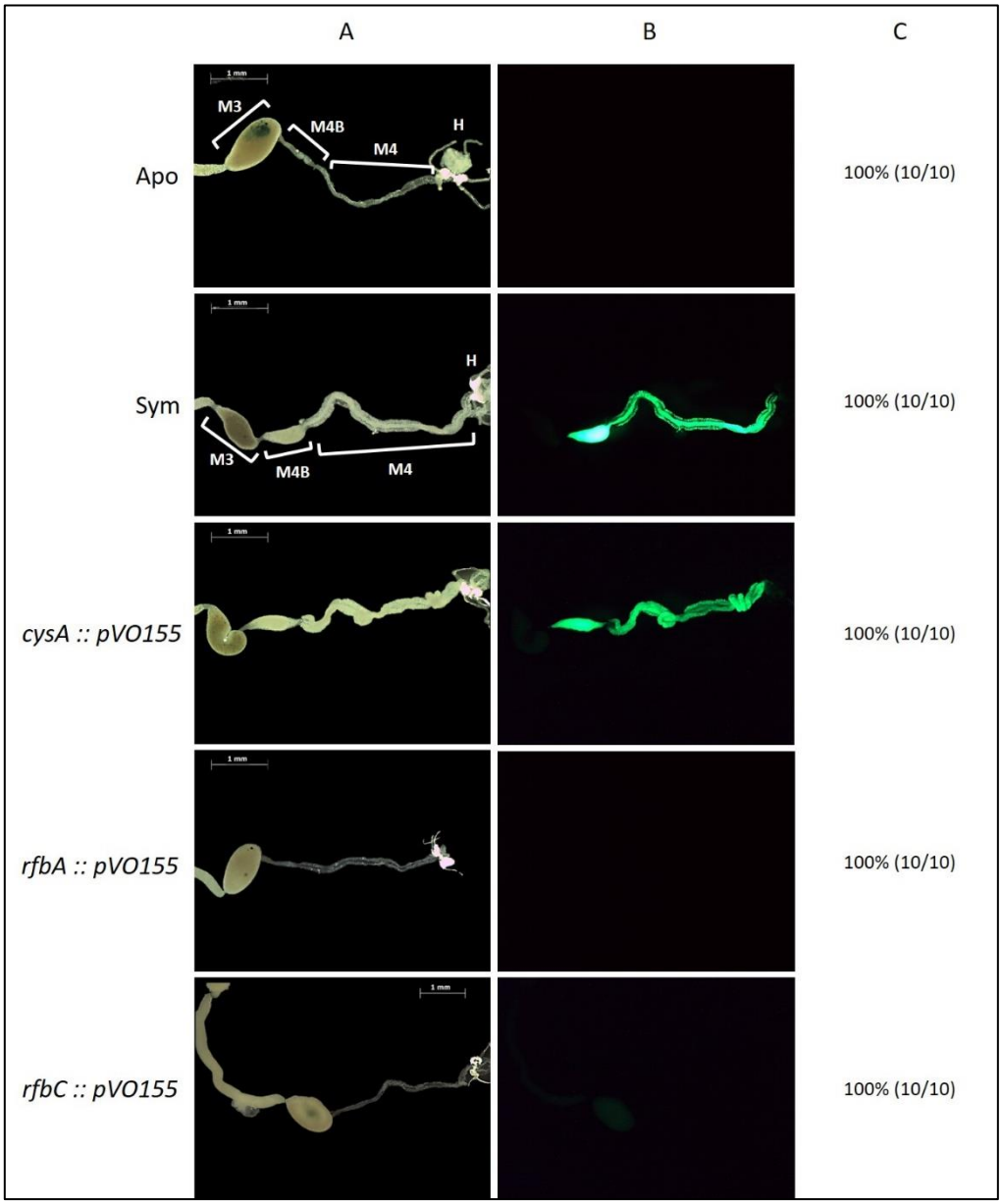


Figure 80: Midgut morphologies of second instar *R. pedestris* insects infected with *rfb* mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

of the insect population being aposymbiotic (**Figure 81**). In addition, the mean proportion of the *rfbC* mutant population was similar as for the *wzm* mutant population, significantly reduced compared to wild-type infected insects (**Figure 79**).

Second instar insects infected with the *tolB* or *tolQ* mutant strains showed a similar profile

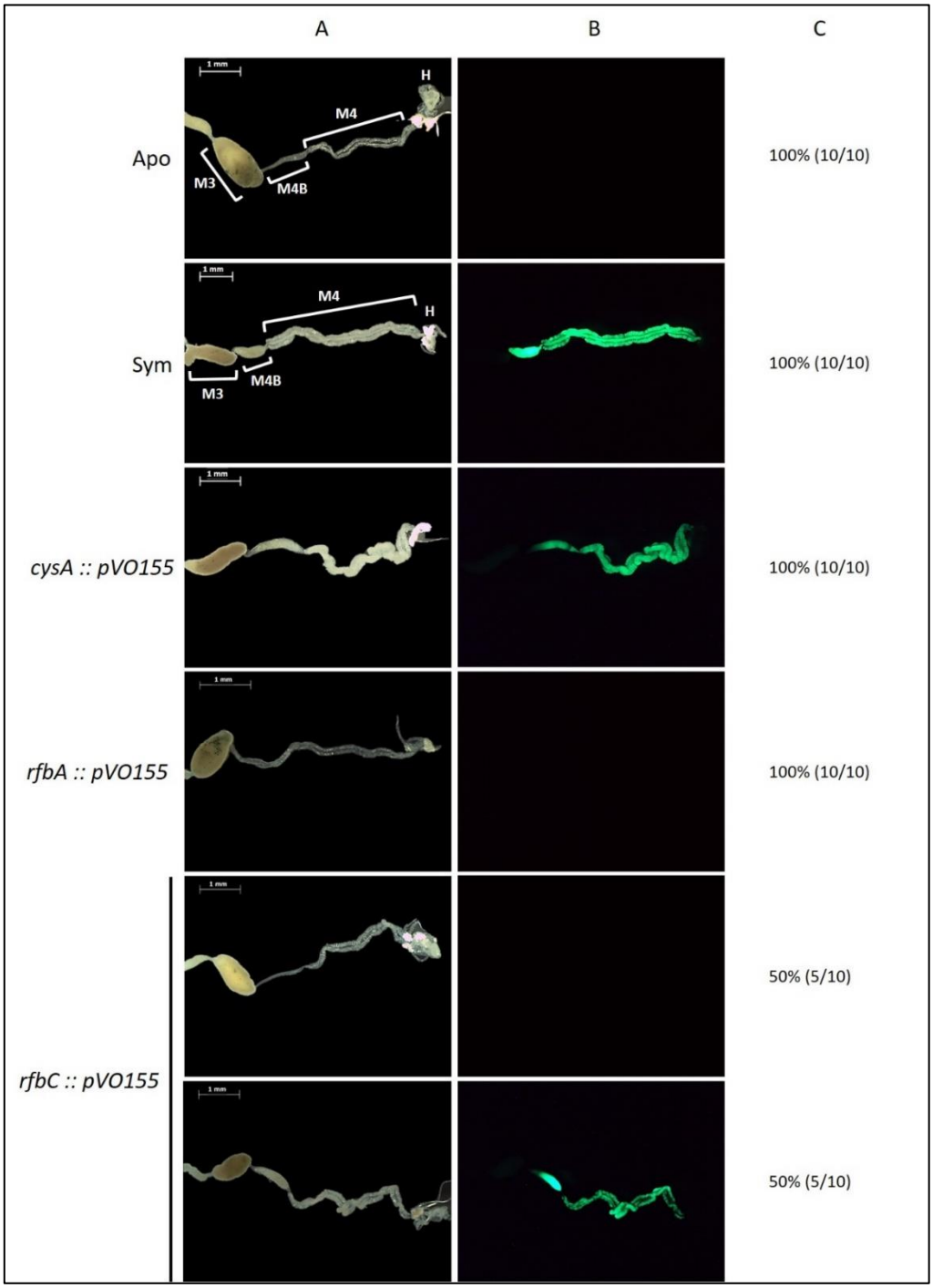


Figure 81: Midgut morphologies of third instar *R. pedestris* insects infected with *rfb* mutant strain of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

than insects infected with the wild-type stain, with an infection rate of 100% (**Figure 82**).

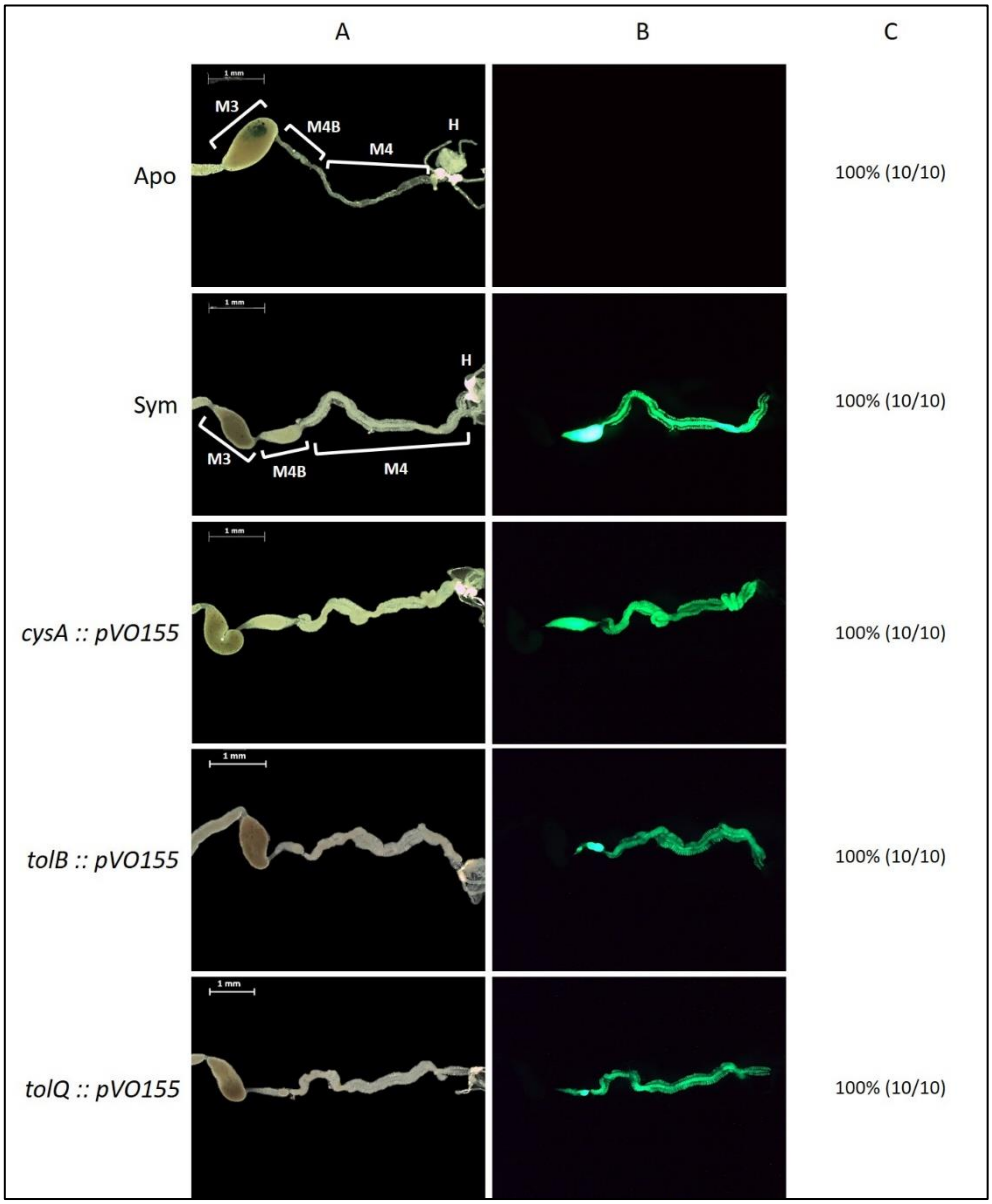


Figure 82: Midgut morphologies of second instar *R. pedestris* insects infected with *tol* mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

However, the number of symbiotic bacteria quantified in the M4 organ in these insects (between 10^4 and 10^5 CFUs per insect) was significantly lower compared to the population present in insects infected by the wild-type strain (≈ 10^6 CFUs per insect) (**Figure 77**). Thus, the *tol* mutants can colonize the M4 region but not as efficiently as the wild-type strain. At the

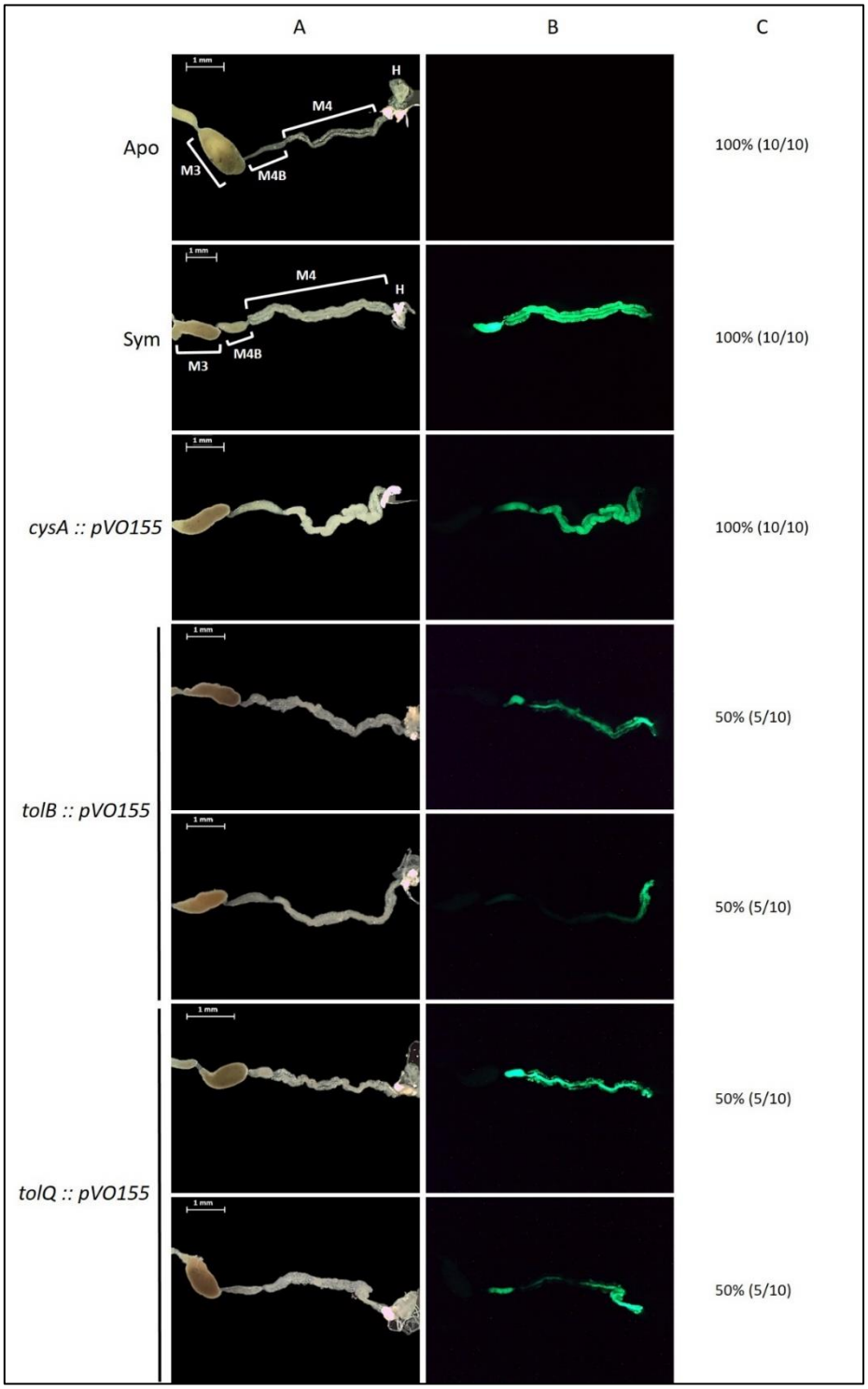


Figure 83: Midgut morphologies of third instar *R. pedestris* insects infected with *tol* mutant strains of *B. insecticola*.

Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and Sym insects. A) Bright field. B) GFP fluorescence. C) Insect proportion associated to the corresponded observation is displayed in percentage with the indicated number of insects associated to that observation out of ten insects. Scale bars (white, upper left corner) represent 1 mm. Abbreviations: Apo: aposymbiotic, Sym: symbiotic.

third instar stage, the *tolB* and *tolQ* mutants persisted in the midgut of all studied insects but interestingly, 50% of the dissected population showed a symbiotic organ that was only partially colonized (**Figure 83**). Indeed, the GFP fluorescence was less intense or absent in the M4B region and in the anterior part of the M4 organ, while the fluorescence persisted in the posterior part of the M4 region close to the hindgut in these insects (**Figure 83**). In addition, the symbiont titers of the two *tol* mutants contained in the M4 region (between 10^5 and 10^6 CFUs per insect) was significantly lower than the symbiont titers of the wild-type strain (≈ 10^7 CFUs per insect) (**Figure 79**). The extinguishment of the GFP fluorescence for the *tol* mutants could be due to a plasmid reversion phenomenon in these symbiotic mutants. To check that possibility, the presence of revertants in the M4 organ of these insects infected by the *tol* mutants was analysed by plating the M4 content, and I did not detect any wild-type bacteria. Thus, the loss of the GFP signal in the third instar insects might correspond to the death of symbiotic bacteria.

In conclusion, all the five mutants showed an impaired *in vivo* colonization of the symbiotic organ. The *rfbA* mutant has entirely lost its colonization abilities. The *wzm* and *rfbC* mutants that displayed an impaired colonization efficiency with a 50% infection rate only. And finally, the *tol* mutants were not able to reach the wild-type population level and progressively lost their capacity to persist in the midgut. Thus these results indicated a correlation between the loss of AMP resistance and colonization defects in these five mutant strains, and confirmed and extended the conclusion obtained by the candidate-gene approach.

4.3. Host colonization competitions experiments

In the previous experiments, I have shown that some bacterial mutants displayed intermediate colonization phenotypes of the symbiotic organ of *R. pedestris*. However, these observations were based on mono-infection experiments (see **sections 4.1 and 4.2.3**), and did not take into account the possible loss of fitness when a mutant strain has to compete for ressources compared to a wild-type bacterial population. Therefore, I have performed competition experiments between the wild-type strain and these mutant strains. To evaluate these competitions, I have made *in vitro* competitions until the bacterial populations reached a stationary phase, and I have performed insect colonization competition experiments and collected the M4 organs at three days post-infection during the second instar stage. To

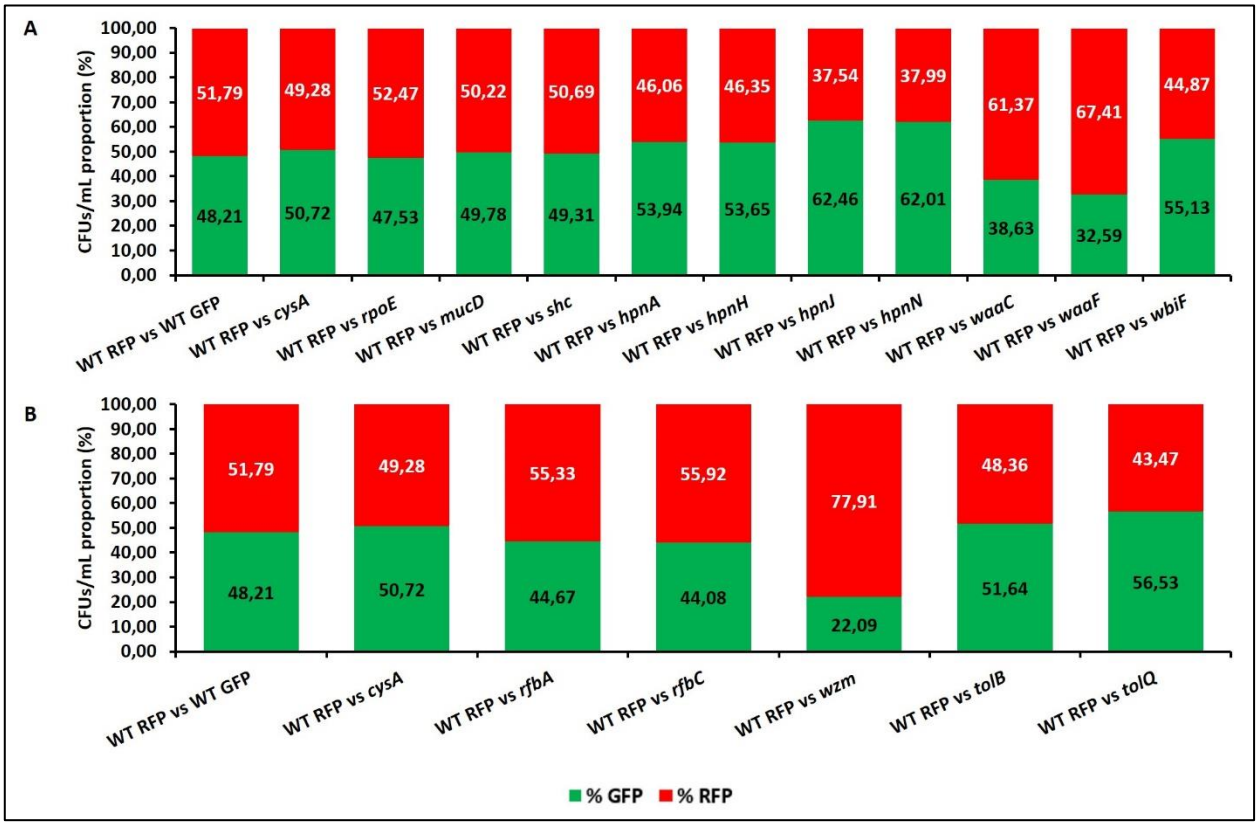


Figure 84: Proportions of wild-type and mutant populations in the inoculum before starting competition experiments.

In each inoculum, the wild-type strain (RFP-labelled) and a mutant strain (GFP-labelled) were theoretically mixed in a 1:1 ratio. The measured percentages of each population were estimated by counting the mean number of CFUs per mL on three deposits for each inoculum. The percentage of each GFP-labelled bacteria (black) and RFP-labelled bacteria (white) is indicated for each inoculum. A) Candidate gene approach targets. B) Tn-seq targets. Abbreviations: WT: wild-type.

distinguish the competitors, I used a RFP-labelled wild-type strain and the GFP-labelled mutants.

As a first step, I have checked the initial proportions of the two bacterial populations that were present in the inoculums with a theoretical ratio of 1:1 (wild-type RFP strain : mutant GFP strain) by CFU counting, that were subsequently used for *in vitro* and *in vivo* competition assays (**Figure 84**). In the inoculums, almost all the conditions with the two mixed bacterial populations were closed to a 1:1 ratio with a few exceptions (**Figure 84**). Indeed, two competiton associations with the *hpnJ* and *hpnN* mutants, contained approximately 60% of mutant population and 40% of wild-type population (**Figure 84**). Additionally, three other coinoculums contained less of the mutant population than expected, with approximately 40% for the *waaC* mutant, 30% for the *waaF* and even 20% for the *wzm* mutant (**Figure 84**).

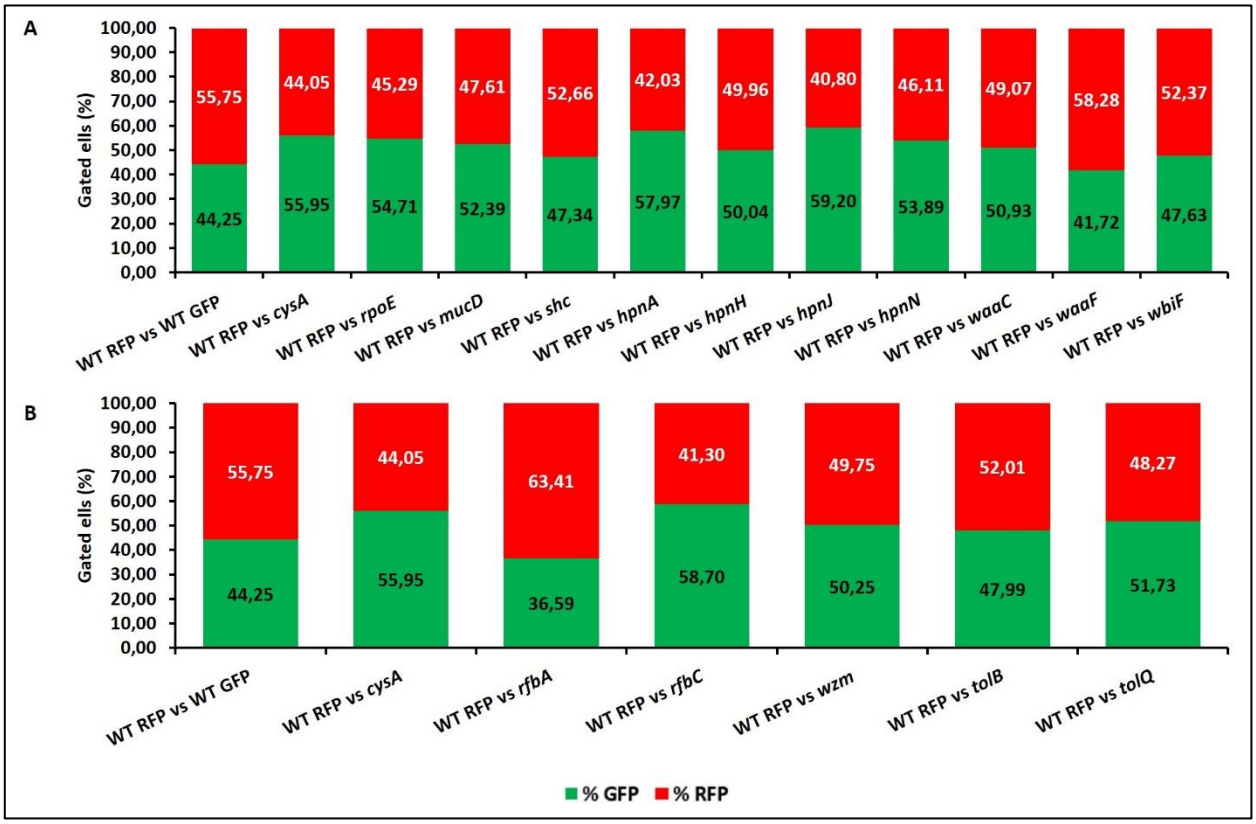


Figure 85: Proportions of wild-type and mutant populations during *in vitro* competitions. Each *in vitro* competition was performed between the wild-type strain (RFP-labelled) and a mutant strain (GFP-labelled) initially mixed at a ratio 1:1. The mean percentage of gated cells that were GFP (green) or RFP (red)-labelled for each competition were obtained from flow cytometry experiments on three independent *in vitro* competitions at stationary phase. The percentage of each GFP-labelled bacteria (black) and RFP-labelled bacteria (white) is indicated for each competition. A) Candidate gene approach targets. B) Tn-seq targets. Abbreviations: WT: wild-type.

Next, I have performed *in vitro* competitions using these initial coinoculums, and determined the final bacterial populations in the stationary phase culture by flow cytometry (**Figure 85**). The mean proportions of each population was close to 50%, except for the *rfbA* mutant population that reached only approximately 40% of the total population (**Figure 85**). From the initial proportions and the output populations, I calculated a competitive index (CI) as was described before (Auerbuch *et al.*, 2001; Gonzalez-Mula *et al.*, 2018; Macho *et al.*, 2016). (**Figure 86**). The CI indicates if one of the two strains is advantaged in the colonization of a specific niche compared to the other strain. Here, when a CI is inferior to 1, the mutant strain is outcompeted by the wild-type strain, whereas a CI superior to 1 indicates that the mutant strain is more competitive than the wild-type strain. For *in vitro* competitions, I noticed that

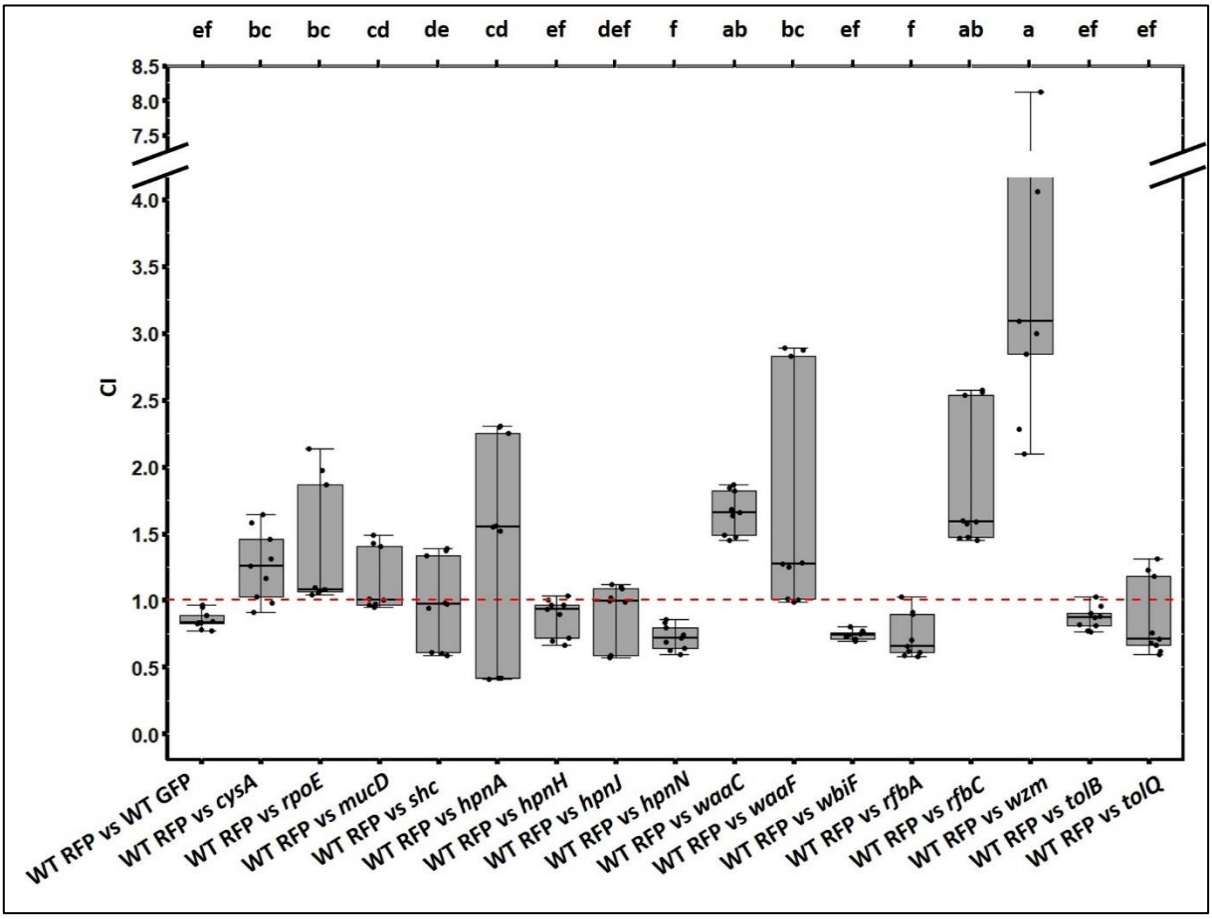


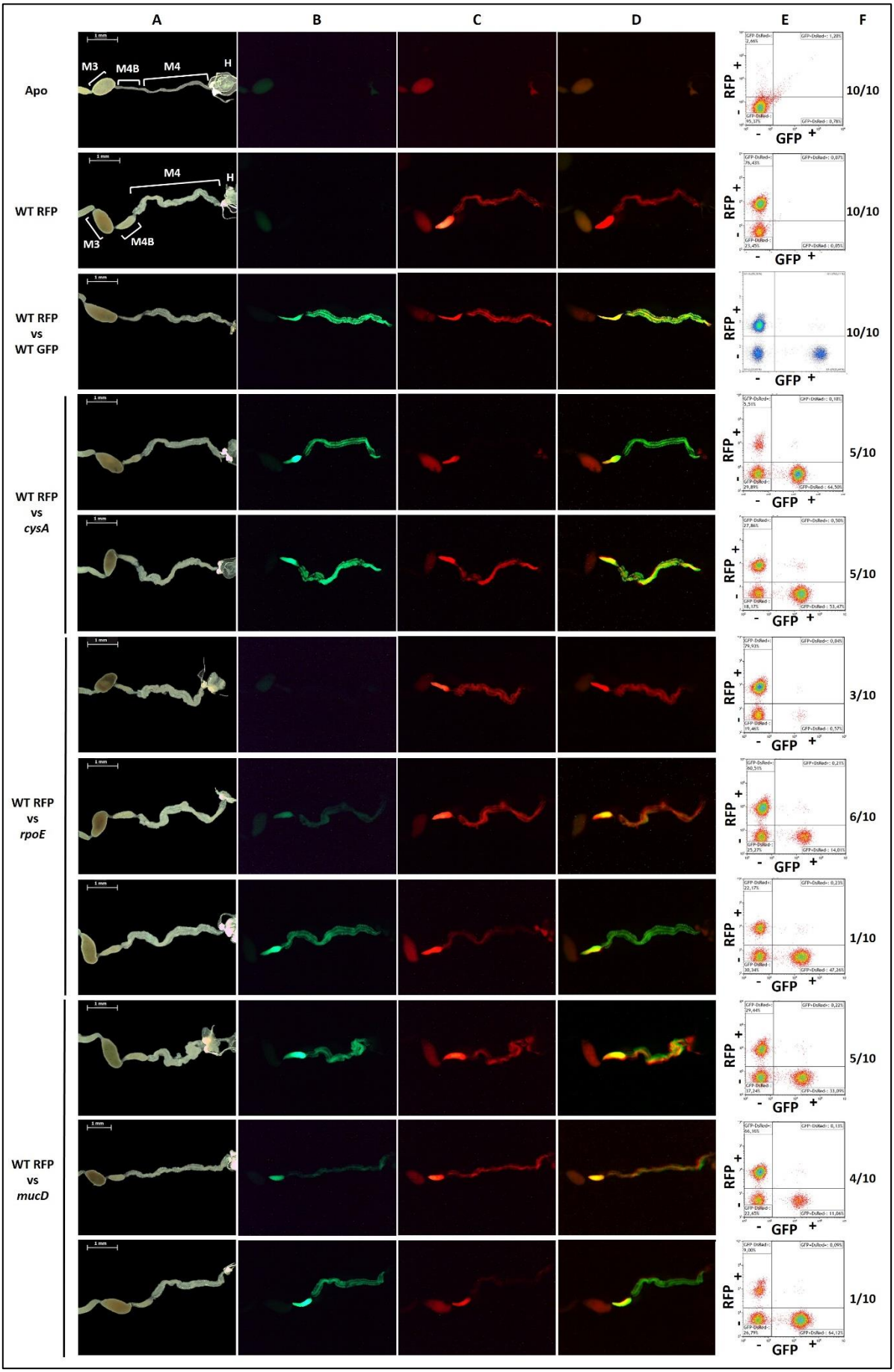
Figure 86: Competitive indexes for *in vitro* competitions.
Competitive indexes (CI) for each competition were calculated and displayed as boxplots. The red dotted line indicates a CI equals to 1, which means no competition between the wild-type strain and the mutant strain. When a CI is inferior to 1, the wild-type strain is more competitive than the mutant strain. When a CI is superior to 1, the mutant strain becomes more competitive than the wild-type strain. Different letters above each boxplot indicate statistically significant differences (p-value < 0.05, Kruskal-Wallis).

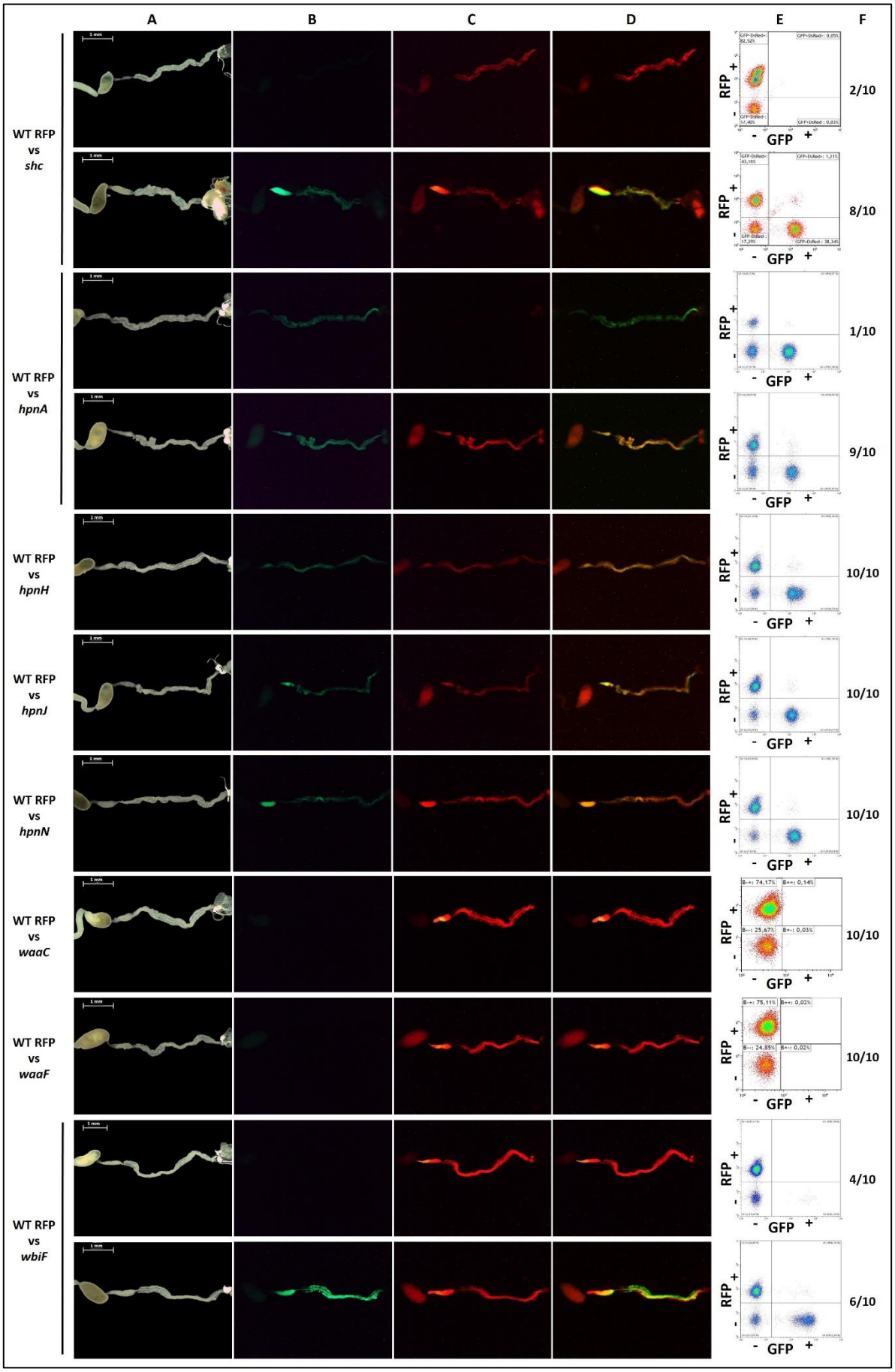
the CIs of a control competition between two wild-type strains RFP and GFP-tagged were close to 1, which confirmed that there are no competition between these two strains during *in vitro* growth (**Figure 86**). Other CIs of *in vitro* competitions containing the mutant populations of the *shc*, *hpnH*, *hpnJ*, *hpnN*, *wbiF*, *rfbA*, *tolB* and *tolQ* mutants were also close to 1, indicating no competition with the wild-type strain for *in vitro* ressources (**Figure 86**). However, CIs were higher than 1 for competition assays that included the *cysA*, *rpoE*, *mucD*, *hpnA*, *waaC*, *waaF*, *wzm* and *rfbC* mutants (**Figure 86**). As the proportions of the mutant populations increased from the initial inoculum to reach almost the same proportions as the wild-type strain (close to 50%), these results could indicate that the estimations of the initial inocula were erroneous, or these mutant strains gained a fitness advantage during *in vitro* growth conditions in a rich

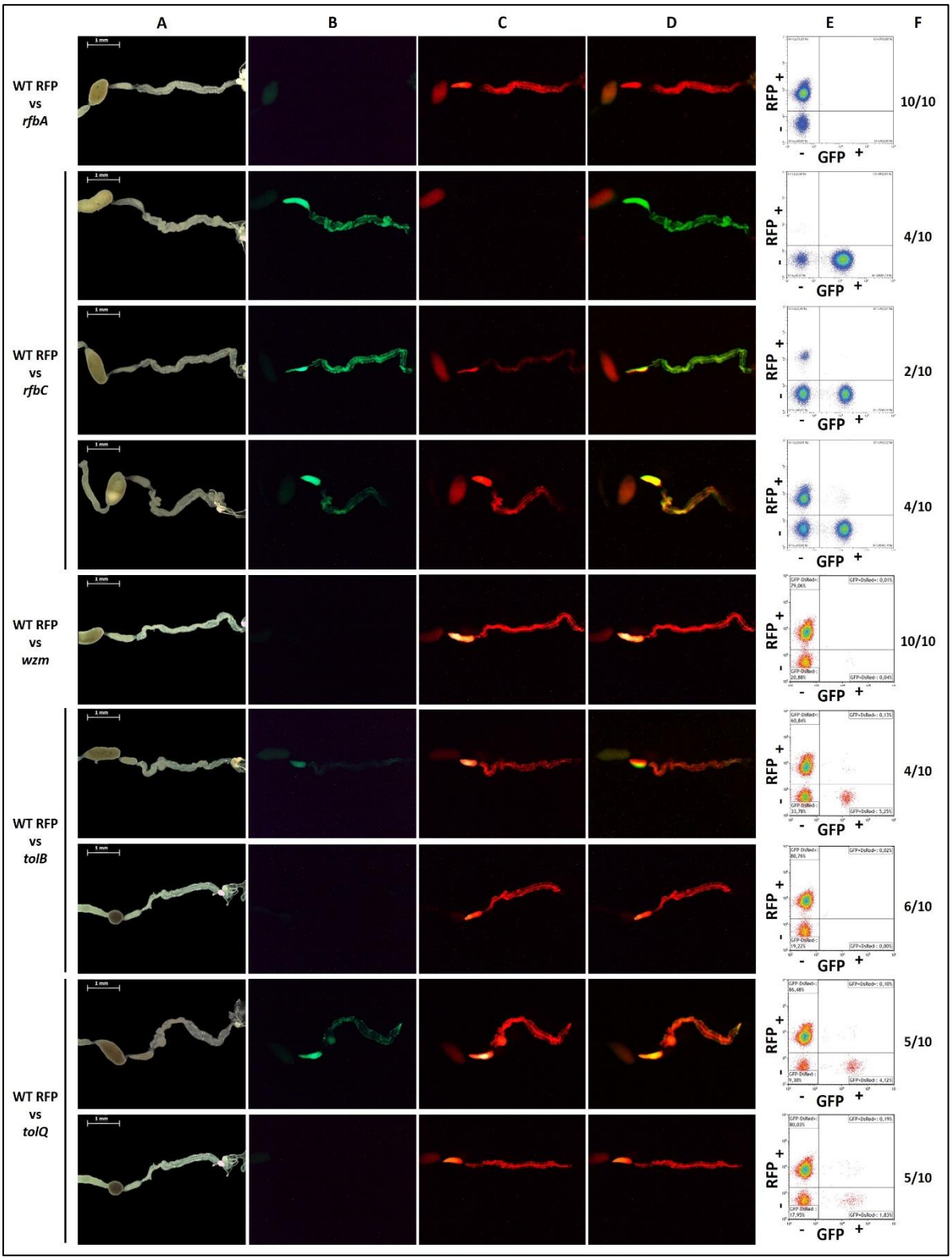
medium to reach a similar population level as the wild-type strain. This was particularly obvious for the *wzm* mutant, which increased from 22.09% in the initial inoculum (**Figure 84**) to reach on average 50.25% of the total bacterial population after *in vitro* competition with the wild-type strain (**Figure 85**).

Next, I have infected young *R. pedestris* insects with these mixed inoculums in similar conditions than the previous mono-infections (see **sections 4.1 and 4.2.3**). After three days post-infection, I collected the M4 organs of ten insects per competition assay. I observed the crypt morphology and the fluorescence patterns in the symbiotic organ for each insect and estimated their content of each bacterial population by flow cytometry (**Figure 87**).

For aposymbiotic insects, there was no detection of GFP or RFP signals in the symbiotic organ, as expected. However, there was some fluorescent signals detectable in the M3 organs (**Figure 87),** which was also previously observed and most likely corresponds to some autofluorescence in this organ. In the flow cytometry analysis of the M4 organs of aposymbiotic insects, cellular debris or mitochondrial content that is both GFP and RFP negative is detected using the forward and side scatter gating parameters for detection of bacteria (**Figure 87E**). This background signal originates from the insect tissues of the symbiotic organs that contain eukaryotic cells and was present in all the M4 samples analyzed in regular proportions (≈ 25 % of the total gated cells) (**Figure 87E**). In a control mono-infections experiment with the RFP wild-type strain, all insects carried 100% of RFP-tagged bacteria in their symbiotic organ (**Figures 87 and 88**). In an additional control infection with GFP-labelled and RFP-labelled wild-type strains, both were able to colonize the M4 organ (**Figure 87**), but the flow cytometry revealed that the RFP population was more abundant (≈ 75%) than the GFP population (≈ 25%) (**Figures 87 and 88**). With the *cysA* mutant that I used as a positive control during mono-infections, I observed that the mutant was able to colonize the symbiotic organ better than the wild-type strain which was absent from the M4 organ in 50% of dissected insects (**Figure 87**). By calculating the mean proportions of the ten samples, it appeared that the *cysA* mutant population was more abundant (≈ 73%) than the wild-type population in the symbiotic organ (≈ 27%) (**Figure 88**). In coinfections with the *rpoE* mutant, insects were almost deprived of GFP fluorescence in the M4 region (**Figure 87**), which were mostly colonized by the RFP wild-type strain (**Figures 87 and 88**).


A
B
C
D
E
F
Apo
M3
M4B
M4
H
10/10
WT RFP
M4
H
M3
M4B
10/10
WT RFP
vs
WT GFP
10/10
WT RFP
vs
cysA
5/10
5/10
WT RFP
vs
rpoE
3/10
6/10
1/10
WT RFP
vs
mucD
5/10
4/10
1/10
RFP
GFP


A
B
C
D
E
F
WT RFP vs shc
2/10
8/10
WT RFP vs hpnA
1/10
9/10
WT RFP vs hpnH
10/10
WT RFP vs hpnJ
10/10
WT RFP vs hpnN
10/10
WT RFP vs waaC
10/10
WT RFP vs waaF
10/10
WT RFP vs wbiF
4/10
6/10
RFP
GFP
1 mm



On the contrary, by coinfecting insects with the *mucD* mutant, the M4 region was more colonized by the mutant strain than the wild-type strain (**Figure 87**), with the mutant population representing approximately 60% of the total symbiont population (**Figure 88**). For

Figure 87: Colonization of the symbiotic organ during *in vivo* competitions.
Each *in vivo* competition was performed between the wild-type strain (RFP-labelled) and a mutant strain (GFP-labelled), that were initially mixed at a ratio 1:1. Pictures are showing the posterior midgut regions, with the M3, M4B, M4 and H regions indicated in white for Apo and WT RFP insects. A) Bright field. Scale bars (white, upper left corner) represent 1 mm. B) GFP fluorescence. C) RFP fluorescence. D) Merged fluorescences. E) Flow cytometry of bacterial cells gated by detection of negative (-) or positive (+) signals of RFP and GFP fluorescences. F) Indicated number of insects which were associated to the observed phenotype, out of ten insects. Abbreviations: Apo: aposymbiotic, WT: wild-type.

coinfections with hopanoid mutants, I found that these five mutants were able to colonize the symbiotic organ in the presence of the wild-type strain (**Figure 87**). Regarding the population sizes, the mutant populations were slightly less abundant than the wild-type population,

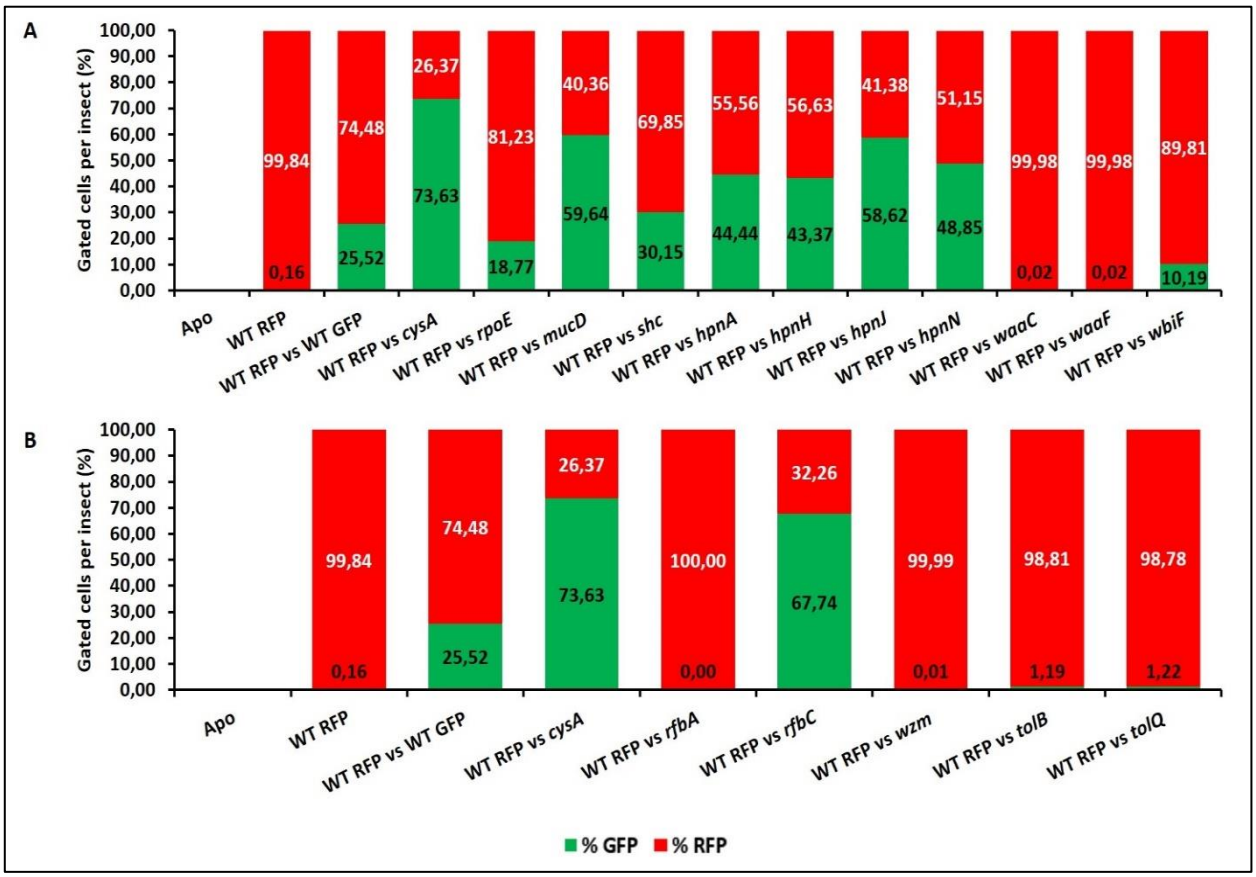


Figure 88: Colonization efficiency of the symbiotic organ during *in vivo* competitions.
Each *in vivo* competition was performed between the wild-type strain (RFP-labelled) and a mutant strain (GFP-labelled), that were initially mixed at a ratio 1:1. The mean percentage of gated cells that were GFP (green) or RFP (red)-labelled for each competition were obtained from flow cytometry experiments on symbiotic organs collected from ten insects (replicates) at three days-post-infection during the second instar stage. The percentage of each GFP-labelled bacteria (black) and RFP-labelled bacteria (white) is indicated for each competition. A) Candidate gene approach targets. B) Tn-seq targets. Abbreviations: Apo: aposymbiotic, WT: wild-type.

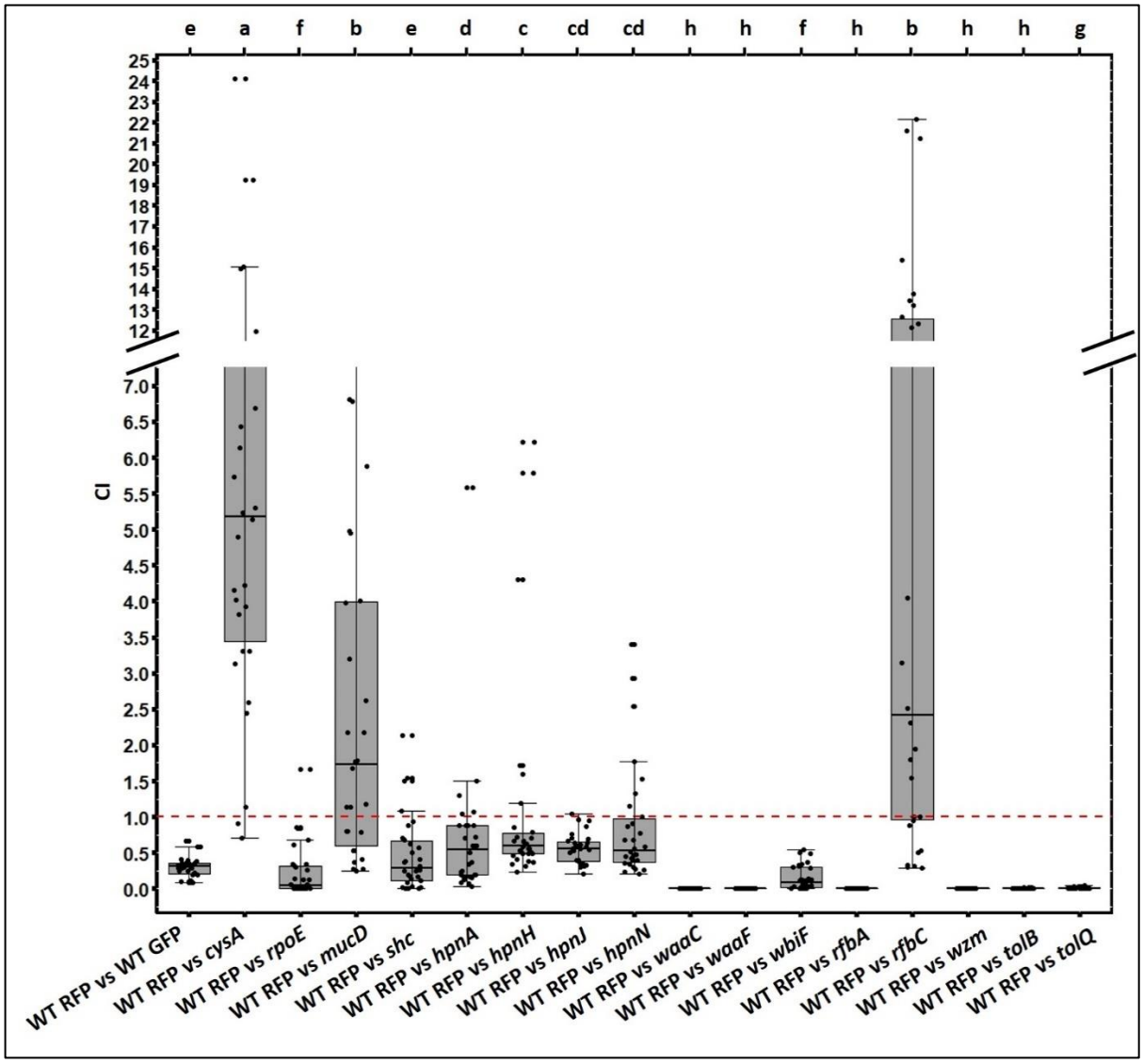


Figure 89: Competitive indexes for *in vivo* competitions.
Competitive indexes (CI) for each competition were calculated and displayed as boxplots. The red dotted line indicates a CI equals to 1, which means no competition between the wild-type strain and the mutant strain. When a CI is inferior to 1, the wild-type strain is more competitive than the mutant strain. When a CI is superior to 1, the mutant strain becomes more competitive than the wild-type strain. Different letters above each boxplot indicate statistically significant differences (p-value < 0.05, Kruskal-Wallis).

especially for the *shc* mutant which represented around 30% of the total bacterial population (**Figure 88**). When insects were coinfected with the wild-type strain and the *waaC* or the *waaF* LPS mutants, I observed that only the wild-type strain was able to colonize the symbiotic organ for each analyzed insect in agreement with the inability of these mutants to infect the M4 in mono-infections as well (**Figures 87 and 88**). Concerning the *wbiF* mutant, the coinfections showed that half of the insect cohort was only colonized by the wild-type strain, whereas the

other half was colonized by both populations (**Figure 87**). However, the flow cytometry results indicated that the proportion of the *wbiF* mutant population was on average less abundant (≈ 10%) than the wild-type population (≈ 90%) (**Figure 88**). Concerning the *in vivo* competitions between the wild-type strain and the *rfbA* or the *wzm* strains, I observed the same results than the *waaC* and *waaF* coinfections, with an exclusive colonization of the symbiotic organ by the wild-type strain (**Figures 87 and 88**). For the *tolB* and *tolQ* mutants, some of the insects were partially colonized by the mutants (**Figure 87**), but the quantification of each population revealed that the wild-type population was extremely abundant and dominated the mutant population in the symbiotic organ (**Figure 88**). Surprisingly, the *rfbC* mutant was more frequently detected in the M4 region than the wild-type strain (**Figure 87**), and the average mutant population quantified was greater (≈ 67%) than the wild-type population (≈ 33%) (**Figure 88**).

By using these *in vivo* quantifications and the initial populations present in the coinoculums, I obtained CI values for each *in vivo* competition assay (**Figure 89**). Concerning the GFP wild-type strain, as it was less abundant in the symbiotic organ the CI values are lower than 1, which means that the GFP wild-type strain was less competitive than the RFP wild-type strain during *in vivo* competition (**Figure 89**). The CI values for competitions between the wild-type strain and hopanoid mutants were close to 1, which indicated that there were no competitions between these strains, except for the *shc* mutant which seems less competitive than the wild-type strain (**Figure 89**). Interestingly, *in vivo* competitions that included the *waaC*, *waaF*, *tolB*, *tolQ*, *wzm* and *rfbA* mutants showed CI values nearly 0, which indicated a significative dominance of the wild-type strain for the colonization of the symbiotic organ (**Figure 89**). Additionally, competitions with the *rpoE* and *wbiF* mutants also demonstrated very low CI values close to 0 which suggested that these two mutants are much less competitive than the wild-type strain to colonize the host's midgut (**Figure 89**). On the contrary, competitions involving the three remaining mutants, *cysA*, *mucD* and *rfbC*, exerted very high CI values superior to 1, especially for the *cysA* mutant with an average CI value close to 5 (**Figure 89**). These CI values indicated that these three mutants were significantly more competitive than the wild-type strain during *in vivo* competitions (**Figure 89**).

In conclusion, these results showed that some bacterial mutants that were able to colonize the host during mono-infections such as the *rpoE*, *wbiF*, *wzm*, *tolB* and *tolQ* mutants lost their

fitness for colonizing *R. pedestris* during mixed infections. Interestingly, I found that the GFP wild-type strain was less competitive *in vivo* than its RFP homologue, which suggest that the insertion of the GFP cassette had an impact on the bacterial fitness. In addition, the *cysA*, *mucD* and *rfbC* mutant strains that showed similar colonization efficiencies than the wild-type strain during mono-infections tended to gain a competitive advantage over the wild-type strain during coinfections. Thus, coinfections are complementary to the mono-infection experiments by unraveling additional fitness traits involved in the competitive abilities to colonize the host.

4.4. Host fitness parameters in infections with AMP-sensitive mutants

During these previous sections, I focused on the characterization of different fitness traits of the *Burkholderia* mutant strains, with both *in vitro* and *in vivo* studies. As this symbiotic interaction is known to confer beneficial effects on the insect host fitness (Kikuchi *et al.*, 2007), I also studied different fitness parameters of the *R. pedestris* adult insects when they were mono-infected by each of the *Burkholderia* mutants. Among these fitness determinants, I have checked the adult emergence rate, the gender, the dry weight, the body size, and the size and width of the abdomen and the thorax parts for each adult *R. pedestris*. Concerning the adult emergence rate, it was previously described that symbiotic insects infected with the wild-type symbiont have a faster developmental rate compared to aposymbiotic insects (Takeshita and Kikuchi, 2017) (see **Chapter I**). My experiments confirmed this and I observed a significant difference of close to four days in the adult emergence rates between symbiotic and aposymbiotic insects (**Figures 90 and 91**). Insects infected with the *cysA* mutant had the same developmental rate than symbiotic insects infected with the wild-type strain, which confirmed that the *cysA* mutant had similar effects than the wild-type symbiont (**Figures 90 and 91**). For the LPS mutants affected in the core oligosaccharide biosynthesis, insects infected with the *waaC* and *waaF* mutants had a similar developing time than aposymbiotic insects, whereas insects fed with the *wbiF* mutant showed a similar developmental rate than symbiotic insects (**Figure 90**). With the hopanoid mutants, only the adult emergence rates of *R. pedestris* insects

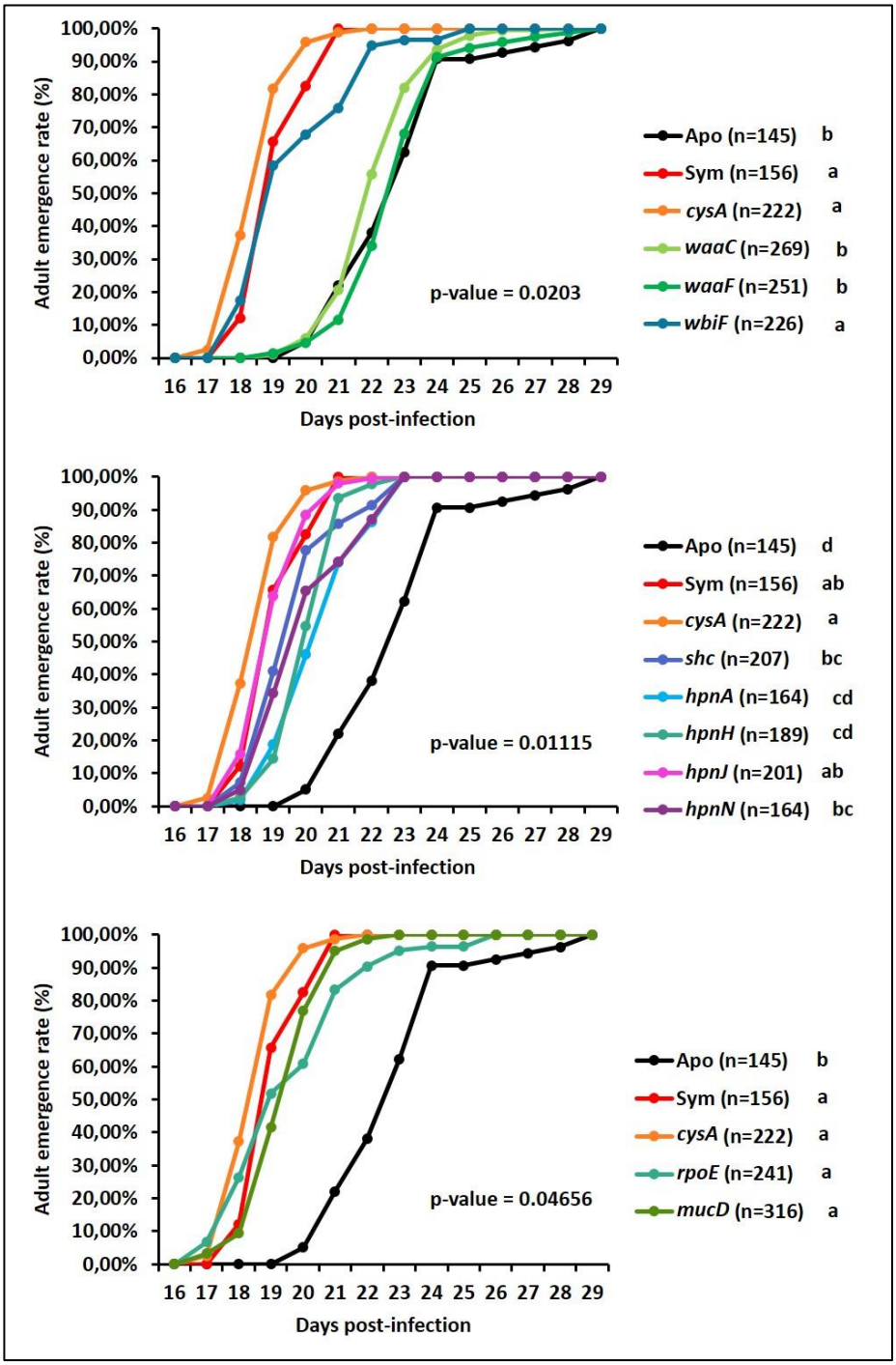


Figure 90: Adult emergence rates of *R. pedestris* insects mono-infected by *Burkholderia* mutant strains of LPS, hopanoids and RpoE ESR pathway.

The number of adult insects was counted each day post-infection for three independent batches of insects infected with each *Burkholderia* mutant, with LPS mutants (upper figure), hopanoid mutants (middle figure) and RpoE pathway mutants (bottom figure). The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters indicated for each condition (on the right) represent statistically significant differences, with the global p-value indicated on each graph (p-value < 0.05, Kurskal-Wallis). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.

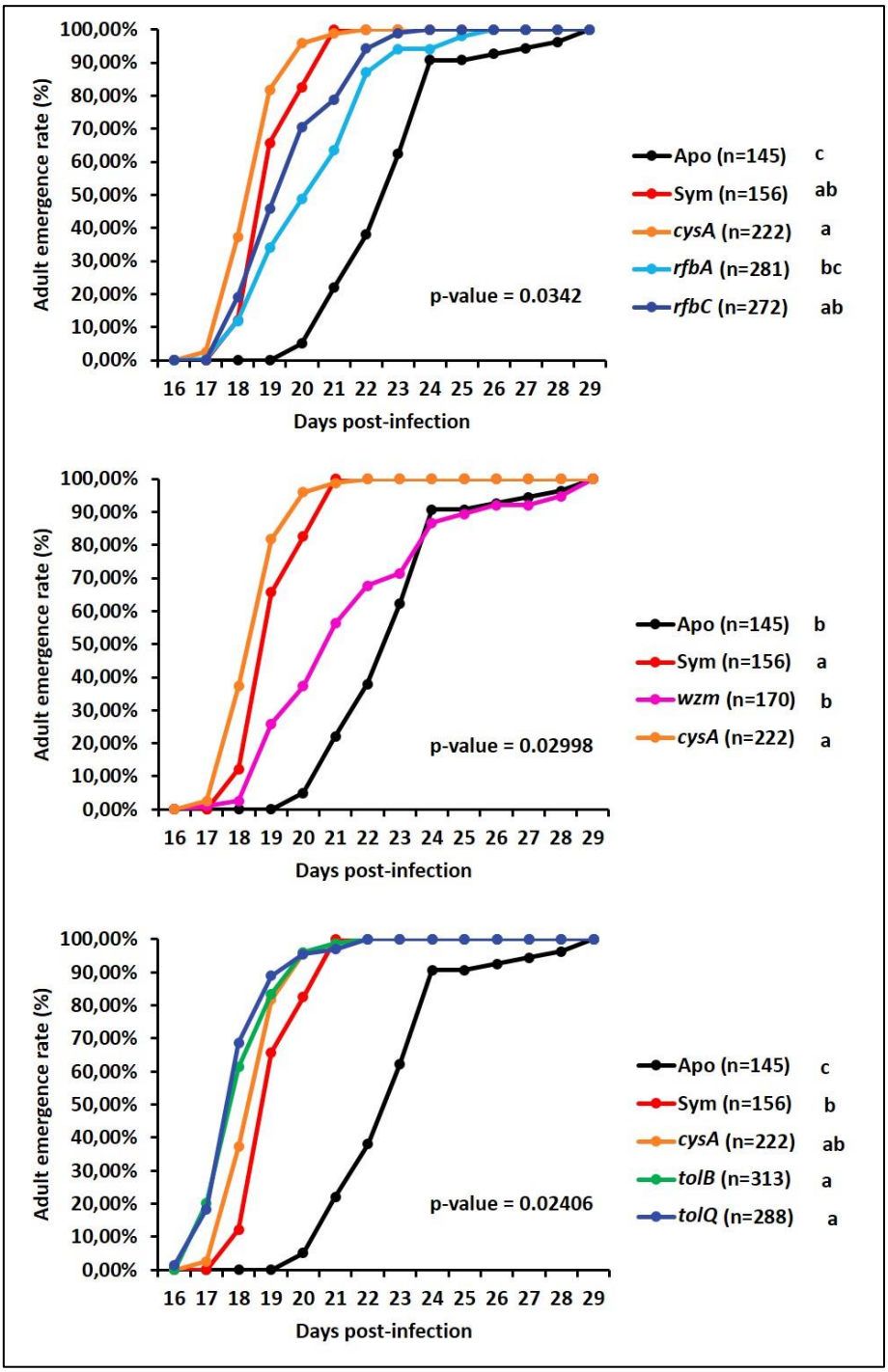


Figure 91: Adult emergence rates of *R. pedestris* insects mono-infected by *Burkholderia* mutant strains of Tn-seq targets.

The number of adult insects was counted each day post-infection for three independent batches of insects infected with each *Burkholderia* mutant, with *rfb* mutants (upper figure), *wzm* mutant (middle figure) and *tol* mutants (bottom figure). The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters indicated for each condition (on the right) represent statistically significant differences, with the global p-value indicated on each graph (p-value < 0.05, Kurskal-Wallis). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.

infected with *hpnA* and *hpnH* mutants were significantly different from insects infected with the wild-type strain, although the delay was small (about one day) (**Figure 90**). However, insects that were fed with the other hopanoid mutants (*shc*, *hpnJ* and *hpnN*) and the RpoE pathway mutants (*rpoE* and *mucD*) showed a fast developing time which was equivalent to symbiotic insects (**Figure 90**). Together, these results matched the previous midgut observations, where insects infected by bacterial mutants that are not able to colonize the M4 region showed an aposymbiotic-like developing time. When insects were infected with the *rfbC*, *tolB* and *tolQ* mutants, they showed the same profile of development than symbiotic insects while *rfbA*-infected insects were slightly delayed in development (**Figure 91**). However, in the presence of the *wzm* mutant, the developmental rate of these insects was close to the adult emergence rate of aposymbiotic insects (**Figure 91**). Thus for these AMP-sensitive strains, it is difficult to make correlations between the ability to colonize the symbiotic organ in the second and the third instar stages and the adult emergence rate, as these mutants showed diverse colonization phenotypes.

In addition, I have checked several morphological parameters that were also measured in previous studies dealing with *Burkholderia* mutants (Kim *et al.*, 2016, 2017). I have checked all the morphometric parameters in male insects and female insects separately (see **Annexes 12 and 13**), as the two genders have their own body mass and body size parameters (Kikuchi *et al.*, 2007; Takeshita and Kikuchi, 2017) (see **Chapter I**). However, the insect populations had equivalent representatives of males and females when I infected them with all the mutant strains (see **Annexe 12**), so I also pooled all the measurements together for male and female insects. The figures below show the data for the pools of males and females, whereas the separate male and female datasets are shown in **Annexe 13**.

I confirmed that symbiotic insects have significant increased morphometric parameters compared to aposymbiotic insects (**Figures 92, 93 and 94, Annexe 13**) (Kikuchi *et al.*, 2007; Takeshita and Kikuchi, 2017). In the presence of mutant strains, the dry weights of insects infected with hopanoid mutants (*shc*, *hpnA*, *hpnH*, *hpnJ*, *hpnN*), *rfb* mutants (*rfbA* and *rfbC*), *tol* mutants (*tolB* and *tolQ*), two LPS mutants (*waaF* and *wbiF*) and *wzm* mutant were equivalent to symbiotic insects (**Figure 92**). Only insects fed with *mucD* and *waaC* mutants presented a significant reduction of their dry weights, which were close to the ones measured for aposymbiotic insects (**Figure 92**). With the *waaC* mutant, this significant decreasing dry

weight was mostly visible on female insects (see **Annexe 13**). Interestingly, adult insects infected with the *rpoE* mutant showed a significant gain of weight compared to wild-type infected insects, which was specifically obvious in female insects (**Figure 92,** see **Annexe 13**). Concerning the entire body size, I noticed that adult insects that were fed with *rpoE*, *hpnJ*, *wbiF*, *wzm* and the two *tol* mutants showed equivalent body size than wild-type infected insects (**Figure 92**). However, insects that were infected with the other mutants, including *mucD*, *shc*, *hpnA*, *hpnH*, *hpnN*, *waaC*, *waaF* and the two *rfb* mutants had similar body sizes than aposymbiotic insects (**Figure 92**). For *rfb* mutants, the reduced size observed within these infected insects was notably observed in the male populations (see **Annexe 13**).

I also checked additional parameters by measuring the size and width of the abdomen and the thorax. For the abdomen size, insects infected with four mutants (*hpnJ*, *wbiF*, *wzm* and *tolQ*) displayed symbiotic fitness traits, whereas *R. pedestris* adults fed with the *mucD*, *hpnN*, *waaC* and *waaF* mutants had similar abdomen sizes than aposymbiotic insects (**Figure 93**). With the other mutants, adult insects had abdomen size values that were intermediate between aposymbiotic and symbiotic insects (**Figure 93**). Interestingly, the abdomen width was significantly reduced in insects infected by the *rpoE*, *mucD* and *rfbA* mutants compared to wild-type infected insects, whereas the other mutants did not change this host fitness parameter (**Figure 93**). Additionally, the decreasing of the abdomen width was noticed in female insects in the presence of the *rfbA* mutant (see **Annexe 13**). Regarding the thorax, hosts that were fed with the *mucD*, *hpnH*, *hpnN*, *waaC*, *waaF* and the two *rfb* mutants exhibited similar sizes than aposymbiotic insects (**Figure 94**). However, the thorax sizes of *R. pedestris* adults infected with *rpoE*, *shc*, *hpnJ*, *wbiF*, *wzm* and the two *tol* mutants were equivalent to the ones measured for insects infected with the wild-type strain (**Figure 94**). Surprisingly, insects infected by the *hpnA* mutant showed a significant reduction of their thorax sizes, notably in females, which were even smaller than aposymbiotic insects (**Figure 94,** see **Annexe 13**). For the other thorax parameter, I observed equivalent thorax width with aposymbiotic insects only for insects infected by the *mucD*, *waaC*, *waaF* and *rfbA* mutants (**Figure 94**). Interestingly, these significant differences of thorax widths were only detected on male populations (see **Annexe 13**).

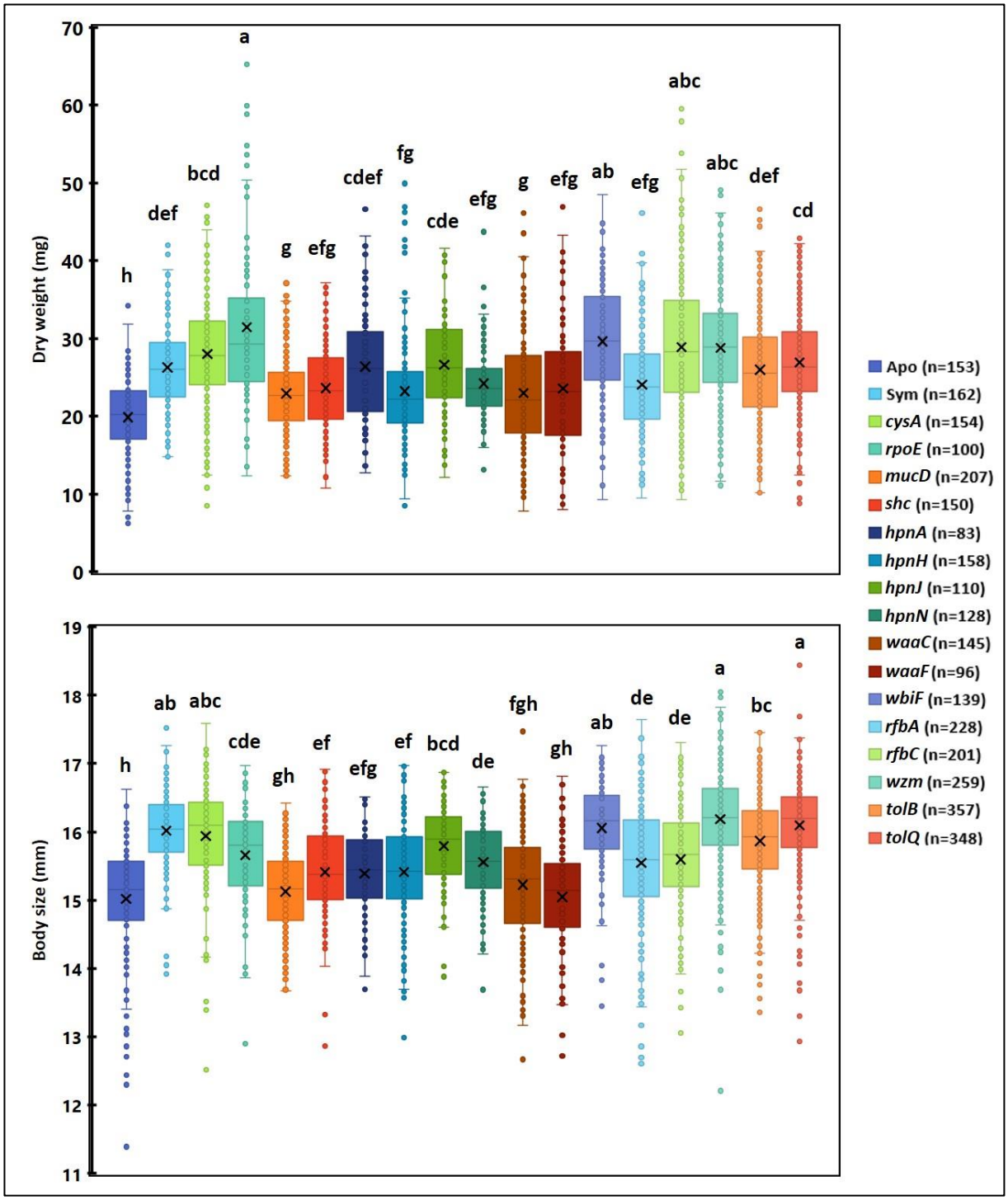


Figure 92: Effects of the *Burkholderia* mutant strains on the body weight and body size of *R. pedestris* adult insects.

Dry weight and body size were measured for each adult insect mono-infected by each *Burkholderia* strain. The mean values are indicated by a black cross on each boxplot. The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters on the top of each boxplot indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey correction). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.

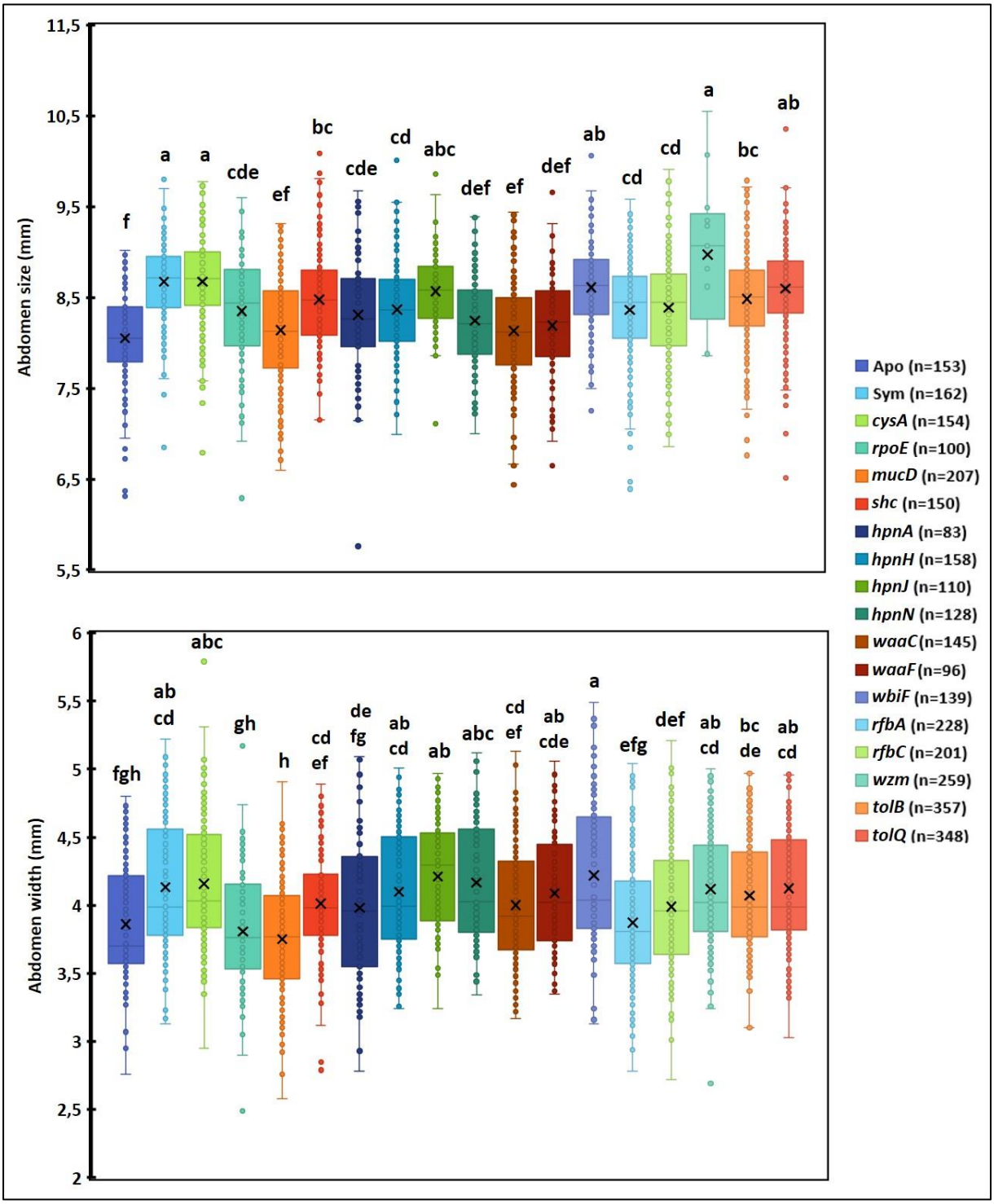


Figure 93: Effects of the *Burkholderia* mutant strains on the abdomen size and width of *R. pedestris* adult insects.

Abdomen size and width were measured for each adult insect mono-infected by each *Burkholderia* strain. The mean values are indicated by a black cross on each boxplot. The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters on the top of each boxplot indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey correction). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.

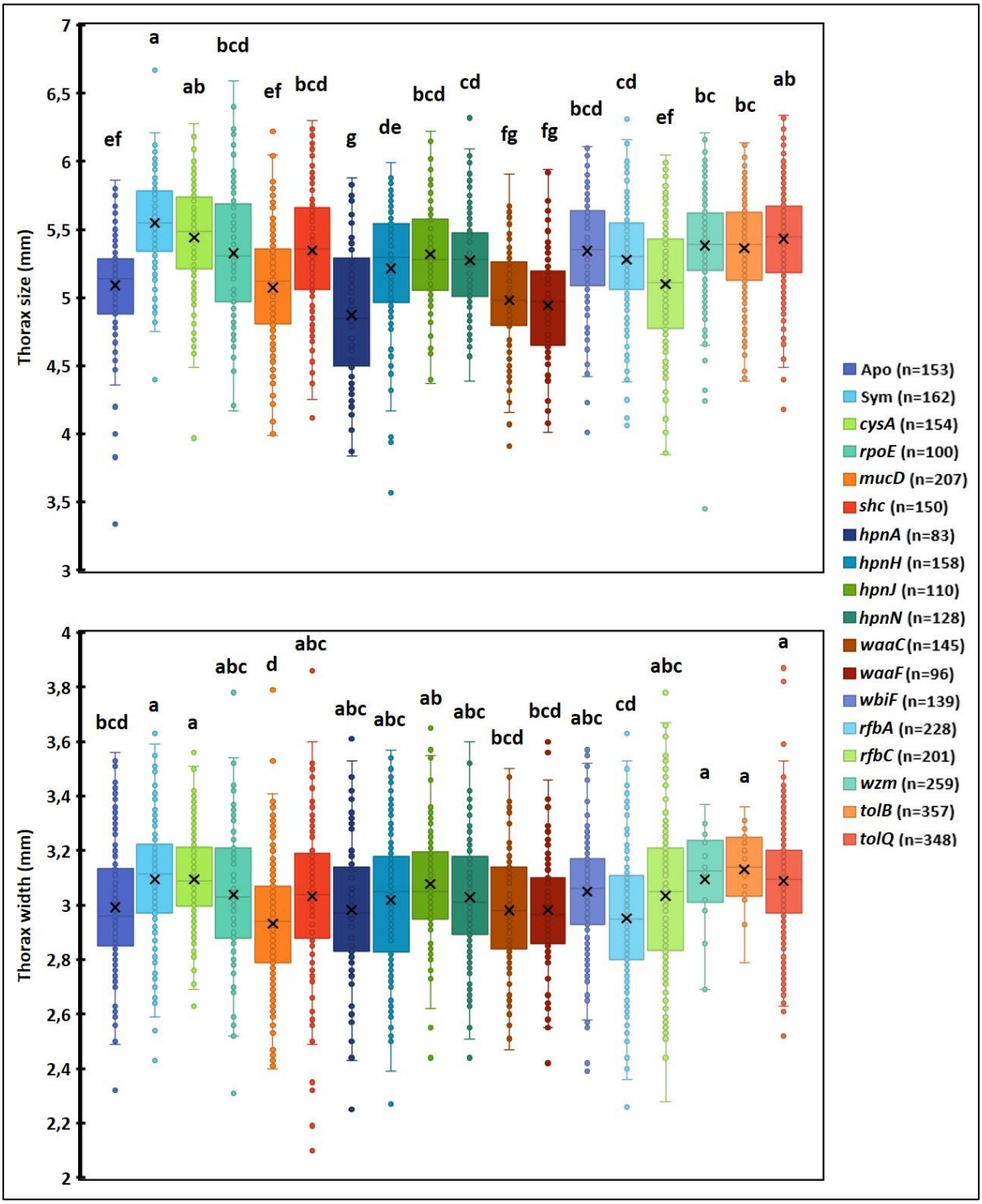


Figure 94: Effects of the *Burkholderia* mutant strains on the thorax size and width of *R. pedestris* adult insects.

Thorax size and width were measured for each adult insect mono-infected by each *Burkholderia* strain. The mean values are indicated by a black cross on each boxplot. The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters on the top of each boxplot indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey correction). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.

In conclusion, I confirmed that the inner core oligosaccharide mutants, *waaC* and *waaF*, triggered the same host fitness parameters than aposymbiotic insects, which was previously demonstrated (Kim *et al.*, 2016, 2017). On the contrary, insects infected with *wbiF* mutant were similar to symbiotic insects, regarding both their development and their morphologies, which was also the case for the *tol* mutants. With the *rfb* mutants, especially for *rfbA*, infected insects had a reduced abdomen width and a smaller body size due to a reduced thorax size, but they have a fast symbiotic-like developmental growth. In the opposite, insects fed with the *wzm* mutant had only an aposymbiotic-like developing time, whereas the other morphometric parameters were not affected. It was generally assumed that the ability of bacterial mutants to colonize efficiently the host was associated to host symbiotic beneficial fitness traits (Kim and Lee, 2015; Kim *et al*., 2016, 2017). However, this statement cannot be applied for the hopanoid and ESR mutants. Even if these mutants are capable to colonize the symbiotic organ, the host showed aposymbiotic fitness traits. For hopanoid mutants, infected insects had smaller body sizes with both reduced abdomen and thorax sizes depending on the mutant. Interestingly, in the presence of the *rpoE* mutant, the insect population showed a fast development similar to symbiotic insects but female hosts gained significant weight despite a reduction of their abdomen widths. This weight gain in female hosts may be explained by an increased production of eggs, however as their abdomen widths looked smaller, this increased body weight might be attributed to another organ. Concerning the last ESR mutant, the *mucD* mutant, it was surprising to see that all the morphometric fitness parameters of the host were aposymbiotic-like, despite a fast developing time and an efficient colonization of the symbiotic organ. Thus, this bacterial mutant may have an impact on the host's growth or on its metabolic activities.

5. Discussion

In the *Burkholderia* genus, many species were reported to be resistant towards a large spectrum of antibiotics, including AMPs (Loutet and Valvano, 2011; Rhodes and Schweizer, 2016). Multiple bacterial functions involving the bacterial membranes were characterized for their protective role towards AMPs (Loutet and Valvano, 2011). As AMPs have also been described to be involved in symbiotic relationships (Mergaert, 2018) and they are abundantly produced in the *R. pedestris* midgut, including in the symbiotic crypts of the M4 region, I

investigated here which bacterial factors were involved in AMPs resistance in the *B. insecticola* symbiont of *R. pedestris*.

In a first approach, I have chosen candidate genes based on previously described bacterial resistance functions in *Burkholderia* strains, and assessed their role in *B. insecticola* with respect to AMP resistance. As AMPs represent a large family of peptides with different physicochemical properties and are produced by various organisms, I have used five AMPs during this study including polymyxin B, LL-37, riptocin and two CCR peptides. Among the selected bacterial targets, there were LPS biosynthesis genes, specifically involved in the synthesis of the core oligosaccharide part (Loutet *et al.*, 2006), genes involved in hopanoid biosynthesis (Malott *et al.*, 2012; Schmerk *et al.*, 2011, 2015) and genes from the ESR RpoE pathway (Flannagan and Valvano, 2008). This candidate gene approach revealed that only mutant strains of the core oligosaccharide of LPS, the *waaC*, *waaF* and *wbiF* mutants, showed a strongly increased sensitivity towards AMPs. Interestingly, the *waaC* and *waaF* mutant strains showed also a decreased motility, which was also previously described in *E. coli* (Wang *et al.*, 2016).

The two other studied pathways, the hopanoid production and ESR, were not involved in AMP resistance in the *Burkholderia* symbiont, except for the ESR pathway that affected riptocin resistance. In addition to confer AMPs resistance in *B. cenocepacia* (Schmerk *et al.*, 2011) and *B. multivorans* (Malott *et al*., 2012), it was shown that hopanoids are required for low pH tolerance and motility (Schmerk *et al.*, 2011; Welander *et al.*, 2012), but this was not the case for *B. insecticola*. Therefore, hopanoids do not seem to play the same protective roles in *B. insecticola* that were described in *B. cenocepacia* and *B. multivorans* (Malott *et al.*, 2012; Schmerk *et al.*, 2011). Concerning the ESR σ^E pathway, it was reported to be required during an osmotic stress and increased temperatures in *B. cenocepacia* (Flannagan and Valvano, 2008). Nevertheless, the RpoE pathway from *B. insecticola* does not have the equivalent functions, and seem to be only required for riptocin resistance. Additionally, as *B. insecticola* also carries genes that are homologous to the genes of the Bae ESR pathway found in *E. coli* (Guest and Raivio, 2016), other types of ESR pathways may be involved in AMP resistance in the *Burkholderia* symbiont. Even if hopanoids and the ESR σ^E pathway are both required for AMPs resistance in *B. cenocepacia* and *B. multivorans*, the results obtained with *B. insecticola* suggest that these bacterial factors may not be considered as general AMP resistance factors

in the *Burkholderia* genus.

By performing mono-infections of *R. pedestris* with these mutant strains, I found that the AMP hypersensitive strains *waaC* and *waaF* were not able to colonize the symbiotic organ in agreement with a previous report (Kim *et al.*, 2017). For the ESR pathway, the *rpoE* mutant which was more sensitive to riptocin only, was perfectly capable to colonize the M4 crypts in mono-infections but was outcompeted by the wild-type in coinfection experiments, demonstrating that this mutant is nevertheless significantly weakened for colonization. Hence, these observations suggest that there is a link between AMP resistance and colonization efficiency. Even if all hopanoid mutants were able to colonize the symbiotic organ, the *hpnH* and *hpnJ* mutant strains showed a decreased proliferation rate in the M4 crypts compared to the wild-type strain. As I could not link hopanoids to AMP resistance in the *B. insecticola* strain, these results suggest other functions of hopanoids in the colonization of the host's midgut.

The candidate gene approach highlighted that the *Burkholderia* symbiont does not share the same bacterial resistance factors with related species from the *Burkholderia* genus. This suggests that several strain-specific functions control AMP resistance in *Burkholderia*. Therefore, to obtain a genome-wide overview of AMPs resistance functions in *B. insecticola*, I have used a Tn-seq approach to identify bacterial genes required for the bacterial fitness in the presence of the five different AMPs that were previously used in this study. By using the Tn-seq methodology, I identified 42, 42, 15, 21 and 39 bacterial fitness genes (including both conditionally-essential and domain conditionally-essential genes) that were required for fitness in the presence of polymyxin B, LL-37, riptocin, CCR179 and CCR480 peptides, respectively. Interestingly, the encoded bacterial functions were mostly related to cell wall and membranes biogenesis, which constitute the main known targets for AMPs (Brogden, 2005; Kumar *et al.*, 2018). In a previous Tn-seq study conducted with *S. meliloti*, a majority of cell wall components were also identified as resistance factors against polymyxin B and one NCR peptide (Arnold *et al.*, 2017). Among the fitness genes found for the *Burkholderia* symbiont, only three of them were shared between the five AMP conditions. They are encoding the three subunits of the Tat system. This transporting system was shown to be involved in biofilm formation, flagellar motility and *in vivo* colonization in *V. cholerae* (Zhang *et al.*, 2009), but also in the maintenance of the outer membrane stability in *E. coli* (Ize *et al.*,

2003). Except this transporter, many other bacterial factors were shared between different sets of AMPs which mostly included the LPS biosynthesis process. As LPS are negatively charged molecules, they constitute the privileged interacting site for cationic AMPs to target bacterial membranes (Kumar *et al.*, 2018; Loutet and Valvano, 2011). Indeed, I have found multiple fitness genes that are encoding for the O-antigen biosynthesis, such as multiple glycosyl transferases, the O-antigen polymerase, the dTDP-L-rhamnose biosynthesis, and the export of O-antigen through an ABC transporting system Wzm/Wzt. Interestingly, an *Erwinia amylovora* mutant strain of the O-antigen polymerase, also known as O-antigen ligase, was reported to be more sensitive towards polymyxin B (Berry *et al.*, 2009). Additionally, fitness genes involved in the core oligosaccharide part of LPS molecules were also identified to be required for AMP resistance, and interestingly I found the *waaC, waaF* and *wbiF* genes that were already selected in the candidate gene approach. In addition, several of the fitness genes obtained by Tn-seq were confirmed for their role in AMP resistance, hence testifying the robustness of the Tn-seq methodology applied on *B. insecticola*. However, I observed that there are exclusive sets of fitness genes required for each AMP, especially for the CCR480 peptide which only shared the *tatABC* genes with the other AMPs. As AMPs have specific physicochemical properties (Brogden, 2005), it was not surprising to obtain gene specificities for each AMP tested. The differences between the fitness gene sets required for CCR179 and CCR480 peptides may be linked to these differences in physicochemical properties. Knowing that there are 97 CCR peptides produced by *R. pedestris* in the M4 crypts (Futahashi *et al.*, 2013), it is not unlikely that the other bacterial functions have to be discovered. Still among the peptide-specific functions, the protein quality control process was specific to the polymyxin B condition, cytochrome c biogenesis proteins were involved in CCR peptides resistance, and one protease known as oligopeptidase B was specific to riptocin resistance. In the presence of both CCR peptides, it appeared that metabolic activities are also involved in CCR peptides resistance, in addition to membrane components. These metabolic functions may indicate that CCR peptides could have both membrane and intracellular targets, and that each CCR peptide possesses its own range of activities. Such a diversity of actions was also proposed and demonstrated in some cases for the legume NCR peptides, which similarly as the CCR peptides, show a very high diversity (Mergaert, 2018).

From these fitness gene lists, I have targeted five genes including one gene from the Wzm/Wzt

O-antigen transport system (*wzm*), two genes involved in dTDP-L-rhamnose biosynthesis (*rfbA*, *rfbC*), and two genes encoding the Tol-Pal complex (*tolB*, *tolQ*) that were commonly identified for several AMPs tested. It appeared that all these bacterial mutants of *B. insecticola* were hypersensitive towards AMPs, hence confirming the Tn-seq analysis. Interestingly, these AMP-sensitive strains showed different defects of host colonization, with the *rfbA* strain that was not able to colonize the symbiotic organ, and the other strains that showed intermediate colonization phenotypes. Among these *in vivo* phenotypes, the *wzm* and *rfbC* mutants displayed a delayed colonization of the M4 region and could infect only 50 to 60% of the total insect population. Interestingly, the *tolB* and *tolQ* mutant strains are able to colonize the total insect population, but after molting to the third instar stage, the mutants were localized only at the posterior part of the M4 region in 50% of the insect population. In a previous experiment, the expression of several CCR peptide genes in the M4 organ were checked by *in situ* hybridization and it was shown that these CCR peptides are less expressed in the same last few posterior crypts of the M4 region where the tol mutant population is remaining in the third instar crypts (Ohbayashi T., unpublished data). This suggests that the collapse of the *tol* mutant population in the majority of the third instar crypts is linked to a local high expression of CCR peptides and the hypersensitivity of these mutants to AMPs. Overall, these results strongly suggest that there is a correlation between the capacity of the symbiont to resist AMPs and its ability to colonize the host with high efficiency.

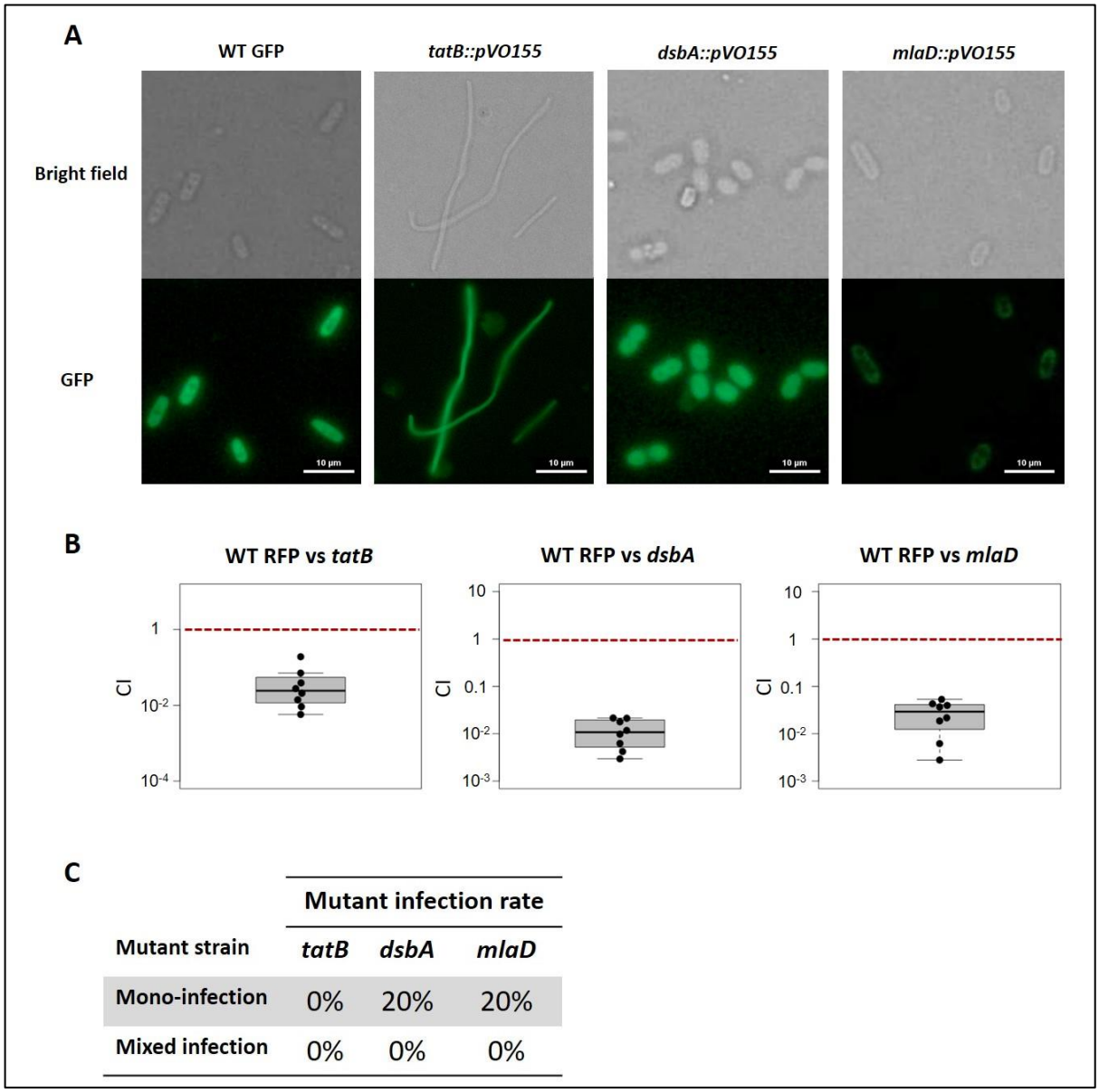
For completeness, the reduced motility of several of the strains discussed above (except the *rfbC* mutant) has to be considered in discussing their symbiotic phenotype. Flagellar motility is indeed a key function for colonization of the M4 crypts, and in particular to enter the symbiotic organ through the constricted region (Lee *et al.*, 2015; Ohbayashi *et al.*, 2015). At later stages of colonization of the M4 crypts, motility is not needed anymore and the symbiotic bacteria lose altogether their motility in the crypts (Ohbayashi *et al.*, 2019). However, the reduced motility of these mutants is not likely to be the cause of their symbiotic defect because some of these reduced motility mutants do initially colonize the crypts (*e.g.* the *tolB* and *tolQ* mutants) and also other mutants with similar reduced motility are not affected in symbiosis (Ohbayashi *et al.*, 2015) (see **Chapter IV**). Only the completely non-motile mutants do not colonize the symbiotic organ (Ohbayashi *et al.*, 2015).

In an environmental context, wild insects have to acquire their symbiont from a broad range

of bacterial species in their environment (Kikuchi *et al.*, 2007; Takeshita and Kikuchi, 2017). Hence, the *Burkholderia* symbiont has to be selected from this diverse bacterial population by the host, which also indicates that the symbiont has to compete with the other bacterial species before colonizing the symbiotic organ. To take into account this competition parameter, I have performed *in vivo* competition experiments to assess the colonization abilities of all the mutant strains in a mixed infection context with the wild-type strain. Consistently, the mutant strains that could not colonize the host during mono-infections, such as *waaC*, *waaF* and *rfbA* mutants, were also unable to colonize it during mixed infections. But more importantly, I showed that mutants that displayed intermediate colonization phenotypes during mono-infections, such as the *wzm*, *tolB* and *tolQ* mutants, were totally outcompeted by the wild-type strain during *in vivo* competitions. However, three mutant strains including *cysA*, *mucD* and *rfbC* mutants showed an *in vivo* fitness gain compared to the wild-type strain. The *cysA* mutant was used as a positive control of infection during mono-infections, especially because this mutant was generated with the same mutation strategy that I have used in this study. Based on mono-infection experiments, this mutant strain was not further studied for its *in vivo* phenotypes as its colonization efficiency looked similar to the wild-type strain. However, the competitiveness of mutant strains enables to describe their behaviours in a population context. As its competitiveness *in vivo* was not assessed before, it was the first observation of a cysteine biosynthesis mutant being capable to outcompete the wild-type strain in the M4 region. It was surprising to notice that *rfbA* and *rfbC* mutants did not exert the same *in vitro* and *in vivo* phenotypes, despite the fact that these two mutated genes belong to the dTDP-L-rhamnose biosynthesis gene cluster. Indeed, the *rfbA* mutant showed a stronger hypersensitivity towards AMPs compared to the *rfbC* mutant, and the *rfbA* mutant could not colonize the host whereas the *rfbC* mutant gained a competitive advantage for the colonization of the symbiotic organ. The *rfbA* gene encodes the glucose-1-phosphate thymidylyltransferase, which is the first enzymatic step of dTDP-L-rhamnose biosynthesis pathway, and the *rfbC* gene encodes the dTDP-4-dehydrorhamnose 3,5-epimerase which corresponds to the third enzymatic step to produce dTDP-L-rhamnose (Tsukioka *et al*., 1997). As an *rfbC* mutant could still produce intermediate compounds of the dTDP-L-rhamnose biosynthesis, such as dTDP-D-glucose, a possibility to explain the phenotype differences between these two *rfb* mutants is that these intermediate molecules produced in the *rfbC*

mutant are used as precursor compounds for the biosynthesis of an alternative O-antigen, which was described in thermophilic and lactic bacteria (Pföstl *et al.*, 2008). Thus, it would certainly be of interest to characterize in the near future biochemically the LPS molecules produced by the *rfbA* and *rfbC* mutants. In addition, none of these genes were reported to play a role in AMPs resistance in the literature, however an *E. coli rfbC* mutant was reported to be more sensitive towards mitomycin C and UV irradiation (Han *et al.*, 2010). These two features could be tested as future perspectives to describe more *in vitro* phenotypes on the *B. insecticola rfbC* mutant.

As this symbiotic association provides benefits for the host's development (Kikuchi *et al*., 2007; Takeshita and Kikuchi, 2017), I also investigated the effects of mono-infections with these mutant strains on the different host fitness traits, including morphometric parameters and the adult emergence rate. With these measurements, I confirmed previous results which showed that insects infected by *waaC* or *waaF* mutants, strains that were unable to colonize the host, have aposymbiotic fitness traits (Kim *et al.*, 2017). Concerning the *wbiF*, *tolB* and *tolQ* mutants, the infected insect populations had the same fitness parameters than the insect population infected with the wild-type strain. Interestingly, even if the *mucD* mutant was able to colonize efficiently the host during mono-infections and also gained a competitive advantage during coinfections to colonize the symbiotic organ, the infected insects displayed aposymbiotic-like morphometric traits but with a fast-developing time equivalent to symbiotic insects. These growth deficiencies of the insects infected with the *mucD* mutant could be attributed to a lack of nutrients or vitamins that are not provided by the symbiotic mutant anymore. In addition to its serine protease activity, the MucD protein was also shown to play a role of chaperone at low temperatures in *E. coli* (Spiess *et al*., 1999). Thus, this *mucD* mutant may have an impaired metabolism which could either provide insufficient nutrient loads for the host insect or produce unfolded proteins that are not well digested and assimilated to sustain the growth of *R. pedestris*. It would be interesting to perform metabolomics analyses on this *mucD* mutant and on the symbiotic organ of these infected insects, and comparing these results with the wild-type strain and symbiotic insects. On the opposite of the *mucD* mutant-infected insects, insects fed with the *wzm* mutant showed a longer developmental rate similar to aposymbiotic insects, but with symbiotic-like growth parameters. Additionally, insects that were infected with hopanoids and *rfb* mutants had a fast adult emergence rate



C

	Mutant infection rate		
Mutant strain	*tatB*	*dsbA*	*mlaD*
Mono-infection	0%	20%	20%
Mixed infection	0%	0%	0%

Figure 95: Preliminary results obtained from *B. insecticola* mutants targeted in fitness genes identified by Tn-seq in the presence of AMPs.

A) Observations of *in vitro* bacterial cell morphologies. The scale bar (white) indicates 10 µm. B) Competitive indexes (CI) for each *in vivo* competition were obtained by counting the number of RFP and GFP fluorescent bacteria in the symbiotic organ by flow cytometry. The red dotted line indicates a CI equals to 1, which means no competition between the wild-type strain and the mutant strain. When a CI is inferior to 1, the wild-type strain is more competitive than the mutant strain. When a CI is superior to 1, the mutant strain becomes more competitive than the wild-type strain. C) Resume table of the mutant infection rates obtained from the observations of GFP fluorescence in symbiotic organs of mono-infections and mixed infections at three days post-infection in ten insects at the second instar stage. Abbreviations: WT: wild-type.

but with smaller body sizes. With the *rpoE* mutant, I observed that female insects gained significantly more weight than symbiotic female insects, which could be due to the increasing

number eggs produced inside the female's body. Recently, it was reported that the transcript levels of two hormones especially produced by *R. pedestris* females, hexamerin and vitellogenin, are increased in the presence of the gut symbiont (Lee *et al.*, 2017). Thus, it could be interesting to check the transcription level of these hormones when insects are infected by the *rpoE* mutant. As the production of eggs is increased in symbiotic insects and constitutes an additional fitness trait (Kikuchi *et al.*, 2007; Lee *et al.*, 2017), it could be interesting to check the number of eggs laid by females infected with the different mutant strains.

In conclusion, the phenotypes of the here analysed mutants, including colonization capacity, level of colonization and competition for M4 colonization, as well as the host morphometric and developmental parameters, did not show an *a priori* expected simple pattern but strongly suggested that these phenotypes are uncoupled during the host's development. Possibly, the different colonization dynamics of each of these mutants combined with altered "nutritional services" provided by these mutants to the host induces, each time, a specific nutritional and hormonal profile in the host that affects in different ways the host development that deviates strongly from the previously known profile of symbiotic and aposymbiotic insects.

During this study, I have evaluated five fitness genes identified from the Tn-seq results with AMPs. However, I have targeted additional *Burkholderia* fitness genes that are shared between AMPs or that are unique for a specific AMP. This work is currently in progress in the laboratory, but Christy Calif, a Master 2 student that I supervised during my thesis, obtained *Burkholderia* mutants in the *tatB* (*BRPE64_RS12015*), *dsbA* (*BRPE64_RS00670*) and *mlaD* (*BRPE64_RS12120*) genes (**Figure 95**). As I mentioned before, the *tatB* gene was one of the fitness genes shared between the five AMPs tested and encodes the TatB subunit of the Tat system that is involved in outer membrane stability (Ochsner *et al*., 2002; Robinson *et al.*, 2011a; Zhang *et al.*, 2009). The *dsbA* gene was involved specifically to polymyxin B resistance and encodes a periplasmic protease involved in protein quality control (Manta *et al.*, 2019; Meehan *et al.*, 2017). The third gene, *mlaD*, that is commonly required for polymyxin B and LL-37 resistance, encodes a subunit of an ABC lipid transporter which is involved in the outer membrane stability (Bernier *et al*., 2018). Interestingly, we found that the *tatB* mutant exerted an exaggerated elongation of cell morphology compared to the rod-shape morphology of the wild-type strain (**Figure 95A**). A similar morphology was previously observed in a *tatC* mutant of *E. coli* which exerted a chain-forming cell morphology (Ize *et al*., 2003). As preliminary

results obtained from *in vivo* colonization experiments, we found that the *tatB* mutant was unable to colonize the host neither during mono-infections nor during coinfections with the wild-type strain, with CI values close to 0 (**Figure 95B-C**). Additionally, we noticed that the *dsbA* and *mlaD* mutants colonize less efficiently the symbiotic organ during mono-infections (**Figure 95C**), and they were completely outcompeted by the wild-type strain during mixed infections, with CI values close to 0 (**Figure 95B-C**). Thus, these preliminary results strongly suggest that these three additional genes are also required for the host colonization in *B. insecticola*. Concerning the AMP sensitivity, we tested only the *dsbA* mutant so far which showed an exclusive hypersensitivity towards polymyxin B, in agreement with the Tn-seq results indicating that this fitness gene was only required for polymyxin B resistance. This work is still ongoing and both *in vitro* and *in vivo* phenotypes are currently being tested for the other *Burkholderia* mutant strains obtained to validate this Tn-seq study. Nevertheless, it seems that the correlation in *B. insecticola* between AMP resistance and the capacity to colonize the midgut of *R. pedestris* in *B. insecticola* is further strengthened by the characterization of these additional mutants.

Chapter IV

Host Colonisation Functions of *Burkholderia insecticola* Identified by *in vivo* Tn-seq

1. Introduction

Similar to adherence and virulence factors in pathogenic bacterial species, symbiotic factors are critical for colonization and maintenance of symbiotic bacterial populations inside their respective host. In *R. pedestris*, different studies were conducted to find symbiotic bacterial genes which participate to the host colonization of the symbiotic organ. Based on screening of Tn*5* transposon mutagenesis libraries of the *Burkholderia* symbiont directly inoculated to the insect host or on a proteomics approach, several *Burkholderia* mutants were shown to have lost partially or entirely their colonization capability. These symbiotic-deficient genes were involved in motility (Lee *et al.*, 2015; Ohbayashi *et al.*, 2015), purine biosynthesis with the *purL* gene (Kim *et al.*, 2014), polyhydroxyalkanoate (PHA) granules biosynthesis with the *phaABC* gene cluster (Jang *et al.*, 2017; Kim *et al.*, 2013a) and the biosynthesis of extracellular elements such as the LPS and the peptidoglycan (Kim *et al.*, 2013b, 2016, 2017) (see **Chapter I**).

Even if some critical bacterial factors for the host colonization were found, the global genetic repertoire required for a successful colonization process in *R. pedestris* remains uncharacterized. In order to identify these colonization factors on a genome-wide scale, the Tn-seq methodology was previously applied *in vivo* on different symbiosis model bacteria, such as *Snodgrassella alvi* in honey bees (Powell *et al.*, 2016), *V. fischeri* in the Hawaiian bobtail squid (Brooks *et al.*, 2014) and very recently *Borrelia burgdorferi*, the causative agent of Lyme disease in humans, in the midgut of *Ixodes* tick vector (Phelan *et al.*, 2019). In honey bees (*Apis mellifera*), the gut microbiota consists of approximately eight different bacterial species with one dominant member identified as *S. alvi* (Powell *et al.*, 2016). Based on *in vivo* Tn-seq method, Powell *et al.*, found that 399 genes (out of 2,226 total genes) were required for honey bee gut colonization, covering cell wall biogenesis functions, metabolic activities of specific amino acids and nucleic acids biosynthesis pathways, and many stress response elements (Powell *et al.*, 2016).

Concerning the *Vibrio*-squid symbiosis model, the *in vivo* Tn-seq approach revealed that 380 genes (out of 3,828 total genes) from *V. fischeri* were characterized as colonization factors (Brooks *et al.*, 2014). Similar to *S. alvi*, it was shown that biofilm formation and stress response mechanisms, including chaperones, are key functions necessary for the establishment of a

successful colonization of the squid light organ (Brooks *et al.*, 2014; Lyell *et al.*, 2017).

B. burgdorferi has a complex biphasic life cycle, which alternates between the midgut of ticks and the blood of vertebrate hosts. A Tn-seq screen on this bacterium during the colonization of the insect gut revealed about 100 genes that were absolutely essential or that affected the bacterial fitness (Phelan *et al.*, 2019). Strikingly, about half of these genes encode proteins of unknown function, with another portion of these genes that encode membrane-associated proteins and also genes important for the resistance to reactive oxygen and nitrogen species (Phelan *et al.*, 2019).

In addition to these symbiosis models, several pathogens were investigated to characterize colonization factors in rodent models, such as *B. pseudomallei* (Guttierez *et al.*, 2016), and there is a similar genetic pattern than in symbiotic bacteria for host establishment (Fu *et al.*, 2013; Gutierrez *et al.*, 2015; Skurnik *et al.*, 2013; Wang *et al.*, 2014).

Based on these previous successful *in vivo* applications, Tn-seq is an attractive strategy to determine the full gene repertoire of *B. insecticola* that is essential or that contributes to the colonization of the *R. pedestris* stinkbug symbiotic organ. However, a limitation of the Tn-seq approach that has to be taken into consideration is its sensitivity to bottlenecks. Biological bottlenecks consist of a sharp constriction of the population size which alters the population composition caused by stochastic sampling of certain genotypes in the population and not by fitness parameters (Abel *et al.*, 2015). Organisms that are subjected to severe bottlenecks are infectious microbes, pathogens or symbionts, that often multiply enormously within their hosts, starting from small founding populations (**Figure 96**) (Abel *et al.*, 2015; Chao *et al.*, 2016). A case of an extreme bottleneck is happening during the infection of legume plants by *Rhizobium* bacteria. Nodules are often pretended to be colonized by a (nearly) clonal population of rhizobia, originating from a single or few founder cells (Goormachtig *et al.*, 2004; Kondorosi *et al.*, 2013; Remigi *et al.*, 2016). These bottlenecks are especially problematic for *in vivo* Tn-seq experiments with transposon mutant libraries, where the sample sizes collected after host colonization may be affected by host barriers and may not reflect the initial composition of the library (Chao *et al.*, 2016). In addition to physical barriers, host-imposed bottlenecks comprise innate and adaptive immune mechanisms, limitations of specific nutrients, the environment availability and accessibility, and competitions with local settled microorganisms (Abel *et al.*, 2015). For example, in the *Vibrio*-squid interaction, the size of the

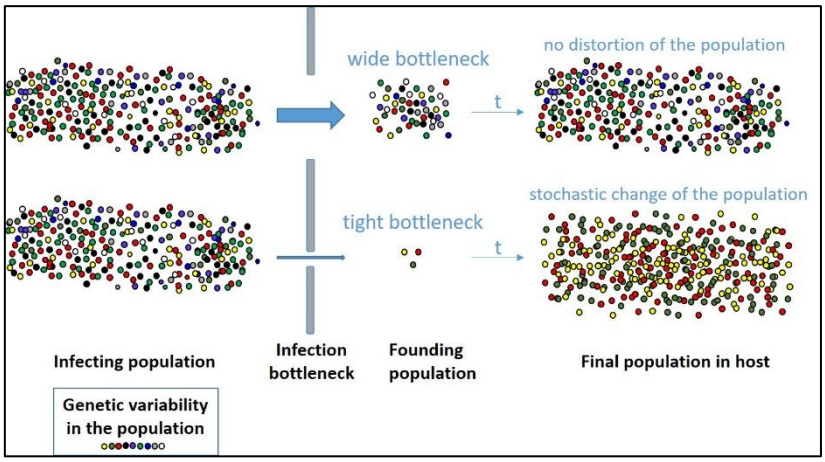


Figure 96: Schematic presentation of bottlenecks and their impact on the genetic diversity of the population.

Individuals in the population are represented by circles and their genetic diversity is indicated by different colors. Bottlenecks will reduce the population size (number of circles) and the founding population will give rise to a new population after replication. Wide bottlenecks lead to limited or no changes in the genetic makeup of the population that developed from the post-bottleneck founding population. In contrast, tight bottlenecks lead to stochastic loss of many markers and substantial changes in the genetic makeup of the final population.

pores in the light organ (Guerrero-Ferreira and Nishiguchi, 2009), the low pH in the stomach and the presence of AMPs constitute inherent bottlenecks for the *Vibrio* symbiont (Heath-Heckman *et al.*, 2014; Mandel and Dunn, 2016).

To estimate these bottleneck sizes, the simplest method is to count the number of bacterial cells recovered after host infection by CFU counting or by microscopy (Abel *et al.*, 2015). However, these methods do not take into account the genetic composition of the bacterial population, which is shaped by these bottlenecks (Abel *et al.*, 2015). With the Tn-seq method, these bottlenecks can be evaluated by determining the number of unique insertions recovered after host colonization with a precise initial inoculum (Brooks *et al.*, 2014; Stephens *et al.*, 2015). In the *Vibrio*-squid model, it was estimated that 250 animals had to be sacrificed in order to sample at least 20,000 mutants for an inoculated population of 2.10^5 CFU.mL^{-1}, based on the Tn-seq method (Brooks *et al.*, 2014).

There are three observations that suggest that the infecting population of *B. insecticola* in the *R. pedestris* midgut may also experience a bottleneck. Firstly, the constricted region that connects the M3 region of the midgut with the symbiotic M4 region, and that fulfils a sorting function by selecting the proper symbiont from the ingested bacteria, is a very narrow channel of only a few micrometers in diameter (Ohbayashi *et al.*, 2015). This suggests that the number

of bacteria that can pass through the constricted region during an infection process is limited. Secondly, the "gate" that is formed by the constricted region is only open for a few hours after the initial infection event (Kikuchi and Fukatsu, 2014). Indeed, it was shown by microscopy that the constricted region and the M4B closes after about 12 to 18 hours after infection (Kikuchi and Fukatsu, 2014; Ohbayashi *et al.*, 2015). This closure was further confirmed by double infection experiments with differently marked strains (*e.g.* GFP-labelled and RFP-labelled strains) that were fed to the insects at different time intervals. About 15 to 18 hours after a first infection, a second infection of the symbiotic organ is not possible anymore (Kikuchi, Ohbayashi, and Mergaert, unpublished data). And thirdly, as mentioned in the previous chapter, the infecting population is confronted with a strong challenge of AMPs when entering the midgut crypts, which could further restrict the founding population for the crypt colonization (see **Chapter III**). Taken together, the parameter of a possible infection bottleneck during *R. pedestris* infection has to be taken into account for the *in vivo* Tn-seq experiments.

Here, I describe the use of Tn-seq to create a full genome picture of the fitness landscape of *B. insecticola* inside the symbiotic organ of *R. pedestris*. I have first estimated the infection bottleneck size of *Burkholderia* symbiotic population after colonizing its host *R. pedestris* for a given precise initial inoculum. Thanks to these data, I have performed an *in vivo* Tn-seq approach to identify bacterial fitness genes involved in the colonization of different midgut compartments, including the symbiotic organ of *R. pedestris*. Hence, this work was able to settle *in vivo* Tn-seq experiments and to pinpoint the bacterial symbiotic functions involved in the *B. insecticola*-*R. pedestris* symbiosis.

2. Contributions

Quality control and sequencing of the Tn-seq samples were performed by the I2BC sequencing platform (CNRS Gif-sur-Yvette, France). The collections of the M4 organs of dissected *R. pedestris* were performed with the help of two members of the team, Tsubasa Ohbayashi (postdoctoral student) and Raynald Cossard (assistant engineer). Tsubasa Ohbayashi further contributed to the bottleneck assessment by mixed infection and the phenotypic characterization of the chemotaxis mutant.

3. Materials and methods

Bacterial strains, growth conditions and insect rearing

The growth of the *B. insecticola* strain RPE64, the growth of the Tn-seq transposon mutant library, the insect rearing and infections of insects were performed essentially as it is described in **Chapters II and III sections 3**.

Bottleneck assessment by Tn-seq

Two days after birth, second instar nymphs of *R. pedestris* were deprived of water for one day to make them thirsty, facilitating subsequent infection. One hundred second instar nymphs of three days-old were transferred individually in 1.5 mL Eppendorf tubes with 1 µL water droplet containing 10^6 CFU of the *B. insecticola* RPE75 Tn-seq library. After five hours, insects that drank the bacterial suspension droplet were transferred in a Petri dish containing soybean seeds and sterile water, and kept at 25°C. At three days post-infection, at the second instar larval stage, 60 insects were sacrificed and dissected under a binocular microscope in sterilized PBS solution (170 mM NaCl, 3.3 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4). M4 organs were harvested individually and collected in 250 µL of sterilized PBS solution. Each M4 organ was homogenized with a plastic pestle and after homogenization, the pestle was washed with 250 µL of sterilized PBS solution. A fraction of each M4 crushed organ suspension was subjected to serial dilutions to assess CFU counting per insect. Each remaining suspension was spread onto YG agar plates and incubated for two days at 28°C to multiply the symbiotic population. As the bacterial load present in one insect is not high enough for an efficient DNA extraction, this *in vitro* growing step was necessary to increase the bacterial biomass for DNA extraction and to obtain the required total amount of bacterial DNA for sequencing. After incubation, the bacteria grown onto these 60 YG plates, corresponding to each dissected insect, were resuspended in 2 mL of YG medium. Each bacterial solution was centrifuged at 4000 g for 15 minutes at room temperature and pellets were stored at -20°C until DNA extraction. DNA extraction, library preparation, sequencing and bioinformatics were done as described before in **Chapter II section 3**. Samples from each insect were marked with a specific barcode during the synthesis of the Illumina sequencing library. The pooled Illumina libraries of all insects were sequenced in one run. The bioinformatics analysis of the sequencing reads

consisted in the counting of the number of different sequence reads identified per sample rather than counting the number of times each sequence was obtained as it is done in the usual Tn-seq analysis (see **Chapter II section 3**). The number of different sequences obtained in each insect corresponds to the minimal number of bacterial infection events in that insect, and thus to the minimal estimate of the bottleneck. The real bottleneck could be higher taking into account that multiple specimens of the same bacterial mutant could have infected the symbiotic organ. However, since the determined values were much lower than the complexity of the library, the latter factor should be low and the number of different mutants identified in the individual insects should be close to the true bottleneck.

Bottleneck assessment by mixed infections

To estimate with an independent method how many symbiont cells can infect the symbiotic region before the midgut closure, co-inoculation of the GFP-labelled strain RPE225 (Rif^R,Km^R) and the non-labelled strain RPE75 (Rif^R) was performed. The non-labelled symbiont and GFP-labelled symbiont were mixed in different ratios ranging from [1:10] to [1:20,000] [GFP-labelled symbiont:non-labelled symbiont]. Symbiont cells were diluted in distilled water to obtain a cell density of 10^5 CFU.μl^{-1} or 10^6 CFU.μl^{-1}, and second instar insects were fed with 1 µl of these suspensions. At three days post-infection, the M4 symbiotic organ was dissected and observed under the microscope for the detection of the GFP signal. The organs were further crushed with a plastic pestle and ten-fold dilution series were plated on YG plates containing kanamycin (Km) 30 µg.ml^{-1} to check whether GFP-labelled symbionts entered the symbiotic organ. Insects were counted as positive for infection with the GFP-labelled symbiont if at least one colony of the GFP-labelled symbiont was detected.

Determination of the bacterial population in the M1, M2, M3, M4B and M4 organs

The size of the bacterial population in the M1, M2 and M3 regions of the midgut was determined one, two and three days after feeding of the insects with *B. insecticola*. To do so, second instar nymphs of *R. pedestris*, two days after birth, were deprived of water. Insects were placed individually in 1.5 mL Eppendorf tubes containing a 1 µL water droplet with 10^6 CFU of the *B. insecticola* RPE75 symbiont. After drinking, insects were further maintained at

25°C in a Petri dish containing soybean seeds and sterile water. Ten insects were dissected at one, two and three days post-infection and M1, M2, M3, M4B and M4 organs were harvested separately for each insect in 100 µL of PBS buffer. Organs were crushed with a plastic pestle, ten-fold dilution series were prepared and 10 µL of each dilution were spotted on YG medium for CFU counting.

In vivo Tn-seq screening

Two days after birth, second instar nymphs of *R. pedestris* were deprived of water for one day to make them thirsty, facilitating subsequent infection.

For the Tn-seq on the M4 organs, two to three hundred second instar nymphs of three days-old were transferred individually in 1.5 mL Eppendorf tubes with 1 µL water droplet containing 10^6 CFU of the *B. insecticola* RPE75 Tn-seq library. After five hours, insects that drank the bacterial suspension droplet were transferred in a Petri dish containing soybean seeds and sterile water, and kept at 25°C. At three days and five days post-infection, corresponding to second instar and third instar larval stages respectively, one hundred insects per experimental replicate were sacrificed and dissected in sterilized PBS solution under a binocular microscope. M4 organs were harvested and pooled together in 250 µL of sterilized PBS solution. The pooled M4 organs were homogenized with a plastic pestle and after homogenization, the plastic pestle was washed with 250 µL of sterilized PBS solution. The total bacterial suspension (500 µL) was centrifuged at 100 g for 10 minutes at 4°C to pellet host cellular debris. The bacterial supernatant was kept and centrifuged at 12,000g for 10 minutes at 4°C and the bacterial pellet was stored at -20°C until DNA extraction. Each experiment was performed in triplicates.

For Tn-seq on the M1 and M3 organs, one hundred second instar nymphs of three days-old were transferred individually in 1.5 mL Eppendorf tubes with 1 µL water droplet containing 10^6 CFU of the *B. insecticola* RPE75 Tn-seq library. After five hours, insects that drank the bacterial suspension droplet were transferred in a Petri dish containing soybean seeds and sterile water, and kept at 25°C. At 24 hours post-infection, at second instar larval stage, twenty insects per experimental replicate were sacrificed and dissected in sterilized PBS solution under a binocular microscope. M1 and M3 organs were harvested separately and pooled per twenty insects in 250 µL of sterilized PBS solution. The pooled M1 and M3 organs were

homogenized with a plastic pestle and after homogenization, the plastic pestle was washed with 250 µL of sterilized PBS solution. A fraction of each crushed organ suspension was subjected to serial dilutions to assess CFU counting per insect. Each remaining suspension was spread onto YG agar plates and incubated for two days at 28°C. After incubation, the bacteria grown onto these YG plates were scrapped and resuspended in 2 mL of YG medium. Each bacterial solution was centrifuged at 4000 g for 15 minutes at room temperature and pellets were stored at -20°C until DNA extraction. Each experiment was performed in triplicates.

DNA extraction, preparation of the libraries, high-throughput sequencing and bioinformatics

DNA extractions, preparation of the Illumina sequencing libraries, sequencing and bioinformatics analysis to identify fitness genes were performed essentially as it is described in **Chapter II section 3**. To evaluate the correlation between each Tn-seq experiment conducted so far (including experiments described in **Chapters II, III and IV**), the Pearson correlation coefficients between replicates of each experiment was calculated from the read counts per gene using the corresponding function in Excel.

4. Results

4.1. Bottleneck size estimation

The infection bottleneck of the *B. insecticola* population when establishing in the crypts of the *R. pedestris* midgut was determined by two independent methods. In the first method, a GFP-labelled strain of *B. insecticola* was mixed in varying proportions with an unlabelled strain. The mixes ranged from 1 in 10 to 1 in 20,000 (GFP-labelled to unlabelled). After establishment of the symbiosis by these symbiont mixes, the insects were dissected and the presence of GFP-labelled bacteria in the crypts was determined by microscopy and by plating the gut content for detecting the GFP strain (Km^R). The rationale is that, if the dilution factor of the GFP strain is below the bottleneck size, the GFP-labelled bacteria should be co-infecting the symbiotic organ with the unlabelled bacteria, so GFP signals should appear as spots in the crypts and the strain should be recovered from the symbiotic organ. When the dilution factor approaches

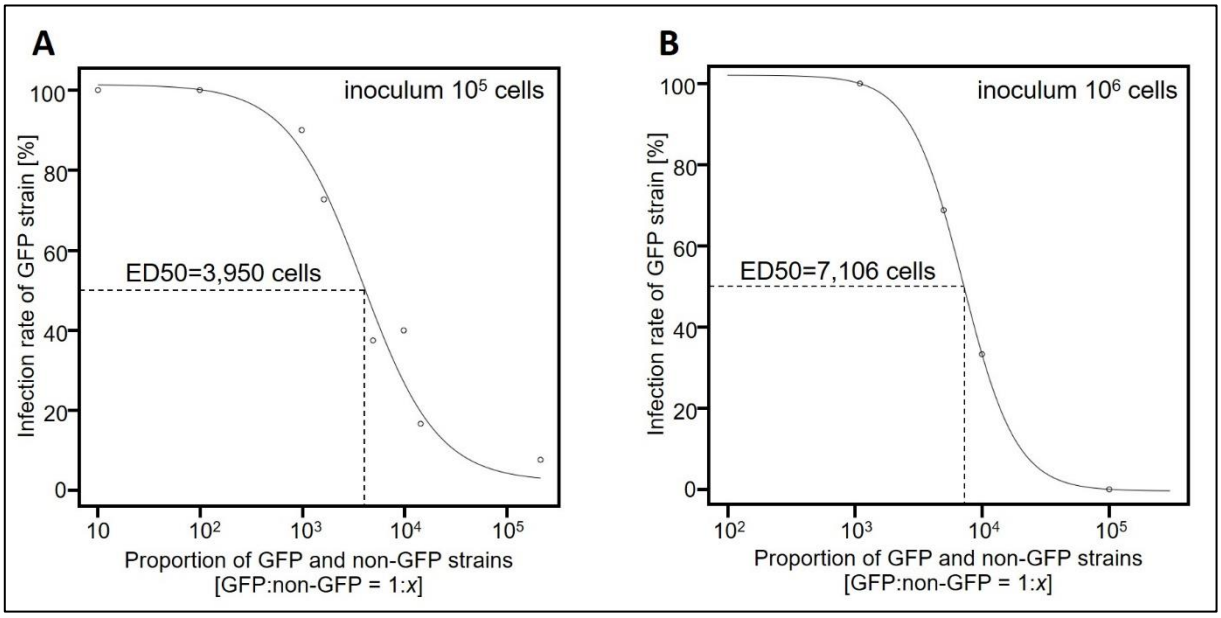


Figure 97: Determination of the infection bottleneck by mixed inoculation.
Insects were infected with a mix of GFP-labeled and unlabeled strains with varying proportions as indicated in the x-axis. The rate of insects infected with the GFP strain in insects was determined by microscopy inspection of the symbiotic organ at three days post-infection and by plating the crypt content on selective medium for the GFP strain. Insects were infected with an inoculum of 10^5 (A) or 10^6 (B) bacteria. The median effective dose (ED50) for infection by the GFP strain is indicated.

the bottleneck size, a portion of the insects should not be infected with the GFP-labelled strain (no sign of the GFP signal and no growth on selective medium) because the GFP-labelled bacteria were by chance not able to pass through the constricted region. When the dilution factor is above the bottleneck size, all insects should not be infected with the GFP strain. With two different inoculum sizes, 10^5 and 10^6 bacteria per insect, the latter corresponding to the inoculum size of the Tn-seq experiment (see below), a bottleneck of respectively 3,950 and 7,106 bacteria was observed (**Figure 97**).

The second estimation of the lower limit of the bottleneck size for the *B. insecticola* population after colonizing its host *R. pedestris* used the Tn-seq method itself. For that purpose, I have inoculated 10^6 CFU of the *B. insecticola* Tn-seq library per insect to a cohort of 60 young insects at the second instar stage (**Figure 98**). From each of these 60 infected insects, I have collected the M4 organ three days after infection, multiplied its bacterial content and extracted the bacterial DNA. Out of the 60 DNA samples, six of them showed a low DNA quality and were discarded for further analysis, hence I obtained sequencing data for 54 insect samples. For each insect replicate, approximately three million reads were obtained and uniquely mapped

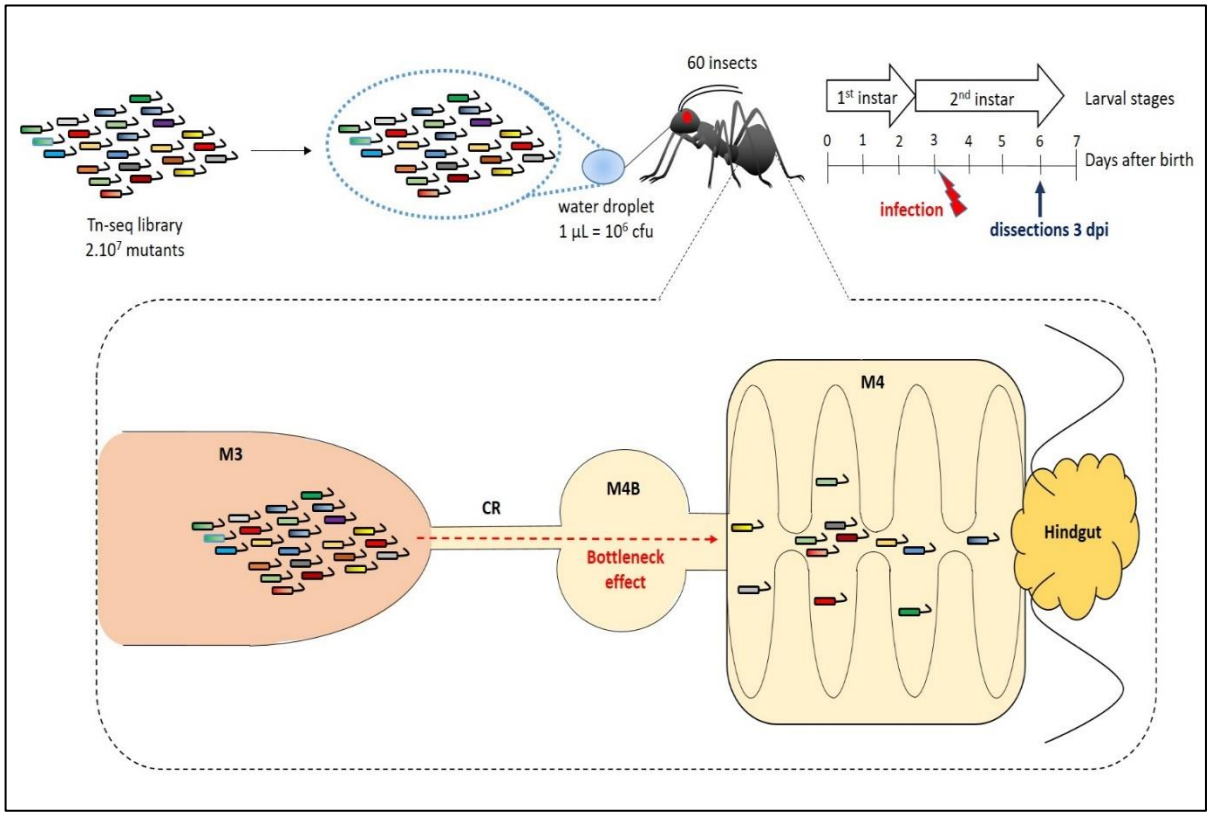


Figure 98: Experimental setup to determine the initial bottleneck bacterial population in the symbiotic organ of *R. pedestris* by Tn-seq.
Abbreviations: M: midgut, CR: constricted region

to the *B. insecticola* genome (**Table 4**). Thanks to the first steps of the ARTIST analysis, I counted the number of TA sites mutated per insect, corresponding to the number of unique mutants sequenced from the recovered bacterial population after colonization (**Table 4**). In addition, I also quantified the number of bacteria per insect in order to calculate the number of bacterial generations for each insect (**Table 4**).

By determining the number of mutated TA sites for each insect sample, I found that the mean number of unique mutants present per symbiotic organ was equal to 10,514, from an initial inoculated population of 10^6 CFU per insect that contained 110,735 potential individual mutants (**Table 4**). This lower limit estimate of 10,514 for the infection bottleneck is in good agreement with the bottleneck estimated by the mixed infection approach described above.

However, this bottleneck size strongly varied between insects, from 1,952 to 20,325 independent bacteria sampled (**Table 4**). By checking the distribution of these numbers of total TA sites mutated across the different insect samples, I observed that these insect samples fall into two subsets (**Figure 99**). These two insect subpopulations contain approximately 7,000

Insect sample	Number of unique mapped reads	Number of TA sites mutated						CFU per insect	Number of generations
		Chrom 1	Chrom 2	Chrom 3	Plasm 1	Plasm 2	Total		
1	2,562,211	7,592	3,097	1,189	2,231	885	14,994	5.00E+06	8.38
2	4,608,961	6,975	2,767	1,100	2,047	753	13,642	2.40E+06	7.46
3	3,426,018	8,346	3,320	1,394	2,382	924	16,366	2.60E+05	3.99
4	4,155,717	8,691	3,457	1,486	2,512	1,026	17,172	3.20E+05	4.22
5	4,408,277	8,558	3,384	1,401	2,532	989	16,864	5.30E+05	4.97
6	4,809,589	3,358	1,223	426	888	353	6,248	3.30E+05	5.72
7	3,960,430	3,677	1,399	468	999	448	6,991	3.10E+05	5.47
8	3,917,728	5,193	2,017	678	1,452	674	10,014	2.50E+05	4.64
9	3,025,229	7,253	2,867	1,109	2,068	809	14,106	4.20E+06	8.22
10	4,673,669	8,309	3,375	1,313	2,472	976	16,445	2.10E+07	10.32
11	4,250,365	9,762	3,937	1,614	2,904	1,122	19,339	1.10E+07	9.15
12	3,998,194	6,903	2,566	1,068	1,925	770	13,232	2.60E+07	10.94
13	2,905,437	3,268	1,206	432	877	387	6,170	1.50E+06	7.93
14	4,727,178	3,698	1,404	546	1,041	411	7,100	5.00E+06	9.46
15	4,657,791	6,521	2,530	999	1,897	720	12,667	1.70E+07	10.39
16	4,334,513	1,831	707	272	534	217	3,561	3.00E+06	9.72
17	4,229,647	1,805	751	268	531	221	3,576	4.90E+06	10.42
18	3,276,174	4,961	1,930	767	1,448	499	9,605	3.50E+06	8.51
19	3,525,110	6,686	2,583	1,073	1,959	742	13,043	5.40E+06	8.69
20	3,533,797	3,951	1,540	554	1,062	450	7,557	5.30E+06	9.45
21	4,009,277	3,315	1,224	490	804	362	6,195	1.30E+06	7.71
22	3,537,854	4,198	1,518	633	1,154	460	7,963	2.40E+07	11.56
23	4,000,544	6,346	2,490	933	1,837	631	12,237	1.70E+07	10.44
24	4,050,283	9,423	3,869	1,483	2,795	1,066	18,636	3.40E+06	7.51
25	1,556,338	7,126	2,971	1,166	2,123	777	14,163	3.00E+06	7.73
26	1,034,102	1,411	499	173	418	172	2,673	2.40E+06	9.81
27	1,990,896	2,965	1,175	435	790	342	5,707	6.40E+06	10.13
28	1,427,909	2,734	1,037	352	709	287	5,119	4.60E+06	9.81
29	1,468,796	6,834	2,702	1,072	1,971	789	13,368	6.30E+06	8.88
30	1,471,479	4,381	1,700	669	1,181	465	8,396	4.60E+06	9.10
31	1,112,901	2,740	1,028	377	782	306	5,233	7.10E+06	10.41
32	1,163,327	5,275	1,997	852	1,499	580	10,203	5.60E+06	9.10
33	585,856	4,044	1,598	644	1,131	522	7,939	2.70E+07	11.73
34	602,561	2,953	1,134	376	804	374	5,641	4.60E+06	9.67
35	972,660	6,084	2,371	901	1,714	671	11,741	4.60E+06	8.61
36	616,670	2,900	1,139	382	840	384	5,645	5.40E+06	9.90
37	434,564	4,432	1,686	615	1,281	581	8,595	2.90E+07	11.72
38	1,651,539	5,669	2,171	919	1,593	617	10,969	2.30E+05	4.39
39	1,691,605	3,158	1,092	478	796	331	5,855	3.80E+05	6.02
40	598,525	1,008	409	119	295	121	1,952	3.60E+06	10.85
41	542,238	1,584	602	222	463	200	3,071	1.80E+06	9.20
42	513,056	3,006	1,210	449	837	380	5,882	1.70E+05	4.85
43	642,414	3,044	1,168	483	819	337	5,851	2.70E+06	8.85
44	577,697	3,962	1,564	610	1,136	455	7,727	3.20E+06	8.69
45	674,683	3,697	1,409	506	1,023	409	7,044	2.70E+06	8.58
46	5,837,131	5,313	2,121	829	1,516	581	10,360	4.10E+05	5.31
47	4,438,231	7,982	3,069	1,171	2,284	951	15,457	3.30E+06	7.74
48	4,946,909	7,685	3,073	1,297	2,189	859	15,103	2.00E+06	7.05
49	4,572,732	10,128	4,175	1,735	3,056	1,231	20,325	2.50E+05	3.62
50	4,094,956	9,646	3,959	1,569	2,970	1,114	19,258	1.20E+05	2.64
51	5,487,335	8,090	3,187	1,336	2,460	905	15,978	2.50E+05	3.97
52	3,775,200	9,063	3,650	1,487	2,841	1,092	18,133	1.70E+06	6.55
53	5,669,257	7,003	2,877	1,130	2,058	798	13,866	1.60E+06	6.85
54	4,946,841	6,468	2,619	1,052	1,923	721	12,783	1.50E+06	6.87
Mean	2,957,082	5,389	2,121	835	1,553	616	10,514	5.54E+06	9.04

Table 4: Bottleneck sizes at the M4 region of second instar *R. pedestris* insects. Each parameter is showed for the 54 insect samples, and the mean value for each parameter is provided below. Abbreviations: Chrom: chromosome, Plasm: plasmid, CFU: colony-forming units.

Insect sample	Frequency of TA sites mutated per replicon (%)				
	Chrom 1	Chrom 2	Chrom 3	Plasm 1	Plasm 2
1	50.63	20.65	7.93	14.88	5.90
2	51.13	20.28	8.06	15.01	5.52
3	51.00	20.29	8.52	14.55	5.65
4	50.61	20.13	8.65	14.63	5.97
5	50.75	20.07	8.31	15.01	5.86
6	53.75	19.57	6.82	14.21	5.65
7	52.60	20.01	6.69	14.29	6.41
8	51.86	20.14	6.77	14.50	6.73
9	51.42	20.32	7.86	14.66	5.74
10	50.53	20.52	7.98	15.03	5.93
11	50.48	20.36	8.35	15.02	5.80
12	52.17	19.39	8.07	14.55	5.82
13	52.97	19.55	7.00	14.21	6.27
14	52.08	19.77	7.69	14.66	5.79
15	51.48	19.97	7.89	14.98	5.68
16	51.42	19.85	7.64	15.00	6.09
17	50.48	21.00	7.49	14.85	6.18
18	51.65	20.09	7.99	15.08	5.20
19	51.26	19.80	8.23	15.02	5.69
20	52.28	20.38	7.33	14.05	5.95
21	53.51	19.76	7.91	12.98	5.84
22	52.72	19.06	7.95	14.49	5.78
23	51.86	20.35	7.62	15.01	5.16
24	50.56	20.76	7.96	15.00	5.72
25	50.31	20.98	8.23	14.99	5.49
26	52.79	18.67	6.47	15.64	6.43
27	51.95	20.59	7.62	13.84	5.99
28	53.41	20.26	6.88	13.85	5.61
29	51.12	20.21	8.02	14.74	5.90
30	52.18	20.25	7.97	14.07	5.54
31	52.36	19.64	7.20	14.94	5.85
32	51.70	19.57	8.35	14.69	5.68
33	50.94	20.13	8.11	14.25	6.58
34	52.35	20.10	6.67	14.25	6.63
35	51.82	20.19	7.67	14.60	5.72
36	51.37	20.18	6.77	14.88	6.80
37	51.56	19.62	7.16	14.90	6.76
38	51.68	19.79	8.38	14.52	5.62
39	53.94	18.65	8.16	13.60	5.65
40	51.64	20.95	6.10	15.11	6.20
41	51.58	19.60	7.23	15.08	6.51
42	51.11	20.57	7.63	14.23	6.46
43	52.03	19.96	8.25	14.00	5.76
44	51.27	20.24	7.89	14.70	5.89
45	52.48	20.00	7.18	14.52	5.81
46	51.28	20.47	8.00	14.63	5.61
47	51.64	19.86	7.58	14.78	6.15
48	50.88	20.35	8.59	14.49	5.69
49	49.83	20.54	8.54	15.04	6.06
50	50.09	20.56	8.15	15.42	5.78
51	50.63	19.95	8.36	15.40	5.66
52	49.98	20.13	8.20	15.67	6.02
53	50.50	20.75	8.15	14.84	5.76
54	50.60	20.49	8.23	15.04	5.64
Mean	51.56	20.10	7.75	14.67	5.92

Table 5: Proportion stability of mutated replicon of *B. insecticola* in the symbiotic organ. The mean mutated proportion for each replicon is displayed below. Abbreviations: Chrom: chromosome, Plasm: plasmid.

and 13,000 bacterial mutants, respectively (**Figure 99**). We hypothesized that this binary distribution might be explained by another binary parameter, the insect's gender, which was

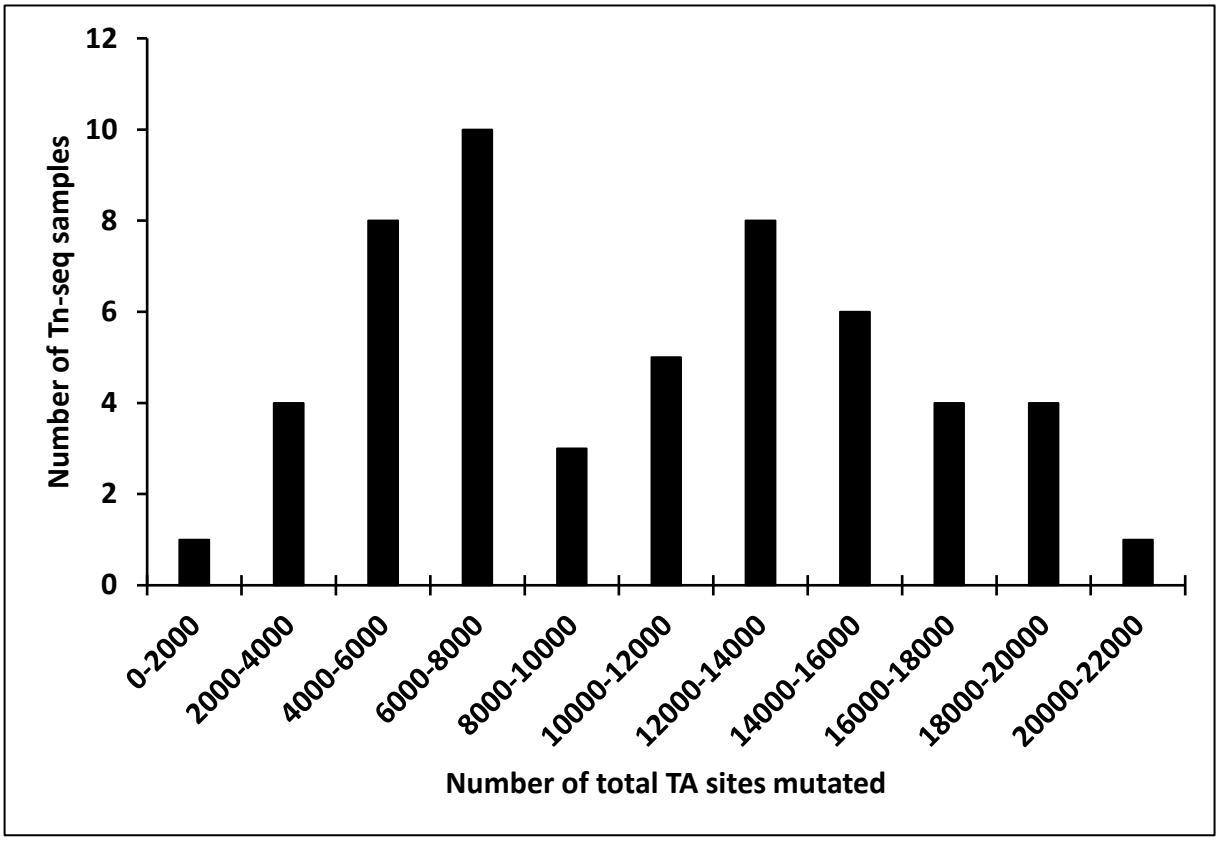


Figure 99: Distribution of measured bottleneck sizes.
Each insect sample was classified according to their number of total TA sites mutated in bins of 2,000 mutated TA sites. The distribution showed two peaks corresponding to the categories of 6,000-8,000 and 12,000-14,000 mutated TA sites, respectively.

however not verified during this experiment.

From these 54 insect samples, there was also a variation of the number of total CFU recovered for each individual, between 10^5 to 10^7 bacteria per insect (**Table 4**). There was no correlation between the number of CFU and the number of total TA sites mutated per insect sample. Additionally, the proportion of TA sites mutated per replicon is similar for each insect sample, with a mean proportion of 51.56% in chromosome 1, 20.1% in chromosome 2, 7.75% in chromosome 3, 14.67% in plasmid 1 and 5.92% in plasmid 2 (**Table 5**). This conserved proportion of TA sites mutated per replicon reflects the reproducibility of bottleneck effects on the symbiont for each insect, with half of the colonizing population bearing mutations in the chromosome 1 (**Table 5**). Compared to the mean M4 infecting population (*i.e.* mean bottleneck size), the mean number of CFU per insect, equal to 5.54×10^6 CFU, showed that the bacterial population is strongly growing in the M4 region (**Table 4**). So, I could not have estimated the bottleneck size with CFU counting method (Abel *et al.*, 2015). However, thanks

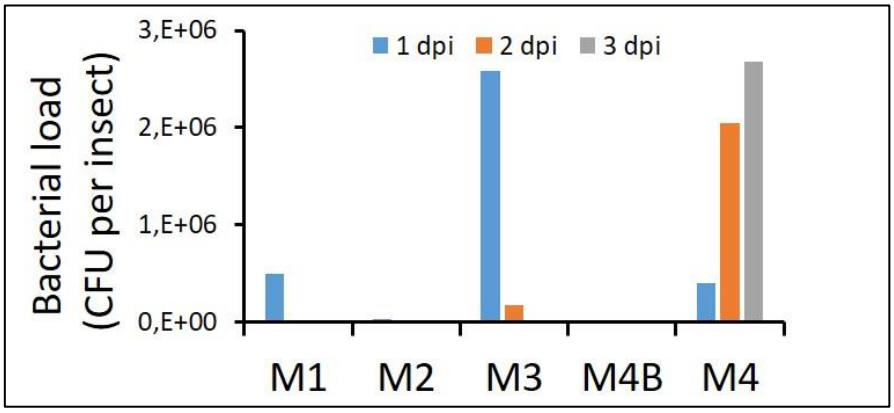


Figure 100: Dynamics of the *B. insecticola* load in the midgut compartments of *R. pedestris.*

Insects were infected with 10^6 bacteria and the bacterial load in the M1, M2, M3, M4B and M4 midgut compartments was determined by CFU counting at 1 day post-infection (dpi), 2 dpi and 3 dpi.

to these quantitative data, I found that the number of bacterial generations also varied between individuals, from 2 to 12 generations with a mean of 9 generations (**Table 4**). This strong multiplication of the bacteria in the M4 is ideal for Tn-seq screens that searches for depleted mutants after growth in the condition of interest.

Based on both methods, for an initial inoculum of 10^6 CFU per insect, the mean bottleneck size corresponds to approximately 10,000 bacteria (**Table 4**). The genome of *B. insecticola* bears 110,735 TA sites, which represent 110,735 possible independent mutants in the Tn-seq library. In order to cover about 10 times all these mutants in the population recovered after colonization, which is approximately 1,000,000 bacterial mutants, we would need to sacrifice about 100 insects per biological replicate.

4.2. Dynamics of the *B. insecticola* population in the *R. pedestris* midgut

In addition to the M4 symbiotic organ, I also aimed to perform Tn-seq on the *B. insecticola* population after its ingestion by the insect and during its passage through the upstream regions of the midgut. To settle the conditions for these Tn-seq experiments, I first determined the dynamics of the bacterial population in these midgut regions. *R. pedestris* second instar insects were fed by a single inoculum in 1 µL with 10^6 CFU of *B. insecticola*. At one, two and three days post-infection, the M1, M2, M3, M4B and M4 organs were harvested and the

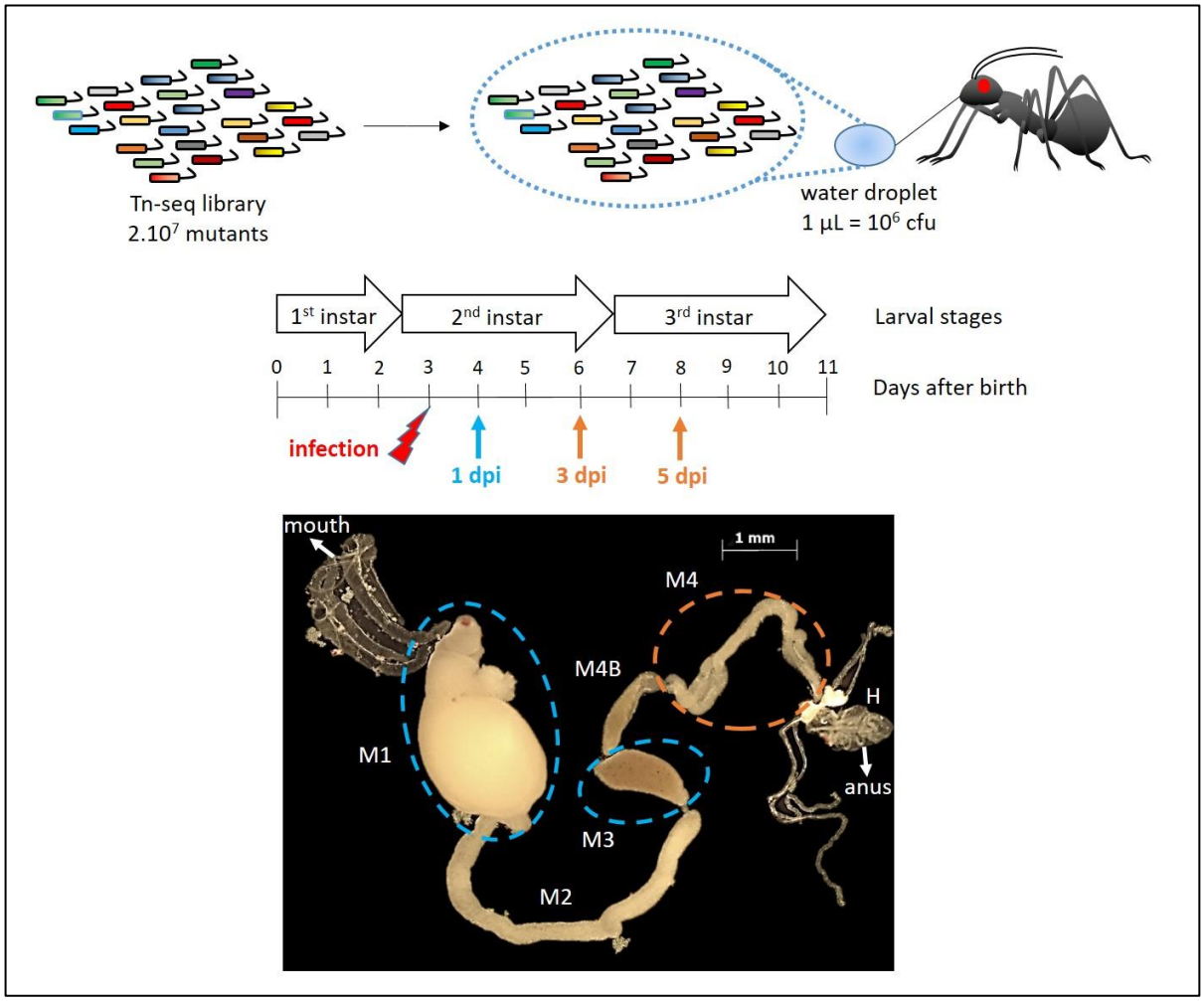


Figure 101: Experimental setup for *in vivo* Tn-seq conditions.
The time points indicated in orange correspond to the collections of the M4 organs, and the time points indicated in blue correspond to the collections of the M1 and the M3 organs. The picture below shows the total midgut dissected from a second instar larval stage *R. pedestris* insect under the binocular microscope. The M4 organ is encircled in orange and the M1 and M3 organs are encircled in blue. The scale bar represents 1 mm.

bacterial load in these midgut regions was determined by CFU counting. This experiment provided several interesting observations (**Figure 100**). It showed that the population in the M2 is very low (**Figure 100**), probably because this organ is a narrow tube in which the bacteria are transiting rapidly. Also, the M4B has a low bacterial load (**Figure 100**), firstly because bacteria are only transiting in this region before arriving in the M4 crypts, and secondly at two and three days post-infection, because the bacteria are digested in the M4B (Ohbayashi *et al.*, 2019). On the other hand, the M1 and M3 organs contain bacteria at one day post-infection, about 500,000 and 2,500,000 CFU per insect, respectively (**Figure 100**). These populations then drop to very low levels from two days post-infection on, indicating that the bacteria

Sample	Nb of Post-Trim reads	% of Alignment against *Burkholderia insecticola* genome	Nb of Mapped reads
M1 replicate 1	2,823,041	62	1,751,617
M1 replicate 2	3,227,869	76.2	2,459,079
M1 replicate 3	2,694,597	71.4	1,924,816
M3 replicate 1	2,167,774	70.2	1,522,780
M3 replicate 2	3,088,791	75.2	2,321,943
M3 replicate 3	2,654,772	75.8	2,011,550
M4 2nd instar replicate 1	2,263,121	86	1,946,496
M4 2nd instar replicate 2	1,421,559	49.8	707,555
M4 2nd instar replicate 3	3,458,206	78.6	2,719,071
M4 3rd instar replicate 1	2,505,169	77.9	1,950,544
M4 3rd instar replicate 2	2,911,121	68.6	1,998,208
M4 3rd instar replicate 3	3,231,227	70.1	2,263,847

Table 6: Sequencing results for *in vivo* Tn-seq conditions.
The number of post-trim reads corresponds to the number of filtered reads after the trimming step. Abbreviations: Nb: number.

cannot stably colonize these organs efficiently, even if the high CFU number at one day post-infection in the M3 indicates that they initially multiply in the midgut (**Figure 100**). Finally, and as already discussed above, the founding population of the M4 of 10,000 multiplies very rapidly, producing 5 new generations already in the first day to reach a population of several millions at three days post-infection (**Figure 100**). Taken together, in light of the bacterial populations present, Tn-seq can be performed on the M1, M3 and M4 organs.

4.3. Bacterial symbiotic functions

Based on the estimated bottleneck size and the population size in the midgut regions, as well as the requirement to recover the equivalent of at least 1,000,000 infecting clones to avoid stochastic changes in the Tn-seq population, I performed an *in vivo* Tn-seq experiment on the M1, M3 and M4 organs. The M2 and M4B regions were not included because of too few resident bacteria. For all conditions, I have inoculated 10^6 CFU of the *B. insecticola* Tn-seq library per insect to a cohort of young insects at the second instar larval stage per biological replicate. The number of insects to infect per biological replicate was chosen based on the estimation of the bottleneck size to collect sufficient mutants in the organs (see **sections 4.1 and 4.2**). For the M4 organ, I have collected the organs at three and five days post-infection, corresponding to the second and the third instar larval stages, respectively (**Figure 101**).

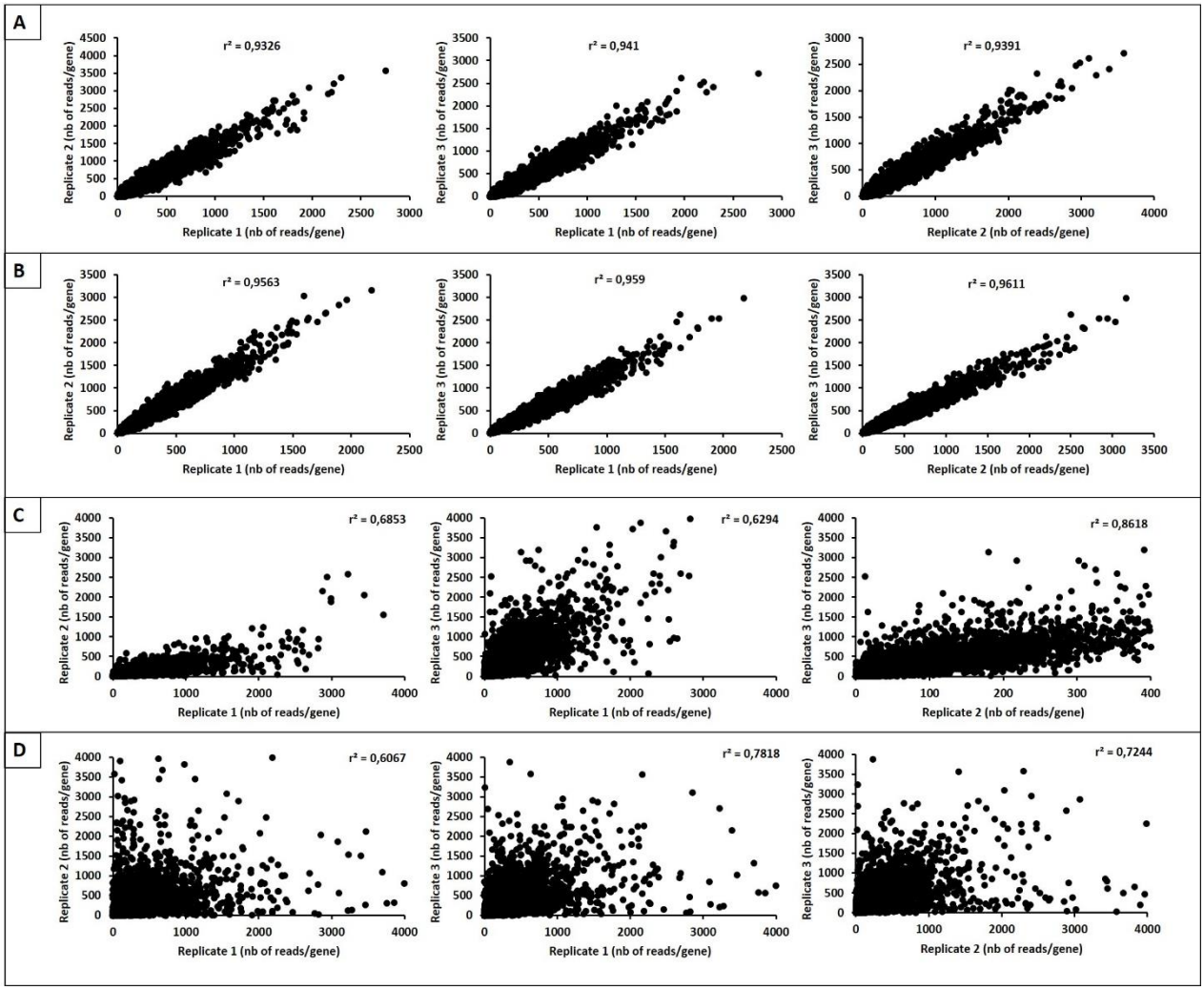


Figure 102: Correlations between read counts distribution in the three replicates of each *in vivo* Tn-seq condition.

Dot plot representations of the comparison of each transposon insertions distribution between the three Tn-seq replicates are shown for each *in vivo* condition. The number of reads per gene is displayed for each replicate. The Pearson correlation coefficient r^2 was calculated for each comparison and indicated on each graph. A) M1 replicates. B) M3 replicates. C) Second instar M4 replicates. D) Third instar M4 replicates.

Concerning the M1 and the M3 regions, organs were collected at 24 hours post-infection (**Figure 101**). Thus in total, I had four *in vivo* conditions in triplicates with symbiotic bacteria collected in the M1 organ (one day post-infection), the M3 organ (one day post-infection), the M4 organ at the second instar larval stage (three days post-infection) and the M4 organ at the third larval stage (five days post-infection) (**Figure 101**).

After processing the samples and sequencing, each replicate of the four *in vivo* conditions contained around two million filtered reads, with approximately 70% of these reads that were aligned to the *B. insecticola* genome (**Table 6**). Compared to the *in vitro* Tn-seq samples, this

Pearson correlation	MMGlu	MMSucc	PMB cmin	PMB cmax	LL37 cmin	LL37 cmax	Riptocin cmin	Riptocin cmax	CCR179 cmin	CCR179 cmax	CCR480 cmin	CCR480 cmax	M1	M3	M4 2nd	M4 3rd
YG	0,95457881	0,93347501	0,93347937	0,92104801	0,93845857	0,93725831	0,94736365	0,93423804	0,95274317	0,95244117	0,94934584	0,95886871	0,96153019	0,96845888	0,8435369	0,67372985
MMGlu		0,95649691	0,95779211	0,93681455	0,96168479	0,96291514	0,97052075	0,95381577	0,97506608	0,97535598	0,97349659	0,97786928	0,95290342	0,95656082	0,85422535	0,68643918
MMSucc			0,94497321	0,93613122	0,95131613	0,93981376	0,94654464	0,92755377	0,95885497	0,957779	0,94714972	0,95155968	0,94633841	0,95136521	0,85724916	0,68706402
PMB cmin				0,95870593	0,97194941	0,96678156	0,9657995	0,93435032	0,96083611	0,95395049	0,95805923	0,95651734	0,93833411	0,94385691	0,857885	0,68211871
PMB cmax					0,95614502	0,94869076	0,94816708	0,92860164	0,94151657	0,9349065	0,94075532	0,94082218	0,93406331	0,9307018	0,84883514	0,67290633
LL37 cmin						0,95681981	0,96626893	0,93567021	0,96310328	0,95950773	0,96363784	0,96143333	0,93615035	0,94310259	0,8527725	0,68141367
LL37 cmax							0,97450939	0,95357362	0,97399396	0,96583095	0,96736559	0,96386695	0,9516028	0,95756257	0,86276057	0,68626424
Riptocin cmin								0,96176795	0,97529742	0,9694184	0,97436176	0,97206752	0,95173443	0,95932425	0,85857486	0,68453898
Riptocin cmax									0,95304157	0,95414203	0,9508476	0,9526329	0,93920407	0,94556032	0,83547588	0,66771051
CCR179 cmin										0,98673159	0,9764459	0,97412037	0,96080111	0,96572687	0,86974262	0,69243256
CCR179 cmax											0,97146157	0,97169953	0,96102441	0,96287887	0,86980115	0,70036498
CCR480 cmin												0,97417121	0,95289698	0,95889594	0,85323153	0,67668506
CCR480 cmax													0,95489468	0,96074336	0,85270171	0,6853639
M1														0,98688097	0,86993349	0,69757069
M3															0,87090672	0,69434466
M4 2nd																0,70326178

Figure 103: Correlation between experiments.
Pairwise Pearson correlation coefficients between experiments were calculated based on the read counts per gene. The blue color indicates values close to 1, the green color indicates values close to 0.9 and yellow values indicates values close to 0.7.

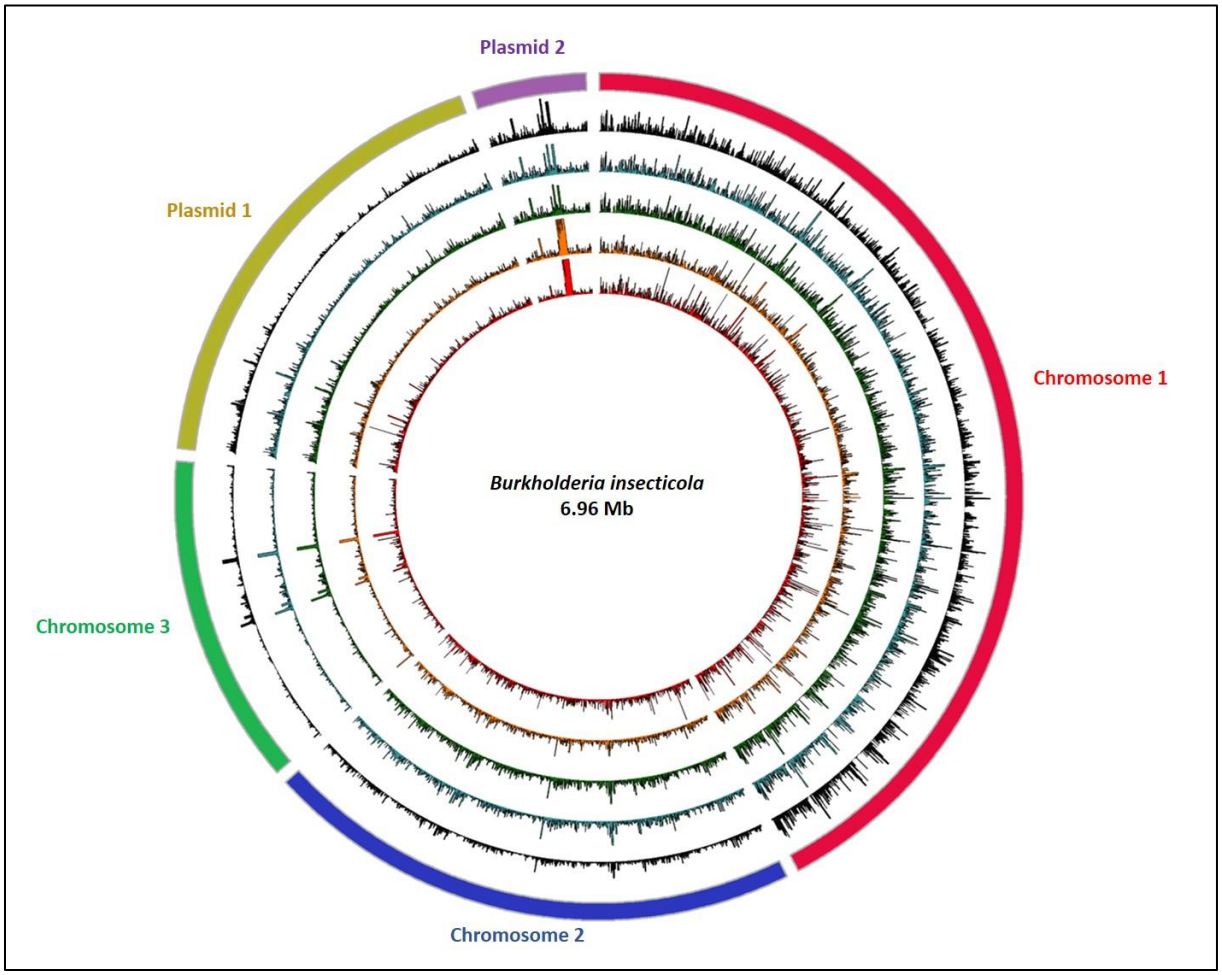


Figure 104: Insertion distributions of *in vivo* Tn-seq conditions across the *B. insecticola* genome.

Circular representation of the *B. insecticola* genome (6.96 Mb) consisting of chromosomes 1, 2 and 3 and plasmids 1 and 2. The read counts per TA site are represented from outer to inner rings: YG rich medium condition (black histograms), M1 *in vivo* condition (blue histograms), M3 *in vivo* condition (green histograms), second instar M4 *in vivo* condition (orange histograms), and third instar M4 *in vivo* condition (red histograms).

reduced percentage of alignment is possibly due to the presence of host DNA in the *in vivo* samples.

The three replicates of the M1 and the M3 experiments showed high correlation coefficients ($r^2 > 0.93$ and $r^2 > 0.95$ for M1 and M3 replicates, respectively) (**Figure 102**). Therefore, I pooled the sequencing data of these three replicates for the M1 and the M3 conditions. However, I observed that the correlation coefficients for the three replicates of the two M4 *in vivo* conditions were lower than for the M1 and M3 replicates ($0.62 < r^2 < 0.86$ and $0.60 < r^2 < 0.78$ for the second and third instar M4 conditions, respectively) (**Figure 102**). Compared to the other midgut compartments and the previous *in vitro* conditions, these lower correlation

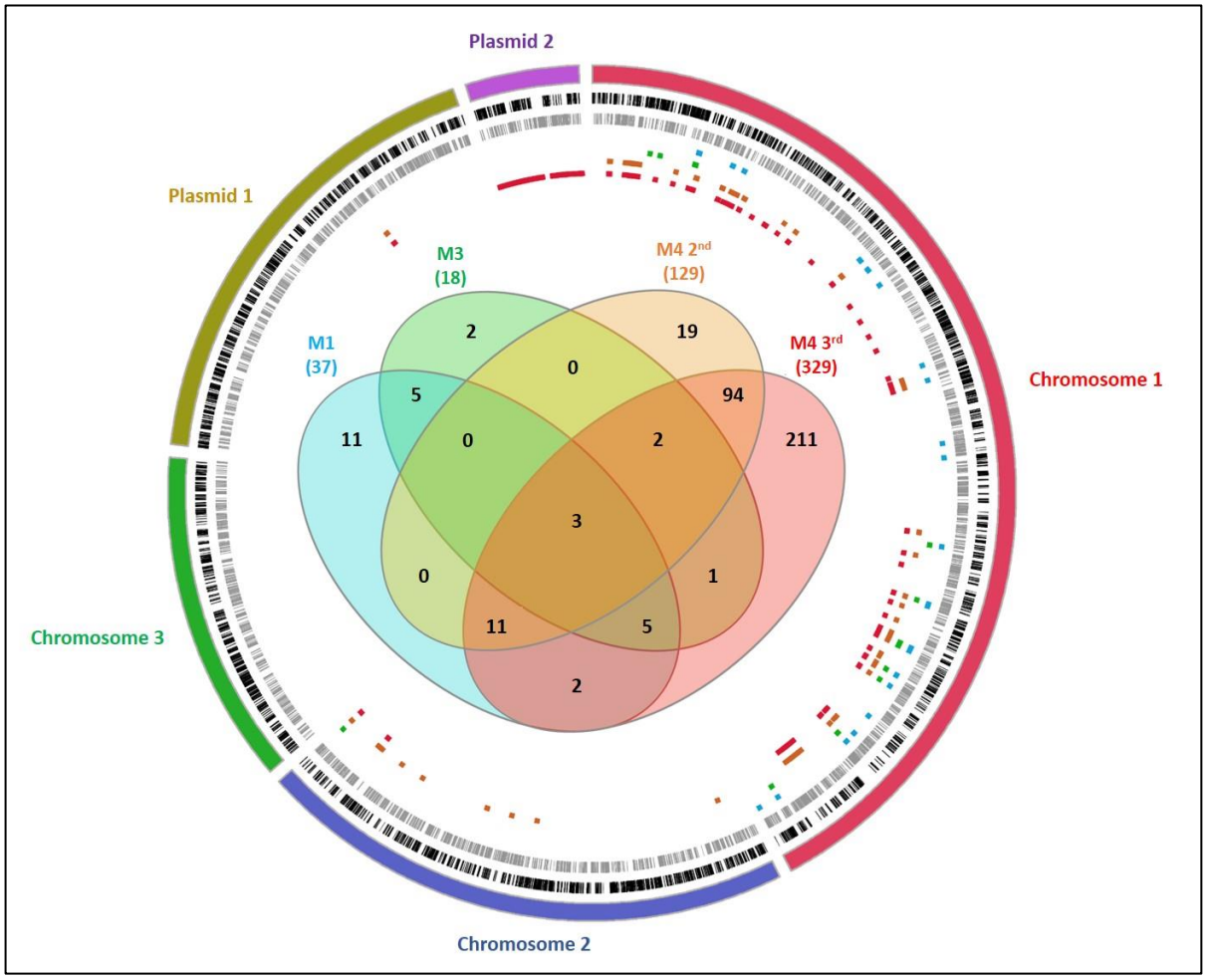


Figure 105: Fitness genes identified by Con-ARTIST analysis in *B. insecticola* for *in vivo* host colonization.

Circular representation of the *B. insecticola* genome consisting of chromosomes 1, 2 and 3 and plasmids 1 and 2. From outer to inner rings: forward CDS (black bars), reverse CDS (grey bars), conditionally-essential genes (Con-ARTIST essentiality score = 2) for M1 (blue dots), conditionally-essential genes for M3 (green dots), conditionally-essential genes (Con-ARTIST essentiality score = 2) for M4 second instar (orange dots), conditionally-essential genes (Con-ARTIST essentiality score = 2) for M4 third instar (red dots), distribution of conditionally-essential genes between the four *in vivo* conditions (Venn diagram).

coefficient values suggest that the population pools mobilized in the M4 organ is much more variable between biological replicates, possibly related to the strong bottleneck effect. I noticed that a specific gene cluster of 19 genes was represented with a very high number of reads in the M4 conditions (up to 320,000 reads for the second and the third instar M4 conditions) compared to the other *in vivo* and *in vitro* conditions (around 20,000 reads) (see **section 4.3.6** for a discussion on these genes). However, removing these values did not modify drastically the correlation coefficients of the M4 replicates. Despite these lower values, the correlation coefficients were still high enough to pool the sequencing data of the three

replicates of the two M4 conditions for further analysis.

Furthermore, the pairwise correlation coefficients between all *in vitro* (see **Chapters II and III**) and *in vivo* experiments revealed an overall high correlation between them (**Figure 103**). Nevertheless, the M4 second instar condition, and particularly the third instar condition, had a clearly lower similarity to the other experiments (**Figure 103**), indicating that the fitness landscape of the bacteria required for survival in the M4 crypts is specific and distinct than for the other tested conditions.

After pooling the sequencing data for each *in vivo* condition, I compared the insertions distribution on the *B. insecticola* genome of these samples with *in vitro* growth in YG rich medium which served as the control condition (**Figure 104**). I observed that the global patterns of insertions across the genome are quite similar between the *in vivo* and the YG rich medium conditions, except for the above-mentioned region of the plasmid 2 for the two M4 conditions where I noticed a strong hotspot of insertions (**Figure 104**).

The Con-ARTIST analysis, which compares the *in vivo* samples with the control condition, identified 37, 18, 129 and 329 conditionally essential genes for the bacterial fitness in the M1, M3, second instar and third instar M4 *in vivo* conditions, respectively (Con-ARTIST essentiality score = 2) (**Figure 105**, see **Annexes 14, 15, 16 and 17** for the gene lists with Con-ARTIST scores for *in vivo* samples). In addition, there were 14, 4, 30 and 46 domain-conditionally essential genes which could be required for the bacterial fitness in the M1, M3, second instar and third instar M4 *in vivo* conditions, respectively (Con-ARTIST essentiality score = 1) (see **Annexes 14, 15, 16 and 17**). It is striking that the number of fitness genes is higher for the colonization of the symbiotic organ compared to the other midgut organs (**Figure 105**).

Interestingly, the Con-ARTIST analysis was also able to identify genes which are enriched in transposon insertions in the *in vivo* conditions compared to the rich medium condition, specifically for the M4 conditions. Thus, I have found 11 and 18 conditionally enriched genes (Con-ARTIST essentiality score = 4), and 3 and 1 domain-conditionally enriched genes (Con-ARTIST essentiality score = 3) in the second and the third instars M4 *in vivo* conditions, respectively (see **Annexes 16 and 17**). These enriched genes were only located in the plasmid 2 (see **section 4.3.6** and **Annexes 16 and 17**).

Concerning the locations of the *in vivo* fitness genes (corresponding to the conditionally essential genes, score 2), they are mostly located in the chromosome 1, and none of them

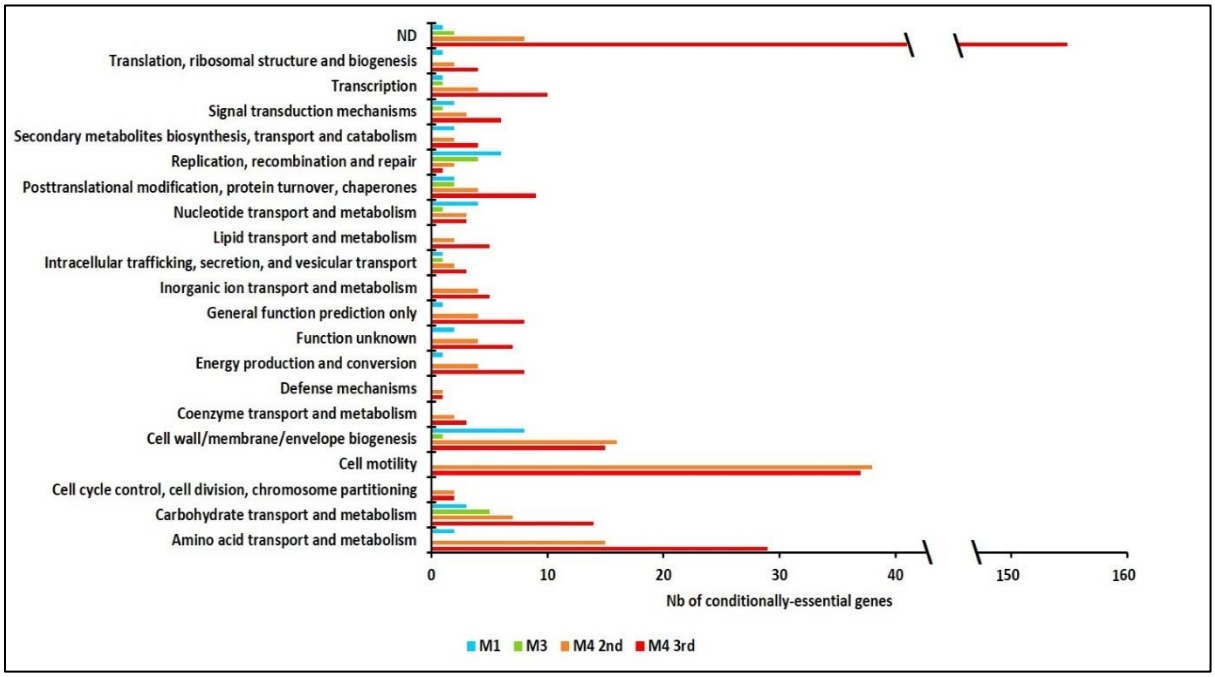


Figure 106: COG categories of *in vivo* bacterial fitness genes.
The numbers of conditionally-essential genes identified for each *in vivo* condition, M1 organ (blue), M3 organ (green), M4 organ at the second instar (orange) and the M4 organ at the third instar (red) are displayed for each COG category. ND: not determined.

were located in the chromosome 3 (**Figure 105**). For the third instar M4 condition, I observed that almost the whole plasmid 2 became essential for the bacterial fitness, excepting the cluster of genes enriched in transposon insertions (**Figure 105**). As I will discuss in **section 4.3.6**, the classification of the plasmid 2 genes as fitness functions for the M4 colonization is an artefact resulting from the fact that the plasmid is largely lost by the symbiotic bacteria.

By comparing the *in vivo* fitness genes of each midgut compartment, I found that only 3 bacterial genes were shared between the M1, M3 and M4 organs (**Figure 105**). These three genes were *BRPE64_RS10560*, *BRPE64_RS11040* and *BRPE6_RS11220* which are encoding the Wzm subunit of the ABC transporter O-antigen exporting system, the TolB protein from the Tol-Pal complex and the transketolase enzyme from the pentose-phosphate glycolysis pathway, respectively (see **Annexes 14, 15, 16 and 17**). The two M4 conditions shared 94 fitness genes, which was close to the total number of fitness genes required for the second instar M4 condition (129 fitness genes, **Figure 105**). Only 19 fitness genes were specifically associated to the symbiotic organ at the second instar (**Figure 105**). The 211 specific fitness genes for the third instar M4 condition mainly corresponded to the genes located in the plasmid 2 (**Figure 105**). For the M1 and the M3 organs, only 11 and 2 specific fitness genes

were respectively identified (**Figure 105**).

According to the COG categories (Tatusov *et al.*, 2000), I noticed that most of the fitness genes for the M1 and the M3 organs belonged to the replication, recombination and repair category (L category) (**Figure 106**). Other biological functions such as the envelope biogenesis (M category), nucleotide and carbohydrate metabolisms (F and G categories, respectively) were also representative among the fitness genes of the M1 and the M3 organs (**Figure 106**). Compared to the non-symbiotic organs, new COG categories appeared in the classification of symbiotic genes with lipid and coenzyme metabolisms (I and H categories, respectively), the defense mechanisms (V category), the inorganic ion transport (P category), and the cell motility (N category) (**Figure 106**). In the symbiotic organ, the most abundant bacterial functions were related to the cell motility (N category), the amino acid metabolism (E category) and the envelope biogenesis (M category) for both developing times (**Figure 106**). The cell motility function, which appeared specific for the M4 organ, contained 38 and 37 fitness genes required for the second and the third instars, respectively (**Figure 106**). The number of symbiotic genes which belonged to the envelope biogenesis category was also quite similar between the second instar (16 genes) and the third instar (15 genes) (**Figure 106**). However, the number of fitness genes from the amino acid metabolism category increased from the second instar (15 genes) to the third instar (29 genes) in the symbiotic organ (**Figure 106**). Additionally, the majority of symbiotic genes at the third instar were encoding hypothetical proteins with unknown functions ("General function prediction only", "Function unknown" and "ND" categories) (**Figure 106**). These unknown functions mainly corresponded to fitness genes localized to the plasmid 2 (**Figure 105**).

From these results, I focused more deeply on the major functional categories represented in the symbiotic organ and in the non-symbiotic organs, which could be divided into five biological functions: DNA repair, stress response, metabolism, envelope biogenesis and cell motility (see the recapitulative functions illustrated in the **Figure 129**). Concerning the plasmid 2 genes which were enriched in transposon insertions in the M4 organ conditions, I identified their functions and checked their presence among the *Burkholderia* genus (see below with **Figures 126 and 127, Table 7**).

4.3.1. DNA repair, transcription and translation modulations

DNA repair

Notably in the M1 and the M3 organs, I found different fitness genes involved in the repair of double-stranded DNA breaks such as the *ruvABC* gene cluster (*BRPE64_RS01640-BRPE64_RS01650*) (**Figure 107**) which is involved in the resolution of Holliday junctions (Wardrope *et al.*, 2009), the *uvrD* gene (*BRPE64_RS09020*) which is encoding a helicase involved in methyl-directed DNA mismatch repair (Tomko and Lohman, 2017), *mutL* (*BRPE64_RS02145*) also involved in mismatch repair (**Figure 108**) and the *recBCD* gene cluster (*BRPE64_RS04640-BRPE64_RS04650*) encoding the exodeoxyribonuclease V involved in foreign DNA degradation and in repair of chromosomal double-stranded DNA breaks (Lohman and Fazio, 2018). Additionally, I identified the *dusB-fis* gene cluster (*BRPE64_RS01665* and *BRPE64_RS01660*) (**Figure 107**) with the *fis* gene encoding the DNA-binding protein Fis and the *dusB* gene encoding a tRNA-dihydrouridine synthase. Together, these two genes were found to be involved in the protection of DNA against oxidative stress such as ROS in *E. coli*

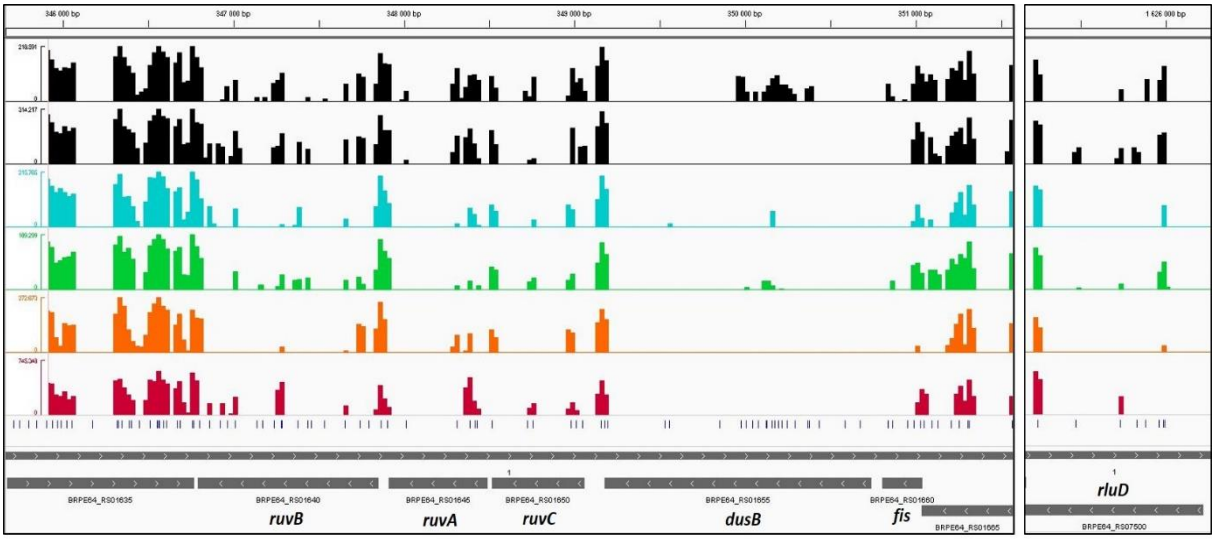


Figure 107: *In vivo* fitness genes involved in DNA repair and rRNA processing.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Left picture: *in vivo* fitness genes involved in DNA repair with *ruvB*, *ruvA*, *ruvC*, *dusB* and *fis* genes. Right picture: *in vivo* fitness gene *rluD* involved in rRNA processing.

(Weinstein-Fischer *et al.*, 2000) and in *Yersinia pseudotuberculosis* (Green *et al.*, 2016). Hence,

these gene functions suggest that symbiotic bacteria may face multiple stresses which can induce DNA damages, especially during the passage of non-symbiotic organs (see the recapitulative functions illustrated in the **Figure 129**).

Transcription and translation modulations

Concerning the transcription regulation, a gene encoding the transcription factor DksA (*BRPE64_RS13660*) was required for the bacterial fitness in the M1 and M3 organs. DksA is known to regulate the expression of a large set of genes during various nutrient starvations, often in synergy with the alarmone ppGpp (Furman *et al.*, 2015; Potrykus and Cashel, 2008). This transcription factor was shown to be more active at low pH values and essential for survival under acidic conditions in *E. coli* (Furman *et al.*, 2015). In addition, regarding the *Rhizobium*-legume symbiosis, it was recently demonstrated that a *Sinorhizobium meliloti dksA* mutant showed a significant delay in nodule development (Wippel and Long, 2016). Hence, this transcription factor might play a role in the transcriptional processes at early steps of symbiosis, in the putative acidic midgut organs.

In the M1 organ only, I identified the gene *BRPE64_RS14010*, which is the homologue of the *mnmE* gene in *E. coli,* and is involved in tRNA processing. It was shown that MnmE activates a

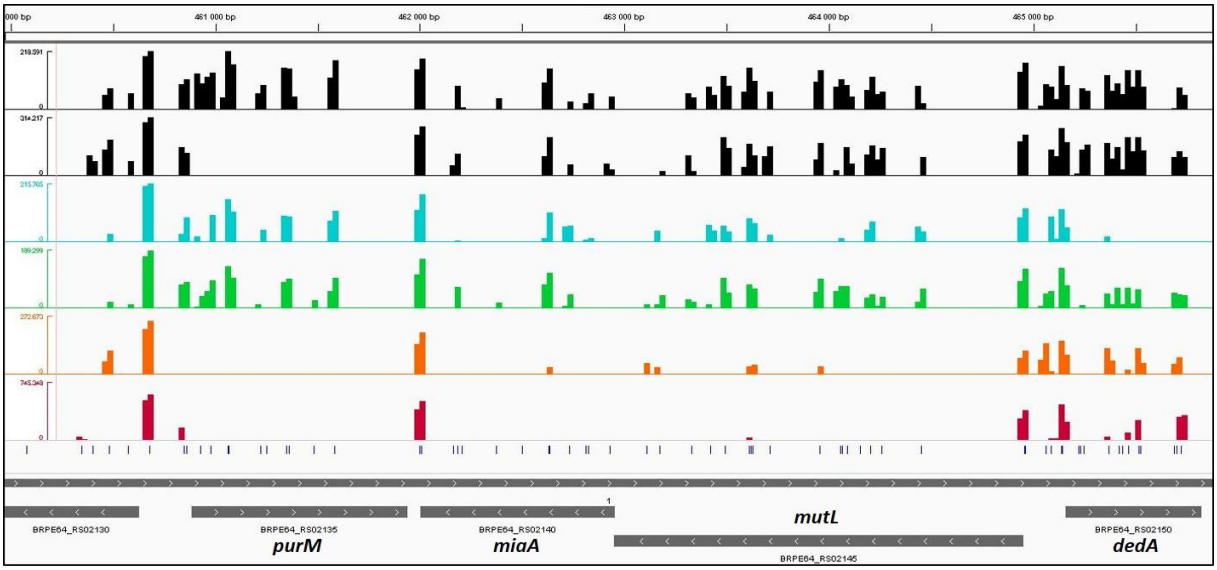


Figure 108: *In vivo* fitness genes related to purine biosynthesis (*purM*), tRNA processing (*miaA-mutL*) and in cell wall biosynthesis (*dedA*).

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey.

transcriptional regulator involved in glutamate-dependent acid resistance in *E. coli* (Gong *et al.*, 2004), and is also involved in the regulation of virulence factors in *Salmonella* (Shippy and Fadl, 2014; Shippy *et al.*, 2013).

In all *in vivo* conditions, I found two fitness genes that are organized together in a cluster called *miaA* (*BRPE64_RS02140*) and *mutL* (*BRPE64_RS02145*) (**Figure 108**). MiaA is involved in tRNA modifications whereas MutL is involved in methyl-directed DNA mismatch repair, however both of these genes were found to be involved in heat shock resistance in *E. coli* (Thompson and Gottesman, 2014; Tsui *et al.*, 1996). Additionally, another gene named *rluD* (*BRPE64_RS07500*) was also identified for *in vivo* colonization of *R. pedestris* (**Figure 107**). This gene encodes the pseudouridine synthase RluD which is involved in replacement of uridine by pseudouridine in 23S rRNA in *E. coli* (Gutgsell *et al.*, 2005; Hamma and Ferré-D'Amaré, 2006). Another set of two genes was found to be required for the colonization of the M4 organ, which are *truA* (*BRPE64_RS20065*) and *BRPE64_RS20070*. The *truA* gene is involved in tRNA modifications, and the gene *BRPE64_RS20070* encodes a hypothetical protein in the *B. insecticola* genome but is homologous to the *fimV* gene in *B. multivorans* (33.54% of identity). In *Pseudomonas aeruginosa*, it was showed that *fimV* and *truA* forms an operon together, but *fimV* is required for twitching motility whereas *truA* is required for the expression of type III secretory system (Ahn *et al.*, 2004). It is striking that there are less insertions in the *truA* gene at the third instar than at the second instar, hence showing that this gene became more important for the bacterial fitness at later developmental stages of the host.

Similar to the *truA* gene, two other genes showed a specific decrease of insertions at the third instar compared to the second instar condition in the symbiotic organ. These genes were encoding the RNAse G (*BRPE64_RS04515*) which degrades mRNA (Deana and Belasco, 2004), and a DEAD-box helicase protein (*BRPE64_RS05570*) which exerts RNA helicase properties and can also act as a chaperone to promote RNA folding reactions (Jarmoskaite and Russell, 2011). Different genes encoding diverse transcription factors were characterized as essential for the colonization of the M4 organ (*BRPE64_RS11205*, *BRPE64_RS13515*, *BRPE64_RS17825*), especially RNA polymerase sigma factors and anti-sigma factors (*rpoE* (*BRPE64_RS09525*, see **Chapter III**), the anti-sigma/sigma factors *BRPE64_RS17825*/*BRPE64_RS17830*, and the sigma factor *BRPE64_RS19700*). These latter two genes are homologous to the *rpoD* gene from *B. pseudomallei*. Also named σ^{70}, RpoD was showed to regulate virulence factors in *Vibrio*

splendidus with a temperature-dependent manner (Zhang *et al.*, 2019). I also found a number of two-component regulators (*BRPE64_RS02370-BRPE64_RS02375*, *BRPE64_RS06475-BRPE64_RS06480*). Taken together, these results suggest that symbiotic bacteria have to modulate their transcriptional and translational activities in the midgut in response to diverse stress and metabolic conditions. They could be involved in the regulation of the profound transcriptional adaptation of the symbiotic bacteria in the midgut as it was described before (Ohbayashi *et al.*, 2019b). Moreover, some of these regulators (*e.g.* the two-component regulators and the sigma-anti-sigma regulators) could be directly involved in sensing the environmental conditions in the midgut organs and the crypts, and regulate the appropriate responses in the symbiotic bacteria. Thus, these identified regulators constitute a rich resource for future functional studies focusing on the regulation of the bacterial adaptation to the midgut environment (see the recapitulative functions illustrated in the **Figure 129**).

4.3.2. Stress response elements

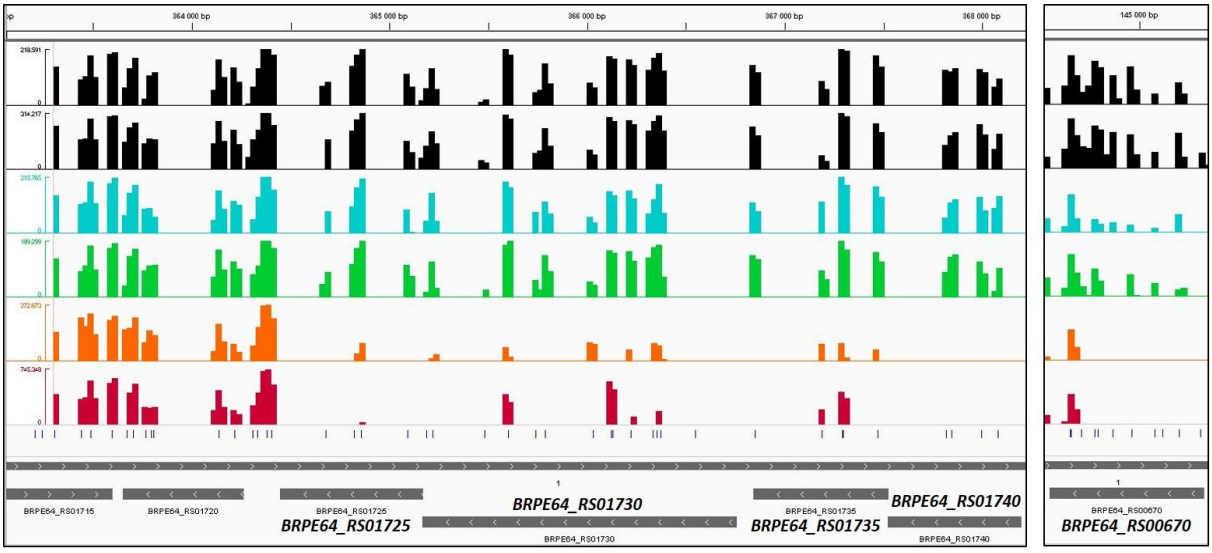


Figure 109: Identification of *in vivo* fitness genes encoding paraquat-inducible proteins and the DsbA protein.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Left picture: *in vivo* fitness genes encoding paraquat-inducible proteins from *BRPE64_RS01725* to *BRPE64_RS01740*. Right picture: *in vivo* fitness gene encoding the DsbA protein (*BRPE64_RS00670*).

In the M1 region, three genes were identified for the *in vivo* bacterial fitness and are encoding a two-component system (*BRPE64_RS04905* and *BRPE64_RS04910*) and a thioredoxin (*BRPE64_RS07700*), which are involved in stress response mechanisms. This two-component system showed homologies with the FixL-FixJ two-component system from *Bradyrhizobium japonicum* which activates the expression of the nitrogen fixation genes only when this two-component system senses low concentrations of O_2 (Wright *et al.*, 2018). The other gene encodes a thioredoxin which acts as a thiol-disulphide interchange protein, and was showed to be activated in response to oxidative stress or to a decrease of thiol-disulphide ratio of proteins (Prieto-Alamo *et al.*, 2000).

Besides the DNA repair functions described above suggesting a genotoxic stress in the midgut, many other bacterial stress response elements were found when symbiotic bacteria were present at the level of the symbiotic organ. Predominant requirements for gut colonization are the protein quality control mechanisms. One of these fitness genes, named as *dsbA* (*BRPE64_RS00670*) (**Figure 109**), is encoding a thiol-disulphide interchange protein involved in disulphide bond formation as proteins emerge into the periplasm (Manta *et al.*, 2019). Interestingly, an *E. coli dsbA* mutant showed a growth defect only during anaerobic conditions (Meehan *et al.*, 2017). In addition, four genes (*BRPE64_RS01725-BRPE64_RS01740*) (**Figure 109**) organized in a cluster are encoding paraquat-inducible proteins according to the genome annotation. These proteins are not well studied, but it was demonstrated that the transcription of these genes in *E. coli* was induced in the presence of paraquat, a ROS generating chemical (Koh and Roe, 1995, 1996). Additionally, these proteins may encode a transporting system which can contribute to the membrane integrity (Nakayama and Zhang-Akiyama, 2017). Another interesting M4 fitness gene cluster included three genes, *clpP*, *clpX* and *lon* (*BRPE64_RS06530-BRPE64_RS06540*) (**Figure 110**) which are encoding ATP-dependent proteases involved in the heat shock response in *E. coli* (Aertsen *et al.*, 2004). These heat shock induced proteins participate in protein recycling or turnover, and it has been demonstrated that these three proteases are required for the pathogen *Salmonella enterica* serovar Typhimurium to optimally colonize chicken ceca (Troxell, 2016). To complete this stress response, two other genes related to the Clp proteases, *clpS* (*BRPE64_RS02725*) and *BRPE64_RS02730* were identified for the M4 bacterial fitness at the third instar condition. In

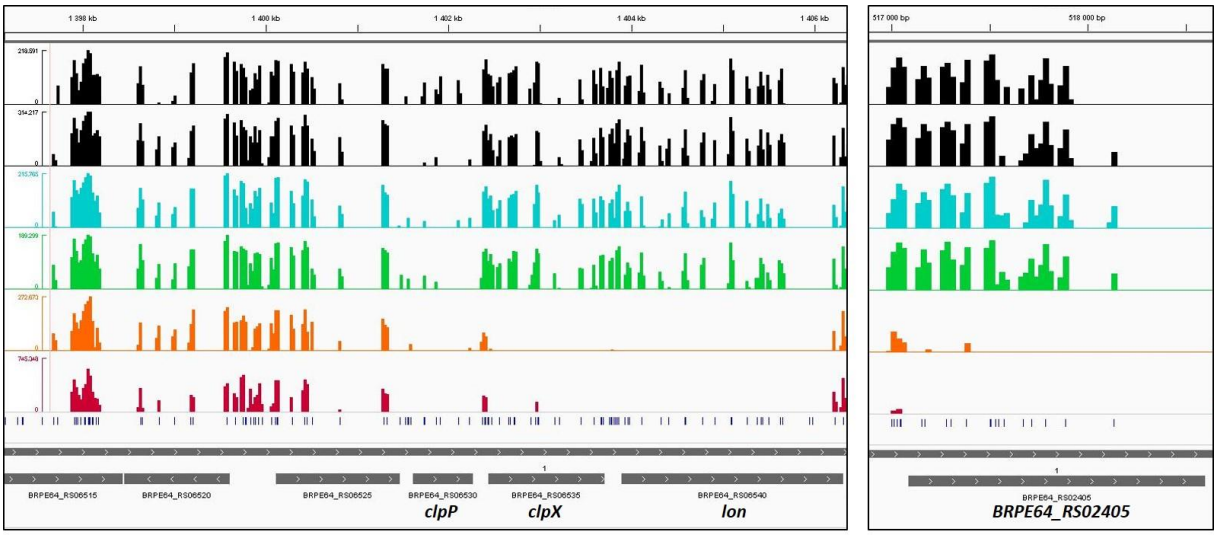


Figure 110: Identification of *in vivo* fitness genes involved in proteolytic cleavages. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Left picture: *in vivo* fitness genes *clpP*, *clpX* and *lon* (from *BRPE64_RS06530* to *BRPE64_RS06540)*. Right picture: *in vivo* fitness gene encoding a protease Do (*BRPE64_RS02405*).

the same category of proteolytic enzymes, one fitness gene (*BRPE64_RS02405*) (**Figure 110**) is encoding a protease Do, which belongs to the peptidase S1C family along with HtrA and MucD proteases. This protease Do may exert both chaperone and protease functions, and it was showed that this protease was essential for the survival of *E. coli* at elevated temperatures (Seol *et al.*, 1991).

Also in the M4 organ, I found another fitness gene named *otsB* (*BRPE64_RS09400*) encoding the trehalose-6-phosphate phosphatase involved in trehalose biosynthesis (Joseph *et al.*, 2010). The accumulation of trehalose inside the bacterial cell is a known mechanism to deal with external osmotic stress (Joseph *et al.*, 2010; Park *et al.*, 2007), hence suggesting that symbiotic bacteria may face some osmotic stresses in the symbiotic organ.

Concerning the second instar condition, I found the *surA* gene (*BRPE64_RS11090*) which encodes for a protein chaperone required for proper protein folding, and also two genes (*BRPE64_RS20095* and *BRPE64_RS20100*) which are homologous to the toxin-antitoxin system MazF/MazE from *B. pseudomallei* (82.93% and 83.02% of identity, respectively). Under stress conditions, especially during antibiotic treatments, the transcription of the *mazEF* genes is reduced which leads to the degradation of the MazE antitoxin by the ClpP and Lon proteases

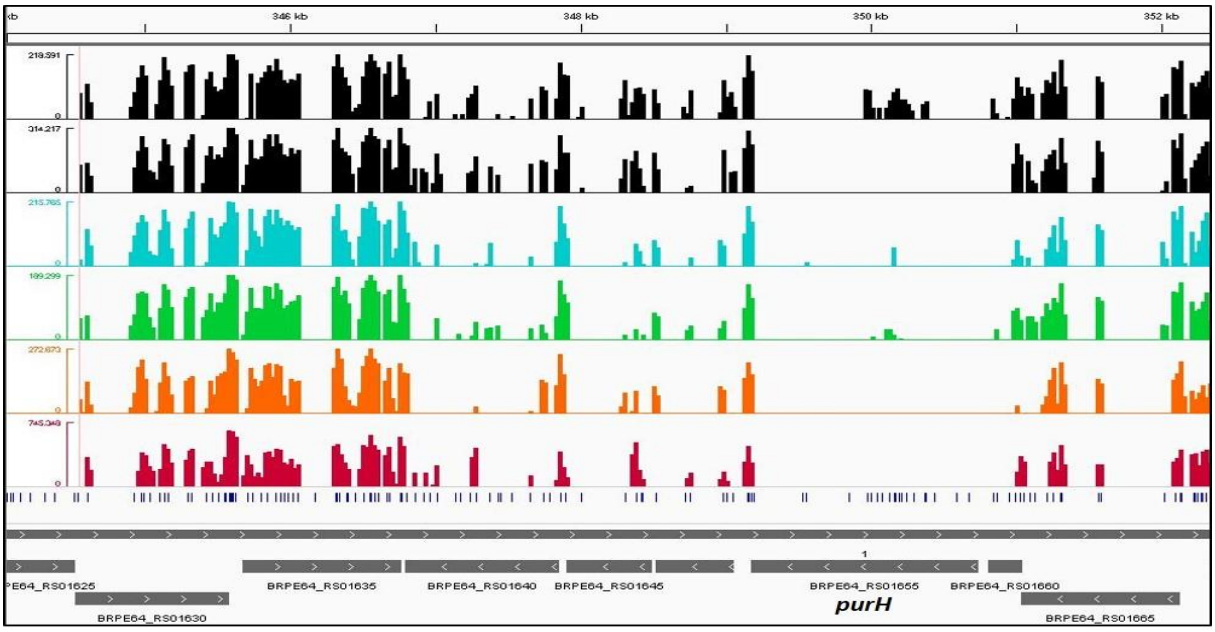


Figure 111: Identification of *in vivo* fitness gene (*purH* or *BRPE64_RS01655*) involved in purine biosynthesis.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey.

in *E. coli* (Tripathi *et al.*, 2014). Interestingly, the ClpP and Lon proteases were also required for the bacterial fitness in the symbiotic organ, as mentioned previously. The released MazF toxin exerts its ribonuclease activity against mRNAs and rRNAs, which blocks the protein biosynthesis and finally ends up to the formation of persister cells (Cho *et al.*, 2017; Curtis *et al.*, 2017). Following this line of reasoning, I can propose the hypothesis that MazF toxin-mediated growth arrest and ClpP/Lon-mediated degradation of the antitoxin MazE are essential for survival of the bacteria in the crypts.

Regarding the third instar condition exclusively, I found a two-component system that senses various external molecules and may be involved in biofilm formation and the regulation of cell motility (*BRPE64_RS02370* and *BRPE64_RS02375*) (Weigel and Demuth, 2016), and another two-component system which is putatively involved in catabolism control (*BRPE64_RS04905* and *BRPE64_RS04910*) (Ohtsubo *et al.*, 2006). In addition, I also found the *BRPE64_RS28700* gene which produces a hypothetical protein which is closely related to the HdeA protein from *B. gladioli* (43.27% of identity). This HdeA protein is a chaperone that is activated upon acidic pH and protects periplasmic proteins against denaturation in *E. coli* (Salmon *et al.*, 2018).

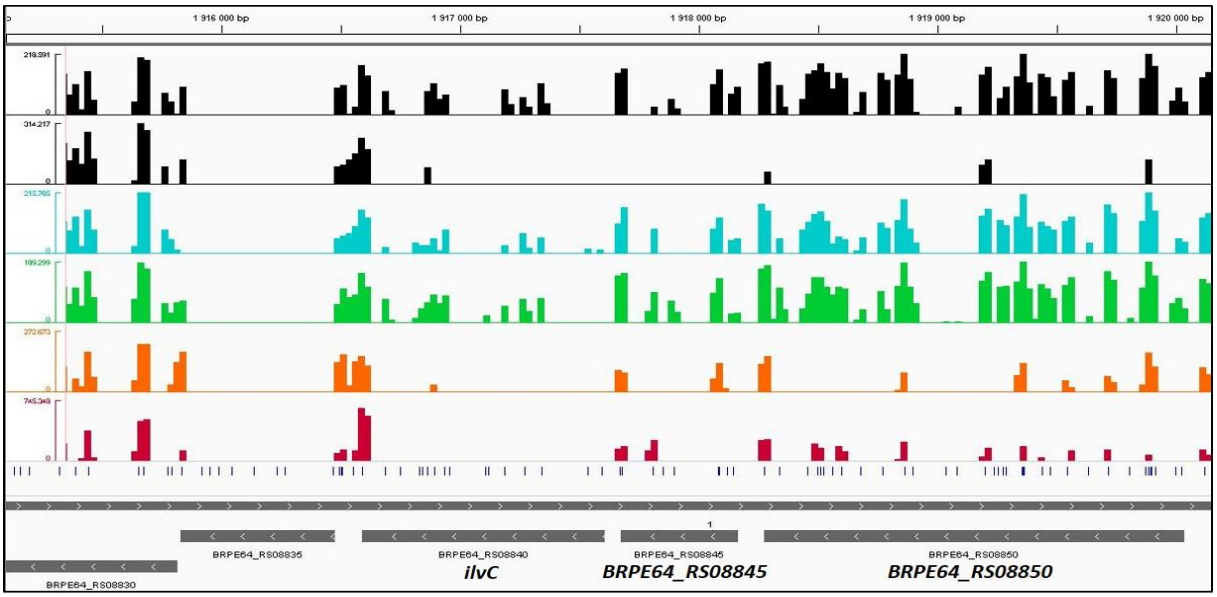


Figure 112: Identification of *in vivo* fitness genes involved in amino acids biosynthesis. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The indicated *in vivo* fitness genes are *ilvC* (*BRPE64_RS08840*), *BRPE64_RS08845* and *BRPE64_RS08850*.

Thus, the *Burkholderia* symbiont mobilizes an arsenal of functions known to be involved in coping with stress to colonize the midgut crypts. Thus, it seems that it faces multiple stress conditions in this environment such as osmotic stress, protein stability and temperature variations, low concentrations of oxygen, oxidative stress, genotoxicity and/or the presence of antibacterial compounds like the AMPs discussed in **Chapter III** (see the recapitulative functions illustrated in the **Figure 129**).

4.3.3. Metabolism

Nucleotide metabolism

In the four *in vivo* conditions, I noticed that multiple genes from the purine *de novo* biosynthesis pathway were required for the bacterial fitness to colonize all the midgut regions (M1, M3 and M4). These identified symbiotic genes were *purH* (*BRPE64_RS01655*) (**Figure 111**), *purM* (*BRPE64_RS02135*) (**Figure 108**), *purC* (*BRPE64_RS02340*), *purE* (*BRPE64_RS02345*), *purK* (*BRPE64_RS02350*), *purL* (*BRPE64_RS06595*), *BRPE64_RS06685*, and *purF* (*BRPE64_RS20020*). In addition to the purine biosynthesis, I found the *cmk* gene (*BRPE64_RS09970*) which produces the cytidylate kinase involved in the pyrimidine

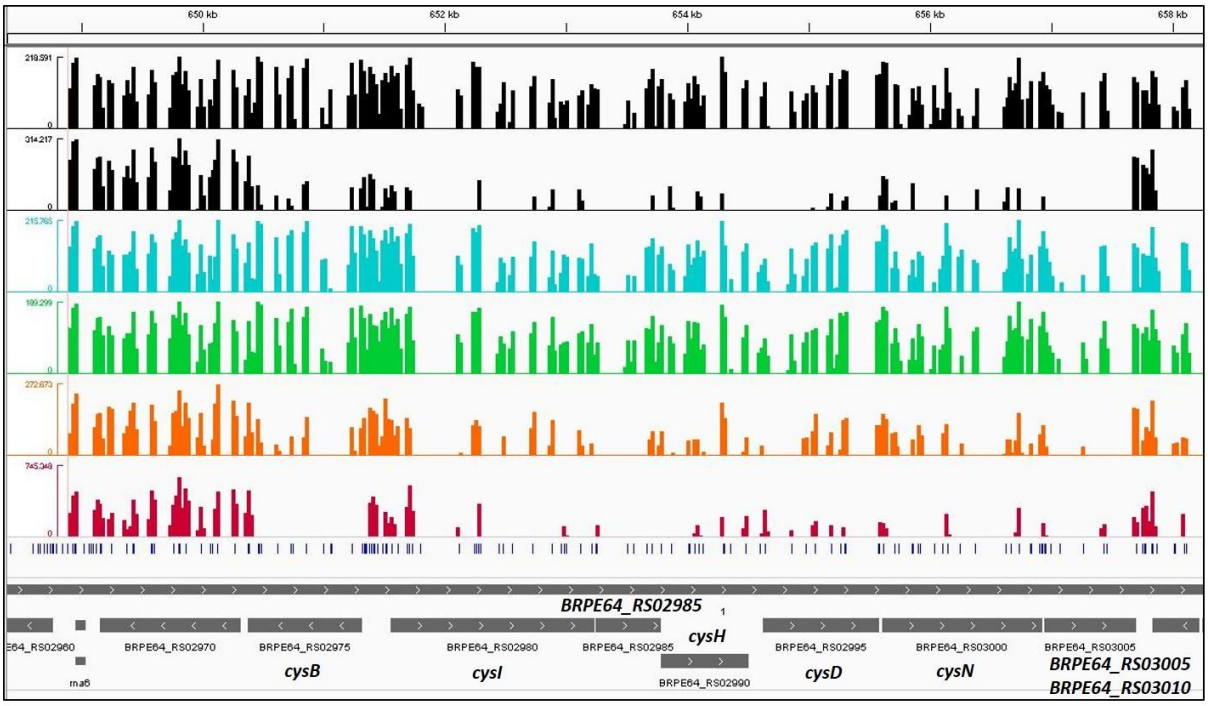


Figure 113: Identification of *in vivo* fitness genes involved in amino acid biosynthesis and vitamin B12 biosynthesis.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The indicated *in vivo* fitness genes are *cysB* (*BRPE64_RS02975*), *cysI* (*BRPE64_RS02980*), *BRPE64_RS02985*, *cysH* (*BRPE64_RS02990*), *cysD* (*BRPE64_RS02995*) and *cysN* (*BRPE64_RS03000*) involved in cysteine biosynthesis, and *BRPE64_RS03005* with *BRPE64_RS03010* are involved in vitamin B12 biosynthesis.

biosynthesis, required only for the M1 region. Thus, the nucleotide biosynthesis pathways are active and required for the *Burkholderia* symbiont during the colonization of the host. This is similar as for growth in the minimal medium and indicates that purines and pyrimidines are not part of the nutrients that the insect provides to its crypt inhabitants. This result is furthermore a beautiful validation of the Tn-seq approach since the *purM* and *purL B. insecticola* mutants were previously reported to be affected in the colonization of the M4 crypts (Kim *et al.*, 2014).

Amino acid metabolism

The biosynthesis pathways of amino acids were only required when symbiotic bacteria were present in the symbiotic organ. Different amino acid biosynthesis genes were identified as essential for the colonization of the M4 region, with the biosynthesis of methionine

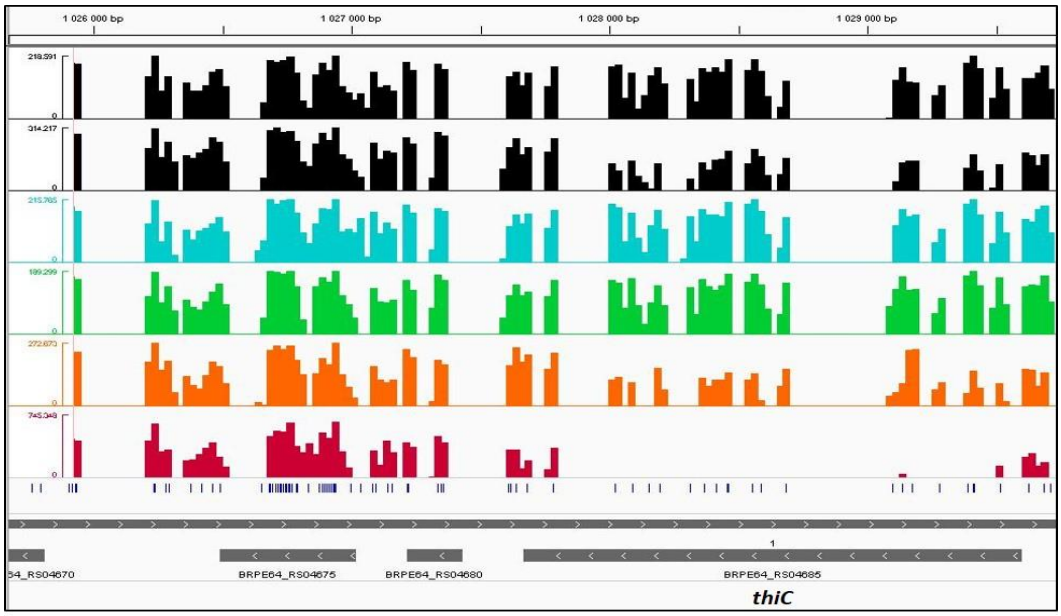


Figure 114: Identification of *in vivo* fitness gene (*thiC* or *BRPE64_RS04695*) involved in vitamin B1 biosynthesis.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey.

(*BRPE64_RS00270*, *BRPE64_RS00275*, *metF*, *BRPE64_RS12995*, *metR*, *metE*), arginine (*argB*, *argH*, *astB*, *BRPE64_RS10830*, *argG*, *argF*), glutamate (*gltD* and *gltB*), lysine (*lysA*), tryptophan (*trpA*, *trpB*, *BRPE64_RS20060*), threonine/methionine (*BRPE64_RS06390*), alanine (*alaT* or *BRPE64_RS06385*) and the branched amino acid valine (*ilvC*, *BRPE64_RS08845*, *BRPE64_RS08850*) (**Figure 112**). Interestingly, genes involved in the biosynthesis of the other two branched amino acids, leucine and isoleucine (*BRPE64_RS01390*, *BRPE64_RS20080*, *BRPE64_RS20085*, *BRPE64_RS20090*), are essential for growth in minimal medium (see **Annexes 2 and 3**) but are not essential for the M4 colonization. This suggests that contrary to the other mentioned amino acids that the crypt symbionts have to produce themselves, the insect host provides leucine and isoleucine as nutrients to the bacterial symbiont. Also the tryptophan biosynthesis pathway is very instructive. In minimal medium, all genes of the pathway (*trpE*, *trpG*, *trpD*, *trpC*, *trpF*, *trpA* and *trpB*) are essential for growth as expected (see **Chapter II and Annexes 2 and 3**). However, the upstream part of the pathway, from chorismate till indole, encoded by *trpE*, *trpG*, *trpD*, *trpC*, and *trpF*, is not required for

colonization of the M4 crypts and only the last step, from indole to tryptophan, encoded by

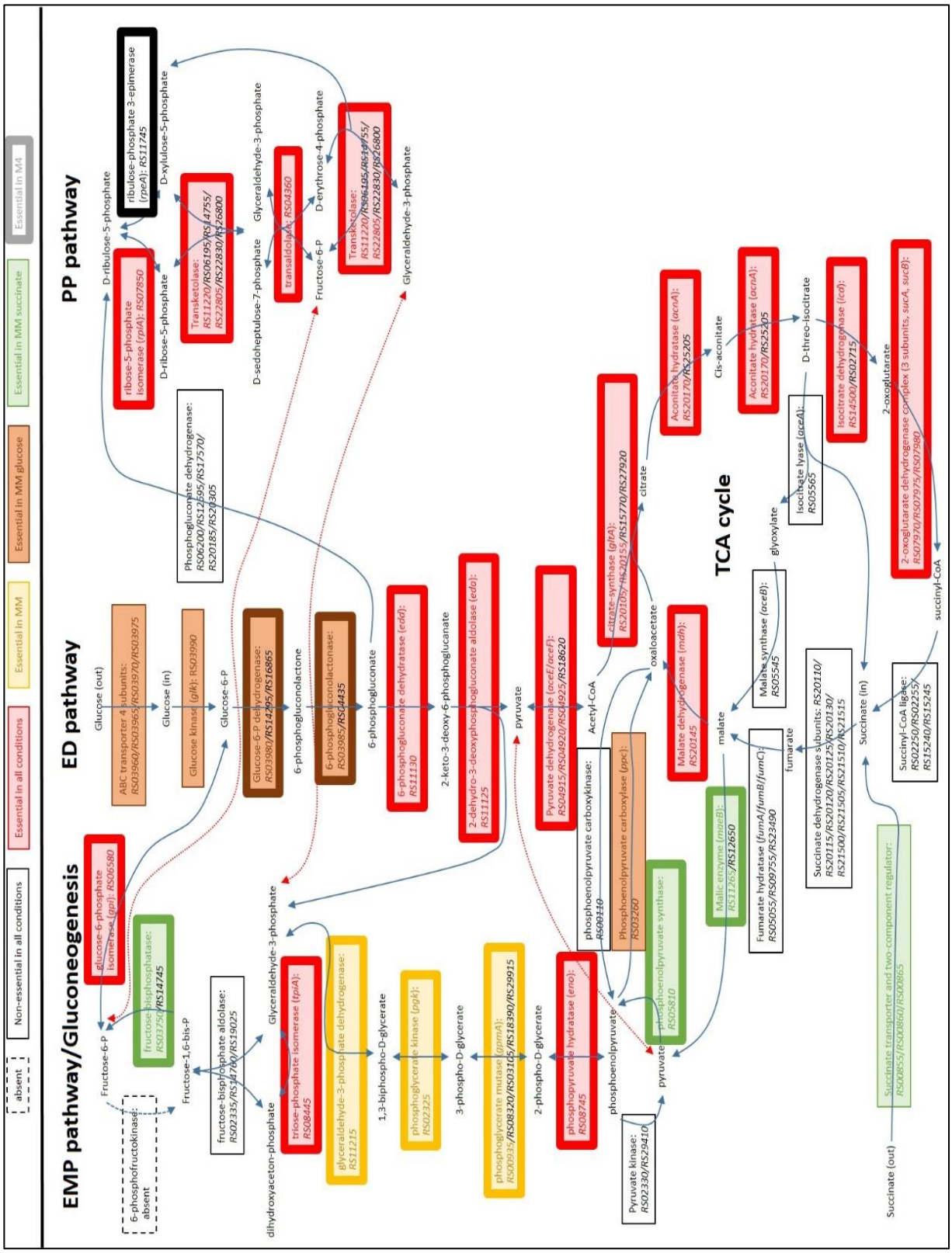


Figure 115: Central metabolism in the M4 crypts.

The representation of central metabolism of *B. insecticola* is as in Figure 35. The genes that are essential for colonization of the M4 crypts are squared in bold.

trpA and *trpB*, is essential also in M4 (see **Annexe 16**). Indole is not part of any other known metabolic pathway in *B. insecticola*. Together, this suggests that the insect is feeding indole to the symbionts. Other genes, related to cysteine biosynthesis were specifically required for the bacterial fitness in the symbiotic organ only at the third instar (*cysB*, *cysI*, *cysH*, *cysD*, *cysN*, *BRPE64_RS19075*) (**Figure 113**).

Vitamin biosynthesis

Several genes involved in vitamin biosynthesis are essential for the bacterial fitness exclusively in the symbiotic organ. At the second instar, only the production of vitamin B6 was required for the symbiont in the symbiotic organ, with the *BRPE64_RS11085* gene, homologous to the *pdxA* gene from *B. cenocepacia* (82.93% of identity). This gene encodes the 4-hydroxythreonine-4-phosphate dehydrogenase. At the third instar, two vitamins are produced by the *Burkholderia* symbiont, the vitamin B1 also known as thiamine (*thiC*, *thiG*) (**Figure 114**) (Palmer and Downs, 2013) and the vitamin B12 also called cobalamin (*BRPE64_RS03005*, *BRPE64_RS03010*) (**Figure 113**) (Fang *et al.*, 2017). As vitamins are required as cofactors to promote the activity of various enzymatic reactions, these biosynthesis

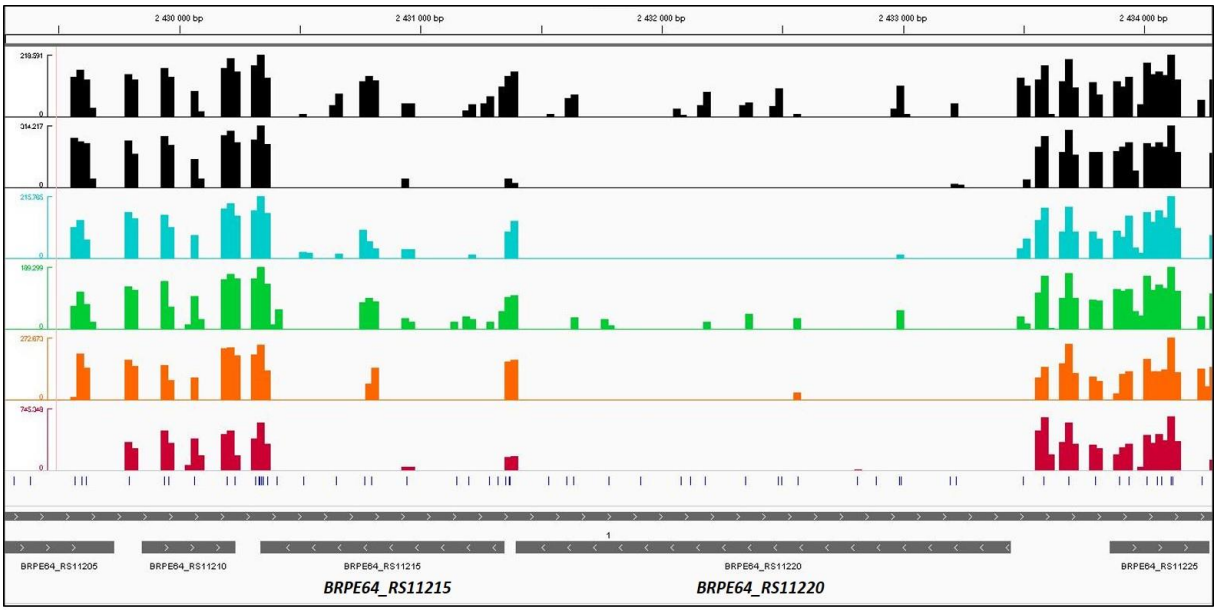


Figure 116: Identification of *in vivo* fitness genes involved in glycolysis.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The indicated *in vivo* fitness genes are *BRPE64_RS11215* and *BRPE64_RS11220.*

pathways identified may be involved in the metabolic activities of the symbiont when it reaches the symbiotic organ. In addition, these vitamins might also be produced by the bacterial symbiont to provide these compounds for its host (Ohbayashi *et al.*, 2019).

Central carbon metabolism

As I mentioned before, I found that the gluconeogenesis and the TCA cycle are essential for the bacterial viability, and that ED pathway was the main glycolysis pathway used by *B. insecticola* in free-living condition to degrade glucose (see **Chapter II Figure 35**). Mapping the fitness genes in the M4 crypts on these pathways (**Figure 115**) demonstrate that the gluconeogenesis, starting from the TCA cycle with the malic enzyme (*BRPE64_RS11265*) and phosphoenolpyruvate synthase (*BRPE64_RS05810*) and ending with the fructose-bisphosphatase (*BRPE64_RS03750*), is essential in the M4 organ (**Figures 115 and 116**). In addition, the ED and PP pathways, and the TCA cycle are essential in the M4 organ (**Figure 115**). On the other hand, the succinate and glucose transporters, as well as the glucose kinase (*BRPE64_RS03990*) are not essential for M4 colonization (**Figure 115**), indicating that these metabolites are not nutrients for the symbiotic bacteria. Taken together, this analysis suggests that the symbiont has to produce its sugars and other carbon skeletons via gluconeogenesis

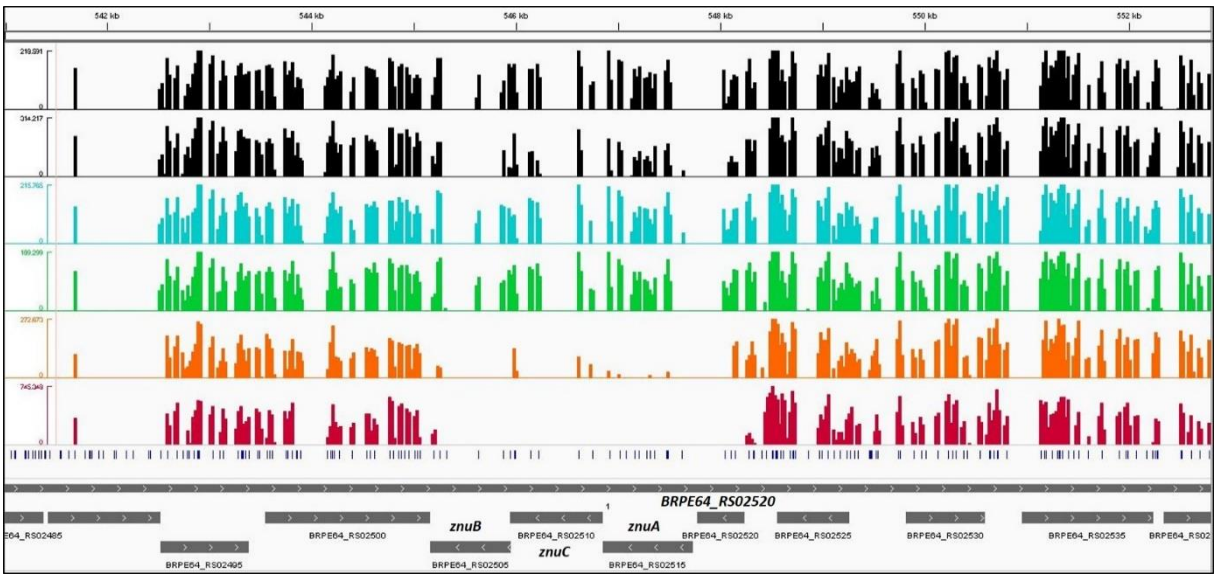


Figure 117: Identification of *in vivo* fitness genes involved in zinc import.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The indicated *in vivo* fitness genes are *znuB* (*BRPE64_RS02505*), *znuC* (*BRPE64_RS02510*), znuA (*BRPE64_RS02515*) and *BRPE64_RS02520*.

and the PP pathways.

In the M3 organ, I exclusively identified one gene (*BRPE64_RS00910*) encoding a carbohydrate transporting system named phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) (Postma *et al.*, 1993). Here, this PTS is annotated as a type IIA PTS component fructose subfamily, which seems to transport gluconeogenic carbon sources inside the bacterial cell.

Inorganic ion transport

I identified a cluster of M4 fitness genes made of the *znuABC* genes (*BRPE64_RS02505-BRPE64_RS02515*) and the *BRPE64_RS02520* gene (homologous to the *zur* gene from *E. coli*) (**Figure 117**). These genes encode the high affinity zinc transporter ZnuABC and the transcriptional repressor Zur that controls the expression of genes in response to zinc availability (Gabbianelli *et al.*, 2011). This transporter is required for the bacterial growth in environments with very low zinc availability in order to maintain the zinc homeostasis (Bhubhanil *et al.*, 2014; Gabbianelli *et al.*, 2011). Additionally, this transporter was shown to be required for virulence and efficient host colonization in pathogenic bacteria such as *V.*

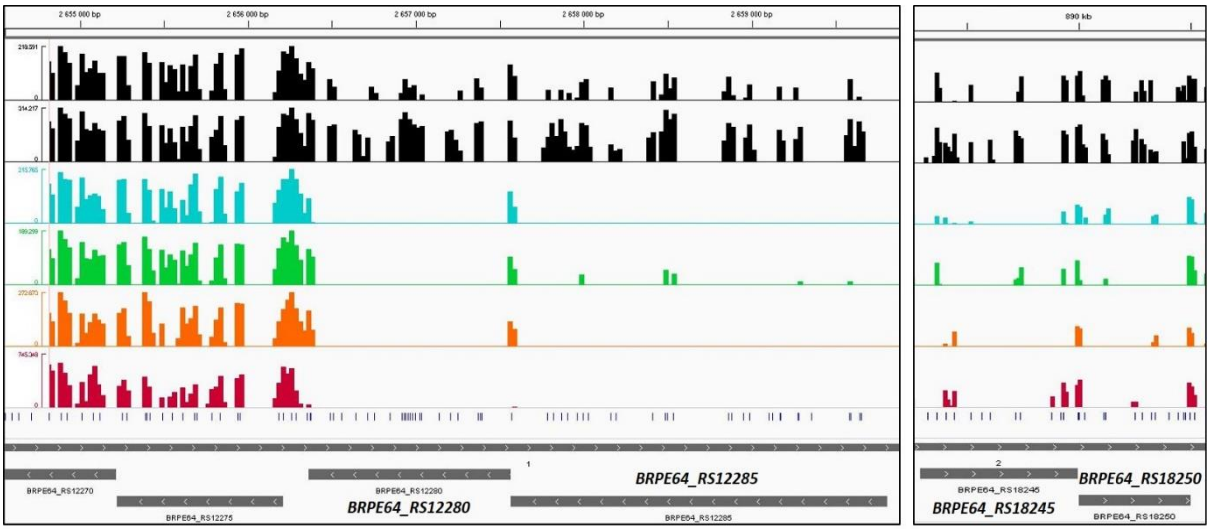


Figure 118: Identification of *in vivo* fitness genes involved in the respiratory chain. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 (left picture) and chromosome 2 (right picture) are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Left picture: the indicated *in vivo* fitness genes are *BRPE64_RS12280* and *BRPE64_RS12285* involved in cytochrome c biogenesis. Right picture: the indicated in vivo fitness genes are *BRPE64_RS18245* and *BRPE64_RS18250*.

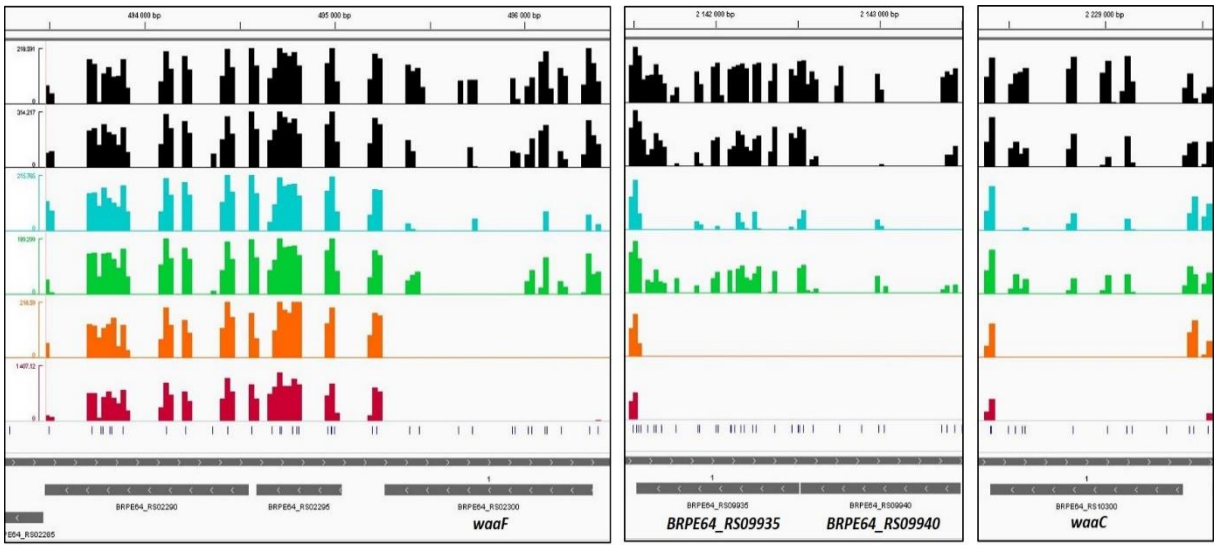


Figure 119: Identification of *in vivo* fitness genes involved in the core oligosaccharide biosynthesis (LPS biosynthesis pathway).

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The indicated *in vivo* fitness genes are *waaF* (*BRPE64_RS02300*), *BRPE64_RS09935*, *BRPE64_RS09940* and *waaC* (*BRPE64_RS10300*).

cholerae and *Yersinia pestis* (Bobrov *et al.*, 2017; Sheng *et al.,* 2015). Another protein belonging to the inorganic ion transport and metabolism COG category was identified as important for the bacterial fitness only at the third instar host stage, which is the MgtC/SapB protein (*BRPE64_RS20660*). The MgtC transporter was reported to be essential for the bacterial growth under magnesium limiting conditions in *B. cenocepacia* (Maloney and Valvano, 2006). Interestingly, it was reported that this transporter was required for intracellular survival of different pathogens inside macrophages, such as *Mycobacterium tuberculosis*, *S. enterica* and *B. cenocepacia* (Belon and Blanc-Potard, 2016; Maloney and Valvano, 2006). Hence, the requirement of these proteins for the symbiont suggest that the symbiotic organ constitutes an environment with limited amounts of inorganic ions such as Mg^{2+} and Zn^{2+}. Interestingly, limiting the availability of nutrient metals such as iron, manganese, magnesium and zinc is a strategy, known as nutritional immunity, employed by hosts to control microbial infections and conversely, microbes express high affinity transporters to steal these metal nutrients from the host (Corbin *et al.*, 2008; Kehl-Fie and Skaar, 2010). Thus, in the light of this concept of nutritional immunity, my results suggest that *R. pedestris* limits the availability of zinc and magnesium specifically in the midgut crypts as a

strategy for controlling the symbiont population, and perhaps as part of a mechanism for symbiont selection as well.

Energy production

Two genes, which encode cytochrome c-type assembly proteins, *nrfE* (*BRPE64_RS12280*) and *BRPE64_RS12285* and two other genes, *BRPE64_RS18245* and *BRPE64_RS18250* that are encoding the two subunits of the cytochrome bd oxidase, are essential for the colonization of all the midgut regions (**Figure 118**). These cytochrome c-type proteins constitute essential components of respiratory electron transfer chains in bacteria in order to produce energy (Ahuja *et al.*, 2009; Le Brun *et al.*, 2000). The latter two genes are closely related to the *cioA* and *cioB* genes from *B. pseudomallei*, respectively, and are also involved in the respiratory process. This specific cytochrome *bd* oxidase is required for sustaining the bacterial growth under microaerobic or anaerobic conditions (Cunningham *et al*., 1997; Fischer *et al.*, 2018).

Apart of these respiratory chain genes, I found another fitness gene named *cyaY* (*BRPE4_RS12255*) which is required only at the third instar. This gene produces the protein CyaY that is playing a role in iron-sulfur [Fe-S] cluster synthesis in bacteria (Layer *et al.*, 2006). These [Fe-S] clusters are known to participate in electron transfer of the respiratory chain, in oxidation-reduction enzymatic reactions, and are primary used as iron-sulfur cellular storage (Johnson *et al.*, 2005). Thus the symbiont activates different routes of respiratory chains during the host development in order to produce sufficient amounts of energy for its biological processes.

4.3.4. Envelope biogenesis functions

A significant proportion of *in vivo* required genes belong to the cell wall biogenesis pathways (**Figure 106**). Among these envelope functions, the majority participates in the biosynthesis of LPS, one of the major cell envelope components of Gram-negative bacteria. As mentioned previously, LPS is made of a lipid anchor called lipid A, an elongated core oligosaccharide, and the most external part constitutes the O-antigen (see **Chapter I**). The identified bacterial genes required for *in vivo* colonization are responsible for the biosynthesis of the core oligosaccharide and the O-antigen of the LPS molecule. For the core oligosaccharide component, I identified the genes *waaC* (*BRPE64_RS10300*) (**Figure 119**) and *waaF* (*BRPE64_RS02300*) (**Figure 119**) that were specific to the M4 organ and are involved in the

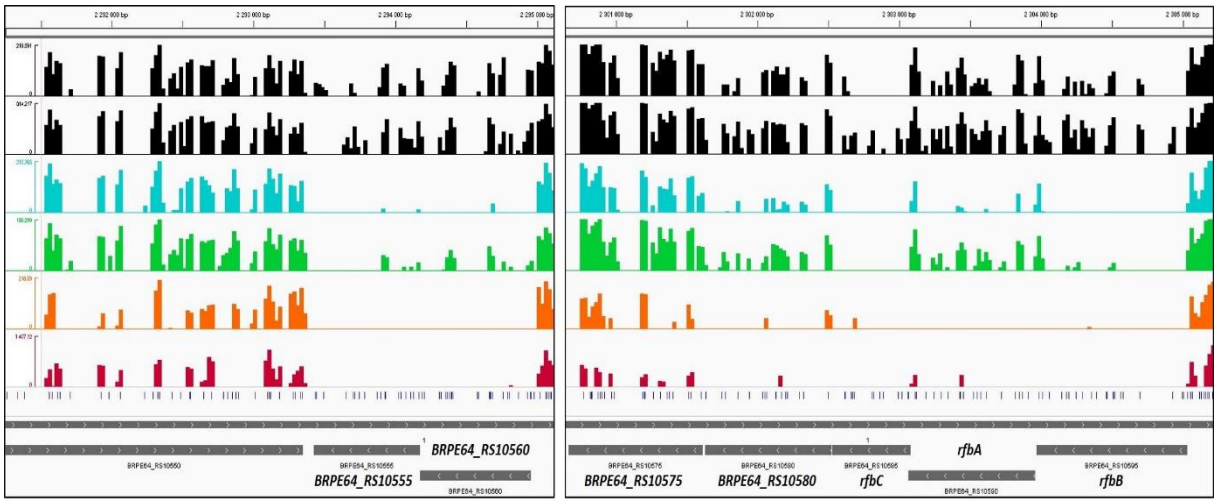


Figure 120: Identification of *in vivo* fitness genes involved in the O-antigen biosynthesis and export (LPS biosynthesis pathway).

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Left picture: the indicated *in vivo* fitness genes are *BRPE64_RS10555* and *BRPE64_RS10560*, involved in the export of the O-antigen. Right picture: the indicated in vivo fitness genes are *BRPE64_RS10575*, *BRPE64_RS10580*, *rfbC* (*BRPE64_RS10585*), *rfbA* (*BRPE64_RS10590*) and *rfbB* (*BRPE64_RS10595*), involved in the O-antigen biosynthesis.

biosynthesis of the inner core oligosaccharide. In addition, three other genes were required for the colonization of the entire midgut regions with *rfaD* (*BRPE64_RS09935*) (**Figure 119**) and the two copies of *rfaE* (*BRPE64_RS00755* and *BRPE64_RS09940*) (**Figure 119**). For the outer core oligosaccharide component, I found four genes *BRPE64_RS10475*, *BRPE64_RS10480*, *BRPE64_RS10485* and *BRPE64_RS10490* organized in a cluster which are homologues of the *wbiI*, *wbiH*, *wbiG* and *wbiF* genes from *B. cenocepacia*, respectively.

Concerning the O-antigen biosynthesis, I found a gene cluster with five genes, *BRPE64_RS10575*, *BRPE64_RS10580*, *rfbC* (*BRPE64_RS10585*), *rfbA* (*BRPE64_RS10590*) and *rfbB* (*BRPE64_RS10595*) which were required for the bacterial fitness in all *in vivo* conditions (**Figure 120**). This gene cluster is responsible for the biosynthesis of dTDP-L-rhamnose, which is a precursor for rhamnose incorporation in the O-antigen (Vinion-Dubiel and Goldberg, 2003). Four other genes that are encoding for glycosyl transferases (*BRPE64_RS04485*, *BRPE64_RS04490*, *BRPE64_RS04495*) and the O-antigen polymerase (*BRPE64_RS04500*) were also identified as *in vivo* fitness genes, and are involved in the transfer of sugar moieties to the O-antigen. After completion of the O-antigen synthesis, it has to cross the bacterial

membranes to be attached to the core oligosaccharide of LPS molecules. I have found the two genes *BRPE64_RS10555* and *BRPE64_RS10560* (**Figure 120**) that encode the Wzm/Wzt O-antigen export system that is important for the symbiont in all *in vivo* conditions tested. Thus, the two external components of LPS molecules, the core oligosaccharide and the O-antigen are very important for symbiotic bacteria during the colonization of the host. Although the statistical analysis qualifies some of the LPS genes only required for colonization of the M4 organ, it is obvious that the number of reads corresponding to these genes decreases gradually from free-living bacteria till the M4 colonization indicating that the LPS structure contributes to the bacterial fitness all along the path of the midgut till colonization of the crypts.

Another member of the cell wall constituents was detected in the *in vivo* Tn-seq which was the Tol-Pal complex. The full complex is made of several subunits, the TolQ, TolR, TolA, TolB and Pal proteins which are encoded by *tolQ* (*BRPE64_RS11025*), *BRPE64_RS11030*, *BRPE64_RS11035*, *tolB* (*BRPE64_RS11040*) and *BRPE64_RS11045*, respectively (**Figure 121**) (Lloubès *et al.*, 2001). These different protein subunits interact together and contribute to the

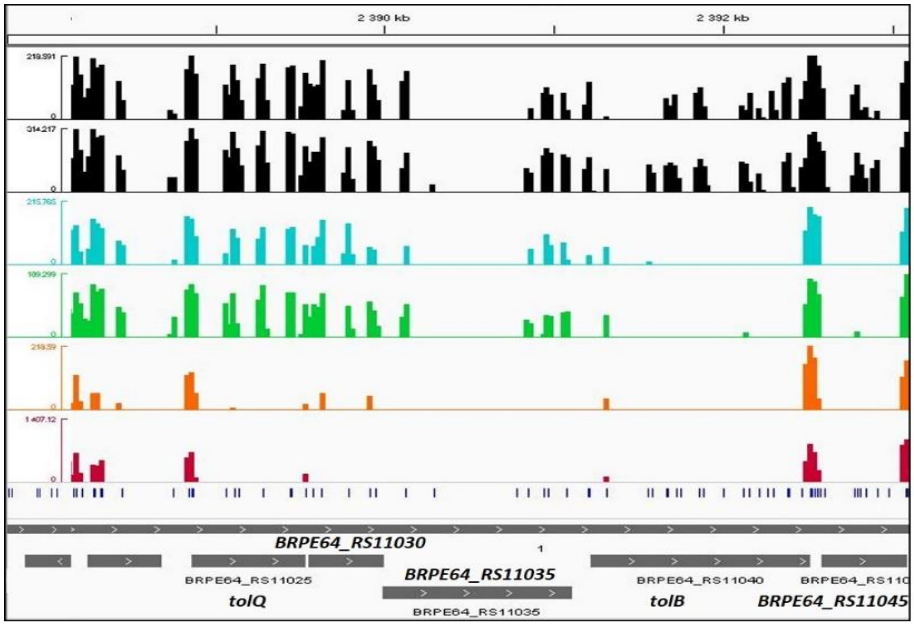


Figure 121: Identification of *in vivo* fitness genes encoding the Tol-Pal complex.
The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The indicated in vivo fitness genes are *tolQ* (*BRPE64_RS11025*), *BRPE64_RS11030*, *BRPE64_RS11035*, *tolB* (*BRPE64_RS11040*) and *BRPE64_RS11045*.

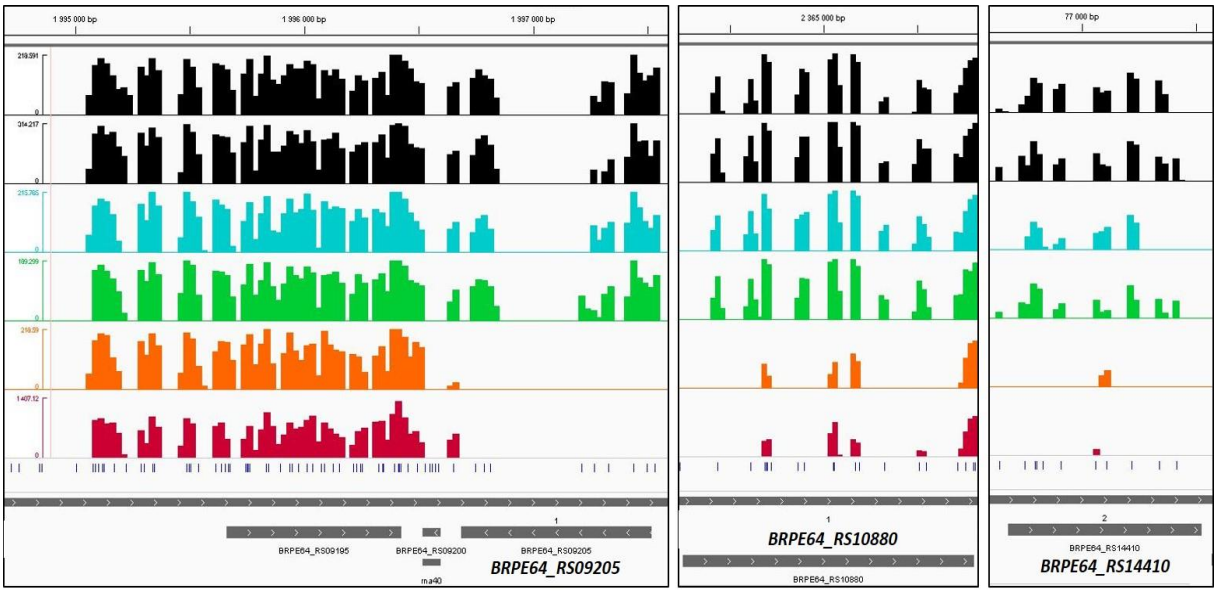


Figure 122: Identification of *in vivo* fitness genes involved in peptidoglycan and hopanoid biosynthesis.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 (left and middle pictures) and chromosome 2 (right picture) are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. From the left to the right: *BRPE64_RS09205* and *BRPE64_RS10880* involved in peptidoglycan biosynthesis, and *BRPE64_RS14410* involved in hopanoid biosynthesis.

membrane maintenance and the export of several membrane constituents (Godlewska *et al.*, 2009; Lazzaroni *et al.*, 1999). Similarly as for the LPS biosynthesis genes, as the symbiont progressively colonizes the different midgut sections, from the M1 to the M4, the number of reads in these genes was progressively decreasing, which demonstrates the requirement for the whole complex in the host colonization process.

Another cluster of genes (*BRPE64_RS12110-BRPE64_RS12130*) encoding the MlaCADEF proteins, constituting a transport complex involved in phospholipid transport and maintenance of lipid asymmetry in the outer membrane was characterized as essential for *in vivo* bacterial fitness in the various midgut sections. These proteins contribute to membrane tolerance by regulating the fluidity of bacterial membranes usually following an osmotic stress (Segura *et al.*, 2012). In agreement, the previously mentioned *otsB*, involved in trehalose accumulation inside the cell and tolerance to osmotic stress (see section **4.3.2**), is also required for the M4 colonization.

Interestingly, I found a specific bacterial fitness gene called *dedA* (*BRPE64_RS02150*) (**Figure**

108) that was strictly required in the M1 organ and contributed to the colonization of the other midgut regions. This gene is encoding a membrane-associated protein that belongs to the ancient DedA membrane protein family (Doerrler *et al.*, 2013). Little is known about this protein, however, some members of this DedA family were reported to participate in the maintenance of the membrane proton motive force, in heat shock response and also in AMP resistance (Doerrler *et al.*, 2013; Kumar and Doerrler, 2014; Tzeng and Stephens, 2015).

In addition to these cell envelope functions, I identified numerous bacterial genes that are specific for the colonization of the M4 organ, at both second and third instar developmental stages, and that are involved in peptidoglycan biosynthesis. These genes include *BRPE64_RS06035* which encodes the penicillin–binding transmembranar protein MrcA, *BRPE64_RS09205* that produces the undecaprenyl-diphosphatase UppP (**Figure 122**), *BRPE64_RS10880* which corresponds to the *amiC* gene that encodes the N-acetylmuramoyl-L-alanine amidase (**Figure 122**), and *BRPE64_RS00675* encoding a peptidoglycan-binding SPOR-domain protein. Additionally, I have also found genes that encode putative lipoproteins and outer membrane proteins that could be potentially contributing to the integrity of bacterial membranes (*BRPE64_RS01090*, *BRPE64_RS01725*, *BRPE64_RS06685*, *BRPE64_RS12260*, *BRPE64_RS19345*). Another gene named *cvpA* (*BRPE64_RS20025*) that is located upstream of the *purF* gene (*BRPE64_RS20020*), was also detected as important for the bacterial fitness in the M4 organ. This gene encodes an inner membrane protein that was previously identified as a biofilm modulator in uropathogenic *E. coli* (Hadjifrangiskou *et al.*, 2012; Shaffer *et al*., 2017). Interestingly, the gene *BRPE64_RS14410* that is encoding a putative squalene/phytoene synthase, closely related to the *hpnD* gene from *Paraburkholderia fungorum* (73.48% of identity), was identified as a symbiotic factor for the M4 colonization (**Figure 122**). This putative squalene synthase was reported to catalyse the formation of squalene from the coupling of two molecules of farnesyl-diphosphate (Pan *et al.*, 2015; Welander *et al.*, 2012). Although squalene is the precursor molecule for hopanoid biosynthesis, sterol components of the symbiont membranes, squalene is also known to be used as a potential carbon source in bacteria (Ghimire *et al.*, 2016).

Thus, in conclusion, it seems that the cell surface of the symbiotic bacteria is of extreme importance for the colonization of the midgut, starting from M1 and then becoming more and more important till the M4. The cell envelope is a compartment that is of a general importance

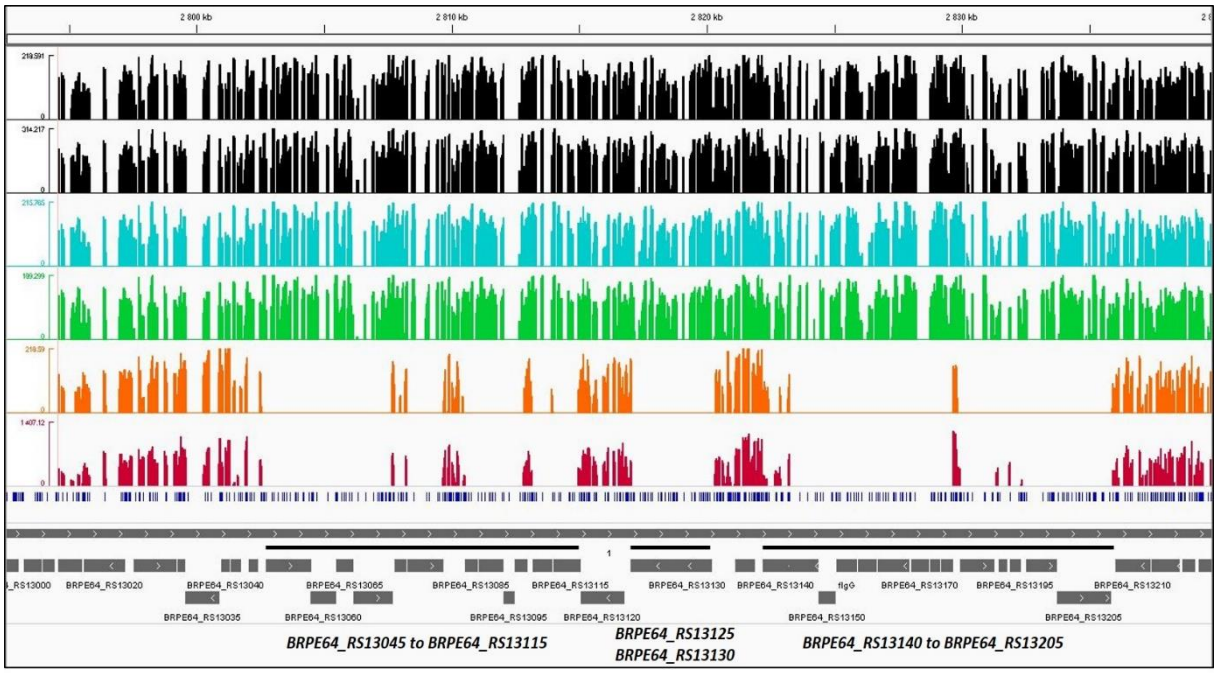


Figure 123: Identification of fitness genes specific for the colonization of the symbiotic organ involved in flagellar motility.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes that are specific for the symbiotic organ colonization are outlined by 3 black lines which correspond to the 3 regions of fitness genes: from *BRPE64_RS13045* to *BRPE64_RS13115*, from *BRPE64_RS13125* to *BRPE64_RS13130*, and from *BRPE64_RS13140* to *BRPE64_RS13205*.

for bacterial viability, but the features that I discovered here are very specific since the concerned genes do not affect the viability of the bacteria in the free-living condition. The necessity of some of these features such as the LPS biosynthesis, the Tol-Pal complex or the MlaCADEF phospholipid transporter can be explained by the very strong challenge of the bacteria with AMPs in the midgut (see **Chapter III and section 4.4**), but other functions such as the peptidoglycan modifying enzymes or the PqiABC transporter that regulates membrane stability do not seem to affect sensitivity towards AMPs. This suggests that, besides the AMPs, other types of stress factors are present in the midgut. It will be of interest in the future to identify these conditions. One exciting possibility to tackle this challenge would be to use again the Tn-seq approach and subject the transposon library *in vitro* to various potential midgut stresses (*e.g.* pH, osmotic stress, temperature, oxidative stress). Fitness genes for growth in these conditions that would be common with the here identified gene set for midgut

colonization could provide indications on the type of stress conditions that are important in the midgut (see the recapitulative functions illustrated in the **Figure 129**).

4.3.5. Motility and chemotaxis

The functional COG category of cell motility was specifically identified from the *in vivo* Tn-seq data in the M4 organ conditions, both at the second and the third instar. The majority of the fitness genes identified in this category are involved in flagellar motility. This type of cell motility includes: the flagellar assembly proteins with the *fliDC* genes (*BRPE64_RS00520*, *BRPE64_RS00525*), the *fliSEFGHIJKLMOPQR* gene cluster (*BRPE64_RS13045-BRPE64_RS13115*) (**Figure 123**), the *flhBA* genes (*BRPE64_RS13200* and *BRPE64_RS13205*) (**Figure 123**), *BRPE64_RS17375* and *BRPE64_RS17380*; the flagellar hook-associated proteins and basal-body rod proteins encoded by the *flgLKJIHGFECBA* gene cluster (*BRPE64_RS13125-BRPE64_RS13185*) (**Figure 123**); and the associated transcription factors and regulatory

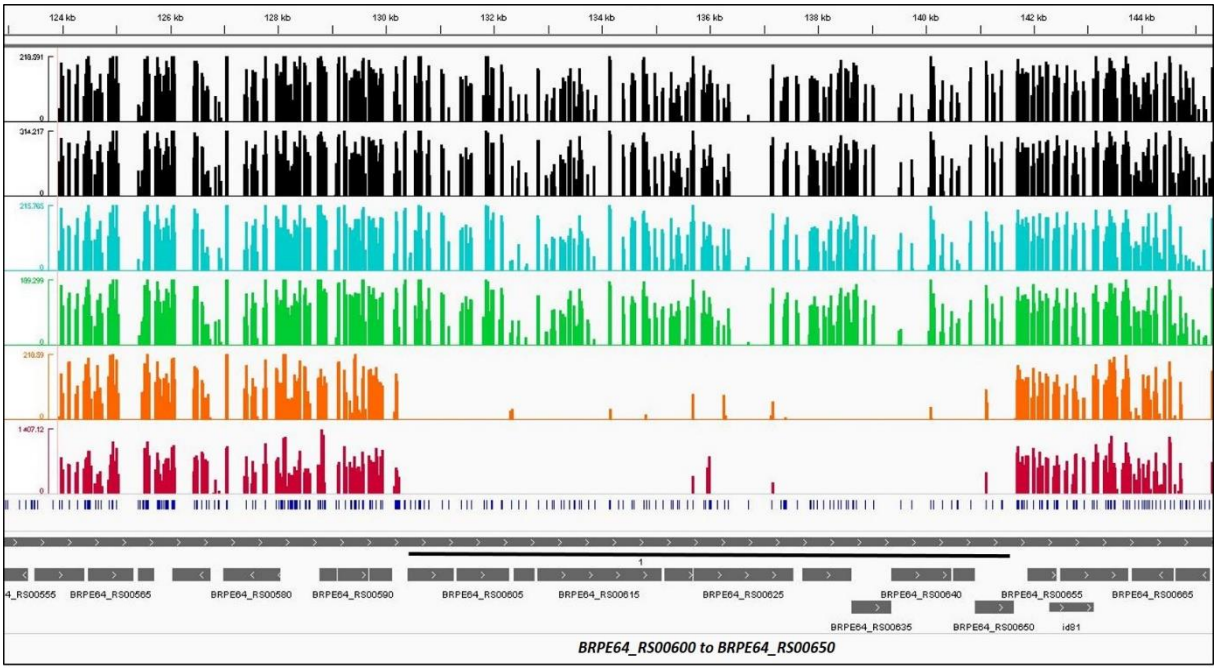


Figure 124: Identification of fitness genes specific for the colonization of the symbiotic organ involved in chemotaxis.

The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. Fitness genes that are specific for the symbiotic organ colonization are outlined by a black line which correspond to fitness genes from *BRPE64_RS00600* to *BRPE64_RS00650*.

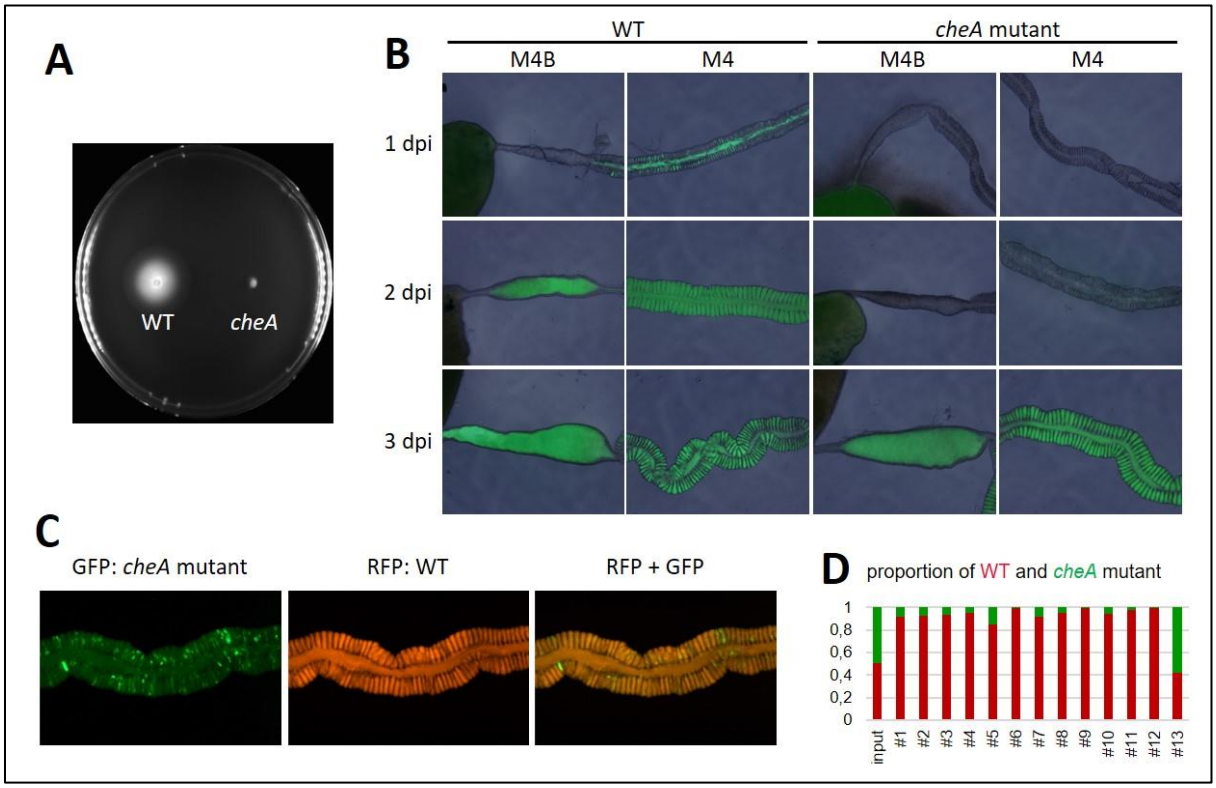


Figure 125: Chemotaxis is essential for M4 crypt colonization.
A) A *cheA* mutant has lost its motility in a soft agar assay. B) In single infections, in the absence of competition, the mutant is delayed in infection but is still able to establish the symbiosis after three days-post-infection (dpi). C) In a mixed infection with the wild-type strain, the *cheA* mutant is outcompeted at three dpi. The wild-type strain was marked by RFP (red) and the mutant strain by GFP (green). D) Quantification by flow cytometry of the *in vivo* competition between the wild-type (red) and *cheA* mutant (green) represented as relative proportions in 13 different insects. The input corresponds to the inoculum of the second instar nymphs. Abbreviations: WT: wild-type.

proteins encoded by *fliA* (*BRPE64_RS13275*), *fliS* (*BRPE64_RS13045*) (**Figure 123**), *fliJ* (*BRPE64_RS13075*) (**Figure 123**), *flgA* (*BRPE64_RS13185*) (**Figure 123**), *BRPE64_RS13195* (**Figure 123**), *BRPE64_RS13280* and *BRPE64_RS13285* (Rajagopala *et al*., 2007). Another cell motility gene *BRPE64_RS20070*, homologous to the *fimV* gene from *B. multivorans* (33.54% of identity) was also detected as a fitness gene for the colonization of the symbiotic organ. This gene forms an operon with the *truA* gene (*BRPE64_RS20065*), and was reported to be involved in twitching motility in *P. aeruginosa* (Ahn *et al.*, 2004; Semmler *et al.*, 2000).

Linked to flagellar motility, I also identified several genes involved in the chemotaxis signalling network which senses a chemical gradient (e.g. chemoattractants or chemorepellents) and stimulates the flagellar proteins to induce bacterial movement towards or away from this gradient (Baker *et al.*, 2006). Among these chemotaxis M4-required genes, there were two

methyl-accepting chemotaxis proteins (MCPs) encoding genes, *tsr* (*BRPE64_RS00625*) (**Figure 124**) and *BRPE64_RS20670* (homologous to the *tar* gene from *B. cenocepacia* with 65.25 % of identity). These membrane-associated proteins act as chemosensors and were reported to detect specific chemical ligands such as aspartate, glutamate and maltose for the Tar protein and serine, alanine and glycine for the Tsr protein (Baker *et al.*, 2006; Callahan and Parkinson, 1985; Hedblom and Adler, 1980; Wang and Koshland, 1980). Besides these two MCP-encoding genes, the *B. insecticola* genome carries 11 other MCP-encoding genes. Once these MCPs are activated by these chemoattractants, they transduce this signal to the chemotaxis Che proteins (Baker *et al.*, 2006), which were also identified as required for the M4 colonization. These proteins correspond to bacterial fitness genes that are organized in a cluster together with the *tsr* gene, and including *BRPE64_RS00610*, *BRPE64_RS00615*, *BRPE64_RS00620*, *BRPE64_RS00630*, *cheD* (*BRPE64_RS00635*), *BRPE64_RS00640*, *BRPE64_RS00645* and *cheZ* (*BRPE64_RS00650*) (**Figure 124**). These intracellular Che proteins transmit this signal to the flagellar motor proteins MotA and MotB which change the rotation directions of the flagella to move the symbiont towards or away from the attractant or from the repellent (Baker *et al.*, 2006). These two MotA and MotB proteins are encoded by the *BRPE64_RS00600* and *BRPE64_RS00605* genes (**Figure 124**) that are part of the chemotaxis gene cluster, and were also required for the bacterial fitness in the symbiotic organ. Hence, cell motility and chemotaxis are among the most important requirements for symbiotic bacteria in order to colonize efficiently the M4 region. This Tn-seq data confirms a previous study that demonstrated the importance of flagellar motility to infect the M4 crypts (Ohbayashi *et al.*, 2019). In addition, the Tn-seq results now demonstrate that the chemotaxis is also important, suggesting that symbiotic bacteria might move towards a specific chemical attractant produced in the M4 region to reach and colonize the symbiotic organ. The identification of two MCPs might set the stage for the discovery of these molecules which, on its turn, might pave the way to interfere with symbiosis (as a pest control strategy) using ligand-mediated saturation of the chemoreceptors.

To verify the newly discovered role of chemotaxis in symbiosis, the phenotype of a *cheA* mutant was analysed in detail (**Figure 125**). This mutant has lost its motility in a soft agar plat assay (**Figure 125A**), compatible with a loss in chemotaxis required for this motility. In a single-strain infection assay, the *cheA* mutant has a strong delay of one to two days compared to the

wild-type strain in infecting the M4 crypts, but ultimately establishes normally in the crypts (**Figure 125B**). However, in mixed infection experiments (see **Chapter III**) in which the insect is co-infected with the *cheA* mutant and the wild-type strain, the mutant is strongly outcompeted by the wild-type (**Figure 125C-D**). Thus, these experiments confirm the conclusions from the Tn-seq analysis (see the recapitulative functions illustrated in the **Figure 129**).

4.3.6. Plasmid 2 genes

One of the most striking difference between the rich medium, M1 and M3 conditions and the M4 organs at the second and third instars conditions is the transposon insertion landscapes of the plasmid 2 (**Figure 104**). As I mentioned previously, these differences were identified as a transposon insertions enrichment in a specific region of the plasmid 2 in the M4 organ conditions (**Figure 126**), and as a drastic increase of bacterial fitness genes of the plasmid 2 at the third instar M4 condition (**Figure 105**).

Essentiality of plasmid 2 in the M4 organ at the third instar

At the third instar host developmental stage, a huge proportion of the plasmid 2 genes were annotated as “conditionally-essential” for the colonization of the M4 organ (183 genes out of 206 total genes in the plasmid 2) (**Figure 105**). According to the Con-ARTIST analysis, this result suggests that the whole plasmid 2 became important for the bacterial fitness in the symbiotic organ when the host molted to the third instar. Interestingly, it was recently demonstrated in the laboratory that the symbiotic bacteria in the M4 region lose their plasmid 2 gradually. In second instar insects, three days after infection, already 50% of the bacteria had lost the plasmid while in fifth instar insects, only 10% of the bacteria had maintained the plasmid (Ohbayashi *et al.*, 2019). Moreover, clones of *B. insecticola* isolated from the M4 which have lost the plasmid were fully capable to establish the symbiosis and showed even an enhanced fitness compared to the wild-type strain containing the plasmid in a co-infection experiment (unpublished data). If symbiotic bacteria lost their plasmid 2 during the third instar host stage, then the plasmid 2 transposon insertions would no longer be sequenced and the plasmid 2 genes wouldn’t be detected in the Tn-seq analysis. Thus, the reduction of transposon insertions in the plasmid 2 at the third instar is probably associated to a loss of this plasmid in symbiotic bacteria, rather than a requirement of this plasmid for the bacterial fitness.

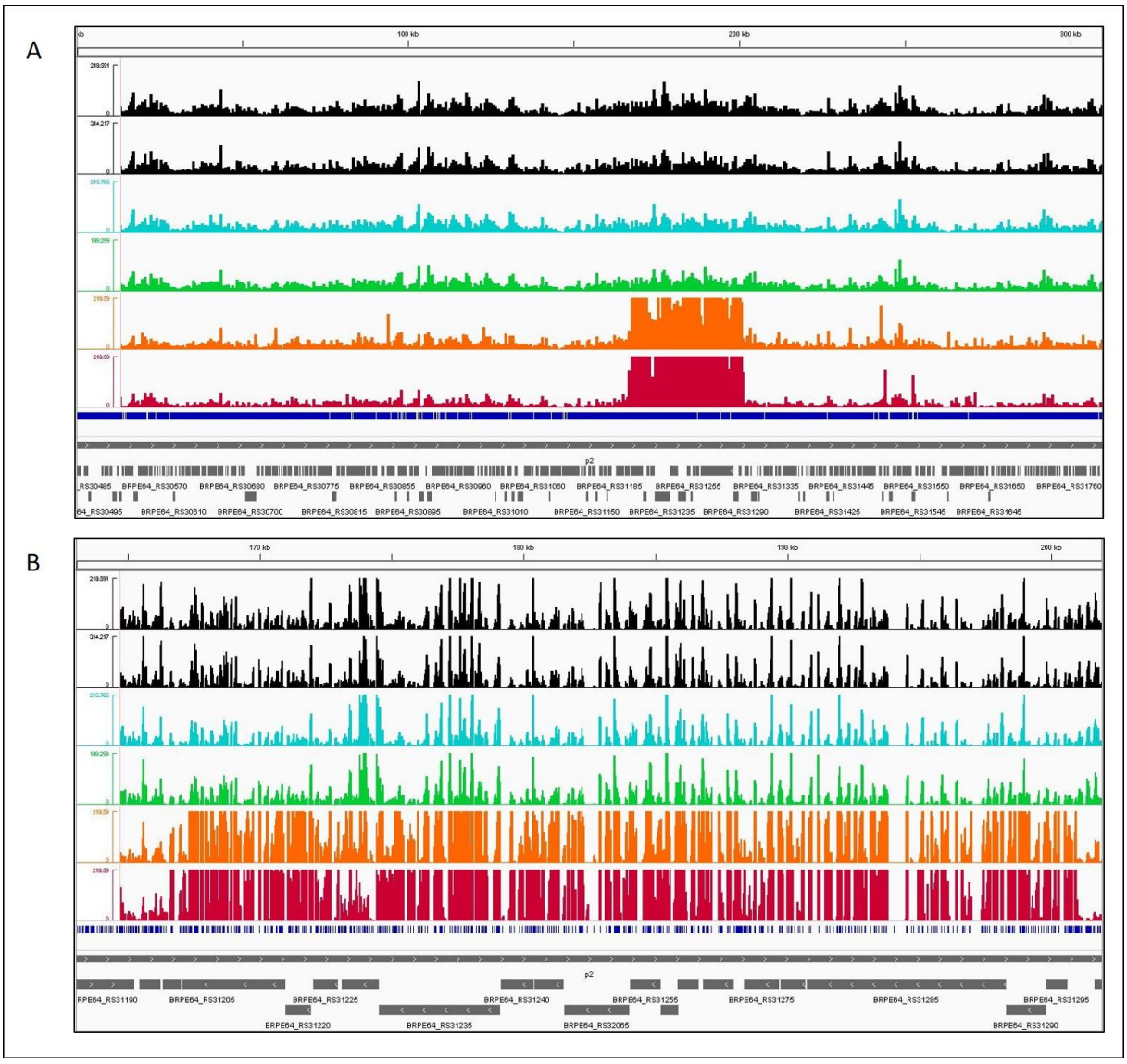


Figure 126: Localization of transposon-enriched genes in the plasmid 2 in the symbiotic organ.

A) Visualization of the whole plasmid 2 sequence. B) Visualization of the plasmid 2 region which contains the 19 enriched transposon genes, from *BRPE64_RS31200* to *BRPE64_RS31295*. The insertion distributions (log^{10} scale) are displayed for each condition: YG and MM with glucose (black bars), M1 (blue bars), M3 (green bars), M4 second instar (orange bars) and M4 third instar (red bars). The different positions on the plasmid 2 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey.

Plasmid 2 enriched genes

Another interesting and remarkable characteristic of the *in vivo* Tn-seq data is that a cluster of genes in the plasmid 2 was highly enriched in transposon insertions in the M4 samples compared to the other conditions (**Figures 104 and 126**). This cluster of genes corresponds to 19 genes (out of a total of 206 genes in the plasmid 2), from *BRPE64_RS31200* to

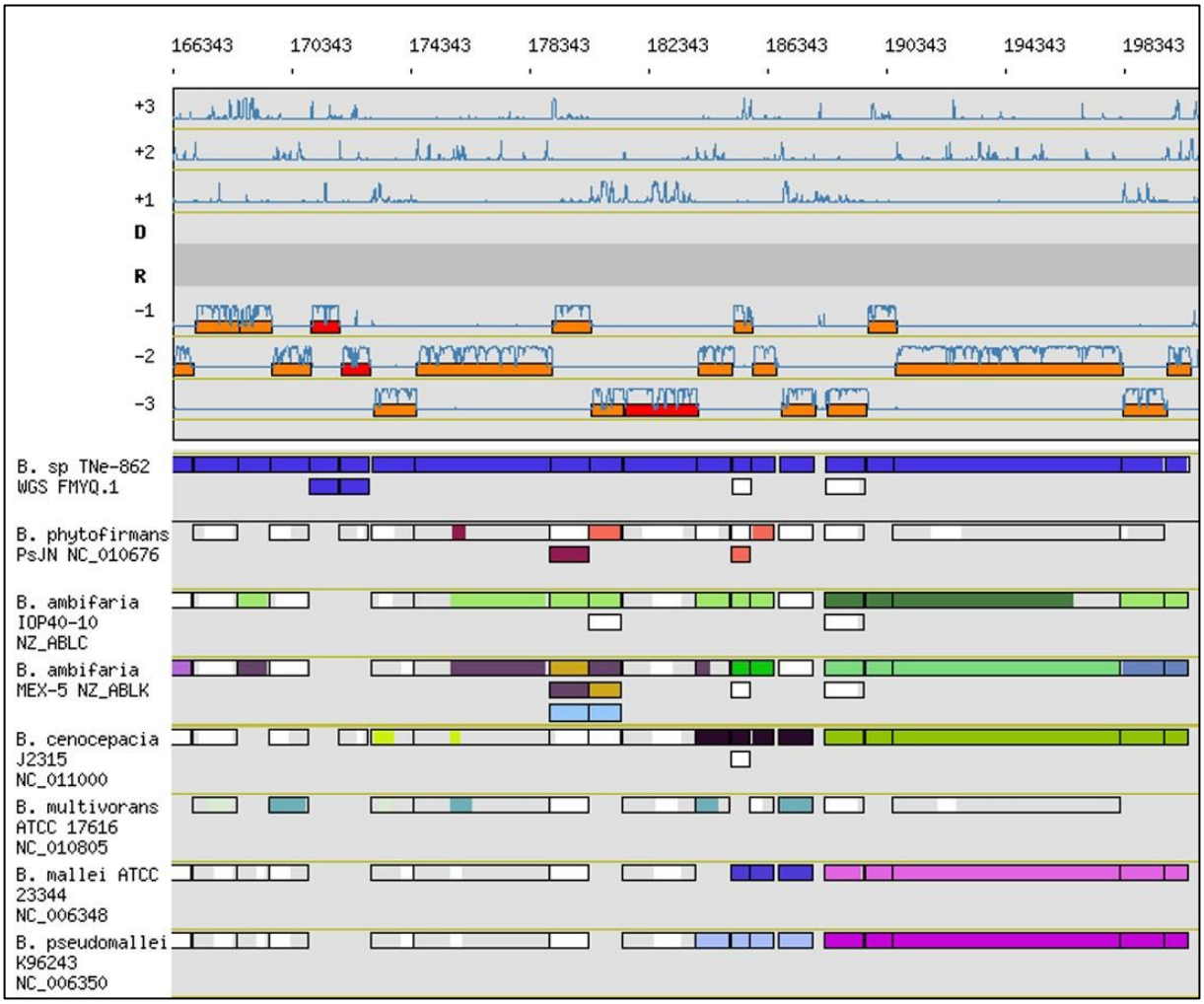


Figure 127: Conservation of this enriched gene cluster from the plasmid 2 of *B. insecticola* in other *Burkholderia* species.

Synteny maps with other *Burkholderia* species was computed with the Genome Browser tool of the MaGe website (MicroScope platform) using the PkGDB database. Positions in the plasmid 2 are indicated above the map in bp. Numbers indicated on the left correspond to the reading frame in the direct strand (D) and the reverse strand (R). The upper picture represents the genetic map of *B. insecticola* with NCBI annotated genes in red and MAGE annotated genes in orange. The blue lines indicate the coding prediction curves. The synteny maps for each *Burkholderia* species are displayed below with a specific color attributed for each species. The darker the color, the more the gene is conserved. White color indicates that there is no synteny conservation.

BRPE64_RS31295, with 18 conditionally enriched genes and 1 domain-conditionally enriched gene identified at the third instar (**Figure 126**). These 19 genes are encoding enzymes involved in the biosynthesis of a polysaccharide component and a lipid anchor (**Table 7**). Concerning the polysaccharide moiety, I found multiple genes involved in polysaccharide biosynthesis (*BRPE64_RS31205*, *BRPE64_RS31220*, *BRPE64_RS31225* and *BRPE64_RS31295*) (**Figure 126 and Table 7**) with a putative cellulose synthase (Römling and Galperin, 2015), in capsular

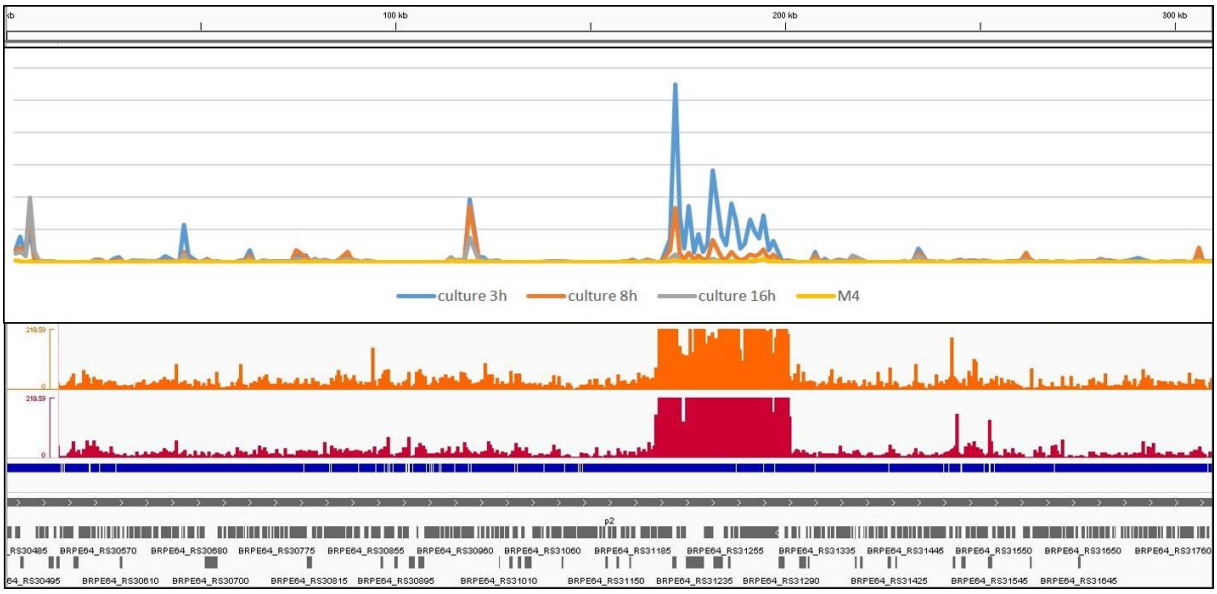


Figure 128: Correlation between the expression level and the transposon-enriched regions in the plasmid 2.

The above graph shows the expression level of plasmid 2 genes in lag-phase (3h), exponential phase (8h) and stationary phase (16h) cultures as well as in the symbiotic organ (M4) obtained from Ohbayashi *et al.,* 2019b. The transposon-insertion distributions (log^{10} scale) are displayed for the M4 second instar (orange bars) and the M4 third instar (red bars). The different positions on the plasmid 2 are indicated above the figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey.

polysaccharide export by an ABC transporter (*BRPE64_RS31255*, *BRPE64_RS31260*, *BRPE64_RS31265*, *BRPE64_RS31270*) (**Figure 126 and Table 7**) (Larue *et al.,* 2011; Nsahlai and Silver, 2003; Rosenow *et al.*, 1995), and in O-antigen and core oligosaccharide biosynthesis (*BRPE64_RS31235*, *BRPE64_RS31240*, *BRPE64_RS31245*) (**Figure 126 and Table 7**), such as the UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase and the UDP-N-acetyl glucosamine-2-epimerase (Burrows *et al.*, 2000; Pradel *et al.*, 1992). The other part of this gene cluster is responsible for the biosynthesis of a fatty acid component, with *BRPE64_RS31275*, *BRPE64_RS31280* and *BRPE64_RS31285* (**Figure 126 and Table 7**) that are encoding respectively an 8-amino-7-oxonanoate synthase (Manandhar and Cronan, 2018), an UDP-3-O-acyl N-acetylglucosamine deacetylase (Barb and Zhou, 2008), and a putative type I polyketide synthase (Trindade-Silva *et al.*, 2013). Interestingly, this putative type I polyketide synthase is encoded by the largest gene of this enriched cluster (approximately 8 kb) (**Figure 126**), and is homologous to the *wcbR* gene of *B. pseudomallei* that was reported to be present in the capsular polysaccharide I coding region (Cuccui *et al.*, 2012).

Gene tag	Start	End	Gene name	Gene product (Uniprot annotation)	Gene product (MAGE annotation)	BLAST *Burkholderia* species (species name and % of identity)	BLAST *E. coli* K-12 (% of identity)	Biological function
BRPE64_RS31200	166343	167047	-	Uncharacterized protein	Conserved protein of unknown function	Acetyltransferase (GNAT) family protein (*Burkholderia* sp. TNe-862, 78.95%)	-	Unknown function
BRPE64_RS31205	167106	168587	-	Glycosyl transferase family 2	Glycosyl transferase family protein	Cellulose synthase (UDP-forming) (*Burkholderia* sp. TNe-862, 84.05%)	*bcsA* (27.25%)	Exopolysaccharide biosynthesis
BRPE64_RS31210	168600	169685	-	Possible spore protein	Spore protein	Spore maturation protein CgeB (*Burkholderia* sp. TNe-862, 86.15%)	Hypothetical protein (34.48%)	Cell wall biosynthesis
BRPE64_RS31215	169691	170989	-	Glycosyl transferase group 1	Glycosyl transferase group 1	Glycosyltransferase involved in cell wall biosynthesis (*Burkholderia* sp. TNe-862, 84.22%)	-	Cell wall biosynthesis
BRPE64_RS31220	170991	171959	-	Uncharacterized protein	Conserved protein of unknown function	Conserved protein of unknown function (*Burkholderia* sp. TNe-862, 88.82%)	Colanic acid biosynthesis glycosyltransferase Wcal (41.67%)	Exopolysaccharide biosynthesis
BRPE64_RS31225	172034	172987	-	Uncharacterized protein	Conserved protein of unknown function	Conserved protein of unknown function (*Burkholderia* sp. TNe-862, 88,86%)	-	Exopolysaccharide biosynthesis
BRPE64_RS31235	174548	179116	-	-	Protein of unknown function	Spore maturation protein CgeB (*Burkholderia* sp. TNe-862, 81.76%)	-	Core oligosaccharide biosynthesis (RfaB domain)
BRPE64_RS31240	179142	180404	- (*rffD*)	UDP-N-acetyl-D-mannosamine dehydrogenase	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase (*Burkholderia* sp. TNe-862, 90.71%)	*wecC* (67.75%)	O-antigen biosynthesis / Enterobacterial common antigen biosynthesis
BRPE64_RS31245	180421	181542	- (*rffE*)	UDP-N-acetylglucosamine-2-epimerase	UDP-N-acetyl glucosamine-2-epimerase	UDP-N-actelglucosamine 2-epimerase (*Burkholderia* sp. TNe-862, 89.61%)	*wecB* (65.05%)	O-antigen biosynthesis / Enterobacterial common antigen biosynthesis
BRPE64_RS32065	181546	184029	-	Uncharacterized protein	Conserved protein of unknown function	Conserved protein of unknown function (*Burkholderia* sp. TNe-862, 67.47%)	-	Unknown function

Gene tag	Start	End	Gene name	Gene product (Uniprot annotation)	Gene product (MAGE annotation)	BLAST *Burkholderia* species (species name and % of identity)	BLAST *E. coli* K-12 (% of identity)	Biological function
BRPE64_RS31255	184049	185200	-	Chain length determinant protein	Chain length determinant protein	Capsular polysaccharide transport system permease protein (*Burkholderia* sp. TNe-862, 87.47%)	-	Capsular polysaccharide export (KpsE domain)
BRPE64_RS31260	185202	185858	- (*kpsT*)	ATPase component ABC-type polysaccharide/polyol phosphate transport system	Polysialic acid transport ATP-binding protein KpsT	Polysialic acid transport ATP-binding protein KpsT (*Burkholderia* sp. TNe-862, 88.99%)	*ccmA* (28.33%)	Capsular polysaccharide export (KpsT domain)
BRPE64_RS31265	185855	186658	-	Transport permease protein	ABC transporter	Transport permease protein (*Burkholderia* sp. TNe-862, 90.64%)	-	Capsular polysaccharide export
BRPE64_RS31270	186814	187971	-	Polysaccharide export protein	Polysaccharide export protein	Capsule polysaccharide export outer membrane protein CtrA (*Burkholderia* sp. TNe-862, 85.42%)	Polysaccharide export protein (30.34%)	Capsular polysaccharide export
BRPE64_RS31275	188371	189696	-	Glycine C-acetyltransferase	8-amino-7-oxononanoate synthase	8-amino-7-oxononanoate synthase (*Burkholderia* sp. TNe-862, 91.04%)	Glycine C-acetyltransferase (36.36%)	Fatty acids biosynthesis
BRPE64_RS31280	189735	190688	- (*lpxC*)	UDP-3-O-acyl N-acetylglycosamine deacetylase	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	UDP-3-O-acyl N-acetylglycosamine deacetylase (*Burkholderia* sp. TNe-862, 85.8%)	UDP-3-O-acyl N-acetylglycosamine deacetylase (32.32%)	Lipid A biosynthesis
BRPE64_RS31285	190739	198283	-	Putative type I polyketide synthase WcbR	KR domain protein	Acyl transferase domain-containing protein (*Burkholderia* sp. TNe-862, 84.34%)	Beta-ketoacyl-ACP synthase II (25.85%)	Capsular polysaccharide fatty acid biosynthesis
BRPE64_RS31290	198286	199815	-	Sulfatase	Sulfatase	Sulfatase (*Burkholderia* sp. TNe-862, 80.63%)	-	Host cell adhesion (mucin degradation)
BRPE64_RS31295	199817	200602	-	Short-chain dehydrogenase/reductase SDR	Oxidoreductase, short-chain dehydrogenase/reductase family protein	Short-chain dehydrogenase (Burkholderia sp. TNe-862, 89.61%)	SDR family oxidoreductase (28.87%)	Polysaccharide biosynthesis

Table 7: List of enriched transposon genes in the plasmid 2 of *B. insecticola*.
The 19 genes identified as enriched transposon genes in the symbiotic organ conditions are listed in this table. For each gene, numerous informations are displayed: the gene tag, the starting and ending positions (in bp), the gene product through the Uniprot and the MaGe annotations, the gene or the protein name of the BLAST result with *Burkholderia* species (with the species name and the percentage of identity indicated), the gene or the protein name of the BLAST result with *E. coli* K-12 (with the percentage of identity indicated), and the biological function attributed to this gene (based on literature search).

Thus, this gene cluster in the plasmid 2 is probably involved in the biosynthesis of a specific capsular polysaccharide (see the recapitulative functions illustrated in the **Figure 129**). This entire gene cluster is conserved in only one other sequenced *Burkholderia* species, namely *Burkholderia lycopersici* TNe-862, a diazotrophic strain isolated in Mexico from the rhizoplane of tomato plants (**Figure 127**) (Caballero-Mellado *et al.*, 2007). However, the last seven genes of this cluster were generally found in other pathogenic *Burkholderia* species, such as *B. ambifaria*, *B. cenocepacia*, *B. mallei* and *B. pseudomallei* (**Figure 127**).

The very large number of insertions in this particular gene cluster in the M4 bacteria, compared to the other conditions, suggests that it has a negative impact on the symbiosis and that the symbiont gains in fitness in the M4 region when the cluster is inactivated by mutation. As bacterial mutants of this specific region seem to have a better colonization capacity, there might be a strong selective pressure applied on the symbiotic population which then might explains why the entire plasmid 2 is lost during the colonization of the symbiotic organ. The gene expression profile of the plasmid 2 genes is further confirming that the capsular polysaccharide cluster is the main cause of the negative selection against the plasmid 2 in the symbiotic bacteria (Ohbayashi *et al.*, 2019). The genes of this cluster are by far the most strongly expressed genes of the plasmid 2, and all the other genes, except for a few others, are only very weakly or not expressed (**Figure 128**). Capsular polysaccharides form an extracellular structure known as capsule that is widely distributed among bacterial species, mostly including pathogens such as *E. coli* (Willis and Whitfield, 2013). For these pathogenic species, capsules are known as virulence factors that are required to evade the host immune system such as phagocytosis, complement-mediated killing and AMPs, and also to promote adherence to the host cells (Willis and Whitfield, 2013). For example, it was reported that the capsular polysaccharide of *B. pseudomallei* is required for the bacterial survival and persistence inside the host by limiting phagocytosis (Reckseidler-Zenteno *et al.*, 2005).

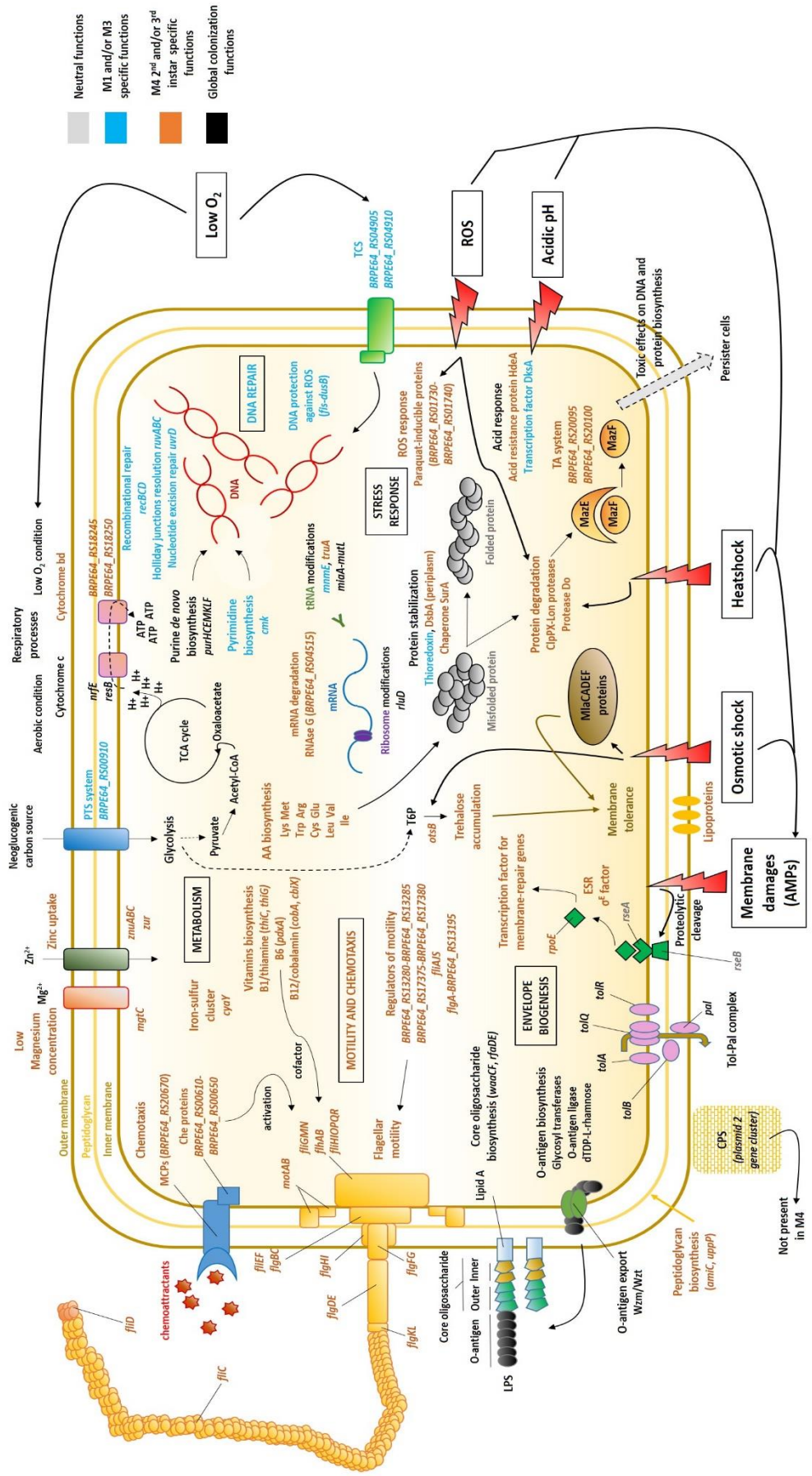

Neutral functions
M1 and/or M3 specific functions
M4 2nd and/or 3rd instar specific functions
Global colonization functions
Low O2
ROS
Acidic pH
TCS
BRPE64_RS04905
BRPE64_RS04910
Toxic effects on DNA and protein biosynthesis
Persister cells
DNA REPAIR
DNA protection against ROS (fis-dusB)
DNA
Holliday junctions resolution ruvABC
Nucleotide excision repair uvrD
Recombinational repair recBCD
STRESS RESPONSE
ROS response
Paraquat-inducible proteins (BRPE64_RS01730 BRPE64_RS01740)
Acid response
Acid resistance protein HdeA
Transcription factor DksA
TA system
BRPE64_RS20095
BRPE64_RS20100
MazE
MazF
Heatshock
Osmotic shock
Membrane damages (AMPs)
Low O2 condition
Respiratory processes
Cytochrome bd
BRPE64_RS18245
BRPE64_RS18250
Aerobic condition
Cytochrome c
nrfE
resB
Purine de novo biosynthesis purHCEMKLF
Pyrimidine biosynthesis cmk
tRNA modifications
mnmE, truA
miaA-mutL
Folded protein
Protein stabilization
Thioredoxin, DsbA (periplasm)
Chaperone SurA
Misfolded protein
Protein degradation
ClpPX-Lon proteases
Protease Do
MlaCADEF proteins
Neoglucogenic carbon source
PTS system
BRPE64_RS00910
TCA cycle
Oxaloacetate
Acetyl-CoA
Glycolysis
Pyruvate
mRNA degradation
RNAse G (BRPE64_RS04515)
mRNA
Ribosome modifications
rluD
AA biosynthesis
Lys Met
Trp Arg
Cys Glu
Leu Val
Ile
T6P
otsB
Trehalose accumulation
Membrane tolerance
Lipoproteins
Zinc uptake
Zn2+
zmuABC
zur
METABOLISM
Vitamins biosynthesis
B1/thiamine (thiC, thiG)
B6 (pdxA)
B12/cobalamin (cobA, cbiX)
Transcription factor for membrane-repair genes
ESR
σE factor
rpoE
rseA
Proteolytic cleavage
rseB
Low Magnesium concentration
Mg2+
mgtC
Iron-sulfur cluster cyaY
cofactor
MOTILITY AND CHEMOTAXIS
Regulators of motility
BRPE64_RS13280-BRPE64_RS13285
BRPE64_RS17375-BRPE64_RS17380
fliAJS
flgA-BRPE64_RS13195
ENVELOPE BIOGENESIS
tolR
tolQ
tolA
tolB
pal
Tol-Pal complex
Outer membrane
Peptidoglycan
Inner membrane
Chemotaxis
MCPs (BRPE64_RS20670)
Che proteins
BRPE64_RS00610-BRPE64_RS00650
activation
motAB
fliGMN
fliAB
fliHIOPQR
Flagellar motility
Core oligosaccharide biosynthesis (waaCF, rfaDE)
Lipid A
O-antigen biosynthesis
Glycosyl transferases
O-antigen ligase
dTDP-L-rhamnose
CPS (plasmid 2 gene cluster)
Not present in M4
Peptidoglycan biosynthesis (amiC, uppP)
chemoattractants
fliD
fliC
fliEF
flgBC
flgHI
flgDE
flgKL
flgFG
Core oligosaccharide
O-antigen
Outer
Inner
LPS
O-antigen export
Wzm/Wzt

Figure 129: Overview of symbiotic factors of *B. insecticola* required for the colonization of *R. pedestris*.
Each gene name and function is coloured according to its in vivo requirement. Genes and functions required for the M1 and/or the M3 colonization are indicated in blue. Genes and functions required for the M4 at the second and/or the third instars are indicated in orange. Genes and functions required for all *in vivo* conditions are indicated in black. Neutral genes and functions are indicated in grey. Abbreviations: AA: amino acid, CPS: capsular polysaccharide, T6P: trehalose-6-phosphate, TA: toxin-antitoxin system, TCS: two-component system.

Here for the *Burkholderia* symbiont, this capsular polysaccharide might be recognized by the insect immune system as a pathogen-associated molecular pattern (PAMP), hence triggering an immune response. Hence, the absence of this capsular polysaccharide would be beneficial for the symbiotic population in order to be recognized as a "pacific" partner and not an invader. Additionally, as capsular polysaccharides are also involved in the adherence to the host cells and biofilm production, their absence might increase the exchange surface available to interact with the host as an extracellular symbiont. As mentioned before, capsules are also able to provide resistance towards AMPs, and as numerous CCR peptides are secreted in the symbiotic organ as symbiotic AMPs, the absence of this capsular polysaccharide might facilitate the activity of these CCR peptides. Interestingly, it was reported that the artificial production of a specific exopolysaccharide in *Salmonella* named PGA (poly-β-1,6-N-acetyl-D-glucosamine), reduced the intracellular survival inside macrophages and also increased the sensitivity towards bile salts and oxidative stress (Echeverz *et al.*, 2017). In contrast to the other enterobacterial species, it was showed that *Salmonella* has lost genes responsible for the production of this PGA exopolysaccharide, which was probably due to the negative impact of PGA on the virulence of *Salmonella* (Echeverz *et al.*, 2017). Hence, PGA exopolysaccharide was characterized as an antivirulence factor for *Salmonella* and its production was lost during the evolution (Echeverz *et al.*, 2017). Here, similar to the PGA loss in *Salmonella*, the loss of the capsular polysaccharide biosynthesis in the symbiont inside the symbiotic organ could be interpreted as a benefit for the bacterial fitness *in vivo*. Given the negative role of the capsular polysaccharide gene cluster and the loss of the plasmid 2 in the symbiotic bacteria, the question arises: why this bacterium maintains these genes? It is even very surprising that the clone that was isolated from a wild captured adult insect and that was chosen as a model strain, has the plasmid 2 at all since in fifth instar nymphs raised in the laboratory, only 10% of the bacteria still have the plasmid (Ohbayashi *et al.*, 2019b). Several testable hypotheses

can be proposed. Perhaps, in insects growing in natural conditions, the gene cluster has a positive rather than a negative fitness effect on the symbiotic bacteria. It is also possible that, even if the gene cluster has a negative impact on the bacterial fitness in the insect gut, it has a postive fitness impact in the other lifestyles of the *B. insecticola* bacterium, in the soil or in the rhizosphere of plants (as suggested by the conservation of this gene cluster in a tomato rhizosphere bacterium (Caballero-Mellado *et al.*, 2007)). Moreover, I can not exclude that the capsular polysaccharide has a postive role in some early stages of the interaction with the insect, for example during the passage in the M1, M2 and M3 midgut regions, or the constricted region, or for the very early stages of the M4 crypt colonization. It will certainly be exciting to figure out in the future what is the role of this plasmid 2 and its intriguing capsular polysaccharide gene cluster.

4.4. Correlation between symbiosis factors and AMP resistance

In the previous chapter, I have identified bacterial genes involved in AMP resistance towards five different AMPs: polymyxin B, LL-37, riptocin, and two CCR peptides (CCR179 and CCR480) (see **Chapter III**). As we hypothesized that AMP resistance may play a role in the specific colonization of the host by the *Burkholderia* symbiont, I looked for common bacterial fitness genes between the *in vivo* and the *in vitro* AMP conditions. Concerning the five AMPs Tn-seq datasets, I found 95 bacterial genes that were required for AMP resistance with a Con-ARTIST essentiality score of 1 or 2 (domain-conditionally essential genes and conditionally essential genes, respectively) for at least one AMP. As the Tn-seq experiment with AMPs was performed in MM (minimal medium) supplemented with glucose, I performed another Con-ARTIST analysis for the *in vivo* Tn-seq data by comparing them with the MM Tn-seq data. This analysis allowed me to attribute correct Con-ARTIST essentiality scores for the comparison between the Tn-seq datasets. When I checked the attributed Con-ARTIST scores for the four *in vivo* conditions (M1, M3, M4 second instar, M4 third instar), I noticed that multiple genes shared the same essentiality scores between the *in vivo* conditions and the *in vitro* AMP conditions (**Figure 130**). By applying a clustering analysis, I have found that the M4 conditions essentiality

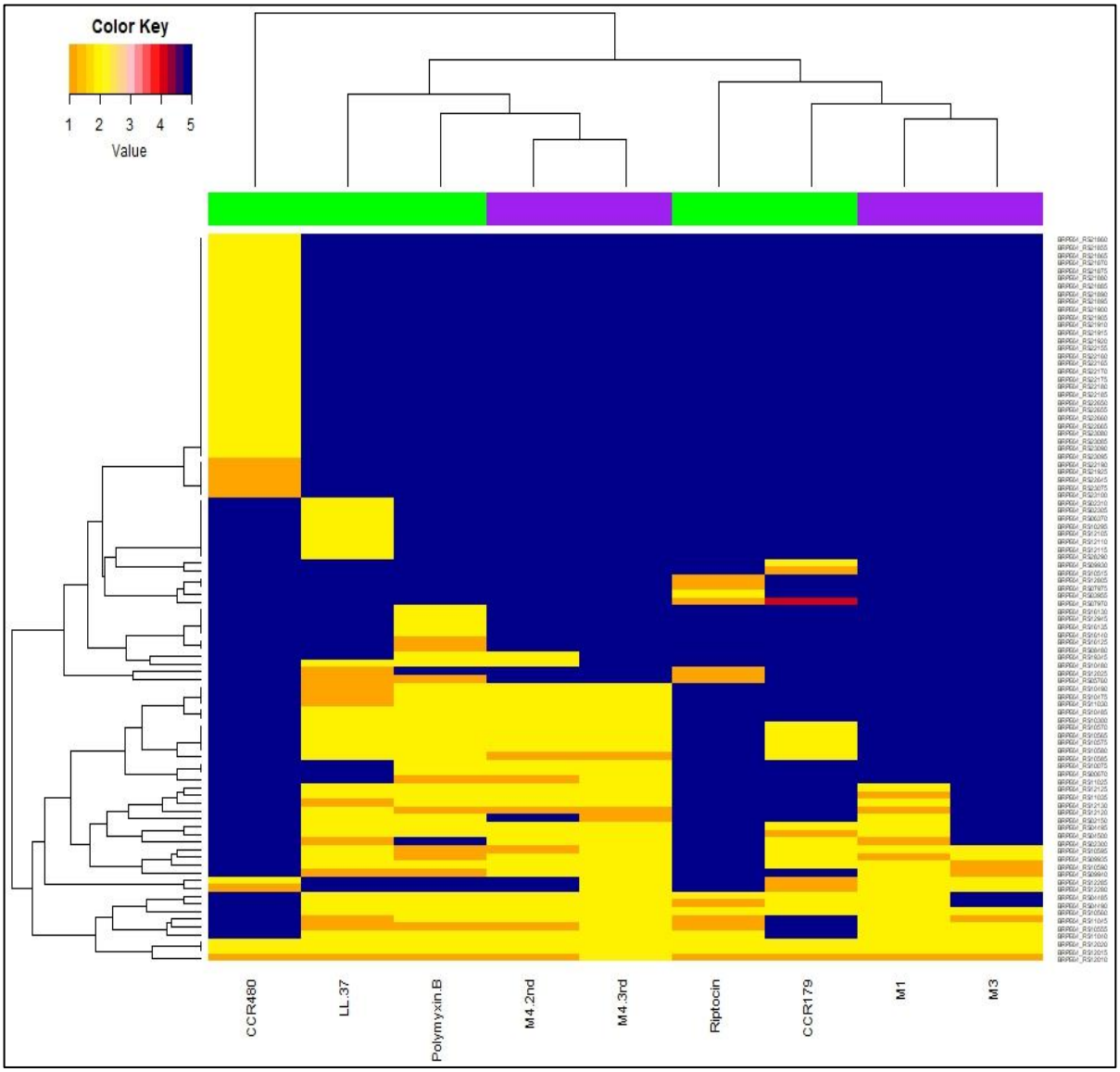


Figure 130: Correlations between *in vivo* fitness genes and AMP resistance genes. This heatmap is representing the Con-ARTIST essentiality scores of previously identified fitness genes involved in AMP resistance for the five AMPs studied (Polymyxin B, LL-37, Riptocin, CCR179 and CCR480 peptides) highlighted in green and for the four *in vivo* conditions (M1, M3, M4 2nd instar and M4 3rd instar) highlighted in purple. In total, there were 95 genes that presented a Con-ARTIST essentiality score of 1 (domain-conditionally essential genes) or 2 (conditionally essential genes) in at least one AMP. These 95 genes were used to generate the heatmap. The color key is indicating the Con-ARTIST essentiality score from 1 to 5: 1 represents domain-conditionally essential genes (orange), 2 represents conditionally essential genes (yellow), 3 represents domain-enriched genes (pink), 4 represents enriched genes (red) and 5 represents neutral genes (blue). The clustering analysis was performed for each condition and organized the 95 genes from the best shared essential genes to the less shared essential genes between all conditions (from the bottom part to the upper part).

profiles were relatively similar to the profiles of polymyxin B and LL-37 conditions (**Figure 130**).

In addition, I observed that there were similarities between the essentiality scores of both the

Heatmap order	Gene tag	Gene name	Gene product	BLAST *Burkholderia* species (species name and % of identity)	BLAST *E. coli* K-12 (% of identity)	Biological function
1	BRPE64_RS12010	*tatC*	Sec-independent protein translocase TatC subunit	*tatC* (*Burkholderia pseudomallei*, 82.63%)	*tatC* (51%)	TatABC transporter system
2	BRPE64_RS12015	*tatB*	Sec-independent protein translocase protein TatB	*tatB* (*Burkholderia pseudomallei*, 64.37%)	*tatB* (27.14%)	TatABC transporter system
3	BRPE64_RS12020	*tatA*	Sec-independent protein translocase protein TatA	*tatA* (*Burkholderia pseudomallei*, 68.66%)	-	TatABC transporter system
4	BRPE64_RS11040	*tolB*	Protein TolB	*tolB* (*Paraburkholderia phytofirmans*, 91.33%)	*tolB* (38.03%)	Cell wall Tol-Pal complex
5	BRPE64_RS10555	-	Hypothetical protein	*wzt* (*Bukholderia cenocepacia*, 42.5%)	Hypothetical protein (29.32%)	O-antigen export (Wzm/Wzt)
6	BRPE64_RS11045	-	Peptidoglycan-associated lipoprotein	*pal* (*Burkholderia multivorans*, 84.83%)	*pal* (46.6%)	Tol-Pal complex
7	BRPE64_RS10560	-	ABC-2 type transporter	*wzm* (*Burkholderia cenocepacia*, 27,03%)	*tagG* (44.02%)	O-antigen export (Wzm/Wzt)
8	BRPE64_RS04490	-	Glycosyl transferase family 2	Glycosyl transferase family 2 (*Burkholderia multivorans*, 78.63%)	Glycosyl transferase family 2 (40%)	LPS biosynthesis
9	BRPE64_RS04485	-	Glycosyl transferase group 1	*rfaB* (*Burkholderia cenocepacia*, 65.34%)	*rfaG* (27,78%)	LPS Core oligosaccharide biosynthesis
10	BRPE64_RS12280	*nrfE*	Cytochrome c assembly protein	*nrfE* (*Burkholderia pseudomallei*, 83%)	-	Cytochrome c biogenesis
11	BRPE64_RS12285	-	ResB family protein	ResB family protein (*Burkholderia pseudomallei*, 71.35%)	-	Cytochrome c biogenesis
12	BRPE64_RS09940	-	RfaE bifunctional protein	*rfaE* (*Paraburkholderia fungorum*, 82.7%)	*hldE* (51.8%)	LPS Core oligosaccharide biosynthesis
13	BRPE64_RS10590	*rfbA*	Glucose-1-phosphate thymidylyltransferase	*rfbA* (*Burkholderia multivorans*, 89.23%)	*rmlA* (73.96%)	LPS O-antigen biosynthesis
14	BRPE64_RS09935	-	ADP-L-glycero-D-manno-heptose-6-epimerase	*rfaD* (*Paraburkholderia xenovorans*, 89.09%)	*rfaD* (57.98%)	LPS Core oligosaccharide biosynthesis
15	BRPE64_RS10595	*rfbB*	dTDP-glucose 4,6-dehydratase	*rfbB* (*Burkholderia multivorans*, 89.8%)	*rmlB* (62.12%)	LPS O-antigen biosynthesis
16	BRPE64_RS02300	-	Lipopolysaccharide heptosyltransferase II	*waaF* (*Paraburkholderia xenovorans*, 81.64%)	*waaF* (41.58%)	LPS Core oligosaccharide biosynthesis
17	BRPE64_RS04500	-	O-antigen polymerase	*rfaL* (*Caballeronia pedi*, 86.82%)	*yjjB* (44.83%)	LPS biosynthesis
18	BRPE64_RS04495	-	Glycosyl transferase family 2	*epsE* (*Caballeronia pedi*, 84.76%)	Glycosyl transferase family 2 (33.64%)	LPS biosynthesis

Heatmap order	Gene tag	Gene name	Gene product	BLAST *Burkholderia* species (species name and % of identity)	BLAST *E. coli* K-12 (% of identity)	Biological function
19	BRPE64_RS02150	*dedA*	Membrane-associated protein	*dedA* (*Burkholderia cepacia*, 83.56%)	*dedA* (52.68%)	Cell wall biosynthesis
20	BRPE64_RS12120	-	Putative signal peptide protein toluene tolerance Ttg2C-like protein	Putative signal peptide protein toluene tolerance Ttg2C-like protein (*Burkholderia pseudomallei*, 78.03%)	*yrbD* (50.31%)	Membrane tolerance
21	BRPE64_RS12130	-	ABC transporter related	ABC transporter related (*Burkholderia pseudomallei*, 85.77%)	*yrbF* (52.87%)	Membrane tolerance
22	BRPE64_RS11035	-	Protein TolA	Protein TolA (*Paraburkholderia phytofirmans*, 65.83%)	-	Tol-Pal complex
23	BRPE64_RS12125	-	Hypothetical protein	Hypothetical protein (*Burkholderia pseudomallei*, 89.8%)	*yrbE* (57.31%)	Membrane tolerance
24	BRPE64_RS11025	*tolQ*	Protein TolQ	*tolQ* (*Burkholderia multivorans*, 92.41%)	*tolQ* (48.4%)	Tol-Pal complex
25	BRPE64_RS00670	*dsbA*	Thiol disulfide interchange protein DsbA	*dsbA* (*Burkholderia cenocepacia*, 73.49%)	*dsbA* (32.26%)	Protein stabilization
26	BRPE64_RS10075	-	Hypothetical protein	Hypothetical protein (*Burkholderia pseudomallei*, 59.5%)	-	Unknown function (TPR repeat protein)
27	BRPE64_RS10585	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	*rfbC* (*Burkholderia multivorans*, 92.35%)	*rmlC* (56.8%)	LPS O-antigen biosynthesis
28	BRPE64_RS10580	-	dTDP-4-dehydrorhamnose reductase	*rfbD* (Burkholderia multivorans, 76.17%)	*rmlD* (46.85%)	LPS O-antigen biosynthesis
29	BRPE64_RS10575	-	Rhamnosyltransferase	Rhamnosyltransferase (*Burkholderia multivorans*, 56.94%)	-	LPS O-antigen biosynthesis
30	BRPE64_RS10565	-	Glycosyl transferase family 2	*wbxF* (*Burkholderia cenocepacia*, 32.71%)	-	LPS Core olgosaccharide biosynthesis
31	BRPE64_RS10570	-	Hypothetical protein	Hypothetical protein (*Burkholderia cenocepacia*, 39.3%)	-	Cell wall biosynthesis (Glycosyl transferase domain)
32	BRPE64_RS10300	*waaC*	Lipopolysaccharide heptosyltransferase I	*waaC* (*Burkholderia cenocepacia*, 83.69%)	*waaC* (39.48%)	LPS Core oligosaccharide biosynthesis
33	BRPE64_RS10485	-	NAD-dependent epimerase/dehydratase	*wbiG* (*Burkholderia cenocepacia*, 53.5%)	*ybjS* (26.09%)	LPS Core oligosaccharide biosynthesis
34	BRPE64_RS11030	-	TolR protein	*tolR* (*Burkholderia multivorans*, 78.1%)	*exbD* (37.5%)	Tol-Pal complex
35	BRPE64_RS10475	-	Polysaccharide biosynthesis protein CapD	*wbiI* (*Burkholderia cenocepacia*, 80.83%)	-	LPS Core oligosaccharide biosynthesis
36	BRPE64_RS10490	-	Putative glycosyl transferase	*wbiF* (*Burkholderia pseudomallei*, 62.31%)	-	LPS Core oligosaccharide biosynthesis

Table 8: List of fitness genes involved in host colonization and AMP resistance.
The 36 genes identified in the heatmap as common fitness genes in both *in vivo* colonization and AMP resistance are listed in this table. For each gene, numerous informations are displayed: the heatmap order of genes (ranked from the most shared essential to the less shared essential gene), the gene tag, the gene name, the gene product, the gene or the protein name of the BLAST result with *Burkholderia* species (with the species name and the percentage of identity indicated), the gene or the protein name of the BLAST result with *E. coli* K-12 (with the percentage of identity indicated), and the biological function attributed to this gene (based on literature search).

M1 and M3 in vivo conditions and the essentiality scores of riptocin and the CCR179 peptide (**Figure 130**). However, the CCR480 peptide essentiality profile was completely different from

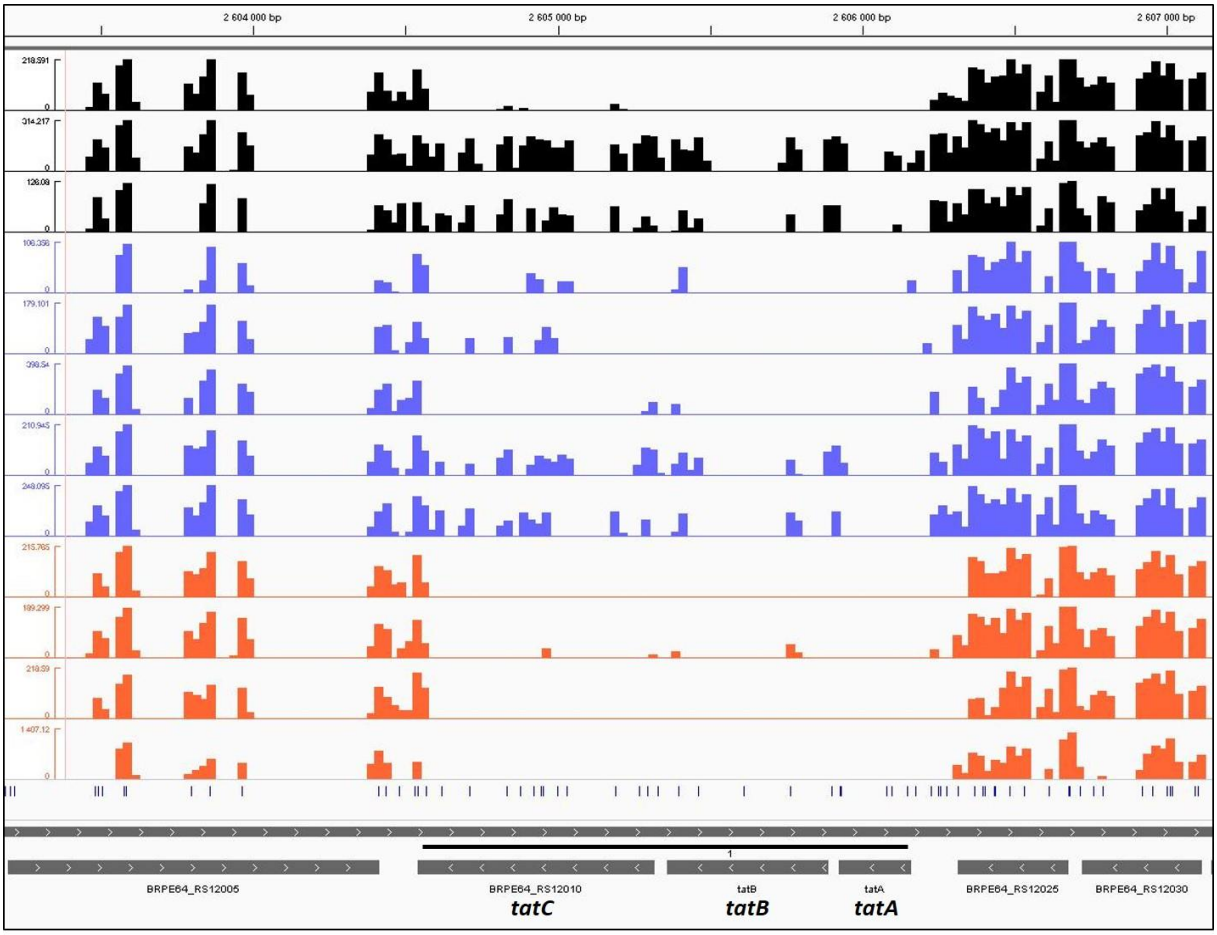


Figure 131: Identification of the Tat system involved in host colonization and AMP resistance.
The insertion distributions (log^{10} scale) are displayed for each condition, from the upper to the bottom layer: YG, MM with glucose and MM with succinate (black bars), polymyxin B, LL-37, riptocin, CCR179 peptide and CCR480 peptide (blue bars), M1, M3, M4 second instar and M4 third instar (orange bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The genes outlined with a black line are *tatC* (*BRPE64_RS12010*), *tatB* (*BRPE64_RS12015*) and *tatA* (*BRPE64_RS12020*).

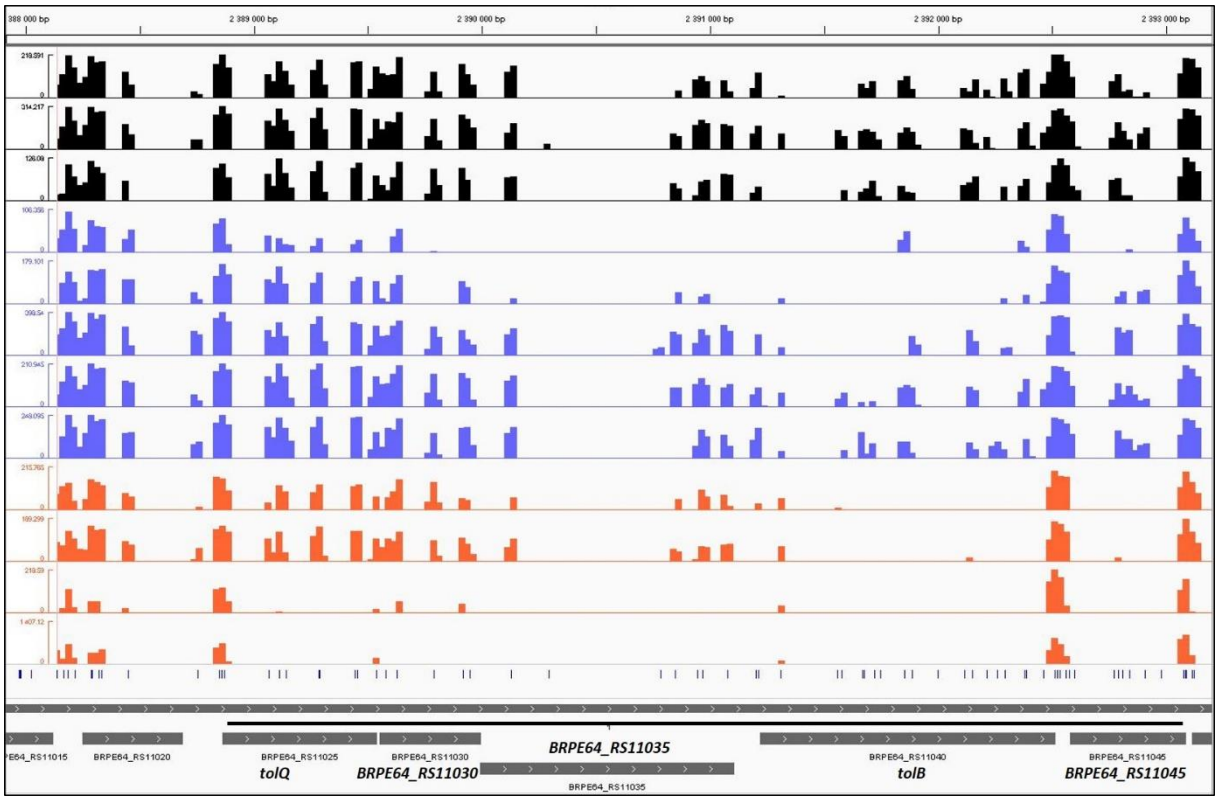


Figure 132: Identification of the Tol-Pal complex involved in host colonization and AMP resistance.

The insertion distributions (log^{10} scale) are displayed for each condition, from the upper to the bottom layer: YG, MM with glucose and MM with succinate (black bars), polymyxin B, LL-37, riptocin, CCR179 peptide and CCR480 peptide (blue bars), M1, M3, M4 second instar and M4 third instar (orange bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The genes outlined with a black line are *tolQ* (*BRPE64_RS11025*), *BRPE64_RS11030*, *BRPE64_RS11035*, *tolB* (*BRPE64_RS11040*) and *BRPE64_RS11045*.

the other conditions, excepting five genes that were also identified as *in vivo* fitness genes (**Figure 130**). Among these 95 genes studied, I identified 36 fitness genes required for both *in vivo* colonization and AMP resistance towards at least one AMP studied (**Figure 130 and Table 8**). As the AMP resistance factors identified in the previous chapter were mostly involved in cell wall biogenesis (see **Chapter III**), it was not surprising to find that these 36 genes are mostly encoding for cell wall functions (**Table 8**). In addition, these 36 genes are only located in the chromosome 1.

Out of these 36 genes, only three genes were required for both *in vivo* colonization and for the resistance towards the five AMPs studied (**Figure 130 and Table 8**). These three genes are *tatA*, *tatB* and *tatC*, of the above-discussed twin-arginine translocation (Tat) system (**Figure 131 and Table 8**), which is responsible for the transmembrane export of folded proteins

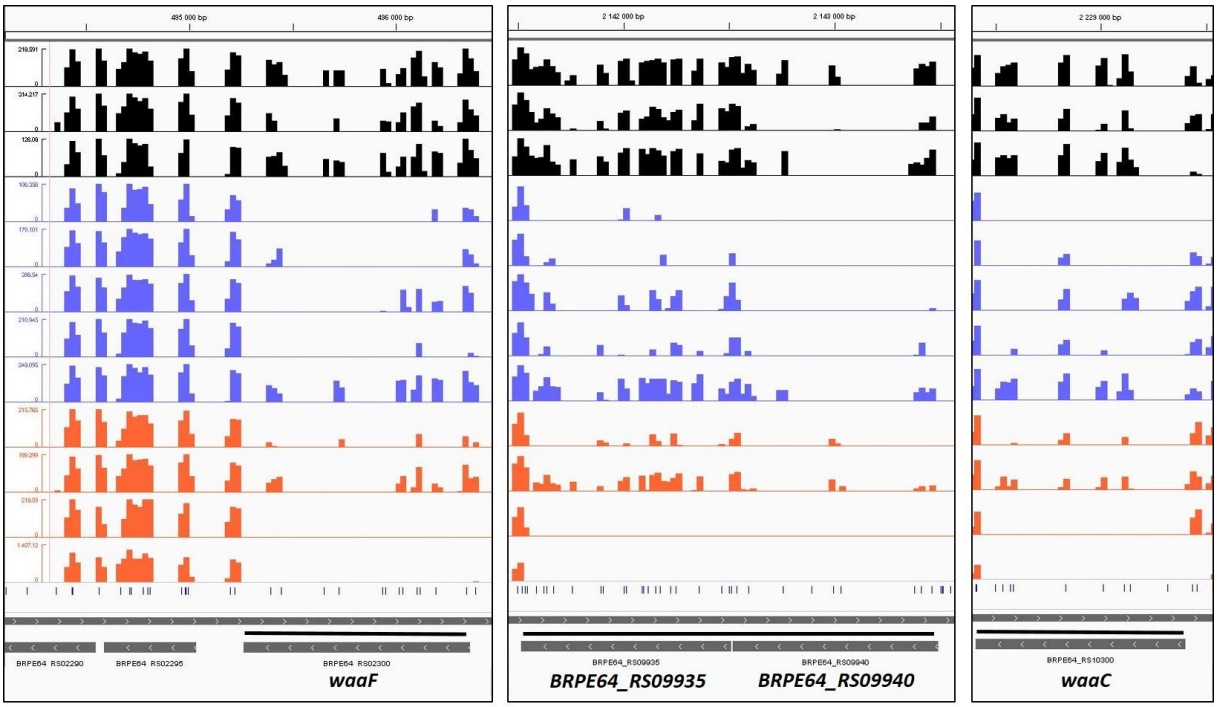


Figure 133: Identification of core oligosaccharide biosynthesis genes involved in host colonization and AMP resistance.

The insertion distributions (log^{10} scale) are displayed for each condition, from the upper to the bottom layer: YG, MM with glucose and MM with succinate (black bars), polymyxin B, LL-37, riptocin, CCR179 peptide and CCR480 peptide (blue bars), M1, M3, M4 second instar and M4 third instar (orange bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The genes outlined with a black line are *waaF* (*BRPE64_RS02300*), *BRPE64_RS09935*, *BRPE64_RS09940* and *waaC* (*BRPE64_RS10300*).

(Robinson *et al.*, 2011a). In addition, I also identified the genes encoding for the whole Tol-Pal complex, with *tolQ* (*BRPE64_RS11025*), *tolR* (*BRPE64_RS11030*), *tolA* (*BRPE64_RS11035*), *tolB* (*BRPE64_RS11040*) and *pal* (*BRPE64_RS11045*) (**Figure 132 and Table 8**). The other fitness genes involved in both *in vivo* colonization and AMP resistance are involved in the biosynthesis of the core oligosaccharide (**Figure 133**) and the O-antigen (**Figure 134**) parts of LPS molecules (**Table 8**). Concerning the core oligosaccharide biosynthesis, the identified genes were *waaC* (*BRPE64_RS10300*), *waaF* (*BRPE64_RS02300*), *BRPE64_RS09935*, *BRPE64_RS09940*, *BRPE64_RS10475* and *BRPE64_RS10490* (**Figure 133 and Table 8**). For the O-antigen component, the identified genes were encoding for the dTDP-L-rhamnose biosynthesis (*BRPE64_RS10575*, *BRPE64_RS10580*, *rfbC*, *rfbA*, *rfbB*) (**Figure 134**), different glycosyl transferases (*BRPE64_RS04485*, *BRPE64_RS04490*, *BRPE64_RS04495*, *BRPE64_RS10565*, *BRPE64_RS10570*) (**Figure 134**), the O-antigen polymerase (*BRPE64_RS04500*) and the O-

antigen export system Wzm/Wzt (*BRPE64_RS10555*, *BRPE64_RS10560*) (**Figure 134**). Thus, the external parts of LPS molecules seem to be required for both AMP resistance and host colonization. Other genes which encode cell wall functions, such as the MlaCADEF phospholipid transporter (*BRPE64_RS12120*, *BRPE64_RS12125*, *BRPE64_RS12130*) (**Figure 135**) and the DedA protein (*dedA*) (**Figure 136**) were also identified as common bacterial factors for AMP resistance and *in vivo* colonization (**Table 8**). Quite similar to the Tol-Pal complex, the MlaCADEF transporter and the DedA family proteins are known to participate to the bacterial membrane integrity (see **section 4.3.4**) (Doerrler *et al.*, 2013; Segura *et al.*, 2012).

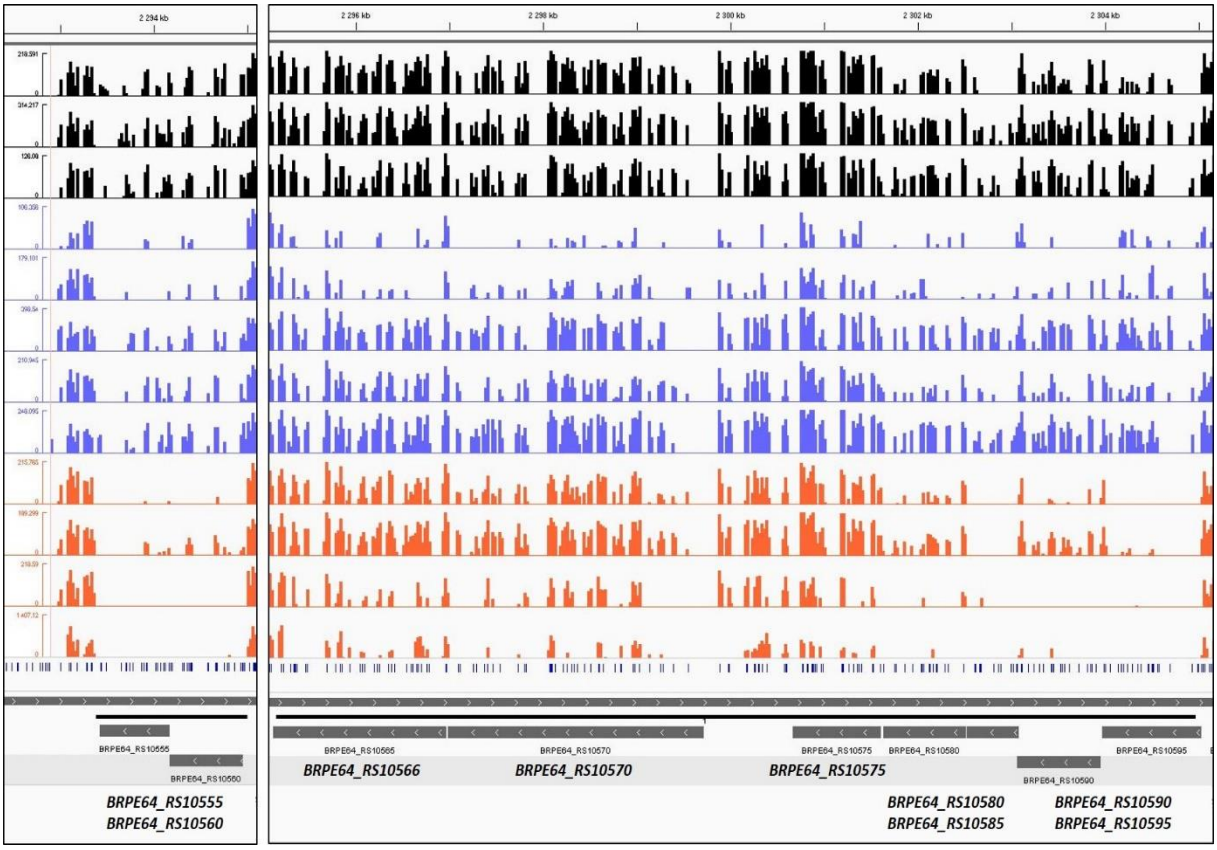


Figure 134: Identification of O-antigen biosynthesis and export genes involved in host colonization and AMP resistance.

The insertion distributions (log^{10} scale) are displayed for each condition, from the upper to the bottom layer: YG, MM with glucose and MM with succinate (black bars), polymyxin B, LL-37, riptocin, CCR179 peptide and CCR480 peptide (blue bars), M1, M3, M4 second instar and M4 third instar (orange bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The genes outlined with a black line are *BRPE64_RS10555* and *BRPE64_RS10560* involved in O-antigen export, *BRPE64_RS10565* and *BRPE64_RS10570* encoding for glycosyl transferases, *BRPE64_RS10575*, *BRPE64_RS10580*, *rfbC* (*BRPE64_RS10585*), *rfbA* (*BRPE64_RS10590*) and *rfbB* (*BRPE64_RS10595*) involved in dTDP-L-rhamnose biosynthesis.

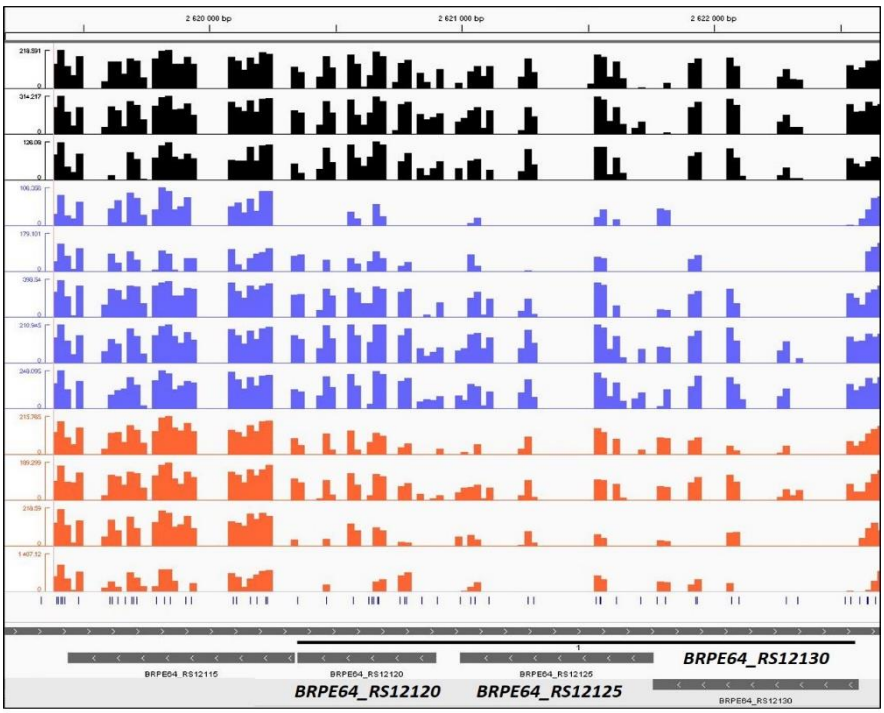


Figure 135: Identification of Mla proteins involved in host colonization and AMP resistance.

The insertion distributions (log^{10} scale) are displayed for each condition, from the upper to the bottom layer: YG, MM with glucose and MM with succinate (black bars), polymyxin B, LL-37, riptocin, CCR179 peptide and CCR480 peptide (blue bars), M1, M3, M4 second instar and M4 third instar (orange bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The genes outlined with a black line are *BRPE64_RS12120*, *BRPE64_RS12125* and *BRPE64_RS12130*.

Among the 36 fitness genes shared between the AMP and *in vivo* conditions, only four genes were not encoding for membrane-related components (**Table 8**). Among these four genes, I identified the *nrfE* gene (*BRPE64_RS12280*) and the *BRPE64_RS12285* gene (**Table 8**) that are playing a role in cytochrome c biogenesis. These two cytochrome c-type assembly proteins participate in the respiration machinery to produce energy for the bacterial metabolic processes (Ahuja *et al.*, 2009; Le Brun *et al.*, 2000). The last two genes, *dsbA* (*BRPE64_RS00670*) and *BRPE64_RS10075* (**Figure 136**) are encoding for the thiol-disulphide interchange protein DsbA and a hypothetical protein, respectively (**Table 8**). As mentioned before, the DsbA protein is involved in protein quality control by promoting the formation of disulphide bonds to stabilize periplasmic proteins (Manta *et al.*, 2019). Regarding the hypothetical protein encoded by the *BRPE64_RS10075* gene (**Figure 136**), this protein contains a tetratricopeptide repeat (TPR) motif (**Table 8**). TPR-containing proteins were reported to be

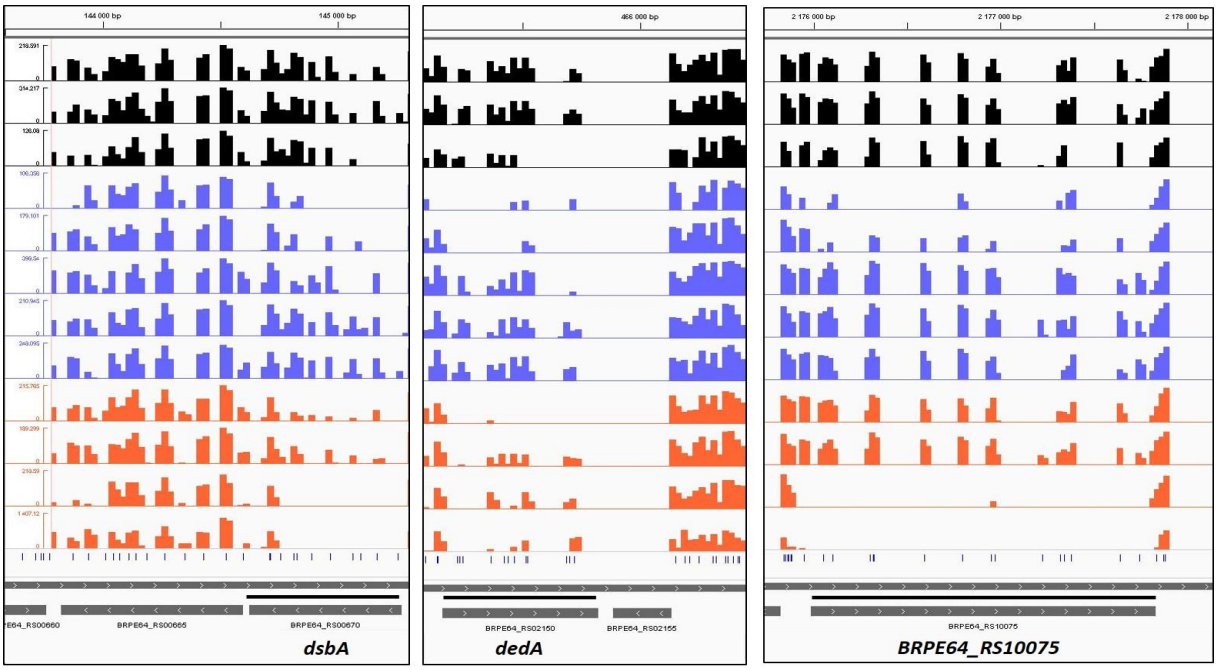


Figure 136: Identification of three genes *dsbA*, *dedA* and *BRPE64_RS10075* involved in host colonization and AMP resistance.

The insertion distributions (log^{10} scale) are displayed for each condition, from the upper to the bottom layer: YG, MM with glucose and MM with succinate (black bars), polymyxin B, LL-37, riptocin, CCR179 peptide and CCR480 peptide (blue bars), M1, M3, M4 second instar and M4 third instar (orange bars). The different positions on the chromosome 1 are indicated above each figure in bp. TA sites are indicated by blue bars under the insertion distributions. Genes are indicated in grey. The genes outlined with a black line are *dsbA* (*BRPE64_RS00670*), *dedA* (*BRPE64_RS02150*) and *BRPE64_RS10075*.

involved in a diverse array of cellular functions such as protein-protein interactions, protein folding, protein transport, chaperone, cell cycle control and transcriptional regulation (Blatch and Lässle, 1999; D'Andrea and Regan, 2003).

Thus, by comparing the Tn-seq data from the different AMP conditions and *in vivo* conditions, I found that 36 fitness genes were required for AMPs resistance and host colonization. These common bacterial factors are encoding for cell wall functions which suggest an important role of the bacterial membranes to resist AMPs and to promote resistance towards various stress conditions during *in vivo* colonization. Compared to the total number of *in vivo* fitness genes required for the colonization of the symbiotic organ (129 genes at the second instar), these 36 genes represent 28% of the total symbiotic functions identified. This proportion of symbiotic factors suggests that the symbiont is effectively facing membrane stress factors such as AMPs during the colonization of its host.

5. Discussion

The symbiotic mechanisms by which beneficial bacteria establish inside their host remain poorly understood. In the *R. pedestris-Burkholderia* symbiosis, even if some bacterial genes were previously identified as symbiotic factors (Jang *et al*., 2017; Kim *et al.*, 2013a, 2013b, 2014, 2016, 2017; Lee *et al.*, 2015; Ohbayashi *et al.*, 2015), these individual bacterial factors do not provide a complete overview of the symbiotic processes required for the host colonization. With Tn-seq, the screening of large libraries of mutant populations by high-throughput sequencing allows to identify bacterial fitness genes required for a specific condition at a genome-wide scale and in a single experiment (Chao *et al.*, 2016; Pritchard *et al.*, 2014). Here, I used a Tn-seq approach to identify bacterial determinants of *B. insecticola* involved in the colonization of its host *R. pedestris*. Prior to start this work, I studied the bottleneck effect applied by the host insect to the symbiotic colonizing population by calculating the initial bottleneck bacterial size with the Tn-seq methodology. In this study, I was able to perform an *in vivo* Tn-seq experiment and to obtain a genome-wide list of symbiotic determinants required for colonization of the symbiotic organ in *R. pedestris*.

During host colonization, bacterial populations have to face different selective pressures and physical barriers which tend to affect the composition of successful populations notably by reducing the number of initial invaders (Abel *et al.*, 2015; Chao *et al.*, 2016). Additionally, host bottlenecks such as constraint available space or stochastic sampling of the bacterial population from the environment are also playing a role in shaping the successful invading population (Abel *et al.,* 2015; Chao *et al.*, 2016). In *R. pedestris*, such bottleneck effect can be applied on the symbiotic population when bacteria have to cross the constricted region (CR) to reach the symbiotic organ (M4 region). This CR constitutes a physicochemical barrier made of microvilli and a mucous matrix rich in polysaccharides (Ohbayashi *et al.*, 2015), and as its name suggests, can possibly constrict the number of successful symbiotic colonizers. In this study, I estimated the initial bottleneck symbiont size from a defined inoculum using the Tn-seq approach. By providing 10^6 bacteria per insect which contained 110,735 individual potential mutants, I found that approximately 10,000 bacterial mutants were able to colonize the symbiotic organ. This estimation of the bottleneck was in good agreement with the number obtained in an independent experiment using mixed infections. This result suggested

that there was a strong bottleneck effect applied on the symbiotic population that could be attributed to the midgut anatomical constraint of the CR as well as the physiological parameters of the midgut environment such as the production of AMPs. By counting the total number of bacteria that colonized the symbiotic organ, I noticed that approximately 4 x 10^5 and 2.5 x 10^6 CFUs were sampled per symbiotic organ one and three days post-infection, respectively, which means that the symbiotic bacteria grow very efficiently inside the symbiotic organ. Particularly during the first day, 5 generations are obtained. Taking into account that the bacteria need 6 hours to reach the CR and that there might be some time required for passing the CR, these 5 generations are formed in less than 18 hours. So, this means that the bacterial doubling time in the M4 organ is close to the doubling time in YG rich medium, which is 2 to 3 hours.

Similar to my Tn-seq approach, Brooks *et al.*, have estimated the initial bottleneck size of a defined *V. fischeri* inoculum that could colonize the squid light organ using INSeq (Brooks *et al.*, 2014; Goodman *et al.*, 2009). They found that, out of a defined 96-mutant library, over 80 mutants were recovered per animal which revealed that a large number of independent bacterial mutants were able to efficiently colonize the squid light organ even with some initial host bottleneck (Brooks *et al.*, 2014). However, no precise bottleneck estimation was performed in this study. In another insect symbiosis involving cicada species and their vertically-transmitted symbionts "*Candidatus* Hodgikinia cicadicola" and "*Candidatus* Sulcia muelleri", it was recently shown that the number of transmitted bacterial cells to each egg strongly varies between cicada species (Campbell *et al.,* 2018). By counting the number of each bacterial cell using fluorescent microscopy, it appeared that approximately 12,000 "*Candidatus* Hodgikinia cicadicola" cells were transmitted to each egg in *Tettigades chilensis* (Campbell *et al.*, 2018), which is quite close to the number of *Burkholderia* cells that were able to colonize each *R. pedestris* midgut found in this work.

Interestingly, it appeared that the bottleneck sizes were distributed in two categories, with approximately 7,000 and 13,000 bacterial mutants, respectively. This bimodal distribution suggests that there are two different insect populations that have their own bottleneck effects on the symbiotic population. One parameter which separates the insect population in two categories is the gender and these two distributions could represent the infection bottleneck sizes in males and in females. Thus, it could be interesting to start a new estimation of the

initial bottleneck sizes in male insects and in female insects. However, other physiological differences between the insects could also be at the basis of this different bottleneck size. For example, the time lap between feeding on the soybeans and the ingestion of the bacteria could influence the midgut content and the passage of the infecting bacteria.

Based on the bottleneck size estimation, I have designed an *in vivo* Tn-seq experiment and the obtained results provided a genome-wide overview of cellular pathways required to establish efficient host colonization by the *Burkholderia* symbiont. By checking different host compartments, I found that the colonization of the different midgut sections involves specific cellular processes, but depends mainly on cell surface components, stress response elements, metabolic activities similar to low nutrition growth conditions, and DNA repair machineries. Similar cellular pathways were already described in previous *in vivo* Tn-seq studies for the colonization of the bee gut ileum by *S. alvi* (Powell *et al.*, 2016) and the colonization of the squid light organ by *V. fischeri* (Brooks *et al.*, 2014; Lyell *et al.*, 2017). Among the diverse Tn-seq publications, a few genome-wide studies were conducted to unravel colonization mechanisms of symbiotic bacteria (Brooks *et al.*, 2014; Lyell *et al.*, 2017; Phelan *et al.*, 2019; Powell *et al.*, 2016; Stephens *et al.*, 2015) and pathogenic bacteria (Fu *et al.*, 2013; Gutierrez *et al.*, 2015; Wang *et al.*, 2014) in their native host. By comparing these colonization factors, some symbiotic functions of *B. insecticola* required to colonize the symbiotic organ of *R. pedestris* are also critical as virulence mechanisms for human pathogens (Fu *et al.,* 2013; Wang *et al.*, 2014) such as DNA breaks repair, LPS modifications with O-antigen biosynthesis, amino acid metabolism and responses to various stressors. Additionally, it was also interesting to notice that nearly all previous characterized symbiotic functions such as purine metabolism (Kim *et al.*, 2014), LPS (Kim *et al.*, 2016, 2017) and peptidoglycan biosynthesis (Kim *et al.*, 2013b; Lee *et al.*, 2015) were also identified as symbiotic factors in this *in vivo* Tn-seq analysis. Only the PHA biosynthesis was not confirmed (Jang *et al*., 2017; Kim *et al*., 2013a). Thus, these correlations demonstrate that the *in vivo* Tn-seq experiments with *B. insecticola* were robust and allowed to detect previous validated bacterial targets involved in host colonization. Regarding the overall metabolic activities, it appears that the symbiont requires to synthesize specific essential amino acids such as arginine, methionine and leucine. In addition to its own metabolism, the *Burkholderia* symbiont may also supply these amino acids to support the insect host's metabolism (Ohbayashi *et al.*, 2019). On the other hand, the Tn-seq data suggest

that the branched chain amino acids leucine and isoleucine and the indole intermediate of tryptophan biosynthesis are nutrients provided by the host to the M4 bacteria. The essentiality of the gluconeogenesis in the M4 organ further suggests that a gluconeogenic carbon source (such as amino acids or lipids) is provided by the insect. The symbiont also seems to produce B vitamins in the symbiotic organ, especially vitamins B1, B6 and B12, which can support diverse bacterial functions (Asakura *et al.*, 2013; Fang *et al.*, 2017; Palmer and Downs, 2013) and may also be provided to the host (Ohbayashi *et al*., 2019). On the contrary, the requirement of the MgtC protein involved in the adaptation to low magnesium environments and the zinc transporter ZnuABC (Gabbianelli *et al.*, 2011; Maloney and Valvano, 2006) for the *in vivo* bacterial fitness may suggest that the host is providing these micronutrients in very limited amounts to the symbiont and possibly uses them to control the bacterial population, similarly as in nutritional immunity.

One of the most striking differences between the colonization of the first midgut regions and the symbiotic organ resides in the exclusive recruitment of cell motility mechanisms for the colonization of the symbiotic organ. The major identified pathways related to cell motility were the flagellar motility and chemotaxis functions. As previous studies have reported that flagellar motility-deficient mutants of *B. insecticola* lost their ability to colonize the host midgut, these Tn-seq results confirmed their requirement for the colonization of the symbiotic organ (Lee *et al.*, 2015; Ohbayashi *et al.*, 2015). Similarly, flagellar motility was showed to be crucial for host colonization in different symbiotic bacteria including *V. fischeri* (Brooks *et al.*, 2014) and *Aeromonas veronii* (Stephens *et al.*, 2015), but also in pathogenic species such as *V. cholerae* (Fu *et al.*, 2013). Linked to the flagellar motility, I found that multiple chemotaxis proteins were required for the bacterial fitness inside the symbiotic organ. These chemotaxis functions, especially with two MCPs, are important colonization factors and suggest that the symbiont is potentially attracted by different compounds inside the symbiotic organ. Once symbiotic bacteria are entering the host's midgut, these chemoattractive molecules might be secreted by the M4 region and direct the bacterial movement towards the M4 section. As it was previously reported that these MCPs are able to detect specific amino acids and carbon sources (Baker *et al.*, 2006; Callahan and Parkinson, 1985; Hedblom and Adler, 1980; Wang and Koshland, 1980), it would be interesting to identify possible chemoattractants of *B. insecticola.* It is further interesting to note that once the symbionts are established in the M4

region, motility and chemotaxis are shut down as it was shown by the transcriptomic analysis of the M4 bacteria (Ohbayashi *et al.*, 2019).

Hence, all these highlighted cellular pathways suggest that the *Burkholderia* symbiont is submitted to a challenging environment inside the symbiotic organ, with fluctuating oxygen supply, specific nutrient availability, acidic variations, osmotic pressures, oxidative stress and the presence of AMPs (**Figure 130**). The specific requirement of extracellular components and stress response elements for the host colonization, such as the LPS core oligosaccharide biosynthesis and the ClpPX proteases which were identified in this study, seem to confirm that the symbiont is facing several stress factors. Although the O-antigen was shown to be absent at the surface of *in vivo B. insecticola* cells (Kim *et al.*, 2017), I found that the O-antigen biosynthesis and its export are key functions for the host colonization. Even if the O-antigen component of LPS seems crucial for the colonization, the structural change of LPS molecules may be impacted by the different stress factors present in the symbiotic organ. However, even with these particular stress conditions, the *Burkholderia* symbiont is able to grow efficiently inside the symbiotic organ (this work; Kikuchi *et al.*, 2011), thus reflecting its specific adaptation towards its host niche. In addition, this stressful environment generated inside the crypts might have an impact on the bacterial cell surface and might explain the rapid morphological changes observed for *in vivo* bacterial cells (Ohbayashi *et al.*, 2019). Interestingly, I found that almost 28% of the identified symbiotic functions also participate to the resistance towards AMPs, which strongly suggest that AMP resistance mechanisms are crucial for the host colonization. As AMPs are known to cause morphological alterations of bacterial cells (Arouri *et al.*, 2011; Su *et al.*, 2012), the observed *in vivo* cell shape might be attributed to the activity of the host AMPs, including the CCR peptides that are massively expressed in the symbiotic organ. Thus, these symbiotic factors, which are also AMP resistance factors, may be particularly important to resist the pressure exerted by the host AMPs during the host colonization process.

To follow this work, it would be imperative to validate experimentally these *in vivo* fitness genes identified in this study to confirm their requirement for the host colonization (ongoing work). These validations are also critical for investigating the role of the plasmid 2 enriched genes. As mentioned before, the cell wall of symbiotic bacteria is strongly altered during *in vivo* colonization (Kim *et al.*, 2017; Ohbayashi *et al.*, 2019). As these genes are putatively

encoding for a capsular polysaccharide and that this compound may negatively affect the colonization efficiency of the symbiont, it could be of interest to identify the capsular polysaccharide produced by *B. insecticola* and study its role in the symbiosis efficiency. As I have noticed multiple differences in the genetic requirements of the symbiont between the second instar and the third instar developmental host stages, it would be useful to study the bacterial requirements for the population maintenance in the other steps of the insect's development. These differences were notably pointed out by the requirement of the entire plasmid 2 at the third instar compared to the second instar nymphal stage. Based on recent experimental data (unpublished data), it was demonstrated that this requirement was actually reflecting the loss of this plasmid in the third and later instars. Thus, by checking two close early steps of the insect's development, there were noticeable dynamic changes of symbiotic factors recruited for the symbiont's fitness inside the symbiotic organ. A Tn-seq study on *B. insecticola* that could cover the whole midgut regions at the different larval stages, including the adult form, would provide a complete overview of the dynamics of the genetic background of the symbiont required to colonize, proliferate and maintain the symbiotic population inside the host. In addition, it could be of interest to perform experiments on *R. pedestris* insects feeding on different seeds than soybeans because this might have an impact on the nutrients that the host is providing to the bacteria, and in nature *R. pedestris* is feeding on soybeans only at the end of the season when soybean plants have set seeds. Another interesting comparison could be made by determining the fitness landscape of the *B. insecticola* bacterium in other stinkbug hosts. Indeed, this bacterium can efficiently colonize related stinkbug species, such as *C. marginatus* or *C. punctiger*. Finally, *B. insecticola* is a bacterium that can adapt different lifestyles which also lives in soils or in the plant rhizosphere. Determining the genetic requirements for the lifestyles in these environments and the comparison with the here determined requirements for the lifestyle in the insect gut could give interesting information on the evolution of this insect symbiotic bacterium.

Chapter V

General Discussion and Perspectives

1. Discussion

Symbiotic associations with microorganisms, especially with bacterial species, are widespread among insect species (Douglas, 2011; López-García *et al.*, 2017). Indeed, insects have specific diets that lack essential nutrients such as vitamins or amino acids, which are provided by their bacterial symbionts to sustain their nutritional requirements (Engel and Moran, 2013). In addition, some microbial partners of insects provide increased resistance to abiotic or biotic stresses. Different biological models are studied to understand these interactions, including the symbiosis between the vertically-transmitted symbiont *B. aphidicola* and its aphid host (Shigenobu and Wilson, 2011). Another insect symbiosis model, based on a horizontal transmission mode, was recently described as a suitable system to study symbiotic interactions, which involves the stinkbug *R. pedestris* and its bacterial symbiont *B. insecticola* (Kikuchi *et al.*, 2007; Takeshita and Kikuchi, 2017; Takeshita *et al.*, 2018). Belonging to the Heteroptera suborder, this stinkbug species, also called the bean bug, is a notorious crop pest located in South-Eastern Asia which feeds preferentially on soybean, pigeon pea, cowpea and chickpea seeds (Bae *et al.*, 2014; Kikuchi *et al.*, 2007). This insect possesses a unique and specific extracellular bacterial symbiont, *B. insecticola*, located in a specific region of the midgut, named the M4 region, which constitutes the symbiotic organ where the symbiont proliferates (Kikuchi *et al.*, 2007). This symbiont is acquired from the environment at early stages of the host's development, and promotes beneficial effects on the host's growth, development and fecundity (Kikuchi *et al.*, 2007). In addition, the symbiont can be cultured *in vitro* and an aposymbiotic lineage of *R. pedestris* can be generated in laboratory conditions, thus constituting a suitable model to study both the host and the symbiont separately (Takeshita and Kikuchi, 2017). Surprisingly, even if *B. insecticola* is a facultative symbiont, each collected wild insect is colonized by this symbiont which suggests a strong and *de facto* obligatory association between these two partners (Takeshita and Kikuchi, 2017). However, the mechanisms by which the *Burkholderia* symbiont is selected by the host is poorly understood. A transcriptomic analysis conducted on the host side revealed that a specific category of antimicrobial peptides or AMPs are expressed in the symbiotic organ, which are the crypt-specific cysteine-rich peptides or CCR peptides (Futahashi *et al.*, 2013). These AMPs were shown to exert antimicrobial properties but do not participate in the immune response

triggered by a septic shock, hence these peptides were considered as symbiotic AMPs (**Figure 21** in **Chapter I**; unpublished data). Interestingly, in other symbiosis systems, host organisms produce symbiotic AMPs which are able to control and maintain the symbiotic bacterial populations (Mergaert, 2018). As it was shown that *Burkholderia* species are particularly resistant towards AMPs (Loutet and Valvano, 2011), we hypothesized that symbiotic CCR AMPs produced by *R. pedestris*, participate in the specific colonization of the symbiotic organ by the *B. insecticola* symbiont. Moreover, immunity-related AMPs such as riptocin, rip-defensin and rip-thanatin can also be expressed in the midgut and the M4 region under particular conditions, including the molting stage, starvation of the insect, hemolymph infection or the presence of commensal bacteria in the midgut (Park *et al.*, 2018; Seong Han Jang and Yoshitomo Kikuchi, unpublished data).

In this work, I have used a Tn-seq approach for the first time on *B. insecticola* to identify bacterial factors required for AMP resistance and for host colonization. The Tn-seq methodology relies on the generation of a bacterial transposon library that is coupled with high-throughput sequencing in order to identify genes involved in the bacterial fitness for a specific condition (Chao *et al.*, 2016). Based on this method, I was able to pinpoint the main bacterial functions involved in the *in vivo* colonization of the symbiotic organ, including some bacterial genes that were previously characterized in independent studies. This work showed that Tn-seq is a robust and powerful genetic tool to screen for bacterial factors involved in a defined condition, including both *in vitro* and *in vivo* settings.

1.1. The essential genome of *B. insecticola* differs from other *Burkholderia* species

As Tn-seq was primarily used to identify essential genomes in bacterial species on rich media (Barquist *et al.*, 2013; Chao *et al.*, 2013; DeJesus *et al.*, 2017; Hooven *et al.*, 2016), including multiple *Burkholderia* species (Baugh *et al.*, 2013; Moule *et al.*, 2014; Wong *et al.*, 2016), I have described the essential genome of *B. insecticola* on a rich medium condition (see **Chapter II**). This method enabled to find 1,080 essential genes in the *B. insecticola* genome that were mostly located in the chromosome 1 and the plasmid 1 (see **Chapter II**). The main essential functions identified by Tn-seq involved the transcription and translation machineries such as the ribosomal subunits, the energy production with the ATP synthase, the transport of amino

acids, and also cell wall functions including the peptidoglycan biosynthesis (see **Chapter II**). These essential functions were previously identified in other bacterial species (Christen *et al.*, 2014), which strongly confirmed the participation of these functions to sustain the bacterial viability of *B. insecticola*. By comparing this essential gene set with the essential genes identified for three other *Burkholderia* species, including *B. cenocepacia* J2315 (Wong *et al.*, 2016), *B. thailandensis* E264 (Baugh *et al.*, 2013) and *B. pseudomallei* K96243 (Moule *et al.*, 2014), it appeared that only 151 essential genes were shared between these four *Burkholderia* species which constitute the core essential genome of *Burkholderia* species (see **Chapter II**). Before obtaining the essential genes of *B. insecticola*, the comparison of the essential gene sets of these three *Burkholderia* species revealed that only 164 essential genes were shared (Wong *et al.*, 2016), which is close to the number I found in this work by adding the *B. insecticola* essential genes. As the proportion of essential genes was higher for *B. insecticola* (17%) than for the other three species (6.1 to 8.5%), this suggests that there are specific sets of essential genes required for the viability of species belonging to different clades among the *Burkholderia* genus. Thus, the actual comparison showed that 715 essential genes were specifically found for the viability of *B. insecticola*, which contained genes that covered complete predicted essential pathways, such as the identification of all the ribosomal subunits (see **Chapter II**). This finding suggests that the Tn-seq method I have used is sufficiently robust to identify overall essential genes of a defined pathway. However, a huge proportion of these specific essential genes encode for hypothetical proteins with unknown functions that were mostly homologous to hypothetical proteins of a closely related species, *Burkholderia sp.* YI23. Knowing that *B. insecticola* and *Burkholderia sp.* YI23 are soil microorganisms, these specific identified essential genes may be attributed to the niche adaptation of these two species. Hence, it would be interesting to identify the essential genome of the *Burkholderia* sp. YI23 with the same Tn-seq settings than *B. insecticola* and find a correlation between the essential gene sets of these two *Burkholderia* species. Finally, it should be noted that essentiality in the different species were defined with different experimental and bioinformatics setups entailing probably methodology-related biases and differences in the stringency to assign essentiality to a gene.

1.2. Is Tn-seq sufficiently robust? A case study to detect bacterial genes involved in carbon source exploitation

As Tn-seq is a genome-wide approach to identify bacterial essential genomes, this method is also used to find bacterial genes required for a specific condition (Chao *et al*., 2016). By comparing the insertion profiles of the bacterial population before and after a specific treatment or condition, genes can be identified which have a differential number of transposon insertions between the compared conditions (Chao *et al.*, 2016; DeJesus and Ioerger, 2013; van Opijnen *et al*., 2009). Thus, these genes represent fitness genes which are predicted to play a role in the specific treatment or condition tested. In order to check the robustness of the Tn-seq method in *B. insecticola*, I have identified bacterial factors involved in the exploitation of two carbon sources, glucose and succinate, in a minimal medium condition (see **Chapter II**). In this work, I identified specific transporting systems for glucose and succinate, respectively, and I revealed that the Entner-Doudoroff glycolysis pathway was the main catabolic route used to degrade glucose in *B. insecticola* (see **Chapter II**). For succinate exploitation, it appeared that this carbon source can be directly incorporated in the TCA cycle to produce energy, or can be assimilated in the gluconeogenesis pathway to generate glucose and the different carbohydrates that are required to build the different cellular constituents (see **Chapter II**). Since this pattern of genes corresponded to what was expected, I conclude that this Tn-seq approach was sufficiently robust to identify multiple bacterial factors required for the use of these two carbon sources. In addition, by comparing the growth in the YG rich medium that contains many cellular building blocks, pre-made for the bacteria to incorporate them in their metabolism, and the growth in the MM in which none of these components are present, I identified tens of genes involved in the anabolic pathways for synthesis of nucleotides, amino acids, vitamins and others. Again, this validates the transposon library and the Tn-seq approach. In conclusion, the robustness of these genome-wide screens performed with different growth media were highly encouraging to set up further *in vitro* and *in vivo* Tn-seq experiments to analyse the *B. insecticola*-*R. pedestris* symbiosis.

1.3. AMP resistance mainly involves cell wall functions and plays a role in host colonization efficiency

The identification of bacterial factors in *B. insecticola* involved in AMP resistance was based on a candidate-gene approach and on the application of Tn-seq. As AMPs constitute a large family, I have chosen five AMPs, including polymyxin B, LL-37 and three AMPs produced by *R. pedestris* (two CCR peptides and riptocin) in order to represent the large diversity of this peptide family. As a first approach, I assessed the roles of previously characterized AMP resistance factors described in *Burkholderia* species, such as LPS (Loutet *et al.*, 2006), hopanoids (Malott *et al.*, 2012; Schmerk *et al*., 2011) and the RpoE ESR pathway (Flannagan and Valvano, 2008) (see **Chapter III**). However, among these membrane targets, I confirmed that the LPS was involved as an AMP resistance factor, but the RpoE factor was only required for riptocin resistance and I did not confirm a role for the hopanoid lipids in the resistance to any of the tested peptides (see **Chapter III**). This suggested that there are considerable differences between species, even between related *Burkholderia* strains, in the mechanisms that they mobilize to resist challenges with AMPs.

Therefore, in a second approach, I have used Tn-seq to identify global and specific resistance factors towards the five selected AMPs (see **Chapter III**). I identified 42, 42, 15, 21 and 39 fitness genes that were required for polymyxin B, LL-37, riptocin, CCR179 and CCR480 peptides, respectively (see **Chapter III**). This analysis revealed further that only three fitness genes were shared between these five AMPs. These three genes encode the Tat transporting system, which was shown to participate in outer membrane integrity in *E. coli* (Ize *et al.*, 2003). It is not likely that the TatABC proteins themselves are responsible for the AMP resistance; it is more likely that this is the responsibility of one or several of the client proteins of the Tat transporter. It will thus be of interest to determine experimentally the Tat-dependent secretome in *B. insecticola* which could lead to the identification of the direct determinant(s) of AMP resistance. Prediction of this secretome has identified a number of interesting candidates such as β-lactamases, the kinase LpxK involved in lipid A biosynthesis, the LPS-assembly protein LptD or the amidase AmiA involved in peptidoglycan synthesis (see **Chapter III**). Interestingly, the latter protein is essential for the M4 colonization according to the *in vivo* Tn-seq data. However, none of the predicted Tat clients was among the identified AMP

resistance genes.

The majority of the resistance factors identified for each AMP was representing cell wall biogenesis functions, including the core oligosaccharide and O-antigen components of LPS with fitness genes that were previously characterized in the candidate-gene approach (see **Chapter III**). As the main cellular target of AMPs is the bacterial membranes (Kumar *et al.*, 2018), it was not surprising to find this biological category in abundance among the functions of these fitness genes. To confirm these Tn-seq results, I created *B. insecticola* mutants from common identified fitness genes such as *dsbA*, *wzm*, *tolB*, *tolQ*, *rfbA* and *rfbC*. I have found that these *B. insecticola* mutant strains were hypersensitive towards AMPs, thus confirming the role of the predictive genes in AMPs resistance identified by Tn-seq.

To confirm my initial hypothesis about the connection between AMP resistance and host colonization, I have tested the colonization efficiencies of the AMP-sensitive mutant strains. Concerning the LPS inner core oligosaccharide mutants *waaC* and *waaF*, I confirmed that these two mutant strains are completely unable to colonize the host (Kim *et al.*, 2017), and I identified an additional mutant in the *rfbA* gene with a similar phenotype. In addition, six mutants constructed based on the Tn-seq approach have intermediate colonization phenotypes which were not described before, especially with the *tol* mutant strains that displayed a specific localization in the posterior region of the symbiotic organ (see **Chapter III**). These results revealed that AMP-sensitive strains showed colonization deficiencies, which confirms indeed that AMP resistance is a key fitness trait of *B. insecticola* necessary to colonize the symbiotic organ of *R. pedestris*.

During these experiments, I showed that some strains that were able (with reduced efficiency) to colonize the host during mono-infections became totally outcompeted by the wild-type strain during *in vivo* competitions, such as the *dsbA*, *mlaD*, *rpoE*, *wbiF*, *tolB* and *tolQ* mutants (see **Chapter III**). Hence, this work suggests that it must be recommended to include competition experiments for the characterization of *in vivo* phenotypes, which seems more representative to describe infection behaviours in a population context.

Additionally, by analyzing the host fitness parameters of insects infected by the mutant strains produced in this work, it appeared that the adult emergence rate and the morphometric parameters are uncoupled host developmental features (see **Chapter III**). Interestingly, insects infected by the RpoE ESR pathway mutant strains, *rpoE* and *mucD*, showed a regular adult

emergence rate compared to symbiotic insects but displayed different morphological parameters which suggest host metabolic deficiencies (see **Chapter III**). Thus, the colonization efficiency of the symbiotic organ by a mutant strain does not ensure that the host fitness traits would exactly follow the attributed phenotype observed in the symbiotic organ.

1.4. Host colonization does not rely only on the capacity to resist AMPs

In addition to the *in vitro* Tn-seq analyses in different media and in the presence of AMPs, I have performed an *in vivo* Tn-seq on *R. pedestris* insects. As there are anatomical constraints imposed by the host midgut morphology, especially due to the constricted region which forms a narrow passage for the symbiotic bacteria to the symbiotic organ (Ohbayashi *et al.*, 2015), and which is open for only a few hours, a prerequisite was to make an estimation of the infection bottleneck of the *Burkholderia* symbiont. By using the Tn-seq library, I determined that approximately 10,000 bacteria were able to colonize the symbiotic organ of one insect, starting from an initial population of 10^6 bacteria (see **Chapter IV**). This estimate was in very good agreement with an independent method based on mixed infections with a GFP-marked strain. These results were crucial to settle the *in vivo* Tn-seq experiments by obtaining a precise number of insects to sacrifice in order to recover a sufficient quantity of bacteria for the sequencing. Interestingly, the measured bottleneck sizes showed a bimodal distribution which suggest that there must be another parameter that influences the bottleneck size of the symbiont population (see **Chapter IV**). An intriguing hypothesis is that the bottleneck size is different according to the gender of *R. pedestris*, but other factors could be responsible as well, such as the time laps between feeding and infection.

Taking into account the bottleneck effect, I have performed an *in vivo* Tn-seq experiment by recovering the symbiotic population inside the symbiotic organ at the second and the third instar stages. In addition, I have also sampled other midgut regions, the M1 and M3 organs, during the second instar stage, for which the symbiotic population is temporarily present during the host colonization. The *in vivo* Tn-seq approach revealed that 37, 18, 129 and 329 conditionally essential genes were required for the bacterial fitness in the M1, M3, M4 at second instar and M4 at third instar *in vivo* conditions, respectively (see **Chapter IV**). Among the fitness genes required for the colonization of the M4 organ, there were five main cellular

functions which were the DNA repair machineries, diverse metabolic activities, responses to stress factors, cell wall biogenesis and the cell motility functions (see **Chapter IV**). Interestingly, these biological categories were previously found in other Tn-seq *in vivo* studies performed on symbiosis systems, such as the bee gut symbiont *S. alvi* (Powell *et al*., 2016) and the *V. fischeri* symbiont of the bobtail squid (Stephens *et al.*, 2015). These bacterial functions suggest that the *Burkholderia* symbiont faces multiple stresses inside the symbiotic organ, such as the presence of AMPs, osmotic and oxidative pressures, and variations of pH, oxygen and temperatures that cause alterations of the metabolism and the cell morphology of symbiotic bacteria (see **Chapter IV**) (Ohbayashi *et al.*, 2019b). Among these bacterial functions, I found genes that were characterized before for their role in the host colonization in independent studies, such as genes from the LPS core oligosaccharide and O-antigen biosynthesis pathways (Kim *et al.*, 2016b, 2017), the biosynthesis of purines (Kim *et al.*, 2014a), the peptidoglycan stability (Lee *et al.*, 2015) and the flagellar motility (Lee *et al.*, 2015; Ohbayashi *et al.*, 2015).

It appeared that approximately 28% of the identified symbiotic functions also participate to AMP resistance (see **Chapter IV**), which suggest that the capacity of *B. insecticola* to resist AMPs is a key feature of *B. insecticola* for the successful colonization of the symbiotic organ. However, there are still a majority of these symbiotic factors that are not related to AMP resistance, which means that the colonization efficiency of *B. insecticola* is not only dependent of its ability to resist AMPs and necessitates many other cellular functions, such as the above mentioned flagellar motility, chemotaxis and specific metabolic requirements; but also many genes, including genes encoding proteins of unknown function, for which the specific contribution in the gut colonization remains obscure. This is illustrated by the hopanoid mutants, previously studied in the candidate-gene approach, which were not involved in AMP resistance but displayed some colonization deficiencies (see **Chapter III**). Thus, this work contributed to the identification on a genome-wide scale of the bacterial factors involved in symbiotic interactions between *B. insecticola* and *R. pedestris* that are required to colonize the symbiotic organ. This dataset is a gold mine for future studies that can analyse the specific roles of these genes.

2. Perspectives

In order to confirm the *in vivo* Tn-seq analysis, it is necessary to target multiple fitness genes involved in the host colonization by directed mutagenesis, especially for the symbiotic organ. Many genes found in this work were previously described in the literature as important for the insect colonization, such as the lipopolysaccharide biosynthesis genes (Kim *et al.*, 2016b, 2017). Hence, it is of interest to focus on different functional categories, such as the *tatABC* and the inorganic ion transporter *znuABC* gene clusters as well as the many function unknown (FUN) genes (see **Chapter IV**), which were not reported as symbiotic factors. Possibly, among the identified genes, there are functions that are important contributors for the very strong specificity between *B. insecticola* (and related species) and *R. pedestris*. An interesting approach could be to focus on the *in vivo* fitness genes that are specifically present in the genomes of *B. insecticola* and allied species but absent in the genomes of distant species that cannot colonize the *R. pedestris* midgut. As chemotaxis genes were also required for the colonization of the symbiotic organ, it would be of interest to find which chemoattractive molecules are produced by the host that are sensed by the *Burkholderia* symbiont.

Additionally, this work highlighted the importance of the plasmid 2 for the symbiont adaptation during the host development (see **Chapter IV**). The investigation of the plasmid 2 genes' functions may be relevant to understand the progressive loss of the entire plasmid 2 at the third instar larval stage, especially by focusing on the specific gene cluster which is highly enriched in transposon insertions (see **Chapter IV**).

The Tn-seq methodology was used to identify symbiotic factors mobilized for the host colonization at early stages of the insect development (see **Chapter IV**). As this *in vivo* Tn-seq method works efficiently and is robust, it is feasible to perform a Tn-seq experiment on symbiotic insects by checking the symbiotic population at the adult stage. Such study would pinpoint bacterial genes implicated in the long-term maintenance of the symbiotic population. In addition, it could be envisaged to use Tn-seq to study the impact of the nutritional status (*e.g.* applying starvation or feeding on different legume seeds) or the immunological status (*e.g.* in the presence of gut microbiota or hemolymph infections) of the insect on the required gene repertoire in *B. insecticola*. The *B. insecticola* strain that was used here can also infect other, closely related insect species, *e.g. C. punctiger*, *C. marginatus* or *L. occidentalis*, which

could be exploited to identify insect-species specific essential genes. *B. insecticola* is a bacterium which is adapted to different lifestyles as it is a soil bacterium and it can also efficiently colonize plant roots (unpublished data). It will certainly be of interest to use the Tn-seq approach to compare the required genes for growth and survival in these different environments. Finally, Tn-seq experiments could be performed on a range of different *in vitro* conditions, including conditions that are suspected to mimic the crypt conditions (similarly as my approach to analyse *in vitro* AMP treatments) as a strategy to pinpoint functions of poorly annotated genes (FUN genes).

I observed that insects infected with some bacterial mutants, which were able to colonize efficiently the symbiotic organ, displayed aposymbiotic host fitness parameters (see **Chapter III**). Such mutants, more specifically the *mucD* mutant, might affect the metabolite production of the symbiont, hence changing the metabolic exchanges with its host. It might be interesting to conduct a metabolomic analysis focused on the symbiotic organ of *R. pedestris* insects by comparing the metabolic profiles of these mutants with the wild-type symbiont. As metabolomic analyses were already performed on bee gut microbiota and on different midgut compartments of bees (Kešnerová *et al.*, 2017; Zheng *et al.*, 2017), it must be feasible to design the same type of experiments on *R. pedestris*. The generation of both bacterial and host metabolomes would spot the key molecules exchanged between these two partners and would help to understand the nutrient requirements to sustain this symbiosis.

References

Abel, S., Abel zur Wiesch, P., Davis, B.M., and Waldor, M.K. (2015). Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog.* 11(6), e1004823. doi: 10.1371/journal.ppat.1004823.

Aertsen, A., Vanoirbeek, K., De Spiegeleer, P., Sermon, J., Hauben, K., Farewell, A., Nyström, T., and Michiels, C.W. (2004). Heat shock protein-mediated resistance to high hydrostatic pressure in *Escherichia coli*. *Appl. Environ. Microbiol.* 70(5), 2660–2666. doi: 10.1128/AEM.70.5.2660-2666.2004.

Agnoli, K., Frauenknecht, C., Freitag, R., Schwager, S., Jenul, C., Vergunst, A., Carlier, A., and Eberl, L. (2014). The third replicon of members of the *Burkholderia cepacia* Complex, plasmid pC3, plays a role in stress tolerance. *Appl Environ Microbiol*. 80(4), 1340-1348. doi: 10.1128/AEM.03330-13.

Ahn, K.-S., Ha, U., Jia, J., Wu, D., and Jin, S. (2004). The *truA* gene of *Pseudomonas aeruginosa* is required for the expression of type III secretory genes. *Microbiology.* 150(Pt 3), 539–547. doi: 10.1099/mic.0.26652-0.

Ahuja, U., Kjelgaard, P., Schulz, B.L., Thöny-Meyer, L., and Hederstedt, L. (2009). Haem-delivery proteins in cytochrome c maturation System II. *Mol. Microbiol.* 73(6), 1058–1071. doi: 10.1111/j.1365-2958.2009.06833.x.

Akiyama, Y., Kanehara, K., and Ito, K. (2004). RseP (YaeL), an *Escherichia coli* RIP protease, cleaves transmembrane sequences. *EMBO J.* 23(22), 4434–4442. doi: 10.1038/sj.emboj.7600449.

Anderson, A.J., and Dawes, E.A. (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol. Rev.* 54(4), 450–472.

Anselme, C., Pérez-Brocal, V., Vallier, A., Vincent-Monegat, C., Charif, D., Latorre, A., Moya, A., and Heddi, A. (2008). Identification of the weevil immune genes and their expression in the bacteriome tissue. *BMC Biol.* 6, 43. doi: 10.1186/1741-7007-6-43.

Aravind, L., Walker, D.R., and Koonin, E.V. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* 27(5), 1223–1242. doi: 10.1093/nar/27.5.1223.

Armbruster, C.E., Forsyth-DeOrnellas, V., Johnson, A.O., Smith, S.N., Zhao, L., Wu, W., and Mobley, H.L.T. (2017). Genome-wide transposon mutagenesis of *Proteus mirabilis*: Essential genes, fitness factors for catheter-associated urinary tract infection, and the impact of polymicrobial infection on fitness requirements. *PLoS Pathog.* 13(6), e1006434. doi: 10.1371/journal.ppat.1006434.

Arnold, M.F.F., Shabab, M., Penterman, J., Boehme, K.L., Griffitts, J.S., and Walker, G.C. (2017). Genome-Wide Sensitivity Analysis of the Microsymbiont *Sinorhizobium meliloti* to Symbiotically Important, Defensin-Like Host Peptides. *MBio* 8(4). pii: e01060-17. doi: 10.1128/mBio.01060-17.

Arouri, A., Kiessling, V., Tamm, L., Dathe, M., and Blume, A. (2011). Morphological Changes Induced by the Action of Antimicrobial Peptides on Supported Lipid Bilayers. *J. Phys. Chem. B* 115(1), 158–167. doi: 10.1021/jp107577k.

Arrese, E.L., and Soulages, J.L. (2010). Insect fat body: energy, metabolism, and regulation. *Annu. Rev. Entomol.* 55, 207–225. doi: 10.1146/annurev-ento-112408-085356.

Asakura, H., Hashii, N., Uema, M., Kawasaki, N., Sugita-Konishi, Y., Igimi, S., and Yamamoto, S. (2013). *Campylobacter jejuni pdxA* affects flagellum-mediated motility to alter host colonization. *PLoS One* 8(8), e70418. doi: 10.1371/journal.pone.0070418.

Attardo, G.M., Lohs, C., Heddi, A., Alam, U.H., Yildirim, S., and Aksoy, S. (2008). Analysis of milk gland structure and function in *Glossina morsitans*: milk protein production, symbiont populations and fecundity. *J. Insect Physiol.* 54(8), 1236–1242. doi: 10.1016/j.jinsphys.2008.06.008.

Auerbuch, V., Lenz, L.L., and Portnoy, D.A. (2001). Development of a Competitive Index Assay To Evaluate the Virulence of *Listeria monocytogenes actA* mutants during Primary and Secondary Infection of Mice. *Infect. Immun.* 69(9), 5953–5957. doi: 10.1128/iai.69.9.5953-5957.2001.

Augustin, R., Anton-Erxleben, F., Jungnickel, S., Hemmrich, G., Spudy, B., Podschun, R., and Bosch, T.C.G. (2009). Activity of the novel peptide arminin against multiresistant human pathogens shows the considerable potential of phylogenetically ancient organisms as drug sources. *Antimicrob. Agents Chemother.* 53(12), 5245–5250. doi: 10.1128/AAC.00826-09.

Augustin, R., Fraune, S., and Bosch, T.C.G. (2010). How Hydra senses and destroys microbes. *Semin. Immunol.* 22(1), 54–58. doi: 10.1016/j.smim.2009.11.002.

Bae, S.D., Kim, H.J., and Mainali, B.P. (2014). Infestation of *Riptortus pedestris* (Fabricius) decreases the nutritional quality and germination potential of soybean seeds. *Journal of Asia-Pacific Entomology* 17(3), 477–481. doi: 10.1016/j.aspen.2014.04.006.

Bahar, A.A., and Ren, D. (2013). Antimicrobial peptides. *Pharmaceuticals (Basel)* 6(12), 1543–1575. doi: 10.3390/ph6121543.

Baker, M.D., Wolanin, P.M., and Stock, J.B. (2006). Signal transduction in bacterial chemotaxis. *Bioessays* 28(1), 9–22. doi: 10.1002/bies.20343.

Barb, A.W., and Zhou, P. (2008). Mechanism and inhibition of LpxC: an essential zinc-dependent deacetylase of bacterial lipid A synthesis. *Curr. Pharm. Biotechnol.* 9(1), 9–15.

Barquist, L., Langridge, G.C., Turner, D.J., Phan, M.-D., Turner, A.K., Bateman, A., Parkhill, J., Wain, J., and Gardner, P.P. (2013). A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res.* 41(8), 4549–4564. doi: 10.1093/nar/gkt148.

Barrière, Q., Guefrachi, I., Gully, D., Lamouche, F., Pierre, O., Fardoux, J., Chaintreuil, C., Alunni, B., Timchenko, T., Giraud, E., Mergaert, P. (2017). Integrated roles of BclA and DD-carboxypeptidase 1 in *Bradyrhizobium* differentiation within NCR-producing and NCR-lacking root nodules. *Sci. Rep.* 7(1), 9063. doi: 10.1038/s41598-017-08830-0.

de Bary, A. (1879). Die Erscheinung Der Symbiose: Vortrag, Gehalten auf der Versammlung Deutscher Naturforscher und Aerzte zu Cassel. 1831-1888.

Baugh, L., Gallagher, L.A., Patrapuvich, R., Clifton, M.C., Gardberg, A.S., Edwards, T.E., Armour, B., Begley, D.W., Dieterich, S.H., Dranow, D.M., *et al.* (2013). Combining Functional and Structural Genomics to Sample the Essential *Burkholderia* Structome. *PLoS One* 8(1), e53851. doi: 10.1371/journal.pone.0053851.

Bechinger, B., and Gorr, S.-U. (2017). Antimicrobial Peptides: Mechanisms of Action and Resistance. *J. Dent. Res.* 96(3), 254–260. doi: 10.1177/0022034516679973.

Belin, B.J., Busset, N., Giraud, E., Molinaro, A., Silipo, A., and Newman, D.K. (2018). Hopanoid lipids: from membranes to plant-bacteria interactions. *Nat. Rev. Microbiol.* 16(5), 304–315. doi: 10.1038/nrmicro.2017.173.

Belon, C., and Blanc-Potard, A.-B. (2016). Intramacrophage Survival for Extracellular Bacterial Pathogens: MgtC As a Key Adaptive Factor. *Front. Cell. Infect. Microbiol.* 6, 52. doi: 10.3389/fcimb.2016.00052.

de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., *et al.* (2008). A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* 4, 174. doi: 10.1038/msb.2008.10.

Bergmann, E.J., Venugopal, P.D., Martinson, H.M., Raupp, M.J., and Shrewsbury, P.M. (2016). Host Plant Use by the Invasive *Halyomorpha halys* (Stål) on Woody Ornamental Trees and Shrubs. *PLoS One* 11(2), e0149975. doi: 10.1371/journal.pone.0149975.

Bernier, S.P., Son, S., and Surette, M.G. (2018). The Mla Pathway Plays an Essential Role in the Intrinsic Resistance of *Burkholderia cepacia* Complex Species to Antimicrobials and Host Innate Components. *J. Bacteriol.* 200(18). pii: e00156-18. doi: 10.1128/JB.00156-18.

Berry, M.C., McGhee, G.C., Zhao, Y., and Sundin, G.W. (2009). Effect of a *waaL* mutation on lipopolysaccharide composition, oxidative stress survival, and virulence in *Erwinia amylovora*. *FEMS Microbiol. Lett.* 291(1), 80–87. doi: 10.1111/j.1574-6968.2008.01438.x.

Besse, A., Peduzzi, J., Rebuffat, S., and Carré-Mlouka, A. (2015). Antimicrobial peptides and proteins in the face of extremes: Lessons from archaeocins. *Biochimie* 118, 344–355. doi: 10.1016/j.biochi.2015.06.004.

Bhubhanil, S., Sittipo, P., Chaoprasid, P., Nookabkaew, S., Sukchawalit, R., and Mongkolsuk, S. (2014). Control of zinc homeostasis in *Agrobacterium tumefaciens* via *zur* and the zinc uptake genes *znuABC* and *zinT*. *Microbiology* 160(Pt 11), 2452–2463. doi: 10.1099/mic.0.082446-0.

Bishop, A.H., Rachwal, P.A., and Vaid, A. (2014). Identification of Genes Required by *Bacillus thuringiensis* for Survival in Soil by Transposon-Directed Insertion Site Sequencing. *Curr. Microbiol.* 68(4), 477–485. doi: 10.1007/s00284-013-0502-7.

Blatch, G.L., and Lässle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* 21(11), 932–939.

Blumenberg, M., Oppermann, B.I., Guyoneaud, R., and Michaelis, W. (2009). Hopanoid production by *Desulfovibrio bastinii* isolated from oilfield formation water. *FEMS Microbiol. Lett.* 293(1), 73–78. doi: 10.1111/j.1574-6968.2009.01520.x.

Bobrov, A.G., Kirillina, O., Fosso, M.Y., Fetherston, J.D., Miller, M.C., VanCleave, T.T., Burlison, J.A., Arnold, W.K., Lawrenz, M.B., Garneau-Tsodikova, S., Perry, R.D. (2017). Zinc transporters YbtX and ZnuABC are required for the virulence of *Yersinia pestis* in bubonic and pneumonic plague in mice. *Metallomics* 9(6), 757–772. doi: 10.1039/c7mt00126f.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15), 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bongrand, C., Koch, E.J., Moriano-Gutierrez, S., Cordero, O.X., McFall-Ngai, M., Polz, M.F., and Ruby, E.G. (2016). A genomic comparison of 13 symbiotic *Vibrio fischeri* isolates from the perspective of their host source and colonization behavior. *ISME J.* 10(12), 2907–2917. doi: 10.1038/ismej.2016.69.

van Borm, S., Buschinger, A., Boomsma, J.J., and Billen, J. (2002). *Tetraponera* ants have gut symbionts related to nitrogen-fixing root-nodule bacteria. *Proc. Biol. Sci.* 269(1504), 2023–2027. doi: 10.1098/rspb.2002.2101.

Brogden, K.A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* 3(3), 238–250. doi: 10.1038/nrmicro1098.

Brooks, J.F., Gyllborg, M.C., Cronin, D.C., Quillin, S.J., Mallama, C.A., Foxall, R., Whistler, C., Goodman, A.L., and Mandel, M.J. (2014). Global discovery of colonization determinants in the squid symbiont *Vibrio fischeri*. *Proc. Natl. Acad. Sci. U.S.A.* 111(48), 17284–17289. doi: 10.1073/pnas.1415957111.

Buchner, P. (1965). Endosymbiosis of Animals with Plant Microorganisms (Interscience Publishers).

Bulet, P., Urge, L., Ohresser, S., Hetru, C., and Otvos, L. Jr. (1996). Enlarged scale chemical synthesis and range of activity of drosocin, an O-glycosylated antibacterial peptide of *Drosophila*. *Eur. J. Biochem.* 238(1), 64–69. doi: 10.1111/j.1432-1033.1996.0064q.x.

Bulet, P., Hetru, C., Dimarcq, J.L., and Hoffmann, D. (1999). Antimicrobial peptides in insects; structure and function. *Dev. Comp. Immunol.* 23(4-5), 329–344.

Burrows, L.L., Pigeon, K.E., and Lam, J.S. (2000). *Pseudomonas aeruginosa* B-band lipopolysaccharide genes *wbpA* and *wbpI* and their *Escherichia coli* homologues *wecC* and *wecB* are not functionally interchangeable. *FEMS Microbiol. Lett.* 189(2), 135–141. doi: 10.1111/j.1574-6968.2000.tb09219.x.

Burtnick, M.N., and Woods, D.E. (1999). Isolation of polymyxin B-susceptible mutants of *Burkholderia pseudomallei* and molecular characterization of genetic loci involved in polymyxin B resistance. *Antimicrob. Agents Chemother.* 43(11), 2648–2656.

Bury-Moné, S., Nomane, Y., Reymond, N., Barbet, R., Jacquet, E., Imbeaud, S., Jacq, A., and Bouloc, P. (2009). Global analysis of extracytoplasmic stress signaling in *Escherichia coli. PLoS Genet.* 5(9), e1000651. doi: 10.1371/journal.pgen.1000651.

Byeon, J.H., Seo, E.S., Lee, J.B., Lee, M.J., Kim, J.K., Yoo, J.W., Jung, Y., and Lee, B.L. (2015). A specific cathepsin-L-like protease purified from an insect midgut shows antibacterial activity against gut symbiotic bacteria. *Dev. Comp. Immunol.* 53(1), 79–84. doi: 10.1016/j.dci.2015.06.003.

Caballero-Mellado, J., Onofre-Lemus, J., Santos, P.E. los, and Martínez-Aguilar, L. (2007). The Tomato Rhizosphere, an Environment Rich in Nitrogen-Fixing *Burkholderia* Species with Capabilities of Interest for Agriculture and Bioremediation. *Appl. Environ. Microbiol.* 73(16), 5308–5319. doi: 10.1128/AEM.00324-07.

Cai, Y., Lee, W., and Kwa, A.L. (2015). Polymyxin B versus colistin: an update. *Expert. Rev. Anti. Infect. Ther.* 13(12), 1481–1497. doi: 10.1586/14787210.2015.1093933.

Callahan, A.M., and Parkinson, J.S. (1985). Genetics of methyl-accepting chemotaxis proteins in *Escherichia coli*: *cheD* mutations affect the structure and function of the Tsr transducer. *J. Bacteriol.* 161(1), 96–104.

Campbell, M.A., Łukasik, P., Meyer, M.C., Buckner, M., Simon, C., Veloso, C., Michalik, A., and McCutcheon, J.P. (2018). Changes in Endosymbiont Complexity Drive Host-Level Compensatory Adaptations in Cicadas. *MBio* 9(6). pii: e02104-18. doi: 10.1128/mBio.02104-18.

Carey, A.F., Rock, J.M., Krieger, I.V., Chase, M.R., Fernandez-Suarez, M., Gagneux, S., Sacchettini, J.C., Ioerger, T.R., and Fortune, S.M. (2018). TnSeq of *Mycobacterium tuberculosis* clinical isolates reveals strain-specific antibiotic liabilities. *PLoS Pathog.* 14(3), e1006939. doi: 10.1371/journal.ppat.1006939.

Carlier, A.L., and Eberl, L. (2012). The eroded genome of a *Psychotria* leaf symbiont: hypotheses about lifestyle and interactions with its plant host. *Environ. Microbiol.* 14(10), 2757–2769. doi: 10.1111/j.1462-2920.2012.02763.x.

Carlier, A.L., Omasits, U., Ahrens, C.H., and Eberl, L. (2013). Proteomics analysis of *Psychotria* leaf nodule symbiosis: improved genome annotation and metabolic predictions. *Mol. Plant Microbe Interact.* 26(11), 1325–1333. doi: 10.1094/MPMI-05-13-0152-R.

Carvalho, F., Atilano, M.L., Pombinho, R., Covas, G., Gallo, R.L., Filipe, S.R., Sousa, S., and Cabanes, D. (2015). L-Rhamnosylation of *Listeria monocytogenes* Wall Teichoic Acids Promotes Resistance to Antimicrobial Peptides by Delaying Interaction with the Membrane. *PLoS Pathog.* 11(5). doi: 10.1371/journal.ppat.1004919.

diCenzo, G.C., and Finan, T.M. (2017). The Divided Bacterial Genome: Structure, Function and Evolution. *Microbiol. Mol. Biol. Rev.* 81(3), pii: e00019-17. doi: 10.1128/MMBR.00019-17.

Chao, M.C., Pritchard, J.R., Zhang, Y.J., Rubin, E.J., Livny, J., Davis, B.M., and Waldor, M.K. (2013). High-resolution definition of the *Vibrio cholerae* essential gene set with hidden Markov model–based analyses of transposon-insertion sequencing data. *Nucleic Acids Res.* 41(19), 9033–9048. doi: 10.1093/nar/gkt654.

Chao, M.C., Abel, S., Davis, B.M., and Waldor, M.K. (2016). The design and analysis of transposon insertion sequencing experiments. *Nat. Rev. Microbiol.* 14(2), 119–128. doi: 10.1038/nrmicro.2015.7.

Charles, H., Heddi, A., and Rahbe, Y. (2001). A putative insect intracellular endosymbiont stem clade, within the Enterobacteriaceae, infered from phylogenetic analysis based on a heterogeneous model of DNA evolution. *C. R. Acad. Sci. III* 324(5), 489–494.

Chaudhary, R., Atamian, H.S., Shen, Z., Briggs, S.P., and Kaloshian, I. (2014). GroEL from the endosymbiont *Buchnera aphidicola* betrays the aphid by triggering plant defense. *Proc. Natl. Acad. Sci. U.S.A.* 111(24), 8919–8924. doi: 10.1073/pnas.1407687111.

Cheong, W.-H., Tan, Y.-C., Yap, S.-J., and Ng, K.-P. (2015). ClicO FS: an interactive web-based service of Circos. *Bioinformatics* 31(22), 3685–3687. doi: 10.1093/bioinformatics/btv433.

Chikindas, M.L., Weeks, R., Drider, D., Chistyakov, V.A., and Dicks, L.M. (2018). Functions and emerging applications of bacteriocins. *Curr. Opin. Biotechnol.* 49, 23–28. doi: 10.1016/j.copbio.2017.07.011.

Cho, J., Carr, A.N., Whitworth, L., Johnson, B., and Wilson, K.S. (2017). MazEF toxin-antitoxin proteins alter *Escherichia coli* cell morphology and infrastructure during persister formation and regrowth. *Microbiology* 163(3), 308–321. doi: 10.1099/mic.0.000436.

Chow, J., Lee, S.M., Shen, Y., Khosravi, A., and Mazmanian, S.K. (2010). Host-bacterial symbiosis in health and disease. *Adv. Immunol.* 107, 243–274. doi: 10.1016/B978-0-12-381300-8.00008-3.

Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Coller, J.A., Fero, M.J., McAdams, H.H., and Shapiro, L. (2011). The essential genome of a bacterium. *Mol. Syst. Biol.* 7, 528–528. doi: 10.1038/msb.2011.58.

Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., Crabos, A., Auguy, F., Franche, C., Gherbi, H., *et al.* (2015). The Casuarina NIN gene is transcriptionally activated throughout Frankia root infection as well as in response to bacterial diffusible signals. *New Phytol.* 208(3), 887–903. doi: 10.1111/nph.13506.

C.M. Nowack, E. (2014). *Paulinella chromatophora* - Rethinking the transition from endosymbiont to organelle. *Acta Societatis Botanicorum Poloniae* 83(4). doi: 10.5586/asbp.2014.049.

Coenye, T., and Vandamme, P. (2003). Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ. Microbiol.* 5(9), 719–729.

Cole, B.J., Feltcher, M.E., Waters, R.J., Wetmore, K.M., Mucyn, T.S., Ryan, E.M., Wang, G., Ul-Hasan, S., McDonald, M., Yoshikuni, Y., *et al.* (2017). Genome-wide identification of bacterial plant colonization genes. *PloS Biol.* 15(9), e2002860. doi: 10.1371/journal.pbio.2002860.

Compant, S., Nowak, J., Coenye, T., Clement, C., and Barka, E.A. (2008). Diversity and occurrence of *Burkholderia* spp. in the natural environment. *FEMS Microbiol. Rev.* 32(4), 607–626. doi: 10.1111/j.1574-6976.2008.00113.x.

Corbin, B.D., Seeley, E.H., Raab, A., Feldmann, J., Miller, M.R., Torres, V.J., Anderson, K.L., Dattilo, B.M., Dunman, P.M., Gerads, R., *et al.* (2008). Metal chelation and inhibition of bacterial growth in tissue abscesses. *Science* 319(5865), 962–965. doi: 10.1126/science.1152449.

Crüsemann, M., Reher, R., Schamari, I., Brachmann, A.O., Ohbayashi, T., Kuschak, M., Malfacini, D., Seidinger, A., Pinto-Carbó, M., Richarz, R., *et al.* (2018). A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Angew Chem. Int. Ed. Engl.* 57(3), 836–840. doi: 10.1002/anie.201707996.

Cuccui, J., Milne, T.S., Harmer, N., George, A.J., Harding, S.V., Dean, R.E., Scott, A.E., Sarkar-Tyson, M., Wren, B.W., Titball, R.W., *et al.* (2012). Characterization of the *Burkholderia pseudomallei* K96243 Capsular Polysaccharide I Coding Region. *Infect. Immun.* 80(3), 1209–1221. doi: 10.1128/IAI.05805-11.

Cudic, M., Bulet, P., Hoffmann, R., Craik, D.J., and Otvos, L. (1999). Chemical synthesis, antibacterial activity and conformation of diptericin, an 82-mer peptide originally isolated from insects. *Eur. J. Biochem.* 266(2), 549–558. doi: 10.1046/j.1432-1327.1999.00894.x.

Cunningham, L., Pitt, M., and Williams, H.D. (1997). The *cioAB* genes from *Pseudomonas aeruginosa* code for a novel cyanide-insensitive terminal oxidase related to the cytochrome bd quinol oxidases. *Mol. Microbiol.* 24(3), 579–591. doi: 10.1046/j.1365-2958.1997.3561728.x.

Curtis, T.D., Takeuchi, I., Gram, L., and Knudsen, G.M. (2017). The Influence of the Toxin/Antitoxin mazEF on Growth and Survival of *Listeria monocytogenes* under Stress. *Toxins (Basel)* 9(1). pii: E31. doi: 10.3390/toxins9010031.

Cvejic, J.H., Putra, S.R., El-Beltagy, A., Hattori, R., Hattori, T., and Rohmer, M. (2000). Bacterial triterpenoids of the hopane series as biomarkers for the chemotaxonomy of *Burkholderia*, *Pseudomonas* and *Ralstonia* spp. *FEMS Microbiol. Lett.* 183(2), 295–299. doi: 10.1111/j.1574-6968.2000.tb08974.x.

Dalebroux, Z.D., and Miller, S.I. (2014). Salmonellae PhoPQ regulation of the outer membrane to resist innate immunity. *Curr. Opin. Microbiol.* 17, 106–113. doi: 10.1016/j.mib.2013.12.005.

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem. Sci.* 28(12), 655–662. doi: 10.1016/j.tibs.2003.10.007.

Dasch, G.A. (1984). Endosymbionts of insects. Bergey's Manual of Systematic Bacteriology 1.

De Gregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M., and Lemaitre, B. (2002). The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J.* 21(11), 2568–2579. doi: 10.1093/emboj/21.11.2568.

Deana, A., and Belasco, J.G. (2004). The function of RNase G in *Escherichia coli* is constrained by its amino and carboxyl termini. *Mol. Microbiol.* 51(4), 1205–1217. doi: 10.1046/j.1365-2958.2003.03905.x.

DeJesus, M.A., and Ioerger, T.R. (2013). A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinformatics* 14, 303. doi: 10.1186/1471-2105-14-303.

DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E., Boutte, C.C., Rubin, E.J., Schnappinger, D., Ehrt, S., Fortune, S.M., Sassetti, C.M., and Ioerger, T.R. (2017). Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *MBio* 8(1). pii: e02133-16. doi: 10.1128/mBio.02133-16.

Dilks, K., Rose, R.W., Hartmann, E., and Pohlschröder, M. (2003). Prokaryotic Utilization of the Twin-Arginine Translocation Pathway: a Genomic Survey. *J. Bacteriol.* 185(4), 1478–1483. doi: 10.1128/JB.185.4.1478-1483.2003

Divan Baldani, V.L., Baldani, J.I., and Döbereiner, J. (2000). Inoculation of rice plants with the endophytic diazotrophs *Herbaspirillum seropedicae* and *Burkholderia* spp. *Biol. Fertil. Soils* 30(5-6), 485–491. doi: 10.1007/s003740050027.

Dobritsa, A.P., and Samadpour, M. (2016). Transfer of eleven species of the genus *Burkholderia* to the genus *Paraburkholderia* and proposal of *Caballeronia* gen. nov. to accommodate twelve species of the genera *Burkholderia* and *Paraburkholderia*. *Int. J. Syst. Evol. Microbiol.* 66(8), 2836–2846. doi: 10.1099/ijsem.0.001065.

Doerrler, W.T., Sikdar, R., Kumar, S., and Boughner, L.A. (2013). New Functions for the Ancient DedA Membrane Protein Family. *J. Bacteriol.* 195(1), 3-11. doi: 10.1128/JB.01006-12.

Doughty, D.M., Coleman, M.L., Hunter, R.C., Sessions, A.L., Summons, R.E., and Newman, D.K. (2011). The RND-family transporter, HpnN, is required for hopanoid localization to the outer membrane of *Rhodopseudomonas palustris* TIE-1. *Proc. Natl. Acad. Sci. U.S.A.* 108(45), E1045-1051. doi: 10.1073/pnas.1104209108.

Douglas, A.E. (2011). Lessons from studying insect symbioses. *Cell Host Microbe* 10(4), 359–367. doi: 10.1016/j.chom.2011.09.001.

Dubarry, N., Pasta, F., and Lane, D. (2006). ParABS systems of the four replicons of *Burkholderia cenocepacia*: new chromosome centromeres confer partition specificity. *J. Bacteriol.* 188(4), 1486–1496. doi: 10.1128/JB.188.4.1489-1496.2006.

Eberl, L., and Vandamme, P. (2016). Members of the genus *Burkholderia*: good and bad guys. F1000Res 5, pii: F1000 Faculty Rev-1007. doi: 10.12688/f1000research.8221.1.

Echeverz, M., García, B., Sabalza, A., Valle, J., Gabaldón, T., Solano, C., and Lasa, I. (2017). Lack of the PGA exopolysaccharide in *Salmonella* as an adaptive trait for survival in the host. *PLoS Genet.* 13(5), e1006816. doi: 10.1371/journal.pgen.1006816.

Eichler, S., and Schaub, G.A. (2002). Development of Symbionts in Triatomine Bugs and the Effects of Infections with Trypanosomatids. *Exp. Parasitol.* 100(1), 17–27. doi: 10.1006/expr.2001.4653.

El-Halfawy, O.M., and Valvano, M.A. (2013). Chemical communication of antibiotic resistance by a highly resistant subpopulation of bacterial cells. *PLoS One* 8(7), e68874. doi: 10.1371/journal.pone.0068874.

Engel, P., and Moran, N.A. (2013). The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol. Rev.* 37(5), 699–735. doi: 10.1111/1574-6976.12025.

Ernst, R.K., Moskowitz, S.M., Emerson, J.C., Kraig, G.M., Adams, K.N., Harvey, M.D., Ramsey, B., Speert, D.P., Burns, J.L., and Miller, S.I. (2007). Unique Lipid A Modifications in *Pseudomonas aeruginosa* Isolated from the Airways of Patients with Cystic Fibrosis. *J. Infect. Dis.* 196(7), 1088–1092. doi: 10.1086/521367.

Esmaeel, Q., Miotto, L., Rondeau, M., Leclère, V., Clément, C., Jacquard, C., Sanchez, L., and Barka, E.A. (2018). *Paraburkholderia phytofirmans* PsJN-Plants Interaction: From Perception to the Induced Mechanisms. *Front. Microbiol.* 9, 2093. doi: 10.3389/fmicb.2018.02093.

Estrada-de Los Santos, P., Palmer, M., Chávez-Ramírez, B., Beukes, C., Steenkamp, E.T., Briscoe, L., Khan, N., Maluk, M., Lafos, M., Humm, E., *et al.* (2018). Whole Genome Analyses Suggests that *Burkholderia* sensu lato Contains Two Additional Novel Genera (*Mycetohabitans* gen. nov., and *Trinickia* gen. nov.): Implications for the Evolution of Diazotrophy and Nodulation in the Burkholderiaceae. *Genes (Basel)* 9(8), pii: E389. doi: 10.3390/genes9080389.

Fabisiak, A., Murawska, N., and Fichna, J. (2016). LL-37: Cathelicidin-related antimicrobial peptide with pleiotropic activity. *Pharmacol. Rep.* 68(4), 802–808. doi: 10.1016/j.pharep.2016.03.015.

Falanga, A., Galdiero, M., and Galdiero, S. (2015). Membranotropic Cell Penetrating Peptides: The Outstanding Journey. *Int. J. Mol. Sci.* 16(10), 25323–25337. doi: 10.3390/ijms161025323.

Fang, H., Kang, J., and Zhang, D. (2017). Microbial production of vitamin B12: a review and future perspectives. *Microb. Cell Fact.* 16, 15. doi: 10.1186/s12934-017-0631-y.

Farkas, A., Maróti, G., Kereszt, A., and Kondorosi, É. (2017). Comparative Analysis of the Bacterial Membrane Disruption Effect of Two Natural Plant Antimicrobial Peptides. *Front. Microbiol.* 8, 51. doi: 10.3389/fmicb.2017.00051.

Farkas, A., Pap, B., Kondorosi, É., and Maróti, G. (2018). Antimicrobial Activity of NCR Plant Peptides Strongly Depends on the Test Assays. *Front. Microbiol.* 9, 2600. doi: 10.3389/fmicb.2018.02600.

Fehlbaum, P., Bulet, P., Michaut, L., Lagueux, M., Broekaert, W.F., Hetru, C., and Hoffmann, J.A. (1994). Insect immunity. Septic injury of *Drosophila* induces the synthesis of a potent antifungal peptide with sequence homology to plant antifungal peptides. *J. Biol. Chem.* 269(52), 33159–33163.

Fischer, M., Falke, D., Naujoks, C., and Sawers, R.G. (2018). Cytochrome bd Oxidase Has an Important Role in Sustaining Growth and Development of *Streptomyces coelicolor* A3(2) under Oxygen-Limiting Conditions. *J. Bacteriol.* 200. pii: e00239-18. doi: 10.1128/JB.00239-18.

Fjell, C.D., Hiss, J.A., Hancock, R.E.W., and Schneider, G. (2011). Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.* 11(1), 37–51. doi: 10.1038/nrd3591.

Flannagan, R.S., and Valvano, M.A. (2008). *Burkholderia cenocepacia* requires RpoE for growth under stress conditions and delay of phagolysosomal fusion in macrophages. *Microbiology* 154(Pt 2), 643–653. doi: 10.1099/mic.0.2007/013714-0.

Flores-Kim, J., and Darwin, A.J. (2014). Regulation of bacterial virulence gene expression by cell envelope stress responses. *Virulence* 5(8), 835–851. doi: 10.4161/21505594.2014.965580.

Flórez, L.V., and Kaltenpoth, M. (2017). Symbiont dynamics and strain diversity in the defensive mutualism between *Lagria* beetles and *Burkholderia*. *Environ. Microbiol.* 19(9), 3674–3688. doi: 10.1111/1462-2920.13868.

Flórez, L.V., Scherlach, K., Gaube, P., Ross, C., Sitte, E., Hermes, C., Rodrigues, A., Hertweck, C., and Kaltenpoth, M. (2017). Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nat. Commun.* 8, 15172. doi: 10.1038/ncomms15172.

Franzenburg, S., Walter, J., Künzel, S., Wang, J., Baines, J.F., Bosch, T.C.G., and Fraune, S. (2013). Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proc. Natl. Acad. Sci. U.S.A.* 110(39), E3730-3738. doi: 10.1073/pnas.1304960110.

Fraune, S., Augustin, R., Anton-Erxleben, F., Wittlieb, J., Gelhaus, C., Klimovich, V.B., Samoilovich, M.P., and Bosch, T.C.G. (2010). In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proc. Natl. Acad. Sci. U.S.A.* 107(42), 18067–18072. doi: 10.1073/pnas.1008573107.

Froussart, E., Bonneau, J., Franche, C., and Bogusz, D. (2016). Recent advances in actinorhizal symbiosis signaling. *Plant Mol. Biol.* 90(6), 613–622. doi: 10.1007/s11103-016-0450-2.

Fu, Y., Waldor, M.K., and Mekalanos, J.J. (2013). Tn-Seq Analysis of *Vibrio cholerae* Intestinal Colonization Reveals a Role for T6SS-Mediated Antibacterial Activity in the Host. *Cell Host Microbe* 14(6), 652–663. doi: 10.1016/j.chom.2013.11.001.

Fukatsu, T., and Hosokawa, T. (2002). Capsule-transmitted gut symbiotic bacterium of the Japanese common plataspid stinkbug, *Megacopta punctatissima*. *Appl. Environ. Microbiol.* 68(1), 389–396. doi: 10.1128/aem.68.1.389-396.2002.

Furman, R., Danhart, E.M., NandyMazumdar, M., Yuan, C., Foster, M.P., and Artsimovitch, I. (2015). pH dependence of the stress regulator DksA. *PLoS One* 10(3), e0120746. doi: 10.1371/journal.pone.0120746.

Futahashi, R., Tanaka, K., Matsuura, Y., Tanahashi, M., Kikuchi, Y., and Fukatsu, T. (2011). Laccase2 is required for cuticular pigmentation in stinkbugs. *Insect Biochem. Mol. Biol.* 41(3), 191–196. doi: 10.1016/j.ibmb.2010.12.003.

Futahashi, R., Tanaka, K., Tanahashi, M., Nikoh, N., Kikuchi, Y., Lee, B.L., and Fukatsu, T. (2013). Gene Expression in Gut Symbiotic Organ of Stinkbug Affected by Extracellular Bacterial Symbiont. *PLoS One* 8(5), e64557. doi: 10.1371/journal.pone.0064557.

Gabbianelli, R., Scotti, R., Ammendola, S., Petrarca, P., Nicolini, L., and Battistoni, A. (2011). Role of ZnuABC and ZinT in *Escherichia coli* O157:H7 zinc acquisition and interaction with epithelial cells. *BMC Microbiol.* 11, 36. doi: 10.1186/1471-2180-11-36.

Gallagher, L.A., Shendure, J., and Manoil, C. (2011). Genome-Scale Identification of Resistance Functions in *Pseudomonas aeruginosa* Using Tn-seq. *MBio* 2(1). doi: 10.1128/mBio.00315-10.

Gao, B., Vorwerk, H., Huber, C., Lara-Tejero, M., Mohr, J., Goodman, A.L., Eisenreich, W., Galán, J.E., and Hofreuter, D. (2017). Metabolic and fitness determinants for *in vitro* growth and intestinal colonization of the bacterial pathogen *Campylobacter jejuni*. *PLoS Biol.* 15(5), e2001390. doi: 10.1371/journal.pbio.2001390.

Garcia, J.R., and Gerardo, N.M. (2014). The symbiont side of symbiosis: do microbes really benefit? *Front. Microbiol.* 5, 510 doi: 10.3389/fmicb.2014.00510.

Gaspar, J.A., Thomas, J.A., Marolda, C.L., and Valvano, M.A. (2000). Surface expression of O-specific lipopolysaccharide in *Escherichia coli* requires the function of the TolA protein. *Mol. Microbiol.* 38(2), 262–275. doi: 10.1046/j.1365-2958.2000.02094.x.

Gawronski, J.D., Wong, S.M.S., Giannoukos, G., Ward, D.V., and Akerley, B.J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc. Natl. Acad. Sci. U.S.A.* 106(38), 16422–16427. doi: 10.1073/pnas.0906627106.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., *et al.* (2003). Experimental Determination and System Level Analysis of Essential Genes in *Escherichia coli* MG1655. *J. Bacteriol.* 185(19), 5673–5684. doi: 10.1128/jb.185.19.5673-5684.2003.

Gerding, M.A., Ogata, Y., Pecora, N.D., Niki, H., and de Boer, P.A.J. (2007). The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli. Mol. Microbiol.* 63(4), 1008–1025. doi: 10.1111/j.1365-2958.2006.05571.x.

Ghimire, G.P., Thuan, N.H., Koirala, N., and Sohng, J.K. (2016). Advances in Biochemistry and Microbial Production of Squalene and Its Derivatives. *J. Microbiol. Biotechnol.* 26(3), 441–451. doi: 10.4014/jmb.1510.10039.

Gil, R., Silva, F.J., Pereto, J., and Moya, A. (2004). Determination of the Core of a Minimal Bacterial Gene Set. *Microbiol. Mol. Biol. Rev.* 68(3), 518–537. doi: 10.1128/MMBR.68.3.518-537.2004.

Gimenez, M.R., Chandra, G., Van Overvelt, P., Voulhoux, R., Bleves, S., and Ize, B. (2018). Genome wide identification and experimental validation of *Pseudomonas aeruginosa* Tat substrates. *Sci. Rep.* 8(1), 11950. doi: 10.1038/s41598-018-30393-x.

Glasgow, H. (1913). The gastric caeca and the caecal bacteria of Heteroptera. *Biol Bull.* 3, 101–171. doi: 10.2307/1536004.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O., and Venter, J.C. (2006). Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 103(2), 425–430. doi: 10.1073/pnas.0510013103.

Gobbo, M., Biondi, L., Filira, F., Gennaro, R., Benincasa, M., Scolaro, B., and Rocchi, R. (2002). Antimicrobial peptides: synthesis and antibacterial activity of linear and cyclic drosocin and apidaecin 1b analogues. *J. Med. Chem.* 45(20), 4494–4504. doi: 10.1021/jm020861d.

Godbey, W.T. (2014). Chapter 16 - Fermentation, Beer, and Biofuels. In An Introduction to Biotechnology, W.T. Godbey, ed. (Woodhead Publishing), pp. 331–351.

Godlewska, R., Wiśniewska, K., Pietras, Z., and Jagusztyn-Krynicka, E.K. (2009). Peptidoglycan-associated lipoprotein (Pal) of Gram-negative bacteria: function, structure, role in pathogenesis and potential application in immunoprophylaxis. *FEMS Microbiol. Lett*. 298(1), 1–11. doi: 10.1111/j.1574-6968.2009.01659.x.

Gong, S., Ma, Z., and Foster, J.W. (2004). The Era-like GTPase TrmE conditionally activates gadE and glutamate-dependent acid resistance in *Escherichia coli*. *Mol. Microbiol.* 54(4), 948–961. doi: 10.1111/j.1365-2958.2004.04312.x.

Gonzalez-Mula, A., Lachat, J., Mathias, L., Naquin, D., Lamouche, F., Mergaert, P., and Faure, D. (2018). The biotroph *Agrobacterium tumefaciens* thrives in tumors by exploiting a wide spectrum of plant host metabolites. *New Phytol.* 222(1), 455–467. doi: 10.1111/nph.15598.

Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host Microbe* 6(3), 279–289. doi: 10.1016/j.chom.2009.08.003.

Goormachtig, S., Capoen, W., and Holsters, M. (2004). *Rhizobium* infection: lessons from the versatile nodulation behaviour of water-tolerant legumes. *Trends Plant Sci.* 9(11), 518–522. doi: 10.1016/j.tplants.2004.09.005.

Gordon, T.L., Haseeb, M., Kanga, L.H.B., and Legaspi, J.C. (2017). Potential of Three Trap Crops in Managing *Nezara viridula* (Hemiptera: Pentatomidae) on Tomatoes in Florida. *J. Econ. Entomol.* 110(6), 2478–2482. doi: 10.1093/jee/tox267.

Gottar, M., Gobert, V., Michel, T., Belvin, M., Duyk, G., Hoffmann, J.A., Ferrandon, D., and Royet, J. (2002). The Drosophila immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature* 416(6881), 640–644. doi: 10.1038/nature734.

Graf, M., Mardirossian, M., Nguyen, F., Seefeldt, A.C., Guichard, G., Scocchi, M., Innis, C.A., and Wilson, D.N. (2017). Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.* 34(7), 702–711. doi: 10.1039/c7np00020k.

Grazziotin, A.L., Vidal, N.M., and Venancio, T.M. (2015). Uncovering major genomic features of essential genes in Bacteria and a methanogenic Archaea. *FEBS J.* 282(17), 3395–3411. doi: 10.1111/febs.13350.

Green, E.R., Clark, S., Crimmins, G.T., Mack, M., Kumamoto, C.A., and Mecsas, J. (2016). Fis Is Essential for *Yersinia pseudotuberculosis* Virulence and Protects against Reactive Oxygen Species Produced by Phagocytic Cells during Infection. *PLoS Pathog.* 12(9), e1005898. doi: 10.1371/journal.ppat.1005898.

Greenfield, L.K., and Whitfield, C. (2012). Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr. Res.* 356, 12–24. doi: 10.1016/j.carres.2012.02.027.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., and Sassetti, C.M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* 7(9), e1002251. doi: 10.1371/journal.ppat.1002251.

Guefrachi, I., Pierre, O., Timchenko, T., Alunni, B., Barrière, Q., Czernic, P., Villaécija-Aguilar, J.-A., Verly, C., Bourge, M., Fardoux, J., *et al.* (2015). *Bradyrhizobium* BclA Is a Peptide Transporter Required for Bacterial Differentiation in Symbiosis with Aeschynomene Legumes. *Mol. Plant Microbe Interact.* 28(11), 1155–1166. doi: 10.1094/MPMI-04-15-0094-R.

Guerrero-Ferreira, R.C., and Nishiguchi, M.K. (2009). ULTRASTRUCTURE OF LIGHT ORGANS OF LOLIGINID SQUIDS AND THEIR BACTERIAL SYMBIONTS: A NOVEL MODEL SYSTEM FOR THE STUDY OF MARINE SYMBIOSES. *Vie Milieu Paris* 59(3-4), 307–313.

Guest, R.L., and Raivio, T.L. (2016). Role of the Gram-Negative Envelope Stress Response in the Presence of Antimicrobial Agents. *Trends Microbiol.* 24(5), 377–390. doi: 10.1016/j.tim.2016.03.001.

Gutgsell, N.S., Deutscher, M.P., and Ofengand, J. (2005). The pseudouridine synthase RluD is required for normal ribosome assembly and function in *Escherichia coli*. *RNA* 11(7), 1141–1152. doi: 10.1261/rna.2550105.

Gutierrez, M.G., Yoder-Himes, D.R., and Warawa, J.M. (2015). Comprehensive identification of virulence factors required for respiratory melioidosis using Tn-seq mutagenesis. *Front. Cell. Infect. Microbiol.* 5, 78. doi: 10.3389/fcimb.2015.00078.

Gyaneshwar, P., Hirsch, A.M., Moulin, L., Chen, W.-M., Elliott, G.N., Bontemps, C., Estrada-de los Santos, P., Gross, E., dos Reis, F.B., Sprent, J.I., Young, J.P., and James, E.K. (2011). Legume-Nodulating Betaproteobacteria: Diversity, Host Range, and Future Prospects. *Mol. Plant Microbe Interact.* 24(11), 1276–1288. doi: 10.1094/MPMI-06-11-0172.

Gyllborg, M.C., Sahl, J.W., Cronin, D.C., Rasko, D.A., and Mandel, M.J. (2012). Draft Genome Sequence of *Vibrio fischeri* SR5, a Strain Isolated from the Light Organ of the Mediterranean Squid *Sepiola robusta*. *J. Bacteriol.* 194(6), 1639–1639. doi: 10.1128/JB.06825-11.

Haag, A.F., Baloban, M., Sani, M., Kerscher, B., Pierre, O., Farkas, A., Longhi, R., Boncompagni, E., Hérouart, D., Dall'angelo, S., *et al.* (2011). Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biol.* 9(10), e1001169. doi: 10.1371/journal.pbio.1001169.

Hadjifrangiskou, M., Gu, A.P., Pinkner, J.S., Kostakioti, M., Zhang, E.W., Greene, S.E., and Hultgren, S.J. (2012). Transposon mutagenesis identifies uropathogenic *Escherichia coli* biofilm factors. *J. Bacteriol.* 194(22), 6195–6205. doi: 10.1128/JB.01012-12.

Hamidi, M. Mosavi Nasab, S.D., Ahmadi, N., Basati, G., Avlad, G.R., Zargar, M. (2013). Synthesis of antimicrobial peptides in bacteria. *Journal of Ilam University of Medical Sciences* 20(5), 149-157

Hamma, T., and Ferré-D'Amaré, A.R. (2006). Pseudouridine Synthases. *Chem. Biol.* 13(11), 1125–1135. doi: 10.1016/j.chembiol.2006.09.009.

Han, X., Dorsey-Oresto, A., Malik, M., Wang, J.-Y., Drlica, K., Zhao, X., and Lu, T. (2010). *Escherichia coli* genes that reduce the lethal effects of stress. *BMC Microbiol.* 10, 35. doi: 10.1186/1471-2180-10-35.

Hansen, A.K., and Moran, N.A. (2011). Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. U.S.A.* 108(7), 2849–2854. doi: 10.1073/pnas.1013465108.

Hanson, M.A., Dostálová, A., Ceroni, C., Poidevin, M., Kondo, S., and Lemaitre, B. (2019). Synergy and remarkable specificity of antimicrobial peptides in vivo using a systematic knockout approach. *ELife* 8, e44341. pii: e44341. doi: 10.7554/eLife.44341.

Harrison, P.W., Lower, R.P., Kim, N.K., and Young, J.P. (2010). Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol.* 18(4), 141–148. doi: 10.1016/j.tim.2009.12.010.

Härtner, T., Straub, K.L., and Kannenberg, E. (2005). Occurrence of hopanoid lipids in anaerobic *Geobacter* species. *FEMS Microbiol. Lett.* 243(1), 59–64. doi: 10.1016/j.femsle.2004.11.039.

Heath-Heckman, E.A.C., Gillette, A.A., Augustin, R., Gillette, M.X., Goldman, W.E., and McFall-Ngai, M.J. (2014). Shaping the microenvironment: evidence for the influence of a host galaxin on symbiont acquisition and maintenance in the squid-*Vibrio* symbiosis. *Environ. Microbiol.* 16(12), 3669–3682. doi: 10.1111/1462-2920.12496.

Hedblom, M.L., and Adler, J. (1980). Genetic and biochemical properties of *Escherichia coli* mutants with defects in serine chemotaxis. *J. Bacteriol.* 144(3), 1048–1060.

Heddi, A., Charles, H., Khatchadourian, C., Bonnot, G., and Nardon, P. (1998). Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G + C content of an endocytobiotic DNA. *J. Mol. Evol.* 47(1), 52–61.

Heidrich, C., Ursinus, A., Berger, J., Schwarz, H., and Höltje, J.-V. (2002). Effects of Multiple Deletions of Murein Hydrolases on Viability, Septum Cleavage, and Sensitivity to Large Toxic Molecules in *Escherichia coli*. *J. Bacteriol.* 184(22), 6093–6099. doi: 10.1128/jb.184.22.6093-6099.2002.

Hemarajata, P., Baghdadi, J.D., Hoffman, R., and Humphries, R.M. (2016). *Burkholderia pseudomallei*: Challenges for the Clinical Microbiology Laboratory. *J. Clin. Microbiol.* 54(12), 2866–2873. doi: 10.1128/JCM.01636-16.

Henry, T.J. (1997). Phylogenetic Analysis of Family Groups within the Infraorder Pentatomomorpha (Hemiptera: Heteroptera), with Emphasis on the Lygaeoidea. *Ann. Entomol. Soc. Am.* 90(3), 275–301. doi: 10.1093/aesa/90.3.275.

Higgins, S., Sanchez-Contreras, M., Gualdi, S., Pinto-Carbó, M., Carlier, A., and Eberl, L. (2017). The Essential Genome of *Burkholderia cenocepacia* H111. *J. Bacteriol.* 199(22), pii: e00260-17. doi: 10.1128/JB.00260-17.

Hoffmann, J.A., and Hetru, C. (1992). Insect defensins: inducible antibacterial peptides. *Immunol. Today* 13(10), 411–415. doi: 10.1016/0167-5699(92)90092-L.

Hofmann, K. (2000). A superfamily of membrane-bound O-acyltransferases with implications for *wnt* signaling. *Trends Biochem. Sci.* 25(3), 111–112.

Hooper, L.V., and Gordon, J.I. (2001). Commensal host-bacterial relationships in the gut. *Science* 292(5519), 1115–1118. doi: 10.1126/science.1058709.

Hooven, T.A., Catomeris, A.J., Akabas, L.H., Randis, T.M., Maskell, D.J., Peters, S.E., Ott, S., Santana-Cruz, I., Tallon, L.J., Tettelin, H., *et al.* (2016). The essential genome of *Streptococcus agalactiae*. *BMC Genomics* 17, 406. doi: 10.1186/s12864-016-2741-z.

Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M., and Fukatsu, T. (2006). Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol.* 4(10), e337. doi: 10.1371/journal.pbio.0040337.

Hosokawa, T., Kikuchi, Y., and Fukatsu, T. (2007). How many symbionts are provided by mothers, acquired by offspring, and needed for successful vertical transmission in an obligate insect-bacterium mutualism? *Mol. Ecol.* 16(24), 5316–5325. doi: 10.1111/j.1365-294X.2007.03592.x.

Hosokawa, T., Kaiwa, N., Matsuura, Y., Kikuchi, Y., and Fukatsu, T. (2015). Infection prevalence of *Sodalis* symbionts among stinkbugs. *Zoological Lett.* 1, 5. doi: 10.1186/s40851-014-0009-5.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., and Bu, W. (2008). Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). *BMC Genomics* 9, 610. doi: 10.1186/1471-2164-9-610.

Hultmark, D., Engström, A., Andersson, K., Steiner, H., Bennich, H., and Boman, H.G. (1983). Insect immunity. Attacins, a family of antibacterial proteins from *Hyalophora cecropia. EMBO J.* 2(4), 571–576.

Itoh, H., Navarro, R., Takeshita, K., Tago, K., Hayatsu, M., Hori, T., and Kikuchi, Y. (2014). Bacterial population succession and adaptation affected by insecticide application and soil spraying history. *Front. Microbiol.* 5, 547. doi: 10.3389/fmicb.2014.00457.

Ize, B., Stanley, N.R., Buchanan, G., and Palmer, T. (2003). Role of the *Escherichia coli* Tat pathway in outer membrane integrity. *Mol. Microbiol.* 48(5), 1183–1193. doi: 10.1046/j.1365-2958.2003.03504.x.

Izquierdo, L., Abitiu, N., Coderch, N., Hita, B., Merino, S., Gavin, R., Tomás, J.M., and Regué, M. (2002). The inner-core lipopolysaccharide biosynthetic *waaE* gene: function and genetic distribution among some Enterobacteriaceae. *Microbiology* 148(Pt 11), 3485–3496. doi: 10.1099/00221287-148-11-3485

Jack, D.L., Yang, N.M., and Saier, M.H. (2001). The drug/metabolite transporter superfamily. *Eur. J. Biochem.* 268(13), 3620–3639. doi: 10.1046/j.1432-1327.2001.02265.x.

Jakes, K.S., and Cramer, W.A. (2012). Border crossings: colicins and transporters. *Annu. Rev. Genet.* 46, 209–231. doi: 10.1146/annurev-genet-110711-155427.

Jang, S.H., Jang, H.A., Lee, J., Kim, J.U., Lee, S.A., Park, K.-E., Kim, B.H., Jo, Y.H., and Lee, B.L. (2017). PhaR, a Negative Regulator of PhaP, Modulates the Colonization of a *Burkholderia* Gut Symbiont in the Midgut of the Host Insect, *Riptortus pedestris. Appl. Environ. Microbiol.* 83(11), pii: e00459-17.. doi: 10.1128/AEM.00459-17.

Jarmoskaite, I., and Russell, R. (2011). DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdiscip. Rev. RNA* 2(1), 135–152. doi: 10.1002/wrna.50.

Johnson, D.C., Dean, D.R., Smith, A.D., and Johnson, M.K. (2005). Structure, function, and formation of biological iron-sulfur clusters. *Annu. Rev. Biochem.* 74, 247–281. doi: 10.1146/annurev.biochem.74.082803.133518.

Joseph, T.C., Rajan, L.A., Thampuran, N., and James, R. (2010). Functional characterization of trehalose biosynthesis genes from *E. coli*: an osmolyte involved in stress tolerance. *Mol. Biotechnol.* 46(1), 20–25. doi: 10.1007/s12033-010-9259-4.

Kaiwa, N., Hosokawa, T., Kikuchi, Y., Nikoh, N., Meng, X.Y., Kimura, N., Ito, M., and Fukatsu, T. (2011). Bacterial symbionts of the giant jewel stinkbug *Eucorysses grandis* (Hemiptera: Scutelleridae). *Zool. Sci.* 28(3), 169–174. doi: 10.2108/zsj.28.169.

Kaiwa, N., Hosokawa, T., Nikoh, N., Tanahashi, M., Moriyama, M., Meng, X.-Y., Maeda, T., Yamaguchi, K., Shigenobu, S., Ito, M., and Fukatsu, T. (2014). Symbiont-supplemented maternal investment underpinning host's ecological adaptation. *Curr. Biol.* 24(20), 2465–2470. doi: 10.1016/j.cub.2014.08.065.

Kamp, H.D., Patimalla-Dipali, B., Lazinski, D.W., Wallace-Gadsden, F., and Camilli, A. (2013). Gene Fitness Landscapes of *Vibrio cholerae* at Important Stages of Its Life Cycle. *PLoS Pathog.* 9(12), e1003800. doi: 10.1371/journal.ppat.1003800.

Kannenberg, E.L., and Poralla, K. (1999). Hopanoid Biosynthesis and Function in Bacteria. *Naturwissenschaften* 86(4), 168–176. doi: 10.1007/s001140050592.

Kanost, M.R. (2009). Chapter 117 - Hemolymph. In Encyclopedia of Insects (Second Edition), V.H. Resh, and R.T. Cardé, eds. (San Diego: Academic Press), pp. 446–449.

Kauffman, W.B., Fuselier, T., He, J., and Wimley, W.C. (2015). Mechanism Matters: A Taxonomy of Cell Penetrating Peptides. *Trends Biochem. Sci*. 40(12), 749–764. doi: 10.1016/j.tibs.2015.10.004.

Kehl-Fie, T.E., and Skaar, E.P. (2010). Nutritional immunity beyond iron: a role for manganese and zinc. *Curr. Opin. Chem. Biol.* 14(2), 218–224. doi: 10.1016/j.cbpa.2009.11.008.

Kešnerová, L., Mars, R.A.T., Ellegaard, K.M., Troilo, M., Sauer, U., and Engel, P. (2017). Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS Biol.* 15(12), e2003467. doi: 10.1371/journal.pbio.2003467.

de Kievit, T.R., and Lam, J.S. (1997). Isolation and characterization of two genes, *waaC* (*rfaC*) and *waaF* (*rfaF*), involved in *Pseudomonas aeruginosa* serotype O5 inner-core biosynthesis. *J. Bacteriol.* 179(11), 3451–3457. doi: 10.1128/jb.179.11.3451-3457.1997.

Kikuchi, Y. (2009). Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes Environ.* 24(3), 195–204. doi: 10.1264/jsme2.me09140s.

Kikuchi, Y., and Fukatsu, T. (2014). Live imaging of symbiosis: spatiotemporal infection dynamics of a GFP-labelled *Burkholderia* symbiont in the bean bug *Riptortus pedestris*. *Mol. Ecol.* 23(6), 1445–1456. doi: 10.1111/mec.12479.

Kikuchi, Y., and Graf, J. (2007). Spatial and temporal population dynamics of a naturally occurring two-species microbial community inside the digestive tract of the medicinal leech. *Appl. Environ. Microbiol.* 73(6), 1984–1991. doi: 10.1128/AEM.01833-06.

Kikuchi, Y., and Yumoto, I. (2013). Efficient Colonization of the Bean Bug *Riptortus pedestris* by an Environmentally Transmitted *Burkholderia* Symbiont. *Appl. Environ. Microbiol.* 79(6), 2088–2091. doi: 10.1128/AEM.03299-12.

Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2007). Insect-Microbe Mutualism without Vertical Transmission: a Stinkbug Acquires a Beneficial Gut Symbiont from the Environment Every Generation. *Appl. Environ. Microbiol.* 73(13), 4308–4316. doi: 10.1128/AEM.00067-07.

Kikuchi, Y., Hosokawa, T., Nikoh, N., Meng, X.-Y., Kamagata, Y., and Fukatsu, T. (2009). Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol.* 7, 2. doi: 10.1186/1741-7007-7-2.

Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2011a). Specific developmental window for establishment of an insect-microbe gut symbiosis. *Appl. Environ. Microbiol.* 77(12), 4075–4081. doi: 10.1128/AEM.00358-11.

Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2011b). An ancient but promiscuous host-symbiont association between *Burkholderia* gut symbionts and their heteropteran hosts. *ISME J.* 5(3), 446–460. doi: 10.1038/ismej.2010.

Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., and Fukatsu, T. (2012). Symbiont-mediated insecticide resistance. *Proc. Natl. Acad. Sci. U.S.A.* 109(22), 8618–8622. doi: 10.1073/pnas.1200231109.

Kim, J.K., and Lee, B.L. (2015). Symbiotic factors in *Burkholderia* essential for establishing an association with the bean bug, *Riptortus pedestris. Arch. Insect Biochem. Physiol.* 88(1), 4–17. doi: 10.1002/arch.21218.

Kim, J.K., Kim, N.H., Am Jang, H., Kikuchi, Y., Kim, C.-H., Fukatsu, T., and Lee, B.L. (2013a). Specific midgut region controlling the symbiont population in an insect-microbe gut symbiotic association. *Appl. Environ. Microbiol.* 79(23), 7229–7233. doi: 10.1128/AEM.02152-13.

Kim, J.K., Won, Y.J., Nikoh, N., Nakayama, H., Han, S.H., Kikuchi, Y., Rhee, Y.H., Park, H.Y., Kwon, J.Y., Kurokawa, K., *et al.* (2013b). Polyester synthesis genes associated with stress resistance are involved in an insect-bacterium symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 110(26), E2381-2389. doi: 10.1073/pnas.1303228110.

Kim, J.K., Lee, H.J., Kikuchi, Y., Kitagawa, W., Nikoh, N., Fukatsu, T., and Lee, B.L. (2013c). Bacterial cell wall synthesis gene *uppP* is required for *Burkholderia* colonization of the Stinkbug Gut. *Appl. Environ. Microbiol.* 79(16), 4879–4886. doi: 10.1128/AEM.01269-13.

Kim, J.K., Jang, H.A., Won, Y.J., Kikuchi, Y., Han, S.H., Kim, C.-H., Nikoh, N., Fukatsu, T., and Lee, B.L. (2014a). Purine biosynthesis-deficient *Burkholderia* mutants are incapable of symbiotic accommodation in the stinkbug. *ISME J.* 8(3), 552–563. doi: 10.1038/ismej.2013.168.

Kim, J.K., Han, S.H., Kim, C.-H., Jo, Y.H., Futahashi, R., Kikuchi, Y., Fukatsu, T., and Lee, B.L. (2014b). Molting-associated suppression of symbiont population and up-regulation of antimicrobial activity in the midgut symbiotic organ of the *Riptortus-Burkholderia* symbiosis. *Dev. Comp. Immunol.* 43(1), 10–14. doi: 10.1016/j.dci.2013.10.010.

Kim, J.K., Son, D.W., Kim, C.-H., Cho, J.H., Marchetti, R., Silipo, A., Sturiale, L., Park, H.Y., Huh, Y.R., Nakayama, H., *et al.* (2015a). Insect Gut Symbiont Susceptibility to Host Antimicrobial Peptides Caused by Alteration of the Bacterial Cell Envelope. *J. Biol. Chem.* 290(34), 21042–21053. doi: 10.1074/jbc.M115.651158.

Kim, J.K., Lee, J.B., Huh, Y.R., Jang, H.A., Kim, C.-H., Yoo, J.W., and Lee, B.L. (2015b). *Burkholderia* gut symbionts enhance the innate immunity of host *Riptortus pedestris. Dev. Comp. Immunol.* 53(1), 265–269. doi: 10.1016/j.dci.2015.07.006.

Kim, J.K., Lee, J.B., Jang, H.A., Han, Y.S., Fukatsu, T., and Lee, B.L. (2016a). Understanding regulation of the host-mediated gut symbiont population and the symbiont-mediated host immunity in the *Riptortus-Burkholderia* symbiosis system. *Dev. Comp. Immunol.* 64, 75–81. doi: 10.1016/j.dci.2016.01.005.

Kim, J.K., Park, H.Y., and Lee, B.L. (2016b). The symbiotic role of O-antigen of *Burkholderia* symbiont in association with host *Riptortus pedestris*. *Dev. Comp. Immunol.* 60, 202–208. doi: 10.1016/j.dci.2016.02.009.

Kim, J.K., Jang, H.A., Kim, M.S., Cho, J.H., Lee, J., Di Lorenzo, F., Sturiale, L., Silipo, A., Molinaro, A., and Lee, B.L. (2017). The lipopolysaccharide core oligosaccharide of *Burkholderia* plays a critical role in maintaining a proper gut symbiosis with the bean bug *Riptortus pedestris. J. Biol. Chem.* 292(47), 19226–19237. doi: 10.1074/jbc.M117.813832.

Kim, K., Lee, S., Lee, K., and Lim, D. (1998). Isolation and characterization of toluene-sensitive mutants from the toluene-resistant bacterium *Pseudomonas putida* GM73. *J. Bacteriol.* 180(14), 3692–3696.

Kim, K.-D., Ahn, J.-H., Kim, T., Park, S.C., Seong, C.N., Song, H.-G., and Ka, J.-O. (2009). Genetic and phenotypic diversity of fenitrothion-degrading bacteria isolated from soils. *J. Microbiol. Biotechnol.* 19(2), 113–120.

Klein, B.A., Tenorio, E.L., Lazinski, D.W., Camilli, A., Duncan, M.J., and Hu, L.T. (2012). Identification of essential genes of the periodontal pathogen *Porphyromonas gingivalis*. *BMC Genomics* 13, 578. doi: 10.1186/1471-2164-13-578.

Kneidinger, B., Marolda, C., Graninger, M., Zamyatina, A., McArthur, F., Kosma, P., Valvano, M.A., and Messner, P. (2002). Biosynthesis pathway of ADP-L-glycero-beta-D-manno-heptose in *Escherichia coli. J. Bacteriol.* 184(2), 363–369. doi: 10.1128/jb.184.2.363-369.2002.

Koga, R., Meng, X.-Y., Tsuchida, T., and Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc. Natl. Acad. Sci. U.S.A.* 109(20), E1230-1237. doi: 10.1073/pnas.1119212109.

Koh, Y.S., and Roe, J.H. (1995). Isolation of a novel paraquat-inducible (*pqi*) gene regulated by the soxRS locus in *Escherichia coli. J. Bacteriol.* 177(10), 2673–2678. doi: 10.1128/jb.177.10.2673-2678.1995.

Koh, Y.S., and Roe, J.H. (1996). Dual regulation of the paraquat-inducible gene *pqi-5* by SoxS and RpoS in *Escherichia coli. Mol. Microbiol.* 22(1), 53–61. doi: 10.1111/j.1365-2958.1996.tb02655.x.

de Kok, A., Hengeveld, A.F., Martin, A., and Westphal, A.H. (1998). The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria. *Biochim. Biophys. Acta* 1385(2), 353–366. doi: 10.1016/s0167-4838(98)00079-x.

Kondorosi, E., Mergaert, P., and Kereszt, A. (2013). A Paradigm for Endosymbiotic Life: Cell Differentiation of *Rhizobium* Bacteria Provoked by Host Plant Factors. *Annu. Rev. Microbiol.* 67, 611–628. doi: 10.1146/annurev-micro-092412-155630.

Korbsrisate, S., Vanaporn, M., Kerdsuk, P., Kespichayawattana, W., Vattanaviboon, P., Kiatpapan, P., and Lertmemongkolchai, G. (2005). The *Burkholderia pseudomallei* RpoE (AlgU) operon is involved in environmental stress tolerance and biofilm formation. *FEMS Microbiol. Lett.* 252(2), 243–249. doi: 10.1016/j.femsle.2005.09.002

Krishnakumar, V., Kim, M., Rosen, B.D., Karamycheva, S., Bidwell, S.L., Tang, H., and Town, C.D. (2015). MTGD: The *Medicago truncatula* genome database. *Plant Cell Physiol.* 56(1), e1. doi: 10.1093/pcp/pcu179.

Küchler, S.M., Dettner, K., and Kehl, S. (2010). Molecular characterization and localization of the obligate endosymbiotic bacterium in the birch catkin bug *Kleidocerys resedae* (Heteroptera: Lygaeidae, Ischnorhynchinae). *FEMS Microbiol. Ecol.* 73(2), 408–418. doi: 10.1111/j.1574-6941.2010.00890.x

Kuechler, S.M., Dettner, K., and Kehl, S. (2011). Characterization of an obligate intracellular bacterium in the midgut epithelium of the bulrush bug *Chilacis typhae* (Heteroptera, Lygaeidae, Artheneinae). *Appl. Environ. Microbiol.* 77(9), 2869–2876. doi: 10.1128/AEM.02983-10.

Kuechler, S.M., Renz, P., Dettner, K., and Kehl, S. (2012). Diversity of symbiotic organs and bacterial endosymbionts of lygaeoid bugs of the families blissidae and lygaeidae (hemiptera: heteroptera: lygaeoidea). *Appl. Environ. Microbiol.* 78(8), 2648–2659. doi: 10.1128/AEM.07191-11.

Kulkarni, G., Busset, N., Molinaro, A., Gargani, D., Chaintreuil, C., Silipo, A., Giraud, E., and Newman, D.K. (2015). Specific hopanoid classes differentially affect free-living and symbiotic states of *Bradyrhizobium diazoefficiens*. *MBio* 6(5), e01251-1215. doi: 10.1128/mBio.01251-15.

Kumar, S., and Doerrler, W.T. (2014). Members of the conserved DedA family are likely membrane transporters and are required for drug resistance in *Escherichia coli. Antimicrob. Agents Chemother.* 58(2), 923–930. doi: 10.1128/AAC.02238-13.

Kumar, P., Kizhakkedathu, J.N., and Straus, S.K. (2018). Antimicrobial Peptides: Diversity, Mechanism of Action and Strategies to Improve the Activity and Biocompatibility *In Vivo*. *Biomolecules* 8(1), pii: E4. doi: 10.3390/biom8010004.

Lai, Y., and Gallo, R.L. (2009). AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends Immunol.* 30(3), 131–141. doi: 10.1016/j.it.2008.12.003.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3), R25. doi: 10.1186/gb-2009-10-3-r25.

Langridge, G.C., Phan, M.-D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., *et al.* (2009). Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res.* 19(12), 2308–2316. doi: 10.1101/gr.097097.109.

Larue, K., Ford, R.C., Willis, L.M., and Whitfield, C. (2011). Functional and structural characterization of polysaccharide co-polymerase proteins required for polymer export in ATP-binding cassette transporter-dependent capsule biosynthesis pathways. *J. Biol. Chem.* 286(19), 16658–16668. doi: 10.1074/jbc.M111.228221.

Layer, G., Ollagnier-de Choudens, S., Sanakis, Y., and Fontecave, M. (2006). Iron-sulfur cluster biosynthesis: characterization of *Escherichia coli* CYaY as an iron donor for the assembly of [2Fe-2S] clusters in the scaffold IscU. *J. Biol. Chem.* 281(24), 16256–16263. doi: 10.1074/jbc.M513569200.

Lazzaroni, J.C., Germon, P., Ray, M.C., and Vianney, A. (1999). The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability. *FEMS Microbiol. Lett.* 177(2), 191–197. doi: 10.1111/j.1574-6968.1999.tb13731.x.

Lazzaroni, J.-C., Dubuisson, J.-F., and Vianney, A. (2002). The Tol proteins of *Escherichia coli* and their involvement in the translocation of group A colicins. *Biochimie* 84(5-6), 391–397.

Le Brun, N.E., Bengtsson, J., and Hederstedt, L. (2000). Genes required for cytochrome c synthesis in *Bacillus subtilis. Mol. Microbiol.* 36(3), 638–650. doi: 10.1046/j.1365-2958.2000.01883.x.

Lee, D.J., Lee, J.B., Jang, H.A., Ferrandon, D., and Lee, B.L. (2017a). An antimicrobial protein of the *Riptortus pedestris* salivary gland was cleaved by a virulence factor of *Serratia marcescens. Dev. Comp. Immunol.* 67, 427–433. doi: 10.1016/j.dci.2016.08.009.

Lee, J.B., Byeon, J.H., Jang, H.A., Kim, J.K., Yoo, J.W., Kikuchi, Y., and Lee, B.L. (2015a). Bacterial cell motility of *Burkholderia* gut symbiont is required to colonize the insect gut. *FEBS Lett.* 589(19 Pt B), 2784–2790. doi: 10.1016/j.febslet.2015.08.022.

Lee, J.B., Park, K.-E., Lee, S.A., Jang, S.H., Eo, H.J., Jang, H.A., Kim, C.-H., Ohbayashi, T., Matsuura, Y., Kikuchi, Y., *et al.* (2017b). Gut symbiotic bacteria stimulate insect growth and egg production by modulating hexamerin and vitellogenin gene expression. *Dev. Comp. Immunol.* 69, 12–22. doi: 10.1016/j.dci.2016.11.019.

Lee, S.A., Gallagher, L.A., Thongdee, M., Staudinger, B.J., Lippman, S., Singh, P.K., and Manoil, C. (2015b). General and condition-specific essential functions of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.* 112(16), 5189–5194. doi: 10.1073/pnas.1422186112.

Lee, T.-H., Hall, K.N., and Aguilar, M.-I. (2016). Antimicrobial Peptide Structure and Mechanism of Action: A Focus on the Role of Membrane Structure. *Curr. Top. Med. Chem.* 16(1), 25–39.

Lemaire, B., Vandamme, P., Merckx, V., Smets, E., and Dessein, S. (2011). Bacterial leaf symbiosis in angiosperms: host specificity without co-speciation. *PLoS One* 6(9), e24430. doi: 10.1371/journal.pone.0024430.

Lemaire, B., Lachenaud, O., Persson, C., Smets, E., and Dessein, S. (2012). Screening for leaf-associated endophytes in the genus Psychotria (Rubiaceae). *FEMS Microbiol. Ecol.* 81(2), 364–372. doi: 10.1111/j.1574-6941.2012.01356.x.

Lemaire, B., Van Cauwenberghe, J., Verstraete, B., Chimphango, S., Stirton, C., Honnay, O., Smets, E., Sprent, J., James, E.K., and Muasya, A.M. (2016). Characterization of the papilionoid-*Burkholderia* interaction in the Fynbos biome: The diversity and distribution of beta-rhizobia nodulating *Podalyria calyptrata* (Fabaceae, Podalyrieae). *Syst. Appl. Microbiol.* 39(1), 41–48. doi: 10.1016/j.syapm.2015.09.006.

Lemaitre, B., and Hoffmann, J. (2007). The Host Defense of *Drosophila melanogaster. Annu. Rev. Immunol*. 25, 697–743. doi: 10.1146/annurev.immunol.25.022106.141615.

Levashina, E.A., Ohresser, S., Bulet, P., Reichhart, J.M., Hetru, C., and Hoffmann, J.A. (1995). Metchnikowin, a novel immune-inducible proline-rich peptide from *Drosophila* with antibacterial and antifungal properties. *Eur. J. Biochem*. 233(2), 694–700. doi: 10.1111/j.1432-1033.1995.694_2.x.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5), 589–595. doi: 10.1093/bioinformatics/btp698.

Li, H.-M., Deng, R.-Q., Wang, J.-W., Chen, Z.-Y., Jia, F.-L., and Wang, X.-Z. (2005). A preliminary phylogeny of the Pentatomomorpha (Hemiptera: Heteroptera) based on nuclear 18S rDNA and mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 37(2), 313–326. doi: 10.1016/j.ympev.2005.07.013.

Li, X., Wang, B., Feng, L., Kang, H., Qi, Y., Wang, J., and Shi, Y. (2009). Cleavage of RseA by RseP requires a carboxyl-terminal hydrophobic amino acid following DegS cleavage. *Proc. Natl. Acad. Sci. U.S.A.* 106(35), 14837–14842. doi: 10.1073/pnas.0903289106.

Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7), 923–930. doi: 10.1093/bioinformatics/btt656.

Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., Villanueva, J., Wei, T., and Ausubel, F.M. (2006). An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl. Acad. Sci. U.S.A.* 103(8), 2833–2838. doi: 10.1073/pnas.0511100103.

Lim, J.S., Choi, B.S., Choi, A.Y., Kim, K.D., Kim, D.I., Choi, I.Y., and Ka, J.-O. (2012). Complete Genome Sequence of the Fenitrothion-Degrading *Burkholderia* sp. Strain YI23. *J. Bacteriol.* 194(4), 896. doi: 10.1128/JB.06479-11.

Lloubès, R., Cascales, E., Walburger, A., Bouveret, E., Lazdunski, C., Bernadac, A., and Journet, L. (2001). The Tol-Pal proteins of the *Escherichia coli* cell envelope: an energized system required for outer membrane integrity? *Res. Microbiol.* 152(6), 523–529.

Login, F.H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., and Heddi, A. (2011). Antimicrobial peptides keep insect endosymbionts under control. *Science* 334(6054), 362–365. doi: 10.1126/science.1209728.

Lohman, T.M., and Fazio, N.T. (2018). How Does a Helicase Unwind DNA? Insights from RecBCD Helicase. *Bioessays* 40(6), e1800009. doi: 10.1002/bies.201800009.

Lopes-Santos, L., Castro, D.B.A., Ferreira-Tonin, M., Corrêa, D.B.A., Weir, B.S., Park, D., Ottoboni, L.M.M., Neto, J.R., and Destéfano, S.A.L. (2017). Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of *Robbsia andropogonis* gen. nov., comb. nov. *Antonie Van Leeuwenhoek* 110(6), 727–736. doi: 10.1007/s10482-017-0842-6.

López-García, P., Eme, L., and Moreira, D. (2017). Symbiosis in eukaryotic evolution. *J. Theor. Biol.* 434, 20–33. doi: 10.1016/j.jtbi.2017.02.031.

Loutet, S.A., and Valvano, M.A. (2011). Extreme antimicrobial Peptide and polymyxin B resistance in the genus *Burkholderia*. *Front. Microbiol.* 2, 159. doi: 10.3389/fmicb.2011.00159.

Loutet, S.A., Flannagan, R.S., Kooi, C., Sokol, P.A., and Valvano, M.A. (2006). A complete lipopolysaccharide inner core oligosaccharide is required for resistance of *Burkholderia cenocepacia* to antimicrobial peptides and bacterial survival *in vivo. J. Bacteriol.* 188(6), 2073–2080. doi: 10.1128/JB.188.6.2073-2080.2006.

Loutet, S.A., Mussen, L.E., Flannagan, R.S., and Valvano, M.A. (2011). A two-tier model of polymyxin B resistance in *Burkholderia cenocepacia. Environ. Microbiol. Rep.* 3(2), 278–285. doi: 10.1111/j.1758-2229.2010.00222.x.

Lu, Y., Chen, M., Reding, K., and Pick, L. (2017). Establishment of molecular genetic approaches to study gene expression and function in an invasive hemipteran, *Halyomorpha halys. Evodevo* 8, 15. doi: 10.1186/s13227-017-0078-6.

Lyell, N.L., Septer, A.N., Dunn, A.K., Duckett, D., Stoudenmire, J.L., and Stabb, E.V. (2017). An Expanded Transposon Mutant Library Reveals that *Vibrio fischeri* δ-Aminolevulinate Auxotrophs Can Colonize *Euprymna scolopes. Appl. Environ. Microbiol.* 83(5), pii: e02470-16. doi: 10.1128/AEM.02470-16.

Macho, A.P., Rufián, J.S., Ruiz-Albert, J., and Beuzón, C.R. (2016). Competitive Index: Mixed Infection-Based Virulence Assays for Genetic Analysis in *Pseudomonas syringae*-Plant Interactions. *Methods Mol. Biol.* 1363, 209–217. doi: 10.1007/978-1-4939-3115-6_17.

Malinverni, J.C., and Silhavy, T.J. (2009). An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proc. Natl. Acad. Sci. U.S.A.* 106(19), 8009–8014. doi: 10.1073/pnas.0903229106.

Maloney, K.E., and Valvano, M.A. (2006). The *mgtC* gene of *Burkholderia cenocepacia* is required for growth under magnesium limitation conditions and intracellular survival in macrophages. *Infect. Immun.* 74(10), 5477–5486. doi: 10.1128/IAI.00798-06.

Malott, R.J., Steen-Kinnaird, B.R., Lee, T.D., and Speert, D.P. (2012). Identification of hopanoid biosynthesis genes involved in polymyxin resistance in *Burkholderia multivorans*. *Antimicrob. Agents Chemother.* 56(1), 464–471. doi: 10.1128/AAC.00602-11.

Malott, R.J., Wu, C.-H., Lee, T.D., Hird, T.J., Dalleska, N.F., Zlosnik, J.E.A., Newman, D.K., and Speert, D.P. (2014). Fosmidomycin decreases membrane hopanoids and potentiates the effects of colistin on *Burkholderia multivorans* clinical isolates. *Antimicrob. Agents Chemother.* 58(9), 5211–5219. doi: 10.1128/AAC.02705-14.

Manandhar, M., and Cronan, J.E. (2018). A Canonical Biotin Synthesis Enzyme, 8-Amino-7-Oxononanoate Synthase (BioF), Utilizes Different Acyl Chain Donors in *Bacillus subtilis* and *Escherichia coli. Appl. Environ. Microbiol.* 84(1), pii: e02084-17. doi: 10.1128/AEM.02084-17.

Mandel, M.J., and Dunn, A.K. (2016). Impact and Influence of the Natural *Vibrio*-Squid Symbiosis in Understanding Bacterial–Animal Interactions. *Front. Microbiol.* 7, 1982. doi: 10.3389/fmicb.2016.01982.

Manta, B., Boyd, D., and Berkmen, M. (2019). Disulfide Bond Formation in the Periplasm of *Escherichia coli. EcoSal. Plus* 8(2). doi: 10.1128/ecosalplus.ESP-0012-2018.

Manzano-Marín, A., Coeur d'Acier, A., Clamens, A.-L., Orvain, C., Cruaud, C., Barbe, V., and Jousselin, E. (2018). A Freeloader? The Highly Eroded Yet large Genome of the *Serratia symbiotica* symbiont of *Cinara strobi*. *Genome Biol. Evol.* 10(9), 2178–2189. doi: 10.1093/gbe/evy173.

Maróti, G., and Kondorosi, E. (2014). Nitrogen-fixing *Rhizobium*-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front. Microbiol.* 5, 326. doi: 10.3389/fmicb.2014.00326.

Maróti, G., Kereszt, A., Kondorosi, E., and Mergaert, P. (2011). Natural roles of antimicrobial peptides in microbes, plants and animals. *Res. Microbiol.* 162(4), 363–374. doi: 10.1016/j.resmic.2011.02.005.

Masilamani, R., Cian, M.B., and Dalebroux, Z.D. (2018). *Salmonella* Tol-Pal Reduces Outer Membrane Glycerophospholipid Levels for Envelope Homeostasis and Survival during Bacteremia. *Infect. Immun.* 86(7), pii: e00173-18. doi: 10.1128/IAI.00173-18.

Matsuura, Y., Kikuchi, Y., Hosokawa, T., Koga, R., Meng, X.-Y., Kamagata, Y., Nikoh, N., and Fukatsu, T. (2012). Evolution of symbiotic organs and endosymbionts in lygaeid stinkbugs. *ISME J.* 6(2), 397–409. doi: 10.1038/ismej.2011.103.

Matsuzaka, K., Sato, D., Ishihara, K., Hashimoto, S., Yoshinari, M., Katakura, A., and Inoue, T. (2006). Age-related differences in localization of beta-defensin-2 in human gingival epithelia. *Bull. Tokyo Dent. Coll.* 47(4), 167–170. doi: 10.2209/tdcpublication.47.167.

Mattiuzzo, M., De Gobba, C., Runti, G., Mardirossian, M., Bandiera, A., Gennaro, R., and Scocchi, M. (2014). Proteolytic activity of *Escherichia coli* oligopeptidase B against proline-rich antimicrobial peptides. *J. Microbiol. Biotechnol.* 24(2), 160–167.

McCutcheon, J.P., and Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol*. 10(1), 13–26. doi: 10.1038/nrmicro2670.

McDonough, J.A., Hacker, K.E., Flores, A.R., Pavelka, M.S., and Braunstein, M. (2005). The twin-arginine translocation pathway of *Mycobacterium smegmatis* is functional and required for the export of mycobacterial beta-lactamases. *J. Bacteriol.* 187(22), 7667–7679. doi: 10.1128/JB.187.22.7667-7679.2005.

McFall-Ngai, M.J. (2014). The Importance of Microbes in Animal Development: Lessons from the Squid-*Vibrio* Symbiosis. *Annu. Rev. Microbiol.* 68, 177–194. doi: 10.1146/annurev-micro-091313-103654.

Médigue, C., Calteau, A., Cruveiller, S., Gachet, M., Gautreau, G., Josso, A., Lajus, A., Langlois, J., Pereira, H., Planel, R., *et al.* (2017). MicroScope-an integrated resource for community expertise of gene functions and comparative analysis of microbial genomic and metabolic data. *Brief. Bioinformatics*. doi: 10.1093/bib/bbx113.

Meehan, B.M., Landeta, C., Boyd, D., and Beckwith, J. (2017). The Disulfide Bond Formation Pathway Is Essential for Anaerobic Growth of *Escherichia coli*. *J. Bacteriol.* 199(16). pii: e00120-17. doi: 10.1128/JB.00120-17.

Mercer, R., Nguyen, O., Ou, Q., McMullen, L., and Gänzle, M.G. (2017). Functional Analysis of Genes Comprising the Locus of Heat Resistance in *Escherichia coli. Appl. Environ. Microbiol.* 83(20). pii: e01400-17. doi: 10.1128/AEM.01400-17.

Mergaert, P. (2018). Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Nat. Prod. Rep.* 35(4), 336–356. doi: 10.1039/c7np00056a.

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., and Kondorosi, E. (2003). A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* 132(1), 161–173. doi: 10.1104/pp.102.018192.

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E., Barloy-Hubler, F., Galibert, F., Kondorosi, A., Kondorosi, E. (2006). Eukaryotic control on bacterial cell cycle

and differentiation in the *Rhizobium*–legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A. 103*, 5230–5235. doi: 10.1073/pnas.0600912103.

Mergaert, P., Kikuchi, Y., Shigenobu, S., and Nowack, E.C.M. (2017). Metabolic Integration of Bacterial Endosymbionts through Antimicrobial Peptides. *Trends Microbiol.* 25(9), 703–712. doi: 10.1016/j.tim.2017.04.007.

Mitchell, A.M., and Silhavy, T.J. (2019). Envelope stress responses: balancing damage repair and toxicity. *Nat. Rev. Microbiol.* 17(7), 417–428. doi: 10.1038/s41579-019-0199-0.

Moran, N.A. (2006). Symbiosis. *Curr. Biol.* 16(20), R866–R871. doi: 10.1016/j.cub.2006.09.019.

Moran, N.A., and Mira, A. (2001). The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola. Genome Biol.* 2(12), research0054.1-research0054.12. doi: 10.1186/gb-2001-2-12-research0054

Morty, R.E., Fülöp, V., and Andrews, N.W. (2002). Substrate recognition properties of oligopeptidase B from *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 184(12), 3329–3337. doi: 10.1128/jb.184.12.3329-3337.2002.

Moule, M.G., Hemsley, C.M., Seet, Q., Guerra-Assuncao, J.A., Lim, J., Sarkar-Tyson, M., Clark, T.G., Tan, P.B.O., Titball, R.W., Cuccui, J., and Wren, B.W. (2014). Genome-Wide Saturation Mutagenesis of *Burkholderia pseudomallei* K96243 Predicts Essential Genes and Novel Targets for Antimicrobial Development. *MBio* 5(1), e00926-13.doi: 10.1128/mBio.00926-13

Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. (2001). Nodulation of legumes by members of the beta-subclass of Proteobacteria. *Nature* 411(6840), 948–950. doi: 10.1038/35082070.

Nakayama, T., and Zhang-Akiyama, Q.M. (2017). *pqiABC* and *yebST*, Putative *mce* Operons of *Escherichia coli,* Encode Transport Pathways and Contribute to Membrane Integrity. *J. Bacteriol.* 199(1) pii: e00606-16. doi: 10.1128/JB.00606-16.

Nelson, M.C., Bomar, L., and Graf, J. (2015). Complete Genome Sequence of the Novel Leech Symbiont *Mucinivorans hirudinis* M3T. *Genome Announc.* 3(1). pii: e01530-14. doi: 10.1128/genomeA.01530-14.

Nishide, Y., Kageyama, D., Yokoi, K., Jouraku, A., Tanaka, H., Futahashi, R., and Fukatsu, T. (2019). Functional crosstalk across IMD and Toll pathways: insight into the evolution of incomplete immune cascades. *Proc. Biol. Sci.* 286(1897), 2018–2207. doi: 10.1098/rspb.2018.2207.

Nishie, M., Nagao, J.-I., and Sonomoto, K. (2012). Antibacterial peptides “bacteriocins”: an overview of their diverse characteristics and applications. *Biocontrol Sci.* 17(1), 1–16.

Nizet, V., Ohtake, T., Lauth, X., Trowbridge, J., Rudisill, J., Dorschner, R.A., Pestonjamasp, V., Piraino, J., Huttner, K., and Gallo, R.L. (2001). Innate antimicrobial peptide protects the skin from invasive bacterial infection. *Nature* 414(6862), 454–457. doi: 10.1038/35106587.

Nsahlai, C.J., and Silver, R.P. (2003). Purification and characterization of KpsT, the ATP-binding component of the ABC-capsule exporter of *Escherichia coli* K1. *FEMS Microbiol. Lett.* 224(1), 113–118. doi: 10.1016/S0378-1097(03)00428-2.

Ochsner, U.A., Snyder, A., Vasil, A.I., and Vasil, M.L. (2002). Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 99(12), 8312–8317. doi: 10.1073/pnas.082238299

Ohbayashi, T., Takeshita, K., Kitagawa, W., Nikoh, N., Koga, R., Meng, X.-Y., Tago, K., Hori, T., Hayatsu, M., Asano, K., *et al.* (2015). Insect's intestinal organ for symbiont sorting. *Proc. Natl. Acad. Sci. U.S.A.* 112(37), E5179–E5188. doi: 10.1073/pnas.1511454112.

Ohbayashi, T., Itoh, H., Lachat, J., Kikuchi, Y., and Mergaert, P. (2019a). *Burkholderia* Gut Symbionts Associated with European and Japanese Populations of the Dock Bug *Coreus marginatus* (Coreoidea: Coreidae). *Microbes Environ.* 34(2), 219–222. doi: 10.1264/jsme2.ME19011.

Ohbayashi, T., Futahashi, R., Terashima, M., Barrière, Q., Lamouche, F., Takeshita, K., Meng, X.-Y., Mitani, Y., Sone, T., Shigenobu, S., *et al.* (2019b). Comparative cytology, physiology and transcriptomics of *Burkholderia insecticola* in symbiosis with the bean bug *Riptortus pedestris* and in culture. *ISME J.* 13(6), 1469–1483. doi: 10.1038/s41396-019-0361-8.

Ohtsubo, Y., Goto, H., Nagata, Y., Kudo, T., and Tsuda, M. (2006). Identification of a response regulator gene for catabolite control from a PCB-degrading beta-proteobacteria, *Acidovorax* sp. KKS102. *Mol. Microbiol.* 60(6), 1563–1575. doi: 10.1111/j.1365-2958.2006.05197.x.

Oliver, K.M., Degnan, P.H., Burke, G.R., and Moran, N.A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu. Rev. Entomol.* 55, 247–266. doi: 10.1146/annurev-ento-112408-085305.

Onchuru, T.O., Javier Martinez, A., Ingham, C.S., and Kaltenpoth, M. (2018). Transmission of mutualistic bacteria in social and gregarious insects. *Curr. Opin. Insect Sci.* 28, 50–58. doi: 10.1016/j.cois.2018.05.002.

van Opijnen, T., Bodi, K.L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* 6(10), 767–772. doi: 10.1038/nmeth.1377.

Ormeño-Orrillo, E., Rosenblueth, M., Luyten, E., Vanderleyden, J., and Martínez-Romero, E. (2008). Mutations in lipopolysaccharide biosynthetic genes impair maize rhizosphere and root colonization of *Rhizobium tropici* CIAT899. *Environ. Microbiol.* 10(5), 1271–1284. doi: 10.1111/j.1462-2920.2007.01541.x

Ortega, X., Hunt, T.A., Loutet, S., Vinion-Dubiel, A.D., Datta, A., Choudhury, B., Goldberg, J.B., Carlson, R., and Valvano, M.A. (2005). Reconstitution of O-specific lipopolysaccharide expression in *Burkholderia cenocepacia* strain J2315, which is associated with transmissible infections in patients with cystic fibrosis. *J. Bacteriol.* 187(4), 1324–1333. doi: 10.1128/JB.187.4.1324-1333.2005.

Ortega, X., Silipo, A., Saldías, M.S., Bates, C.C., Molinaro, A., and Valvano, M.A. (2009). Biosynthesis and structure of the *Burkholderia cenocepacia* K56-2 lipopolysaccharide core oligosaccharide: truncation of the core oligosaccharide leads to increased binding and sensitivity to polymyxin B. *J. Biol. Chem.* 284(32), 21738–21751. doi: 10.1074/jbc.M109.008532.

Ortega, X.P., Cardona, S.T., Brown, A.R., Loutet, S.A., Flannagan, R.S., Campopiano, D.J., Govan, J.R.W., and Valvano, M.A. (2007). A Putative Gene Cluster for Aminoarabinose Biosynthesis Is Essential for *Burkholderia cenocepacia* Viability. *J. Bacteriol.* 189(9), 3639–3644. doi: 10.1128/JB.00153-07.

Oulhen, N., Schulz, B.J., and Carrier, T.J. (2016). English translation of Heinrich Anton de Bary's 1878 speech, "Die Erscheinung der Symbiose" ("De la symbiose"). *Symbiosis* 69(3), 131–139.

Ourisson, G., and Albrecht, P. (1992). Hopanoids. 1. Geohopanoids: the most abundant natural products on Earth? *Acc. Chem. Res.* 25(9), 398–402. doi: 10.1021/ar00021a003.

Paik, C.-H., Choi, M.-Y., Seo, H.-Y., and Kim, J.-D. (2007). Effects of Temperature on The Development and Reproduction of *Cletus punctiger* Dallas and *Cletus schmidti* Kiritshenko (Heteroptera : Coreidae) on Rice. *Korean Journal of Applied Entomology* 46(1), 51–56.

Palmer, L.D., and Downs, D.M. (2013). The thiamine biosynthetic enzyme ThiC catalyzes multiple turnovers and is inhibited by S-adenosylmethionine (AdoMet) metabolites. *J. Biol. Chem.* 288(42), 30693–30699. doi: 10.1074/jbc.M113.500280.

Pan, J.-J., Solbiati, J.O., Ramamoorthy, G., Hillerich, B.S., Seidel, R.D., Cronan, J.E., Almo, S.C., and Poulter, C.D. (2015). Biosynthesis of Squalene from Farnesyl Diphosphate in Bacteria: Three Steps Catalyzed by Three Enzymes. *ACS Cent. Sci.* 1(2), 77–82. doi: 10.1021/acscentsci.5b00115.

Pankey, M.S., Minin, V.N., Imholte, G.C., Suchard, M.A., and Oakley, T.H. (2014). Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proc. Natl. Acad. Sci. U.S.A.* 111(44), E4736-4742. doi: 10.1073/pnas.1416574111.

Park, H.C., Bae, Y.U., Cho, S.D., Kim, S.A., Moon, J.Y., Ha, K.C., Kim, D.W., Lee, K., Jeong, Y.K., Kwack, D.O., *et al.* (2007). Toluene-induced accumulation of trehalose by *Pseudomonas* sp. BCNU 106 through the expression of *otsA* and *otsB* homologues. *Lett. Appl. Microbiol.* 44(1), 50–55. doi: 10.1111/j.1472-765X.2006.02036.x.

Park, K.-E., Jang, S.H., Lee, J., Lee, S.A., Kikuchi, Y., Seo, Y.-S., and Lee, B.L. (2018). The roles of antimicrobial peptide, rip-thanatin, in the midgut of *Riptortus pedestris*. *Dev. Comp. Immunol.* 78, 83–90. doi: 10.1016/j.dci.2017.09.009.

Pearson, A., Flood Page, S.R., Jorgenson, T.L., Fischer, W.W., and Higgins, M.B. (2007). Novel hopanoid cyclases from the environment. *Environ. Microbiol.* 9(9), 2175–2188. doi: 10.1111/j.1462-2920.2007.01331.x.

Pechter, K.B., Gallagher, L., Pyles, H., Manoil, C.S., and Harwood, C.S. (2016). Essential Genome of the Metabolically Versatile Alphaproteobacterium *Rhodopseudomonas palustris*. *J. Bacteriol.* 198(5), 867–876. doi: 10.1128/JB.00771-15.

Peng, Q., Yang, M., Wang, W., Han, L., Wang, G., Wang, P., Zhang, J., and Song, F. (2014). Activation of *gab* cluster transcription in *Bacillus thuringiensis* by γ-aminobutyric acid or succinic semialdehyde is mediated by the Sigma 54-dependent transcriptional activator GabR. *BMC Microbiol.* 14, 306. doi: 10.1186/s12866-014-0306-3.

Peng, Q., Wang, G., Liu, G., Zhang, J., and Song, F. (2015). Identification of metabolism pathways directly regulated by sigma54 factor in *Bacillus thuringiensis*. *Front. Microbiol.* 6, 407. doi: 10.3389/fmicb.2015.00407.

Pföstl, A., Zayni, S., Hofinger, A., Kosma, P., Schäffer, C., and Messner, P. (2008). Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-glucose. *Biochem. J.* 410(1), 187–194. doi: 10.1042/BJ20071044.

Phelan, J.P., Kern, A., Ramsey, M.E., Lundt, M.E., Sharma, B., Lin, T., Gao, L., Norris, S.J., Hyde, J.A., Skare, J.T., and Hu, L.T. (2019). Genome-wide screen identifies novel genes required for *Borrelia burgdorferi* survival in its Ixodes tick vector. *PLoS Pathog.* 15(5), e1007644. doi: 10.1371/journal.ppat.1007644.

Pickard, J.M., Zeng, M.Y., Caruso, R., and Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol. Rev.* 279(1), 70–89. doi: 10.1111/imr.12567.

Pinto-Carbó, M., Gademann, K., Eberl, L., and Carlier, A. (2018). Leaf nodule symbiosis: function and transmission of obligate bacterial endophytes. *Curr. Opin. Plant Biol.* 44, 23–31. doi: 10.1016/j.pbi.2018.01.001.

Pinto-Carbó, M., Sieber, S., Dessein, S., Wicker, T., Verstraete, B., Gademann, K., Eberl, L., and Carlier, A. (2016). Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *ISME J.* 10(9), 2092–2105. doi: 10.1038/ismej.2016.27.

Poblete-Castro, I., Escapa, I.F., Jäger, C., Puchalka, J., Chi Lam, C.M., Schomburg, D., Prieto, M.A., and Martins dos Santos, V.A. (2012). The metabolic response of *P. putida* KT2442 producing high levels of polyhydroxyalkanoate under single- and multiple-nutrient-limited growth: Highlights from a multi-level omics approach. *Microb. Cell Fact.* 11, 34. doi: 10.1186/1475-2859-11-34.

Poralla, K., Muth, G., and Härtner, T. (2000). Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* 189(1), 93–95. doi: 10.1111/j.1574-6968.2000.tb09212.x.

Post, D.M.B., Yu, L., Krasity, B.C., Choudhury, B., Mandel, M.J., Brennan, C.A., Ruby, E.G., McFall-Ngai, M.J., Gibson, B.W., and Apicella, M.A. (2012). O-antigen and core carbohydrate of *Vibrio fischeri* lipopolysaccharide: composition and analysis of their role in *Euprymna scolopes* light organ colonization. *J. Biol. Chem.* 287(11), 8515–8530. doi: 10.1074/jbc.M111.324012.

Postma, P.W., Lengeler, J.W., and Jacobson, G.R. (1993). Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* 57(3), 543–594.

Potrykus, K., and Cashel, M. (2008). (p)ppGpp: still magical? *Annu. Rev. Microbiol*. 62, 35–51. doi: 10.1146/annurev.micro.62.081307.162903.

Powell, J.E., Leonard, S.P., Kwong, W.K., Engel, P., and Moran, N.A. (2016). Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. *Proc. Natl. Acad. Sci. U.S.A.* 113(48), 13887–13892. doi: 10.1073/pnas.1610856113.

Pradel, E., Parker, C.T., and Schnaitman, C.A. (1992). Structures of the *rfaB*, *rfaI*, *rfaJ*, and *rfaS* genes of *Escherichia coli* K-12 and their roles in assembly of the lipopolysaccharide core. *J. Bacteriol.* 174(14), 4736–4745. doi: 10.1128/jb.174.14.4736-4745.1992.

Prieto-Àlamo, M.J., Jurado, J., Gallardo-Madueño, R., Monje-Casas, F., Holmgren, A., and Pueyo, C. (2000). Transcriptional regulation of glutaredoxin and thioredoxin pathways and related enzymes in response to oxidative stress. *J. Biol. Chem.* 275, 13398–13405. doi: 10.1074/jbc.275.18.13398.

Primon-Barros, M., and José Macedo, A. (2017). Animal Venom Peptides: Potential for New Antimicrobial Agents. *Curr. Top. Med. Chem.* 17(10), 1119–1156. doi: 10.2174/1568026616666160930151242.

Pritchard, J.R., Chao, M.C., Abel, S., Davis, B.M., Baranowski, C., Zhang, Y.J., Rubin, E.J., and Waldor, M.K. (2014). ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet.* 10(11), e1004782. doi: 10.1371/journal.pgen.1004782.

Qiu, D. (2012). [Essential genes, minimal genome and synthetic cell of bacteria: a review]. *Sheng Wu Gong Cheng Xue Bao* 28(5), 540–549.

Raetz, C.R.H., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* 71, 635–700. doi: 10.1146/annurev.biochem.71.110601.135414.

Raivio, T.L. (2005). Envelope stress responses and Gram-negative bacterial pathogenesis. *Mol. Microbiol.* 56(5), 1119–1128. doi: 10.1111/j.1365-2958.2005.04625.x.

Rajagopala, S.V., Titz, B., Goll, J., Parrish, J.R., Wohlbold, K., McKevitt, M.T., Palzkill, T., Mori, H., Finley, R.L., and Uetz, P. (2007). The protein network of bacterial motility. *Mol. Syst. Biol.* 3, 128. doi: 10.1038/msb4100166.

Reckseidler-Zenteno, S.L., DeVinney, R., and Woods, D.E. (2005). The capsular polysaccharide of *Burkholderia pseudomallei* contributes to survival in serum by reducing complement factor C3b deposition. *Infect. Immun*. 73(2), 1106–1115. doi: 10.1128/IAI.73.2.1106-1115.2005.

Reeve, W., O'Hara, G., Chain, P., Ardley, J., Bräu, L., Nandesena, K., Tiwari, R., Copeland, A., Nolan, M., Han, C., *et al.* (2010). Complete genome sequence of *Rhizobium leguminosarum* bv. trifolii strain WSM1325, an effective microsymbiont of annual Mediterranean clovers. *Stand. Genomic Sci.* 2(3), 347–356. doi: 10.4056/sigs.852027.

Remigi, P., Zhu, J., Young, J.P.W., and Masson-Boivin, C. (2016). Symbiosis within Symbiosis: Evolving Nitrogen-Fixing Legume Symbionts. *Trends Microbiol.* 24(1), 63–75. doi: 10.1016/j.tim.2015.10.007.

Rhodes, K.A., and Schweizer, H.P. (2016). Antibiotic resistance in *Burkholderia* species. *Drug Resist. Updat.* 28, 82–90. doi: 10.1016/j.drup.2016.07.003.

Robinson, C., Matos, C.F.R.O., Beck, D., Ren, C., Lawrence, J., Vasisht, N., and Mendel, S. (2011a). Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochim. Biophys. Acta* 1808(3), 876–884. doi: 10.1016/j.bbamem.2010.11.023.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011b). Integrative genomics viewer. *Nat. Biotechnol.* 29(1), 24–26. doi: 10.1038/nbt.1754.

Rocchetta, H.L., and Lam, J.S. (1997). Identification and functional characterization of an ABC transport system involved in polysaccharide export of A-band lipopolysaccharide in *Pseudomonas aeruginosa. J. Bacteriol.* 179(15), 4713–4724. doi: 10.1128/jb.179.15.4713-4724.1997.

Römling, U., and Galperin, M.Y. (2015). Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends Microbiol.* 23(9), 545–557. doi: 10.1016/j.tim.2015.05.005.

Rose, R.W., Brüser, T., Kissinger, J.C., and Pohlschröder, M. (2002). Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol. Microbiol.* 45(4), 943–950. doi: 10.1046/j.1365-2958.2002.03090.x.

Rosenow, C., Esumeh, F., Roberts, I.S., and Jann, K. (1995). Characterization and localization of the KpsE protein of *Escherichia coli* K5, which is involved in polysaccharide export. *J. Bacteriol.* 177(5), 1137–1143. doi: 10.1128/jb.177.5.1137-1143.1995.

Ruby, E.G., Urbanowski, M., Campbell, J., Dunn, A., Faini, M., Gunsalus, R., Lostroh, P., Lupp, C., McCann, J., Millikan, D., *et al.* (2005). Complete genome sequence of *Vibrio fischeri*: a symbiotic

bacterium with pathogenic congeners. *Proc. Natl. Acad. Sci. U.S.A.* 102(8), 3004–3009. doi: 10.1073/pnas.0409900102.

Rutschmann, S., Kilinc, A., and Ferrandon, D. (2002). Cutting edge: the toll pathway is required for resistance to gram-positive bacterial infections in *Drosophila*. *J. Immunol.* 168(4), 1542–1546. doi: 10.4049/jimmunol.168.4.1542.

Sahm, H., Rohmer, M., Bringer-Meyer, S., Sprenger, G.A., and Welle, R. (1993). Biochemistry and Physiology of Hopanoids in Bacteria. *Adv.Microb. Physiol.* 35, 247–273.

Saikh, K.U., and Mott, T.M. (2017). Innate immune response to *Burkholderia mallei*. *Curr. Opin. Infect. Dis.* 30(3), 297–302. doi: 10.1097/QCO.0000000000000362.

Salem, H., Kreutzer, E., Sudakaran, S., and Kaltenpoth, M. (2013). Actinobacteria as essential symbionts in firebugs and cotton stainers (Hemiptera, Pyrrhocoridae). *Environ. Microbiol.* 15(7), 1956–1968. doi: 10.1111/1462-2920.12001.

Salem, H., Florez, L., Gerardo, N., and Kaltenpoth, M. (2015). An out-of-body experience: the extracellular dimension for the transmission of mutualistic bacteria in insects. *Proc. Biol. Sci.* 282(1804), 20142957. doi: 10.1098/rspb.2014.2957.

Salmon, L., Stull, F., Sayle, S., Cato, C., Akgül, Ş., Foit, L., Ahlstrom, L.S., Eisenmesser, E.Z., Al-Hashimi, H.M., Bardwell, J.C.A., and Horowitz, S. (2018). The Mechanism of HdeA Unfolding and Chaperone Activation. *J. Mol. Biol.* 430(1), 33–40. doi: 10.1016/j.jmb.2017.11.002.

Salzman, N.H., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjöberg, J., Amir, E., Teggatz, P., Barman, M., Hayward, M., Eastwood, D., *et al.* (2010). Enteric defensins are essential regulators of intestinal microbial ecology. *Nat. Immunol.* 11(1), 76–83. doi: 10.1038/ni.1825.

Sawana, A., Adeolu, M., and Gupta, R.S. (2014). Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front. Genet.* 5, 429. doi: 10.3389/fgene.2014.00429.

Sazama, E.J., Ouellette, S.P., and Wesner, J.S. (2019). Bacterial Endosymbionts Are Common Among, but not Necessarily Within, Insect Species. *Environ. Entomol.* 48(1), 127–133. doi: 10.1093/ee/nvy188.

Schaefer, C., and Panizzi, A. (2000). Economic importance of Heteroptera: a general view. p. 3-8. In: Schaefer, C. W. & Panizzi, A.R. (eds.). Heteroptera of Economic Importance, *CRC Press*, Boca Raton, FL., USA, 828 pp. p.

Schmerk, C.L., Bernards, M.A., and Valvano, M.A. (2011). Hopanoid Production Is Required for Low-pH Tolerance, Antimicrobial Resistance, and Motility in *Burkholderia cenocepacia*. *J. Bacteriol.* 193(23), 6712–6723. doi: 10.1128/JB.05979-11.

Schmerk, C.L., Welander, P.V., Hamad, M.A., Bain, K.L., Bernards, M.A., Summons, R.E., and Valvano, M.A. (2015). Elucidation of the *Burkholderia cenocepacia* hopanoid biosynthesis pathway uncovers functions for conserved proteins in hopanoid-producing bacteria. *Environ. Microbiol.* 17(3), 735–750. doi: 10.1111/1462-2920.12509.

Scocchi, M., Mardirossian, M., Runti, G., and Benincasa, M. (2016). Non-Membrane Permeabilizing Modes of Action of Antimicrobial Peptides on Bacteria. *Curr. Top. Med. Chem.* 16(1), 76–88.

Scoffone, V.C., Chiarelli, L.R., Trespidi, G., Mentasti, M., Riccardi, G., and Buroni, S. (2017). *Burkholderia cenocepacia* Infections in Cystic Fibrosis Patients: Drug Resistance and Therapeutic Approaches. *Front. Microbiol. 8,* 1592. doi: 10.3389/fmicb.2017.01592.

Segura, A., Molina, L., Fillet, S., Krell, T., Bernal, P., Muñoz-Rojas, J., and Ramos, J.-L. (2012). Solvent tolerance in Gram-negative bacteria. *Curr. Opin. Biotechnol.* 23(3), 415–421. doi: 10.1016/j.copbio.2011.11.015.

Semmler, A.B., Whitchurch, C.B., Leech, A.J., and Mattick, J.S. (2000). Identification of a novel gene, fimV, involved in twitching motility in *Pseudomonas aeruginosa*. *Microbiology* 146 (Pt 6), 1321–1332. doi: 10.1099/00221287-146-6-1321.

Seol, J.H., Woo, S.K., Jung, E.M., Yoo, S.J., Lee, C.S., Kim, K.J., Tanaka, K., Ichihara, A., Ha, D.B., and Chung, C.H. (1991). Protease Do is essential for survival of *Escherichia coli* at high temperatures: its identity with the htrA gene product. *Biochem. Biophys. Res. Commun.* 176(2), 730–736. doi: 10.1016/s0006-291x(05)80245-1.

Sessitsch, A., Coenye, T., Sturz, A.V., Vandamme, P., Barka, E.A., Salles, J.F., Van Elsas, J.D., Faure, D., Reiter, B., Glick, B.R., Wang-Pruski, G., and Nowak, J. (2005). *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. *Int. J. Syst. Evol. Microbiol.* 55(Pt 3), 1187–1192. doi: 10.1099/ijs.0.63149-0.

Sfeir, M.M. (2018). *Burkholderia cepacia* complex infections: More complex than the bacterium name suggest. *J. Infect.* 77(3), 166–170. doi: 10.1016/j.jinf.2018.07.006

Shafer, W.M., Casey, S.G., and Spitznagel, J.K. (1984). Lipid A and resistance of *Salmonella typhimurium* to antimicrobial granule proteins of human neutrophil granulocytes. *Infect. Immun.* 43(3), 834–838.

Shaffer, C.L., Zhang, E.W., Dudley, A.G., Dixon, B.R.E.A., Guckes, K.R., Breland, E.J., Floyd, K.A., Casella, D.P., Algood, H.M.S., Clayton, D.B., and Hadjifrangiskou, M. (2017). Purine Biosynthesis Metabolically Constrains Intracellular Survival of Uropathogenic *Escherichia coli. Infect. Immun.* 85(1), pii: e00471-16.. doi: 10.1128/IAI.00471-16.

Sheng, Y., Fan, F., Jensen, O., Zhong, Z., Kan, B., Wang, H., and Zhu, J. (2015). Dual Zinc Transporter Systems in *Vibrio cholerae* Promote Competitive Advantages over Gut Microbiome. *Infect. Immun.* 83(10), 3902–3908. doi: 10.1128/IAI.00447-15.

Shi, Y., Cromie, M.J., Hsu, F.-F., Turk, J., and Groisman, E.A. (2004). PhoP-regulated *Salmonella* resistance to the antimicrobial peptides magainin 2 and polymyxin B. *Mol. Microbiol.* 53(1), 229–241. doi: 10.1111/j.1365-2958.2004.04107.x.

Shibata, T.F., Maeda, T., Nikoh, N., Yamaguchi, K., Oshima, K., Hattori, M., Nishiyama, T., Hasebe, M., Fukatsu, T., Kikuchi, Y., and Shigenobu, S. (2013). Complete Genome Sequence of *Burkholderia* sp. Strain RPE64, Bacterial Symbiont of the Bean Bug *Riptortus pedestris*. *Genome Announc.* 1(4), pii: e00441-13. doi: 10.1128/genomeA.00441-13.

Shigenobu, S., and Stern, D.L. (2013). Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *Proc. Biol. Sci.* 280(1750), 20121952. doi: 10.1098/rspb.2012.1952.

Shigenobu, S., and Wilson, A.C. (2011). Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cell. Mol. Life Sci.* 68(8), 1297–1309. doi: 10.1007/s00018-011-0645-2.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407(6800), 81–86. doi: 10.1038/35024074.

Shippy, D.C., and Fadl, A.A. (2014). tRNA modification enzymes GidA and MnmE: potential role in virulence of bacterial pathogens. *Int. J. Mol. Sci.* 15(10), 18267–18280. doi: 10.3390/ijms151018267.

Shippy, D.C., Eakley, N.M., Lauhon, C.T., Bochsler, P.N., and Fadl, A.A. (2013). Virulence characteristics of *Salmonella* following deletion of genes encoding the tRNA modification enzymes GidA and MnmE. *Microb. Pathog.* 57, 1–9. doi: 10.1016/j.micpath.2013.01.004.

Sieber, S., Carlier, A., Neuburger, M., Grabenweger, G., Eberl, L., and Gademann, K. (2015). Isolation and Total Synthesis of Kirkamide, an Aminocyclitol from an Obligate Leaf Nodule Symbiont. *Angew. Chem. Int. Ed. Engl.* 54(27), 7968–7970. doi: 10.1002/anie.201502696.

Singer, A., Poschmann, G., Mühlich, C., Valadez-Cano, C., Hänsch, S., Hüren, V., Rensing, S.A., Stühler, K., and Nowack, E.C.M. (2017). Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba *Paulinella chromatophora*. *Curr. Biol.* 27(18), 2763–2773.e5. doi: 10.1016/j.cub.2017.08.010.

Sinnesael, A., Leroux, O., Janssens, S.B., Smets, E., Panis, B., and Verstraete, B. (2019). Is the bacterial leaf nodule symbiosis obligate for *Psychotria umbellata*? The development of a *Burkholderia*-free host plant. *PloS One* 14(7), e0219863. doi: 10.1371/journal.pone.0219863.

Skurnik, D., Roux, D., Aschard, H., Cattoir, V., Yoder-Himes, D., Lory, S., and Pier, G.B. (2013). A Comprehensive Analysis of *In Vitro* and *In Vivo* Genetic Fitness of *Pseudomonas aeruginosa* Using High-Throughput Sequencing of Transposon Libraries. *PLoS Pathog.* 9(9), e1003582. doi: 10.1371/journal.ppat.1003582.

Söhling, B., and Gottschalk, G. (1996). Molecular analysis of the anaerobic succinate degradation pathway in *Clostridium kluyveri. J. Bacteriol.* 178(3), 871–880. doi: 10.1128/jb.178.3.871-880.1996.

Spiess, C., Beil, A., and Ehrmann, M. (1999). A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* 97(3), 339–347. doi: 10.1016/s0092-8674(00)80743-6.

Steiner, H., Hultmark, D., Engström, A., Bennich, H., and Boman, H.G. (1981). Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* 292(5820), 246–248. doi: 10.1038/292246a0.

Steiner, K., Hagelueken, G., Messner, P., Schäffer, C., and Naismith, J.H. (2010). Structural Basis of Substrate Binding in WsaF, a Rhamnosyltransferase from *Geobacillus stearothermophilus*. *J. Mol. Biol.* 397(2), 436–447. doi: 10.1016/j.jmb.2010.01.035.

Stephens, W.Z., Wiles, T.J., Martinez, E.S., Jemielita, M., Burns, A.R., Parthasarathy, R., Bohannan, B.J.M., and Guillemin, K. (2015). Identification of Population Bottlenecks and Colonization Factors during Assembly of Bacterial Communities within the Zebrafish Intestine. *MBio* 6(6), e01163-1115. doi: 10.1128/mBio.01163-15.

Su, H.-N., Chen, Z.-H., Song, X.-Y., Chen, X.-L., Shi, M., Zhou, B.-C., Zhao, X., and Zhang, Y.-Z. (2012). Antimicrobial Peptide Trichokonin VI-Induced Alterations in the Morphological and Nanomechanical Properties of *Bacillus subtilis. PLoS One* 7(9), e45818. doi: 10.1371/journal.pone.0045818.

Sudakaran, S., Salem, H., Kost, C., and Kaltenpoth, M. (2012). Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs, *Pyrrhocoris apterus* (Hemiptera, Pyrrhocoridae). *Mol. Ecol.* 21(24), 6134–6151. doi: 10.1111/mec.12027.

Sudakaran, S., Retz, F., Kikuchi, Y., Kost, C., and Kaltenpoth, M. (2015). Evolutionary transition in symbiotic syndromes enabled diversification of phytophagous insects on an imbalanced diet. *ISME J.* 9(12), 2587–2604. doi: 10.1038/ismej.2015.75.

Suzuki, T., Murai, T., Fukuda, I., Tobe, T., Yoshikawa, M., and Sasakawa, C. (1994). Identification and characterization of a chromosomal virulence gene, vacJ, required for intercellular spreading of *Shigella flexneri. Mol. Microbiol.* 11(1), 31–41. doi: 10.1111/j.1365-2958.1994.tb00287.x.

Takeshita, K., and Kikuchi, Y. (2017). *Riptortus pedestris* and *Burkholderia* symbiont: an ideal model system for insect-microbe symbiotic associations. *Res. Microbiol.* 168(3), 175–187. doi: 10.1016/j.resmic.2016.11.005.

Takeshita, K., Matsuura, Y., Itoh, H., Navarro, R., Hori, T., Sone, T., Kamagata, Y., Mergaert, P., and Kikuchi, Y. (2015). *Burkholderia* of Plant-Beneficial Group are Symbiotically Associated with Bordered Plant Bugs (Heteroptera: Pyrrhocoroidea: Largidae). *Microbes Environ.* 30(4), 321–329. doi: 10.1264/jsme2.ME15153.

Takeshita, K., Tamaki, H., Ohbayashi, T., Meng, X.-Y., Sone, T., Mitani, Y., Peeters, C., Kikuchi, Y., and Vandamme, P. (2018). *Burkholderia insecticola* sp. nov., a gut symbiotic bacterium of the bean bug *Riptortus pedestris*. *Int. J. Syst. Evol. Microbiol.* 68(7), 2370–2374. doi: 10.1099/ijsem.0.002848.

Tasiemski, A., Massol, F., Cuvillier-Hot, V., Boidin-Wichlacz, C., Roger, E., Rodet, F., Fournier, I., Thomas, F., and Salzet, M. (2015). Reciprocal immune benefit based on complementary production of antibiotics by the leech *Hirudo verbana* and its gut symbiont *Aeromonas veronii. Sci. Rep.* 5, 17498. doi: 10.1038/srep17498.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28(1), 33–36. doi: 10.1093/nar/28.1.33.

Thompson, K.M., and Gottesman, S. (2014). The MiaA tRNA modification enzyme is necessary for robust RpoS expression in *Escherichia coli. J. Bacteriol.* 196(4), 754–761. doi: 10.1128/JB.01013-13.

Thong, S., Ercan, B., Torta, F., Fong, Z.Y., Wong, H.Y.A., Wenk, M.R., and Chng, S.-S. (2016). Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry. *Elife* 5, pii: e19042. doi: 10.7554/eLife.19042.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14(2), 178–192. doi: 10.1093/bib/bbs017.

Thursby, E., and Juge, N. (2017). Introduction to the human gut microbiota. *Biochem. J.* 474(11), 1823–1836. doi: 10.1042/BCJ20160510.

Titball, R.W., Burtnick, M.N., Bancroft, G.J., and Brett, P. (2017). *Burkholderia pseudomallei* and *Burkholderia mallei* vaccines: Are we close to clinical trials? *Vaccine* 35(44), 5981–5989. doi: 10.1016/j.vaccine.2017.03.022.

Tomko, E.J., and Lohman, T.M. (2017). Modulation of *Escherichia coli* UvrD Single-Stranded DNA Translocation by DNA Base Composition. *Biophys. J.* 113(7), 1405–1415. doi: 10.1016/j.bpj.2017.08.023.

Trindade-Silva, A.E., Rua, C.P.J., Andrade, B.G.N., Vicente, A.C.P., Silva, G.G.Z., Berlinck, R.G.S., and Thompson, F.L. (2013). Polyketide Synthase Gene Diversity within the Microbiome of the Sponge *Arenosclera brasiliensis*, Endemic to the Southern Atlantic Ocean. *Appl. Environ. Microbiol.* 79(5), 1598–1605. doi: 10.1128/AEM.03354-12.

Tripathi, A., Dewan, P.C., Siddique, S.A., and Varadarajan, R. (2014). MazF-induced growth inhibition and persister generation in *Escherichia coli. J. Biol. Chem.* 289(7), 4191–4205. doi: 10.1074/jbc.M113.510511.

Troxell, B. (2016). *Salmonella enterica* serovar Typhimurium utilizes the ClpPX and Lon proteases for optimal fitness in the ceca of chickens. *Poult. Sci.* 95(7), 1617–1623. doi: 10.3382/ps/pew103.

Tsui, H.C., Feng, G., and Winkler, M.E. (1996). Transcription of the *mutL* repair, *miaA* tRNA modification, *hfq* pleiotropic regulator, and *hflA* region protease genes of *Escherichia coli* K-12 from clustered Esigma32-specific promoters during heat shock. *J. Bacteriol.* 178(19), 5719–5731. doi: 10.1128/jb.178.19.5719-5731.1996.

Tsukioka, Y., Yamashita, Y., Oho, T., Nakano, Y., and Koga, T. (1997). Biological function of the dTDP-rhamnose synthesis pathway in *Streptococcus mutans. J. Bacteriol.* 179(4), 1126–1134. doi: 10.1128/jb.179.4.1126-1134.1997.

Tzeng, Y.-L., and Stephens, D.S. (2015). Antimicrobial peptide resistance in *Neisseria meningitidis*. *Biochim. Biophys. Acta* 1848(11 Pt B), 3026–3031. doi: 10.1016/j.bbamem.2015.05.006.

Uchi, N., Fukudome, M., Nozaki, N., Suzuki, M., Osuki, K.-I., Shigenobu, S., and Uchiumi, T. (2019). Antimicrobial Activities of Cysteine-rich Peptides Specific to Bacteriocytes of the Pea Aphid *Acyrthosiphon pisum*. *Microbes Environ* 34(2), 155-160. doi: 10.1264/jsme2.ME18148.

Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science* 334(6053), 255–258. doi: 10.1126/science.1209791.

Valentini, M., Storelli, N., and Lapouge, K. (2011). Identification of C4-Dicarboxylate Transport Systems in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* 193(17), 4307–4316. doi: 10.1128/JB.05074-11.

Valentino, M.D., Foulston, L., Sadaka, A., Kos, V.N., Villet, R.A., Santa Maria, J., Lazinski, D.W., Camilli, A., Walker, S., Hooper, D.C., and Gilmore, M.S. (2014). Genes Contributing to *Staphylococcus aureus* Fitness in Abscess- and Infection-Related Ecologies. *MBio* 5(5), e01729–14. doi: 10.1128/mBio.01729-14.

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., *et al.* (2010). Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis. *Science* 327(5969), 1122–1126. doi: 10.1126/science.1184057.

Vanaporn, M., Vattanaviboon, P., Thongboonkerd, V., and Korbsrisate, S. (2008). The rpoE operon regulates heat stress response in *Burkholderia pseudomallei*. *FEMS Microbiol. Lett.* 284(2), 191–196. doi: 10.1111/j.1574-6968.2008.01216.x.

Vandamme, P., Peeters, C., De Smet, B., Price, E.P., Sarovich, D.S., Henry, D.A., Hird, T.J., Zlosnik, J.E.A., Mayo, M., Warner, J., *et al.* (2017). Comparative Genomics of *Burkholderia singularis* sp. nov., a Low G+C Content, Free-Living Bacterium That Defies Taxonomic Dissection of the Genus *Burkholderia*. *Front. Microbiol.* 8, 1679. doi: 10.3389/fmicb.2017.01679.

Vinion-Dubiel, A.D., and Goldberg, J.B. (2003). Lipopolysaccharide of *Burkholderia cepacia* complex. *J. Endotoxin Res.* 9(4), 201–213. doi: 10.1179/096805103225001404.

de Vries, S.P.W., Linn, A., Macleod, K., MacCallum, A., Hardy, S.P., Douce, G., Watson, E., Dagleish, M.P., Thompson, H., Stevenson, A., *et al.* (2017). Analysis of *Campylobacter jejuni* infection in the gnotobiotic piglet and genome-wide identification of bacterial factors required for infection. *Sci. Rep.* 7, 44283. doi: 10.1038/srep44283.

Wang, G. (2014). Human antimicrobial peptides and proteins. *Pharmaceuticals (Basel)* 7(5), 545–594. doi: 10.3390/ph7050545.

Wang, E.A., and Koshland, D.E. (1980). Receptor structure in the bacterial sensing system. *Proc. Natl. Acad. Sci. U.S.A.* 77(12), 7157–7161. doi: 10.1073/pnas.77.12.7157.

Wang, G., Mishra, B., Lau, K., Lushnikova, T., Golla, R., and Wang, X. (2015). Antimicrobial peptides in 2014. *Pharmaceuticals (Basel)* 8(1), 123–150. doi: 10.3390/ph8010123.

Wang, J., Weiss, B.L., and Aksoy, S. (2013). Tsetse fly microbiota: form and function. *Front. Cell. Infect. Microbiol*. 3, 69. doi: 10.3389/fcimb.2013.00069.

Wang, N., Ozer, E.A., Mandel, M.J., and Hauser, A.R. (2014). Genome-wide identification of *Acinetobacter baumannii* genes necessary for persistence in the lung. *MBio* 5(3), e01163-1114. doi: 10.1128/mBio.01163-14.

Wang, Y., Wang, L., Sun, Y., Chen, Y., Zhu, L., Guo, L., Luo, B., and Wang, H. (2007). Disrupted OmpC causes osmosis sensitivity of *Escherichia coli* in alkaline medium. *J. Genet. Genomics* 34(12), 1131–1138. doi: 10.1016/S1673-8527(07)60129-5.

Wang, Z., Wang, J., Ren, G., Li, Y., and Wang, X. (2016). Deletion of the genes *waaC*, *waaF*, or *waaG* in *Escherichia coli* W3110 disables the flagella biosynthesis. *J. Basic Microbiol.* 56(9), 1021–1035. doi: 10.1002/jobm.201600065.

Wardrope, L., Okely, E., and Leach, D. (2009). Resolution of joint molecules by RuvABC and RecG following cleavage of the *Escherichia coli* chromosome by EcoKI. *PLoS One* 4(8), e6542. doi: 10.1371/journal.pone.0006542.

Webster, R.E. (1991). The tol gene products and the import of macromolecules into *Escherichia coli*. *Mol. Microbiol.* 5(5), 1005–1011. doi: 10.1111/j.1365-2958.1991.tb01873.x.

Weerdenburg, E.M., Abdallah, A.M., Rangkuti, F., Abd El Ghany, M., Otto, T.D., Adroub, S.A., Molenaar, D., Ummels, R., ter Veen, K., van Stempvoort, G., *et al.* (2015). Genome-Wide Transposon Mutagenesis Indicates that *Mycobacterium marinum* Customizes Its Virulence Mechanisms for Survival and Replication in Different Hosts. *Infect. Immun.* 83(5), 1778–1788. doi: 10.1128/IAI.03050-14.

Weigel, W.A., and Demuth, D.R. (2016). QseBC, a two-component bacterial adrenergic receptor and global regulator of virulence in Enterobacteriaceae and Pasteurellaceae. *Mol. Oral Microbiol.* 31(5), 379–397. doi: 10.1111/omi.12138.

Weinstein-Fischer, D., Elgrably-Weiss, M., and Altuvia, S. (2000). *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. *Mol. Microbiol.* 35(6), 1413–1420. doi: 10.1046/j.1365-2958.2000.01805.x.

Welander, P.V., Doughty, D.M., Wu, C.-H., Mehay, S., Summons, R.E., and Newman, D.K. (2012). Identification and characterization of *Rhodopseudomonas palustris* TIE-1 hopanoid biosynthesis mutants. *Geobiology* 10(2), 163–177. doi: 10.1111/j.1472-4669.2011.00314.x.

Werren, J.H., Baldo, L., and Clark, M.E. (2008). *Wolbachia*: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6(10), 741–751. doi: 10.1038/nrmicro1969.

Wetmore, K.M., Price, M.N., Waters, R.J., Lamson, J.S., He, J., Hoover, C.A., Blow, M.J., Bristow, J., Butland, G., Arkin, A.P., and Deutschbauer, A. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* 6(3), e00306-315. doi: 10.1128/mBio.00306-15.

Wier, A.M., Nyholm, S.V., Mandel, M.J., Massengo-Tiassé, R.P., Schaefer, A.L., Koroleva, I., Splinter-Bondurant, S., Brown, B., Manzella, L., Snir, E., *et al.* (2010). Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 107(5), 2259–2264. doi: 10.1073/pnas.0909712107.

Wilde, A.D., Snyder, D.J., Putnam, N.E., Valentino, M.D., Hammer, N.D., Lonergan, Z.R., Hinger, S.A., Aysanoa, E.E., Blanchard, C., Dunman, P.M., *et al.* (2015). Bacterial Hypoxic Responses Revealed as Critical Determinants of the Host-Pathogen Outcome by TnSeq Analysis of *Staphylococcus aureus* Invasive Infection. *PLoS Pathog.* 11(12), e1005341. doi: 10.1371/journal.ppat.1005341.

Wiles, T.J., Norton, J.P., Russell, C.W., Dalley, B.K., Fischer, K.F., and Mulvey, M.A. (2013). Combining Quantitative Genetic Footprinting and Trait Enrichment Analysis to Identify Fitness Determinants of a Bacterial Pathogen. *PLoS Genet.* 9(8), e1003716. doi: 10.1371/journal.pgen.1003716.

Willis, L.M., and Whitfield, C. (2013). Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydr. Res.* 378, 35–44. doi: 10.1016/j.carres.2013.05.007.

Wilson, A.C.C., and Duncan, R.P. (2015). Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc. Natl. Acad. Sci. U.S.A.* 112(33), 10255–10261. doi: 10.1073/pnas.1423305112.

Wippel, K., and Long, S.R. (2016). Contributions of *Sinorhizobium meliloti* Transcriptional Regulator DksA to Bacterial Growth and Efficient Symbiosis with *Medicago sativa*. *J. Bacteriol.* 198(9), 1374–1383. doi: 10.1128/JB.00013-16.

Wong, Y.-C., Abd El Ghany, M., Naeem, R., Lee, K.-W., Tan, Y.-C., Pain, A., and Nathan, S. (2016). Candidate Essential Genes in *Burkholderia cenocepacia* J2315 Identified by Genome-Wide TraDIS. *Front. Microbiol.* 7, 1288. doi: 10.3389/fmicb.2016.01288.

Worthen, P.L., Gode, C.J., and Graf, J. (2006). Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis. *Appl. Environ. Microbiol.* 72(7), 4775–4781. doi: 10.1128/AEM.00356-06.

Wright, G.S.A., Saeki, A., Hikima, T., Nishizono, Y., Hisano, T., Kamaya, M., Nukina, K., Nishitani, H., Nakamura, H., Yamamoto, M., *et al.* (2018). Architecture of the complete oxygen-sensing FixL-FixJ

two-component signal transduction system. *Sci. Signal.* 11(525), pii: eaaq0825. doi: 10.1126/scisignal.aaq0825.

Wu, Q., Patočka, J., and Kuča, K. (2018). Insect Antimicrobial Peptides, a Mini Review. *Toxins (Basel)* 10(11), pii: E461. doi: 10.3390/toxins10110461.

Xu, Q., Mengin-Lecreulx, D., Liu, X.W., Patin, D., Farr, C.L., Grant, J.C., Chiu, H.-J., Jaroszewski, L., Knuth, M.W., Godzik, A., *et al.* (2015). Insights into Substrate Specificity of NlpC/P60 Cell Wall Hydrolases Containing Bacterial SH3 Domains. *MBio* 6(5), e02327-2314. doi: 10.1128/mBio.02327-14.

Yeh, Y.-C., Comolli, L.R., Downing, K.H., Shapiro, L., and McAdams, H.H. (2010). The caulobacter Tol-Pal complex is essential for outer membrane integrity and the positioning of a polar localization factor. *J. Bacteriol.* 192(19), 4847–4858. doi: 10.1128/JB.00607-10.

Yi, H.-Y., Chowdhury, M., Huang, Y.-D., and Yu, X.-Q. (2014). Insect antimicrobial peptides and their applications. *Appl. Microbiol. Biotechnol.* 98(13), 5807–5822. doi: 10.1007/s00253-014-5792-6.

York, G.M., Junker, B.H., Stubbe, J.A., and Sinskey, A.J. (2001). Accumulation of the PhaP phasin of Ralstonia eutropha is dependent on production of polyhydroxybutyrate in cells. *J. Bacteriol.* 183(14), 4217–4226. doi: 10.1128/JB.183.14.4217-4226.2001.

Yu, Z., Qin, W., Lin, J., Fang, S., and Qiu, J. (2015). Antibacterial mechanisms of polymyxin and bacterial resistance. *Biomed. Res. Int.* 2015, ID 679109. doi: 10.1155/2015/679109.

Yuen, C.-W., Ong, E.B.B., Mohamad, S., Manaf, U.A., and Najimudin, N. (2012). Construction and characterization of a *Burkholderia pseudomallei wzm* deletion mutant. *J. Microbiol. Biotechnol.* 22(10), 1336–1342.

Zhang, L., and Gallo, R.L. (2016). Antimicrobial peptides. *Curr. Biol.* 26(1), R14–R19. doi: 10.1016/j.cub.2015.11.017.

Zhang, C., Zhang, W., Liang, W., Shao, Y., Zhao, X., and Li, C. (2019). A sigma factor RpoD negatively regulates temperature-dependent metalloprotease expression in a pathogenic *Vibrio splendidus. Microb. Pathog.* 128, 311–316. doi: 10.1016/j.micpath.2019.01.021.

Zhang, L., Zhu, Z., Jing, H., Zhang, J., Xiong, Y., Yan, M., Gao, S., Wu, L.-F., Xu, J., and Kan, B. (2009). Pleiotropic effects of the twin-arginine translocation system on biofilm formation, colonization, and virulence in *Vibrio cholerae. BMC Microbiol.* 9, 114. doi: 10.1186/1471-2180-9-114.

Zhang, Y., Morar, M., and Ealick, S.E. (2008). Structural biology of the purine biosynthetic pathway. *Cell. Mol. Life Sci.* 65(23), 3699–3724. doi: 10.1007/s00018-008-8295-8.

Zhao, F., Lan, X.-Q., Du, Y., Chen, P.-Y., Zhao, J., Zhao, F., Lee, W.-H., and Zhang, Y. (2018). King cobra peptide OH-CATH30 as a potential candidate drug through clinic drug-resistant isolates. *Zool. Res.* 39(2), 87–96. doi: 10.24272/j.issn.2095-8137.2018.025.

Zheng, H., Powell, J.E., Steele, M.I., Dietrich, C., and Moran, N.A. (2017). Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proc. Natl. Acad. Sci. U.S.A.* 114(18), 4775–4780. doi: 10.1073/pnas.1701819114.

Zhu, S., and Gao, B. (2013). Evolutionary origin of β-defensins. *Dev. Comp. Immunol.* 39(1-2), 79–84. doi: 10.1016/j.dci.2012.02.011.

Zilber-Rosenberg, I., and Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* 32(5), 723–735. doi: 10.1111/j.1574-6976.2008.00123.x.

Publications

The biotroph *Agrobacterium tumefaciens* thrives in tumors by exploiting a wide spectrum of plant host metabolites

Almudena Gonzalez-Mula, Joy Lachat, Léo Mathias, Delphine Naquin, Florian Lamouche, Peter Mergaert and Denis Faure

Institute for Integrative Biology of the Cell (I2BC), CNRS CEA University Paris-Sud, University Paris-Saclay, Gif-sur-Yvette F-91190, France

Author for correspondence:
Denis Faure
Tel: +33 169823498
Email: *denis.faure@i2bc.paris-saclay.fr*

Received: *22 August 2018*
Accepted: *3 November 2018*

New Phytologist (2019)
doi: 10.1111/nph.15598

Key words: biotroph, ecological niche, fitness, plant pathogen, transcriptomics, transposon-sequencing (Tn-seq).

Summary

- *Agrobacterium tumefaciens* is a niche-constructing biotroph that exploits host plant metabolites.
- We combined metabolomics, transposon-sequencing (Tn-seq), transcriptomics, and reverse genetics to characterize *A. tumefaciens* pathways involved in the exploitation of resources from the *Solanum lycopersicum* host plant.
- Metabolomics of healthy stems and plant tumors revealed the common (e.g. sucrose, glutamate) and enriched (e.g. opines, γ-aminobutyric acid (GABA), γ-hydroxybutyric acid (GHB), pyruvate) metabolites that *A. tumefaciens* could use as nutrients. Tn-seq and transcriptomics pinpointed the genes that are crucial and/or upregulated when the pathogen grew on either sucrose (*pgi*, *kdgA*, *pycA*, *cisY*) or GHB (*blcAB*, *pckA*, *eno*, *gpsA*) as a carbon source. While sucrose assimilation involved the Entner–Doudoroff and tricarboxylic acid (TCA) pathways, GHB degradation required the *blc* genes, TCA cycle, and gluconeogenesis. The tumor-enriched metabolite pyruvate is at the node connecting these pathways. Using reverse genetics, we showed that the *blc*, *pckA*, and *pycA* loci were important for aggressiveness (tumor weight), proliferation (bacterial charge), and/or fitness (competition between the constructed mutants and wild-type) of *A. tumefaciens* in plant tumors.
- This work highlighted how a biotroph mobilizes its central metabolism for exploiting a wide diversity of resources in a plant host. It further shows the complementarity of functional genome-wide scans by transcriptomics and Tn-seq to decipher the lifestyle of a plant pathogen.

Introduction

Hosts and microbes evolved a wide spectrum of biological interactions, ranging from pathogenesis to symbiosis. To succeed in their lifestyle, host-interacting microbes are able to escape host defense, overcome competition with other microbiota members, and exploit nutrients available in the hosts. Ecological niche construction, which ensures a preferred access to host-derived resources, represents a recurrent strategy in pathogens and symbionts (Kylafis & Loreau, 2011; McNally & Brown, 2015; Martin *et al.*, 2017; Poole *et al.*, 2018). Identifying metabolic pathways involved in the exploitation of resources and evaluating their involvement in the fitness of microbes represent important issues in ecology and evolution for understanding adaptation of microbes to the hosts, with applied perspectives in plant, animal, and human health.

Different strategies have emerged to identify the microbial pathways involved in resource exploitation. They basically employ a two-step methodology. The first step is the identification of candidate genes and pathways by different genome-wide scans (functional screening of individual mutants, transcriptomics, genomics, genome wide association, etc.) in microbes that exploit a given resource using, in some instances, a comparison with microbes that do not exploit it. The second step is the validation of a fitness trait by confronting microbes carrying allelic variation (natural or constructed variants) in those candidate genes and pathways. Because of its relative simplicity, transposon sequencing (Tn-seq), which combines transposon insertional mutagenesis with massively parallel sequencing of the transposon insertion sites in transposon mutant populations grown in control and test conditions, seemed an attractive approach to examine ecologically important genes and pathways in prokaryotic and eukaryotic microbes (van Opijnen & Camilli, 2013). In this study, we combined plant metabolomics and two functional genome-wide scans (transcriptomics and Tn-seq) for identifying genes and pathways involved in the exploitation of the *Solanum lycopersicum* host by the biotrophic pathogen *Agrobacterium tumefaciens*.

A. tumefaciens is a niche-constructing pathogen that genetically modifies the plant host genome by transferring a part (the transfer DNA (T-DNA)) of its virulence Ti plasmid (Barton *et al.*, 2018; Dessaux & Faure, 2018). When expressed into the plant cell nucleus, the T-DNA genes divert the host hormonal and

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

metabolic pathways to provoke the development of galls or plant tumors (Deeken *et al.*, 2006). In previous work, we paid attention to specific metabolites, the opines, that accumulate in the *A. tumefaciens*-infected plant tumors (Lang *et al.*, 2014; El Sahili *et al.*, 2015; Marty *et al.*, 2016; Tannières *et al.*, 2017; Lang *et al.*, 2017; Vigouroux *et al.*, 2017). Opines, such as agrocinopines, mannopine, nopaline, and octopine, result from the condensation of sugars and amino and organic acids (Dessaux *et al.*, 1993). According to chemical and genome databases, the opines synthesized by *Agrobacterium* T-DNA-encoded enzymes are almost exclusively produced by host plants infected by *A. tumefaciens*. To our knowledge, the only reported exception is the opine octopine that is also produced in the muscle of the marine animal octopus (Fields *et al.*, 1976). The biosynthesis of opines in the chimeric plant cells expressing bacterial T-DNA could be considered as a biological innovation resulting from the holobiont assembly (Faure *et al.*, 2018). In the cases of nopaline and octopine, we showed that opine assimilation confers a selective advantage when *A. tumefaciens* populations colonize the plant tumors (Lang *et al.*, 2014, 2017; Vigouroux *et al.*, 2017). Aside from opines, diverse metabolites accumulate in the tumors on *Arabidopsis thaliana* and *Brassica rapa* (Deeken *et al.*, 2006; Simoh *et al.*, 2009; Lang *et al.*, 2016), but their contribution to *A. tumefaciens* fitness and proliferation is poorly documented. Recently, the transcriptome of *A. tumefaciens* C58 living in *A. thaliana* tumors highlighted considerable changes in gene expression profile compared with a culture in a synthetic medium (González-Mula *et al.*, 2018). In addition to the opines, the transcriptomic data suggested the exploitation of a wide diversity of resources by *A. tumefaciens*, but direct evidence of the contribution of these different metabolites to the *Agrobacterium* fitness in plant tumors was still missing.

In this work, metabolomics indeed revealed the presence of a wide spectrum of potential resources in *S. lycopersicum* tumors, including metabolites that were enriched compared with uninfected stems. We combined Tn-seq and transcriptomics to investigate the *A. tumefaciens* pathways for exploiting the three metabolites sucrose, γ-hydroxybutyrate, and γ-aminobutyrate that accumulated at different levels in plant tumors. Finally, we used reverse genetics and host plant infections to measure the aggressiveness, proliferation, and competitive fitness conferred by assimilation of these metabolites when *A. tumefaciens* colonized the plant tumor niche. We showed that the ecological success of the *A. tumefaciens* biotroph resulted from its capacity to exploit a wider spectrum of host metabolites than the sole opines. This work also highlighted the strength and limits of Tn-seq and transcriptomics to decipher the microbial genetic determinants that are involved in ecological niche exploitation.

Materials and Methods

Bacterial strains and culture conditions

We used *A. tumefaciens* C58, the genome of which was sequenced in 2001 (Goodner *et al.*, 2001; Wood *et al.*, 2001). The kanamycin (Km)-resistance and gentamicin (Gm)-resistance cassettes (Dennis & Zylstra, 1998) were used for the construction of the knockout (KO) mutants. The *atu0035* (*pckA*), *atu2726* (*pycA*), *atu3706*, and *atu4761* genes were cloned into the pGEM-T Easy vector (Promega), and the mutated alleles were created by inserting an antibiotic-resistance cassette in a unique restriction site of the open reading frame. The constructed plasmids were electroporated in *A. tumefaciens* C58. Marker exchange was selected using Gm or Km resistance and verified by PCR. Previously constructed *A. tumefaciens* C58 mutants were also used in this study: the derivatives C107-Gm and C107-Km in which the Gm and Km cassettes were cloned in a noncoding region of the Ti plasmid (Haudecoeur *et al.*, 2009a) and the △*blcRABC* mutant in which the *blcRABC* operon was replaced by the Km-resistance cassette (Carlier *et al.*, 2004).

A. tumefaciens was cultivated at 28°C in TY medium (Bacto tryptone, 5 g l^{-1}; yeast extract, 3 g l^{-1}; agar, 15 g l^{-1}) or *Agrobacterium* broth (AB) minimal medium (dipotassium hydrogen phosphate, 3 g l^{-1}; sodium dihydrogen phosphate, 1 g l^{-1}; magnesium sulfate heptahydrate, 0.3 g l^{-1}; potassium chloride, 0.15 g l^{-1}; calcium chloride, 0.01 g l^{-1}; ferrous sulfate heptahydrate, 2.5 mg l^{-1}; pH 7) (Chilton *et al.*, 1974) supplemented with sucrose or γ-hydroxybutyric acid (GHB) at 10 mM as carbon (C) source, and ammonium chloride (NH_4Cl) or γ-aminobutyric acid (GABA) at 20 mM as nitrogen (N) source. *Escherichia coli* MFD*pir* harboring the pSAM_DGm plasmid (Skurnik *et al.*, 2013), auxotroph for diaminopimelic acid, was used as transposon donor for mutagenesis. *E. coli* DH5α was the routine host for cloning. *E. coli* strains were cultivated at 37°C in lysogenic broth modified medium (LBm; 10 g l^{-1} peptone, 5 g l^{-1} yeast extract, sodium chloride (NaCl) 5 g l^{-1}). Media were supplemented when appropriate with Gm (25 μg ml^{-1}), ampicillin (50 μg ml^{-1}), rifampicin (100 μg ml^{-1}), and diaminopimelic acid (300 μg ml^{-1}).

Plant culture, metabolomics, and infection assays

S. lycopersicum (Dona hybrid F_1, Vilmorin, France) plants were cultivated in a glasshouse under long day conditions and controlled temperature (24–26°C). Four-week-old plants were incised with a scalpel between the first and second nodes and infected by *c.* 10^7 *A. tumefaciens* cells as described by Planamente *et al.* (2010). Plant tumors were collected 4 wk after infection.

For plant metabolomics, tumors and wounded but not infected stems were directly frozen in liquid N_2, crushed, extracted and analyzed by gas chromatography–time of flight mass spectrometry (GC–TOF-MS) at the *Plateforme de Chimie du Végétal* (Versailles, France). The method was previously described in detail by Lang *et al.* (2016). Approx. 150 compounds were searched and 130 compounds, including the opines nopaline and agrocinopine A, and GABA and GHB were detected and quantified in three biological replicates of plant tumors and uninfected stems.

For virulence and fitness assays, eight plant tumors were crushed in a 0.8% NaCl solution to recover the agrobacteria, which were then spotted onto selective agar media to enumerate colony forming units (CFU). In the case of mixed infections, the

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

proportions of the genotypes (wild-type and KO alleles) were measured by testing *c.* 96 CFU. Using appropriate primers (Supporting Information Table S1), length of the PCR products distinguished wild-type allele from the KO-alleles in which the resistance gene cassette was inserted. This permitted calculation of competitive index values as previously described (Macho *et al.*, 2010). Two independent assays (eight plants for each of the assays) were carried out for each virulence and fitness assays. A Mann–Whitney test was used to analyze the values from the two independent experiments (the null hypothesis postulates that both experiments were comparable). If no difference was detected, the values were pooled and a nonparametric Kruskal–Wallis test ($P < 0.05$) coupled with a post-hoc Tukey test ($P < 0.05$) was performed.

Transposon library construction and use

A. tumefaciens C58 was mutagenized using a *Himar1* mariner transposon carrying a Gm resistance cassette. The pSAM_DGm plasmid donor *E. coli* MFDpir and *A. tumefaciens* C58 rifampicin-resistant recipient were cultivated separately: *A. tumefaciens* C58 Rif^R overnight in TY medium and *E. coli* MFDpir (pSAM_DGm) for 4 h in LBm supplemented with 300 μg ml^{-1} diaminopimelic acid. Both cultures were centrifuged and adjusted to 1 unit of OD_{600}. Equivalent volumes (0.4 ml) of cell suspensions were mixed, centrifuged, and suspended in TY with diaminopimelic acid. The cell mixture (0.2 ml) was deposited on a nitrocellulose filter (0.45 μm diameter, Millipore) on a TY agar plate and incubated overnight at 28°C. Bacterial cells were removed from the filter, suspended in 0.8% NaCl solution, and then plated on TY medium supplemented with rifampicin and Gm. Serial dilutions and plating were performed to determine the number of mutants obtained. After 72 h of incubation, mutants were collected. The mutant population was homogenized, aliquoted, and stored at −80°C in 25% (v/v) glycerol.

Four aliquots of the *A. tumefaciens* mutant library were thawed and cultured in liquid TY medium (4 h at 28°C) to revive them. Bacteria were washed twice with 0.8% NaCl solution and used to inoculate AB medium (10 ml) at an initial OD_{600} of 0.05. AB medium was supplemented with three combinations of C and N sources: sucrose–ammonium (NH_4), GHB–NH_4, and sucrose–GABA. After growth at 28°C for 24 h, bacterial cells were centrifuged and stored at −20°C for further DNA manipulation.

Transposon library sequencing and Artist analysis

Genomic DNA of mutant populations grown in TY medium or AB medium with the C and N sources tested was extracted using the DNeasy Blood & Tissue Kit (Qiagen). DNA (2 μg) was digested with the *Mme*I type II restriction–modification enzyme (BioLabs, Evry, France) for 1 h at 37°C. Digested DNA was incubated for 1 h at 37°C with FastAP thermosensitive alkaline phosphatase (Thermo Scientific, Waltham, MA, USA) followed by enzyme inactivation by heating at 75°C for 5 min. Digested DNA samples were purified using a QIAquick PCR purification kit (Qiagen) and were ligated to the p-adapters (Table S1) in the presence of Thermo Scientific T4 DNA ligase (16 h at 16°C). The p-adaptors contain a five-nucleotide long barcode that is specific for each experiment. The ligation products were used as templates to perform a PCR amplification with Illumina-primers P7 and P5 (Table S1). The PCR products of *c.* 130 base pairs, which contain the transposon insertion site, were separated on agarose gel and purified with the QIAquick gel extraction kit (Qiagen). The final samples were mixed in equimolar amounts and sequenced on an Illumina NextSeq 500 instrument (Illumina, San Diego, CA, USA), in a paired-end 2 × 75 run at the I2BC-sequencing platform (Gif-sur-Yvette, France).

The experiment-specific barcodes enabled the attribution of each sequence read to the corresponding experiment. The data generated were demultiplexed using bcl2fastq2 v.2.15.0 (Illumina) and Fastx-Toolkit software (http://hannonlab.cshl.edu/fastx_toolkit/). Only read 1 from each sequenced fragment has been used. The 3′ transposon sequence was trimmed using Trimmomatic (Bolger *et al.*, 2014), and reads with a length of 75 nucleotides were removed (reads without the transposon insertion). After the trimming step, reads with a length between 19 and 23 bp were reverse-complemented and only the reads starting with TA were mapped using Bowtie (Bowtie-1.1.2) (Langmead *et al.*, 2009) to the genome of *A. tumefaciens* C58. The *.bam output files were sorted with SAMtools (http://www.htslib.org/). FeatureCounts (Liao *et al.*, 2014) was used to evaluate the number of reads by gene or by coding sequence.

The mapping results (*.bam files) were analyzed by the Artist pipeline (Pritchard *et al.*, 2014) using Matlab software (The MathWorks, Natick, MA, USA). Artist compares the observed (reads) and predicted numbers of transposons at each of the 115 525 insertion sites (TA dinucleotides) along the *A. tumefaciens* C58 genome. Two different analyses were carried out: El-Artist (essential loci analysis) and Con-Artist (conditionally essential loci analysis). El-Artist searches for a nonrandom distribution of transposon insertions in the constructed mutant library in TY medium. Hence, it identifies all loci that are required for an optimal growth in the initial culture condition. A gene is annotated as 'essential' when there are a low number of transposon insertions (reads) or no associated transposon insertions within the entire gene. In El-Artist, 0.03 is the *P*-value threshold for calling a region significantly underrepresented in reads. Then, Con-Artist was applied to compare the distribution of transposon insertions between the initial TY culture condition and each of the three AB medium conditions. In Con-Artist, 0.01 is the *P*-value cutoff in the Mann–Whitney *U* test for defining genes with significantly different read numbers. This allowed the identification of *A. tumefaciens* genes and pathways that were required for efficient proliferation in the presence of sucrose–NH_4, GHB–NH_4, and sucrose–GABA as nutrients.

Transcriptomics and DESeq2 analysis

An overnight culture of *A. tumefaciens* C58 was grown in AB medium with sucrose and NH_4 as sources of C and N

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

respectively. This culture was washed twice with NaCl 0.8% and served to inoculate AB medium supplemented with the three different combinations of C and N sources as earlier for the Tn-seq experiments: sucrose–NH_4, GHB–NH_4, and sucrose–GABA. Inoculations (at $OD_{600} = 0.05$) were performed in triplicate. At exponential phase (at $OD_{600} = 0.30$), bacterial cultures were centrifuged and RNA extracted with the MasterPure™ Complete DNA and RNA Purification Kit according to the supplier's instructions. RNA-sequencing (RNA-seq) libraries were constructed using the Ribo-Zero and ScriptSeq-V2 kits (Illumina). Libraries were sequenced on a NextSeq 500 instrument (Illumina) at the I2BC platform (Gif-sur-Yvette, France) using the 75-cycles NextSeq 500 High Output Kit. Count tables have been filtered to retain only genes with a gene count over 1 count per million in half of the samples of the dataset. Normalization and differential analyses were performed using generalized linear models as described in the DESeq2 package (v.1.12.4) (Love *et al.*, 2014). The cutoff chosen for differentially expressed genes are a false discovery rate < 0.01 and a $\log_2$ fold change > 2. RNA-seq data from this article were deposited at https://www.ncbi.nlm.nih.gov/geo/info/linking.html using accession GSE121889, according to Minimum Information About a Microarray Experiment standards.

Gene expression was also measured by quantitative PCR (qPCR) using dedicated primers (Table S1). The cDNA was prepared from 1 µg of bacterial RNA using the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas, Saint-Remy-les-Chevreuses, France) following the manufacturer's instructions. The qPCRs were performed with a Lightcycler 96 (Roche) apparatus. The data were processed using the $2^{-\Delta\Delta C_T}$ method (Livak & Schmittgen, 2001) and compared with the expression profile acquired from the RNA-seq transcriptome. The internal controls used were the *atu1789* (for the GHB condition) and *nocT* (*atu6027*, for the GABA condition) genes.

Data accessibility

All the data are available in Supporting Information for the article.

Results

Metabolic resources in the plant tumor niche

The abundance of 130 compounds was quantified in uninfected stems of *S. lycopersicum* and tumors induced by *A. tumefaciens* C58 (Fig. 1; Table S2). In plant tumors, the 13 most abundant compounds accounted for 97% of the relative abundance of all the compounds quantified. In decreasing order, these are propanediol, glucose, malate, dehydroascorbate, fructose, phosphate, sucrose, glutamine, glutamate, *myo*-inositol, citrate, asparagine, and aspartate. They were also present in healthy stems at a similar level. Using a threshold fold change value (≥ 4), 24 compounds were enriched in plant tumors compared with uninfected stems (Fig. 1). As expected, the tumor-enriched compounds encompassed the two opines nopaline and agrocinopine, but some other metabolites were also remarkable. Six metabolites, agrocinopine, *trans*-ferulate, nicotianamine, pyruvate, spermidine, and succinic semialdehyde (SSA), exhibited an enrichment by four to six orders in plant tumors. Their accumulation reflected some well-known characteristics of the plant tumors: obviously, the opine synthesis driven by T-DNA; but also the activation of plant defense as revealed by accumulation of *trans*-ferulate phenolics as antimicrobial compounds and nicotianamine as iron chelator (Deeken *et al.*, 2006; Aznar *et al.*, 2015); a response to abiotic (hypoxia and drought) and biotic stresses, as shown by the accumulation of SSA and spermidine that are related to the GABA pathway (Lang *et al.*, 2016; Podlešáková *et al.*, 2019); and a shift to an anaerobic and heterotrophic metabolism, as suggested by pyruvate accumulation (Deeken *et al.*, 2006).

GABA, SSA, and GHB were all enriched metabolites in plant tumors (Fig. 1). They are metabolically connected and, together with proline, are involved in regulation of the quorum-sensing signal-degrading lactonase BlcC in *A. tumefaciens* (Carlier *et al.*, 2004; Chevrot *et al.*, 2006; Chai *et al.*, 2007; Haudecoeur *et al.*, 2009b; Lang *et al.*, 2016). Noticeably, GABA and proline were the two enriched metabolites that accumulated at the highest concentrations (Fig. 1). As osmoprotectants, proline and sucrose (the latter was not enriched but very abundant) were proposed to contribute to desiccation resistance in plant tumors (Wächter *et al.*, 2003).

In the next part of this work, we focused on compounds other than the opines, and investigated how *A. tumefaciens* could be able to use them as a resource. Because of its role in tumor development, we chose sucrose as a representative of the abundant class of metabolites. We also studied two structurally and functionally related metabolites, GABA and GHB, which were enriched in plant tumors, either at a high concentration (GABA) or at a lower concentration (GHB).

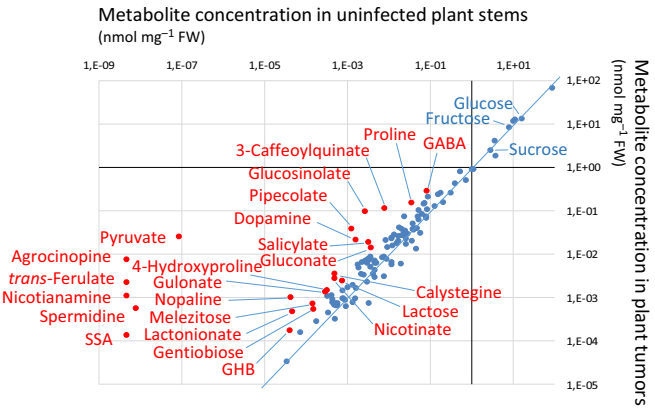


Fig. 1 Metabolome of *Solanum lycopersicum* healthy stems vs *Agrobacterium tumefaciens* C58 tumors. The graphic shows the relative abundance of compounds found in *A. tumefaciens* tumors when compared with uninfected tissue. Red dots indicate metabolites enriched in plant tumors compared with uninfected stems (fold change value ≥ 4); blue dots are the other plant metabolites. GABA, γ-aminobutyric acid; GHB, γ-hydroxybutyric acid; SSA, succinic semialdehyde.

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

Genome-wide mutant library of *A. tumefaciens*

The *A. tumefaciens* C58 genome contained 115 525 TA dinucleotides that are potential insertion sites of the *Himar1* mariner transposon. They are positioned along the circular chromosome (55 348 TA within 2 841 580 bp), linear chromosome (41 503 TA within 2 075 577 bp), and the two plasmids, the pAt (13 084 TA within 542 868 bp) and pTi (5 590 TA within 214 233 bp). From 55 matings between the *E. coli* transposon donor and *A. tumefaciens* C58 recipient, we collected 1.1×10^6 mutant colonies on TY medium supplemented with rifampicin and Gm, hence *c.* 10-fold more than the TA site number in *A. tumefaciens*. After homogenization of all mutant colonies, around 3×10^{10} individuals were kept in each frozen aliquot.

For analyzing the constructed transposon mutant population, four mutant library aliquots were cultivated for 4 h in TY medium. Total DNA was extracted and transposon insertion sites were sequenced. Between 4 million and 8 million filtered reads were obtained for each replicate. When the replicates were compared, transposon distribution revealed a high homogeneity ($r^2 > 0.98$); hence, all the sequencing reads were analyzed together by EL-ARTIST. Mutants in most genes (4730) (Fig. 2a; Table S3.1) were unaffected in their fitness ('nonessential gene' according to the EL-ARTIST classification). Mutants in 513 genes ('essential genes' according to the EL-ARTIST classification) were impaired in their fitness for growth in the rich TY medium (Fig. 2c; Table S3.1). Some other genes (105) (Fig. 2b; Table S3.1) contained a domain in which transposon insertion provoked a decrease of the fitness (genes with an 'essential domain' according to the EL-ARTIST classification). The 513 fitness genes in TY medium represented *c.* 10% of the total genes of *A. tumefaciens* C58, a similar percentage as reported in other bacteria (Christen *et al.*, 2011; DeJesus & Ioerger, 2013). Most of them (428, hence 83%) were located on the circular chromosome (Table S3.1). Using a Tn*5* mutant library, Curtis & Brun (2014) reported 372 essential genes in *A. tumefaciens* C58. Even if the two approaches were different in the choice of transposons, growth condition, library sequencing, and data analysis, most of the Tn*5*-picked essential genes (307 of 372) were also present in the list of the *Himar1*-identified mutants (Table S3; Fig. S1), hence consolidating the two approaches.

The 513 fitness genes were classified according to clusters of orthologous genes (COG) (Tatusov *et al.*, 2000). The functional category most represented was that of translation, ribosomal structure, and biogenesis (Fig. S2; Table S3.1). Genes coding for some ribosomal proteins (*atu1928–atu1951*) exemplified this COG category. Genes involved in the COG category energy production and conversion were also found to be abundant. This is the case of the *nuoABCDEFGHIJKLMN* (*atu1268–atu1283*) operon involved in respiration. Noticeably, the Tn-seq approach revealed some genes that are not essential for cell viability but essential for the maintenance of the *A. tumefaciens* C58 plasmids, such as the operons *repABC* in the At (*atu5000–atu5002*) and pTi (*atu6043–atu6045*) plasmids. This is explained by a transposon insertion in these replicative functions causing the loss of the respective plasmid and, hence, after growth, the disappearance of these mutants in the mutagenized population.

A. tumefaciens key-genes for exploiting sucrose, GHB, and GABA

The transposon mutant population was cultivated in a minimal medium for searching the genes associated with exploitation of either sucrose or GHB as a C source (with NH_4 as a N source) and GABA as a N source (with sucrose as a C source). For each of the bacterial culture replicates, between 3 million and 6 million filtered reads were obtained and analyzed by ARTIST. Among replicates of a same condition, transposon distribution was highly correlated ($r^2 > 0.93$); hence, reads of a same condition were pooled. By comparing the transposon distribution between the initial growth condition in TY rich medium and the three culture conditions in minimal media, CON-ARTIST revealed 69, 37, and 47 genes of which mutants were impaired for growth in the presence of sucrose–NH_4, GHB–NH_4, and sucrose–GABA, respectively (Fig. 3; Table S3.2–S3.4). Most of them are involved in amino acid and nucleobase biosynthesis and were shared between the conditions tested. This was expected because cultures in AB minimal medium were compared with an initial culture in TY rich medium. We focused on genes that are specific to each of the growth conditions: there were 28 genes identified in the sucrose–NH_4 condition, 11 in the GHB–NH_4 condition, and nine in the sucrose–GABA condition.

In the presence of sucrose and NH_4 as nutrients, the noticeable genes were *pgi* (*atu0404*) coding for glucose-6-phosphate isomerase, *pycA* (*atu2726*) allowing conversion of pyruvate into oxaloacetate, *cisY* (*atu1392*) for conversion of oxaloacetate into citrate, and the *sdhCDA* (*atu2643–atu2645*) genes for malate conversion and energy production (Table S3.2). These genes are pivotal for the entry of C compounds into the Entner–Doudoroff pathway and the tricarboxylic acid (TCA) cycle. When *A. tumefaciens* grew on GHB–NH_4, key fitness genes were *blcAB* (*atu5137–atu5138*) coding for the conversion of GHB into succinate, as well as *sdhDC*, *pckA* (*atu0035*) and *eno* (*atu1426*) for connecting the TCA cycle and gluconeogenesis, and *gpsA* (*atu2650*) that links gluconeogenesis and lipid biosynthesis (Table S3.3). In the presence of GABA as an N source, we did not identify the expected GABA transaminase key gene that could convert GABA into succinic semialdehyde, probably because of functional redundancy.

A. tumefaciens transcriptomes during growth on sucrose, GHB, and GABA

Using the same minimal media supplemented with sucrose and NH_4, GHB and NH_4, and sucrose and GABA, we produced RNA-seq transcriptomes of *A. tumefaciens* C58 under exponential growth culture condition. RNA-seq trancriptomic data were validated by qPCR assays on a set of nine genes (Fig. S3). In the GHB–NH_4 vs sucrose–NH_4 transcriptome comparison (Fig. 4; Table S4.1), 203 genes were differentially expressed ($\log_2$ fold change > 2; $P < 0.05$). Among them, 109 genes were upregulated

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

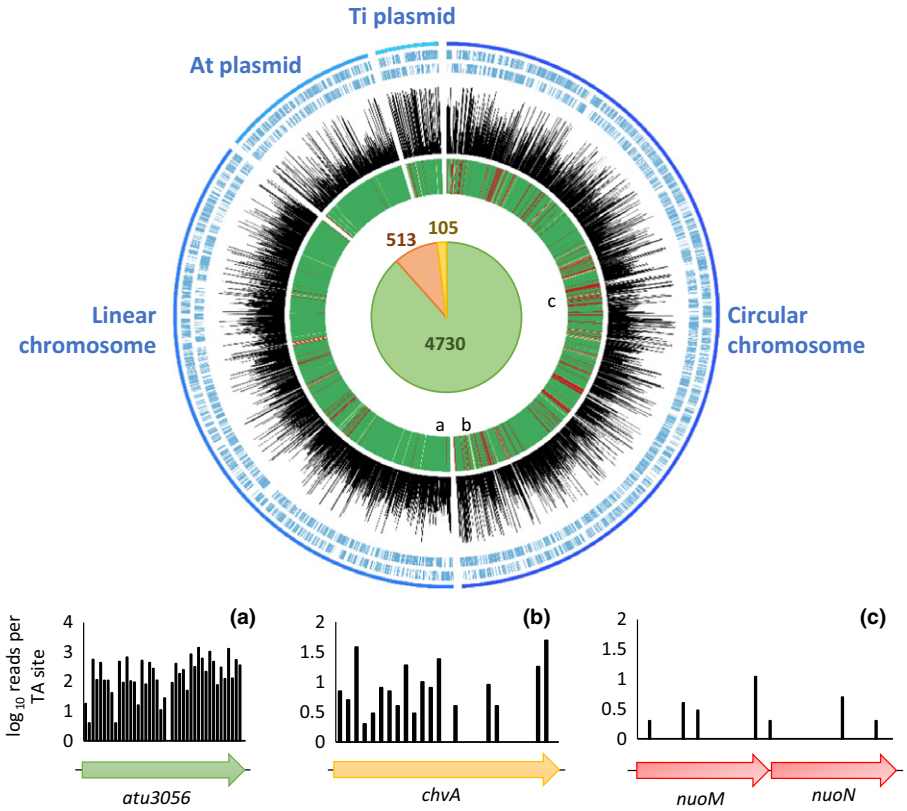


Fig. 2 Genome of *Agrobacterium tumefaciens* C58 with locations of transposon insertions in the constructed transposon library. From the outside to the inside, the tracks represent: forward and reverse coding sequences (in blue), number of transposon insertions per TA site for each gene expressed in $\log_{10}$ (in black), and EL-ARTIST analysis in which nonessential genes are in green, genes with an essential domain are in yellow, and essential genes are in red. The circle chart shows the total number of nonessential genes (in green), essential genes (in red), and genes with a domain essential (in yellow). Panels (a), (b) and (c) exemplify these three categories of genes, showing the number of transposon insertions per TA sites in each of the genes *atu3056*, *chvA*, *nuoM* and *nuoN*.

and 94 were downregulated in the GHB condition. In the GHB condition, the top 10 of the highest upregulated genes ($\log_2$ fold change between 6.44 and 4.17) encompassed the *blcABC* operon and *pckA* gene, which, except for the lactonase-encoding *blcC* gene, were all also identified by Tn-seq as crucial under GHB assimilation. The *blcC* gene encodes a lactonase that is involved in γ-butyrolactone cleavage, but not in GHB degradation (Carlier *et al.*, 2004; Chai *et al.*, 2007). Other remarkable upregulated genes were *sdhCD* (also revealed by Tn-seq), *atu3740* and *pfp* (*atu2115*) encoding two successive steps converting glyceraldehyde 3-phosphate to fructose 6-phosphate in gluconeogenesis, and *dctA* (*atu3298*) coding for a transporter of C_4-dicarboxylic acids. Most of the other upregulated genes belong to the COG category of energy production and conversion, including oxidative phosphorylation pathways (*cyd* and *fix* genes) and nitrate reductase (*nap* genes).

Considering the upregulated genes in the sucrose condition, the most remarkable gene was *kdgA* (*atu4494*) that is coding for the last step of the Entner–Doudoroff pathway. Some others were involved in sugar uptake, such as the *agl* genes (*atu0590–atu0594* coding for a transcriptional regulator, a sugar ATP-binding cassette transporter and a glucosidase) and the *rbs* genes (*atu4369–atu4372* coding a sugar ATP-binding cassette transporter). Most of the other upregulated genes belonged to the COG category inorganic ion transport and metabolism, including iron siderophore synthesis (*atu3670–atu3673* and *atu3675–atu3685*) and uptake (*atu5311–atu5316*) and copper resistance genes (*atu3990–atu3992*). The differentially expressed genes related to C conversion from GHB and sucrose were positioned in a simplified scheme of metabolic pathways (Fig. 5).

When the sucrose–GABA vs sucrose–NH_4 transcriptomes were compared, 163 genes were differentially expressed ($\log_2$ fold change > 2; $P < 0.05$). Most of them (109) were upregulated in the GABA–sucrose condition. In the top five of the highest upregulated genes ($\log_2$ fold change between 5.4 and 4.8) were the aforementioned *blcABC* genes (Fig. 4; Table S4.2). Among the upregulated genes, we searched for putative GABA-transaminase genes coding for the conversion of GABA to SSA. We found two candidate genes, *atu4761* and *atu3407*, highlighting a potential redundancy of this activity in *A. tumefaciens* C58. The gene *atu4761* was co-expressed with *atu4762*, a *blcA* paralogous gene coding for an SSA dehydrogenase. Most of the other upregulated genes belonged to COG category amino acid transport and metabolism, including several transporters (*amtB* = *atu2758*; genes *atu1387–atu1391* and *atu3903–atu3905*) and regulatory proteins (*glnK* = *atu2757*). Remarkably, expression of the genes coding catalase KatE (*atu5491*) and superoxide dismutase SodB (*atu4726*) was also enhanced, indicating a response to an oxidative stress.

Validation of *A. tumefaciens* fitness traits when exploiting the host plant

We constructed single and double mutants of the genes *atu4761* and *atu3407* coding for the putative GABA transaminases. None of these mutants was impaired for growth on GABA as a sole N source (Fig. S4), suggesting that either they are not coding for

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

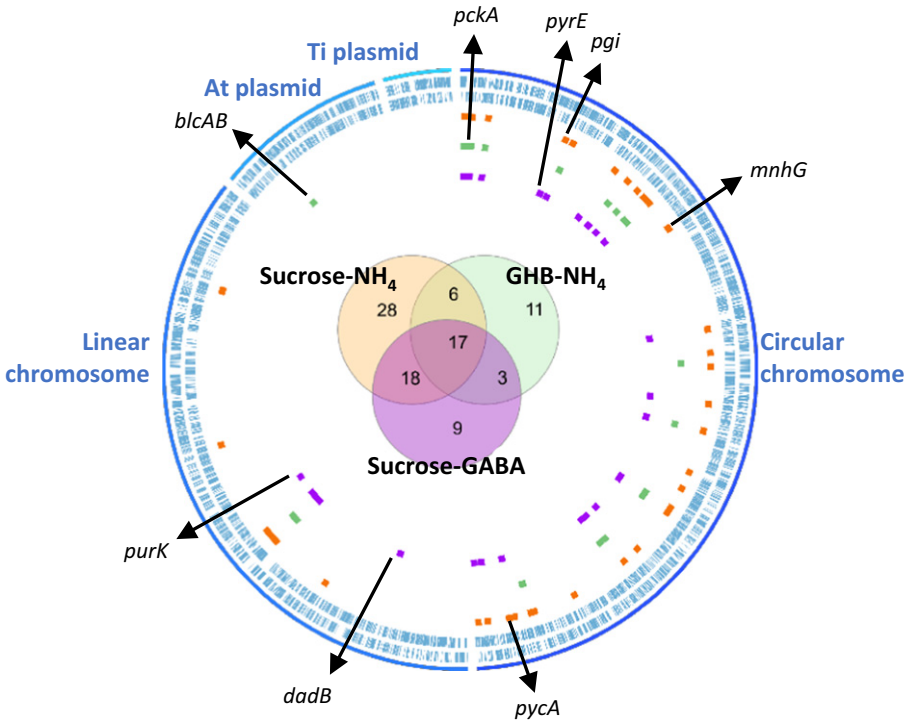


Fig. 3 *Agrobacterium tumefaciens* C58 fitness genes for growth with different carbon and nitrogen sources. From the outside to the inside, the tracks represent: forward and reverse coding sequence (in blue) and fitness genes ('essential genes' according to Con-Artist analysis) when *A. tumefaciens* grew in the presence of sucrose and ammonium (NH_4, as chloride; in orange), γ-hydroxybutyric acid (GHB) and NH_4 (in green), and sucrose and γ-aminobutyric acid (GABA; in purple). The Venn diagram represents a comparison of fitness genes in the three conditions. Some examples of fitness genes are indicated.

GABA transaminase or their mutation was compensated by one (or more) other genes expressing GABA transaminase activity.

We pursued our investigations on C metabolism by comparing GHB and sucrose pathways as archetypes of two major C entries (the TCA cycle and Entner–Doudoroff pathway) in *A. tumefaciens* when it lives on the host plant. Based on the Tn-seq and transcriptomics data, we constructed two *A. tumefaciens* C58 mutants, *pckA::Gm* and *pycA::Gm*, which are affected in pivotal reactions connecting the TCA cycle to gluconeogenesis and Entner–Doudoroff pathways. We also used an already constructed mutant *blcRCAB::Km* deleted for the *blcRCAB* gene cluster (Carlier *et al.*, 2004). Two other *A. tumefaciens* C58 derivatives, 107-Km and 107-Gm, carrying a Km- or Gm-resistance cassette, respectively, in the same noncoding region were used as controls (Haudecoeur *et al.*, 2009a). We verified that the *pycA::Gm* was impaired for growth on sucrose, fructose, and glucose, and the *pckA::Gm* mutant on GHB, succinate, and nopaline as C source. Growth of both mutants was impaired on pyruvate. The mutant *blcRCAB::Km* was only impaired in the GHB assimilation (Fig. 6a). The control strains 107-Km and 107-Gm grew on all C sources.

All these mutants were tested for aggressiveness (tumor weight), proliferation (bacterial charge), and competitive fitness (against wild-type allele) on the tomato host plant. For single strain inoculation experiments, the weight and bacterial charge of 5-wk-old tumors were measured (Fig. 6b,c). The *A. tumefaciens* derivatives 107-Km and 107-Gm exhibited similar traits on the host plant and were used as control conditions. When *blcRCAB::Km* and 107-Km derivatives, which harbor the same Km-resistance cassette, were compared, a decrease of both tumor weight and pathogen charge were observed in the *blcRCAB* mutant. When *pycA::Gm*, *pckA::Gm*, and 107-Gm mutants were compared, a decrease of tumor weight and bacterial charge was observed in the *pycA::Gm* mutant only. A previous study reported a decreased aggressiveness of a *pckA* mutant (Liu *et al.*, 2005), but the virulence assay conditions, and hence resource availability, could explain this discrepancy: stem of entire tomato plants (our study) vs tobacco leaf disks (Liu *et al.*, 2005).

Dual competitions were performed for evaluating the fitness of the *blcRCAB::Km*, *pycA::Gm*, and *pckA::Gm* mutants compared with the control derivatives 107-Km or 107-Gm. All the three mixed populations reached a bacterial charge of 10^6 CFU per tumor (Fig. 7a). The *blcRCAB::Km*, *pycA::Gm*, and *pckA::Gm* mutants were impaired in competitive fitness (Fig. 7b). Finally, we performed competitions between the *pycA::Gm* and *pckA::Gm* mutants to ascertain whether one of the two pathways (Entner–Doudoroff or gluconeogenesis) could be a major contributor to bacterial fitness in plant tumors. The competitive index was close to 1, showing that the two pathways contributed equally to the tumor niche exploitation by *A. tumefaciens*. Noticeably, the mixed population composed of the *pycA::Gm* and *pckA::Gm* mutants colonized the plant tumors less efficiently (bacterial charge in Fig. 7a) compared with the other mixed populations, especially the *pycA::Gm* and C107-Gm mix. This result suggested that *A. tumefaciens* could gain an advantage in the simultaneous expression of the Entner–Doudoroff pathway and gluconeogenesis.

Discussion

The biotrophic pathogen *A. tumefaciens* diverts the plant development and metabolism for constructing and exploiting a privileged ecological niche: the plant tumor. Numerous studies have deciphered the tumor niche construction process by studying the

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

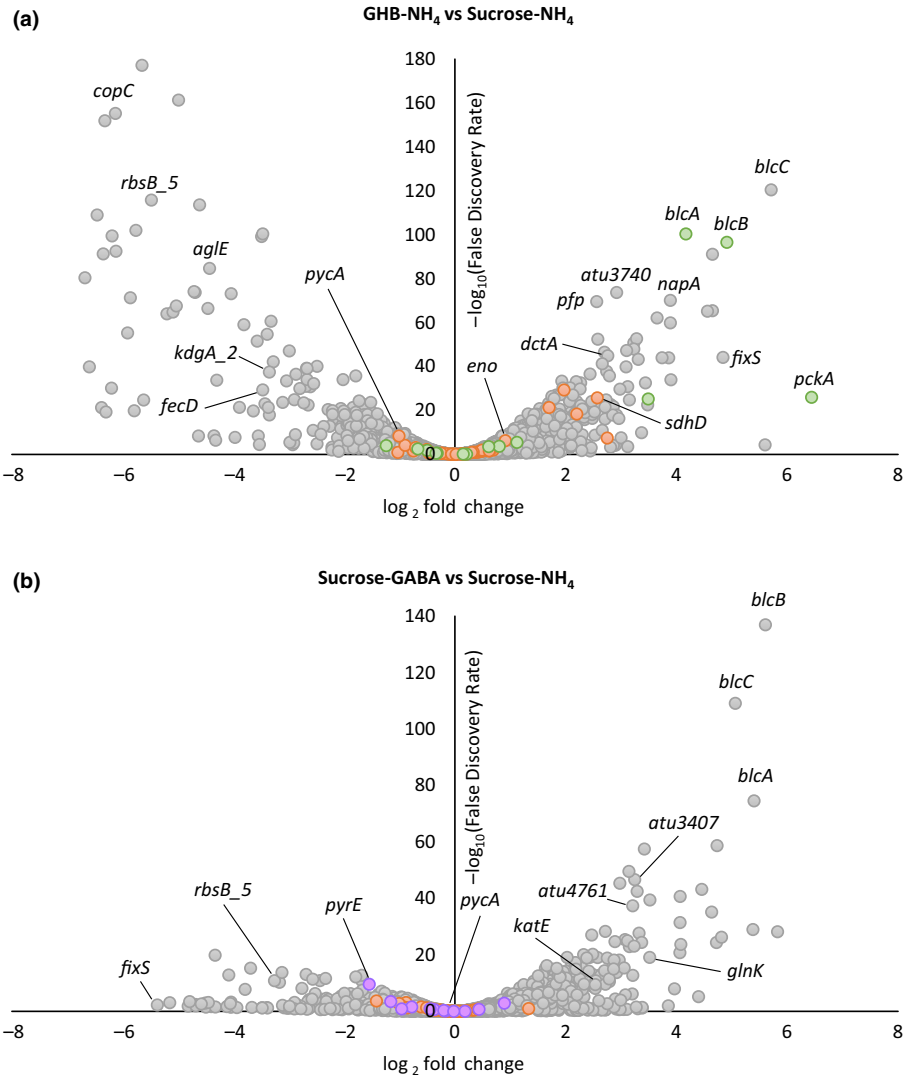


Fig. 4 Volcano plot of transcriptomic data. The data for all genes are plotted as $\log_2$ fold change vs the $-\log_{10}$ of the adjusted *P*-value. (a) Differentially expressed genes between γ-hydroxybutyric acid (GHB)–ammonium (NH_4, as chloride) and sucrose–NH_4 growth conditions. The fitness genes (identified by transposon-sequencing (Tn-seq)) in the GHB–NH_4 growth condition are colored in green, and those in the sucrose–NH_4 growth condition are presented in orange. (b) Differential expressed genes between sucrose–γ-aminobutyric acid (GABA) and sucrose–NH_4 growth conditions. The essential genes (identified by Tn-seq) in the sucrose–GABA condition are in purple, and those in the sucrose–NH_4 condition are presented in orange. The identities of some genes are indicated.

T-DNA transfer and expression in the host plant, as well as mechanisms to escape plant defense (Gohlke & Deeken, 2014; Gelvin, 2017). In this study, we combined different omics (metabolomics, transcriptomics, and Tn-seq) to uncover the role of *A. tumefaciens* genes and pathways in tumor niche exploitation.

Metabolomics of tumor tissues induced on *S. lycopersicum* by *A. tumefaciens* revealed a wide variety of metabolites (e.g. sugars, polyols, amino acids, organic acids, phenolics). They are potential nutrients supporting the proliferation of *A. tumefaciens* that reached 10^6 CFU g^{-1} of fresh tumor tissues. Most of the quantified metabolites (106 of 130) were accumulated at a quite similar concentration in uninfected stems and tumors (Fig. 1). The most abundant metabolites were also the common ones in the two tissues (e.g. glucose, sucrose, malate, glutamate), as well as in tomato seeds and root exudates (Kamilova *et al.*, 2006). These plant metabolites could support the growth of *A. tumefaciens* when it colonizes either asymptomatic or symptomatic plants.

The 24 other metabolites that we quantified, such as GABA, proline, pyruvate, GHB, SSA, and opines, were enriched at least four times in plant tumors compared with healthy stems. Some of them (e.g. agrocinopine, pyruvate, SSA) were increased in plant tumors by several orders of magnitude. These enriched compounds are chemical signatures of the tumor niche: *A. tumefaciens* was expected to have evolved pathways for detoxifying and exploiting these compounds as nutrients and signals. This paradigm is well exemplified by the two opines nopaline and agrocinopine and ferulic derivatives. Nopaline confers a selective growth advantage to a nopaline-assimilating *A. tumefaciens* in *S. lycopersicum* tumors (Lang *et al.*, 2014). The agrocinopine is cleaved into sucrose and arabinose-2-phosphate, which plays an important signaling role: arabinose-2-phosphate enhances the quorum-sensing, which in turn activates the horizontal transfer of the Ti plasmid, and hence the dissemination of the virulence genes (El Sahili *et al.*, 2015). *Agrobacterium* detoxifies ferulic derivatives using different pathways (Brencic *et al.*, 2004; Campillo *et al.*, 2014).

By combining transcriptomics, Tn-seq, and plant infection assays, we investigated the degradative pathways of one common metabolite, sucrose, and two tumor-enriched metabolites, GHB as a C source and GABA as an N source. In the case of sucrose, the combination of Tn-seq and transcriptomics led us to identify the assimilative circuit that starts by the conversion of glucose and fructose into glucose 6-phosphate (*pgi* as a fitness

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

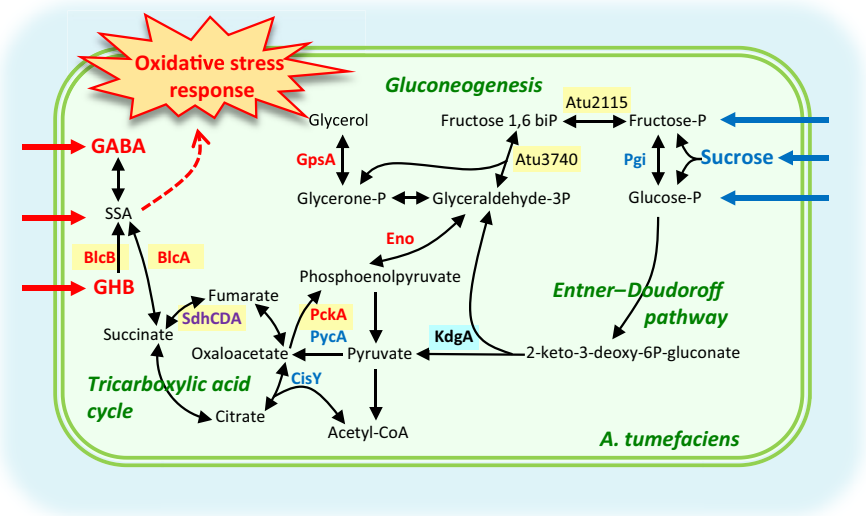


Fig. 5 *Agrobacterium tumefaciens* key pathways for exploiting sucrose and γ-hydroxybutyric acid (GHB). This scheme combines the transposon-sequencing (Tn-seq) and transcriptomics data. Tn-seq revealed fitness genes for growing on either GHB (*blcAB*, *pckA*, *eno*, *gpsA* in red) or sucrose (*pgi*, *pycA*, *cisY* in blue), and on both carbon sources (*sdhCDA* in purple). Transcriptomics revealed response genes when *A. tumefaciens* was grown on either GHB (*blcAB*, *sdhCDA*, *pckA*, *atu3740*, *atu2115* in yellow) or sucrose (*kdgA* in blue). GABA, γ-aminobutyric acid; SSA, succinic semialdehyde.

gene), then its conversion into pyruvate via the Entner–Doudoroff pathway (*kdgA* as an upregulated gene) before entering into the TCA cycle (*pycA*, *cisY*, and *sdhCDA* as fitness genes). These pathways are consistent with a previous metabolic study showing the absence of glycolysis in *A. tumefaciens* and demonstrating the Entner–Doudoroff pathway as a main road of sugar degradation (Fuhrer *et al.*, 2005). The same work also pointed to a high C flux (almost 100%) between the Entner–Doudoroff pathway and the TCA cycle. These two metabolic characteristics are shared by Rhizobiaceae such as *A. tumefaciens* and the legume symbiont *Sinorhizobium meliloti*, and contrasted to sugar assimilation in other bacteria such as *E. coli*, *Bacillus subtilis*, and *Pseudomonas fluorescens* (Fuhrer *et al.*, 2005). The importance of the TCA cycle in a complete exploitation of sugars as resource by *A. tumefaciens* was also supported by our Tn-seq data, with *pycA*, *cisY*, and *sdhCDA* as key fitness genes in the presence of sucrose.

These coherent results led us to evaluate the role of the C flows into and from the TCA cycle in the *A. tumefaciens*–host plant interaction using a reverse genetics approach. In co-infection assays in plant tumors, the *pycA* and *pckA* mutants were outcompeted by a wild-type genotype, but they showed a similar relative fitness when they were competed together. A wild-type *A. tumefaciens* could be considered as a generalist for the assimilation of a wide spectrum of metabolites. By contrast, the constructed *pycA* and *pckA* mutants could be considered as specialists for a restricted range of metabolites that are assimilated by either the gluconeogenesis or the Entner–Doudoroff pathway. A fitness decrease of each mutant in competition with the wild-type highlighted the advantage that the biotrophic pathogen gained by assimilating a wide spectrum of plant metabolites. Moreover, the co-existence of the two *pycA* and *pckA* mutants revealed that the two types of resources were abundant enough and/or differentially distributed to sustain the growth of these two specialists in plant tumors. A remarkable study reported an increase of the sucrose concentration in tumors according to the age of the tumors, as well as in a gradient from the center to the periphery of the tumors on *Ricinus communis* (Wächter *et al.*, 2003). In further studies, the spatial and temporal distribution of the metabolites should be considered as an important parameter driving the resource exploitation strategy of *A. tumefaciens* when it colonizes the heterogeneous environment that the plant tumors are.

Our data also revealed that a mixture of the two specialists (the *pycA* and *pckA* mutants) was less efficient for exploiting the host in terms of bacterial charge than a mixture containing the generalist (the wild-type) and one of the two specialists (Fig. 7). This suggested that *A. tumefaciens* could take advantage of the simultaneous expression of the Entner–Doudoroff pathway and gluconeogenesis in the same individual, quite simply because these two pathways are connected for allowing the recycling of several metabolites such as glyceraldehyde 3-phosphate and phosphoenolpyruvate (Fig. 5). As already discussed, some other explanations related to spatial and temporal distribution of the resource cannot be excluded. The transcriptome of *A. tumefaciens* living in tumors on the *A. thaliana* host plant is consistent with a simultaneous expression of pathways to exploit of a wide diversity of C and N sources (González-Mula *et al.*, 2018). In *A. tumefaciens* and in some other Rhizobiaceae, the separation of C flows incoming and outgoing into and from the TCA cycle using the Entner–Doudoroff pathway and gluconeogenesis (instead of a unique, reversible glycolysis/gluconeogenesis pathway) could be a biological innovation contributing to an optimal exploitation of the diversified resources available in the plant hosts. The capacity of the microbial pathogens to activate and regulate their C assimilative pathways is crucial for survival and invasion in plant and animal hosts (Alteri *et al.*, 2009; Brock, 2009; Basu *et al.*, 2018).

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

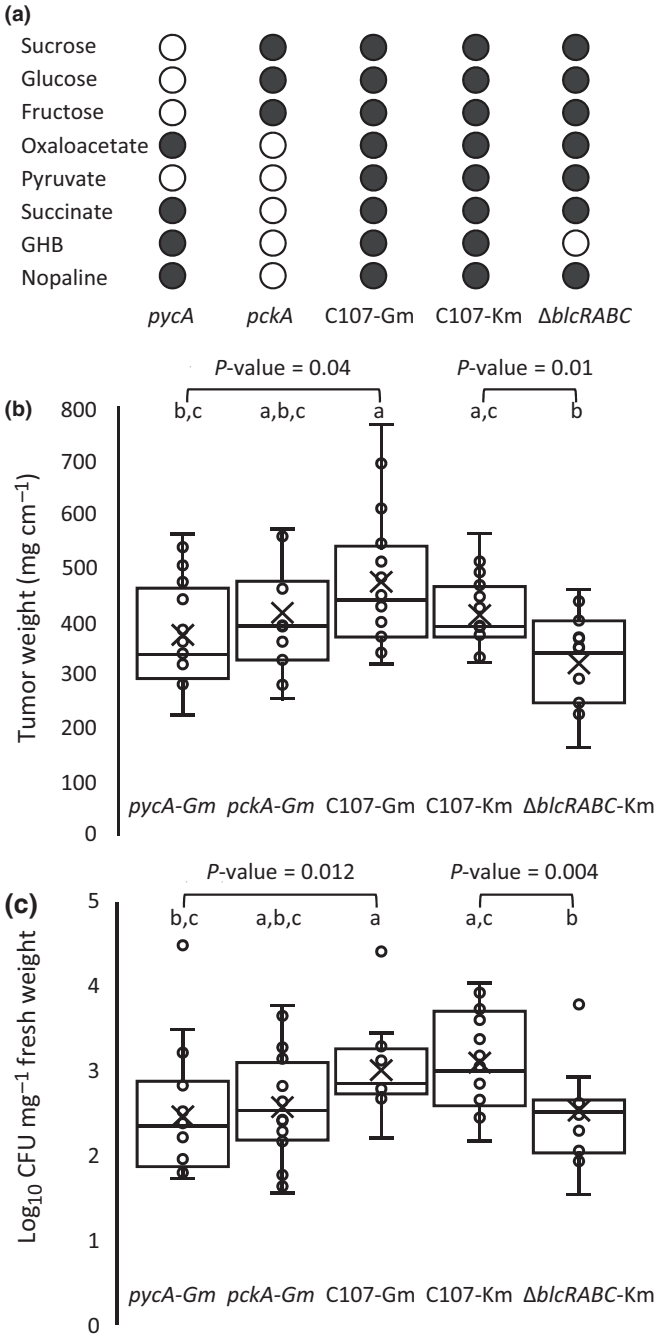


Fig. 6 Metabolic capacity, aggressiveness, and colonization of the constructed *Agrobacterium tumefaciens* mutants. (a) *A. tumefaciens* knockout (KO) mutants and control strains growing in AB medium supplemented with different carbon sources. Open circles represent the absence of growth ($OD_{600} < 0.05$). (b) Fresh weight (FW) of tomato tumors induced with KO mutants and control strains. (c) Colonization efficiency (bacterial numeration, $\log_{10}$ CFU mg^{-1} FW) in tomato tumors. Mean values (indicated by a cross), median (horizontal line), and standard deviations (SD) of two independent experiments are presented. Nonparametric Kruskal–Wallis and post-hoc Tukey tests ($n = 16$; $P < 0.05$) were used, and different letters indicate statistical significance.

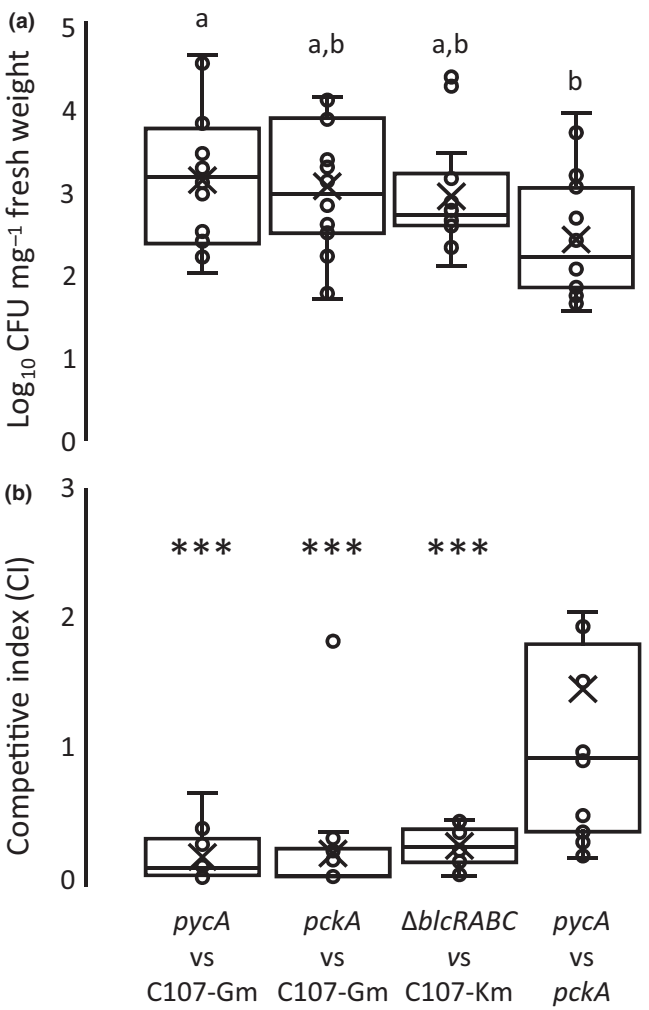


Fig. 7 Competitive fitness of the constructed *Agrobacterium tumefaciens* mutants. (a) Colonization efficiency (total bacterial numeration (mutant plus control), $\log_{10}$ CFU mg^{-1} FW) in tomato tumors. Mean values (indicated by a cross), median (horizontal line), and standard deviations (SD) of two independent experiments are presented. Nonparametric Kruskal–Wallis and post-hoc Tukey tests ($n = 16$; $P < 0.05$) were used and different letters indicate statistical significance. (b) Relative abundance of knockout mutant and control strain was compared at infection time and in tumors. A competitive index value < 1 indicates a fitness loss of the mutant strain in plant tumors. Average (indicated by a cross), median (horizontal line), and SD were calculated from two independent experiments ($n = 16$). Significant fitness loss of mutants is noted by a triple asterisk (Wilcoxon signed rank test $P < 0.001$).

The tumor-enriched metabolites GHB, GABA, and SSA are metabolically connected. GABA is the highest abundant nonprotein amino acid in tumor tissues of *A. thaliana* and *S. lycopersicum* (Deeken *et al.*, 2006; Lang *et al.*, 2016; this work). In the host plant, GABA is mainly produced from glutamate (by GABA decarboxylase) and then degraded into SSA (by GABA transaminase), which is in turn converted into succinate (by SSA dehydrogenase) or GHB (by GHB reductase) (Bown & Shelp, 2016). SSA is a toxic metabolite provoking an oxidative stress in plants and other organisms and microorganisms, including *A. tumefaciens* (Bouché *et al.*, 2003; Ludewig *et al.*, 2008; Wang *et al.*, 2016). In plant tumors, *A. tumefaciens* may exploit plant GABA and GHB as N and C sources but has to face toxic SSA – either exogenous SSA resulting from plant metabolism or

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

endogenous SSA as an intermediate of the *A. tumefaciens* GABA and GHB degradation pathways. The published transcriptome of *A. tumefaciens* in plant tumors showed that the pathogen responded to the presence of GABA and its derived metabolites SSA and GHB, since the *blc* genes, as well as the *atu4761* gene coding for a putative GABA transaminase were upregulated (González-Mula *et al.*, 2018).

When *A. tumefaciens* was grown on GABA as an N source, the Tn-seq approach failed to identify any genes coding for a putative GABA transaminase, nor an SSA dehydrogenase that would be involved in the degradation of GABA and the detoxification of SSA. This may be explained by a redundancy of genes coding these two enzymatic activities. Transcriptomics supported this hypothesis. Two genes coding for putative transaminases (*atu3407* and *atu4761*) and two others for SSA dehydrogenases (*atu4762* and *blcA*) were upregulated in *A. tumefaciens* growing on GABA. The role of the SSA dehydrogenases in stress response and quorum-sensing signal decay was previously studied by Wang *et al.* (2006). In our study, simple and double-KO mutants of the two transaminases still grew on GABA, suggesting the presence of at least a third gene encoding a GABA transaminase in *A. tumefaciens*. In the related species *Rhizobium leguminosarum*, three GABA transaminases are involved in the degradation of GABA (Prell *et al.*, 2009).

When *A. tumefaciens* was grown in the presence of GHB as a C source, a combination of Tn-seq and transcriptomics identified the *blcAB* genes, which are required for the conversion of GHB to succinate (Carlier *et al.*, 2004; Chai *et al.*, 2007). This approach also permitted connecting this particular pathway to the central metabolism by highlighting genes of the TCA cycle, gluconeogenesis, and synthesis of lipid precursors (*sdhDC*, *pckA*, *eno*, *atu3740*, *atu2115* and *gpsA*), . The *A. tumefaciens* mutants defective in *blc* or *pckA* genes were unable to grow on GHB as a nutrient, validating the data collected from Tn-seq. Tn-seq and transcriptomics appeared as complementary for deciphering microbial pathways.

Besides a potential growth advantage related to nutrient exploitation, the GABA- and GHB-transcriptomes highlighted an oxidative stress response in *A. tumefaciens*. In the presence of GABA, the upregulated genes concerned were, for instance, *katE* and *sodB*, coding for catalase and superoxide dismutase, respectively. In the presence of GHB, the stress response's upregulated genes were the *cyd* and *fix* genes coding for oxidative phosphorylation pathways, whereas the downregulated genes were involved in siderophore synthesis and uptake (Fig. 4a). A decrease of iron uptake would contribute to reduce the production of highly deleterious hydroxyl radicals via the Fenton reaction. In culture assays, Wang *et al.* (2016) showed that a pre-exposure of *A. tumefaciens* to extracellular SSA induces an oxidative stress response and increases resistance of *A. tumefaciens* to hydrogen peroxide. Noticeably, this effect was lost in a *blcABC* KO-mutant (Wang *et al.*, 2016). Our study showed that a *blc* KO-mutant was impaired for inducing tumors on tomato stems, as well as colonizing plant tumors and competing with a wild-type strain in the plant tumors. Two nonexclusive explanations of this selective advantage could be proposed: an impaired assimilation of GHB or GABA as nutrients, and an impaired SSA-mediated activation of the oxidative stress response to face plant defense.

Different arguments supported the impaired oxidative stress response as an important cause of the decreased aggressiveness and fitness in the *blc* mutant: first, *A. tumefaciens* mutants of catalase and superoxide dismutase were impaired in virulence, highlighting oxidative stress response as an important trait during plant infection (Xu & Pan, 2000; Saenkham *et al.*, 2007); second, whereas the *blc* mutant was affected in aggressiveness (tumor weight), bacterial invasion (bacterial charge), and fitness (competition vs wild-type allele), the *pckA* mutant was impaired in competitive fitness only, suggesting that the *blc* genes conferred an advantage that could not be exclusively explained by nutrition.

In *A. tumefaciens*, the *blc* genes are carried by the dispensable At plasmid, which reaches a size of 0.5 Mb in *A. tumefaciens* C58 (Goodner, 2001; Wood *et al.*, 2001). Several studies have highlighted the fitness cost imposed by maintenance and expression of At plasmid genes (Morton *et al.*, 2013; Platt *et al.*, 2014; González-Mula *et al.*, 2018). This study showed the fitness gains conferred by *blc* genes in plant host infection. The selective advantage conferred by the *blc* operon would not be restricted to *Agrobacterium* pathogens, as data mining analysis revealed its presence in the genome of several host-interacting bacteria, such as *Rhizobium etli*, *Burkholderia phenoliruptrix*, and *Pantoea* sp.

Beyond the use of opines, our study expanded the ecological traits supporting exploitation of tumor niche by *A. tumefaciens*, highlighting novel targets for controlling its virulence and proliferation.

Acknowledgements

We thank Julien Lang, Anthony Kwasiborski and Fabienne Pierre (I2BC) for their contribution to metabolomics, transcriptomics, and reverse genetics, and Erwan Gueguen (MAP, Université Lyon) for kindly providing plasmid pSAM_DGm. This work has benefited from the facilities and expertise of the high-throughput sequencing platform and the plant culture facilities of the I2BC and the metabolomics platform of the Plant Observatory–Chemistry and Metabolomics (Versailles, France). This work was supported by CNRS (SE2016-2017), University Paris-Sud (PhD grant to AG-M) and LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS). The authors declare no conflict of interest.

Author contributions

PM and DF designed research; AG-M, JL, LM and DN performed research; AG-M, DF, JL and FL analyzed data; AG-M, DN, PM and DF wrote the paper.

ORCID

Denis Faure https://orcid.org/0000-0002-5379-8867
Florian Lamouche https://orcid.org/0000-0001-6232-5829
Peter Mergaert https://orcid.org/0000-0002-5919-7317

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

References

Alteri CJ, Smith SN, Mobley HLT. 2009. Fitness of *Escherichia coli* during urinary tract infection requires gluconeogenesis and the TCA cycle. *PLoS Pathogens* **5**: e1000448.

Aznar A, Chen NW, Thomine S, Dellagi A. 2015. Immunity to plant pathogens and iron homeostasis. *Plant Science* **240**: 90–97.

Barton IS, Fuqua C, Platt TG. 2018. Ecological and evolutionary dynamics of a model facultative pathogen: *Agrobacterium* and crown gall disease of plants. *Environmental Microbiology* **20**: 16–29.

Basu P, Sandhu N, Bhatt A, Singh A, Balhana R, Gobe I, Crowhurst NA, Mendum TA, Gao L, Ward JL *et al.* 2018. The anaplerotic node is essential for the intracellular survival of *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* **293**: 5695–5704.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bouché N, Fait A, Bouchez D, Møller SG, Fromm H. 2003. Mitochondrial succinic-semialdehyde dehydrogenase of the gamma-aminobutyrate shunt is required to restrict levels of reactive oxygen intermediates in plants. *Proceedings of the National Academy of Sciences, USA* **100**: 6843–6848.

Bown AW, Shelp BJ. 2016. Plant GABA: not just a metabolite. *Trends in Plant Science* **21**: 811–813.

Brencic A, Eberhard A, Winans SC. 2004. Signal quenching, detoxification and mineralization of *vir* gene-inducing phenolics by the VirH2 protein of *Agrobacterium tumefaciens*. *Molecular Microbiology* **51**: 1103–1115.

Brock M. 2009. Fungal metabolism in host niches. *Current Opinion in Microbiology* **12**: 371–376.

Campillo T, Renoud S, Kerzaon I, Vial L, Baude J, Gaillard V, Bellvert F, Chamignon C, Comte G, Nesme X *et al.* 2014. Analysis of hydroxycinnamic acid degradation in *Agrobacterium fabrum* reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway. *Applied and Environmental Microbiology* **80**: 3341–3349.

Carlier A, Chevrot R, Dessaux Y, Faure D. 2004. The assimilation of γ-butyrolactone in *Agrobacterium tumefaciens* C58 interferes with the accumulation of the *N*-acyl-homoserine lactone signal. *Molecular Plant–Microbe Interactions* **17**: 951–957.

Chai Y, Ching ST, Cho H, Winans SC. 2007. Reconstitution of the biochemical activities of the AttJ repressor and the AttK, AttL, and AttM catabolic enzymes of *Agrobacterium tumefaciens*. *Journal of Bacteriology* **189**: 3674–3679.

Chevrot R, Rosen R, Haudecoeur E, Cirou A, Shelp BJ, Ron E, Faure D. 2006. GABA controls the level of quorum-sensing signal in *Agrobacterium tumefaciens*. *Proceedings of the National Academy of Sciences, USA* **103**: 7460–7464.

Chilton MD, Currier TC, Farrand SK, Bendich AJ, Gordon MP, Nester EW. 1974. *Agrobacterium tumefaciens* DNA and PS8 bacteriophage DNA not detected in crown gall tumors. *Proceedings of the National Academy of Sciences, USA* **71**: 3672–3676.

Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiro L. 2011. The essential genome of a bacterium. *Molecular Systems Biology* **7**: E528.

Curtis PD, Brun YV. 2014. Identification of essential alphaproteobacterial genes reveals operational variability in conserved developmental and cell cycle systems. *Molecular Microbiology* **93**: 713–735.

Deeken R, Engelmann JC, Efetova M, Czirjak T, Muller T, Kaiser WM, Tietz O, Krischke M, Mueller MJ, Palme K *et al.* 2006. An integrated view of gene expression and solute profiles of *Arabidopsis* tumors: a genome-wide approach. *Plant Cell* **18**: 3617–3634.

DeJesus MA, Ioerger TR. 2013. A hidden Markov model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinformatics* **14**: e303.

Dennis JJ, Zylstra GJ. 1998. Plasposons: modular self-cloning minitransposon derivatives for rapid genetic analysis of Gram-negative bacterial genomes. *Applied and Environmental Microbiology* **64**: 2710–2715.

Dessaux Y, Faure D. 2018. Quorum sensing and quorum quenching in *Agrobacterium*: a 'go/no go system'? *Genes* **9**: E210.

Dessaux Y, Petit A, Tempe J. 1993. Chemistry and biochemistry of opines, chemical mediators of parasitism. *Phytochemistry* **34**: 31–38.

El Sahili A, Li S-Z, Lang J, Virus C, Planamente S, Ahmar M, Guimaraes BG, Aumont-Nicaise M, Vigouroux A, Soulère L *et al.* 2015. A pyranose-2-phosphate motif is responsible for both antibiotic import and quorum-sensing regulation in *Agrobacterium tumefaciens*. *PLoS Pathogens* **11**: e1005071.

Faure D, Simon JC, Heulin T. 2018. Holobiont: a conceptual framework to explore the eco-evolutionary and functional implications of host–microbiota interactions in all ecosystems. *New Phytologist* **218**: 1321–1324.

Fields JH, Baldwin J, Hochachka PW. 1976. On the role of octopine dehydrogenase in cephalopod mantle muscle metabolism. *Canadian Journal of Zoology* **54**: 871–878.

Fuhrer T, Fischer E, Sauer U. 2005. Experimental identification and quantification of glucose metabolism in seven bacterial species. *Society* **187**: 1581–1590.

Gelvin SB. 2017. Integration of *Agrobacterium* T-DNA into the plant genome. *Annual Review of Genetics* **82**: 5542–5552.

Gohlke J, Deeken R. 2014. Plant responses to *Agrobacterium tumefaciens* and crown gall development. *Frontiers in Plant Science* **5**: 155.

González-Mula A, Lang J, Grandclément C, Naquin D, Ahmar M, Soulère L, Queneau Y, Dessaux Y, Faure D. 2018. Lifestyle of the biotroph *Agrobacterium tumefaciens* in the ecological niche constructed on its host plant. *New Phytologist* **219**: 350–362.

Goodner B. 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**: 2323–2328.

Haudecoeur E, Planamente S, Cirou A, Tannières M, Shelp BJ, Moréra S, Faure D. 2009b. Proline antagonizes GABA-induced quenching of quorum-sensing in *Agrobacterium tumefaciens*. *Proceedings of the National Academy of Sciences, USA* **106**: 14587–14592.

Haudecoeur E, Tannieres M, Cirou A, Raffoux A, Dessaux Y, Faure D. 2009a. Different regulation and roles of lactonases AiiB and AttM in *Agrobacterium tumefaciens* C58. *Molecular Plant–Microbe Interactions* **22**: 529–537.

Kamilova F, Kravchenko LV, Shaposhnikov AI, Azarova T, Makarova N, Lugtenberg B. 2006. Organic acids, sugars, and L-tryptophane in exudates of vegetables growing on stonewool and their effects on activities of rhizosphere bacteria. *Molecular Plant–Microbe Interactions* **19**: 250–256.

Kylafis G, Loreau M. 2011. Niche construction in the light of niche theory. *Ecology Letters* **14**: 82–90.

Lang J, Gonzalez-Mula A, Taconnat L, Clement G, Faure D. 2016. The plant GABA signaling downregulates horizontal transfer of the *Agrobacterium tumefaciens* virulence plasmid. *New Phytologist* **210**: 974–983.

Lang J, Vigouroux A, El Sahili A, Kwasiborski A, Aumont-Nicaise M, Dessaux Y, Shykoff JA, Moréra S, Faure D, El Sahili A *et al.* 2017. Fitness costs restrict niche expansion by generalist niche-constructing pathogens. *The ISME Journal* **11**: 374–385.

Lang J, Vigouroux A, Planamente S, El Sahili A, Blin P, Aumont-Nicaise M, Dessaux Y, Moréra S, Faure D. 2014. *Agrobacterium* uses a unique ligand-binding mode for trapping opines and acquiring a competitive advantage in the niche construction on plant host. *PLoS Pathogens* **10**: e1004444.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Liu P, Wood D, Nester EW. 2005. Phosphoenolpyruvate carboxykinase is an acid-induced, chromosomally encoded virulence factor in *Agrobacterium tumefaciens*. *Journal of Bacteriology* **187**: 6039–6045.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**: 402–408.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: e550.

Ludewig F, Hüser A, Fromm H, Beauclair L, Bouché N. 2008. Mutants of GABA transaminase (POP2) suppress the severe phenotype of succinic semialdehyde dehydrogenase (ssadh) mutants in *Arabidopsis*. *PLoS ONE* **3**: e3383.

Macho AP, Guidot A, Barberis P, Beuzón CR, Genin S. 2010. A competitive index assay identifies several *Ralstonia solanacearum* type III effector mutant strains with reduced fitness in host plants. *Molecular Plant–Microbe Interactions* **23**: 1197–1205.

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

Martin FM, Uroz S, Barker DG. 2017. Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.

Marty L, Vigouroux A, Aumont-Nicaise M, Dessaux Y, Faure D, Moréra S. 2016. Structural basis for high specificity of Amadori compound and mannopine opine binding in bacterial pathogens. *Journal of Biological Chemistry* **291**: 22638–22649.

McNally L, Brown SP. 2015. Building the microbiome in health and disease: niche construction and social conflict in bacteria. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **370**: e20140298.

Morton ER, Merritt PM, Bever JD, Fuqua C. 2013. Large deletions in the pAtC58 megaplasmid of *Agrobacterium tumefaciens* can confer reduced carriage cost and increased expression of virulence genes. *Genome Biology and Evolution* **5**: 1453–1464.

van Opijnen T, Camilli A. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology* **11**: 435–442.

Planamente S, Vigouroux A, Mondy S, Nicaise M, Faure D, Moréra S. 2010. A conserved mechanism of GABA binding and antagonism is revealed by structure–function analysis of the periplasmic binding protein Atu2422 in *Agrobacterium tumefaciens*. *The Journal of Biological Chemistry* **285**: 30294–30303.

Platt TG, Morton ER, Barton IS, Bever JD, Fuqua C. 2014. Ecological dynamics and complex interactions of *Agrobacterium* megaplasmids. *Frontiers in Plant Science* **5**: e635.

Podlešáková K, Ugena L, Spíchal L, Doležal K, De Diego N. 2019. Phytohormones and polyamines regulate plant stress responses by altering GABA pathway. *New Biotechnology* **48**: 53–65.

Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nature Reviews Microbiology* **16**: 291–303.

Prell J, Bourdés A, Karunakaran R, Lopez-Gomez M, Poole P. 2009. Pathway of γ-aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. *Journal of Bacteriology* **191**: 2177–2186.

Pritchard JR, Chao MC, Abel S, Davis BM, Baranowski C, Zhang YJ, Rubin EJ, Waldor MK. 2014. ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. *PLoS Genetics* **10**: e1004782.

Saenkham P, Eiamphungporn W, Farrand SK, Vattanaviboon P, Mongkolsuk S. 2007. Multiple superoxide dismutases in *Agrobacterium tumefaciens*: functional analysis, gene regulation, and influence on tumorigenesis. *Journal of Bacteriology* **189**: 8807–8817.

Simoh S, Quintana N, Kim HK, Choi YH, Verpoorte R. 2009. Metabolic changes in *Agrobacterium tumefaciens*-infected *Brassica rapa*. *Journal of Plant Physiology* **166**: 1005–1014.

Skurnik D, Roux D, Aschard H, Cattoir V, Yoder-Himes D, Lory S, Pier GB. 2013. A comprehensive analysis of in vitro and in vivo genetic fitness of *Pseudomonas aeruginosa* using high-throughput sequencing of transposon libraries. *PLoS Pathogens* **9**: e1003582.

Tannières M, Lang J, Barnier C, Shykoff JA, Faure D. 2017. Quorum-quenching limits quorum-sensing exploitation by signal-negative invaders. *Scientific Reports* **7**: 40126.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28**: 33–36.

Vigouroux A, El Sahili A, Lang J, Aumont-Nicaise M, Dessaux Y, Faure D, Moréra S. 2017. Structural basis for high specificity of octopine binding in the plant pathogen *Agrobacterium tumefaciens*. *Scientific Reports* **7**: e18033.

Wächter R, Langhans M, Aloni R, Götz S, Weilmünster A, Koops A, Temguia L, Mistrik I, Pavlovkin J, Rascher U *et al.* 2003. Vascularization, high-volume solution flow, and localized roles for enzymes of sucrose metabolism during tumorigenesis by *Agrobacterium tumefaciens*. *Plant Physiology* **133**: 1024–1037.

Wang C, Tang D, Gao YG, Zhang LH. 2016. Succinic semialdehyde promotes prosurvival capability of *Agrobacterium tumefaciens*. *Journal of Bacteriology* **198**: 930–940.

Wang C, Zhang HB, Wang LH, Zhang LH. 2006. Succinic semialdehyde couples stress response to quorum-sensing signal decay in *Agrobacterium tumefaciens*. *Molecular Microbiology* **62**: 45–56.

Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF *et al.* 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**: 2317–2323.

Xu XQ, Pan SQ. 2000. An *Agrobacterium* catalase is a virulence factor involved in tumorigenesis. *Molecular Microbiology* **35**: 407–414.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 The *A. tumefaciens* C58 essential genes identified in the Tn*5* and *Himar1* transposon libraries.

Fig. S2 Classification of essential genes by COG categories.

Fig. S3 RT-qPCR and transcriptome comparative gene expression.

Fig. S4 Growth on GABA as a sole nitrogen source.

Table S1 Primer list.

Table S2 Plant metabolomics.

Table S3 Tn-seq data.

Table S4 Transcriptomics.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

© 2018 The Authors
New Phytologist © 2018 New Phytologist Trust

Microbes Environ. Vol. 34, No. 2, 219-222, 2019
https://www.jstage.jst.go.jp/browse/jsme2 doi:10.1264/jsme2.ME19011

Short Communication

Burkholderia Gut Symbionts Associated with European and Japanese Populations of the Dock Bug *Coreus marginatus* (Coreoidea: Coreidae)

Tsubasa Ohbayashi[1], Hideomi Itoh[2]*, Joy Lachat[1], Yoshitomo Kikuchi[2,3], and Peter Mergaert[1]

[1]*Institute for Integrative Biology of the Cell, UMR9198, CNRS, Université Paris-Sud, CEA, Gif-sur-Yvette, France;* [2]*Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Hokkaido Center, 2–17–2–1 Tsukisamu-higashi, Toyohira-ku, Sapporo 062–8517, Japan; and* [3]*Computational Bio Big Data Open Innovation Laboratory (CBBDOIL), AIST, Hokkaido Center, 2–17–2–1 Tsukisamu-higashi, Toyohira-ku, Sapporo 062–8517, Japan*

(Received January 21, 2019—Accepted March 19, 2019—Published online June 6, 2019)

Insects of the heteropteran superfamilies Coreoidea and Lygaeoidea are consistently associated with symbionts of a specific group of the genus *Burkholderia*, called the "stinkbug-associated beneficial and environmental (SBE)" group. The symbiosis is maintained by the environmental transmission of symbionts. We investigated European and Japanese populations of the dock bug *Coreus marginatus* (Coreoidea: Coreidae). High nymphal mortality in reared aposymbiotic insects suggested an obligate host-symbiont association in this species. Molecular phylogenetic analyses based on 16S rRNA gene sequences revealed that all 173 individuals investigated were colonized by *Burkholderia*, which were further assigned to different subgroups of the SBE in a region-dependent pattern.

Key words: *Burkholderia*, stinkbug, obligate gut symbiosis, region-dependent symbionts

The suborder Heteroptera is a diverse taxonomic group in insects and consists of 42,300 described species (9). Phytophagous members commonly possess symbiotic bacteria inside their bodies (1, 15, 19). While some heteropteran species harbor symbionts intracellularly (10, 20–22, 24), the majority of phytophagous species possess symbiotic bacteria extracellularly in the lumen of sac-like tissues, called "crypts", in the posterior midgut (2, 8, 25). Members of the superfamily Pentatomoidea harbor specific bacterial symbionts belonging to *Gammaproteobacteria* (19). These symbionts are essential for host growth and reproduction and are vertically transmitted from mother to offspring. In contrast, most members of the superfamilies Lygaeoidea and Coreoidea are associated with betaproteobacterial symbionts of a specific clade in the genus *Burkholderia*, called the "stinkbug-associated beneficial and environmental (SBE)" group (7, 16, 30). The coreoid and lygaeoid species not vertically transmit *Burkholderia* symbionts, but they acquire them from environmental soil every generation (14, 17). At this stage, the biological effects of the *Burkholderia* symbiont have only been reported in the bean bug *Riptortus pedestris* (superfamily Coreoidea: family Alydidae), in which the symbiont is not essential, but significantly enhances the growth rate, body size, and fecundity of the bean bug host (14, 18).

We previously investigated 22 species of Coreoidea and Lygaeoidea, all of which were collected in Japan and harbored the SBE group *Burkholderia* (13, 16). Six species of American Coreoidea and Lygaeoidea were also examined and the symbiotic organs of these species were also dominated by SBE-group *Burkholderia* (1, 7, 26), although other groups of *Burkholderia* were also detected in some cases (1). A recent study on European and Japanese species of the spurge bug, *Dicranocephalus* spp. (superfamily Coreoidea: family Stenocephalidae), revealed that while the Japanese species are consistently associated with the SBE group *Burkholderia*, European species are more likely to harbor a distinct lineage of *Burkholderia*, tentatively named "Stenocephalidae-clade" *Burkholderia* (23). This finding suggests the geographical divergence of the stinkbug-*Burkholderia* association. However, it currently remains unclear whether the case of the spurge bug is exceptional.

The dock bug *Coreus marginatus* (superfamily Coreoidea: family Coreidae) (Fig. 1A) is broadly distributed in the Northern Hemisphere, from Europe over central Asia to Japan (11, 12). It feeds on the leaves and seeds of *Rumex* plants (Fig. 1B), and is a serious pest of *Rumex* herbs, such as sorrel (11). In the present study, we investigated the symbiotic bacteria of *C. marginatus*, which belong to the SBE, and examined their fitness effects on the host insect. We further clarified whether a region-dependent divergence of symbionts exists between European and Japanese host populations.

The dock bug possessed numerous crypts in the posterior region of the midgut. These crypts were white and arranged in two rows (Fig. 1C). To investigate the prevalence of *Burkholderia* in this species, wild populations collected in diverse locations of Europe and Japan were assessed by diagnostic PCR with a *Burkholderia*-specific primer set (29). The insects examined in the present study are listed in Table S1. The crypt region was dissected out by forceps under a binocular, and the symbiotic organ (M4 in Fig. 1C) was subjected to DNA extraction and diagnostic PCR as previously described (16). A total of 163 individuals from 16 European populations and ten individuals from two Japanese populations were investigated, all of which were positive for *Burkholderia* (Table S1). In contrast, no *Burkholderia* infection was observed in the egg samples of reared insects (positive/total tested=0/8), strongly suggesting that *C. marginatus* does not transmit the symbiont vertically, but acquires it from the

* Corresponding author. E-mail: hideomi-itou@aist.go.jp;
Tel: +81–11–857–8979; Fax: +81–11–857–8915.

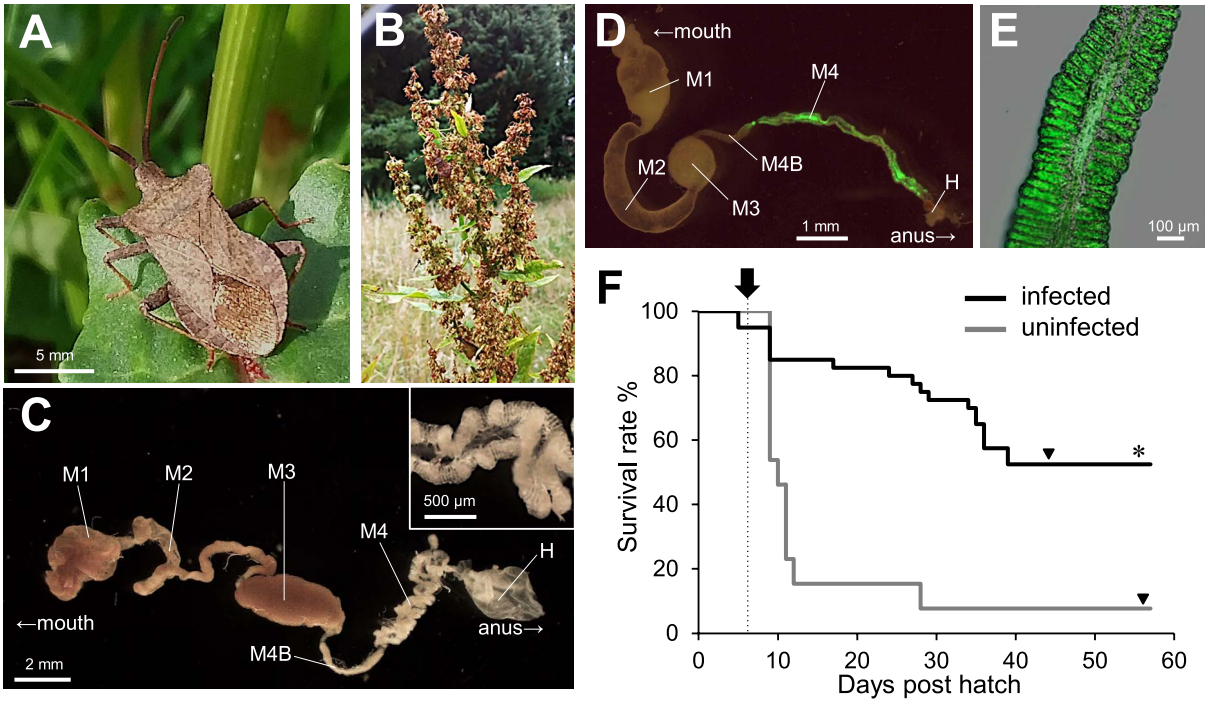


Fig. 1. The dock bug *Coreus marginatus* and its gut symbiotic association. (A) An adult female of *C. marginatus*. (B) *Rumex* host plant. (C) A dissected midgut of an adult male. M1 midgut first section, M2 midgut second section, M3 midgut third section, M4B midgut fourth section with bulb, M4 midgut fourth section with crypts (symbiotic organ), H hindgut. The inset shows an enlarged image of crypt-bearing M4. (D) A dissected midgut of a 3rd instar nymph infected with a GFP-labeled symbiont. (E) An enlarged image of crypts in M4 colonized by GFP-labeled *Burkholderia*. (F) Survival rate of *C. marginatus* infected with *Burkholderia* (black line, n=40) or uninfected (gray line, n=13). An inoculation was performed at 6 d post hatch (arrow with dotted line). Symbiotic insects (21 survivors) molted to adults at 44.2±4.0 d post hatch, and aposymbiotic ones (only 1 survivor) at 57 d post hatch (arrowheads). The survival rate of symbiotic insects was significantly higher than that of aposymbiotic ones (*P<0.01, Fisher's exact test).

environment, similar to other coreoid stinkbugs.

The *Burkholderia* symbiont of the dock bug was successfully isolated from the midgut crypts of an insect collected in Crèche Belle-Image, the CNRS campus, Gif-sur-Yvette, France on 24th May 2017 by culturing the crypt content on a YG (yeast-glucose) agar plate, as previously described (16). A green fluorescence protein (GFP)-expressing derivative, constructed from this isolate as previously described (18) and fed to second instar nymphs that descended from wild insects collected at the same location (Crèche Belle-Image, CNRS-campus, Gif-sur-Yvette, France in 2017), showed a specific localization in the midgut crypts (Fig. 1D and E), confirming the gut symbiotic association between *Burkholderia* and the dock bug. Using this cultured strain, the fitness effects of the *Burkholderia* symbiont were investigated. Second instar nymphs were fed cultured *Burkholderia* 6 d after hatching and maintained in a clean plastic cup at 25°C under a long day regimen (16 h light, 8 h dark) by feeding on roasted pistachio and peanut seeds (*Pistacia vera* and *Arachis hypogaea*, respectively) and distilled water containing 0.05% ascorbic acid. While uninfected insects showed a survival rate of only 7.7% (survived/total=1/13), insect survival significantly improved to 52.5% (21/40) in infected insects (Fig. 1F), strongly suggesting an obligate host-symbiont relationship in the dock bug. In the case of the bean bug *R. pedestris*, the *Burkholderia* association is facultative: the symbiont does not strongly affect host survival, but does influence the growth and fecundity of the insect host (14, 18). Although the biological function of the *Burkholderia* symbiont remains unclear, metabolic dependency on the symbiont appears to differ between stinkbug species that feed on different host plants.

To clarify the phylogenetic placement of *Burkholderia* symbionts associated with dock bugs, selected individuals from the European and Japanese populations were subjected to a clone library analysis of a 1.5-kb fragment of the bacterial 16S rRNA gene, as previously described (16). Ten and four insects representing ten European and two Japanese populations, respectively, were investigated (Table S1). A total of 110 clones were sequenced and subjected to a BLAST search. The top BLAST hits of all sequences were the 16S rRNA gene sequences of *Burkholderia* species. The 110 sequences were classified into five OTUs (Table S2 and S3) based on the UCLUST clustering method with a 99% sequence identity threshold in QIIME (3). These results indicated that (i) 11 and three individuals were infected with single and multiple *Burkholderia* OTUs, respectively, and (ii) OTU3 was the most frequently detected and present in all European individuals and two out of four Japanese specimens (Table S2). Although the clone library analysis demonstrated that the *Burkholderia* composition is simple in the dock bug, this result needs to be confirmed in a more comprehensive analysis using deep sequencing of the bacterial content in midgut crypts.

The genus *Burkholderia* is grouped into three phylogenetically and ecologically distinct clades (6, 32). The first clade consists of many human, animal, and plant pathogens, including *B. cepacia*, *B. pseudomallei*, and *B. mallei*, designated as the "*Burkholderia cepacia* complex (BCC)" group. The second clade includes a number of plant growth-promoting rhizobac-

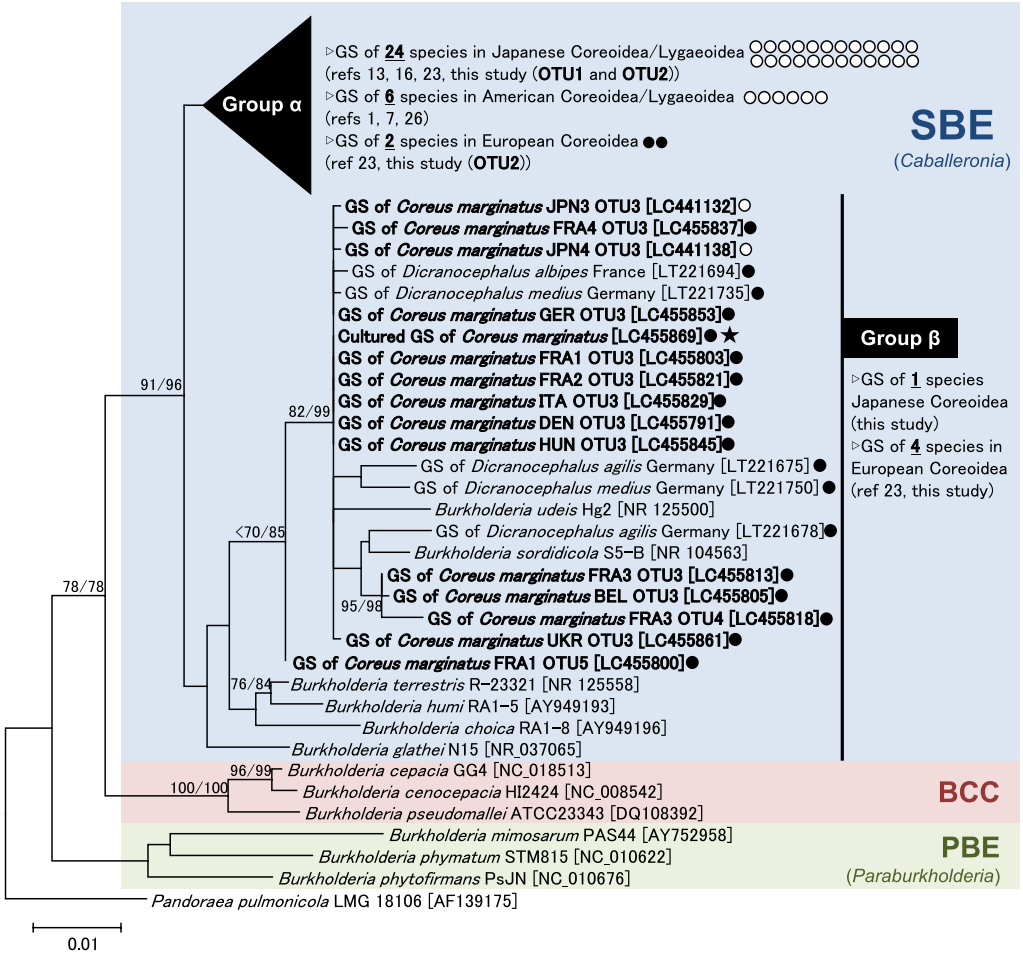


Fig. 2. Molecular phylogeny of gut symbiotic *Burkholderia* of the dock bug shown by a neighbor-joining tree based on 1,332 aligned nucleotide sites of the 16S rRNA gene. The major *Burkholderia* clades (SBE, BCC, and PBE) as well as "SBE Group α" and "SBE Group β" are indicated. SBE Group α is a large group containing the gut symbionts of most Japanese and American species of the Coreoidea/Lygaeoidea (1, 7, 13, 16, 23, 26). An uncompressed tree of this group is shown in Fig. S1. SBE Group β was described as the "Stenocephalidae clade" in a previous study (23) and includes *B. glathei*, *B. sordidicola*, and most of the OTUs detected from European populations of the dock bug. Accession numbers in the DNA database (DDBJ/EMBL/GenBank) are shown in square brackets. Bootstrap values higher than 70% are indicated at the nodes in the order of maximum likelihood/neighbor-joining (1,000 replicates). Maximum likelihood phylogeny was estimated using the neighbor-joining tree as an initial guide tree. OTUs examined in the present study are shown in bold case. Closed circles: symbionts detected from European stinkbug populations. Open circles: symbionts detected from Japanese and American stinkbug populations. Asterisk: a cultured strain isolated from *C. marginatus* collected in Crèche Belle-Image, the CNRS campus, Gif-sur-Yvette, France. GS: gut symbiont.

teria and nodule-forming plant symbionts, assigned as the "plant-associated beneficial and environmental (PBE)" group, which was recently nominated as a novel genus "*Paraburkholderia*" (28). The third clade mainly consists of gut symbionts of the Coreoidea and Lygaeoidea stinkbugs, assigned as SBE or the "*Burkholderia glathei* clade (BGC)", for which the novel genus "*Caballeronia*" has been proposed (5). A recent genome-based phylogenetic study strongly suggested that the *Caballeronia* genus is subdivided into at least two clades: a clade consisting of stinkbug symbionts and leaf-nodule symbionts, and a second clade consisting of *B. glathei*, *B. sordidicola*, and their allied species (31, 32). The former and latter clades are named here as "SBE Group α" and "SBE Group β", respectively. Symbionts of the European spurge bugs are mostly grouped into SBE Group β (23) (Fig. 2).

The phylogenetic placement of the *Burkholderia* OTUs detected from the dock bug is shown in Fig. 2. OTU1 and OTU2, detected in two Japanese populations and one French population of the dock bug, were placed in SBE group α, in which OTUs were clustered with *Burkholderia* detected from Japanese and American coreoid and lygaeoid stinkbugs (Fig. 2, Table S2). The three other OTUs, including OTU3 detected in most European dock bug populations, were placed in SBE group β (Fig. 2). It is important to note that all of the ten insects investigated in seven European countries (France, Germany, Belgium, Italy, Hungary, Denmark, and Ukraine) were almost exclusively associated with *Burkholderia* of SBE group β (Table S2). Based on our previous findings on spurge bugs (23), it is plausible that coreoid stinkbugs inhabiting Europe are consistently associated with this specific clade of *Burkholderia*. Recent worldwide surveys revealed a "region-dependent pattern" of soil microbiota (4, 27), which may affect the region-dependent *Burkholderia* infection of stinkbugs. To clarify this point, further worldwide surveys on both soils and inhabiting stinkbugs are needed.

The nucleotide sequence data of the 16S rRNA gene obtained in the present study have been deposited in the DDBJ/EMBL/GenBank public databases with the accession

numbers LC441114–LC441145 and LC455791–LC455869 (summarized in Table S1).

Acknowledgements

We thank J. Shykoff for insect sampling and R. Cossard for assistance with insect rearing. Fluorescent microscopic observations were performed at the Imagerie-Gif facilities (http://www.i2bc.paris-saclay.fr/spip.php?rubrique184). This study was supported by Japan Society for the Promotion of Science (JSPS) overseas Research fellowship (grant number 20170276) to TO and JSPS-CNRS Bilateral Open Partnership Joint Research Project to PM and YK.

References

1. Boucias, D.G., A. Garcia-Maruniak, R. Cherry, H. Lu, J.E. Maruniak, and V.U. Lietze. 2012. Detection and characterization of bacterial symbionts in the Heteropteran, *Blissus insularis*. FEMS Microbiol. Ecol. 82:629–641.
2. Buchner, P. 1965. Endosymbiosis of Animals with Plant Microorganisms. Interscience, New York.
3. Caporaso, J.G., J. Kuczynski, J. Stombaugh, *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7:335.
4. Delgado-Baquerizo, M., A.M. Oliverio, T.E. Brewer, A. Benavent-González, D.J. Eldridge, R.D. Bardgett, F.T. Maestre, B.K. Singh, and N. Fierer. 2018. A global atlas of the dominant bacteria found in soil. Science 359:320–325.
5. Dobritsa, A.P., and M. Samadpour. 2016. Transfer of eleven species of the genus *Burkholderia* to the genus *Paraburkholderia* and proposal of *Caballeronia* gen. nov. to accommodate twelve species of the genera *Burkholderia* and *Paraburkholderia*. Int. J. Syst. Evol. Microbiol. 66:2836–2846.
6. Estrada-de los Santos, P., F.U. Rojas-Rojas, E.Y. Tapia-García, M.S. Vásquez-Murrieta, and A.M. Hirsch. 2016. To split or not to split: an opinion on dividing the genus *Burkholderia*. Ann. Microbiol. 66:1303–1314.
7. Garcia, J.R., A.M. Laughton, Z. Malik, B.J. Parker, C. Trincot, S.S.L. Chiang, E. Chung, and N.M. Gerardo. 2014. Partner associations across sympatric broad-headed bug species and their environmentally acquired bacterial symbionts. Mol. Ecol. 23:1333–1347.
8. Glasgow, H. 1914. The gastric caeca and the caecal bacteria of the Heteroptera. Biol. Bull. 26:101–171.
9. Henry, T.J. 2009. Biodiversity of Heteroptera, p. 223–263. *In* R.G. Foottit, and P.H. Adler (ed.), Insect Biodiversity: Science and Society. Wiley-Blackwell, Oxford.
10. Hosokawa, T., R. Koga, Y. Kikuchi, X.Y. Meng, and T. Fukatsu. 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc. Natl. Acad. Sci. U.S.A. 107:769–774.
11. Hrušková, M., A. Honěk, and S. Pekár. 2005. *Coreus marginatus* (Heteroptera: Coreidae) as a natural enemy of *Rumex obtusifolius* (Polygonaceae). Acta Oecol. (Montrouge) 28:281–287.
12. Ishikawa, T., M. Takai, and T. Yasunaga. 2012. A Field Guide to Japanese Bugs, Vol. 3. Zenkoku Noson Kyoiku Kyokai, Tokyo.
13. Itoh, H., M. Aita, A. Nagayama, X. Meng, Y. Kamagata, R. Navarro, T. Hori, S. Ohgiya, and Y. Kikuchi 2014. Evidence of environmental and vertical transmission of *Burkholderia* symbionts in the oriental chinch bug, *Cavelerius saccharivorus* (Heteroptera: Blissidae). Appl. Environ. Microbiol. 80:5974–5083.
14. Kikuchi, Y., T. Hosokawa, and T. Fukatsu. 2007. Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. Appl. Environ. Microbiol. 73:4308–4316.
15. Kikuchi, Y. 2009. Endosymbiotic bacteria in insects: their diversity and culturability. Microbes Environ. 24:195–204.
16. Kikuchi, Y., T. Hosokawa, and T. Fukatsu. 2011. An ancient but promiscuous host-symbiont association between *Burkholderia* gut symbionts and their heteropteran hosts. ISME J. 5:446–460.
17. Kikuchi, Y., T. Hosokawa, and T. Fukatsu. 2011. Specific developmental window for establishment of an insect-microbe gut symbiosis. Appl. Environ. Microbiol. 77:4075–4081.
18. Kikuchi, Y., and T. Fukatsu. 2014. Live imaging of symbiosis: spatio-temporal infection dynamics of a GFP-labelled *Burkholderia* symbiont in the bean bug *Riptortus pedestris*. Mol. Ecol. 23:1445–1456.
19. Kikuchi, Y., S.S. Prado, and T.M. Jenkins. 2018. Symbiotic microorganisms associated with Pentatomoidea, p. 643–674. *In* J.E. McPherson (ed.), Invasive Stink Bugs and Related Species (Pentatomoidea): Biology, Higher Systematics, Semiochemistry, and Management. CRC Press, Boca Raton.
20. Kuechler, S.M., K. Dettner, and S. Kehl. 2010. Molecular characterization and localization of the obligate endosymbiotic bacterium in the birch catkin bug *Kleidocerys resedae* (Heteroptera: Lygaeidae, Ischnorhynchinae). FEMS Microbiol. Ecol. 73:408–418.
21. Kuechler, S.M., K. Dettner, and S. Kehl. 2011. Characterization of an obligate intracellular bacterium in the midgut epithelium of the bulrush bug *Chilacis typhae* (Heteroptera, Lygaeidae, Artheneinae). Appl. Environ. Microbiol. 77:2869–2876.
22. Kuechler, S.M., P. Renz, K. Dettner, and S. Kehl. 2012. Diversity of symbiotic organs and bacterial endosymbionts of lygaeoid bugs of the families Blissidae and Lygaeidae (Hemiptera: Heteroptera: Lygaeoidea). Appl. Environ. Microbiol. 78:2648–2659.
23. Kuechler, S.M., Y. Matsuura, K. Dettner, and Y. Kikuchi. 2016. Phylogenetically diverse *Burkholderia* associated with midgut crypts of spurge bugs, *Dicranocephalus* spp. (Heteroptera: Stenocephalidae). Microbes Environ. 31:145–153.
24. Matsuura, Y., Y. Kikuchi, T. Hosokawa, R. Koga, X.Y. Meng, Y. Kamagata, N. Nikoh, and T. Fukatsu. 2012. Evolution of symbiotic organs and endosymbionts in lygaeid stinkbugs. ISME J. 6:397–409.
25. Miyamoto, S. 1961. Comparative morphology of alimentary organs of Heteroptera, with the phylogenetic consideration. Sieboldia 2:197–259.
26. Olivier-Espejel, S., Z.L. Sabree, K. Noge, and J.X. Becerra. 2011. Gut microbiota in nymph and adults of the giant mesquite bug (*Thasus neocalifornicus*) (Heteroptera: Coreidae) is dominated by *Burkholderia* acquired *de novo* every generation. Environ. Entomol. 40:1102–1110.
27. Ramirez, K.S., C.G. Knight, M. De Hollander, F.Q. Brearley, B. Constantinides, A. Cotton, S. Creer, T.W. Crowther, J. Davison, and M. Delgado-Baquerizo. 2018. Detecting macroecological patterns in bacterial communities across independent studies of global soils. Nat. Microbiol. 3:189.
28. Sawana, A., M. Adeolu, and R.S. Gupta. 2014. Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. Front. Genet. 5:429.
29. Tago, K., H. Itoh, Y. Kikuchi, *et al.* 2014. A fine-scale phylogenetic analysis of free-living *Burkholderia* species in sugarcane field soil. Microbes Environ. 29:434–437.
30. Takeshita, K., and Y. Kikuchi. 2017. *Riptortus pedestris* and *Burkholderia* symbiont: an ideal model system for insect-microbe symbiotic associations. Res. Microbiol. 168:175–187.
31. Takeshita, K., H. Tamaki, T. Ohbayashi, X.Y. Meng, T. Sone, Y. Mitani, C. Peeters, Y. Kikuchi, and P. Vandamme 2018. *Burkholderia insecticola* sp. nov., a gut symbiotic bacterium of the bean bug *Riptortus pedestris*. Int. J. Syst. Evol. Microbiol. 68:2370–2374.
32. Vandamme, P., C. Peeters, B. De Smet, *et al.* 2017. Comparative genomics of *Burkholderia singularis* sp. nov., a low G+C content, free-living bacterium that defies taxonomic dissection of the genus *Burkholderia*. Front. Microbiol. 8:1679.

Annexes

Annexe 1: List of essential genes in *B. insecticola* identified by El-ARTIST.

		NC_021287.1	NC_021294.1	NC_021288.1	NC_021289.1	NC_021295.1	
	Essentiality score	Chromosome 1	Chromosome 2	Chromosome 3	Plasmid 1	Plasmid 2	Total
Non-essential genes	1	2165	1112	650	792	247	4966
Essential genes	2	479	152	116	323	10	1080
Domain-essential genes	3	74	55	22	42	5	198

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS00005	2	NC_021287.1	354	1952	*dnaA*	chromosomal replication initiator protein DnaA	Replication, recombination and repair
BRPE64_RS00010	2	NC_021287.1	2204	3307	*dnaN*	DNA polymerase III subunit beta	Replication, recombination and repair
BRPE64_RS00015	2	NC_021287.1	3420	5903	*gyrB*	DNA gyrase subunit B	Replication, recombination and repair
BRPE64_RS00120	2	NC_021287.1	25357	28002	*topB*	DNA topoisomerase	Replication, recombination and repair
BRPE64_RS00140	2	NC_021287.1	30535	31527	*fmt*	methionyl-tRNA formyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS00170	2	NC_021287.1	37898	38614	-	Response regulator containing CheY-like receiver	Signal transduction mechanisms
BRPE64_RS00210	2	NC_021287.1	45077	46225	*mrdB*	Rod shape-determining protein RodA	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS00215	2	NC_021287.1	46251	48560	*mrdA*	penicillin-binding protein 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS00220	2	NC_021287.1	48652	49164	*mreD*	Rod shape-determining protein MreD	Cell wall/membrane/envelope biogenesis
BRPE64_RS00225	2	NC_021287.1	49161	50327	*mreC*	Rod shape-determining protein MreC	Cell wall/membrane/envelope biogenesis
BRPE64_RS00230	2	NC_021287.1	50526	51569	*mreB*	Rod shape-determining protein MreB	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS00235	2	NC_021287.1	51945	52244	-	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C	Translation, ribosomal structure and biogenesis
BRPE64_RS00240	2	NC_021287.1	52306	53802	-	glutamyl-tRNA(Gln) amidotransferase subunit A	Translation, ribosomal structure and biogenesis
BRPE64_RS00245	2	NC_021287.1	53805	55274	-	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	Translation, ribosomal structure and biogenesis
BRPE64_RS00295	2	NC_021287.1	63461	64759	-	integral membrane sensor signal transduction histidine kinase	Signal transduction mechanisms

BRPE64_RS00300	2	NC_021287.1	64804	65346	-	two component transcriptional regulator Fis family	Transcription
BRPE64_RS00680	2	NC_021287.1	146373	148157	*argS*	arginine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS00765	2	NC_021287.1	164340	165161	-	type III pantothenate kinase	Transcription
BRPE64_RS00770	2	NC_021287.1	165158	166060	-	biotin--acetyl-CoA-carboxylase ligase	Coenzyme transport and metabolism
BRPE64_RS00845	2	NC_021287.1	177360	178184	*lipB*	octanoyltransferase	Coenzyme transport and metabolism
BRPE64_RS00850	2	NC_021287.1	178177	179175	*lipA*	lipoyl synthase	Coenzyme transport and metabolism
BRPE64_RS00930	2	NC_021287.1	196900	198486	-	carboxyl-terminal protease	Cell wall/membrane/envelope biogenesis
BRPE64_RS00975	2	NC_021287.1	203916	204896	*bioC*	malonyl-CoA O-methyltransferase BioC	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS01085	2	NC_021287.1	225891	226826	*rpoH*	RNA polymerase sigma factor	Transcription
BRPE64_RS01125	2	NC_021287.1	233678	234985	*ftsY*	signal recognition particle receptor FtsY	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS01135	2	NC_021287.1	235940	236443	*coaD*	phosphopantetheine adenylyltransferase	Coenzyme transport and metabolism
BRPE64_RS01150	2	NC_021287.1	237903	238502	*pth*	peptidyl-tRNA hydrolase	Translation, ribosomal structure and biogenesis
BRPE64_RS01155	2	NC_021287.1	238634	239254	*rplY*	50S ribosomal protein L25	Translation, ribosomal structure and biogenesis
BRPE64_RS01160	2	NC_021287.1	239406	240359	-	ribose-phosphate pyrophosphokinase	Amino acid transport and metabolism
BRPE64_RS01175	2	NC_021287.1	240653	241534	*ispE*	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	Lipid transport and metabolism
BRPE64_RS01180	2	NC_021287.1	241564	242190	*lolB*	outer membrane lipoprotein LolB	Cell wall/membrane/envelope biogenesis
BRPE64_RS01185	2	NC_021287.1	242190	244049	-	TPR repeat-containing protein	General function prediction only
BRPE64_RS01210	2	NC_021287.1	247667	248635	-	HPr kinase/phosphorylase	Signal transduction mechanisms
BRPE64_RS01215	2	NC_021287.1	248822	249277	*ptsN*	putative PTS IIA-like nitrogen-regulatory protein PtsN	Carbohydrate transport and metabolism
BRPE64_RS01230	2	NC_021287.1	251966	252748	-	ABC transporter related protein	General function prediction only
BRPE64_RS01235	2	NC_021287.1	252745	253440	*yhbN*	hypothetical protein	Function unknown
BRPE64_RS01240	2	NC_021287.1	253471	254073	-	hypothetical protein	Function unknown
BRPE64_RS01245	2	NC_021287.1	254076	254612	-	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase YrbI family	General function prediction only
BRPE64_RS01250	2	NC_021287.1	254639	255622	-	sugar isomerase KpsF/GutQ family	Cell wall/membrane/envelope biogenesis

BRPE64_RS01295	2	NC_021287.1	266689	267222	*ssb*	single-stranded DNA-binding protein	Replication, recombination and repair
BRPE64_RS01410	2	NC_021287.1	295415	296146	-	ubiquinone/menaquinone biosynthesis methyltransferase ubiE	Coenzyme transport and metabolism
BRPE64_RS01420	2	NC_021287.1	297350	298078	-	hypothetical protein	Function unknown
BRPE64_RS01425	2	NC_021287.1	298099	299673	-	probable ubiquinone biosynthesis protein UbiB	General function prediction only
BRPE64_RS01435	2	NC_021287.1	300451	300849	-	hypothetical protein	Function unknown
BRPE64_RS01440	2	NC_021287.1	301022	301684	-	hypothetical protein	Function unknown
BRPE64_RS01445	2	NC_021287.1	301778	303580	*aspS*	aspartate--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS01555	2	NC_021287.1	329198	330268	*pyrC*	dihydroorotase	Nucleotide transport and metabolism
BRPE64_RS01570	2	NC_021287.1	332662	333090	*rplM*	50S ribosomal protein L13	Translation, ribosomal structure and biogenesis
BRPE64_RS01575	2	NC_021287.1	333102	333494	*rpsI*	30S ribosomal protein S9	Translation, ribosomal structure and biogenesis
BRPE64_RS01580	2	NC_021287.1	333704	334069	-	putative iron-sulfur cluster insertion protein ErpA 1	Function unknown
BRPE64_RS01590	2	NC_021287.1	335648	336904	*tyrS*	tyrosine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS01655	2	NC_021287.1	349176	350741	*purH*	bifunctional purine biosynthesis protein PurH	Nucleotide transport and metabolism
BRPE64_RS01670	2	NC_021287.1	352350	353573	*ubiH*	ubiquinone biosynthesis hydroxylase UbiH/UbiF/VisC/COQ6	Energy production and conversion
BRPE64_RS01705	2	NC_021287.1	360423	361553	-	tRNA-specific 2-thiouridylase MnmA	Translation, ribosomal structure and biogenesis
BRPE64_RS01760	2	NC_021287.1	370623	371573	*secF*	protein translocase subunit SecF	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS01765	2	NC_021287.1	371600	373666	*secD*	protein translocase subunit SecD	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS01770	2	NC_021287.1	373804	374133	*yajC*	preprotein translocase YajC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS01805	2	NC_021287.1	382751	383608	*ubiA*	4-hydroxybenzoate octaprenyltransferase	Coenzyme transport and metabolism
BRPE64_RS01885	2	NC_021287.1	407731	408627	-	probable inorganic polyphosphate/ATP-NAD kinase	Carbohydrate transport and metabolism
BRPE64_RS01895	2	NC_021287.1	409972	411039	*hemH*	ferrochelatase	Coenzyme transport and metabolism
BRPE64_RS01915	2	NC_021287.1	413093	415048	*dnaK*	chaperone protein DnaK	Posttranslational modification, protein turnover, chaperones

BRPE64_RS01920	2	NC_021287.1	415301	416437	-	chaperone protein DnaJ	Posttranslational modification, protein turnover, chaperones
BRPE64_RS01970	2	NC_021287.1	425810	426367	-	D,D-heptose 1,7-bisphosphate phosphatase	Amino acid transport and metabolism
BRPE64_RS01975	2	NC_021287.1	426377	428476	*glyS*	glycine--tRNA ligase beta subunit	Translation, ribosomal structure and biogenesis
BRPE64_RS01980	2	NC_021287.1	428620	429522	-	hypothetical protein	-
BRPE64_RS01985	2	NC_021287.1	429719	431410	*lnt*	apolipoprotein N-acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS02005	2	NC_021287.1	433804	434925	-	PhoH family protein	Signal transduction mechanisms
BRPE64_RS02115	2	NC_021287.1	456859	457440	*folK*	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	Coenzyme transport and metabolism
BRPE64_RS02125	2	NC_021287.1	459118	459804	-	HAD-superfamily subfamily IB hydrolase TIGR01490	Amino acid transport and metabolism
BRPE64_RS02340	2	NC_021287.1	504058	504948	*purC*	phosphoribosylaminoimidazole-succinocarboxamide synthase	Nucleotide transport and metabolism
BRPE64_RS02345	2	NC_021287.1	505002	505523	*purE*	N5-carboxyaminoimidazole ribonucleotide mutase	Nucleotide transport and metabolism
BRPE64_RS02355	2	NC_021287.1	506834	507865	-	Sua5/YciO/YrdC/YwlC family protein	Translation, ribosomal structure and biogenesis
BRPE64_RS02490	2	NC_021287.1	541419	542525	-	S-(Hydroxymethyl)glutathione dehydrogenase/class III alcohol dehydrogenase	Energy production and conversion
BRPE64_RS02740	2	NC_021287.1	603803	604249	*dut*	deoxyuridine 5'-triphosphate nucleotidohydrolase	Nucleotide transport and metabolism
BRPE64_RS02750	2	NC_021287.1	605267	606478	-	phosphopantothenoylcysteine decarboxylase/phosphopantothenate--cysteine ligase	Coenzyme transport and metabolism
BRPE64_RS02755	2	NC_021287.1	606558	607064	*lspA*	lipoprotein signal peptidase	Cell wall/membrane/envelope biogenesis
BRPE64_RS02760	2	NC_021287.1	607065	609896	*ileS*	isoleucine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS02765	2	NC_021287.1	609998	610993	*ribF*	FMN adenylyltransferase	Coenzyme transport and metabolism
BRPE64_RS02785	2	NC_021287.1	614758	615954	-	fatty acid desaturase	Lipid transport and metabolism
BRPE64_RS03015	2	NC_021287.1	658245	659393	-	permease YjgP/YjgQ family protein	General function prediction only
BRPE64_RS03020	2	NC_021287.1	659398	660489	-	permease YjgP/YjgQ family protein	General function prediction only
BRPE64_RS03055	2	NC_021287.1	666308	667210	*lgt*	prolipoprotein diacylglyceryl transferase	Cell wall/membrane/envelope biogenesis

BRPE64_RS03150	2	NC_021287.1	683058	683891	*panC*	pantothenate synthetase	Coenzyme transport and metabolism
BRPE64_RS03155	2	NC_021287.1	683961	684788	-	segregation and condensation protein A	Function unknown
BRPE64_RS03160	2	NC_021287.1	684859	685044	-	hypothetical protein	-
BRPE64_RS03205	2	NC_021287.1	697198	699315	*metG*	methionine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS03225	2	NC_021287.1	702561	703130	*dcd*	deoxycytidine triphosphate deaminase	Nucleotide transport and metabolism
BRPE64_RS03230	2	NC_021287.1	703248	705554	-	ornithine decarboxylase	Amino acid transport and metabolism
BRPE64_RS03270	2	NC_021287.1	714126	715133	*hemC*	porphobilinogen deaminase	Coenzyme transport and metabolism
BRPE64_RS03310	2	NC_021287.1	722776	723774	-	thioredoxin reductase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS03350	2	NC_021287.1	733005	734303	*serS*	serine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS03390	2	NC_021287.1	738600	738854	*minE*	cell division topological specificity factor	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS03680	2	NC_021287.1	800827	801780	-	hypothetical protein	Function unknown
BRPE64_RS03685	2	NC_021287.1	801790	802464	*gmk*	guanylate kinase	Nucleotide transport and metabolism
BRPE64_RS03690	2	NC_021287.1	802535	802738	*rpoZ*	DNA-directed RNA polymerase subunit omega	Transcription
BRPE64_RS03695	2	NC_021287.1	802884	805208	*spoT*	(P)ppGpp synthetase I (GTP pyrophosphokinase) SpoT/RelA	Transcription
BRPE64_RS03720	2	NC_021287.1	806860	808017	-	outer membrane porin protein 32	Cell wall/membrane/envelope biogenesis
BRPE64_RS03900	2	NC_021287.1	846241	847866	-	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Coenzyme transport and metabolism
BRPE64_RS03915	2	NC_021287.1	850914	851351	*nusB*	N utilization substance protein B homolog	Transcription
BRPE64_RS03920	2	NC_021287.1	851348	851863	-	6,7-dimethyl-8-ribityllumazine synthase	Coenzyme transport and metabolism
BRPE64_RS03925	2	NC_021287.1	851955	853118	-	bifunctional riboflavin biosynthesis protein RibBA	Coenzyme transport and metabolism
BRPE64_RS03930	2	NC_021287.1	853282	853896	-	riboflavin synthase alpha subunit	Coenzyme transport and metabolism
BRPE64_RS03935	2	NC_021287.1	853928	855049	*ribD*	riboflavin biosynthesis protein RibD	Coenzyme transport and metabolism
BRPE64_RS03940	2	NC_021287.1	855068	856351	*hemL*	glutamate-1-semialdehyde 2,1-aminomutase	Coenzyme transport and metabolism
BRPE64_RS04055	2	NC_021287.1	882440	884203	-	binding-protein-dependent transport systems	Inorganic ion transport and metabolism

						inner membrane component	
BRPE64_RS04060	2	NC_021287.1	884225	885571	-	ABC nitrate/sulfonate/bicarbonate family transporter ATPase subunit	Inorganic ion transport and metabolism
BRPE64_RS04240	2	NC_021287.1	916210	918189	*parE*	DNA topoisomerase	Replication, recombination and repair
BRPE64_RS04245	2	NC_021287.1	918228	920546	*parC*	DNA topoisomerase IV A subunit	Replication, recombination and repair
BRPE64_RS04360	2	NC_021287.1	943748	944701	-	transaldolase	Carbohydrate transport and metabolism
BRPE64_RS04480	2	NC_021287.1	971297	974824	*dnaE*	DNA polymerase III alpha subunit	Replication, recombination and repair
BRPE64_RS04505	2	NC_021287.1	979335	981119	*msbA*	lipid A ABC exporter fused ATPase and inner membrane subunits MsbA	Defense mechanisms
BRPE64_RS04535	2	NC_021287.1	985546	986259	*nadD*	probable nicotinate-nucleotide adenylyltransferase	Coenzyme transport and metabolism
BRPE64_RS04540	2	NC_021287.1	986250	987209	*hemF*	coproporphyrinogen -III oxidase aerobic	Coenzyme transport and metabolism
BRPE64_RS04545	2	NC_021287.1	987358	988647	*purD*	phosphoribosylamine--glycine ligase	Nucleotide transport and metabolism
BRPE64_RS04550	2	NC_021287.1	988916	989644	-	probable transcriptional regulatory protein Bphy_2064	Function unknown
BRPE64_RS04835	2	NC_021287.1	1057904	1059319	*glnA*	glutamine synthetase	Amino acid transport and metabolism
BRPE64_RS04890	2	NC_021287.1	1071880	1072734	*folD*	bifunctional protein FolD	Coenzyme transport and metabolism
BRPE64_RS04910	2	NC_021287.1	1076365	1078893	-	multi-sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS04915	2	NC_021287.1	1079186	1081882	*aceE*	pyruvate dehydrogenase E1 component	Energy production and conversion
BRPE64_RS04920	2	NC_021287.1	1081962	1083599	*aceF*	pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase	Energy production and conversion
BRPE64_RS05045	2	NC_021287.1	1107638	1108519	*murI*	glutamate racemase	Cell wall/membrane/envelope biogenesis
BRPE64_RS05350	2	NC_021287.1	1164880	1166553	-	electron-transferring-flavoproteindehydrogenase	Energy production and conversion
BRPE64_RS05400	2	NC_021287.1	1174483	1176390	*thrS*	threonine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS05405	2	NC_021287.1	1176492	1176965	*infC*	translation initiation factor IF-3	Translation, ribosomal structure and biogenesis
BRPE64_RS05410	2	NC_021287.1	1177209	1177406	*rpmI*	50S ribosomal protein L35	Translation, ribosomal structure and biogenesis

BRPE64_RS05415	2	NC_021287.1	1177435	1177794	-	hypothetical protein	-
BRPE64_RS05420	2	NC_021287.1	1177989	1179002	*pheS*	phenylalanine--tRNA ligase alpha subunit	Translation, ribosomal structure and biogenesis
BRPE64_RS05425	2	NC_021287.1	1179081	1181513	*pheT*	phenylalanine--tRNA ligase beta subunit	Translation, ribosomal structure and biogenesis
BRPE64_RS05580	2	NC_021287.1	1206754	1207569	-	peptidase M22 glycoprotease	Posttranslational modification, protein turnover, chaperones
BRPE64_RS05585	2	NC_021287.1	1207566	1208057	*rimI*	ribosomal-protein-alanine acetyltransferase	General function prediction only
BRPE64_RS05590	2	NC_021287.1	1208047	1209066	-	phage SPO1 DNA polymerase-related protein	Replication, recombination and repair
BRPE64_RS05615	2	NC_021287.1	1212426	1213733	-	major facilitator superfamily (MFS) transporter	Amino acid transport and metabolism
BRPE64_RS05670	2	NC_021287.1	1224868	1225695	*dapD*	2,3 4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	Amino acid transport and metabolism
BRPE64_RS05685	2	NC_021287.1	1228362	1231877	-	chromosome partition protein Smc	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS05695	2	NC_021287.1	1233523	1235595	*ligA*	DNA ligase	Replication, recombination and repair
BRPE64_RS05725	2	NC_021287.1	1242861	1243613	*rpsB*	30S ribosomal protein S2	Translation, ribosomal structure and biogenesis
BRPE64_RS05730	2	NC_021287.1	1243756	1244637	*tsf*	elongation factor Ts	Translation, ribosomal structure and biogenesis
BRPE64_RS05735	2	NC_021287.1	1244875	1245588	*pyrH*	uridylate kinase	Nucleotide transport and metabolism
BRPE64_RS05740	2	NC_021287.1	1245681	1246241	*frr*	ribosome-recycling factor	Translation, ribosomal structure and biogenesis
BRPE64_RS05745	2	NC_021287.1	1246324	1247106	-	isoprenyl transferase	Lipid transport and metabolism
BRPE64_RS05750	2	NC_021287.1	1247100	1247912	-	phosphatidate cytidylyltransferase	Lipid transport and metabolism
BRPE64_RS05755	2	NC_021287.1	1247931	1249148	*dxr*	1-deoxy-D-xylulose 5-phosphate reductoisomerase	Lipid transport and metabolism
BRPE64_RS05760	2	NC_021287.1	1249156	1250538	-	membrane-associated zinc metalloprotease	Cell wall/membrane/envelope biogenesis
BRPE64_RS05765	2	NC_021287.1	1250619	1252931	-	outer membrane protein assembly factor BamA	Cell wall/membrane/envelope biogenesis
BRPE64_RS05770	2	NC_021287.1	1253014	1253559	-	outer membrane chaperone Skp	Cell wall/membrane/envelope biogenesis
BRPE64_RS05775	2	NC_021287.1	1253590	1254681	*lpxD*	UDP-3-O-acylglucosamine N-acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS05780	2	NC_021287.1	1254875	1255333	*fabZ*	3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ	Lipid transport and metabolism
BRPE64_RS05785	2	NC_021287.1	1255396	1256184	*lpxA*	acyl-[acyl-carrier-protein]--UDP-N-	Cell wall/membrane/envelope biogenesis

						acetylglucosamine O-acyltransferase	
BRPE64_RS05790	2	NC_021287.1	1256197	1257363	*lpxB*	lipid-A-disaccharide synthase	Cell wall/membrane/envelope biogenesis
BRPE64_RS05830	2	NC_021287.1	1264719	1265156	-	streptomyces cyclase/dehydrase superfamily	Lipid transport and metabolism
BRPE64_RS05845	2	NC_021287.1	1266671	1268131	*guaB*	inosine-5'-monophosphate dehydrogenase	Nucleotide transport and metabolism
BRPE64_RS31850	2	NC_021287.1	1268171	1269001	-	hypothetical protein	-
BRPE64_RS05855	2	NC_021287.1	1269219	1270802	*guaA*	GMP synthase	Nucleotide transport and metabolism
BRPE64_RS05900	2	NC_021287.1	1279823	1280215	-	DnaJ-like subfamily C member 28 conserved domain protein	Cell motility
BRPE64_RS06290	2	NC_021287.1	1349868	1350239	*rpsF*	30S ribosomal protein S6	Translation, ribosomal structure and biogenesis
BRPE64_RS06295	2	NC_021287.1	1350285	1350584	-	putative primosomal replication protein N PriB	Replication, recombination and repair
BRPE64_RS06310	2	NC_021287.1	1351469	1352857	*dnaB*	replicative DNA helicase	Replication, recombination and repair
BRPE64_RS06345	2	NC_021287.1	1359698	1360594	-	hypothetical protein	-
BRPE64_RS06350	2	NC_021287.1	1360607	1361653	-	hypothetical protein	-
BRPE64_RS06355	2	NC_021287.1	1361650	1362636	-	formyl transferase domain protein	Translation, ribosomal structure and biogenesis
BRPE64_RS06360	2	NC_021287.1	1362633	1363670	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS06365	2	NC_021287.1	1363695	1364864	*yfbE*	DegT/DnrJ/EryC1/StrS aminotransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS06375	2	NC_021287.1	1365413	1367086	-	glycosyl transferase family 39	Cell wall/membrane/envelope biogenesis
BRPE64_RS06385	2	NC_021287.1	1368315	1369553	-	aminotransferase AlaT	Amino acid transport and metabolism
BRPE64_RS06390	2	NC_021287.1	1369578	1370900	-	homoserine dehydrogenase	Amino acid transport and metabolism
BRPE64_RS06395	2	NC_021287.1	1370916	1372361	*thrC*	threonine synthase	Amino acid transport and metabolism
BRPE64_RS06455	2	NC_021287.1	1387959	1388693	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	Lipid transport and metabolism
BRPE64_RS06460	2	NC_021287.1	1388727	1389218	*ispF*	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	Lipid transport and metabolism
BRPE64_RS06580	2	NC_021287.1	1411528	1413162	*pgi*	glucose-6-phosphate isomerase	Carbohydrate transport and metabolism
BRPE64_RS06615	2	NC_021287.1	1424873	1425409	-	probable intracellular septation protein A	Cell cycle control, cell division,

							chromosome partitioning
BRPE64_RS06690	2	NC_021287.1	1440297	1440917	*tmk*	thymidylate kinase	Nucleotide transport and metabolism
BRPE64_RS06695	2	NC_021287.1	1440930	1441970	*holB*	DNA polymerase III delta prime subunit	Replication, recombination and repair
BRPE64_RS07105	2	NC_021287.1	1535284	1535808	-	phenylacetic acid degradation protein PaaD	General function prediction only
BRPE64_RS07110	2	NC_021287.1	1535809	1536618	-	phenylacetate-CoA oxygenase PaaI subunit	Function unknown
BRPE64_RS07115	2	NC_021287.1	1536632	1536916	-	phenylacetate-CoA oxygenase PaaH subunit	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS07505	2	NC_021287.1	1626448	1627281	-	outer membrane protein assembly factor BamD	General function prediction only
BRPE64_RS07530	2	NC_021287.1	1633539	1635431	-	probable potassium transport system protein kup	Inorganic ion transport and metabolism
BRPE64_RS07540	2	NC_021287.1	1636365	1637711	*purA*	adenylosuccinate synthetase	Nucleotide transport and metabolism
BRPE64_RS07545	2	NC_021287.1	1637875	1639026	-	ATP phosphoribosyltransferase regulatory subunit	Amino acid transport and metabolism
BRPE64_RS07550	2	NC_021287.1	1639153	1639344	-	hypothetical protein	Function unknown
BRPE64_RS07555	2	NC_021287.1	1639396	1640298	*hflC*	band 7 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS07560	2	NC_021287.1	1640310	1641710	*hflK*	protease FtsH subunit HflK	Posttranslational modification, protein turnover, chaperones
BRPE64_RS07570	2	NC_021287.1	1643126	1643362	-	hypothetical protein	-
BRPE64_RS07575	2	NC_021287.1	1643548	1644888	*der*	GTPase Der	General function prediction only
BRPE64_RS07580	2	NC_021287.1	1645290	1646435	-	outer membrane protein assembly factor BamB	Function unknown
BRPE64_RS07585	2	NC_021287.1	1646568	1647197	-	hypothetical protein	Function unknown
BRPE64_RS07590	2	NC_021287.1	1647298	1648647	*hisS*	histidine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS07595	2	NC_021287.1	1648654	1649970	*ispG*	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	Lipid transport and metabolism
BRPE64_RS07600	2	NC_021287.1	1650105	1651253	-	transcriptional regulator XRE family	Function unknown
BRPE64_RS07605	2	NC_021287.1	1651435	1652577	-	dual-specificity RNA methyltransferase RlmN	General function prediction only
BRPE64_RS07610	2	NC_021287.1	1652732	1653157	*ndk*	nucleoside diphosphate kinase	Nucleotide transport and metabolism
BRPE64_RS07695	2	NC_021287.1	1668399	1670768	*dnaX*	DNA polymerase III subunits gamma and tau	Replication, recombination and repair

BRPE64_RS07705	2	NC_021287.1	1672020	1673285	*rho*	transcription termination factor Rho	Transcription
BRPE64_RS07715	2	NC_021287.1	1674408	1674662	-	50S ribosomal protein L31 type B	Translation, ribosomal structure and biogenesis
BRPE64_RS07720	2	NC_021287.1	1674947	1676710	-	putative inner membrane protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS07795	2	NC_021287.1	1696996	1698018	*pyrD*	dihydroorotate dehydrogenase	Nucleotide transport and metabolism
BRPE64_RS07970	2	NC_021287.1	1735999	1737438	-	dihydrolipoyl dehydrogenase	Energy production and conversion
BRPE64_RS07975	2	NC_021287.1	1737536	1738837	*sucB*	2-oxoglutarate dehydrogenase E2 subunit dihydrolipoamide succinyltransferase	Energy production and conversion
BRPE64_RS07980	2	NC_021287.1	1738934	1741792	*sucA*	2-oxoglutarate dehydrogenase E1 subunit	Energy production and conversion
BRPE64_RS07985	2	NC_021287.1	1742123	1743949	-	GTP-binding protein TypA	Signal transduction mechanisms
BRPE64_RS08015	2	NC_021287.1	1750266	1750679	*rbfA*	ribosome-binding factor A	Translation, ribosomal structure and biogenesis
BRPE64_RS08020	2	NC_021287.1	1750775	1753765	*infB*	translation initiation factor IF-2	Translation, ribosomal structure and biogenesis
BRPE64_RS08025	2	NC_021287.1	1753859	1755334	*nusA*	NusA antitermination factor	Transcription
BRPE64_RS08030	2	NC_021287.1	1755331	1755789	-	ribosome maturation factor RimP	Function unknown
BRPE64_RS08040	2	NC_021287.1	1758224	1759231	-	chromosome segregation and condensation protein ScpB	Transcription
BRPE64_RS08120	2	NC_021287.1	1768446	1769855	*gltX*	glutamate--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS08155	2	NC_021287.1	1776307	1776648	-	ferredoxin 2Fe-2S type ISC system	Energy production and conversion
BRPE64_RS08160	2	NC_021287.1	1776676	1778541	*hscA*	chaperone protein HscA homolog	Posttranslational modification, protein turnover, chaperones
BRPE64_RS08165	2	NC_021287.1	1778578	1779108	*hscB*	Co-chaperone protein HscB homolog	Posttranslational modification, protein turnover, chaperones
BRPE64_RS08170	2	NC_021287.1	1779192	1779515	*iscA*	iron-sulfur cluster assembly protein IscA	Function unknown
BRPE64_RS08175	2	NC_021287.1	1779597	1780025	*iscU*	FeS cluster assembly scaffold IscU	Energy production and conversion
BRPE64_RS08180	2	NC_021287.1	1780074	1781297	*iscS*	cysteine desulfurase	Amino acid transport and metabolism
BRPE64_RS08185	2	NC_021287.1	1781381	1781908	*iscR*	transcriptional regulator BadM/Rrf2 family	Transcription
BRPE64_RS08365	2	NC_021287.1	1816686	1818158	*nuoN*	NADH-quinone oxidoreductase subunit N	Energy production and conversion

BRPE64_RS08370	2	NC_021287.1	1818184	1819683	*nuoM*	NADH dehydrogenase I chain M	Energy production and conversion
BRPE64_RS08375	2	NC_021287.1	1819697	1821766	*nuoL*	proton-translocating NADH-quinone oxidoreductase chain L	Energy production and conversion
BRPE64_RS08380	2	NC_021287.1	1821784	1822089	*nuoK*	NADH-quinone oxidoreductase subunit K	Energy production and conversion
BRPE64_RS08385	2	NC_021287.1	1822110	1822787	*nuoJ*	NADH dehydrogenase subunit J	Energy production and conversion
BRPE64_RS08390	2	NC_021287.1	1822972	1823460	*nuoI*	NADH-quinone oxidoreductase subunit I	Energy production and conversion
BRPE64_RS08395	2	NC_021287.1	1823485	1824549	*nuoH*	NADH-quinone oxidoreductase subunit H	Energy production and conversion
BRPE64_RS08400	2	NC_021287.1	1824552	1826891	*nuoG*	NADH dehydrogenase subunit G	Energy production and conversion
BRPE64_RS08405	2	NC_021287.1	1826944	1828266	*nuoF*	NADH dehydrogenase subunit F	Energy production and conversion
BRPE64_RS08410	2	NC_021287.1	1828263	1828748	*nuoE*	NADH dehydrogenase I chain E	Energy production and conversion
BRPE64_RS08415	2	NC_021287.1	1828902	1830155	-	NADH-quinone oxidoreductase subunit D	Energy production and conversion
BRPE64_RS08420	2	NC_021287.1	1830165	1830767	*nuoC*	NADH-quinone oxidoreductase subunit C	Energy production and conversion
BRPE64_RS08425	2	NC_021287.1	1830801	1831280	*nuoB*	NADH-quinone oxidoreductase subunit B	Energy production and conversion
BRPE64_RS08430	2	NC_021287.1	1831341	1831700	-	hypothetical protein	-
BRPE64_RS08440	2	NC_021287.1	1832125	1832499	*secG*	preprotein translocase SecG subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS08445	2	NC_021287.1	1832581	1833354	*tpiA*	triosephosphate isomerase	Carbohydrate transport and metabolism
BRPE64_RS08455	2	NC_021287.1	1834738	1836876	*pnp*	polyribonucleotide nucleotidyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS08460	2	NC_021287.1	1837191	1837460	*rpsO*	30S ribosomal protein S15	Translation, ribosomal structure and biogenesis
BRPE64_RS08505	2	NC_021287.1	1848419	1849216	*fabI*	enoyl-[acyl-carrier-protein] reductase	Lipid transport and metabolism
BRPE64_RS08580	2	NC_021287.1	1860974	1862224	-	aspartokinase	Amino acid transport and metabolism
BRPE64_RS08585	2	NC_021287.1	1862565	1864004	*tilS*	tRNA(Ile)-lysidine synthase	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS08590	2	NC_021287.1	1864038	1865009	*accA*	acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	Lipid transport and metabolism
BRPE64_RS08600	2	NC_021287.1	1866261	1867658	*cysS*	cysteine--tRNA ligase	Translation, ribosomal structure and biogenesis

BRPE64_RS08605	2	NC_021287.1	1868146	1868946	-	tetratricopeptide TPR_2 repeat protein	General function prediction only
BRPE64_RS08620	2	NC_021287.1	1870186	1870995	*lpxH*	UDP-2,3-diacylglucosamine hydrolase	Function unknown
BRPE64_RS08635	2	NC_021287.1	1873063	1873866	*suhB*	inositol monophosphatase	Carbohydrate transport and metabolism
BRPE64_RS08685	2	NC_021287.1	1884934	1885839	*dapA*	dihydrodipicolinate synthase	Amino acid transport and metabolism
BRPE64_RS08695	2	NC_021287.1	1886532	1887734	*trpS*	tryptophan--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS08700	2	NC_021287.1	1887739	1888404	-	peptidase M50	General function prediction only
BRPE64_RS08745	2	NC_021287.1	1895087	1896370	*eno*	enolase	Carbohydrate transport and metabolism
BRPE64_RS08750	2	NC_021287.1	1896689	1897543	*kdsA*	2-dehydro-3-deoxyphosphooctonate aldolase	Cell wall/membrane/envelope biogenesis
BRPE64_RS08755	2	NC_021287.1	1897540	1899234	*pyrG*	CTP synthase	Nucleotide transport and metabolism
BRPE64_RS08775	2	NC_021287.1	1903655	1904368	*lolD*	lipoprotein releasing system ATP-binding protein	Defense mechanisms
BRPE64_RS08780	2	NC_021287.1	1904361	1905614	-	lipoprotein releasing system transmembrane protein LolC/E family	Cell wall/membrane/envelope biogenesis
BRPE64_RS08805	2	NC_021287.1	1910619	1910702	-	hypothetical protein	-
BRPE64_RS08810	2	NC_021287.1	1910821	1911723	-	hypothetical protein	-
BRPE64_RS08815	2	NC_021287.1	1911812	1913338	-	lysine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS08835	2	NC_021287.1	1915830	1916474	-	phosphatidylserine decarboxylase proenzyme	Lipid transport and metabolism
BRPE64_RS09020	2	NC_021287.1	1954275	1956629	*uvrD*	UvrD/REP helicase	Replication, recombination and repair
BRPE64_RS09025	2	NC_021287.1	1956804	1959671	*valS*	valine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS09030	2	NC_021287.1	1959748	1960629	-	UTP-glucose-1-phosphate uridylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS09035	2	NC_021287.1	1960728	1960955	-	SirA-like protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS09125	2	NC_021287.1	1978780	1981404	*alaS*	alanine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS09145	2	NC_021287.1	1983886	1985595	*glnS*	glutamine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS09290	2	NC_021287.1	2013318	2014679	*glmM*	phosphoglucosamine mutase	Carbohydrate transport and metabolism
BRPE64_RS09300	2	NC_021287.1	2015824	2017710	-	ATP-dependent zinc metalloprotease FtsH	Posttranslational modification,

							protein turnover, chaperones
BRPE64_RS09305	2	NC_021287.1	2017896	2018558	-	hypothetical protein	-
BRPE64_RS09325	2	NC_021287.1	2020560	2023814	*carB*	carbamoyl-phosphate synthase large chain	Amino acid transport and metabolism
BRPE64_RS09330	2	NC_021287.1	2023858	2025000	*carA*	carbamoyl-phosphate synthase small chain	Amino acid transport and metabolism
BRPE64_RS09350	2	NC_021287.1	2029403	2030236	-	methyltransferase type 11	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS09360	2	NC_021287.1	2030738	2031475	*dnaQ*	DNA polymerase III epsilon subunit	Replication, recombination and repair
BRPE64_RS09435	2	NC_021287.1	2042384	2042980	*pgsA*	CDP-diacylglycerol/glycerol-3-phosphate 3-phosphatidyltransferase	Lipid transport and metabolism
BRPE64_RS09445	2	NC_021287.1	2045380	2046540	-	hypothetical protein	Function unknown
BRPE64_RS09450	2	NC_021287.1	2046780	2047337	*efp*	elongation factor P	Translation, ribosomal structure and biogenesis
BRPE64_RS09470	2	NC_021287.1	2050684	2051088	*acpS*	holo-[acyl-carrier-protein] synthase	Lipid transport and metabolism
BRPE64_RS09475	2	NC_021287.1	2051100	2051873	*pdxJ*	pyridoxine 5'-phosphate synthase	Coenzyme transport and metabolism
BRPE64_RS09480	2	NC_021287.1	2051870	2052772	*recO*	DNA repair protein RecO	Replication, recombination and repair
BRPE64_RS09485	2	NC_021287.1	2052789	2053688	*era*	GTPase Era	General function prediction only
BRPE64_RS09490	2	NC_021287.1	2053861	2054868	*rnc*	ribonuclease 3	Transcription
BRPE64_RS09495	2	NC_021287.1	2055034	2055927	*lepB*	signal peptidase I	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS09500	2	NC_021287.1	2055994	2057784	*lepA*	elongation factor 4	Cell wall/membrane/envelope biogenesis
BRPE64_RS09535	2	NC_021287.1	2062755	2063993	*fabF*	3-oxoacyl-[acyl-carrier-protein] synthase 2	Lipid transport and metabolism
BRPE64_RS09540	2	NC_021287.1	2064151	2064390	*acpP*	acyl carrier protein	Lipid transport and metabolism
BRPE64_RS09550	2	NC_021287.1	2065381	2066313	*fabD*	malonyl CoA-acyl carrier protein transacylase	Lipid transport and metabolism
BRPE64_RS09555	2	NC_021287.1	2066445	2067434	*fabH*	3-oxoacyl-[acyl-carrier-protein] synthase 3	Lipid transport and metabolism
BRPE64_RS09560	2	NC_021287.1	2067434	2068603	*plsX*	phosphate acyltransferase	Lipid transport and metabolism
BRPE64_RS09565	2	NC_021287.1	2068742	2068921	*rpmF*	50S ribosomal protein L32	Translation, ribosomal structure and biogenesis
BRPE64_RS09570	2	NC_021287.1	2069139	2069759	-	hypothetical protein	General function prediction only
BRPE64_RS09605	2	NC_021287.1	2075337	2078696	*rne*	ribonuclease E	Translation, ribosomal structure and biogenesis

BRPE64_RS09645	2	NC_021287.1	2085347	2086759	-	putative arsenite-anitmonite efflux pump ArsB family	Inorganic ion transport and metabolism
BRPE64_RS09650	2	NC_021287.1	2086765	2087436	-	hypothetical protein	Function unknown
BRPE64_RS09660	2	NC_021287.1	2089045	2089368	-	putative ferredoxin	Energy production and conversion
BRPE64_RS09775	2	NC_021287.1	2111724	2112695	*thyA*	thymidylate synthase	Nucleotide transport and metabolism
BRPE64_RS09790	2	NC_021287.1	2115184	2115681	*folA*	dihydrofolate reductase	Coenzyme transport and metabolism
BRPE64_RS09810	2	NC_021287.1	2118520	2119074	*orn*	oligoribonuclease	RNA processing and modification
BRPE64_RS09815	2	NC_021287.1	2119272	2120540	-	putative peptidase M48 family	Posttranslational modification, protein turnover, chaperones
BRPE64_RS09820	2	NC_021287.1	2120537	2121472	*rsgA*	putative ribosome biogenesis GTPase RsgA	General function prediction only
BRPE64_RS09845	2	NC_021287.1	2124175	2124945	*trmD*	tRNA (guanine-N(1)-)-methyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS09850	2	NC_021287.1	2124973	2125719	*rimM*	ribosome maturation factor RimM	Translation, ribosomal structure and biogenesis
BRPE64_RS09855	2	NC_021287.1	2125790	2126044	*rpsP*	30S ribosomal protein S16	Translation, ribosomal structure and biogenesis
BRPE64_RS09885	2	NC_021287.1	2131910	2132845	-	electron transfer flavoprotein alpha subunit	Energy production and conversion
BRPE64_RS09890	2	NC_021287.1	2132861	2133610	-	electron transfer flavoprotein alpha/beta-subunit	Energy production and conversion
BRPE64_RS09925	2	NC_021287.1	2139873	2140775	*cysM*	cysteine synthase	Amino acid transport and metabolism
BRPE64_RS09945	2	NC_021287.1	2143458	2144861	-	nucleotide sugar dehydrogenase	Cell wall/membrane/envelope biogenesis
BRPE64_RS09950	2	NC_021287.1	2144944	2146119	-	tetratricopeptide TPR_2 repeat protein	Carbohydrate transport and metabolism
BRPE64_RS09955	2	NC_021287.1	2146162	2146455	-	hypothetical protein	Function unknown
BRPE64_RS09960	2	NC_021287.1	2146771	2147100	*ihfB*	integration host factor subunit beta	Replication, recombination and repair
BRPE64_RS09965	2	NC_021287.1	2147123	2148835	*rpsA*	30S ribosomal protein S1	Translation, ribosomal structure and biogenesis
BRPE64_RS09970	2	NC_021287.1	2149002	2149682	*cmk*	cytidylate kinase	Nucleotide transport and metabolism
BRPE64_RS09975	2	NC_021287.1	2149758	2151062	*aroA*	3-phosphoshikimate 1-carboxyvinyltransferase	Amino acid transport and metabolism
BRPE64_RS09980	2	NC_021287.1	2151073	2151999	-	prephenate dehydrogenase	Amino acid transport and metabolism
BRPE64_RS09985	2	NC_021287.1	2152076	2153158	*pheA*	chorismate mutase	Amino acid transport and metabolism
BRPE64_RS09990	2	NC_021287.1	2153198	2154280	*serC*	phosphoserine aminotransferase	Amino acid transport and metabolism

BRPE64_RS09995	2	NC_021287.1	2154478	2155086	-	hypothetical protein	Function unknown
BRPE64_RS10000	2	NC_021287.1	2155263	2157917	*gyrA*	DNA gyrase subunit A	Replication, recombination and repair
BRPE64_RS10005	2	NC_021287.1	2158616	2159278	-	OmpA/MotB domain protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS10010	2	NC_021287.1	2159473	2160171	*ubiG*	3-demethylubiquinone-9,3-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS10125	2	NC_021287.1	2187950	2190004	-	NAD synthetase/Glutamine amidotransferase chain of NAD synthetase	Coenzyme transport and metabolism
BRPE64_RS10260	2	NC_021287.1	2217771	2218031	-	GP29	Function unknown
BRPE64_RS10275	2	NC_021287.1	2223364	2223789	-	hypothetical protein	-
BRPE64_RS10290	2	NC_021287.1	2226073	2227422	-	putatove 3-deoxy-D-manno-octulosonic-acid transferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10295	2	NC_021287.1	2227455	2228339	-	hypothetical protein	-
BRPE64_RS10305	2	NC_021287.1	2229557	2230957	-	phosphomannomutase	Carbohydrate transport and metabolism
BRPE64_RS10555	2	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10585	2	NC_021287.1	2302528	2303079	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10625	2	NC_021287.1	2309981	2310427	-	putative Holliday junction resolvase	Replication, recombination and repair
BRPE64_RS10665	2	NC_021287.1	2317125	2318765	-	60 kDa chaperonin	Posttranslational modification, protein turnover, chaperones
BRPE64_RS10670	2	NC_021287.1	2318861	2319151	*groS*	10 kDa chaperonin	Posttranslational modification, protein turnover, chaperones
BRPE64_RS10775	2	NC_021287.1	2344707	2345288	*sodB*	superoxide dismutase	Inorganic ion transport and metabolism
BRPE64_RS10785	2	NC_021287.1	2347394	2348410	*lpxK*	tetraacyldisaccharide 4'-kinase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10790	2	NC_021287.1	2348391	2348597	-	UPF0434 protein BamMC406_2464	Function unknown
BRPE64_RS10795	2	NC_021287.1	2348616	2349407	*kdsB*	3-deoxy-manno-octulosonate cytidylyltransferase 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS10800	2	NC_021287.1	2349697	2350359	*adk*	adenylate kinase	Nucleotide transport and metabolism
BRPE64_RS10815	2	NC_021287.1	2352257	2353807	*mviN*	integral membrane protein MviN	General function prediction only
BRPE64_RS10820	2	NC_021287.1	2354197	2354475	*rpsT*	30S ribosomal protein S20	Translation, ribosomal structure and biogenesis
BRPE64_RS10825	2	NC_021287.1	2354678	2355013	-	hypothetical protein	-
BRPE64_RS10840	2	NC_021287.1	2357996	2359027	*murB*	UDP-N-acetylenolpyruvoylg	Cell wall/membrane/envelope biogenesis

						lucosamine reductase	
BRPE64_RS10850	2	NC_021287.1	2359737	2360363	-	glycerol-3-phosphate acyltransferase	Function unknown
BRPE64_RS10875	2	NC_021287.1	2363693	2364271	-	hypothetical protein	General function prediction only
BRPE64_RS10910	2	NC_021287.1	2370460	2371101	*pdxH*	pyridoxine/pyridoxamine 5'-phosphate oxidase	Coenzyme transport and metabolism
BRPE64_RS11010	2	NC_021287.1	2385657	2386904	-	serine hydroxymethyltransferase	Amino acid transport and metabolism
BRPE64_RS11090	2	NC_021287.1	2402511	2403872	*surA*	chaperone SurA	Posttranslational modification, protein turnover, chaperones
BRPE64_RS11095	2	NC_021287.1	2404009	2406399	-	LPS-assembly protein LptD	Cell wall/membrane/envelope biogenesis
BRPE64_RS11110	2	NC_021287.1	2408469	2409857	*purB*	adenylosuccinate lyase	Nucleotide transport and metabolism
BRPE64_RS11140	2	NC_021287.1	2415949	2416365	-	hypothetical protein	Function unknown
BRPE64_RS11145	2	NC_021287.1	2416424	2417695	*proA*	gamma-glutamyl phosphate reductase	Amino acid transport and metabolism
BRPE64_RS11150	2	NC_021287.1	2417802	2418881	*holA*	DNA polymerase III delta subunit	Replication, recombination and repair
BRPE64_RS11155	2	NC_021287.1	2418901	2419503	-	rare lipoprotein B	Cell wall/membrane/envelope biogenesis
BRPE64_RS11160	2	NC_021287.1	2419503	2422097	*leuS*	leucine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS11175	2	NC_021287.1	2423466	2424263	*dapB*	dihydrodipicolinate reductase	Amino acid transport and metabolism
BRPE64_RS11180	2	NC_021287.1	2424366	2425160	-	SmpA/OmlA domain protein	Translation, ribosomal structure and biogenesis
BRPE64_RS11185	2	NC_021287.1	2425336	2425764	*fur*	ferric uptake regulator Fur family	Inorganic ion transport and metabolism
BRPE64_RS11235	2	NC_021287.1	2435185	2435748	-	hypothetical protein	-
BRPE64_RS11270	2	NC_021287.1	2445477	2446475	*thiL*	thiamine-monophosphate kinase	Coenzyme transport and metabolism
BRPE64_RS11285	2	NC_021287.1	2447558	2448370	-	orotidine 5'-phosphate decarboxylase	Nucleotide transport and metabolism
BRPE64_RS11350	2	NC_021287.1	2461104	2463173	*rnb*	ribonuclease II	Transcription
BRPE64_RS11375	2	NC_021287.1	2467141	2467614	*accB*	biotin carboxyl carrier protein	Lipid transport and metabolism
BRPE64_RS11380	2	NC_021287.1	2467779	2469146	*accC*	acetyl-CoA carboxylase biotin carboxylase	Lipid transport and metabolism
BRPE64_RS11400	2	NC_021287.1	2472207	2473145	-	PfkB domain protein	Carbohydrate transport and metabolism
BRPE64_RS11415	2	NC_021287.1	2475066	2476313	*nrdB*	ribonucleoside-diphosphate reductase subunit beta	Nucleotide transport and metabolism

BRPE64_RS11435	2	NC_021287.1	2478941	2481976	*nrdA*	ribonucleoside-diphosphate reductase	Nucleotide transport and metabolism
BRPE64_RS11455	2	NC_021287.1	2484583	2485950	*ffh*	signal recognition particle protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11470	2	NC_021287.1	2487290	2489026	*proS*	proline--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS11490	2	NC_021287.1	2491939	2493033	-	GTPase obg	General function prediction only
BRPE64_RS11495	2	NC_021287.1	2493208	2493471	-	hypothetical protein	-
BRPE64_RS11500	2	NC_021287.1	2493507	2493818	*rplU*	50S ribosomal protein L21	Translation, ribosomal structure and biogenesis
BRPE64_RS11505	2	NC_021287.1	2494092	2495084	*ispB*	octylprenyl-diphosphate synthase	Coenzyme transport and metabolism
BRPE64_RS11535	2	NC_021287.1	2500104	2500709	*coaE*	dephospho-CoA kinase	Coenzyme transport and metabolism
BRPE64_RS11565	2	NC_021287.1	2504712	2507525	*secA*	protein translocase subunit SecA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11570	2	NC_021287.1	2507815	2508333	-	hypothetical protein	-
BRPE64_RS11575	2	NC_021287.1	2508384	2509301	*lpxC*	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11585	2	NC_021287.1	2510356	2511555	*ftsZ*	cell division protein FtsZ	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS11590	2	NC_021287.1	2511680	2512912	*ftsA*	cell division protein ftsA	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS11595	2	NC_021287.1	2512939	2513691	*ftsQ*	cell division protein FtsQ	Cell wall/membrane/envelope biogenesis
BRPE64_RS11600	2	NC_021287.1	2513755	2514696	-	D-alanine--D-alanine ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11605	2	NC_021287.1	2514728	2516119	*murC*	UDP-N-acetylmuramate--L-alanine ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11610	2	NC_021287.1	2516122	2517255	*murG*	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11615	2	NC_021287.1	2517252	2518547	*ftsW*	lipid II flippase FtsW	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS11620	2	NC_021287.1	2518544	2520055	*murD*	UDP-N-acetylmuramoylalanine--D-glutamate ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11625	2	NC_021287.1	2520117	2521286	*mraY*	phospho-N-acetylmuramoyl-pentapeptide-transferase	Cell wall/membrane/envelope biogenesis

BRPE64_RS11630	2	NC_021287.1	2521310	2522740	*murF*	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11635	2	NC_021287.1	2522737	2524281	*murE*	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2, 6-diaminopimelate ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11640	2	NC_021287.1	2524278	2526134	-	peptidoglycan glycosyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11645	2	NC_021287.1	2526131	2526475	-	cell division protein FtsL	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS11650	2	NC_021287.1	2526472	2527425	-	ribosomal RNA small subunit methyltransferase H	Cell wall/membrane/envelope biogenesis
BRPE64_RS11655	2	NC_021287.1	2527442	2527870	*mraZ*	protein MraZ	Function unknown
BRPE64_RS11800	2	NC_021287.1	2560202	2561374	-	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase	Energy production and conversion
BRPE64_RS11845	2	NC_021287.1	2570044	2571348	*hemA*	glutamyl-tRNA reductase	Coenzyme transport and metabolism
BRPE64_RS11850	2	NC_021287.1	2571445	2572527	*prfA*	peptide chain release factor 1	Translation, ribosomal structure and biogenesis
BRPE64_RS11855	2	NC_021287.1	2572524	2573360	*prmC*	release factor glutamine methyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS11860	2	NC_021287.1	2573447	2573758	-	glutaredoxin	Posttranslational modification, protein turnover, chaperones
BRPE64_RS11865	2	NC_021287.1	2573770	2574378	-	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Coenzyme transport and metabolism
BRPE64_RS11980	2	NC_021287.1	2598727	2599338	*sspA*	glutathione S-transferase domain protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12010	2	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12035	2	NC_021287.1	2607149	2607517	-	phosphoribosyl-ATP pyrophosphatase	Amino acid transport and metabolism
BRPE64_RS12040	2	NC_021287.1	2607514	2607918	*hisI*	phosphoribosyl-AMP cyclohydrolase	Amino acid transport and metabolism
BRPE64_RS12045	2	NC_021287.1	2607922	2608695	*hisF*	imidazole glycerol phosphate synthase subunit HisF	Amino acid transport and metabolism

BRPE64_RS12050	2	NC_021287.1	2608812	2609564	*hisA*	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	Amino acid transport and metabolism
BRPE64_RS12055	2	NC_021287.1	2609716	2610357	*hisH*	imidazole glycerol phosphate synthase subunit HisH	Amino acid transport and metabolism
BRPE64_RS12065	2	NC_021287.1	2611015	2611602	*hisB*	imidazoleglycerol-phosphate dehydratase	Amino acid transport and metabolism
BRPE64_RS12075	2	NC_021287.1	2612782	2614104	-	histidinol dehydrogenase	Amino acid transport and metabolism
BRPE64_RS12080	2	NC_021287.1	2614151	2614843	*hisG*	ATP phosphoribosyltransferase	Amino acid transport and metabolism
BRPE64_RS12085	2	NC_021287.1	2614840	2616111	*murA*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS12090	2	NC_021287.1	2616266	2616505	-	BolA family protein	Transcription
BRPE64_RS12095	2	NC_021287.1	2616523	2617278	-	ABC-2 type transporter	Defense mechanisms
BRPE64_RS12100	2	NC_021287.1	2617275	2618201	-	ABC multidrug efflux pump ATPase subunit	Defense mechanisms
BRPE64_RS12220	2	NC_021287.1	2643932	2645008	*aroB*	3-dehydroquinate synthase	Amino acid transport and metabolism
BRPE64_RS12280	2	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12285	2	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12295	2	NC_021287.1	2660987	2661649	-	probable GTP-binding protein EngB	General function prediction only
BRPE64_RS12300	2	NC_021287.1	2661916	2662914	*hemB*	delta-aminolevulinic acid dehydratase	Coenzyme transport and metabolism
BRPE64_RS12305	2	NC_021287.1	2662968	2664854	-	thiol disulfide interchange protein DsbD	Energy production and conversion
BRPE64_RS12315	2	NC_021287.1	2665293	2665688	*rplQ*	50S ribosomal protein L17	Translation, ribosomal structure and biogenesis
BRPE64_RS12320	2	NC_021287.1	2665847	2666824	*rpoA*	DNA-directed RNA polymerase subunit alpha	Transcription
BRPE64_RS12325	2	NC_021287.1	2666971	2667594	*rpsD*	30S ribosomal protein S4	Translation, ribosomal structure and biogenesis
BRPE64_RS12330	2	NC_021287.1	2667754	2668158	*rpsK*	30S ribosomal protein S11	Translation, ribosomal structure and biogenesis
BRPE64_RS12335	2	NC_021287.1	2668187	2668552	-	hypothetical protein	-
BRPE64_RS12345	2	NC_021287.1	2668731	2668949	-	hypothetical protein	-
BRPE64_RS12350	2	NC_021287.1	2668958	2670304	*secY*	protein translocase subunit SecY	Intracellular trafficking,

							secretion, and vesicular transport
BRPE64_RS12355	2	NC_021287.1	2670346	2670780	*rplO*	50S ribosomal protein L15	Translation, ribosomal structure and biogenesis
BRPE64_RS12360	2	NC_021287.1	2670811	2670993	*rpmD*	50S ribosomal protein L30	Translation, ribosomal structure and biogenesis
BRPE64_RS12365	2	NC_021287.1	2671008	2671526	*rpsE*	30S ribosomal protein S5	Translation, ribosomal structure and biogenesis
BRPE64_RS12370	2	NC_021287.1	2671541	2671906	*rplR*	50S ribosomal protein L18	Translation, ribosomal structure and biogenesis
BRPE64_RS12375	2	NC_021287.1	2671919	2672452	*rplF*	50S ribosomal protein L6	Translation, ribosomal structure and biogenesis
BRPE64_RS12380	2	NC_021287.1	2672471	2672866	*rpsH*	30S ribosomal protein S8	Translation, ribosomal structure and biogenesis
BRPE64_RS12385	2	NC_021287.1	2672881	2673186	*rpsN*	30S ribosomal protein S14	Translation, ribosomal structure and biogenesis
BRPE64_RS12390	2	NC_021287.1	2673194	2673733	*rplE*	50S ribosomal protein L5	Translation, ribosomal structure and biogenesis
BRPE64_RS12395	2	NC_021287.1	2673748	2674056	*rplX*	50S ribosomal protein L24	Translation, ribosomal structure and biogenesis
BRPE64_RS12400	2	NC_021287.1	2674066	2674434	*rplN*	50S ribosomal protein L14	Translation, ribosomal structure and biogenesis
BRPE64_RS12405	2	NC_021287.1	2674734	2675006	*rpsQ*	30S ribosomal protein S17	Translation, ribosomal structure and biogenesis
BRPE64_RS12415	2	NC_021287.1	2675208	2675624	*rplP*	50S ribosomal protein L16	Translation, ribosomal structure and biogenesis
BRPE64_RS12420	2	NC_021287.1	2675627	2676421	*rpsC*	30S ribosomal protein S3	Translation, ribosomal structure and biogenesis
BRPE64_RS12425	2	NC_021287.1	2676433	2676762	*rplV*	50S ribosomal protein L22	Translation, ribosomal structure and biogenesis
BRPE64_RS12430	2	NC_021287.1	2676775	2677050	*rpsS*	30S ribosomal protein S19	Translation, ribosomal structure and biogenesis
BRPE64_RS12435	2	NC_021287.1	2677061	2677888	*rplB*	50S ribosomal protein L2	Translation, ribosomal structure and biogenesis
BRPE64_RS12440	2	NC_021287.1	2677891	2678205	*rplW*	50S ribosomal protein L23	Translation, ribosomal structure and biogenesis
BRPE64_RS12445	2	NC_021287.1	2678202	2678822	*rplD*	50S ribosomal protein L4	Translation, ribosomal structure and biogenesis
BRPE64_RS12450	2	NC_021287.1	2678822	2679481	*rplC*	50S ribosomal protein L3	Translation, ribosomal structure and biogenesis
BRPE64_RS12460	2	NC_021287.1	2680128	2681318	-	elongation factor Tu	Translation, ribosomal structure and biogenesis
BRPE64_RS12465	2	NC_021287.1	2681383	2683485	-	elongation factor G	Translation, ribosomal structure and biogenesis

BRPE64_RS12470	2	NC_021287.1	2683615	2684085	*rpsG*	30S ribosomal protein S7	Translation, ribosomal structure and biogenesis
BRPE64_RS12475	2	NC_021287.1	2684287	2684667	-	hypothetical protein	-
BRPE64_RS12485	2	NC_021287.1	2686912	2691153	*rpoC*	DNA-directed RNA polymerase subunit beta'	Transcription
BRPE64_RS12490	2	NC_021287.1	2691175	2695281	*rpoB*	DNA-directed RNA polymerase subunit beta	Transcription
BRPE64_RS12495	2	NC_021287.1	2695650	2696024	*rplL*	50S ribosomal protein L7/L12	Translation, ribosomal structure and biogenesis
BRPE64_RS12500	2	NC_021287.1	2696108	2696605	*rplJ*	50S ribosomal protein L10	Translation, ribosomal structure and biogenesis
BRPE64_RS12505	2	NC_021287.1	2696909	2697607	*rplA*	50S ribosomal protein L1	Translation, ribosomal structure and biogenesis
BRPE64_RS12510	2	NC_021287.1	2697608	2698039	*rplK*	50S ribosomal protein L11	Translation, ribosomal structure and biogenesis
BRPE64_RS12515	2	NC_021287.1	2698186	2698743	*nusG*	transcription antitermination protein nusG	Transcription
BRPE64_RS12520	2	NC_021287.1	2698745	2699125	-	hypothetical protein	-
BRPE64_RS12530	2	NC_021287.1	2699304	2700494	-	elongation factor Tu	Translation, ribosomal structure and biogenesis
BRPE64_RS12595	2	NC_021287.1	2713477	2713761	-	phenylacetate-CoA oxygenase PaaH subunit	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12600	2	NC_021287.1	2713775	2714584	-	phenylacetate-CoA oxygenase PaaI subunit	Function unknown
BRPE64_RS12655	2	NC_021287.1	2728109	2728795	-	orotate phosphoribosyltransferase	Nucleotide transport and metabolism
BRPE64_RS12815	2	NC_021287.1	2754308	2756125	*glmS*	glutamine--fructose-6-phosphate aminotransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS12820	2	NC_021287.1	2756207	2757571	*glmU*	bifunctional protein GlmU	Cell wall/membrane/envelope biogenesis
BRPE64_RS12830	2	NC_021287.1	2758620	2759021	-	dihydroneopterin aldolase	Coenzyme transport and metabolism
BRPE64_RS12840	2	NC_021287.1	2759914	2761119	-	hypothetical protein	Function unknown
BRPE64_RS12855	2	NC_021287.1	2763107	2764351	*cca*	polynucleotide adenylyltransferase/ metal dependent phosphohydrolase	Translation, ribosomal structure and biogenesis
BRPE64_RS12860	2	NC_021287.1	2764348	2764968	-	glutathione S-transferase domain	Posttranslational modification, protein turnover, chaperones
BRPE64_RS13315	2	NC_021287.1	2864418	2866658	*priA*	primosomal protein N'	Replication, recombination and repair
BRPE64_RS13320	2	NC_021287.1	2866655	2866864	-	hypothetical protein	-
BRPE64_RS13325	2	NC_021287.1	2867118	2868203	*hemE*	uroporphyrinogen decarboxylase	Coenzyme transport and metabolism
BRPE64_RS13345	2	NC_021287.1	2872821	2873246	*atpC*	ATP synthase epsilon chain	Energy production and conversion

BRPE64_RS13350	2	NC_021287.1	2873326	2874720	*atpD*	ATP synthase subunit beta 2	Energy production and conversion
BRPE64_RS13355	2	NC_021287.1	2874765	2875649	*atpG*	ATP synthase gamma chain	Energy production and conversion
BRPE64_RS13360	2	NC_021287.1	2875722	2877263	*atpA*	ATP synthase subunit alpha 1	Energy production and conversion
BRPE64_RS13365	2	NC_021287.1	2877322	2877864	*atpH*	ATP synthase subunit delta	Energy production and conversion
BRPE64_RS13375	2	NC_021287.1	2878473	2878742	*atpE*	ATP synthase F0 C subunit	Energy production and conversion
BRPE64_RS13380	2	NC_021287.1	2878833	2879684	*atpB*	ATP synthase subunit a	Energy production and conversion
BRPE64_RS13385	2	NC_021287.1	2879860	2880393	-	ATP synthase I chain	Energy production and conversion
BRPE64_RS13395	2	NC_021287.1	2881765	2882667	-	ParB-like partition protein	Transcription
BRPE64_RS13400	2	NC_021287.1	2882713	2883492	-	cobyrinic acid ac-diamide synthase	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS13410	2	NC_021287.1	2884231	2886192	-	tRNA uridine 5-carboxymethylamin omethyl modification enzyme MnmG	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS13685	2	NC_021287.1	2945985	2946857	*dapF*	diaminopimelate epimerase	Amino acid transport and metabolism
BRPE64_RS13690	2	NC_021287.1	2946908	2947792	-	lipid A biosynthesis acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS13695	2	NC_021287.1	2948069	2949256	*metK*	S-adenosylmethionine synthase	Coenzyme transport and metabolism
BRPE64_RS13730	2	NC_021287.1	2954924	2955694	*fpr*	oxidoreductase FAD-binding domain protein	Energy production and conversion
BRPE64_RS13840	2	NC_021287.1	2978403	2978705	-	histone family protein nucleoid-structuring protein H-NS	General function prediction only
BRPE64_RS14010	2	NC_021287.1	3008347	3009735	-	tRNA modification GTPase MnmE	General function prediction only
BRPE64_RS14020	2	NC_021287.1	3010559	3012229	*yidC*	membrane protein insertase YidC	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS14025	2	NC_021287.1	3012238	3012555	-	putative membrane protein insertion efficiency factor	Function unknown
BRPE64_RS14030	2	NC_021287.1	3012626	3013129	*rnpA*	ribonuclease P protein component	Translation, ribosomal structure and biogenesis
BRPE64_RS14035	2	NC_021287.1	3013212	3013346	*rpmH*	50S ribosomal protein L34	Translation, ribosomal structure and biogenesis
BRPE64_RS14050	2	NC_021294.1	446	1108	-	chromosome partitioning protein	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS14055	2	NC_021294.1	1133	2194	-	ParB family protein	Transcription
BRPE64_RS14060	2	NC_021294.1	2257	3612	-	hypothetical protein	-
BRPE64_RS14180	2	NC_021294.1	29731	31152	-	hopanoid biosynthesis associated radical SAM protein HpnJ	Energy production and conversion

BRPE64_RS14185	2	NC_021294.1	31175	32389	-	putative glycosyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS14480	2	NC_021294.1	92383	93537	-	hopanoid biosynthesis associated radical SAM protein HpnH	General function prediction only
BRPE64_RS14500	2	NC_021294.1	96993	99245	-	isocitrate dehydrogenase NADP-dependent	Energy production and conversion
BRPE64_RS14735	2	NC_021294.1	144261	144677	-	hypothetical protein	-
BRPE64_RS14740	2	NC_021294.1	144693	145625	-	CbbX protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS14775	2	NC_021294.1	151601	152149	-	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	Energy production and conversion
BRPE64_RS14890	2	NC_021294.1	173576	174190	-	transcriptional regulator TetR family	Transcription
BRPE64_RS14895	2	NC_021294.1	174257	175489	-	major facilitator family (MFS) transporter	Amino acid transport and metabolism
BRPE64_RS14900	2	NC_021294.1	176063	176656	-	hypothetical protein	-
BRPE64_RS14905	2	NC_021294.1	176898	177641	-	hypothetical protein	-
BRPE64_RS14985	2	NC_021294.1	197919	198971	-	hypothetical protein	-
BRPE64_RS15085	2	NC_021294.1	217949	218308	-	hypothetical protein	-
BRPE64_RS15090	2	NC_021294.1	218493	220382	-	sensory transduction protein kinase	Signal transduction mechanisms
BRPE64_RS15470	2	NC_021294.1	293402	294448	-	putative patatin-like phospholipase	General function prediction only
BRPE64_RS15645	2	NC_021294.1	334904	335716	-	ABC transporter inner membrane subunit	Amino acid transport and metabolism
BRPE64_RS15655	2	NC_021294.1	338689	340572	-	hypothetical protein	-
BRPE64_RS15720	2	NC_021294.1	353657	354334	-	transcriptional regulator TetR family	Transcription
BRPE64_RS15765	2	NC_021294.1	364421	365872	-	2-methylcitrate dehydratase	General function prediction only
BRPE64_RS15770	2	NC_021294.1	365939	367108	-	citrate synthase	Energy production and conversion
BRPE64_RS15990	2	NC_021294.1	411237	412562	-	hypothetical protein	-
BRPE64_RS16030	2	NC_021294.1	421707	422009	-	hypothetical protein	-
BRPE64_RS16105	2	NC_021294.1	437221	437532	-	hypothetical protein	-
BRPE64_RS16110	2	NC_021294.1	437776	438336	-	histone family protein nucleoid-structuring protein H-NS	General function prediction only
BRPE64_RS16120	2	NC_021294.1	440131	440922	-	methyltransferase type 11	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS16125	2	NC_021294.1	440928	441707	-	cationic amino acid ABC transporter periplasmic binding protein	Amino acid transport and metabolism
BRPE64_RS16130	2	NC_021294.1	441906	442310	-	heat shock protein Hsp20	Posttranslational modification,

							protein turnover, chaperones
BRPE64_RS16135	2	NC_021294.1	442323	442757	-	heat shock protein Hsp20	Posttranslational modification, protein turnover, chaperones
BRPE64_RS16140	2	NC_021294.1	442756	442953	-	hypothetical protein	-
BRPE64_RS16145	2	NC_021294.1	442982	443362	-	hypothetical protein	Function unknown
BRPE64_RS16275	2	NC_021294.1	467100	468488	-	FAD/FMN-containing dehydrogenase	Energy production and conversion
BRPE64_RS16280	2	NC_021294.1	468546	468902	-	hypothetical protein	Cell motility
BRPE64_RS16285	2	NC_021294.1	469026	469283	-	hypothetical protein	-
BRPE64_RS16380	2	NC_021294.1	495864	496415	*fimA*	major type 1 subunit fimbrin	Cell motility
BRPE64_RS16385	2	NC_021294.1	496521	497264	-	fimbrial assembly chaperone	Cell motility
BRPE64_RS16530	2	NC_021294.1	528629	528976	*nirD*	nitrite reductase (NAD(P)H) small subunit	Inorganic ion transport and metabolism
BRPE64_RS16535	2	NC_021294.1	528961	530244	-	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Energy production and conversion
BRPE64_RS16540	2	NC_021294.1	530270	533005	-	molybdopterin oxidoreductase	Energy production and conversion
BRPE64_RS16545	2	NC_021294.1	533634	533981	-	hypothetical protein	-
BRPE64_RS16550	2	NC_021294.1	534107	534454	-	hypothetical protein	-
BRPE64_RS16570	2	NC_021294.1	537962	538675	-	Two component transcriptional regulator winged helix family	Transcription
BRPE64_RS16575	2	NC_021294.1	538772	540229	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS16580	2	NC_021294.1	540226	541461	-	efflux transporter RND family MFP subunit	Cell wall/membrane/envelope biogenesis
BRPE64_RS16585	2	NC_021294.1	541483	544668	-	cation/multidrug efflux pump AcrB/AcrD/AcrF family RND superfamily	Defense mechanisms
BRPE64_RS16590	2	NC_021294.1	544911	545177	-	hypothetical protein	-
BRPE64_RS16595	2	NC_021294.1	545404	545670	-	hypothetical protein	-
BRPE64_RS16600	2	NC_021294.1	545963	547294	-	divalent metal cation transporter MntH	Inorganic ion transport and metabolism
BRPE64_RS16785	2	NC_021294.1	583614	584627	-	rhamnose ABC transporter periplasmic rhamnose-binding protein	Carbohydrate transport and metabolism
BRPE64_RS16920	2	NC_021294.1	611015	611371	-	hypothetical protein	-
BRPE64_RS16970	2	NC_021294.1	622361	622846	-	hypothetical protein	-
BRPE64_RS16975	2	NC_021294.1	623039	623920	-	alpha/beta hydrolase fold protein	General function prediction only
BRPE64_RS16980	2	NC_021294.1	624333	625457	-	hypothetical protein	-
BRPE64_RS16985	2	NC_021294.1	625460	626296	-	hypothetical protein	-

BRPE64_RS17060	2	NC_021294.1	641737	641934	-	hypothetical protein	-
BRPE64_RS17130	2	NC_021294.1	660515	660706	-	hypothetical protein	-
BRPE64_RS17460	2	NC_021294.1	730334	730816	-	hypothetical protein	-
BRPE64_RS17465	2	NC_021294.1	730833	731351	-	hypothetical protein	-
BRPE64_RS17470	2	NC_021294.1	731824	732714	-	LysR family transcriptional regulator	Transcription
BRPE64_RS17475	2	NC_021294.1	732819	733211	-	acyloate catabolism-like protein	General function prediction only
BRPE64_RS17480	2	NC_021294.1	733281	734381	-	mandelate racemase/muconate lactonizing enzyme	Cell wall/membrane/envelope biogenesis
BRPE64_RS17485	2	NC_021294.1	734462	735220	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS17490	2	NC_021294.1	735306	736667	-	Bll0889 protein	Amino acid transport and metabolism
BRPE64_RS17495	2	NC_021294.1	736720	737091	-	hypothetical protein	General function prediction only
BRPE64_RS17500	2	NC_021294.1	737430	737861	-	transcriptional regulator MarR family protein	Transcription
BRPE64_RS17505	2	NC_021294.1	737881	738318	-	hypothetical protein	-
BRPE64_RS17610	2	NC_021294.1	755964	756485	-	hypothetical protein	-
BRPE64_RS17615	2	NC_021294.1	756526	757170	-	methylamine dehydrogenase accessory protein MauD	Energy production and conversion
BRPE64_RS17620	2	NC_021294.1	757167	757727	-	hypothetical protein	-
BRPE64_RS17625	2	NC_021294.1	757737	758894	-	aralkylamine dehydrogenase	Function unknown
BRPE64_RS17630	2	NC_021294.1	759132	760073	-	transcriptional regulator AraC family	Transcription
BRPE64_RS17635	2	NC_021294.1	760363	760902	-	4-hydroxyphenylacetate 3-monooxygenase reductase subunit	General function prediction only
BRPE64_RS17640	2	NC_021294.1	761136	761543	-	hypothetical protein	-
BRPE64_RS17645	2	NC_021294.1	761592	762116	-	putative uricase	Function unknown
BRPE64_RS17650	2	NC_021294.1	762258	763205	-	regulatory protein LysR	Transcription
BRPE64_RS17665	2	NC_021294.1	765424	765762	-	transthyretin	General function prediction only
BRPE64_RS17670	2	NC_021294.1	766010	766519	-	histone family protein nucleoid-structuring protein H-NS	General function prediction only
BRPE64_RS17675	2	NC_021294.1	766525	766929	-	hypothetical protein	-
BRPE64_RS17680	2	NC_021294.1	767009	767434	-	peroxiredoxin Ohr subfamily	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17690	2	NC_021294.1	770016	770411	-	hypothetical protein	-
BRPE64_RS17715	2	NC_021294.1	776231	776944	-	peptide methionine sulfoxide reductase MsrA	Posttranslational modification, protein turnover, chaperones

BRPE64_RS17720	2	NC_021294.1	776908	778035	-	aminotransferase class I and II	Amino acid transport and metabolism
BRPE64_RS17725	2	NC_021294.1	778132	779025	-	transcriptional regulator LysR family	Transcription
BRPE64_RS17735	2	NC_021294.1	779535	779876	-	hypothetical protein	-
BRPE64_RS17740	2	NC_021294.1	779967	780860	-	2-dehydropantoate 2-reductase	Coenzyme transport and metabolism
BRPE64_RS17745	2	NC_021294.1	781069	781356	-	hypothetical protein	-
BRPE64_RS17750	2	NC_021294.1	781353	782063	-	GntR domain protein	Transcription
BRPE64_RS17775	2	NC_021294.1	785815	787647	-	predicted carbamoyl transferase NodU family	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17780	2	NC_021294.1	787653	788654	-	hypothetical protein	Amino acid transport and metabolism
BRPE64_RS17785	2	NC_021294.1	788708	789838	-	hypothetical protein	Nucleotide transport and metabolism
BRPE64_RS17885	2	NC_021294.1	809675	812734	-	cyclic nucleotide-regulated ABC bacteriocin/lantibiotic exporter	Defense mechanisms
BRPE64_RS17890	2	NC_021294.1	812752	813504	-	PpiC-type peptidyl-prolyl cis-trans isomerase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17895	2	NC_021294.1	813577	814950	-	ABC efflux pump membrane fusion protein HlyD subfamily	Cell wall/membrane/envelope biogenesis
BRPE64_RS17900	2	NC_021294.1	814928	817453	-	putative forkhead-associated protein	Signal transduction mechanisms
BRPE64_RS18245	2	NC_021294.1	888588	889997	-	cytochrome bd ubiquinol oxidase subunit I	Energy production and conversion
BRPE64_RS18250	2	NC_021294.1	890002	891003	-	cytochrome d ubiquinol oxidase subunit II	Energy production and conversion
BRPE64_RS18515	2	NC_021294.1	947411	947611	-	hypothetical protein	-
BRPE64_RS18535	2	NC_021294.1	951329	952354	-	ectoine utilization protein EutE	General function prediction only
BRPE64_RS18540	2	NC_021294.1	952359	953576	-	ectoine utilization protein EutD	Amino acid transport and metabolism
BRPE64_RS18545	2	NC_021294.1	953588	954598	-	ectoine utilization protein EutC	Amino acid transport and metabolism
BRPE64_RS18550	2	NC_021294.1	954603	955568	-	ectoine utilization protein EutB	Amino acid transport and metabolism
BRPE64_RS18555	2	NC_021294.1	955649	957046	-	transcriptional regulator GntR family with aminotransferase domain	Amino acid transport and metabolism
BRPE64_RS18560	2	NC_021294.1	957281	958135	-	ectoine/hydroxyectoine ABC transporter solute-binding protein	Amino acid transport and metabolism
BRPE64_RS18565	2	NC_021294.1	958216	958869	-	beta tubulin autoregulation binding site	Amino acid transport and metabolism

BRPE64_RS18575	2	NC_021294.1	959587	960432	-	ectoine/hydroxyectoine ABC transporter ATP-binding protein	Amino acid transport and metabolism
BRPE64_RS18580	2	NC_021294.1	960508	962070	-	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	General function prediction only
BRPE64_RS18585	2	NC_021294.1	962089	962478	-	hypothetical protein	Carbohydrate transport and metabolism
BRPE64_RS18625	2	NC_021294.1	972311	972841	-	transcriptional regulator AsnC family	Transcription
BRPE64_RS18675	2	NC_021294.1	980714	982222	-	putative ABC transporter solute-binding protein	Carbohydrate transport and metabolism
BRPE64_RS18680	2	NC_021294.1	982276	983013	-	transcriptional regulator GntR family	Transcription
BRPE64_RS18685	2	NC_021294.1	983144	984442	-	mandelate racemase/muconate lactonizing protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS18690	2	NC_021294.1	984430	985344	-	hypothetical protein	-
BRPE64_RS18695	2	NC_021294.1	985399	986427	-	HtrA2 peptidase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS18755	2	NC_021294.1	995570	997108	*hsdM*	N-6 DNA methylase	Defense mechanisms
BRPE64_RS18760	2	NC_021294.1	997513	998286	-	hypothetical protein	-
BRPE64_RS18785	2	NC_021294.1	1005617	1006285	*ribA*	GTP cyclohydrolase-2	Coenzyme transport and metabolism
BRPE64_RS19020	2	NC_021294.1	1064447	1064641	-	hypothetical protein	-
BRPE64_RS19170	2	NC_021294.1	1094882	1095988	-	S-(Hydroxymethyl)glutathione dehydrogenase/class III alcohol dehydrogenase	Energy production and conversion
BRPE64_RS19550	2	NC_021294.1	1183640	1185085	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS19645	2	NC_021294.1	1205963	1206874	-	peptidase M23	Cell wall/membrane/envelope biogenesis
BRPE64_RS19650	2	NC_021294.1	1207224	1208438	-	hypothetical protein	-
BRPE64_RS19655	2	NC_021294.1	1208905	1209834	-	hypothetical protein	-
BRPE64_RS19665	2	NC_021294.1	1213937	1214449	-	hypothetical protein	-
BRPE64_RS19675	2	NC_021294.1	1215746	1217248	-	type VI secretion protein VC_A0110 family	Function unknown
BRPE64_RS19705	2	NC_021294.1	1222154	1224037	*dnaG*	DNA primase	Replication, recombination and repair
BRPE64_RS19710	2	NC_021294.1	1224058	1224504	-	hypothetical protein	Function unknown
BRPE64_RS19720	2	NC_021294.1	1226074	1227102	-	probable tRNA threonylcarbamoyladenosine biosynthesis protein Gcp	Posttranslational modification, protein turnover, chaperones
BRPE64_RS19740	2	NC_021294.1	1230044	1230847	-	GTP cyclohydrolase folE2	Function unknown

BRPE64_RS19745	2	NC_021294.1	1230986	1232902	*dxs*	1-deoxy-D-xylulose-5-phosphate synthase	Coenzyme transport and metabolism
BRPE64_RS19750	2	NC_021294.1	1232983	1233870	*ispA*	farnesyl-diphosphate synthase	Coenzyme transport and metabolism
BRPE64_RS19755	2	NC_021294.1	1233867	1234202	*xseB*	exodeoxyribonuclease 7 small subunit	Replication, recombination and repair
BRPE64_RS19780	2	NC_021294.1	1239240	1241978	*polA*	DNA polymerase I	Replication, recombination and repair
BRPE64_RS19970	2	NC_021294.1	1293715	1293903	-	hypothetical protein	-
BRPE64_RS19975	2	NC_021294.1	1294451	1295605	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS19980	2	NC_021294.1	1295616	1296698	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS20035	2	NC_021294.1	1309082	1310392	*folC*	FolC bifunctional protein	Coenzyme transport and metabolism
BRPE64_RS20040	2	NC_021294.1	1310481	1311353	*accD*	acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	Lipid transport and metabolism
BRPE64_RS20075	2	NC_021294.1	1318091	1319212	*asd*	aspartate-semialdehyde dehydrogenase	Amino acid transport and metabolism
BRPE64_RS20105	2	NC_021294.1	1323878	1325179	*gltA*	citrate synthase	Energy production and conversion
BRPE64_RS20110	2	NC_021294.1	1325346	1325630	-	hypothetical protein	Function unknown
BRPE64_RS20115	2	NC_021294.1	1325644	1326345	-	succinate dehydrogenase iron-sulfur protein	Energy production and conversion
BRPE64_RS20120	2	NC_021294.1	1326371	1328146	-	succinate dehydrogenase flavoprotein subunit	Energy production and conversion
BRPE64_RS20125	2	NC_021294.1	1328152	1328520	-	succinate dehydrogenase subunit D	Energy production and conversion
BRPE64_RS20130	2	NC_021294.1	1328536	1328946	-	succinate dehydrogenase cytochrome b556 subunit	Energy production and conversion
BRPE64_RS20145	2	NC_021294.1	1330507	1331493	*mdh*	malate dehydrogenase	Energy production and conversion
BRPE64_RS20150	2	NC_021294.1	1331681	1332691	-	HpcH/HpaI aldolase	Carbohydrate transport and metabolism
BRPE64_RS20155	2	NC_021294.1	1332800	1333972	-	citrate synthase	Energy production and conversion
BRPE64_RS20160	2	NC_021294.1	1334060	1334569	-	hypothetical protein	-
BRPE64_RS20165	2	NC_021294.1	1334602	1336053	-	2-methylcitrate dehydratase	General function prediction only
BRPE64_RS20170	2	NC_021294.1	1336090	1338807	*acnA*	aconitate hydratase 1	Energy production and conversion
BRPE64_RS20740	2	NC_021288.1	1573	2979	-	hypothetical protein	-
BRPE64_RS20745	2	NC_021288.1	3626	4837	-	cobyrinic acid ac-diamide synthase	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS20750	2	NC_021288.1	4834	5811	-	ParB-like partition protein	Transcription
BRPE64_RS20860	2	NC_021288.1	28654	28899	-	hypothetical protein	-

BRPE64_RS20895	2	NC_021288.1	35576	37138	*glpD*	glycerol-3-phosphate dehydrogenase 2	Energy production and conversion
BRPE64_RS31955	2	NC_021288.1	60447	60899	-	hypothetical protein	Energy production and conversion
BRPE64_RS21130	2	NC_021288.1	94175	94687	-	type VI secretion protein VC_A0107 family	Function unknown
BRPE64_RS21505	2	NC_021288.1	164934	165302	-	succinate dehydrogenase subunit D	Energy production and conversion
BRPE64_RS21510	2	NC_021288.1	165308	167083	-	succinate dehydrogenase flavoprotein subunit	Energy production and conversion
BRPE64_RS21515	2	NC_021288.1	167109	167810	-	succinate dehydrogenase iron-sulfur protein	Energy production and conversion
BRPE64_RS21780	2	NC_021288.1	220297	221226	-	transcriptional regulator LysR family	Transcription
BRPE64_RS21785	2	NC_021288.1	221383	221829	-	activator of Hsp90 ATPase 1 family protein	Function unknown
BRPE64_RS21790	2	NC_021288.1	221892	223001	-	hypothetical protein	-
BRPE64_RS21795	2	NC_021288.1	223353	224108	-	NmrA family protein	Carbohydrate transport and metabolism
BRPE64_RS21800	2	NC_021288.1	224123	225031	-	hypothetical protein	-
BRPE64_RS21805	2	NC_021288.1	225028	225576	-	putative MxaK-like protein	General function prediction only
BRPE64_RS21810	2	NC_021288.1	225573	226553	-	putative MxaC-like protein	General function prediction only
BRPE64_RS21815	2	NC_021288.1	226558	227232	-	hypothetical protein	-
BRPE64_RS21820	2	NC_021288.1	227460	228323	-	putative MxaS-like protein	General function prediction only
BRPE64_RS21875	2	NC_021288.1	238338	238988	-	response regulator receiver and ANTAR domain protein	Signal transduction mechanisms
BRPE64_RS21880	2	NC_021288.1	239019	239783	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS21885	2	NC_021288.1	239944	240759	-	transcriptional regulator DeoR family	Carbohydrate transport and metabolism
BRPE64_RS21890	2	NC_021288.1	240831	241796	-	PfkB domain protein	Carbohydrate transport and metabolism
BRPE64_RS21895	2	NC_021288.1	241789	243072	*kbaZ*	putative tagatose 6-phosphate kinase	Carbohydrate transport and metabolism
BRPE64_RS21900	2	NC_021288.1	243379	244284	-	probable sugar ABC transporter permease protein	Carbohydrate transport and metabolism
BRPE64_RS21905	2	NC_021288.1	244359	245306	-	putative sugar (D-ribose) ABC transporter	Carbohydrate transport and metabolism
BRPE64_RS21910	2	NC_021288.1	245354	246895	-	ABC transporter related	Carbohydrate transport and metabolism
BRPE64_RS21915	2	NC_021288.1	246960	247832	-	xylose isomerase domain-containing protein TIM barrel	Carbohydrate transport and metabolism
BRPE64_RS21920	2	NC_021288.1	247829	248458	-	NUDIX hydrolase	Nucleotide transport and metabolism
BRPE64_RS21925	2	NC_021288.1	248477	249229	-	hypothetical protein	-

BRPE64_RS21930	2	NC_021288.1	249275	249460	-	hypothetical protein	-
BRPE64_RS21940	2	NC_021288.1	251347	252135	-	hypothetical protein	Function unknown
BRPE64_RS21945	2	NC_021288.1	252353	253501	-	outer membrane porin OmpC family	Cell wall/membrane/envelope biogenesis
BRPE64_RS21950	2	NC_021288.1	253610	254251	-	putative transcriptional regulator	Transcription
BRPE64_RS21955	2	NC_021288.1	254495	255832	-	acyl-CoA dehydrogenase domain protein	Lipid transport and metabolism
BRPE64_RS21960	2	NC_021288.1	255899	257239	-	acyl-CoA dehydrogenase domain protein	Lipid transport and metabolism
BRPE64_RS21965	2	NC_021288.1	257262	258269	-	hypothetical protein	-
BRPE64_RS21970	2	NC_021288.1	258397	259158	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS21975	2	NC_021288.1	259148	260032	-	NmrA-like protein	Carbohydrate transport and metabolism
BRPE64_RS21980	2	NC_021288.1	260213	261085	-	AraC family transcriptional regulator	Transcription
BRPE64_RS21985	2	NC_021288.1	261237	261986	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS21990	2	NC_021288.1	262078	262965	-	transcriptional regulator LysR family	Transcription
BRPE64_RS21995	2	NC_021288.1	263159	263488	-	hypothetical cytosolic protein	Function unknown
BRPE64_RS22000	2	NC_021288.1	263726	263986	-	hypothetical protein	-
BRPE64_RS22005	2	NC_021288.1	263999	265078	-	hypothetical protein	Energy production and conversion
BRPE64_RS22010	2	NC_021288.1	265341	265781	-	putative HTH-type transcriptional regulator ywnA	Transcription
BRPE64_RS22015	2	NC_021288.1	265907	266554	-	HAD-superfamily hydrolase subfamily IA variant 3	General function prediction only
BRPE64_RS22020	2	NC_021288.1	266593	267474	-	transcriptional regulator LysR family	Transcription
BRPE64_RS22025	2	NC_021288.1	267515	268351	*ylbA*	hypothetical protein	General function prediction only
BRPE64_RS22030	2	NC_021288.1	268747	269589	-	amidohydrolase 2	General function prediction only
BRPE64_RS22035	2	NC_021288.1	269765	270712	-	Blr7068 protein	General function prediction only
BRPE64_RS22040	2	NC_021288.1	270768	271286	-	hypothetical protein	-
BRPE64_RS22155	2	NC_021288.1	291742	292509	-	NAD-dependent epimerase/dehydratase	Lipid transport and metabolism
BRPE64_RS22160	2	NC_021288.1	292855	293760	-	N-acetylneuraminate lyase	Amino acid transport and metabolism
BRPE64_RS22165	2	NC_021288.1	293807	294868	-	putative Glu/Leu/Phe/Val dehydrogenase	Amino acid transport and metabolism
BRPE64_RS22170	2	NC_021288.1	294922	295749	-	transcriptional regulator AraC family	Transcription
BRPE64_RS22175	2	NC_021288.1	295941	297044	-	putative transcriptional regulator Fis family	Transcription

BRPE64_RS22180	2	NC_021288.1	297041	297610	-	TetR family transcriptional regulator	Transcription
BRPE64_RS22185	2	NC_021288.1	297811	298542	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS22190	2	NC_021288.1	298625	299338	-	NAD(P)H dehydrogenase quinone family	General function prediction only
BRPE64_RS22225	2	NC_021288.1	305635	306225	-	2-hydroxychromene-2-carboxylate isomerase-like protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS22230	2	NC_021288.1	306249	306959	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS22235	2	NC_021288.1	307137	307868	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS22240	2	NC_021288.1	307881	308552	-	Gst13 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS22245	2	NC_021288.1	308612	309808	-	L-carnitine dehydratase/bile acid-inducible protein F	Energy production and conversion
BRPE64_RS22250	2	NC_021288.1	309877	310803	-	2-dehydropantoate 2-reductase	Coenzyme transport and metabolism
BRPE64_RS22255	2	NC_021288.1	310825	311454	-	glutathione S-transferase domain	Posttranslational modification, protein turnover, chaperones
BRPE64_RS22260	2	NC_021288.1	311529	312935	*galP*	galactose-proton symport	Amino acid transport and metabolism
BRPE64_RS22370	2	NC_021288.1	338117	339652	-	drug resistance transporter EmrB/QacA subfamily	Amino acid transport and metabolism
BRPE64_RS22375	2	NC_021288.1	339649	340581	*argC*	N-acetyl-gamma-glutamyl-phosphate reductase	Amino acid transport and metabolism
BRPE64_RS22445	2	NC_021288.1	353457	354428	-	transcriptional regulator AraC family protein	Transcription
BRPE64_RS22600	2	NC_021288.1	386776	387939	-	PrpF protein involved in 2-methylcitrate cycle	Function unknown
BRPE64_RS22605	2	NC_021288.1	388009	388890	-	hypothetical protein	-
BRPE64_RS22610	2	NC_021288.1	389168	389557	-	hypothetical protein	-
BRPE64_RS22615	2	NC_021288.1	389914	390792	-	hypothetical protein	Function unknown
BRPE64_RS22620	2	NC_021288.1	390846	391340	-	hypothetical protein	Function unknown
BRPE64_RS22625	2	NC_021288.1	391353	392069	-	hypothetical protein	Amino acid transport and metabolism
BRPE64_RS22630	2	NC_021288.1	392983	393543	-	hypothetical protein	-
BRPE64_RS22635	2	NC_021288.1	393706	393888	-	hypothetical protein	-
BRPE64_RS22640	2	NC_021288.1	394400	395260	-	hypothetical protein	-
BRPE64_RS22645	2	NC_021288.1	395257	396297	-	hypothetical protein	-
BRPE64_RS22650	2	NC_021288.1	396290	397363	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis

BRPE64_RS22655	2	NC_021288.1	397620	397844	-	hypothetical protein	-
BRPE64_RS22660	2	NC_021288.1	398148	398390	-	hypothetical protein	-
BRPE64_RS22665	2	NC_021288.1	398558	398986	-	Bll4598 protein	Function unknown
BRPE64_RS22760	2	NC_021288.1	434009	434311	-	hypothetical protein	-
BRPE64_RS23020	2	NC_021288.1	501285	502913	-	AMP-dependent synthetase and ligase	Lipid transport and metabolism
BRPE64_RS23065	2	NC_021288.1	511330	511917	-	hypothetical protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS23070	2	NC_021288.1	512029	512700	-	hypothetical protein	-
BRPE64_RS23075	2	NC_021288.1	512714	514834	-	glycogen debranching enzyme GlgX	Carbohydrate transport and metabolism
BRPE64_RS23080	2	NC_021288.1	514803	517304	-	phosphorylase	Carbohydrate transport and metabolism
BRPE64_RS23085	2	NC_021288.1	517492	518175	-	putative signal-transduction protein with CBS domains	General function prediction only
BRPE64_RS23090	2	NC_021288.1	518414	520756	-	small conductance mechanosensitive channel ion channel	Cell wall/membrane/envelope biogenesis
BRPE64_RS23095	2	NC_021288.1	521707	522336	-	hypothetical protein	General function prediction only
BRPE64_RS23100	2	NC_021288.1	522364	524328	-	hypothetical protein	-
BRPE64_RS23105	2	NC_021288.1	524329	524766	-	hypothetical protein	Energy production and conversion
BRPE64_RS23110	2	NC_021288.1	524905	525477	-	PEBP family protein	General function prediction only
BRPE64_RS23115	2	NC_021288.1	526327	526518	-	hypothetical protein	-
BRPE64_RS23330	2	NC_021288.1	574929	576413	-	aldehyde Dehydrogenase	Energy production and conversion
BRPE64_RS23335	2	NC_021288.1	576423	577385	-	transcriptional regulator LysR family	Transcription
BRPE64_RS23430	2	NC_021288.1	595527	598463	-	type III restriction protein res subunit	Transcription
BRPE64_RS23525	2	NC_021288.1	637314	638576	-	efflux transporter RND family MFP subunit	Cell wall/membrane/envelope biogenesis
BRPE64_RS23530	2	NC_021288.1	639363	639701	-	transport-associated protein	General function prediction only
BRPE64_RS31990	2	NC_021288.1	727705	729990	-	hypothetical protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS23940	2	NC_021288.1	730076	730309	-	hypothetical protein	-
BRPE64_RS23945	2	NC_021288.1	730330	731085	-	hypothetical protein	-
BRPE64_RS23950	2	NC_021288.1	731578	731886	-	hypothetical protein	-
BRPE64_RS24235	2	NC_021288.1	795574	796626	-	hypothetical protein	-
BRPE64_RS24335	2	NC_021288.1	817885	818145	-	GP29	Function unknown
BRPE64_RS24340	2	NC_021288.1	818522	818926	-	hypothetical protein	-
BRPE64_RS24345	2	NC_021288.1	818932	820779	-	hypothetical protein	-
BRPE64_RS24350	2	NC_021288.1	820769	821656	-	hypothetical protein	-
BRPE64_RS24430	2	NC_021288.1	836342	837277	-	transcriptional regulator LysR family	Transcription

BRPE64_RS24435	2	NC_021288.1	837303	838526	-	general substrate transporter	Amino acid transport and metabolism
BRPE64_RS24440	2	NC_021288.1	838618	839379	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS24565	2	NC_021288.1	866842	867864	-	hypothetical protein	Function unknown
BRPE64_RS24690	2	NC_021289.1	104	1282	-	Soj protein	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS24695	2	NC_021289.1	1397	2254	-	stage 0 sporulation protein J	Transcription
BRPE64_RS24700	2	NC_021289.1	2559	3896	-	hypothetical protein	-
BRPE64_RS24705	2	NC_021289.1	4220	6217	-	RNA polymerase sigma factor	Transcription
BRPE64_RS25135	2	NC_021289.1	90664	92046	-	histidine kinase	Signal transduction mechanisms
BRPE64_RS32010	2	NC_021289.1	92426	93082	-	hypothetical protein	-
BRPE64_RS25145	2	NC_021289.1	95272	97056	*ptsG*	PTS system glucose-specific IIBC subunit	Carbohydrate transport and metabolism
BRPE64_RS25150	2	NC_021289.1	97070	99601	-	phosphoenolpyruvate-protein phosphotransferase	Carbohydrate transport and metabolism
BRPE64_RS25210	2	NC_021289.1	111050	111853	-	transglutaminase-like domain protein	Amino acid transport and metabolism
BRPE64_RS25215	2	NC_021289.1	111908	112780	-	transglutaminase-like domain protein	Amino acid transport and metabolism
BRPE64_RS25220	2	NC_021289.1	113713	114156	-	alkylhydroperoxidase like protein AhpD family	Function unknown
BRPE64_RS25225	2	NC_021289.1	114258	115679	-	transcriptional regulator GntR family with aminotransferase domain	Amino acid transport and metabolism
BRPE64_RS25230	2	NC_021289.1	116049	117164	-	porin Gram-negative type	Cell wall/membrane/envelope biogenesis
BRPE64_RS25240	2	NC_021289.1	117838	118371	-	formaldehyde-activating enzyme	Function unknown
BRPE64_RS25450	2	NC_021289.1	162778	163779	-	4-hydroxythreonine-4-phosphate dehydrogenase	Coenzyme transport and metabolism
BRPE64_RS25505	2	NC_021289.1	176007	177086	-	methylthioribose-1-phosphate isomerase	Translation, ribosomal structure and biogenesis
BRPE64_RS25510	2	NC_021289.1	177083	177754	-	class II aldolase/adducin family protein	Carbohydrate transport and metabolism
BRPE64_RS25515	2	NC_021289.1	177836	178948	-	monosaccharide ABC transporter substrate-binding protein CUT2 family	Carbohydrate transport and metabolism
BRPE64_RS25520	2	NC_021289.1	179007	180542	-	ribose import ATP-binding protein RbsA 1	Carbohydrate transport and metabolism
BRPE64_RS25525	2	NC_021289.1	180539	181540	-	inner-membrane translocator	Carbohydrate transport and metabolism
BRPE64_RS25530	2	NC_021289.1	181601	182611	-	dihydroxyacetone kinase DhaK subunit	Carbohydrate transport and metabolism

BRPE64_RS25535	2	NC_021289.1	182624	183250	*dhaL*	dihydroxyacetone kinase DhaL subunit	Carbohydrate transport and metabolism
BRPE64_RS25540	2	NC_021289.1	183642	183911	-	hypothetical protein	-
BRPE64_RS25545	2	NC_021289.1	183937	184962	-	transcriptional regulator AraC family	Transcription
BRPE64_RS25550	2	NC_021289.1	185093	186031	-	hypothetical protein	Function unknown
BRPE64_RS25555	2	NC_021289.1	186076	187413	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS25560	2	NC_021289.1	187400	189007	-	extracellular solute-binding protein family 5	Amino acid transport and metabolism
BRPE64_RS25565	2	NC_021289.1	189124	190761	-	AMP-dependent synthetase and ligase	Lipid transport and metabolism
BRPE64_RS25570	2	NC_021289.1	190774	191169	-	endoribonuclease L-PSP	Translation, ribosomal structure and biogenesis
BRPE64_RS25575	2	NC_021289.1	191166	191615	-	thioesterase superfamily protein	General function prediction only
BRPE64_RS25580	2	NC_021289.1	191612	192793	-	acyl-CoA dehydrogenase domain protein	Lipid transport and metabolism
BRPE64_RS25585	2	NC_021289.1	192795	193643	-	enoyl-CoA hydratase	Lipid transport and metabolism
BRPE64_RS25590	2	NC_021289.1	193655	194215	-	transcriptional regulator MarR family	Transcription
BRPE64_RS25595	2	NC_021289.1	194212	194985	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS25600	2	NC_021289.1	194988	197372	-	NADH flavin oxidoreductase/NADH oxidase	Energy production and conversion
BRPE64_RS25715	2	NC_021289.1	221852	223603	-	ABC-type siderophore export system fused ATPase and permease components	Inorganic ion transport and metabolism
BRPE64_RS25720	2	NC_021289.1	223866	233405	-	amino acid adenylation domain protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS25725	2	NC_021289.1	233431	238398	-	amino acid adenylation domain protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS25730	2	NC_021289.1	238412	239776	-	L-ornithine 5-monooxygenase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS25880	2	NC_021289.1	270671	271447	-	putative NAD(P)H dehydrogenase	General function prediction only
BRPE64_RS25885	2	NC_021289.1	271540	272436	-	transcriptional regulator LysR family	Transcription
BRPE64_RS25890	2	NC_021289.1	272598	273863	-	extracellular ligand-binding receptor	Amino acid transport and metabolism

BRPE64_RS25895	2	NC_021289.1	273946	274821	-	inner-membrane translocator	Amino acid transport and metabolism
BRPE64_RS25900	2	NC_021289.1	274824	275750	-	inner-membrane translocator	Amino acid transport and metabolism
BRPE64_RS25905	2	NC_021289.1	275747	277261	-	ABC transporter related protein	Amino acid transport and metabolism
BRPE64_RS25910	2	NC_021289.1	277521	278408	-	hypothetical protein	-
BRPE64_RS32015	2	NC_021289.1	278411	279238	-	hypothetical protein	-
BRPE64_RS25920	2	NC_021289.1	279235	280314	-	hypothetical protein	-
BRPE64_RS25965	2	NC_021289.1	293746	293946	-	hypothetical protein	-
BRPE64_RS25970	2	NC_021289.1	294081	295040	-	inner-membrane translocator	Carbohydrate transport and metabolism
BRPE64_RS25975	2	NC_021289.1	295049	296029	-	inner-membrane translocator	Carbohydrate transport and metabolism
BRPE64_RS25980	2	NC_021289.1	296019	297581	-	ABC transporter related protein	Carbohydrate transport and metabolism
BRPE64_RS26080	2	NC_021289.1	320699	320875	-	Flp/Fap pilin component	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS26215	2	NC_021289.1	348674	349708	-	transcriptional regulator AraC family	Transcription
BRPE64_RS26220	2	NC_021289.1	349605	349970	-	hypothetical protein	-
BRPE64_RS26225	2	NC_021289.1	350138	350320	-	Mlr0331 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS26230	2	NC_021289.1	350322	351401	-	putative ABC transporter substrate-binding protein	Inorganic ion transport and metabolism
BRPE64_RS26235	2	NC_021289.1	351457	352479	-	putative aldoketoreductase	Energy production and conversion
BRPE64_RS26240	2	NC_021289.1	352501	353844	-	DDVA O-demethylase	Inorganic ion transport and metabolism
BRPE64_RS26245	2	NC_021289.1	353980	354783	-	putative transcriptional regulator IclR family	Transcription
BRPE64_RS26250	2	NC_021289.1	354826	355791	-	putative oxidoreductase	Energy production and conversion
BRPE64_RS26255	2	NC_021289.1	355861	357132	-	major facilitator superfamily (MFS) transporter	Amino acid transport and metabolism
BRPE64_RS26275	2	NC_021289.1	360346	360672	*hcaC*	rieske (2Fe-2S) domain protein	Inorganic ion transport and metabolism
BRPE64_RS26280	2	NC_021289.1	360698	361633	-	cobalamin synthesis protein/P47K family protein	General function prediction only
BRPE64_RS26285	2	NC_021289.1	361646	362359	-	class II aldolase/adducin family protein	Carbohydrate transport and metabolism
BRPE64_RS26290	2	NC_021289.1	362385	363185	-	putative taurine transport system permease protein	Inorganic ion transport and metabolism
BRPE64_RS26295	2	NC_021289.1	363215	364060	-	ABC transporter ATP-binding protein	Inorganic ion transport and metabolism

BRPE64_RS26300	2	NC_021289.1	364070	365092	-	ABC transporter substrate-binding protein	Inorganic ion transport and metabolism
BRPE64_RS26305	2	NC_021289.1	365313	366449	-	outer membrane porin OmpC family	Cell wall/membrane/envelope biogenesis
BRPE64_RS26310	2	NC_021289.1	366934	367740	-	transcriptional regulator IclR family protein	Transcription
BRPE64_RS26315	2	NC_021289.1	367767	368807	-	ferredoxin Oxidoreductase FAD/NAD(P)-binding Oxidoreductase FAD-binding region	Energy production and conversion
BRPE64_RS26320	2	NC_021289.1	368881	369873	-	hypothetical protein	-
BRPE64_RS26325	2	NC_021289.1	369905	371026	-	hypothetical protein	Function unknown
BRPE64_RS26330	2	NC_021289.1	371155	372144	-	phthalate 4,5-dioxygenase reductase subunit	Energy production and conversion
BRPE64_RS26335	2	NC_021289.1	372192	373379	-	oxidoreductase-like protein	General function prediction only
BRPE64_RS26340	2	NC_021289.1	373376	373930	-	transcriptional regulator MarR family	Transcription
BRPE64_RS26370	2	NC_021289.1	380998	381606	-	chromate transporter	Inorganic ion transport and metabolism
BRPE64_RS26375	2	NC_021289.1	381611	382138	-	probable transmembrane protein	Inorganic ion transport and metabolism
BRPE64_RS26380	2	NC_021289.1	382143	383147	-	alcohol dehydrogenase zinc-binding domain protein	Energy production and conversion
BRPE64_RS26385	2	NC_021289.1	383247	384116	-	transcriptional regulator LysR family	Transcription
BRPE64_RS26455	2	NC_021289.1	398570	399598	-	hypothetical protein	Function unknown
BRPE64_RS26475	2	NC_021289.1	406975	407871	-	transglutaminase domain protein	Amino acid transport and metabolism
BRPE64_RS26480	2	NC_021289.1	407978	408955	-	oxidoreductase (Aldo/keto reductase) protein	Energy production and conversion
BRPE64_RS26590	2	NC_021289.1	432863	433048	-	hypothetical protein	-
BRPE64_RS26595	2	NC_021289.1	433198	434418	-	cytochrome c class I	Amino acid transport and metabolism
BRPE64_RS26600	2	NC_021289.1	434518	436254	-	putative alcohol dehydrogenase	Carbohydrate transport and metabolism
BRPE64_RS26605	2	NC_021289.1	436482	436763	-	hypothetical protein	-
BRPE64_RS26610	2	NC_021289.1	437094	439034	-	GAF modulated sigma54 specific transcriptional regulator Fis family	Transcription
BRPE64_RS26690	2	NC_021289.1	457869	458261	-	hypothetical protein	-
BRPE64_RS26695	2	NC_021289.1	458290	458661	-	hypothetical protein	-
BRPE64_RS26700	2	NC_021289.1	458819	460300	-	phosphoesterase	Cell wall/membrane/envelope biogenesis
BRPE64_RS26705	2	NC_021289.1	460699	460989	-	hypothetical protein	-

BRPE64_RS26710	2	NC_021289.1	460986	461510	-	putative GCN5-related N-acetyltransferase	Transcription
BRPE64_RS26715	2	NC_021289.1	461820	462830	-	hypothetical protein	Energy production and conversion
BRPE64_RS26740	2	NC_021289.1	469810	470769	-	MoxR-like ATPase putative transcriptional regulator C1 metabolism	General function prediction only
BRPE64_RS26745	2	NC_021289.1	470814	471773	-	hypothetical protein	General function prediction only
BRPE64_RS26750	2	NC_021289.1	471770	472285	-	hypothetical protein	-
BRPE64_RS26755	2	NC_021289.1	472272	473288	-	hypothetical protein	General function prediction only
BRPE64_RS26760	2	NC_021289.1	473290	474867	-	hypothetical TPR domain protein	General function prediction only
BRPE64_RS26820	2	NC_021289.1	486774	487370	-	uncharacterized peroxidase-related enzyme	Function unknown
BRPE64_RS26825	2	NC_021289.1	487447	490056	-	PAS/PAC sensor hybrid histidine kinase	Signal transduction mechanisms
BRPE64_RS26850	2	NC_021289.1	495291	496172	-	RNA polymerase sigma-24 subunit ECF subfamily	Transcription
BRPE64_RS26855	2	NC_021289.1	496325	496804	-	alkylhydroperoxidase like protein AhpD family	Function unknown
BRPE64_RS26860	2	NC_021289.1	496843	497250	-	cupin 2 conserved barrel domain protein	Function unknown
BRPE64_RS26865	2	NC_021289.1	497349	498362	-	transcriptional regulator LysR family	Transcription
BRPE64_RS26870	2	NC_021289.1	498603	499121	-	OsmC family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS26875	2	NC_021289.1	499134	500522	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS26880	2	NC_021289.1	500586	501356	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26885	2	NC_021289.1	501370	501708	-	hypothetical protein	-
BRPE64_RS26890	2	NC_021289.1	501705	503450	-	fumarate reductase/succinate dehydrogenase flavoprotein domain protein	Energy production and conversion
BRPE64_RS26895	2	NC_021289.1	503505	504344	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26900	2	NC_021289.1	504480	505373	-	transcriptional regulator AraC family	Transcription
BRPE64_RS26905	2	NC_021289.1	505496	507145	*treA*	alpha alpha-trehalase	Carbohydrate transport and metabolism
BRPE64_RS26910	2	NC_021289.1	507160	507696	-	cytochrome c class I	Energy production and conversion
BRPE64_RS26915	2	NC_021289.1	507689	508900	-	oxidoreductase molybdopterin binding protein	General function prediction only

BRPE64_RS26920	2	NC_021289.1	509169	509405	-	4-oxalocrotonate tautomerase family enzyme	General function prediction only
BRPE64_RS26925	2	NC_021289.1	509454	510164	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26975	2	NC_021289.1	519019	520215	-	acyltransferase 3	Lipid transport and metabolism
BRPE64_RS26980	2	NC_021289.1	520217	521581	-	multi antimicrobial extrusion protein MatE	Defense mechanisms
BRPE64_RS27050	2	NC_021289.1	537592	539115	-	sugar ABC transporter ATPase component	Carbohydrate transport and metabolism
BRPE64_RS27055	2	NC_021289.1	539121	540095	-	periplasmic binding protein/LacI transcriptional regulator	Carbohydrate transport and metabolism
BRPE64_RS27060	2	NC_021289.1	540196	541257	-	alanine racemase	Transcription
BRPE64_RS27065	2	NC_021289.1	541403	541987	-	hypothetical protein	Function unknown
BRPE64_RS27070	2	NC_021289.1	542145	543554	-	hypothetical protein	Function unknown
BRPE64_RS27075	2	NC_021289.1	543575	543832	-	hypothetical protein	-
BRPE64_RS27080	2	NC_021289.1	543874	544713	-	hypothetical protein	Carbohydrate transport and metabolism
BRPE64_RS27085	2	NC_021289.1	544825	545460	-	transcriptional regulator TetR family	Transcription
BRPE64_RS27090	2	NC_021289.1	545486	546568	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS27095	2	NC_021289.1	547804	548856	-	transcriptional regulator LacI family	Carbohydrate transport and metabolism
BRPE64_RS27255	2	NC_021289.1	579480	579740	-	hypothetical protein	-
BRPE64_RS27260	2	NC_021289.1	579947	580921	-	aliphatic sulfonates family ABC transporter periplasmic ligand-binding protein	Inorganic ion transport and metabolism
BRPE64_RS27375	2	NC_021289.1	602604	602831	-	hypothetical protein	-
BRPE64_RS27380	2	NC_021289.1	602852	603976	-	alanine racemase domain protein	Amino acid transport and metabolism
BRPE64_RS27385	2	NC_021289.1	603973	604626	-	HAD-superfamily hydrolase subfamily IA variant 2	General function prediction only
BRPE64_RS27390	2	NC_021289.1	604726	605631	-	transcriptional regulator LysR family	Transcription
BRPE64_RS27395	2	NC_021289.1	605638	606540	-	hypothetical protein	Amino acid transport and metabolism
BRPE64_RS27400	2	NC_021289.1	606730	607035	-	ferredoxin	Inorganic ion transport and metabolism
BRPE64_RS27405	2	NC_021289.1	607058	608302	*hcaD*	3-phenylpropionate dioxygenase ferredoxin-NAD(+) reductase component	General function prediction only
BRPE64_RS27410	2	NC_021289.1	608289	609581	-	probable Ring hydroxylating alpha subunit	Inorganic ion transport and metabolism
BRPE64_RS27415	2	NC_021289.1	609582	610091	-	hypothetical protein	-

BRPE64_RS27420	2	NC_021289.1	610114	612546	-	xanthine dehydrogenase molybdenum binding subunit apoprotein	Energy production and conversion
BRPE64_RS27425	2	NC_021289.1	612543	613379	-	oxidoreductase medium chain	Energy production and conversion
BRPE64_RS27430	2	NC_021289.1	613390	614580	-	putative iron-sulfur binding protein	Energy production and conversion
BRPE64_RS27435	2	NC_021289.1	614689	615654	-	amidohydrolase family protein	General function prediction only
BRPE64_RS27440	2	NC_021289.1	615688	616440	-	putative MALEATE CIS-TRANS ISOMERASE	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS27445	2	NC_021289.1	616452	617294	-	hydrolase or acyltransferase alpha/beta hydrolase superfamily	General function prediction only
BRPE64_RS27450	2	NC_021289.1	617281	618315	-	putative peptidase M29 family	Amino acid transport and metabolism
BRPE64_RS27455	2	NC_021289.1	618334	618957	-	isochorismatase family protein 7	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS27460	2	NC_021289.1	618985	620580	-	probable transporter	Amino acid transport and metabolism
BRPE64_RS27465	2	NC_021289.1	620595	621041	-	probable MarR-family transcriptional regulator	Transcription
BRPE64_RS27485	2	NC_021289.1	624479	626593	-	hypothetical protein	-
BRPE64_RS27490	2	NC_021289.1	627066	627488	-	hypothetical protein	-
BRPE64_RS32025	2	NC_021289.1	627485	630376	-	tyrosinase	Amino acid transport and metabolism
BRPE64_RS27500	2	NC_021289.1	630567	630785	-	hypothetical protein	-
BRPE64_RS27505	2	NC_021289.1	630789	631745	-	LysR family regulatory protein	Transcription
BRPE64_RS27510	2	NC_021289.1	631902	633212	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS27515	2	NC_021289.1	633247	634503	-	metallo peptidase family M20 unassigned	Amino acid transport and metabolism
BRPE64_RS27520	2	NC_021289.1	634514	635539	-	putative aminohydrolase	Chromatin structure and dynamics
BRPE64_RS27525	2	NC_021289.1	635701	636525	-	lipid A biosynthesis lauroyl acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS27535	2	NC_021289.1	637376	638260	-	periplasmic protein-like protein	Function unknown
BRPE64_RS27540	2	NC_021289.1	638309	638605	-	hypothetical protein	-
BRPE64_RS27545	2	NC_021289.1	639206	640537	-	PAS/PAC sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS27550	2	NC_021289.1	640534	641178	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms

BRPE64_RS27555	2	NC_021289.1	641318	641704	-	response regulator receiver protein	Signal transduction mechanisms
BRPE64_RS27830	2	NC_021289.1	702926	703306	-	hypothetical protein	-
BRPE64_RS27835	2	NC_021289.1	703660	704820	-	alpha-methylacyl-CoA racemase	Energy production and conversion
BRPE64_RS27840	2	NC_021289.1	705512	707179	-	hypothetical protein	-
BRPE64_RS27940	2	NC_021289.1	728742	729446	-	hypothetical protein	-
BRPE64_RS27945	2	NC_021289.1	729500	730708	-	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Energy production and conversion
BRPE64_RS27950	2	NC_021289.1	730857	731243	-	hypothetical protein	-
BRPE64_RS27955	2	NC_021289.1	731445	732320	-	transcriptional regulator IclR family	Transcription
BRPE64_RS27960	2	NC_021289.1	732615	733706	-	ABC transporter related	Amino acid transport and metabolism
BRPE64_RS27965	2	NC_021289.1	733699	734616	-	binding-protein-dependent transport systems inner membrane component	Amino acid transport and metabolism
BRPE64_RS27970	2	NC_021289.1	734636	735430	-	binding-protein-dependent transport systems inner membrane component	Amino acid transport and metabolism
BRPE64_RS28000	2	NC_021289.1	740141	740749	-	hypothetical protein	-
BRPE64_RS28005	2	NC_021289.1	740929	741246	-	hypothetical protein	-
BRPE64_RS28010	2	NC_021289.1	741486	742478	-	4,5-dihydroxyphthalate decarboxylase	Inorganic ion transport and metabolism
BRPE64_RS28015	2	NC_021289.1	742482	743696	-	oxidoreductase-like protein	General function prediction only
BRPE64_RS28225	2	NC_021289.1	784477	785094	-	3-isopropylmalate dehydratase small subunit	Amino acid transport and metabolism
BRPE64_RS28230	2	NC_021289.1	785091	786497	-	3-isopropylmalate dehydratase large subunit	Amino acid transport and metabolism
BRPE64_RS28235	2	NC_021289.1	786500	787273	-	ABC nitrate/sulfonate/bicarbonate family transporter inner membrane subunit	Inorganic ion transport and metabolism
BRPE64_RS28240	2	NC_021289.1	787270	788028	-	ABC nitrate/sulfonate/bicarbonate family transporter inner membrane subunit	Inorganic ion transport and metabolism
BRPE64_RS28245	2	NC_021289.1	788025	788885	-	ABC nitrate/sulfonate/bicarbonate family transporter ATPase subunit	Inorganic ion transport and metabolism
BRPE64_RS28250	2	NC_021289.1	788893	789879	-	ABC nitrate/sulfonate/bicarbonate family transporter periplasmic ligand binding protein	Inorganic ion transport and metabolism
BRPE64_RS28255	2	NC_021289.1	790124	790933	-	hypothetical protein	-
BRPE64_RS28260	2	NC_021289.1	791000	791695	-	transcriptional regulator GntR family	Transcription

BRPE64_RS28265	2	NC_021289.1	791706	792206	-	UspA domain protein	Signal transduction mechanisms
BRPE64_RS28270	2	NC_021289.1	792372	793640	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28275	2	NC_021289.1	793757	794425	-	O-methyltransferase family protein	General function prediction only
BRPE64_RS28280	2	NC_021289.1	794528	795115	-	hypothetical protein	-
BRPE64_RS28285	2	NC_021289.1	795175	795693	-	RNA polymerase sigma-24 subunit ECF subfamily	Transcription
BRPE64_RS28290	2	NC_021289.1	795690	796514	-	putative transmembrane anti-sigma factor	Transcription
BRPE64_RS28295	2	NC_021289.1	796519	798078	-	sulphate transporter	Inorganic ion transport and metabolism
BRPE64_RS28300	2	NC_021289.1	798164	798811	-	carbonic anhydrase	Inorganic ion transport and metabolism
BRPE64_RS28305	2	NC_021289.1	798876	799058	-	hypothetical protein	-
BRPE64_RS28310	2	NC_021289.1	799057	799884	-	alpha/beta hydrolase fold protein	General function prediction only
BRPE64_RS28315	2	NC_021289.1	799881	800735	-	polysaccharide deacetylase	Carbohydrate transport and metabolism
BRPE64_RS28320	2	NC_021289.1	800732	801817	-	putative glycosyl transferase group 1 family	Cell wall/membrane/envelope biogenesis
BRPE64_RS28355	2	NC_021289.1	809993	811060	-	hypothetical protein	Translation, ribosomal structure and biogenesis
BRPE64_RS28360	2	NC_021289.1	811057	812061	-	hypothetical protein	-
BRPE64_RS28365	2	NC_021289.1	812131	814596	-	glycoside hydrolase family 2 immunoglobulin domain protein beta-sandwich	Carbohydrate transport and metabolism
BRPE64_RS28370	2	NC_021289.1	814603	815541	-	dehydrogenase	General function prediction only
BRPE64_RS28375	2	NC_021289.1	815588	816517	-	dihydrodipicolinate synthase putative	Amino acid transport and metabolism
BRPE64_RS28380	2	NC_021289.1	816556	817902	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS28385	2	NC_021289.1	818017	819750	-	dihydroxy-acid dehydratase	Amino acid transport and metabolism
BRPE64_RS28390	2	NC_021289.1	819863	820843	-	transcriptional regulator LysR family putative	Transcription
BRPE64_RS28395	2	NC_021289.1	820807	824349	-	indolepyruvate ferredoxin oxidoreductase	Energy production and conversion
BRPE64_RS28400	2	NC_021289.1	824452	825357	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28405	2	NC_021289.1	825494	828739	-	hydrophobe/amphiphile efflux pump RND family	Defense mechanisms
BRPE64_RS28410	2	NC_021289.1	828748	829659	-	transcriptional regulator LysR family	Transcription

BRPE64_RS28415	2	NC_021289.1	829862	830500	-	putative glutathionine S-transferase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS28420	2	NC_021289.1	830744	831004	-	hypothetical protein	-
BRPE64_RS28425	2	NC_021289.1	831224	832006	-	hypothetical protein	-
BRPE64_RS28430	2	NC_021289.1	832439	832699	-	hypothetical protein	-
BRPE64_RS28435	2	NC_021289.1	832742	833023	-	hypothetical protein	-
BRPE64_RS28440	2	NC_021289.1	833066	833347	-	hypothetical protein	-
BRPE64_RS28445	2	NC_021289.1	833390	833650	-	hypothetical protein	-
BRPE64_RS28450	2	NC_021289.1	833711	834664	-	putative hydrolase	General function prediction only
BRPE64_RS28455	2	NC_021289.1	834692	835618	-	amidohydrolase 2	General function prediction only
BRPE64_RS28460	2	NC_021289.1	835615	836031	-	thioesterase/thiol ester dehydrase-isomerase	General function prediction only
BRPE64_RS28465	2	NC_021289.1	836035	837933	-	TRAP C4-dicarboxylate transport system permease DctM subunit	Carbohydrate transport and metabolism
BRPE64_RS28470	2	NC_021289.1	837947	838978	-	Blr4511 protein	Carbohydrate transport and metabolism
BRPE64_RS28475	2	NC_021289.1	839030	840229	-	putative formyl-coenzyme A transferase (Formyl-CoA transferase) Frc	Energy production and conversion
BRPE64_RS28480	2	NC_021289.1	840231	841010	-	enoyl-CoA hydratase EchA	Lipid transport and metabolism
BRPE64_RS28485	2	NC_021289.1	841031	842170	-	4-hydroxybutyrate dehydrogenase protein	Energy production and conversion
BRPE64_RS28490	2	NC_021289.1	842350	843294	-	LysR family transcription regulator protein	Transcription
BRPE64_RS28495	2	NC_021289.1	843428	844012	-	hypothetical protein	General function prediction only
BRPE64_RS28500	2	NC_021289.1	844009	844911	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28505	2	NC_021289.1	845119	845925	-	2,5-didehydrogluconate reductase	General function prediction only
BRPE64_RS28620	2	NC_021289.1	872107	873327	-	hypothetical protein	Energy production and conversion
BRPE64_RS28625	2	NC_021289.1	873349	875499	-	fusaric acid resistance protein conserved region	Function unknown
BRPE64_RS28630	2	NC_021289.1	875514	876437	-	aromatic acid efflux system membrane fusion protein EmrA subfamily	Defense mechanisms
BRPE64_RS28635	2	NC_021289.1	876509	876988	-	hypothetical protein	Function unknown
BRPE64_RS28640	2	NC_021289.1	877000	877821	-	alpha/beta hydrolase fold	General function prediction only
BRPE64_RS28645	2	NC_021289.1	878107	878679	-	hypothetical protein	Function unknown
BRPE64_RS28660	2	NC_021289.1	880703	880981	-	muconolactone delta-isomerase	Secondary metabolites biosynthesis, transport and catabolism

BRPE64_RS28665	2	NC_021289.1	881049	882176	-	muconate and chloromuconate cycloisomerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS28670	2	NC_021289.1	882280	883203	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28675	2	NC_021289.1	883462	884397	-	catechol 1,2-dioxygenase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS28680	2	NC_021289.1	884506	885864	-	rieske (2Fe-2S) domain protein	Inorganic ion transport and metabolism
BRPE64_RS28815	2	NC_021289.1	911880	912140	-	hypothetical protein	-
BRPE64_RS28820	2	NC_021289.1	912378	913442	-	molybdenum cofactor biosynthesis protein A	Coenzyme transport and metabolism
BRPE64_RS28870	2	NC_021289.1	924126	930071	-	PAS sensor protein	General function prediction only
BRPE64_RS28875	2	NC_021289.1	930207	931046	-	enoyl-CoA hydratase/isomerase	Lipid transport and metabolism
BRPE64_RS28880	2	NC_021289.1	931074	931751	-	ThiJ/PfpI domain protein	General function prediction only
BRPE64_RS28885	2	NC_021289.1	932347	932997	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms
BRPE64_RS28890	2	NC_021289.1	933146	933538	-	hypothetical protein	-
BRPE64_RS28895	2	NC_021289.1	933554	934567	-	putative hydrolase	General function prediction only
BRPE64_RS28900	2	NC_021289.1	934705	935001	-	putative transcriptional regulator	Transcription
BRPE64_RS28905	2	NC_021289.1	935020	936822	-	acetolactate synthase	Amino acid transport and metabolism
BRPE64_RS28910	2	NC_021289.1	937273	938301	-	transcriptional regulator AraC family	Transcription
BRPE64_RS28915	2	NC_021289.1	938294	938815	-	uracil-DNA glycosylase superfamily	Replication, recombination and repair
BRPE64_RS28920	2	NC_021289.1	938820	939767	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28925	2	NC_021289.1	940092	940766	-	antibiotic biosynthesis monooxygenase	General function prediction only
BRPE64_RS28930	2	NC_021289.1	941147	942721	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS28935	2	NC_021289.1	942734	943810	-	secretion protein HlyD family protein	Defense mechanisms
BRPE64_RS28940	2	NC_021289.1	943807	945345	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS28945	2	NC_021289.1	945368	945745	-	hypothetical protein	-
BRPE64_RS28950	2	NC_021289.1	945788	946711	-	transcriptional regulator AraC family	Transcription
BRPE64_RS28955	2	NC_021289.1	947010	950204	-	acriflavin resistance protein	Defense mechanisms

BRPE64_RS28960	2	NC_021289.1	950215	951378	-	efflux transporter RND family MFP subunit	Cell wall/membrane/envelope biogenesis
BRPE64_RS28965	2	NC_021289.1	951392	952852	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS29175	2	NC_021289.1	995247	995876	-	hypothetical protein	General function prediction only
BRPE64_RS29180	2	NC_021289.1	996021	996872	-	alpha/beta hydrolase fold protein	General function prediction only
BRPE64_RS29185	2	NC_021289.1	996918	997889	-	alcohol dehydrogenase zinc-binding domain protein	Energy production and conversion
BRPE64_RS29190	2	NC_021289.1	997905	998675	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS29200	2	NC_021289.1	999006	999782	-	oxidoreductase molybdopterin binding protein	General function prediction only
BRPE64_RS29205	2	NC_021289.1	999802	1000431	-	putative transmembrane hydrogenase cytochrome b-type subunit	Energy production and conversion
BRPE64_RS29210	2	NC_021289.1	1000774	1001298	-	hypothetical protein	-
BRPE64_RS29215	2	NC_021289.1	1001295	1001936	-	hypothetical protein	-
BRPE64_RS29220	2	NC_021289.1	1001937	1002560	-	transcriptional regulator TetR family protein	Transcription
BRPE64_RS29225	2	NC_021289.1	1002576	1003280	-	glutathione S-transferase domain	Posttranslational modification, protein turnover, chaperones
BRPE64_RS29230	2	NC_021289.1	1003366	1004385	-	putative dehydrogenase	Energy production and conversion
BRPE64_RS29235	2	NC_021289.1	1004382	1005704	-	C4-dicarboxylate transport protein	Energy production and conversion
BRPE64_RS29240	2	NC_021289.1	1005857	1007032	-	altronate dehydratase	Carbohydrate transport and metabolism
BRPE64_RS29245	2	NC_021289.1	1007082	1007351	-	hypothetical protein	-
BRPE64_RS29250	2	NC_021289.1	1007619	1008371	-	transcriptional regulator GntR family	Transcription
BRPE64_RS29255	2	NC_021289.1	1008404	1008784	-	hypothetical protein	-
BRPE64_RS29260	2	NC_021289.1	1008820	1009278	-	MEKHLA domain protein	Signal transduction mechanisms
BRPE64_RS29265	2	NC_021289.1	1009280	1010140	-	NmrA family protein	Carbohydrate transport and metabolism
BRPE64_RS29270	2	NC_021289.1	1010262	1011170	-	transcriptional regulator	Transcription
BRPE64_RS29275	2	NC_021289.1	1011280	1011893	-	hypothetical protein	-
BRPE64_RS29280	2	NC_021289.1	1012163	1014901	-	hypothetical protein	-
BRPE64_RS29285	2	NC_021289.1	1014909	1016366	-	succinate semialdehyde dehydrogenase	Energy production and conversion
BRPE64_RS29290	2	NC_021289.1	1016500	1018170	*actP*	SSS sodium solute transporter superfamily	General function prediction only
BRPE64_RS29295	2	NC_021289.1	1018167	1018472	-	hypothetical protein	Function unknown

BRPE64_RS29635	2	NC_021289.1	1094891	1095277	-	hypothetical protein	General function prediction only
BRPE64_RS29640	2	NC_021289.1	1095281	1096051	-	aliphatic sulfonate import ATP-binding protein SsuB 2	Inorganic ion transport and metabolism
BRPE64_RS29645	2	NC_021289.1	1096065	1096847	-	binding-protein-dependent transport systems inner membrane component	Inorganic ion transport and metabolism
BRPE64_RS29650	2	NC_021289.1	1096859	1097695	-	binding-protein-dependent transport systems inner membrane component	Inorganic ion transport and metabolism
BRPE64_RS29655	2	NC_021289.1	1097717	1098781	-	ABC nitrate/sulfonate/bicarbonate family transporter periplasmic ligand binding protein	Inorganic ion transport and metabolism
BRPE64_RS29660	2	NC_021289.1	1098849	1099787	-	taurine dioxygenase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS29665	2	NC_021289.1	1099917	1101008	-	2-nitropropane dioxygenase	General function prediction only
BRPE64_RS29670	2	NC_021289.1	1101038	1101952	-	transcriptional regulator LysR family	Transcription
BRPE64_RS29695	2	NC_021289.1	1106655	1107551	-	fructose-bisphosphate aldolase	Carbohydrate transport and metabolism
BRPE64_RS29700	2	NC_021289.1	1107730	1108164	-	hypothetical protein	-
BRPE64_RS29720	2	NC_021289.1	1111956	1113017	-	squalene/phytoene synthase family protein	Lipid transport and metabolism
BRPE64_RS32030	2	NC_021289.1	1113334	1113807	-	hypothetical protein	-
BRPE64_RS29730	2	NC_021289.1	1113954	1114886	-	UDP-glucose pyrophosphorylase	Cell wall/membrane/envelope biogenesis
BRPE64_RS29735	2	NC_021289.1	1115311	1116450	-	acyltransferase 3	Lipid transport and metabolism
BRPE64_RS29740	2	NC_021289.1	1116464	1118935	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS29745	2	NC_021289.1	1118932	1120431	-	polysaccharide biosynthesis protein	General function prediction only
BRPE64_RS29750	2	NC_021289.1	1120449	1122653	-	hypothetical protein	-
BRPE64_RS29755	2	NC_021289.1	1122863	1123906	-	GDP-mannose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS29760	2	NC_021289.1	1123914	1124831	-	NAD-dependent epimerase/dehydratase	Carbohydrate transport and metabolism
BRPE64_RS29765	2	NC_021289.1	1124871	1126049	-	glycosyl transferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS29770	2	NC_021289.1	1126085	1127269	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS29775	2	NC_021289.1	1127232	1128758	-	hypothetical protein	-
BRPE64_RS29780	2	NC_021289.1	1128755	1129888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis

BRPE64_RS29785	2	NC_021289.1	1129885	1130856	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS29790	2	NC_021289.1	1130884	1133097	-	capsular exopolysaccharide family	Cell wall/membrane/envelope biogenesis
BRPE64_RS29795	2	NC_021289.1	1133117	1134301	*wza*	polysaccharide export protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS29800	2	NC_021289.1	1134280	1134723	-	protein tyrosine phosphatase	Signal transduction mechanisms
BRPE64_RS29805	2	NC_021289.1	1134732	1136129	-	nucleotide sugar dehydrogenase	Cell wall/membrane/envelope biogenesis
BRPE64_RS29810	2	NC_021289.1	1136164	1137561	-	undecaprenyl-phosphate glucose phosphotransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS29830	2	NC_021289.1	1142044	1143165	-	acyltransferase 3	Lipid transport and metabolism
BRPE64_RS29835	2	NC_021289.1	1143239	1143706	-	transcriptional regulator MarR-family	Transcription
BRPE64_RS30100	2	NC_021289.1	1198538	1198867	-	hypothetical protein	-
BRPE64_RS30105	2	NC_021289.1	1198966	1200720	-	hypothetical protein	-
BRPE64_RS30150	2	NC_021289.1	1208068	1208370	-	hypothetical protein	-
BRPE64_RS30155	2	NC_021289.1	1208404	1208835	-	hypothetical protein	-
BRPE64_RS30160	2	NC_021289.1	1209233	1209418	-	putative periplasmic nitrate reductase NapE	Energy production and conversion
BRPE64_RS30165	2	NC_021289.1	1209454	1209792	-	periplasmic nitrate reductase chaperone NapD	Inorganic ion transport and metabolism
BRPE64_RS30485	2	NC_021295.1	41	1423	-	hypothetical protein	-
BRPE64_RS30490	2	NC_021295.1	2211	3419	-	cobyrinic acid ac-diamide synthase	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS30495	2	NC_021295.1	3416	4387	-	ParB-like partition protein	Transcription
BRPE64_RS30615	2	NC_021295.1	29840	30175	-	hypothetical protein	-
BRPE64_RS30620	2	NC_021295.1	30884	31384	-	hypothetical protein	-
BRPE64_RS31095	2	NC_021295.1	145393	146151	-	hypothetical protein	-
BRPE64_RS31385	2	NC_021295.1	218747	219211	-	hypothetical protein	-
BRPE64_RS31585	2	NC_021295.1	263445	263879	-	hypothetical protein	-
BRPE64_RS31590	2	NC_021295.1	263922	264458	-	hypothetical protein	-
BRPE64_RS31595	2	NC_021295.1	264617	264910	-	hypothetical protein	-

Annexe 2: List of fitness genes required in MM supplemented with glucose in *B. insecticola* identified by Con-ARTIST.

		NC_021287.1	NC_021294.1	NC_021288.1	NC_021289.1	NC_021295.1	
	Essentiality score	Chromosome 1	Chromosome 2	Chromosome 3	Plasmid 1	Plasmid 2	Total
Conditionally-essential genes	1	18	0	0	0	0	18
Conditionally-essential domains	2	53	0	0	0	0	53
Conditionally-enriched genes	3	8	0	0	0	0	8
Conditionally-enriched domains	4	16	0	0	0	0	16
Neutral genes	5	2623	1319	788	1157	262	6149

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS00935	1	NC_021287.1	198756	199502	*gpmA*	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	Carbohydrate transport and metabolism
BRPE64_RS01390	1	NC_021287.1	288271	289794	*ilvA*	L-threonine ammonia-lyase	Amino acid transport and metabolism
BRPE64_RS01655	1	NC_021287.1	349176	350741	*purH*	bifunctional purine biosynthesis protein PurH	Nucleotide transport and metabolism
BRPE64_RS02340	1	NC_021287.1	504058	504948	*purC*	phosphoribosylaminoimidazole-succinocarboxamidesynthase	Nucleotide transport and metabolism
BRPE64_RS02795	1	NC_021287.1	617201	618067	*nadC*	nicotinate-nucleotide pyrophosphorylase	Coenzyme transport and metabolism
BRPE64_RS02975	1	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS03960	1	NC_021287.1	860619	861737	-	ABC transporter related	Carbohydrate transport and metabolism
BRPE64_RS03990	1	NC_021287.1	867946	869868	*glk*	bifunctional protein glk	Carbohydrate transport and metabolism
BRPE64_RS04905	1	NC_021287.1	1075733	1076368	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms
BRPE64_RS05345	1	NC_021287.1	1163692	1164798	*aroC*	chorismate synthase	Amino acid transport and metabolism
BRPE64_RS06390	1	NC_021287.1	1369578	1370900	-	homoserine dehydrogenase	Amino acid transport and metabolism
BRPE64_RS06600	1	NC_021287.1	1421063	1423570	-	beta-N-acetylhexosaminidase	Carbohydrate transport and metabolism
BRPE64_RS07565	1	NC_021287.1	1641770	1643038	*hflX*	GTPase HflX	General function prediction only
BRPE64_RS09940	1	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane/envelope biogenesis

BRPE64_RS11130	1	NC_021287.1	2412821	2414770	*edd*	6-phosphogluconate dehydratase	Amino acid transport and metabolism
BRPE64_RS11215	1	NC_021287.1	2430339	2431349	-	glyceraldehyde-3-phosphate dehydrogenase type I	Carbohydrate transport and metabolism
BRPE64_RS11490	1	NC_021287.1	2491939	2493033	-	GTPase obg	General function prediction only
BRPE64_RS12160	1	NC_021287.1	2628391	2629857	*gltD*	glutamate synthase (NADH) small subunit	Amino acid transport and metabolism
BRPE64_RS00270	2	NC_021287.1	59268	59876	-	methionine biosynthesis protein MetW	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS00275	2	NC_021287.1	59873	61018	-	homoserine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS00285	2	NC_021287.1	62059	62958	-	acetylglutamate kinase	Amino acid transport and metabolism
BRPE64_RS01395	2	NC_021287.1	290326	294432	-	FAD linked oxidase domain protein	Energy production and conversion
BRPE64_RS01660	2	NC_021287.1	350806	351039	*fis*	DNA-binding protein Fis	Transcription
BRPE64_RS01815	2	NC_021287.1	385035	385847	*proC*	pyrroline-5-carboxylate reductase	Amino acid transport and metabolism
BRPE64_RS02135	2	NC_021287.1	460885	461940	*purM*	phosphoribosylformyl glycinamidine cyclo-ligase	Nucleotide transport and metabolism
BRPE64_RS02345	2	NC_021287.1	505002	505523	*purE*	N5-carboxyaminoimidazole ribonucleotide mutase	Nucleotide transport and metabolism
BRPE64_RS02350	2	NC_021287.1	505612	506799	*purK*	phosphoribosylaminoimidazole carboxylase ATPase subunit	Nucleotide transport and metabolism
BRPE64_RS02790	2	NC_021287.1	616077	617204	*nadA*	quinolinate synthase A	Coenzyme transport and metabolism
BRPE64_RS02980	2	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS02985	2	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS02990	2	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS02995	2	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS03000	2	NC_021287.1	655612	656925	*cysN*	sulfate adenylyltransferase large subunit	Inorganic ion transport and metabolism
BRPE64_RS03005	2	NC_021287.1	656943	657698	-	uroporphyrin-III C-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS03010	2	NC_021287.1	657835	658215	-	cobalamin (Vitamin B12) biosynthesis CbiX protein	Function unknown
BRPE64_RS03245	2	NC_021287.1	707442	708845	*argH*	argininosuccinate lyase	Amino acid transport and metabolism
BRPE64_RS03260	2	NC_021287.1	710699	713785	*ppc*	phosphoenolpyruvate carboxylase	Energy production and conversion

BRPE64_RS03965	2	NC_021287.1	861787	862632	-	binding-protein-dependent transport systems inner membrane component	Carbohydrate transport and metabolism
BRPE64_RS03970	2	NC_021287.1	862634	863572	-	carbohydrate ABC transporter membrane protein 1 CUT1 family	Carbohydrate transport and metabolism
BRPE64_RS03975	2	NC_021287.1	863764	865017	-	extracellular solute-binding protein family 1	Carbohydrate transport and metabolism
BRPE64_RS03980	2	NC_021287.1	865641	867101	-	glucose-6-phosphate 1-dehydrogenase	Carbohydrate transport and metabolism
BRPE64_RS03985	2	NC_021287.1	867246	867965	-	6-phosphogluconolactonase	Carbohydrate transport and metabolism
BRPE64_RS04790	2	NC_021287.1	1045695	1047077	*argA*	amino-acid acetyltransferase	Amino acid transport and metabolism
BRPE64_RS04910	2	NC_021287.1	1076365	1078893	-	multi-sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS05675	2	NC_021287.1	1225755	1227020	-	succinyldiaminopimelate aminotransferase apoenzyme	Amino acid transport and metabolism
BRPE64_RS06385	2	NC_021287.1	1368315	1369553	-	aminotransferase AlaT	Amino acid transport and metabolism
BRPE64_RS06595	2	NC_021287.1	1416832	1420869	*purL*	phosphoribosylformylglycinamidine synthase	Nucleotide transport and metabolism
BRPE64_RS07445	2	NC_021287.1	1612480	1613319	*serB*	phosphoserine phosphatase SerB	Amino acid transport and metabolism
BRPE64_RS07490	2	NC_021287.1	1621985	1624027	-	poly(R)-hydroxyalkanoic acid synthase class I	Lipid transport and metabolism
BRPE64_RS07550	2	NC_021287.1	1639153	1639344	-	hypothetical protein	Function unknown
BRPE64_RS07555	2	NC_021287.1	1639396	1640298	*hflC*	band 7 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS07560	2	NC_021287.1	1640310	1641710	*hflK*	protease FtsH subunit HflK	Posttranslational modification, protein turnover, chaperones
BRPE64_RS08840	2	NC_021287.1	1916590	1917606	*ilvC*	ketol-acid reductoisomerase	Amino acid transport and metabolism
BRPE64_RS08845	2	NC_021287.1	1917676	1918167	-	acetolactate synthase small subunit	Amino acid transport and metabolism
BRPE64_RS08850	2	NC_021287.1	1918274	1920037	-	acetolactate synthase large subunit biosynthetic type	Amino acid transport and metabolism
BRPE64_RS09985	2	NC_021287.1	2152076	2153158	*pheA*	chorismate mutase	Amino acid transport and metabolism
BRPE64_RS10830	2	NC_021287.1	2355573	2356502	-	ornithine carbamoyltransferase	Amino acid transport and metabolism
BRPE64_RS10835	2	NC_021287.1	2356655	2357884	*argG*	argininosuccinate synthase	Amino acid transport and metabolism

BRPE64_RS11220	2	NC_021287.1	2431398	2433449	-	transketolase 1	Carbohydrate transport and metabolism
BRPE64_RS11485	2	NC_021287.1	2490771	2491889	*proB*	glutamate 5-kinase	Amino acid transport and metabolism
BRPE64_RS11755	2	NC_021287.1	2551158	2552654	*trpE*	anthranilate synthase component I	Amino acid transport and metabolism
BRPE64_RS11760	2	NC_021287.1	2552667	2553254	-	glutamine amidotransferase of anthranilate synthase	Amino acid transport and metabolism
BRPE64_RS11765	2	NC_021287.1	2553259	2554293	*trpD*	anthranilate phosphoribosyltransferase	Amino acid transport and metabolism
BRPE64_RS11770	2	NC_021287.1	2554305	2555090	*trpC*	indole-3-glycerol phosphate synthase	Amino acid transport and metabolism
BRPE64_RS12165	2	NC_021287.1	2629958	2634661	*gltB*	glutamate synthase	Amino acid transport and metabolism
BRPE64_RS12260	2	NC_021287.1	2652894	2653202	-	hypothetical protein	-
BRPE64_RS12265	2	NC_021287.1	2653248	2654510	*lysA*	diaminopimelate decarboxylase	Amino acid transport and metabolism
BRPE64_RS12985	2	NC_021287.1	2789444	2790283	*metF*	methylenetetrahydrofolate reductase	Amino acid transport and metabolism
BRPE64_RS12990	2	NC_021287.1	2790324	2790677	-	hypothetical protein	Function unknown
BRPE64_RS12995	2	NC_021287.1	2790753	2792174	-	adenosylhomocysteinase	Coenzyme transport and metabolism
BRPE64_RS13660	2	NC_021287.1	2940847	2941263	*dksA*	transcriptional regulator TraR/DksA family	Signal transduction mechanisms
BRPE64_RS01850	3	NC_021287.1	393018	394703	-	hypothetical protein	-
BRPE64_RS05760	3	NC_021287.1	1249156	1250538	-	membrane-associated zinc metalloprotease	Cell wall/membrane/envelope biogenesis
BRPE64_RS07975	3	NC_021287.1	1737536	1738837	*sucB*	2-oxoglutarate dehydrogenase E2 subunit dihydrolipoamide succinyltransferase	Energy production and conversion
BRPE64_RS31880	3	NC_021287.1	2218036	2222010	-	type VI secretion system Vgr family protein	Function unknown
BRPE64_RS10270	3	NC_021287.1	2221955	2222902	-	hypothetical protein	-
BRPE64_RS12010	3	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12280	3	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12305	3	NC_021287.1	2662968	2664854	-	thiol disulfide interchange protein DsbD	Energy production and conversion
BRPE64_RS02740	4	NC_021287.1	603803	604249	*dut*	deoxyuridine 5'-triphosphate nucleotidohydrolase	Nucleotide transport and metabolism

BRPE64_RS03230	4	NC_021287.1	703248	705554	-	ornithine decarboxylase	Amino acid transport and metabolism
BRPE64_RS03900	4	NC_021287.1	846241	847866	-	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Coenzyme transport and metabolism
BRPE64_RS05615	4	NC_021287.1	1212426	1213733	-	major facilitator superfamily (MFS) transporter	Amino acid transport and metabolism
BRPE64_RS07530	4	NC_021287.1	1633539	1635431	-	probable potassium transport system protein kup	Inorganic ion transport and metabolism
BRPE64_RS07970	4	NC_021287.1	1735999	1737438	-	dihydrolipoyl dehydrogenase	Energy production and conversion
BRPE64_RS07985	4	NC_021287.1	1742123	1743949	-	GTP-binding protein TypA	Signal transduction mechanisms
BRPE64_RS09570	4	NC_021287.1	2069139	2069759	-	hypothetical protein	General function prediction only
BRPE64_RS10275	4	NC_021287.1	2223364	2223789	-	hypothetical protein	-
BRPE64_RS10555	4	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10560	4	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS11860	4	NC_021287.1	2573447	2573758	-	glutaredoxin	Posttranslational modification, protein turnover, chaperones
BRPE64_RS11865	4	NC_021287.1	2573770	2574378	-	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Coenzyme transport and metabolism
BRPE64_RS12015	4	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	4	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12285	4	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification, protein turnover, chaperones

Annexe 3: List of fitness genes required in MM supplemented with succinate in *B. insecticola* identified by Con-ARTIST.

		NC_021287.1	NC_021294.1	NC_021288.1	NC_021289.1	NC_021295.1	
	Essentiality score	Chromosome 1	Chromosome 2	Chromosome 3	Plasmid 1	Plasmid 2	Total
Conditionally-essential genes	1	18	5	0	0	0	23
Conditionally-essential domains	2	45	7	0	2	0	54
Conditionally-enriched genes	3	7	0	0	1	0	8
Conditionally-enriched domains	4	6	0	0	3	0	9
Neutral genes	5	2642	1307	788	1151	262	6150

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS00935	1	NC_021287.1	198756	199502	*gpmA*	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	Carbohydrate transport and metabolism
BRPE64_RS01655	1	NC_021287.1	349176	350741	*purH*	bifunctional purine biosynthesis protein PurH	Nucleotide transport and metabolism
BRPE64_RS01665	1	NC_021287.1	351036	352130	*dusB*	tRNA-dihydrouridine synthase	Translation, ribosomal structure and biogenesis
BRPE64_RS01850	1	NC_021287.1	393018	394703	-	hypothetical protein	-
BRPE64_RS02145	1	NC_021287.1	462950	464950	*mutL*	DNA mismatch repair protein MutL	Replication, recombination and repair
BRPE64_RS02340	1	NC_021287.1	504058	504948	*purC*	phosphoribosylaminoimidazole-succinocarboxamidesynthase	Nucleotide transport and metabolism
BRPE64_RS02975	1	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS03750	1	NC_021287.1	812442	813461	-	fructose-1,6-bisphosphatase class 1,1	Carbohydrate transport and metabolism
BRPE64_RS04795	1	NC_021287.1	1047085	1048452	-	hypothetical protein	-
BRPE64_RS04905	1	NC_021287.1	1075733	1076368	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms
BRPE64_RS06390	1	NC_021287.1	1369578	1370900	-	homoserine dehydrogenase	Amino acid transport and metabolism
BRPE64_RS06600	1	NC_021287.1	1421063	1423570	-	beta-N-acetylhexosaminidase	Carbohydrate transport and metabolism
BRPE64_RS11215	1	NC_021287.1	2430339	2431349	-	glyceraldehyde-3-phosphate dehydrogenase type I	Carbohydrate transport and metabolism
BRPE64_RS11780	1	NC_021287.1	2555788	2556585	*ung*	uracil-DNA glycosylase	Replication, recombination and repair
BRPE64_RS12160	1	NC_021287.1	2628391	2629857	*gltD*	glutamate synthase (NADH) small subunit	Amino acid transport and metabolism

BRPE64_RS12255	1	NC_021287.1	2652520	2652837	*cyaY*	protein CyaY	Inorganic ion transport and metabolism
BRPE64_RS12270	1	NC_021287.1	2654557	2655219	*yedZ*	sulfoxide reductase heme-binding subunit YedZ	Function unknown
BRPE64_RS13225	1	NC_021287.1	2839274	2842204	*gcvP*	glycine dehydrogenase	Amino acid transport and metabolism
BRPE64_RS19700	1	NC_021294.1	1219620	1221668	-	RNA polymerase sigma factor	Transcription
BRPE64_RS20030	1	NC_021294.1	1308190	1309038	-	sporulation domain protein	Function unknown
BRPE64_RS20045	1	NC_021294.1	1311424	1312224	*trpA*	tryptophan synthase alpha chain	Amino acid transport and metabolism
BRPE64_RS20065	1	NC_021294.1	1315108	1315911	*truA*	tRNA pseudouridine synthase A	Translation, ribosomal structure and biogenesis
BRPE64_RS20080	1	NC_021294.1	1319609	1320676	-	3-isopropylmalate dehydrogenase	Energy production and conversion
BRPE64_RS00275	2	NC_021287.1	59873	61018	-	homoserine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS00285	2	NC_021287.1	62059	62958	-	acetylglutamate kinase	Amino acid transport and metabolism
BRPE64_RS00855	2	NC_021287.1	179620	180909	-	C4-dicarboxylate transport protein	Energy production and conversion
BRPE64_RS00860	2	NC_021287.1	181071	182939	-	histidine kinase	Signal transduction mechanisms
BRPE64_RS00865	2	NC_021287.1	182961	184295	-	Two component sigma54 specific transcriptional regulator Fis family	Signal transduction mechanisms
BRPE64_RS01225	2	NC_021287.1	250285	251802	-	RNA polymerase sigma-54 factor	Transcription
BRPE64_RS01390	2	NC_021287.1	288271	289794	*ilvA*	L-threonine ammonia-lyase	Amino acid transport and metabolism
BRPE64_RS01395	2	NC_021287.1	290326	294432	-	FAD linked oxidase domain protein	Energy production and conversion
BRPE64_RS01660	2	NC_021287.1	350806	351039	*fis*	DNA-binding protein Fis	Transcription
BRPE64_RS01815	2	NC_021287.1	385035	385847	*proC*	pyrroline-5-carboxylate reductase	Amino acid transport and metabolism
BRPE64_RS01820	2	NC_021287.1	385861	386568	-	alanine racemase domain protein	General function prediction only
BRPE64_RS02135	2	NC_021287.1	460885	461940	*purM*	phosphoribosylformylglycinamidine cyclo-ligase	Nucleotide transport and metabolism
BRPE64_RS02140	2	NC_021287.1	462006	462953	*miaA*	tRNA dimethylallyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS02325	2	NC_021287.1	499591	500787	*pgk*	phosphoglycerate kinase	Carbohydrate transport and metabolism
BRPE64_RS02350	2	NC_021287.1	505612	506799	*purK*	phosphoribosylaminoimidazole carboxylase ATPase subunit	Nucleotide transport and metabolism

BRPE64_RS02980	2	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS02985	2	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS02990	2	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS02995	2	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS03000	2	NC_021287.1	655612	656925	*cysN*	sulfate adenylyltransferase large subunit	Inorganic ion transport and metabolism
BRPE64_RS03005	2	NC_021287.1	656943	657698	-	uroporphyrin-III C-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS03245	2	NC_021287.1	707442	708845	*argH*	argininosuccinate lyase	Amino acid transport and metabolism
BRPE64_RS04790	2	NC_021287.1	1045695	1047077	*argA*	amino-acid acetyltransferase	Amino acid transport and metabolism
BRPE64_RS04910	2	NC_021287.1	1076365	1078893	-	multi-sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS05810	2	NC_021287.1	1260186	1262603	-	phosphoenolpyruvate synthase	Carbohydrate transport and metabolism
BRPE64_RS06385	2	NC_021287.1	1368315	1369553	-	aminotransferase AlaT	Amino acid transport and metabolism
BRPE64_RS06595	2	NC_021287.1	1416832	1420869	*purL*	phosphoribosylformylglycinamidine synthase	Nucleotide transport and metabolism
BRPE64_RS08840	2	NC_021287.1	1916590	1917606	*ilvC*	ketol-acid reductoisomerase	Amino acid transport and metabolism
BRPE64_RS08845	2	NC_021287.1	1917676	1918167	-	acetolactate synthase small subunit	Amino acid transport and metabolism
BRPE64_RS08850	2	NC_021287.1	1918274	1920037	-	acetolactate synthase large subunit biosynthetic type	Amino acid transport and metabolism
BRPE64_RS10830	2	NC_021287.1	2355573	2356502	-	ornithine carbamoyltransferase	Amino acid transport and metabolism
BRPE64_RS10835	2	NC_021287.1	2356655	2357884	*argG*	argininosuccinate synthase	Amino acid transport and metabolism
BRPE64_RS11220	2	NC_021287.1	2431398	2433449	-	transketolase 1	Carbohydrate transport and metabolism
BRPE64_RS11265	2	NC_021287.1	2442927	2445221	*maeB*	malate dehydrogenase (Oxaloacetate-decarboxylating) (NADP(+)) Phosphate acetyltransferase	Energy production and conversion
BRPE64_RS11755	2	NC_021287.1	2551158	2552654	*trpE*	anthranilate synthase component I	Amino acid transport and metabolism
BRPE64_RS11760	2	NC_021287.1	2552667	2553254	-	glutamine amidotransferase of anthranilate synthase	Amino acid transport and metabolism
BRPE64_RS11765	2	NC_021287.1	2553259	2554293	*trpD*	anthranilate phosphoribosyltransferase	Amino acid transport and metabolism

BRPE64_RS11770	2	NC_021287.1	2554305	2555090	*trpC*	indole-3-glycerol phosphate synthase	Amino acid transport and metabolism
BRPE64_RS11775	2	NC_021287.1	2555112	2555759	-	adenylate cyclase	Function unknown
BRPE64_RS12165	2	NC_021287.1	2629958	2634661	*gltB*	glutamate synthase	Amino acid transport and metabolism
BRPE64_RS12260	2	NC_021287.1	2652894	2653202	-	hypothetical protein	-
BRPE64_RS12265	2	NC_021287.1	2653248	2654510	*lysA*	diaminopimelate decarboxylase	Amino acid transport and metabolism
BRPE64_RS12985	2	NC_021287.1	2789444	2790283	*metF*	methylenetetrahydrofolate reductase	Amino acid transport and metabolism
BRPE64_RS12990	2	NC_021287.1	2790324	2790677	-	hypothetical protein	Function unknown
BRPE64_RS12995	2	NC_021287.1	2790753	2792174	-	adenosylhomocysteinase	Coenzyme transport and metabolism
BRPE64_RS20020	2	NC_021294.1	1305948	1307495	*purF*	amidophosphoribosyltransferase	Nucleotide transport and metabolism
BRPE64_RS20025	2	NC_021294.1	1307649	1308182	*cvpA*	putative bacteriocin production related protein	General function prediction only
BRPE64_RS20050	2	NC_021294.1	1312310	1313188	-	DNA methylase N-4/N-6 domain protein	Replication, recombination and repair
BRPE64_RS20055	2	NC_021294.1	1313200	1314393	*trpB*	tryptophan synthase beta chain	Amino acid transport and metabolism
BRPE64_RS20060	2	NC_021294.1	1314443	1315111	-	N-(5'-phosphoribosyl)anthranilate isomerase	Amino acid transport and metabolism
BRPE64_RS20085	2	NC_021294.1	1320759	1321409	-	3-isopropylmalate dehydratase small subunit	Amino acid transport and metabolism
BRPE64_RS20090	2	NC_021294.1	1321440	1322849	-	3-isopropylmalate dehydratase large subunit	Amino acid transport and metabolism
BRPE64_RS28705	2	NC_021289.1	888916	889830	*metR*	transcriptional regulator LysR family	Transcription
BRPE64_RS28710	2	NC_021289.1	889942	892233	*metE*	5-methyltetrahydropteroyl triglutamate--homocysteine methyltransferase	Amino acid transport and metabolism
BRPE64_RS03230	3	NC_021287.1	703248	705554	-	ornithine decarboxylase	Amino acid transport and metabolism
BRPE64_RS05050	3	NC_021287.1	1108524	1109000	-	bacterioferritin	Inorganic ion transport and metabolism
BRPE64_RS31880	3	NC_021287.1	2218036	2222010	-	type VI secretion system Vgr family protein	Function unknown
BRPE64_RS10275	3	NC_021287.1	2223364	2223789	-	hypothetical protein	-
BRPE64_RS12010	3	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12280	3	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein

							turnover, chaperones
BRPE64_RS12305	3	NC_021287.1	2662968	2664854	-	thiol disulfide interchange protein DsbD	Energy production and conversion
BRPE64_RS24700	3	NC_021289.1	2559	3896	-	hypothetical protein	-
BRPE64_RS05045	4	NC_021287.1	1107638	1108519	*murI*	glutamate racemase	Cell wall/membrane /envelope biogenesis
BRPE64_RS07530	4	NC_021287.1	1633539	1635431	-	probable potassium transport system protein kup	Inorganic ion transport and metabolism
BRPE64_RS10595	4	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane /envelope biogenesis
BRPE64_RS12015	4	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	4	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12285	4	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS24690	4	NC_021289.1	104	1282	-	Soj protein	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS24695	4	NC_021289.1	1397	2254	-	stage 0 sporulation protein J	Transcription
BRPE64_RS29720	4	NC_021289.1	1111956	1113017	-	squalene/phytoene synthase family protein	Lipid transport and metabolism

Annexe 4: List of essential genes shared between *Burkholderia* species.

		Number of genes
Chromosome 1	NC_021287.1	148
Chromosome 2	NC_021294.1	1
Chromosome 3	NC_021288.1	1
Plasmid 1	NC_021289.1	1
Plasmid 2	NC_021295.1	0
	Total	151

Gene tag	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS00005	NC_021287.1	354	1952	*dnaA*	chromosomal replication initiator protein DnaA	Replication, recombination and repair
BRPE64_RS00010	NC_021287.1	2204	3307	*dnaN*	DNA polymerase III subunit beta	Replication, recombination and repair
BRPE64_RS00680	NC_021287.1	146373	148157	*argS*	arginine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS00765	NC_021287.1	164340	165161	-	type III pantothenate kinase	Transcription
BRPE64_RS00770	NC_021287.1	165158	166060	-	biotin--acetyl-CoA-carboxylase ligase	Coenzyme transport and metabolism
BRPE64_RS00850	NC_021287.1	178177	179175	*lipA*	lipoyl synthase	Coenzyme transport and metabolism
BRPE64_RS01125	NC_021287.1	233678	234985	*ftsY*	signal recognition particle receptor FtsY	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS01160	NC_021287.1	239406	240359	-	ribose-phosphate pyrophosphokinase	Amino acid transport and metabolism
BRPE64_RS01210	NC_021287.1	247667	248635	-	HPr kinase/phosphorylase	Signal transduction mechanisms
BRPE64_RS01230	NC_021287.1	251966	252748	-	ABC transporter related protein	General function prediction only
BRPE64_RS01235	NC_021287.1	252745	253440	*yhbN*	hypothetical protein	Function unknown
BRPE64_RS01295	NC_021287.1	266689	267222	*ssb*	single-stranded DNA-binding protein	Replication, recombination and repair
BRPE64_RS01425	NC_021287.1	298099	299673	-	probable ubiquinone biosynthesis protein UbiB	General function prediction only
BRPE64_RS01445	NC_021287.1	301778	303580	*aspS*	aspartate--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS01570	NC_021287.1	332662	333090	*rplM*	50S ribosomal protein L13	Translation, ribosomal structure and biogenesis
BRPE64_RS01670	NC_021287.1	352350	353573	*ubiH*	ubiquinone biosynthesis hydroxylase UbiH/UbiF/VisC/COQ6	Energy production and conversion
BRPE64_RS01765	NC_021287.1	371600	373666	*secD*	protein translocase subunit SecD	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS01885	NC_021287.1	407731	408627	-	probable inorganic polyphosphate/ATP-NAD kinase	Carbohydrate transport and metabolism
BRPE64_RS01975	NC_021287.1	426377	428476	*glyS*	glycine--tRNA ligase beta subunit	Translation, ribosomal structure and biogenesis
BRPE64_RS02355	NC_021287.1	506834	507865	-	Sua5/YciO/YrdC/YwlC family protein	Translation, ribosomal structure and biogenesis

BRPE64_RS02750	NC_021287.1	605267	606478	-	phosphopantothenoylcysteine decarboxylase/phosphopantothenate--cysteine ligase	Coenzyme transport and metabolism
BRPE64_RS03015	NC_021287.1	658245	659393	-	permease YjgP/YjgQ family protein	General function prediction only
BRPE64_RS03020	NC_021287.1	659398	660489	-	permease YjgP/YjgQ family protein	General function prediction only
BRPE64_RS03225	NC_021287.1	702561	703130	*dcd*	deoxycytidine triphosphate deaminase	Nucleotide transport and metabolism
BRPE64_RS03270	NC_021287.1	714126	715133	*hemC*	porphobilinogen deaminase	Coenzyme transport and metabolism
BRPE64_RS03350	NC_021287.1	733005	734303	*serS*	serine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS03685	NC_021287.1	801790	802464	*gmk*	guanylate kinase	Nucleotide transport and metabolism
BRPE64_RS03940	NC_021287.1	855068	856351	*hemL*	glutamate-1-semialdehyde 2,1-aminomutase	Coenzyme transport and metabolism
BRPE64_RS04480	NC_021287.1	971297	974824	*dnaE*	DNA polymerase III alpha subunit	Replication, recombination and repair
BRPE64_RS04505	NC_021287.1	979335	981119	*msbA*	lipid A ABC exporter fused ATPase and inner membrane subunits MsbA	Defense mechanisms
BRPE64_RS04540	NC_021287.1	986250	987209	*hemF*	coproporphyrinogen-III oxidase aerobic	Coenzyme transport and metabolism
BRPE64_RS04835	NC_021287.1	1057904	1059319	*glnA*	glutamine synthetase	Amino acid transport and metabolism
BRPE64_RS04920	NC_021287.1	1081962	1083599	*aceF*	pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase	Energy production and conversion
BRPE64_RS05615	NC_021287.1	1212426	1213733	-	major facilitator superfamily (MFS) transporter	Amino acid transport and metabolism
BRPE64_RS05685	NC_021287.1	1228362	1231877	-	chromosome partition protein Smc	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS05730	NC_021287.1	1243756	1244637	*tsf*	elongation factor Ts	Translation, ribosomal structure and biogenesis
BRPE64_RS05735	NC_021287.1	1244875	1245588	*pyrH*	uridylate kinase	Nucleotide transport and metabolism
BRPE64_RS05745	NC_021287.1	1246324	1247106	-	isoprenyl transferase	Lipid transport and metabolism
BRPE64_RS05750	NC_021287.1	1247100	1247912	-	phosphatidate cytidylyltransferase	Lipid transport and metabolism
BRPE64_RS05755	NC_021287.1	1247931	1249148	*dxr*	1-deoxy-D-xylulose 5-phosphate reductoisomerase	Lipid transport and metabolism
BRPE64_RS05765	NC_021287.1	1250619	1252931	-	outer membrane protein assembly factor BamA	Cell wall/membrane/envelope biogenesis
BRPE64_RS05770	NC_021287.1	1253014	1253559	-	outer membrane chaperone Skp	Cell wall/membrane/envelope biogenesis
BRPE64_RS05775	NC_021287.1	1253590	1254681	*lpxD*	UDP-3-O-acylglucosamine N-acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS05785	NC_021287.1	1255396	1256184	*lpxA*	acyl-[acyl-carrier-protein]--UDP-N- acetylglucosamine O-acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS05790	NC_021287.1	1256197	1257363	*lpxB*	lipid-A-disaccharide synthase	Cell wall/membrane/envelope biogenesis
BRPE64_RS05845	NC_021287.1	1266671	1268131	*guaB*	inosine-5'-monophosphate dehydrogenase	Nucleotide transport and metabolism

BRPE64_RS05855	NC_021287.1	1269219	1270802	*guaA*	GMP synthase	Nucleotide transport and metabolism
BRPE64_RS06310	NC_021287.1	1351469	1352857	*dnaB*	replicative DNA helicase	Replication, recombination and repair
BRPE64_RS32260	NC_021287.1	1359698	1361653	-	hypothetical protein	-
BRPE64_RS06360	NC_021287.1	1362633	1363670	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS06365	NC_021287.1	1363695	1364864	*yfbE*	DegT/DnrJ/EryC1/StrS aminotransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS06375	NC_021287.1	1365413	1367086	-	glycosyl transferase family 39	Cell wall/membrane/envelope biogenesis
BRPE64_RS06390	NC_021287.1	1369578	1370900	-	homoserine dehydrogenase	Amino acid transport and metabolism
BRPE64_RS06395	NC_021287.1	1370916	1372361	*thrC*	threonine synthase	Amino acid transport and metabolism
BRPE64_RS06455	NC_021287.1	1387959	1388693	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	Lipid transport and metabolism
BRPE64_RS06690	NC_021287.1	1440297	1440917	*tmk*	thymidylate kinase	Nucleotide transport and metabolism
BRPE64_RS07590	NC_021287.1	1647298	1648647	*hisS*	histidine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS07595	NC_021287.1	1648654	1649970	*ispG*	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	Lipid transport and metabolism
BRPE64_RS07610	NC_021287.1	1652732	1653157	*ndk*	nucleoside diphosphate kinase	Nucleotide transport and metabolism
BRPE64_RS07695	NC_021287.1	1668399	1670768	*dnaX*	DNA polymerase III subunits gamma and tau	Replication, recombination and repair
BRPE64_RS07705	NC_021287.1	1672020	1673285	*rho*	transcription termination factor Rho	Transcription
BRPE64_RS07980	NC_021287.1	1738934	1741792	*sucA*	2-oxoglutarate dehydrogenase E1 subunit	Energy production and conversion
BRPE64_RS08020	NC_021287.1	1750775	1753765	*infB*	translation initiation factor IF-2	Translation, ribosomal structure and biogenesis
BRPE64_RS08365	NC_021287.1	1816686	1818158	*nuoN*	NADH-quinone oxidoreductase subunit N	Energy production and conversion
BRPE64_RS08370	NC_021287.1	1818184	1819683	*nuoM*	NADH dehydrogenase I chain M	Energy production and conversion
BRPE64_RS08375	NC_021287.1	1819697	1821766	*nuoL*	proton-translocating NADH-quinone oxidoreductase chain L	Energy production and conversion
BRPE64_RS08390	NC_021287.1	1822972	1823460	*nuoI*	NADH-quinone oxidoreductase subunit I	Energy production and conversion
BRPE64_RS08395	NC_021287.1	1823485	1824549	*nuoH*	NADH-quinone oxidoreductase subunit H	Energy production and conversion
BRPE64_RS08400	NC_021287.1	1824552	1826891	*nuoG*	NADH dehydrogenase subunit G	Energy production and conversion
BRPE64_RS08405	NC_021287.1	1826944	1828266	*nuoF*	NADH dehydrogenase subunit F	Energy production and conversion
BRPE64_RS08410	NC_021287.1	1828263	1828748	*nuoE*	NADH dehydrogenase I chain E	Energy production and conversion
BRPE64_RS08415	NC_021287.1	1828902	1830155	-	NADH-quinone oxidoreductase subunit D	Energy production and conversion
BRPE64_RS08420	NC_021287.1	1830165	1830767	*nuoC*	NADH-quinone oxidoreductase subunit C	Energy production and conversion
BRPE64_RS08425	NC_021287.1	1830801	1831280	*nuoB*	NADH-quinone oxidoreductase subunit B	Energy production and conversion
BRPE64_RS08585	NC_021287.1	1862565	1864004	*tilS*	tRNA(Ile)-lysidine synthase	Cell cycle control, cell division, chromosome partitioning

BRPE64_RS08600	NC_021287.1	1866261	1867658	*cysS*	cysteine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS08620	NC_021287.1	1870186	1870995	*lpxH*	UDP-2,3-diacylglucosamine hydrolase	Function unknown
BRPE64_RS08635	NC_021287.1	1873063	1873866	*suhB*	inositol monophosphatase	Carbohydrate transport and metabolism
BRPE64_RS08755	NC_021287.1	1897540	1899234	*pyrG*	CTP synthase	Nucleotide transport and metabolism
BRPE64_RS08780	NC_021287.1	1904361	1905614	-	lipoprotein releasing system transmembrane protein LolC/E family	Cell wall/membrane/envelope biogenesis
BRPE64_RS08815	NC_021287.1	1911812	1913338	-	lysine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS08835	NC_021287.1	1915830	1916474	-	phosphatidylserine decarboxylase proenzyme	Lipid transport and metabolism
BRPE64_RS09290	NC_021287.1	2013318	2014679	*glmM*	phosphoglucosamine mutase	Carbohydrate transport and metabolism
BRPE64_RS09300	NC_021287.1	2015824	2017710	-	ATP-dependent zinc metalloprotease FtsH	Posttranslational modification, protein turnover, chaperones
BRPE64_RS09360	NC_021287.1	2030738	2031475	*dnaQ*	DNA polymerase III epsilon subunit	Replication, recombination and repair
BRPE64_RS09775	NC_021287.1	2111724	2112695	*thyA*	thymidylate synthase	Nucleotide transport and metabolism
BRPE64_RS09845	NC_021287.1	2124175	2124945	*trmD*	tRNA (guanine-N(1)-)-methyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS09885	NC_021287.1	2131910	2132845	-	electron transfer flavoprotein alpha subunit	Energy production and conversion
BRPE64_RS09890	NC_021287.1	2132861	2133610	-	electron transfer flavoprotein alpha/beta-subunit	Energy production and conversion
BRPE64_RS09950	NC_021287.1	2144944	2146119	-	tetratricopeptide TPR_2 repeat protein	Carbohydrate transport and metabolism
BRPE64_RS09965	NC_021287.1	2147123	2148835	*rpsA*	30S ribosomal protein S1	Translation, ribosomal structure and biogenesis
BRPE64_RS09970	NC_021287.1	2149002	2149682	*cmk*	cytidylate kinase	Nucleotide transport and metabolism
BRPE64_RS10010	NC_021287.1	2159473	2160171	*ubiG*	3-demethylubiquinone-9,3-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS10305	NC_021287.1	2229557	2230957	-	phosphomannomutase	Carbohydrate transport and metabolism
BRPE64_RS10665	NC_021287.1	2317125	2318765	-	60 kDa chaperonin	Posttranslational modification, protein turnover, chaperones
BRPE64_RS10785	NC_021287.1	2347394	2348410	*lpxK*	tetraacyldisaccharide 4'-kinase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10800	NC_021287.1	2349697	2350359	*adk*	adenylate kinase	Nucleotide transport and metabolism
BRPE64_RS10850	NC_021287.1	2359737	2360363	-	glycerol-3-phosphate acyltransferase	Function unknown
BRPE64_RS11095	NC_021287.1	2404009	2406399	-	LPS-assembly protein LptD	Cell wall/membrane/envelope biogenesis
BRPE64_RS11110	NC_021287.1	2408469	2409857	*purB*	adenylosuccinate lyase	Nucleotide transport and metabolism
BRPE64_RS11160	NC_021287.1	2419503	2422097	*leuS*	leucine--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS11185	NC_021287.1	2425336	2425764	*fur*	ferric uptake regulator Fur family	Inorganic ion transport and metabolism
BRPE64_RS11270	NC_021287.1	2445477	2446475	*thiL*	thiamine-monophosphate kinase	Coenzyme transport and metabolism
BRPE64_RS11415	NC_021287.1	2475066	2476313	*nrdB*	ribonucleoside-diphosphate reductase subunit beta	Nucleotide transport and metabolism

BRPE64_RS11435	NC_021287.1	2478941	2481976	*nrdA*	ribonucleoside-diphosphate reductase	Nucleotide transport and metabolism
BRPE64_RS11455	NC_021287.1	2484583	2485950	*ffh*	signal recognition particle protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11470	NC_021287.1	2487290	2489026	*proS*	proline--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS11505	NC_021287.1	2494092	2495084	*ispB*	octylprenyl-diphosphate synthase	Coenzyme transport and metabolism
BRPE64_RS11575	NC_021287.1	2508384	2509301	*lpxC*	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11585	NC_021287.1	2510356	2511555	*ftsZ*	cell division protein FtsZ	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS11590	NC_021287.1	2511680	2512912	*ftsA*	cell division protein ftsA	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS11595	NC_021287.1	2512939	2513691	*ftsQ*	cell division protein FtsQ	Cell wall/membrane/envelope biogenesis
BRPE64_RS11600	NC_021287.1	2513755	2514696	-	D-alanine--D-alanine ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11610	NC_021287.1	2516122	2517255	*murG*	UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11615	NC_021287.1	2517252	2518547	*ftsW*	lipid II flippase FtsW	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS11620	NC_021287.1	2518544	2520055	*murD*	UDP-N-acetylmuramoylalanine--D-glutamate ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11625	NC_021287.1	2520117	2521286	*mraY*	phospho-N-acetylmuramoyl-pentapeptide-transferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11630	NC_021287.1	2521310	2522740	*murF*	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11635	NC_021287.1	2522737	2524281	*murE*	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2, 6-diaminopimelate ligase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11980	NC_021287.1	2598727	2599338	*sspA*	glutathione S-transferase domain protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12320	NC_021287.1	2665847	2666824	*rpoA*	DNA-directed RNA polymerase subunit alpha	Transcription
BRPE64_RS12325	NC_021287.1	2666971	2667594	*rpsD*	30S ribosomal protein S4	Translation, ribosomal structure and biogenesis
BRPE64_RS12350	NC_021287.1	2668958	2670304	*secY*	protein translocase subunit SecY	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12355	NC_021287.1	2670346	2670780	*rplO*	50S ribosomal protein L15	Translation, ribosomal structure and biogenesis
BRPE64_RS12365	NC_021287.1	2671008	2671526	*rpsE*	30S ribosomal protein S5	Translation, ribosomal structure and biogenesis
BRPE64_RS12375	NC_021287.1	2671919	2672452	*rplF*	50S ribosomal protein L6	Translation, ribosomal structure and biogenesis
BRPE64_RS12390	NC_021287.1	2673194	2673733	*rplE*	50S ribosomal protein L5	Translation, ribosomal structure and biogenesis
BRPE64_RS12420	NC_021287.1	2675627	2676421	*rpsC*	30S ribosomal protein S3	Translation, ribosomal structure and biogenesis

BRPE64_RS12435	NC_021287.1	2677061	2677888	*rplB*	50S ribosomal protein L2	Translation, ribosomal structure and biogenesis
BRPE64_RS12445	NC_021287.1	2678202	2678822	*rplD*	50S ribosomal protein L4	Translation, ribosomal structure and biogenesis
BRPE64_RS12450	NC_021287.1	2678822	2679481	*rplC*	50S ribosomal protein L3	Translation, ribosomal structure and biogenesis
BRPE64_RS12470	NC_021287.1	2683615	2684085	*rpsG*	30S ribosomal protein S7	Translation, ribosomal structure and biogenesis
BRPE64_RS12485	NC_021287.1	2686912	2691153	*rpoC*	DNA-directed RNA polymerase subunit beta'	Transcription
BRPE64_RS12490	NC_021287.1	2691175	2695281	*rpoB*	DNA-directed RNA polymerase subunit beta	Transcription
BRPE64_RS12500	NC_021287.1	2696108	2696605	*rplJ*	50S ribosomal protein L10	Translation, ribosomal structure and biogenesis
BRPE64_RS12510	NC_021287.1	2697608	2698039	*rplK*	50S ribosomal protein L11	Translation, ribosomal structure and biogenesis
BRPE64_RS12515	NC_021287.1	2698186	2698743	*nusG*	transcription antitermination protein nusG	Transcription
BRPE64_RS12840	NC_021287.1	2759914	2761119	-	hypothetical protein	Function unknown
BRPE64_RS13315	NC_021287.1	2864418	2866658	*priA*	primosomal protein N'	Replication, recombination and repair
BRPE64_RS13325	NC_021287.1	2867118	2868203	*hemE*	uroporphyrinogen decarboxylase	Coenzyme transport and metabolism
BRPE64_RS13350	NC_021287.1	2873326	2874720	*atpD*	ATP synthase subunit beta 2	Energy production and conversion
BRPE64_RS13355	NC_021287.1	2874765	2875649	*atpG*	ATP synthase gamma chain	Energy production and conversion
BRPE64_RS13360	NC_021287.1	2875722	2877263	*atpA*	ATP synthase subunit alpha 1	Energy production and conversion
BRPE64_RS13380	NC_021287.1	2878833	2879684	*atpB*	ATP synthase subunit a	Energy production and conversion
BRPE64_RS13395	NC_021287.1	2881765	2882667	-	ParB-like partition protein	Transcription
BRPE64_RS13695	NC_021287.1	2948069	2949256	*metK*	S-adenosylmethionine synthase	Coenzyme transport and metabolism
BRPE64_RS13730	NC_021287.1	2954924	2955694	*fpr*	oxidoreductase FAD-binding domain protein	Energy production and conversion
BRPE64_RS14020	NC_021287.1	3010559	3012229	*yidC*	membrane protein insertase YidC	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS20105	NC_021294.1	1323878	1325179	*gltA*	citrate synthase	Energy production and conversion
BRPE64_RS21510	NC_021288.1	165308	167083	-	succinate dehydrogenase flavoprotein subunit	Energy production and conversion
BRPE64_RS29805	NC_021289.1	1134732	1136129	-	nucleotide sugar dehydrogenase	Cell wall/membrane/envelope biogenesis

Annexe 5: List of essential genes specific to *B. insecticola*.

		Number of genes
Chromosome 1	NC_021287.1	148
Chromosome 2	NC_021294.1	138
Chromosome 3	NC_021288.1	108
Plasmid 1	NC_021289.1	311
Plasmid 2	NC_021295.1	10
		715

Gene tag	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS00170	NC_021287.1	37898	38614	-	Response regulator containing CheY-like receiver	Signal transduction mechanisms
BRPE64_RS00215	NC_021287.1	46251	48560	*mrdA*	penicillin-binding protein 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS00220	NC_021287.1	48652	49164	*mreD*	Rod shape-determining protein MreD	Cell wall/membrane/envelope biogenesis
BRPE64_RS00245	NC_021287.1	53805	55274	-	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	Translation, ribosomal structure and biogenesis
BRPE64_RS00295	NC_021287.1	63461	64759	-	integral membrane sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS00300	NC_021287.1	64804	65346	-	two component transcriptional regulator Fis family	Transcription
BRPE64_RS00930	NC_021287.1	196900	198486	-	carboxyl-terminal protease	Cell wall/membrane/envelope biogenesis
BRPE64_RS01155	NC_021287.1	238634	239254	*rplY*	50S ribosomal protein L25	Translation, ribosomal structure and biogenesis
BRPE64_RS01185	NC_021287.1	242190	244049	-	TPR repeat-containing protein	General function prediction only
BRPE64_RS01245	NC_021287.1	254076	254612	-	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase YrbI family	General function prediction only
BRPE64_RS01420	NC_021287.1	297350	298078	-	hypothetical protein	Function unknown
BRPE64_RS01915	NC_021287.1	413093	415048	*dnaK*	chaperone protein DnaK	Posttranslational modification, protein turnover, chaperones
BRPE64_RS01980	NC_021287.1	428620	429522	-	hypothetical protein	-
BRPE64_RS02005	NC_021287.1	433804	434925	-	PhoH family protein	Signal transduction mechanisms
BRPE64_RS02125	NC_021287.1	459118	459804	-	HAD-superfamily subfamily IB hydrolase TIGR01490	Amino acid transport and metabolism
BRPE64_RS02340	NC_021287.1	504058	504948	*purC*	phosphoribosylaminoimidazole-succinocarboxamidesynthase	Nucleotide transport and metabolism

BRPE64_RS02490	NC_021287.1	541419	542525	-	S-(Hydroxymethyl)glutathione dehydrogenase/class III alcohol dehydrogenase	Energy production and conversion
BRPE64_RS02740	NC_021287.1	603803	604249	*dut*	deoxyuridine 5'-triphosphate nucleotidohydrolase	Nucleotide transport and metabolism
BRPE64_RS02785	NC_021287.1	614758	615954	-	fatty acid desaturase	Lipid transport and metabolism
BRPE64_RS03160	NC_021287.1	684859	685044	-	hypothetical protein	-
BRPE64_RS03230	NC_021287.1	703248	705554	-	ornithine decarboxylase	Amino acid transport and metabolism
BRPE64_RS03680	NC_021287.1	800827	801780	-	hypothetical protein	Function unknown
BRPE64_RS03720	NC_021287.1	806860	808017	-	outer membrane porin protein 32	Cell wall/membrane/envelope biogenesis
BRPE64_RS03900	NC_021287.1	846241	847866	-	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Coenzyme transport and metabolism
BRPE64_RS03925	NC_021287.1	851955	853118	-	bifunctional riboflavin biosynthesis protein RibBA	Coenzyme transport and metabolism
BRPE64_RS04055	NC_021287.1	882440	884203	-	binding-protein-dependent transport systems inner membrane component	Inorganic ion transport and metabolism
BRPE64_RS04060	NC_021287.1	884225	885571	-	ABC nitrate/sulfonate/bicarbonate family transporter ATPase subunit	Inorganic ion transport and metabolism
BRPE64_RS04240	NC_021287.1	916210	918189	*parE*	DNA topoisomerase	Replication, recombination and repair
BRPE64_RS04360	NC_021287.1	943748	944701	-	transaldolase	Carbohydrate transport and metabolism
BRPE64_RS04545	NC_021287.1	987358	988647	*purD*	phosphoribosylamine--glycine ligase	Nucleotide transport and metabolism
BRPE64_RS04550	NC_021287.1	988916	989644	-	probable transcriptional regulatory protein Bphy_2064	Function unknown
BRPE64_RS04910	NC_021287.1	1076365	1078893	-	multi-sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS04915	NC_021287.1	1079186	1081882	*aceE*	pyruvate dehydrogenase E1 component	Energy production and conversion
BRPE64_RS05045	NC_021287.1	1107638	1108519	*murI*	glutamate racemase	Cell wall/membrane/envelope biogenesis
BRPE64_RS05350	NC_021287.1	1164880	1166553	-	electron-transferring-flavoproteindehydrogenase	Energy production and conversion
BRPE64_RS05415	NC_021287.1	1177435	1177794	-	hypothetical protein	-
BRPE64_RS05585	NC_021287.1	1207566	1208057	*rimI*	ribosomal-protein-alanine acetyltransferase	General function prediction only
BRPE64_RS05695	NC_021287.1	1233523	1235595	*ligA*	DNA ligase	Replication, recombination and repair
BRPE64_RS05900	NC_021287.1	1279823	1280215	-	DnaJ-like subfamily C member 28 conserved domain protein	Cell motility

BRPE64_RS06385	NC_021287.1	1368315	1369553	-	aminotransferase AlaT	Amino acid transport and metabolism
BRPE64_RS07105	NC_021287.1	1535284	1535808	-	phenylacetic acid degradation protein PaaD	General function prediction only
BRPE64_RS07110	NC_021287.1	1535809	1536618	-	phenylacetate-CoA oxygenase PaaI subunit	Function unknown
BRPE64_RS07115	NC_021287.1	1536632	1536916	-	phenylacetate-CoA oxygenase PaaH subunit	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS07530	NC_021287.1	1633539	1635431	-	probable potassium transport system protein kup	Inorganic ion transport and metabolism
BRPE64_RS07545	NC_021287.1	1637875	1639026	-	ATP phosphoribosyltransferase regulatory subunit	Amino acid transport and metabolism
BRPE64_RS07550	NC_021287.1	1639153	1639344	-	hypothetical protein	Function unknown
BRPE64_RS07555	NC_021287.1	1639396	1640298	*hflC*	band 7 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS07560	NC_021287.1	1640310	1641710	*hflK*	protease FtsH subunit HflK	Posttranslational modification, protein turnover, chaperones
BRPE64_RS07570	NC_021287.1	1643126	1643362	-	hypothetical protein	-
BRPE64_RS07585	NC_021287.1	1646568	1647197	-	hypothetical protein	Function unknown
BRPE64_RS07600	NC_021287.1	1650105	1651253	-	transcriptional regulator XRE family	Function unknown
BRPE64_RS07605	NC_021287.1	1651435	1652577	-	dual-specificity RNA methyltransferase RlmN	General function prediction only
BRPE64_RS07970	NC_021287.1	1735999	1737438	-	dihydrolipoyl dehydrogenase	Energy production and conversion
BRPE64_RS07975	NC_021287.1	1737536	1738837	*sucB*	2-oxoglutarate dehydrogenase E2 subunit dihydrolipoamide succinyltransferase	Energy production and conversion
BRPE64_RS07985	NC_021287.1	1742123	1743949	-	GTP-binding protein TypA	Signal transduction mechanisms
BRPE64_RS08015	NC_021287.1	1750266	1750679	*rbfA*	ribosome-binding factor A	Translation, ribosomal structure and biogenesis
BRPE64_RS08025	NC_021287.1	1753859	1755334	*nusA*	NusA antitermination factor	Transcription
BRPE64_RS08040	NC_021287.1	1758224	1759231	-	chromosome segregation and condensation protein ScpB	Transcription
BRPE64_RS08120	NC_021287.1	1768446	1769855	*gltX*	glutamate--tRNA ligase	Translation, ribosomal structure and biogenesis
BRPE64_RS08155	NC_021287.1	1776307	1776648	-	ferredoxin 2Fe-2S type ISC system	Energy production and conversion
BRPE64_RS08160	NC_021287.1	1776676	1778541	*hscA*	chaperone protein HscA homolog	Posttranslational modification, protein turnover, chaperones
BRPE64_RS08430	NC_021287.1	1831341	1831700	-	hypothetical protein	-
BRPE64_RS08605	NC_021287.1	1868146	1868946	-	tetratricopeptide TPR_2 repeat protein	General function prediction only

BRPE64_RS08700	NC_021287.1	1887739	1888404	-	peptidase M50	General function prediction only
BRPE64_RS08775	NC_021287.1	1903655	1904368	*lolD*	lipoprotein releasing system ATP-binding protein	Defense mechanisms
BRPE64_RS08805	NC_021287.1	1910619	1910702	-	hypothetical protein	-
BRPE64_RS09030	NC_021287.1	1959748	1960629	-	UTP-glucose-1-phosphate uridylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS09305	NC_021287.1	2017896	2018558	-	hypothetical protein	-
BRPE64_RS09445	NC_021287.1	2045380	2046540	-	hypothetical protein	Function unknown
BRPE64_RS09470	NC_021287.1	2050684	2051088	*acpS*	holo-[acyl-carrier-protein] synthase	Lipid transport and metabolism
BRPE64_RS09480	NC_021287.1	2051870	2052772	*recO*	DNA repair protein RecO	Replication, recombination and repair
BRPE64_RS09490	NC_021287.1	2053861	2054868	*rnc*	ribonuclease 3	Transcription
BRPE64_RS09500	NC_021287.1	2055994	2057784	*lepA*	elongation factor 4	Cell wall/membrane/envelope biogenesis
BRPE64_RS09535	NC_021287.1	2062755	2063993	*fabF*	3-oxoacyl-[acyl-carrier-protein] synthase 2	Lipid transport and metabolism
BRPE64_RS09550	NC_021287.1	2065381	2066313	*fabD*	malonyl CoA-acyl carrier protein transacylase	Lipid transport and metabolism
BRPE64_RS09570	NC_021287.1	2069139	2069759	-	hypothetical protein	General function prediction only
BRPE64_RS09605	NC_021287.1	2075337	2078696	*rne*	ribonuclease E	Translation, ribosomal structure and biogenesis
BRPE64_RS09650	NC_021287.1	2086765	2087436	-	hypothetical protein	Function unknown
BRPE64_RS09660	NC_021287.1	2089045	2089368	-	putative ferredoxin	Energy production and conversion
BRPE64_RS09815	NC_021287.1	2119272	2120540	-	putative peptidase M48 family	Posttranslational modification, protein turnover, chaperones
BRPE64_RS09945	NC_021287.1	2143458	2144861	-	nucleotide sugar dehydrogenase	Cell wall/membrane/envelope biogenesis
BRPE64_RS09955	NC_021287.1	2146162	2146455	-	hypothetical protein	Function unknown
BRPE64_RS09960	NC_021287.1	2146771	2147100	*ihfB*	integration host factor subunit beta	Replication, recombination and repair
BRPE64_RS09980	NC_021287.1	2151073	2151999	-	prephenate dehydrogenase	Amino acid transport and metabolism
BRPE64_RS09985	NC_021287.1	2152076	2153158	*pheA*	chorismate mutase	Amino acid transport and metabolism
BRPE64_RS09990	NC_021287.1	2153198	2154280	*serC*	phosphoserine aminotransferase	Amino acid transport and metabolism
BRPE64_RS10000	NC_021287.1	2155263	2157917	*gyrA*	DNA gyrase subunit A	Replication, recombination and repair
BRPE64_RS10125	NC_021287.1	2187950	2190004	-	NAD synthetase/Glutamine amidotransferase chain of NAD synthetase	Coenzyme transport and metabolism
BRPE64_RS10260	NC_021287.1	2217771	2218031	-	GP29	Function unknown

BRPE64_RS10275	NC_021287.1	2223364	2223789	-	hypothetical protein	-
BRPE64_RS10295	NC_021287.1	2227455	2228339	-	hypothetical protein	-
BRPE64_RS10555	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10775	NC_021287.1	2344707	2345288	*sodB*	superoxide dismutase	Inorganic ion transport and metabolism
BRPE64_RS10790	NC_021287.1	2348391	2348597	-	UPF0434 protein BamMC406_2464	Function unknown
BRPE64_RS10825	NC_021287.1	2354678	2355013	-	hypothetical protein	-
BRPE64_RS11010	NC_021287.1	2385657	2386904	-	serine hydroxymethyltransferase	Amino acid transport and metabolism
BRPE64_RS11090	NC_021287.1	2402511	2403872	*surA*	chaperone SurA	Posttranslational modification, protein turnover, chaperones
BRPE64_RS11155	NC_021287.1	2418901	2419503	-	rare lipoprotein B	Cell wall/membrane/envelope biogenesis
BRPE64_RS11180	NC_021287.1	2424366	2425160	-	SmpA/OmlA domain protein	Translation, ribosomal structure and biogenesis
BRPE64_RS11235	NC_021287.1	2435185	2435748	-	hypothetical protein	-
BRPE64_RS11285	NC_021287.1	2447558	2448370	-	orotidine 5'-phosphate decarboxylase	Nucleotide transport and metabolism
BRPE64_RS11375	NC_021287.1	2467141	2467614	*accB*	biotin carboxyl carrier protein	Lipid transport and metabolism
BRPE64_RS11380	NC_021287.1	2467779	2469146	*accC*	acetyl-CoA carboxylase biotin carboxylase	Lipid transport and metabolism
BRPE64_RS11400	NC_021287.1	2472207	2473145	-	PfkB domain protein	Carbohydrate transport and metabolism
BRPE64_RS11490	NC_021287.1	2491939	2493033	-	GTPase obg	General function prediction only
BRPE64_RS11495	NC_021287.1	2493208	2493471	-	hypothetical protein	-
BRPE64_RS11570	NC_021287.1	2507815	2508333	-	hypothetical protein	-
BRPE64_RS11640	NC_021287.1	2524278	2526134	-	peptidoglycan glycosyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11655	NC_021287.1	2527442	2527870	*mraZ*	protein MraZ	Function unknown
BRPE64_RS11850	NC_021287.1	2571445	2572527	*prfA*	peptide chain release factor 1	Translation, ribosomal structure and biogenesis
BRPE64_RS11855	NC_021287.1	2572524	2573360	*prmC*	release factor glutamine methyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS11860	NC_021287.1	2573447	2573758	-	glutaredoxin	Posttranslational modification, protein turnover, chaperones
BRPE64_RS11865	NC_021287.1	2573770	2574378	-	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Coenzyme transport and metabolism
BRPE64_RS12010	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12015	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport

BRPE64_RS12020	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12035	NC_021287.1	2607149	2607517	-	phosphoribosyl-ATP pyrophosphatase	Amino acid transport and metabolism
BRPE64_RS12040	NC_021287.1	2607514	2607918	*hisI*	phosphoribosyl-AMP cyclohydrolase	Amino acid transport and metabolism
BRPE64_RS12045	NC_021287.1	2607922	2608695	*hisF*	imidazole glycerol phosphate synthase subunit HisF	Amino acid transport and metabolism
BRPE64_RS12050	NC_021287.1	2608812	2609564	*hisA*	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	Amino acid transport and metabolism
BRPE64_RS12055	NC_021287.1	2609716	2610357	*hisH*	imidazole glycerol phosphate synthase subunit HisH	Amino acid transport and metabolism
BRPE64_RS12065	NC_021287.1	2611015	2611602	*hisB*	imidazoleglycerol-phosphate dehydratase	Amino acid transport and metabolism
BRPE64_RS12075	NC_021287.1	2612782	2614104	-	histidinol dehydrogenase	Amino acid transport and metabolism
BRPE64_RS12080	NC_021287.1	2614151	2614843	*hisG*	ATP phosphoribosyltransferase	Amino acid transport and metabolism
BRPE64_RS12085	NC_021287.1	2614840	2616111	*murA*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS12090	NC_021287.1	2616266	2616505	-	BolA family protein	Transcription
BRPE64_RS12095	NC_021287.1	2616523	2617278	-	ABC-2 type transporter	Defense mechanisms
BRPE64_RS12100	NC_021287.1	2617275	2618201	-	ABC multidrug efflux pump ATPase subunit	Defense mechanisms
BRPE64_RS12280	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12335	NC_021287.1	2668187	2668552	-	hypothetical protein	-
BRPE64_RS12345	NC_021287.1	2668731	2668949	-	hypothetical protein	-
BRPE64_RS12460	NC_021287.1	2680128	2681318	-	elongation factor Tu	Translation, ribosomal structure and biogenesis
BRPE64_RS12465	NC_021287.1	2681383	2683485	-	elongation factor G	Translation, ribosomal structure and biogenesis
BRPE64_RS12475	NC_021287.1	2684287	2684667	-	hypothetical protein	-
BRPE64_RS12520	NC_021287.1	2698745	2699125	-	hypothetical protein	-
BRPE64_RS12530	NC_021287.1	2699304	2700494	-	elongation factor Tu	Translation, ribosomal structure and biogenesis
BRPE64_RS12595	NC_021287.1	2713477	2713761	-	phenylacetate-CoA oxygenase PaaH subunit	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12600	NC_021287.1	2713775	2714584	-	phenylacetate-CoA oxygenase PaaI subunit	Function unknown
BRPE64_RS12655	NC_021287.1	2728109	2728795	-	orotate phosphoribosyltransferase	Nucleotide transport and metabolism

BRPE64_RS12860	NC_021287.1	2764348	2764968	-	glutathione S-transferase domain	Posttranslational modification, protein turnover, chaperones
BRPE64_RS13320	NC_021287.1	2866655	2866864	-	hypothetical protein	-
BRPE64_RS13345	NC_021287.1	2872821	2873246	*atpC*	ATP synthase epsilon chain	Energy production and conversion
BRPE64_RS13400	NC_021287.1	2882713	2883492	-	cobyrinic acid ac-diamide synthase	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS13685	NC_021287.1	2945985	2946857	*dapF*	diaminopimelate epimerase	Amino acid transport and metabolism
BRPE64_RS13690	NC_021287.1	2946908	2947792	-	lipid A biosynthesis acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS13840	NC_021287.1	2978403	2978705	-	histone family protein nucleoid-structuring protein H-NS	General function prediction only
BRPE64_RS14025	NC_021287.1	3012238	3012555	-	putative membrane protein insertion efficiency factor	Function unknown
BRPE64_RS14030	NC_021287.1	3012626	3013129	*rnpA*	ribonuclease P protein component	Translation, ribosomal structure and biogenesis
BRPE64_RS14060	NC_021294.1	2257	3612	-	hypothetical protein	-
BRPE64_RS14180	NC_021294.1	29731	31152	-	hopanoid biosynthesis associated radical SAM protein HpnJ	Energy production and conversion
BRPE64_RS14185	NC_021294.1	31175	32389	-	putative glycosyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS14500	NC_021294.1	96993	99245	-	isocitrate dehydrogenase NADP-dependent	Energy production and conversion
BRPE64_RS14735	NC_021294.1	144261	144677	-	hypothetical protein	-
BRPE64_RS14740	NC_021294.1	144693	145625	-	CbbX protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS14775	NC_021294.1	151601	152149	-	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	Energy production and conversion
BRPE64_RS14890	NC_021294.1	173576	174190	-	transcriptional regulator TetR family	Transcription
BRPE64_RS14895	NC_021294.1	174257	175489	-	major facilitator family (MFS) transporter	Amino acid transport and metabolism
BRPE64_RS14900	NC_021294.1	176063	176656	-	hypothetical protein	-
BRPE64_RS14905	NC_021294.1	176898	177641	-	hypothetical protein	-
BRPE64_RS14985	NC_021294.1	197919	198971	-	hypothetical protein	-
BRPE64_RS15085	NC_021294.1	217949	218308	-	hypothetical protein	-
BRPE64_RS15090	NC_021294.1	218493	220382	-	sensory transduction protein kinase	Signal transduction mechanisms
BRPE64_RS15470	NC_021294.1	293402	294448	-	putative patatin-like phospholipase	General function prediction only
BRPE64_RS15645	NC_021294.1	334904	335716	-	ABC transporter inner membrane subunit	Amino acid transport and metabolism
BRPE64_RS15655	NC_021294.1	338689	340572	-	hypothetical protein	-

BRPE64_RS15720	NC_021294.1	353657	354334	-	transcriptional regulator TetR family	Transcription
BRPE64_RS15765	NC_021294.1	364421	365872	-	2-methylcitrate dehydratase	General function prediction only
BRPE64_RS15770	NC_021294.1	365939	367108	-	citrate synthase	Energy production and conversion
BRPE64_RS15990	NC_021294.1	411237	412562	-	hypothetical protein	-
BRPE64_RS16030	NC_021294.1	421707	422009	-	hypothetical protein	-
BRPE64_RS16105	NC_021294.1	437221	437532	-	hypothetical protein	-
BRPE64_RS16110	NC_021294.1	437776	438336	-	histone family protein nucleoid-structuring protein H-NS	General function prediction only
BRPE64_RS16120	NC_021294.1	440131	440922	-	methyltransferase type 11	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS16125	NC_021294.1	440928	441707	-	cationic amino acid ABC transporter periplasmic binding protein	Amino acid transport and metabolism
BRPE64_RS16130	NC_021294.1	441906	442310	-	heat shock protein Hsp20	Posttranslational modification, protein turnover, chaperones
BRPE64_RS16135	NC_021294.1	442323	442757	-	heat shock protein Hsp20	Posttranslational modification, protein turnover, chaperones
BRPE64_RS16140	NC_021294.1	442756	442953	-	hypothetical protein	-
BRPE64_RS16145	NC_021294.1	442982	443362	-	hypothetical protein	Function unknown
BRPE64_RS16275	NC_021294.1	467100	468488	-	FAD/FMN-containing dehydrogenase	Energy production and conversion
BRPE64_RS16280	NC_021294.1	468546	468902	-	hypothetical protein	Cell motility
BRPE64_RS16285	NC_021294.1	469026	469283	-	hypothetical protein	-
BRPE64_RS16380	NC_021294.1	495864	496415	*fimA*	major type 1 subunit fimbrin	Cell motility
BRPE64_RS16385	NC_021294.1	496521	497264	-	fimbrial assembly chaperone	Cell motility
BRPE64_RS16530	NC_021294.1	528629	528976	*nirD*	nitrite reductase (NAD(P)H) small subunit	Inorganic ion transport and metabolism
BRPE64_RS16535	NC_021294.1	528961	530244	-	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Energy production and conversion
BRPE64_RS16540	NC_021294.1	530270	533005	-	molybdopterin oxidoreductase	Energy production and conversion
BRPE64_RS16545	NC_021294.1	533634	533981	-	hypothetical protein	-
BRPE64_RS16550	NC_021294.1	534107	534454	-	hypothetical protein	-
BRPE64_RS16570	NC_021294.1	537962	538675	-	Two component transcriptional regulator winged helix family	Transcription
BRPE64_RS16575	NC_021294.1	538772	540229	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS16580	NC_021294.1	540226	541461	-	efflux transporter RND family MFP subunit	Cell wall/membrane/envelope biogenesis

BRPE64_RS16585	NC_021294.1	541483	544668	-	cation/multidrug efflux pump AcrB/AcrD/AcrF family RND superfamily	Defense mechanisms
BRPE64_RS16590	NC_021294.1	544911	545177	-	hypothetical protein	-
BRPE64_RS16595	NC_021294.1	545404	545670	-	hypothetical protein	-
BRPE64_RS16600	NC_021294.1	545963	547294	-	divalent metal cation transporter MntH	Inorganic ion transport and metabolism
BRPE64_RS16785	NC_021294.1	583614	584627	-	rhamnose ABC transporter periplasmic rhamnose-binding protein	Carbohydrate transport and metabolism
BRPE64_RS16920	NC_021294.1	611015	611371	-	hypothetical protein	-
BRPE64_RS16970	NC_021294.1	622361	622846	-	hypothetical protein	-
BRPE64_RS16975	NC_021294.1	623039	623920	-	alpha/beta hydrolase fold protein	General function prediction only
BRPE64_RS16980	NC_021294.1	624333	625457	-	hypothetical protein	-
BRPE64_RS16985	NC_021294.1	625460	626296	-	hypothetical protein	-
BRPE64_RS17060	NC_021294.1	641737	641934	-	hypothetical protein	-
BRPE64_RS17130	NC_021294.1	660515	660706	-	hypothetical protein	-
BRPE64_RS17460	NC_021294.1	730334	730816	-	hypothetical protein	-
BRPE64_RS17465	NC_021294.1	730833	731351	-	hypothetical protein	-
BRPE64_RS17470	NC_021294.1	731824	732714	-	LysR family transcriptional regulator	Transcription
BRPE64_RS17475	NC_021294.1	732819	733211	-	acyloate catabolism-like protein	General function prediction only
BRPE64_RS17480	NC_021294.1	733281	734381	-	mandelate racemase/muconate lactonizing enzyme	Cell wall/membrane/envelope biogenesis
BRPE64_RS17485	NC_021294.1	734462	735220	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS17490	NC_021294.1	735306	736667	-	Bll0889 protein	Amino acid transport and metabolism
BRPE64_RS17495	NC_021294.1	736720	737091	-	hypothetical protein	General function prediction only
BRPE64_RS17500	NC_021294.1	737430	737861	-	transcriptional regulator MarR family protein	Transcription
BRPE64_RS17505	NC_021294.1	737881	738318	-	hypothetical protein	-
BRPE64_RS17610	NC_021294.1	755964	756485	-	hypothetical protein	-
BRPE64_RS17615	NC_021294.1	756526	757170	-	methylamine dehydrogenase accessory protein MauD	Energy production and conversion
BRPE64_RS17620	NC_021294.1	757167	757727	-	hypothetical protein	-
BRPE64_RS17625	NC_021294.1	757737	758894	-	aralkylamine dehydrogenase	Function unknown
BRPE64_RS17630	NC_021294.1	759132	760073	-	transcriptional regulator AraC family	Transcription
BRPE64_RS17635	NC_021294.1	760363	760902	-	4-hydroxyphenylacetate 3-monooxygenase reductase subunit	General function prediction only
BRPE64_RS17640	NC_021294.1	761136	761543	-	hypothetical protein	-

BRPE64_RS17645	NC_021294.1	761592	762116	-	putative uricase	Function unknown
BRPE64_RS17650	NC_021294.1	762258	763205	-	regulatory protein LysR	Transcription
BRPE64_RS17665	NC_021294.1	765424	765762	-	transthyretin	General function prediction only
BRPE64_RS17670	NC_021294.1	766010	766519	-	histone family protein nucleoid-structuring protein H-NS	General function prediction only
BRPE64_RS17675	NC_021294.1	766525	766929	-	hypothetical protein	-
BRPE64_RS17680	NC_021294.1	767009	767434	-	peroxiredoxin Ohr subfamily	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17690	NC_021294.1	770016	770411	-	hypothetical protein	-
BRPE64_RS17715	NC_021294.1	776231	776944	-	peptide methionine sulfoxide reductase MsrA	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17720	NC_021294.1	776908	778035	-	aminotransferase class I and II	Amino acid transport and metabolism
BRPE64_RS17725	NC_021294.1	778132	779025	-	transcriptional regulator LysR family	Transcription
BRPE64_RS17735	NC_021294.1	779535	779876	-	hypothetical protein	-
BRPE64_RS17740	NC_021294.1	779967	780860	-	2-dehydropantoate 2-reductase	Coenzyme transport and metabolism
BRPE64_RS17745	NC_021294.1	781069	781356	-	hypothetical protein	-
BRPE64_RS17750	NC_021294.1	781353	782063	-	GntR domain protein	Transcription
BRPE64_RS17775	NC_021294.1	785815	787647	-	predicted carbamoyl transferase NodU family	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17780	NC_021294.1	787653	788654	-	hypothetical protein	Amino acid transport and metabolism
BRPE64_RS17785	NC_021294.1	788708	789838	-	hypothetical protein	Nucleotide transport and metabolism
BRPE64_RS17885	NC_021294.1	809675	812734	-	cyclic nucleotide-regulated ABC bacteriocin/lantibiotic exporter	Defense mechanisms
BRPE64_RS17890	NC_021294.1	812752	813504	-	PpiC-type peptidyl-prolyl cis-trans isomerase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17895	NC_021294.1	813577	814950	-	ABC efflux pump membrane fusion protein HlyD subfamily	Cell wall/membrane/envelope biogenesis
BRPE64_RS17900	NC_021294.1	814928	817453	-	putative forkhead-associated protein	Signal transduction mechanisms
BRPE64_RS18245	NC_021294.1	888588	889997	-	cytochrome bd ubiquinol oxidase subunit I	Energy production and conversion
BRPE64_RS18250	NC_021294.1	890002	891003	-	cytochrome d ubiquinol oxidase subunit II	Energy production and conversion
BRPE64_RS18515	NC_021294.1	947411	947611	-	hypothetical protein	-
BRPE64_RS18535	NC_021294.1	951329	952354	-	ectoine utilization protein EutE	General function prediction only
BRPE64_RS18540	NC_021294.1	952359	953576	-	ectoine utilization protein EutD	Amino acid transport and metabolism
BRPE64_RS18545	NC_021294.1	953588	954598	-	ectoine utilization protein EutC	Amino acid transport and metabolism

BRPE64_RS18550	NC_021294.1	954603	955568	-	ectoine utilization protein EutB	Amino acid transport and metabolism
BRPE64_RS18555	NC_021294.1	955649	957046	-	transcriptional regulator GntR family with aminotransferase domain	Amino acid transport and metabolism
BRPE64_RS18560	NC_021294.1	957281	958135	-	ectoine/hydroxyectoine ABC transporter solute-binding protein	Amino acid transport and metabolism
BRPE64_RS18565	NC_021294.1	958216	958869	-	beta tubulin autoregulation binding site	Amino acid transport and metabolism
BRPE64_RS18575	NC_021294.1	959587	960432	-	ectoine/hydroxyectoine ABC transporter ATP-binding protein	Amino acid transport and metabolism
BRPE64_RS18580	NC_021294.1	960508	962070	-	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	General function prediction only
BRPE64_RS18585	NC_021294.1	962089	962478	-	hypothetical protein	Carbohydrate transport and metabolism
BRPE64_RS18625	NC_021294.1	972311	972841	-	transcriptional regulator AsnC family	Transcription
BRPE64_RS18675	NC_021294.1	980714	982222	-	putative ABC transporter solute-binding protein	Carbohydrate transport and metabolism
BRPE64_RS18680	NC_021294.1	982276	983013	-	transcriptional regulator GntR family	Transcription
BRPE64_RS18685	NC_021294.1	983144	984442	-	mandelate racemase/muconate lactonizing protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS18690	NC_021294.1	984430	985344	-	hypothetical protein	-
BRPE64_RS18695	NC_021294.1	985399	986427	-	HtrA2 peptidase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS18755	NC_021294.1	995570	997108	*hsdM*	N-6 DNA methylase	Defense mechanisms
BRPE64_RS18760	NC_021294.1	997513	998286	-	hypothetical protein	-
BRPE64_RS18785	NC_021294.1	1005617	1006285	*ribA*	GTP cyclohydrolase-2	Coenzyme transport and metabolism
BRPE64_RS19020	NC_021294.1	1064447	1064641	-	hypothetical protein	-
BRPE64_RS19170	NC_021294.1	1094882	1095988	-	S-(Hydroxymethyl)glutathione dehydrogenase/class III alcohol dehydrogenase	Energy production and conversion
BRPE64_RS19550	NC_021294.1	1183640	1185085	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS19645	NC_021294.1	1205963	1206874	-	peptidase M23	Cell wall/membrane/envelope biogenesis
BRPE64_RS19650	NC_021294.1	1207224	1208438	-	hypothetical protein	-
BRPE64_RS19655	NC_021294.1	1208905	1209834	-	hypothetical protein	-
BRPE64_RS19665	NC_021294.1	1213937	1214449	-	hypothetical protein	-
BRPE64_RS19705	NC_021294.1	1222154	1224037	*dnaG*	DNA primase	Replication, recombination and repair

BRPE64_RS19710	NC_021294.1	1224058	1224504	-	hypothetical protein	Function unknown
BRPE64_RS19740	NC_021294.1	1230044	1230847	-	GTP cyclohydrolase folE2	Function unknown
BRPE64_RS19755	NC_021294.1	1233867	1234202	*xseB*	exodeoxyribonuclease 7 small subunit	Replication, recombination and repair
BRPE64_RS19780	NC_021294.1	1239240	1241978	*polA*	DNA polymerase I	Replication, recombination and repair
BRPE64_RS19970	NC_021294.1	1293715	1293903	-	hypothetical protein	-
BRPE64_RS19975	NC_021294.1	1294451	1295605	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS19980	NC_021294.1	1295616	1296698	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS20115	NC_021294.1	1325644	1326345	-	succinate dehydrogenase iron-sulfur protein	Energy production and conversion
BRPE64_RS20120	NC_021294.1	1326371	1328146	-	succinate dehydrogenase flavoprotein subunit	Energy production and conversion
BRPE64_RS20125	NC_021294.1	1328152	1328520	-	succinate dehydrogenase subunit D	Energy production and conversion
BRPE64_RS20130	NC_021294.1	1328536	1328946	-	succinate dehydrogenase cytochrome b556 subunit	Energy production and conversion
BRPE64_RS20145	NC_021294.1	1330507	1331493	*mdh*	malate dehydrogenase	Energy production and conversion
BRPE64_RS20150	NC_021294.1	1331681	1332691	-	HpcH/HpaI aldolase	Carbohydrate transport and metabolism
BRPE64_RS20155	NC_021294.1	1332800	1333972	-	citrate synthase	Energy production and conversion
BRPE64_RS20165	NC_021294.1	1334602	1336053	-	2-methylcitrate dehydratase	General function prediction only
BRPE64_RS20740	NC_021288.1	1573	2979	-	hypothetical protein	-
BRPE64_RS20745	NC_021288.1	3626	4837	-	cobyrinic acid ac-diamide synthase	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS20750	NC_021288.1	4834	5811	-	ParB-like partition protein	Transcription
BRPE64_RS20860	NC_021288.1	28654	28899	-	hypothetical protein	-
BRPE64_RS20895	NC_021288.1	35576	37138	*glpD*	glycerol-3-phosphate dehydrogenase 2	Energy production and conversion
BRPE64_RS31955	NC_021288.1	60447	60899	-	hypothetical protein	Energy production and conversion
BRPE64_RS21130	NC_021288.1	94175	94687	-	type VI secretion protein VC_A0107 family	Function unknown
BRPE64_RS21780	NC_021288.1	220297	221226	-	transcriptional regulator LysR family	Transcription
BRPE64_RS21785	NC_021288.1	221383	221829	-	activator of Hsp90 ATPase 1 family protein	Function unknown
BRPE64_RS21790	NC_021288.1	221892	223001	-	hypothetical protein	-
BRPE64_RS21795	NC_021288.1	223353	224108	-	NmrA family protein	Carbohydrate transport and metabolism
BRPE64_RS21800	NC_021288.1	224123	225031	-	hypothetical protein	-

BRPE64_RS21805	NC_021288.1	225028	225576	-	putative MxaK-like protein	General function prediction only
BRPE64_RS21810	NC_021288.1	225573	226553	-	putative MxaC-like protein	General function prediction only
BRPE64_RS21815	NC_021288.1	226558	227232	-	hypothetical protein	-
BRPE64_RS21820	NC_021288.1	227460	228323	-	putative MxaS-like protein	General function prediction only
BRPE64_RS21875	NC_021288.1	238338	238988	-	response regulator receiver and ANTAR domain protein	Signal transduction mechanisms
BRPE64_RS21880	NC_021288.1	239019	239783	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS21885	NC_021288.1	239944	240759	-	transcriptional regulator DeoR family	Carbohydrate transport and metabolism
BRPE64_RS21890	NC_021288.1	240831	241796	-	PfkB domain protein	Carbohydrate transport and metabolism
BRPE64_RS21895	NC_021288.1	241789	243072	*kbaZ*	putative tagatose 6-phosphate kinase	Carbohydrate transport and metabolism
BRPE64_RS21900	NC_021288.1	243379	244284	-	probable sugar ABC transporter permease protein	Carbohydrate transport and metabolism
BRPE64_RS21905	NC_021288.1	244359	245306	-	putative sugar (D-ribose) ABC transporter	Carbohydrate transport and metabolism
BRPE64_RS21910	NC_021288.1	245354	246895	-	ABC transporter related	Carbohydrate transport and metabolism
BRPE64_RS21915	NC_021288.1	246960	247832	-	xylose isomerase domain-containing protein TIM barrel	Carbohydrate transport and metabolism
BRPE64_RS21920	NC_021288.1	247829	248458	-	NUDIX hydrolase	Nucleotide transport and metabolism
BRPE64_RS21925	NC_021288.1	248477	249229	-	hypothetical protein	-
BRPE64_RS21930	NC_021288.1	249275	249460	-	hypothetical protein	-
BRPE64_RS21940	NC_021288.1	251347	252135	-	hypothetical protein	Function unknown
BRPE64_RS21945	NC_021288.1	252353	253501	-	outer membrane porin OmpC family	Cell wall/membrane/envelope biogenesis
BRPE64_RS21950	NC_021288.1	253610	254251	-	putative transcriptional regulator	Transcription
BRPE64_RS21955	NC_021288.1	254495	255832	-	acyl-CoA dehydrogenase domain protein	Lipid transport and metabolism
BRPE64_RS21960	NC_021288.1	255899	257239	-	acyl-CoA dehydrogenase domain protein	Lipid transport and metabolism
BRPE64_RS21965	NC_021288.1	257262	258269	-	hypothetical protein	-
BRPE64_RS21970	NC_021288.1	258397	259158	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS21975	NC_021288.1	259148	260032	-	NmrA-like protein	Carbohydrate transport and metabolism
BRPE64_RS21980	NC_021288.1	260213	261085	-	AraC family transcriptional regulator	Transcription
BRPE64_RS21985	NC_021288.1	261237	261986	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism

BRPE64_RS21990	NC_021288.1	262078	262965	-	transcriptional regulator LysR family	Transcription
BRPE64_RS21995	NC_021288.1	263159	263488	-	hypothetical cytosolic protein	Function unknown
BRPE64_RS22000	NC_021288.1	263726	263986	-	hypothetical protein	-
BRPE64_RS22005	NC_021288.1	263999	265078	-	hypothetical protein	Energy production and conversion
BRPE64_RS22010	NC_021288.1	265341	265781	-	putative HTH-type transcriptional regulator ywnA	Transcription
BRPE64_RS22015	NC_021288.1	265907	266554	-	HAD-superfamily hydrolase subfamily IA variant 3	General function prediction only
BRPE64_RS22020	NC_021288.1	266593	267474	-	transcriptional regulator LysR family	Transcription
BRPE64_RS22025	NC_021288.1	267515	268351	*ylbA*	hypothetical protein	General function prediction only
BRPE64_RS22030	NC_021288.1	268747	269589	-	amidohydrolase 2	General function prediction only
BRPE64_RS22035	NC_021288.1	269765	270712	-	Blr7068 protein	General function prediction only
BRPE64_RS22040	NC_021288.1	270768	271286	-	hypothetical protein	-
BRPE64_RS22155	NC_021288.1	291742	292509	-	NAD-dependent epimerase/dehydratase	Lipid transport and metabolism
BRPE64_RS22160	NC_021288.1	292855	293760	-	N-acetylneuraminate lyase	Amino acid transport and metabolism
BRPE64_RS22165	NC_021288.1	293807	294868	-	putative Glu/Leu/Phe/Val dehydrogenase	Amino acid transport and metabolism
BRPE64_RS22170	NC_021288.1	294922	295749	-	transcriptional regulator AraC family	Transcription
BRPE64_RS22175	NC_021288.1	295941	297044	-	putative transcriptional regulator Fis family	Transcription
BRPE64_RS22180	NC_021288.1	297041	297610	-	TetR family transcriptional regulator	Transcription
BRPE64_RS22185	NC_021288.1	297811	298542	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS22190	NC_021288.1	298625	299338	-	NAD(P)H dehydrogenase quinone family	General function prediction only
BRPE64_RS22225	NC_021288.1	305635	306225	-	2-hydroxychromene-2-carboxylate isomerase-like protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS22230	NC_021288.1	306249	306959	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS22235	NC_021288.1	307137	307868	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS22240	NC_021288.1	307881	308552	-	Gst13 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS22245	NC_021288.1	308612	309808	-	L-carnitine dehydratase/bile acid-inducible protein F	Energy production and conversion

BRPE64_RS22250	NC_021288.1	309877	310803	-	2-dehydropantoate 2-reductase	Coenzyme transport and metabolism
BRPE64_RS22255	NC_021288.1	310825	311454	-	glutathione S-transferase domain	Posttranslational modification, protein turnover, chaperones
BRPE64_RS22260	NC_021288.1	311529	312935	*galP*	galactose-proton symport	Amino acid transport and metabolism
BRPE64_RS22370	NC_021288.1	338117	339652	-	drug resistance transporter EmrB/QacA subfamily	Amino acid transport and metabolism
BRPE64_RS22375	NC_021288.1	339649	340581	*argC*	N-acetyl-gamma-glutamyl-phosphate reductase	Amino acid transport and metabolism
BRPE64_RS22600	NC_021288.1	386776	387939	-	PrpF protein involved in 2-methylcitrate cycle	Function unknown
BRPE64_RS22605	NC_021288.1	388009	388890	-	hypothetical protein	-
BRPE64_RS22610	NC_021288.1	389168	389557	-	hypothetical protein	-
BRPE64_RS22615	NC_021288.1	389914	390792	-	hypothetical protein	Function unknown
BRPE64_RS22620	NC_021288.1	390846	391340	-	hypothetical protein	Function unknown
BRPE64_RS22625	NC_021288.1	391353	392069	-	hypothetical protein	Amino acid transport and metabolism
BRPE64_RS22630	NC_021288.1	392983	393543	-	hypothetical protein	-
BRPE64_RS22635	NC_021288.1	393706	393888	-	hypothetical protein	-
BRPE64_RS22640	NC_021288.1	394400	395260	-	hypothetical protein	-
BRPE64_RS22645	NC_021288.1	395257	396297	-	hypothetical protein	-
BRPE64_RS22650	NC_021288.1	396290	397363	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS22655	NC_021288.1	397620	397844	-	hypothetical protein	-
BRPE64_RS22660	NC_021288.1	398148	398390	-	hypothetical protein	-
BRPE64_RS22665	NC_021288.1	398558	398986	-	Bll4598 protein	Function unknown
BRPE64_RS22760	NC_021288.1	434009	434311	-	hypothetical protein	-
BRPE64_RS23020	NC_021288.1	501285	502913	-	AMP-dependent synthetase and ligase	Lipid transport and metabolism
BRPE64_RS23065	NC_021288.1	511330	511917	-	hypothetical protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS23070	NC_021288.1	512029	512700	-	hypothetical protein	-
BRPE64_RS23075	NC_021288.1	512714	514834	-	glycogen debranching enzyme GlgX	Carbohydrate transport and metabolism
BRPE64_RS23080	NC_021288.1	514803	517304	-	phosphorylase	Carbohydrate transport and metabolism
BRPE64_RS23085	NC_021288.1	517492	518175	-	putative signal-transduction protein with CBS domains	General function prediction only
BRPE64_RS23090	NC_021288.1	518414	520756	-	small conductance mechanosensitive channel ion channel	Cell wall/membrane/envelope biogenesis
BRPE64_RS23110	NC_021288.1	524905	525477	-	PEBP family protein	General function prediction only
BRPE64_RS23115	NC_021288.1	526327	526518	-	hypothetical protein	-

BRPE64_RS23330	NC_021288.1	574929	576413	-	aldehyde Dehydrogenase	Energy production and conversion
BRPE64_RS23335	NC_021288.1	576423	577385	-	transcriptional regulator LysR family	Transcription
BRPE64_RS23430	NC_021288.1	595527	598463	-	type III restriction protein res subunit	Transcription
BRPE64_RS23525	NC_021288.1	637314	638576	-	efflux transporter RND family MFP subunit	Cell wall/membrane/envelope biogenesis
BRPE64_RS23530	NC_021288.1	639363	639701	-	transport-associated protein	General function prediction only
BRPE64_RS31990	NC_021288.1	727705	729990	-	hypothetical protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS23940	NC_021288.1	730076	730309	-	hypothetical protein	-
BRPE64_RS23945	NC_021288.1	730330	731085	-	hypothetical protein	-
BRPE64_RS23950	NC_021288.1	731578	731886	-	hypothetical protein	-
BRPE64_RS24235	NC_021288.1	795574	796626	-	hypothetical protein	-
BRPE64_RS24335	NC_021288.1	817885	818145	-	GP29	Function unknown
BRPE64_RS24340	NC_021288.1	818522	818926	-	hypothetical protein	-
BRPE64_RS24345	NC_021288.1	818932	820779	-	hypothetical protein	-
BRPE64_RS24350	NC_021288.1	820769	821656	-	hypothetical protein	-
BRPE64_RS24430	NC_021288.1	836342	837277	-	transcriptional regulator LysR family	Transcription
BRPE64_RS24435	NC_021288.1	837303	838526	-	general substrate transporter	Amino acid transport and metabolism
BRPE64_RS24440	NC_021288.1	838618	839379	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS24690	NC_021289.1	104	1282	-	Soj protein	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS24695	NC_021289.1	1397	2254	-	stage 0 sporulation protein J	Transcription
BRPE64_RS24700	NC_021289.1	2559	3896	-	hypothetical protein	-
BRPE64_RS25135	NC_021289.1	90664	92046	-	histidine kinase	Signal transduction mechanisms
BRPE64_RS32010	NC_021289.1	92426	93082	-	hypothetical protein	-
BRPE64_RS25145	NC_021289.1	95272	97056	*ptsG*	PTS system glucose-specific IIBC subunit	Carbohydrate transport and metabolism
BRPE64_RS25150	NC_021289.1	97070	99601	-	phosphoenolpyruvate-protein phosphotransferase	Carbohydrate transport and metabolism
BRPE64_RS25210	NC_021289.1	111050	111853	-	transglutaminase-like domain protein	Amino acid transport and metabolism
BRPE64_RS25215	NC_021289.1	111908	112780	-	transglutaminase-like domain protein	Amino acid transport and metabolism
BRPE64_RS25220	NC_021289.1	113713	114156	-	alkylhydroperoxidase like protein AhpD family	Function unknown

BRPE64_RS25225	NC_021289.1	114258	115679	-	transcriptional regulator GntR family with aminotransferase domain	Amino acid transport and metabolism
BRPE64_RS25230	NC_021289.1	116049	117164	-	porin Gram-negative type	Cell wall/membrane/envelope biogenesis
BRPE64_RS25240	NC_021289.1	117838	118371	-	formaldehyde-activating enzyme	Function unknown
BRPE64_RS25450	NC_021289.1	162778	163779	-	4-hydroxythreonine-4-phosphate dehydrogenase	Coenzyme transport and metabolism
BRPE64_RS25505	NC_021289.1	176007	177086	-	methylthioribose-1-phosphate isomerase	Translation, ribosomal structure and biogenesis
BRPE64_RS25510	NC_021289.1	177083	177754	-	class II aldolase/adducin family protein	Carbohydrate transport and metabolism
BRPE64_RS25515	NC_021289.1	177836	178948	-	monosaccharide ABC transporter substrate-binding protein CUT2 family	Carbohydrate transport and metabolism
BRPE64_RS25520	NC_021289.1	179007	180542	-	ribose import ATP-binding protein RbsA 1	Carbohydrate transport and metabolism
BRPE64_RS25525	NC_021289.1	180539	181540	-	inner-membrane translocator	Carbohydrate transport and metabolism
BRPE64_RS25530	NC_021289.1	181601	182611	-	dihydroxyacetone kinase DhaK subunit	Carbohydrate transport and metabolism
BRPE64_RS25535	NC_021289.1	182624	183250	*dhaL*	dihydroxyacetone kinase DhaL subunit	Carbohydrate transport and metabolism
BRPE64_RS25540	NC_021289.1	183642	183911	-	hypothetical protein	-
BRPE64_RS25545	NC_021289.1	183937	184962	-	transcriptional regulator AraC family	Transcription
BRPE64_RS25550	NC_021289.1	185093	186031	-	hypothetical protein	Function unknown
BRPE64_RS25555	NC_021289.1	186076	187413	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS25560	NC_021289.1	187400	189007	-	extracellular solute-binding protein family 5	Amino acid transport and metabolism
BRPE64_RS25565	NC_021289.1	189124	190761	-	AMP-dependent synthetase and ligase	Lipid transport and metabolism
BRPE64_RS25570	NC_021289.1	190774	191169	-	endoribonuclease L-PSP	Translation, ribosomal structure and biogenesis
BRPE64_RS25575	NC_021289.1	191166	191615	-	thioesterase superfamily protein	General function prediction only
BRPE64_RS25580	NC_021289.1	191612	192793	-	acyl-CoA dehydrogenase domain protein	Lipid transport and metabolism
BRPE64_RS25585	NC_021289.1	192795	193643	-	enoyl-CoA hydratase	Lipid transport and metabolism
BRPE64_RS25590	NC_021289.1	193655	194215	-	transcriptional regulator MarR family	Transcription
BRPE64_RS25595	NC_021289.1	194212	194985	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism

BRPE64_RS25600	NC_021289.1	194988	197372	-	NADH flavin oxidoreductase/NADH oxidase	Energy production and conversion
BRPE64_RS25720	NC_021289.1	223866	233405	-	amino acid adenylation domain protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS25725	NC_021289.1	233431	238398	-	amino acid adenylation domain protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS25730	NC_021289.1	238412	239776	-	L-ornithine 5-monooxygenase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS25880	NC_021289.1	270671	271447	-	putative NAD(P)H dehydrogenase	General function prediction only
BRPE64_RS25885	NC_021289.1	271540	272436	-	transcriptional regulator LysR family	Transcription
BRPE64_RS25890	NC_021289.1	272598	273863	-	extracellular ligand-binding receptor	Amino acid transport and metabolism
BRPE64_RS25895	NC_021289.1	273946	274821	-	inner-membrane translocator	Amino acid transport and metabolism
BRPE64_RS25900	NC_021289.1	274824	275750	-	inner-membrane translocator	Amino acid transport and metabolism
BRPE64_RS25905	NC_021289.1	275747	277261	-	ABC transporter related protein	Amino acid transport and metabolism
BRPE64_RS25910	NC_021289.1	277521	278408	-	hypothetical protein	-
BRPE64_RS32015	NC_021289.1	278411	279238	-	hypothetical protein	-
BRPE64_RS25920	NC_021289.1	279235	280314	-	hypothetical protein	-
BRPE64_RS25965	NC_021289.1	293746	293946	-	hypothetical protein	-
BRPE64_RS25970	NC_021289.1	294081	295040	-	inner-membrane translocator	Carbohydrate transport and metabolism
BRPE64_RS25975	NC_021289.1	295049	296029	-	inner-membrane translocator	Carbohydrate transport and metabolism
BRPE64_RS25980	NC_021289.1	296019	297581	-	ABC transporter related protein	Carbohydrate transport and metabolism
BRPE64_RS26215	NC_021289.1	348674	349708	-	transcriptional regulator AraC family	Transcription
BRPE64_RS26220	NC_021289.1	349605	349970	-	hypothetical protein	-
BRPE64_RS26225	NC_021289.1	350138	350320	-	Mlr0331 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS26230	NC_021289.1	350322	351401	-	putative ABC transporter substrate-binding protein	Inorganic ion transport and metabolism
BRPE64_RS26235	NC_021289.1	351457	352479	-	putative aldoketoreductase	Energy production and conversion
BRPE64_RS26240	NC_021289.1	352501	353844	-	DDVA O-demethylase	Inorganic ion transport and metabolism
BRPE64_RS26245	NC_021289.1	353980	354783	-	putative transcriptional regulator IclR family	Transcription

BRPE64_RS26250	NC_021289.1	354826	355791	-	putative oxidoreductase	Energy production and conversion
BRPE64_RS26255	NC_021289.1	355861	357132	-	major facilitator superfamily (MFS) transporter	Amino acid transport and metabolism
BRPE64_RS26275	NC_021289.1	360346	360672	*hcaC*	rieske (2Fe-2S) domain protein	Inorganic ion transport and metabolism
BRPE64_RS26280	NC_021289.1	360698	361633	-	cobalamin synthesis protein/P47K family protein	General function prediction only
BRPE64_RS26285	NC_021289.1	361646	362359	-	class II aldolase/adducin family protein	Carbohydrate transport and metabolism
BRPE64_RS26290	NC_021289.1	362385	363185	-	putative taurine transport system permease protein	Inorganic ion transport and metabolism
BRPE64_RS26295	NC_021289.1	363215	364060	-	ABC transporter ATP-binding protein	Inorganic ion transport and metabolism
BRPE64_RS26300	NC_021289.1	364070	365092	-	ABC transporter substrate-binding protein	Inorganic ion transport and metabolism
BRPE64_RS26305	NC_021289.1	365313	366449	-	outer membrane porin OmpC family	Cell wall/membrane/envelope biogenesis
BRPE64_RS26310	NC_021289.1	366934	367740	-	transcriptional regulator IcIR family protein	Transcription
BRPE64_RS26315	NC_021289.1	367767	368807	-	ferredoxin Oxidoreductase FAD/NAD(P)-binding Oxidoreductase FAD-binding region	Energy production and conversion
BRPE64_RS26320	NC_021289.1	368881	369873	-	hypothetical protein	-
BRPE64_RS26325	NC_021289.1	369905	371026	-	hypothetical protein	Function unknown
BRPE64_RS26330	NC_021289.1	371155	372144	-	phthalate 4,5-dioxygenase reductase subunit	Energy production and conversion
BRPE64_RS26335	NC_021289.1	372192	373379	-	oxidoreductase-like protein	General function prediction only
BRPE64_RS26340	NC_021289.1	373376	373930	-	transcriptional regulator MarR family	Transcription
BRPE64_RS26370	NC_021289.1	380998	381606	-	chromate transporter	Inorganic ion transport and metabolism
BRPE64_RS26375	NC_021289.1	381611	382138	-	probable transmembrane protein	Inorganic ion transport and metabolism
BRPE64_RS26380	NC_021289.1	382143	383147	-	alcohol dehydrogenase zinc-binding domain protein	Energy production and conversion
BRPE64_RS26385	NC_021289.1	383247	384116	-	transcriptional regulator LysR family	Transcription
BRPE64_RS26455	NC_021289.1	398570	399598	-	hypothetical protein	Function unknown
BRPE64_RS26475	NC_021289.1	406975	407871	-	transglutaminase domain protein	Amino acid transport and metabolism
BRPE64_RS26480	NC_021289.1	407978	408955	-	oxidoreductase (Aldo/keto reductase) protein	Energy production and conversion
BRPE64_RS26590	NC_021289.1	432863	433048	-	hypothetical protein	-

BRPE64_RS26595	NC_021289.1	433198	434418	-	cytochrome c class I	Amino acid transport and metabolism
BRPE64_RS26600	NC_021289.1	434518	436254	-	putative alcohol dehydrogenase	Carbohydrate transport and metabolism
BRPE64_RS26605	NC_021289.1	436482	436763	-	hypothetical protein	-
BRPE64_RS26610	NC_021289.1	437094	439034	-	GAF modulated sigma54 specific transcriptional regulator Fis family	Transcription
BRPE64_RS26690	NC_021289.1	457869	458261	-	hypothetical protein	-
BRPE64_RS26695	NC_021289.1	458290	458661	-	hypothetical protein	-
BRPE64_RS26700	NC_021289.1	458819	460300	-	phosphoesterase	Cell wall/membrane/envelope biogenesis
BRPE64_RS26705	NC_021289.1	460699	460989	-	hypothetical protein	-
BRPE64_RS26710	NC_021289.1	460986	461510	-	putative GCN5-related N-acetyltransferase	Transcription
BRPE64_RS26715	NC_021289.1	461820	462830	-	hypothetical protein	Energy production and conversion
BRPE64_RS26740	NC_021289.1	469810	470769	-	MoxR-like ATPase putative transcriptional regulator C1 metabolism	General function prediction only
BRPE64_RS26745	NC_021289.1	470814	471773	-	hypothetical protein	General function prediction only
BRPE64_RS26750	NC_021289.1	471770	472285	-	hypothetical protein	-
BRPE64_RS26755	NC_021289.1	472272	473288	-	hypothetical protein	General function prediction only
BRPE64_RS26760	NC_021289.1	473290	474867	-	hypothetical TPR domain protein	General function prediction only
BRPE64_RS26820	NC_021289.1	486774	487370	-	uncharacterized peroxidase-related enzyme	Function unknown
BRPE64_RS26825	NC_021289.1	487447	490056	-	PAS/PAC sensor hybrid histidine kinase	Signal transduction mechanisms
BRPE64_RS26850	NC_021289.1	495291	496172	-	RNA polymerase sigma-24 subunit ECF subfamily	Transcription
BRPE64_RS26855	NC_021289.1	496325	496804	-	alkylhydroperoxidase like protein AhpD family	Function unknown
BRPE64_RS26860	NC_021289.1	496843	497250	-	cupin 2 conserved barrel domain protein	Function unknown
BRPE64_RS26865	NC_021289.1	497349	498362	-	transcriptional regulator LysR family	Transcription
BRPE64_RS26870	NC_021289.1	498603	499121	-	OsmC family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS26875	NC_021289.1	499134	500522	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS26880	NC_021289.1	500586	501356	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26885	NC_021289.1	501370	501708	-	hypothetical protein	-

BRPE64_RS26890	NC_021289.1	501705	503450	-	fumarate reductase/succinate dehydrogenase flavoprotein domain protein	Energy production and conversion
BRPE64_RS26895	NC_021289.1	503505	504344	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26900	NC_021289.1	504480	505373	-	transcriptional regulator AraC family	Transcription
BRPE64_RS26905	NC_021289.1	505496	507145	*treA*	alpha alpha-trehalase	Carbohydrate transport and metabolism
BRPE64_RS26910	NC_021289.1	507160	507696	-	cytochrome c class I	Energy production and conversion
BRPE64_RS26915	NC_021289.1	507689	508900	-	oxidoreductase molybdopterin binding protein	General function prediction only
BRPE64_RS26920	NC_021289.1	509169	509405	-	4-oxalocrotonate tautomerase family enzyme	General function prediction only
BRPE64_RS26925	NC_021289.1	509454	510164	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26975	NC_021289.1	519019	520215	-	acyltransferase 3	Lipid transport and metabolism
BRPE64_RS26980	NC_021289.1	520217	521581	-	multi antimicrobial extrusion protein MatE	Defense mechanisms
BRPE64_RS27050	NC_021289.1	537592	539115	-	sugar ABC transporter ATPase component	Carbohydrate transport and metabolism
BRPE64_RS27055	NC_021289.1	539121	540095	-	periplasmic binding protein/LacI transcriptional regulator	Carbohydrate transport and metabolism
BRPE64_RS27060	NC_021289.1	540196	541257	-	alanine racemase	Transcription
BRPE64_RS27065	NC_021289.1	541403	541987	-	hypothetical protein	Function unknown
BRPE64_RS27070	NC_021289.1	542145	543554	-	hypothetical protein	Function unknown
BRPE64_RS27075	NC_021289.1	543575	543832	-	hypothetical protein	-
BRPE64_RS27080	NC_021289.1	543874	544713	-	hypothetical protein	Carbohydrate transport and metabolism
BRPE64_RS27085	NC_021289.1	544825	545460	-	transcriptional regulator TetR family	Transcription
BRPE64_RS27090	NC_021289.1	545486	546568	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS27095	NC_021289.1	547804	548856	-	transcriptional regulator LacI family	Carbohydrate transport and metabolism
BRPE64_RS27255	NC_021289.1	579480	579740	-	hypothetical protein	-
BRPE64_RS27260	NC_021289.1	579947	580921	-	aliphatic sulfonates family ABC transporter periplasmic ligand-binding protein	Inorganic ion transport and metabolism
BRPE64_RS27375	NC_021289.1	602604	602831	-	hypothetical protein	-
BRPE64_RS27380	NC_021289.1	602852	603976	-	alanine racemase domain protein	Amino acid transport and metabolism
BRPE64_RS27385	NC_021289.1	603973	604626	-	HAD-superfamily hydrolase subfamily IA variant 2	General function prediction only

BRPE64_RS27390	NC_021289.1	604726	605631	-	transcriptional regulator LysR family	Transcription
BRPE64_RS27395	NC_021289.1	605638	606540	-	hypothetical protein	Amino acid transport and metabolism
BRPE64_RS27400	NC_021289.1	606730	607035	-	ferredoxin	Inorganic ion transport and metabolism
BRPE64_RS27405	NC_021289.1	607058	608302	*hcaD*	3-phenylpropionate dioxygenase ferredoxin-NAD(+) reductase component	General function prediction only
BRPE64_RS27415	NC_021289.1	609582	610091	-	hypothetical protein	-
BRPE64_RS27420	NC_021289.1	610114	612546	-	xanthine dehydrogenase molybdenum binding subunit apoprotein	Energy production and conversion
BRPE64_RS27425	NC_021289.1	612543	613379	-	oxidoreductase medium chain	Energy production and conversion
BRPE64_RS27430	NC_021289.1	613390	614580	-	putative iron-sulfur binding protein	Energy production and conversion
BRPE64_RS27435	NC_021289.1	614689	615654	-	amidohydrolase family protein	General function prediction only
BRPE64_RS27440	NC_021289.1	615688	616440	-	putative MALEATE CIS-TRANS ISOMERASE	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS27445	NC_021289.1	616452	617294	-	hydrolase or acyltransferase alpha/beta hydrolase superfamily	General function prediction only
BRPE64_RS27450	NC_021289.1	617281	618315	-	putative peptidase M29 family	Amino acid transport and metabolism
BRPE64_RS27455	NC_021289.1	618334	618957	-	isochorismatase family protein 7	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS27460	NC_021289.1	618985	620580	-	probable transporter	Amino acid transport and metabolism
BRPE64_RS27465	NC_021289.1	620595	621041	-	probable MarR-family transcriptional regulator	Transcription
BRPE64_RS27485	NC_021289.1	624479	626593	-	hypothetical protein	-
BRPE64_RS27490	NC_021289.1	627066	627488	-	hypothetical protein	-
BRPE64_RS32025	NC_021289.1	627485	630376	-	tyrosinase	Amino acid transport and metabolism
BRPE64_RS27500	NC_021289.1	630567	630785	-	hypothetical protein	-
BRPE64_RS27505	NC_021289.1	630789	631745	-	LysR family regulatory protein	Transcription
BRPE64_RS27510	NC_021289.1	631902	633212	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS27515	NC_021289.1	633247	634503	-	metallo peptidase family M20 unassigned	Amino acid transport and metabolism
BRPE64_RS27520	NC_021289.1	634514	635539	-	putative aminohydrolase	Chromatin structure and dynamics

BRPE64_RS27525	NC_021289.1	635701	636525	-	lipid A biosynthesis lauroyl acyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS27535	NC_021289.1	637376	638260	-	periplasmic protein-like protein	Function unknown
BRPE64_RS27540	NC_021289.1	638309	638605	-	hypothetical protein	-
BRPE64_RS27545	NC_021289.1	639206	640537	-	PAS/PAC sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS27550	NC_021289.1	640534	641178	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms
BRPE64_RS27555	NC_021289.1	641318	641704	-	response regulator receiver protein	Signal transduction mechanisms
BRPE64_RS27830	NC_021289.1	702926	703306	-	hypothetical protein	-
BRPE64_RS27835	NC_021289.1	703660	704820	-	alpha-methylacyl-CoA racemase	Energy production and conversion
BRPE64_RS27840	NC_021289.1	705512	707179	-	hypothetical protein	-
BRPE64_RS27940	NC_021289.1	728742	729446	-	hypothetical protein	-
BRPE64_RS27945	NC_021289.1	729500	730708	-	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Energy production and conversion
BRPE64_RS27950	NC_021289.1	730857	731243	-	hypothetical protein	-
BRPE64_RS27955	NC_021289.1	731445	732320	-	transcriptional regulator IcIR family	Transcription
BRPE64_RS27960	NC_021289.1	732615	733706	-	ABC transporter related	Amino acid transport and metabolism
BRPE64_RS27965	NC_021289.1	733699	734616	-	binding-protein-dependent transport systems inner membrane component	Amino acid transport and metabolism
BRPE64_RS27970	NC_021289.1	734636	735430	-	binding-protein-dependent transport systems inner membrane component	Amino acid transport and metabolism
BRPE64_RS28000	NC_021289.1	740141	740749	-	hypothetical protein	-
BRPE64_RS28005	NC_021289.1	740929	741246	-	hypothetical protein	-
BRPE64_RS28010	NC_021289.1	741486	742478	-	4,5-dihydroxyphthalate decarboxylase	Inorganic ion transport and metabolism
BRPE64_RS28015	NC_021289.1	742482	743696	-	oxidoreductase-like protein	General function prediction only
BRPE64_RS28225	NC_021289.1	784477	785094	-	3-isopropylmalate dehydratase small subunit	Amino acid transport and metabolism
BRPE64_RS28230	NC_021289.1	785091	786497	-	3-isopropylmalate dehydratase large subunit	Amino acid transport and metabolism
BRPE64_RS28235	NC_021289.1	786500	787273	-	ABC nitrate/sulfonate/bicarbonate family transporter inner membrane subunit	Inorganic ion transport and metabolism

BRPE64_RS28240	NC_021289.1	787270	788028	-	ABC nitrate/sulfonate/bicarbonate family transporter inner membrane subunit	Inorganic ion transport and metabolism
BRPE64_RS28250	NC_021289.1	788893	789879	-	ABC nitrate/sulfonate/bicarbonate family transporter periplasmic ligand binding protein	Inorganic ion transport and metabolism
BRPE64_RS28255	NC_021289.1	790124	790933	-	hypothetical protein	-
BRPE64_RS28260	NC_021289.1	791000	791695	-	transcriptional regulator GntR family	Transcription
BRPE64_RS28265	NC_021289.1	791706	792206	-	UspA domain protein	Signal transduction mechanisms
BRPE64_RS28270	NC_021289.1	792372	793640	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28275	NC_021289.1	793757	794425	-	O-methyltransferase family protein	General function prediction only
BRPE64_RS28280	NC_021289.1	794528	795115	-	hypothetical protein	-
BRPE64_RS28285	NC_021289.1	795175	795693	-	RNA polymerase sigma-24 subunit ECF subfamily	Transcription
BRPE64_RS28290	NC_021289.1	795690	796514	-	putative transmembrane anti-sigma factor	Transcription
BRPE64_RS28295	NC_021289.1	796519	798078	-	sulphate transporter	Inorganic ion transport and metabolism
BRPE64_RS28300	NC_021289.1	798164	798811	-	carbonic anhydrase	Inorganic ion transport and metabolism
BRPE64_RS28305	NC_021289.1	798876	799058	-	hypothetical protein	-
BRPE64_RS28310	NC_021289.1	799057	799884	-	alpha/beta hydrolase fold protein	General function prediction only
BRPE64_RS28315	NC_021289.1	799881	800735	-	polysaccharide deacetylase	Carbohydrate transport and metabolism
BRPE64_RS28320	NC_021289.1	800732	801817	-	putative glycosyl transferase group 1 family	Cell wall/membrane/envelope biogenesis
BRPE64_RS28355	NC_021289.1	809993	811060	-	hypothetical protein	Translation, ribosomal structure and biogenesis
BRPE64_RS28360	NC_021289.1	811057	812061	-	hypothetical protein	-
BRPE64_RS28365	NC_021289.1	812131	814596	-	glycoside hydrolase family 2 immunoglobulin domain protein beta-sandwich	Carbohydrate transport and metabolism
BRPE64_RS28370	NC_021289.1	814603	815541	-	dehydrogenase	General function prediction only
BRPE64_RS28375	NC_021289.1	815588	816517	-	dihydrodipicolinate synthase putative	Amino acid transport and metabolism
BRPE64_RS28380	NC_021289.1	816556	817902	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism

BRPE64_RS28385	NC_021289.1	818017	819750	-	dihydroxy-acid dehydratase	Amino acid transport and metabolism
BRPE64_RS28390	NC_021289.1	819863	820843	-	transcriptional regulator LysR family putative	Transcription
BRPE64_RS28395	NC_021289.1	820807	824349	-	indolepyruvate ferredoxin oxidoreductase	Energy production and conversion
BRPE64_RS28400	NC_021289.1	824452	825357	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28405	NC_021289.1	825494	828739	-	hydrophobe/amphiphile efflux pump RND family	Defense mechanisms
BRPE64_RS28410	NC_021289.1	828748	829659	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28415	NC_021289.1	829862	830500	-	putative glutathionine S-transferase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS28420	NC_021289.1	830744	831004	-	hypothetical protein	-
BRPE64_RS28425	NC_021289.1	831224	832006	-	hypothetical protein	-
BRPE64_RS28430	NC_021289.1	832439	832699	-	hypothetical protein	-
BRPE64_RS28435	NC_021289.1	832742	833023	-	hypothetical protein	-
BRPE64_RS28440	NC_021289.1	833066	833347	-	hypothetical protein	-
BRPE64_RS28445	NC_021289.1	833390	833650	-	hypothetical protein	-
BRPE64_RS28450	NC_021289.1	833711	834664	-	putative hydrolase	General function prediction only
BRPE64_RS28455	NC_021289.1	834692	835618	-	amidohydrolase 2	General function prediction only
BRPE64_RS28460	NC_021289.1	835615	836031	-	thioesterase/thiol ester dehydrase-isomerase	General function prediction only
BRPE64_RS28465	NC_021289.1	836035	837933	-	TRAP C4-dicarboxylate transport system permease DctM subunit	Carbohydrate transport and metabolism
BRPE64_RS28470	NC_021289.1	837947	838978	-	Blr4511 protein	Carbohydrate transport and metabolism
BRPE64_RS28475	NC_021289.1	839030	840229	-	putative formyl-coenzyme A transferase (Formyl-CoA transferase) Frc	Energy production and conversion
BRPE64_RS28480	NC_021289.1	840231	841010	-	enoyl-CoA hydratase EchA	Lipid transport and metabolism
BRPE64_RS28485	NC_021289.1	841031	842170	-	4-hydroxybutyrate dehydrogenase protein	Energy production and conversion
BRPE64_RS28490	NC_021289.1	842350	843294	-	LysR family transcription regulator protein	Transcription
BRPE64_RS28495	NC_021289.1	843428	844012	-	hypothetical protein	General function prediction only
BRPE64_RS28500	NC_021289.1	844009	844911	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28505	NC_021289.1	845119	845925	-	2,5-didehydrogluconate reductase	General function prediction only
BRPE64_RS28620	NC_021289.1	872107	873327	-	hypothetical protein	Energy production and conversion

BRPE64_RS28630	NC_021289.1	875514	876437	-	aromatic acid efflux system membrane fusion protein EmrA subfamily	Defense mechanisms
BRPE64_RS28635	NC_021289.1	876509	876988	-	hypothetical protein	Function unknown
BRPE64_RS28640	NC_021289.1	877000	877821	-	alpha/beta hydrolase fold	General function prediction only
BRPE64_RS28645	NC_021289.1	878107	878679	-	hypothetical protein	Function unknown
BRPE64_RS28660	NC_021289.1	880703	880981	-	muconolactone delta-isomerase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS28665	NC_021289.1	881049	882176	-	muconate and chloromuconate cycloisomerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS28670	NC_021289.1	882280	883203	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28675	NC_021289.1	883462	884397	-	catechol 1,2-dioxygenase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS28680	NC_021289.1	884506	885864	-	rieske (2Fe-2S) domain protein	Inorganic ion transport and metabolism
BRPE64_RS28815	NC_021289.1	911880	912140	-	hypothetical protein	-
BRPE64_RS28820	NC_021289.1	912378	913442	-	molybdenum cofactor biosynthesis protein A	Coenzyme transport and metabolism
BRPE64_RS28870	NC_021289.1	924126	930071	-	PAS sensor protein	General function prediction only
BRPE64_RS28875	NC_021289.1	930207	931046	-	enoyl-CoA hydratase/isomerase	Lipid transport and metabolism
BRPE64_RS28880	NC_021289.1	931074	931751	-	ThiJ/PfpI domain protein	General function prediction only
BRPE64_RS28885	NC_021289.1	932347	932997	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms
BRPE64_RS28890	NC_021289.1	933146	933538	-	hypothetical protein	-
BRPE64_RS28895	NC_021289.1	933554	934567	-	putative hydrolase	General function prediction only
BRPE64_RS28900	NC_021289.1	934705	935001	-	putative transcriptional regulator	Transcription
BRPE64_RS28905	NC_021289.1	935020	936822	-	acetolactate synthase	Amino acid transport and metabolism
BRPE64_RS28910	NC_021289.1	937273	938301	-	transcriptional regulator AraC family	Transcription
BRPE64_RS28915	NC_021289.1	938294	938815	-	uracil-DNA glycosylase superfamily	Replication, recombination and repair
BRPE64_RS28920	NC_021289.1	938820	939767	-	transcriptional regulator LysR family	Transcription
BRPE64_RS28925	NC_021289.1	940092	940766	-	antibiotic biosynthesis monooxygenase	General function prediction only
BRPE64_RS28930	NC_021289.1	941147	942721	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism

BRPE64_RS28935	NC_021289.1	942734	943810	-	secretion protein HlyD family protein	Defense mechanisms
BRPE64_RS28940	NC_021289.1	943807	945345	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS28945	NC_021289.1	945368	945745	-	hypothetical protein	-
BRPE64_RS28950	NC_021289.1	945788	946711	-	transcriptional regulator AraC family	Transcription
BRPE64_RS28955	NC_021289.1	947010	950204	-	acriflavin resistance protein	Defense mechanisms
BRPE64_RS28960	NC_021289.1	950215	951378	-	efflux transporter RND family MFP subunit	Cell wall/membrane/envelope biogenesis
BRPE64_RS28965	NC_021289.1	951392	952852	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane/envelope biogenesis
BRPE64_RS29175	NC_021289.1	995247	995876	-	hypothetical protein	General function prediction only
BRPE64_RS29180	NC_021289.1	996021	996872	-	alpha/beta hydrolase fold protein	General function prediction only
BRPE64_RS29185	NC_021289.1	996918	997889	-	alcohol dehydrogenase zinc-binding domain protein	Energy production and conversion
BRPE64_RS29190	NC_021289.1	997905	998675	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS29200	NC_021289.1	999006	999782	-	oxidoreductase molybdopterin binding protein	General function prediction only
BRPE64_RS29205	NC_021289.1	999802	1000431	-	putative transmembrane hydrogenase cytochrome b-type subunit	Energy production and conversion
BRPE64_RS29210	NC_021289.1	1000774	1001298	-	hypothetical protein	-
BRPE64_RS29215	NC_021289.1	1001295	1001936	-	hypothetical protein	-
BRPE64_RS29220	NC_021289.1	1001937	1002560	-	transcriptional regulator TetR family protein	Transcription
BRPE64_RS29225	NC_021289.1	1002576	1003280	-	glutathione S-transferase domain	Posttranslational modification, protein turnover, chaperones
BRPE64_RS29230	NC_021289.1	1003366	1004385	-	putative dehydrogenase	Energy production and conversion
BRPE64_RS29235	NC_021289.1	1004382	1005704	-	C4-dicarboxylate transport protein	Energy production and conversion
BRPE64_RS29240	NC_021289.1	1005857	1007032	-	altronate dehydratase	Carbohydrate transport and metabolism
BRPE64_RS29245	NC_021289.1	1007082	1007351	-	hypothetical protein	-
BRPE64_RS29250	NC_021289.1	1007619	1008371	-	transcriptional regulator GntR family	Transcription
BRPE64_RS29255	NC_021289.1	1008404	1008784	-	hypothetical protein	-
BRPE64_RS29260	NC_021289.1	1008820	1009278	-	MEKHLA domain protein	Signal transduction mechanisms

BRPE64_RS29265	NC_021289.1	1009280	1010140	-	NmrA family protein	Carbohydrate transport and metabolism
BRPE64_RS29270	NC_021289.1	1010262	1011170	-	transcriptional regulator	Transcription
BRPE64_RS29275	NC_021289.1	1011280	1011893	-	hypothetical protein	-
BRPE64_RS29280	NC_021289.1	1012163	1014901	-	hypothetical protein	-
BRPE64_RS29285	NC_021289.1	1014909	1016366	-	succinate semialdehyde dehydrogenase	Energy production and conversion
BRPE64_RS29290	NC_021289.1	1016500	1018170	*actP*	SSS sodium solute transporter superfamily	General function prediction only
BRPE64_RS29295	NC_021289.1	1018167	1018472	-	hypothetical protein	Function unknown
BRPE64_RS29635	NC_021289.1	1094891	1095277	-	hypothetical protein	General function prediction only
BRPE64_RS29640	NC_021289.1	1095281	1096051	-	aliphatic sulfonate import ATP-binding protein SsuB 2	Inorganic ion transport and metabolism
BRPE64_RS29645	NC_021289.1	1096065	1096847	-	binding-protein-dependent transport systems inner membrane component	Inorganic ion transport and metabolism
BRPE64_RS29650	NC_021289.1	1096859	1097695	-	binding-protein-dependent transport systems inner membrane component	Inorganic ion transport and metabolism
BRPE64_RS29655	NC_021289.1	1097717	1098781	-	ABC nitrate/sulfonate/bicarbonate family transporter periplasmic ligand binding protein	Inorganic ion transport and metabolism
BRPE64_RS29660	NC_021289.1	1098849	1099787	-	taurine dioxygenase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS29665	NC_021289.1	1099917	1101008	-	2-nitropropane dioxygenase	General function prediction only
BRPE64_RS29670	NC_021289.1	1101038	1101952	-	transcriptional regulator LysR family	Transcription
BRPE64_RS29695	NC_021289.1	1106655	1107551	-	fructose-bisphosphate aldolase	Carbohydrate transport and metabolism
BRPE64_RS29700	NC_021289.1	1107730	1108164	-	hypothetical protein	-
BRPE64_RS32030	NC_021289.1	1113334	1113807	-	hypothetical protein	-
BRPE64_RS29735	NC_021289.1	1115311	1116450	-	acyltransferase 3	Lipid transport and metabolism
BRPE64_RS29740	NC_021289.1	1116464	1118935	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS29750	NC_021289.1	1120449	1122653	-	hypothetical protein	-
BRPE64_RS29755	NC_021289.1	1122863	1123906	-	GDP-mannose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS29760	NC_021289.1	1123914	1124831	-	NAD-dependent epimerase/dehydratase	Carbohydrate transport and metabolism
BRPE64_RS29765	NC_021289.1	1124871	1126049	-	glycosyl transferase	Cell wall/membrane/envelope biogenesis

BRPE64_RS29770	NC_021289.1	1126085	1127269	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS29780	NC_021289.1	1128755	1129888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS29785	NC_021289.1	1129885	1130856	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS29795	NC_021289.1	1133117	1134301	*wza*	polysaccharide export protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS29800	NC_021289.1	1134280	1134723	-	protein tyrosine phosphatase	Signal transduction mechanisms
BRPE64_RS29810	NC_021289.1	1136164	1137561	-	undecaprenyl-phosphate glucose phosphotransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS29830	NC_021289.1	1142044	1143165	-	acyltransferase 3	Lipid transport and metabolism
BRPE64_RS29835	NC_021289.1	1143239	1143706	-	transcriptional regulator MarR-family	Transcription
BRPE64_RS30100	NC_021289.1	1198538	1198867	-	hypothetical protein	-
BRPE64_RS30105	NC_021289.1	1198966	1200720	-	hypothetical protein	-
BRPE64_RS30150	NC_021289.1	1208068	1208370	-	hypothetical protein	-
BRPE64_RS30155	NC_021289.1	1208404	1208835	-	hypothetical protein	-
BRPE64_RS30160	NC_021289.1	1209233	1209418	-	putative periplasmic nitrate reductase NapE	Energy production and conversion
BRPE64_RS30165	NC_021289.1	1209454	1209792	-	periplasmic nitrate reductase chaperone NapD	Inorganic ion transport and metabolism
BRPE64_RS30485	NC_021295.1	41	1423	-	hypothetical protein	-
BRPE64_RS30490	NC_021295.1	2211	3419	-	cobyrinic acid ac-diamide synthase	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS30495	NC_021295.1	3416	4387	-	ParB-like partition protein	Transcription
BRPE64_RS30615	NC_021295.1	29840	30175	-	hypothetical protein	-
BRPE64_RS30620	NC_021295.1	30884	31384	-	hypothetical protein	-
BRPE64_RS31095	NC_021295.1	145393	146151	-	hypothetical protein	-
BRPE64_RS31385	NC_021295.1	218747	219211	-	hypothetical protein	-
BRPE64_RS31585	NC_021295.1	263445	263879	-	hypothetical protein	-
BRPE64_RS31590	NC_021295.1	263922	264458	-	hypothetical protein	-
BRPE64_RS31595	NC_021295.1	264617	264910	-	hypothetical protein	-

Annexe 6: Lists of fitness genes involved in polymyxin B resistance in *B. insecticola* identified by Con-ARTIST.

Annexe 6.1: Genes identified for the lowest concentration (1.5 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS10555	1	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS11045	1	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane/envelope biogenesis
BRPE64_RS15370	1	NC_021294.1	274872	275996	*ydiK*	hypothetical protein	General function prediction only
BRPE64_RS16125	1	NC_021294.1	440928	441707	-	cationic amino acid ABC transporter periplasmic binding protein	Amino acid transport and metabolism
BRPE64_RS26610	1	NC_021289.1	437094	439034	-	GAF modulated sigma54 specific transcriptional regulator Fis family	Transcription
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS08480	2	NC_021287.1	1841708	1842364	-	NLP/P60 protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10565	2	NC_021287.1	2295122	2296966	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS10570	2	NC_021287.1	2296988	2300452	-	hypothetical protein	-
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS16130	2	NC_021294.1	441906	442310	-	heat shock protein Hsp20	Posttranslational modification, protein turnover, chaperones
BRPE64_RS20665	2	NC_021294.1	1449187	1450284	-	N-acylglucosamine 2-epimerase	Carbohydrate transport and metabolism
BRPE64_RS22640	2	NC_021288.1	394400	395260	-	hypothetical protein	-
BRPE64_RS22645	2	NC_021288.1	395257	396297	-	hypothetical protein	-
BRPE64_RS22650	2	NC_021288.1	396290	397363	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS23085	2	NC_021288.1	517492	518175	-	putative signal-transduction protein with CBS domains	General function prediction only
BRPE64_RS23090	2	NC_021288.1	518414	520756	-	small conductance mechanosensitive channel ion channel	Cell wall/membrane/envelope biogenesis

BRPE64_RS23430	2	NC_021288.1	595527	598463	-	type III restriction protein res subunit	Transcription
BRPE64_RS26595	2	NC_021289.1	433198	434418	-	cytochrome c class I	Amino acid transport and metabolism
BRPE64_RS26600	2	NC_021289.1	434518	436254	-	putative alcohol dehydrogenase	Carbohydrate transport and metabolism
BRPE64_RS26605	2	NC_021289.1	436482	436763	-	hypothetical protein	-
BRPE64_RS07490	3	NC_021287.1	1621985	1624027	-	poly(R)-hydroxyalkanoic acid synthase class I	Lipid transport and metabolism

Annexe 6.2: Genes identified for the highest concentration (12.5 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS05760	1	NC_021287.1	1249156	1250538	-	membrane-associated zinc metalloprotease	Cell wall/membrane/envelope biogenesis
BRPE64_RS08480	1	NC_021287.1	1841708	1842364	-	NLP/P60 protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS09935	1	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS09940	1	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS10555	1	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10595	1	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11025	1	NC_021287.1	2388863	2389540	*tolQ*	protein TolQ	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12010	1	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12120	1	NC_021287.1	2620349	2620900	-	putative signal peptide protein toluene tolerance Ttg2C-like protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS16125	1	NC_021294.1	440928	441707	-	cationic amino acid ABC transporter periplasmic binding protein	Amino acid transport and metabolism
BRPE64_RS00670	2	NC_021287.1	144624	145271	*dsbA*	thiol disulfide interchange protein DsbA	Energy production and conversion
BRPE64_RS02150	2	NC_021287.1	465159	465821	*dedA*	membrane-associated protein	Function unknown
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04500	2	NC_021287.1	977887	979164	-	O-antigen polymerase	Cell wall/membrane/envelope biogenesis

BRPE64_RS10075	2	NC_021287.1	2175984	2177834	-	hypothetical protein	General function prediction only
BRPE64_RS10300	2	NC_021287.1	2228407	2229402	*waaC*	lipopolysaccharide heptosyltransferase I	Cell wall/membrane/envelope biogenesis
BRPE64_RS10475	2	NC_021287.1	2273911	2275809	-	polysaccharide biosynthesis protein CapD	Carbohydrate transport and metabolism
BRPE64_RS10480	2	NC_021287.1	2275822	2276847	-	glycosyl transferase family 4	Cell wall/membrane/envelope biogenesis
BRPE64_RS10485	2	NC_021287.1	2277059	2278042	-	NAD-dependent epimerase/dehydratase	Carbohydrate transport and metabolism
BRPE64_RS10490	2	NC_021287.1	2278039	2278902	-	putative glycosyl transferase	General function prediction only
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10565	2	NC_021287.1	2295122	2296966	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS10570	2	NC_021287.1	2296988	2300452	-	hypothetical protein	-
BRPE64_RS10575	2	NC_021287.1	2300667	2301611	-	rhamnosyltransferase	General function prediction only
BRPE64_RS10580	2	NC_021287.1	2301632	2302519	-	dTDP-4-dehydrorhamnose reductase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10585	2	NC_021287.1	2302528	2303079	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10590	2	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11030	2	NC_021287.1	2389555	2389998	-	TolR protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11035	2	NC_021287.1	2389995	2391107	-	protein TolA	Cell wall/membrane/envelope biogenesis
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11045	2	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane/envelope biogenesis
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12125	2	NC_021287.1	2620994	2621761	-	hypothetical protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12130	2	NC_021287.1	2621758	2622573	-	ABC transporter related	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12945	2	NC_021287.1	2779557	2779739	-	hypothetical protein	-
BRPE64_RS16130	2	NC_021294.1	441906	442310	-	heat shock protein Hsp20	Posttranslational modification, protein turnover, chaperones
BRPE64_RS16135	2	NC_021294.1	442323	442757	-	heat shock protein Hsp20	Posttranslational modification, protein turnover, chaperones

BRPE64_RS16140	2	NC_021294.1	442756	442953	-	hypothetical protein	-
BRPE64_RS19345	2	NC_021294.1	1134818	1135984	-	outer membrane porin OmpC family	Cell wall/membrane/envelope biogenesis
BRPE64_RS02785	3	NC_021287.1	614758	615954	-	fatty acid desaturase	Lipid transport and metabolism
BRPE64_RS02795	3	NC_021287.1	617201	618067	*nadC*	nicotinate-nucleotide pyrophosphorylase	Coenzyme transport and metabolism
BRPE64_RS02975	3	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS01390	4	NC_021287.1	288271	289794	*ilvA*	L-threonine ammonia-lyase	Amino acid transport and metabolism
BRPE64_RS02790	4	NC_021287.1	616077	617204	*nadA*	quinolinate synthase A	Coenzyme transport and metabolism
BRPE64_RS02980	4	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS02985	4	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS02990	4	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS02995	4	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS03965	4	NC_021287.1	861787	862632	-	binding-protein-dependent transport systems inner membrane component	Carbohydrate transport and metabolism
BRPE64_RS03970	4	NC_021287.1	862634	863572	-	carbohydrate ABC transporter membrane protein 1 CUT1 family	Carbohydrate transport and metabolism
BRPE64_RS03975	4	NC_021287.1	863764	865017	-	extracellular solute-binding protein family 1	Carbohydrate transport and metabolism
BRPE64_RS07485	4	NC_021287.1	1620724	1621905	-	acetyl-CoA acetyltransferase	Lipid transport and metabolism
BRPE64_RS07490	4	NC_021287.1	1621985	1624027	-	poly(R)-hydroxyalkanoic acid synthase class I	Lipid transport and metabolism

Annexe 7: List of fitness genes involved in LL-37 resistance in *B. insecticola* identified by Con-ARTIST (only for the highest concentration (12.5 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS02300	1	NC_021287.1	495263	496360	-	lipopolysaccharide heptosyltransferase II	Cell wall/membrane/envelope biogenesis
BRPE64_RS05760	1	NC_021287.1	1249156	1250538	-	membrane-associated zinc metalloprotease	Cell wall/membrane/envelope biogenesis
BRPE64_RS09940	1	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS10475	1	NC_021287.1	2273911	2275809	-	polysaccharide biosynthesis protein CapD	Carbohydrate transport and metabolism
BRPE64_RS10490	1	NC_021287.1	2278039	2278902	-	putative glycosyl transferase	General function prediction only
BRPE64_RS10555	1	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS11030	1	NC_021287.1	2389555	2389998	-	TolR protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11045	1	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane/envelope biogenesis
BRPE64_RS12010	1	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12025	1	NC_021287.1	2606315	2606677	-	adenosine 5'-monophosphoramidase / Guanosine 5'-monophosphoramidase	Nucleotide transport and metabolism
BRPE64_RS12130	1	NC_021287.1	2621758	2622573	-	ABC transporter related	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS02150	2	NC_021287.1	465159	465821	*dedA*	membrane-associated protein	Function unknown
BRPE64_RS02305	2	NC_021287.1	496532	496729	-	hypothetical protein	Function unknown
BRPE64_RS02310	2	NC_021287.1	496806	497726	*ilvE*	branched-chain amino acid aminotransferase	Amino acid transport and metabolism
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04500	2	NC_021287.1	977887	979164	-	O-antigen polymerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS06370	2	NC_021287.1	1364944	1365315	-	drug/metabolite transporter (DMT) superfamily permease	Amino acid transport and metabolism
BRPE64_RS09935	2	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS10295	2	NC_021287.1	2227455	2228339	-	hypothetical protein	-

BRPE64_RS10300	2	NC_021287.1	2228407	2229402	*waaC*	lipopolysaccharide heptosyltransferase I	Cell wall/membrane/envelope biogenesis
BRPE64_RS10480	2	NC_021287.1	2275822	2276847	-	glycosyl transferase family 4	Cell wall/membrane/envelope biogenesis
BRPE64_RS10485	2	NC_021287.1	2277059	2278042	-	NAD-dependent epimerase/dehydratase	Carbohydrate transport and metabolism
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10565	2	NC_021287.1	2295122	2296966	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS10570	2	NC_021287.1	2296988	2300452	-	hypothetical protein	-
BRPE64_RS10575	2	NC_021287.1	2300667	2301611	-	rhamnosyltransferase	General function prediction only
BRPE64_RS10580	2	NC_021287.1	2301632	2302519	-	dTDP-4-dehydrorhamnose reductase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10585	2	NC_021287.1	2302528	2303079	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10590	2	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10595	2	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11035	2	NC_021287.1	2389995	2391107	-	protein TolA	Cell wall/membrane/envelope biogenesis
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12105	2	NC_021287.1	2618403	2618690	-	hypothetical protein	-
BRPE64_RS12110	2	NC_021287.1	2618687	2619316	-	toluene tolerance family protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12115	2	NC_021287.1	2619442	2620338	-	VacJ family lipoprotein	Cell wall/membrane/envelope biogenesis
BRPE64_RS12120	2	NC_021287.1	2620349	2620900	-	putative signal peptide protein toluene tolerance Ttg2C-like protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12125	2	NC_021287.1	2620994	2621761	-	hypothetical protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS28290	2	NC_021289.1	795690	796514	-	putative transmembrane anti-sigma factor	Transcription

BRPE64_RS28705	3	NC_021289.1	888916	889830	*metR*	transcriptional regulator LysR family	Transcription
BRPE64_RS28710	4	NC_021289.1	889942	892233	*metE*	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	Amino acid transport and metabolism

Annexe 8: Lists of fitness genes involved in riptocin resistance in *B. insecticola* identified by Con-ARTIST.

Annexe 8.1: Genes identified for the lowest concentration (100 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS12010	1	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12025	1	NC_021287.1	2606315	2606677	-	adenosine 5'-monophosphoramidase / Guanosine 5'-monophosphoramidase	Nucleotide transport and metabolism
BRPE64_RS26745	1	NC_021289.1	470814	471773	-	hypothetical protein	General function prediction only
BRPE64_RS26765	1	NC_021289.1	474864	476186	-	hypothetical protein	-
BRPE64_RS03955	2	NC_021287.1	858398	860509	*ptrB*	prolyl endopeptidase	Amino acid transport and metabolism
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS26750	2	NC_021289.1	471770	472285	-	hypothetical protein	-
BRPE64_RS26755	2	NC_021289.1	472272	473288	-	hypothetical protein	General function prediction only
BRPE64_RS26760	2	NC_021289.1	473290	474867	-	hypothetical TPR domain protein	General function prediction only
BRPE64_RS01390	3	NC_021287.1	288271	289794	*ilvA*	L-threonine ammonia-lyase	Amino acid transport and metabolism
BRPE64_RS01395	3	NC_021287.1	290326	294432	-	FAD linked oxidase domain protein	Energy production and conversion
BRPE64_RS02785	3	NC_021287.1	614758	615954	-	fatty acid desaturase	Lipid transport and metabolism
BRPE64_RS02795	3	NC_021287.1	617201	618067	*nadC*	nicotinate-nucleotide pyrophosphorylase	Coenzyme transport and metabolism
BRPE64_RS02975	3	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS02980	3	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS03960	3	NC_021287.1	860619	861737	-	ABC transporter related	Carbohydrate transport and metabolism
BRPE64_RS00270	4	NC_021287.1	59268	59876	-	methionine biosynthesis protein MetW	Secondary metabolites biosynthesis, transport and catabolism

BRPE64_RS00275	4	NC_021287.1	59873	61018	-	homoserine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS02790	4	NC_021287.1	616077	617204	*nadA*	quinolinate synthase A	Coenzyme transport and metabolism
BRPE64_RS03965	4	NC_021287.1	861787	862632	-	binding-protein-dependent transport systems inner membrane component	Carbohydrate transport and metabolism
BRPE64_RS03970	4	NC_021287.1	862634	863572	-	carbohydrate ABC transporter membrane protein 1 CUT1 family	Carbohydrate transport and metabolism
BRPE64_RS03975	4	NC_021287.1	863764	865017	-	extracellular solute-binding protein family 1	Carbohydrate transport and metabolism
BRPE64_RS03980	4	NC_021287.1	865641	867101	-	glucose-6-phosphate 1-dehydrogenase	Carbohydrate transport and metabolism
BRPE64_RS07490	4	NC_021287.1	1621985	1624027	-	poly(R)-hydroxyalkanoic acid synthase class I	Lipid transport and metabolism
BRPE64_RS28705	4	NC_021289.1	888916	889830	*metR*	transcriptional regulator LysR family	Transcription
BRPE64_RS28710	4	NC_021289.1	889942	892233	*metE*	5-methyltetrahydropteroyltriglutamate-- homocysteine methyltransferase	Amino acid transport and metabolism

Annexe 8.2: Genes identified for the highest concentration (200 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS04490	1	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS05760	1	NC_021287.1	1249156	1250538	-	membrane-associated zinc metalloprotease	Cell wall/membrane/envelope biogenesis
BRPE64_RS07970	1	NC_021287.1	1735999	1737438	-	dihydrolipoyl dehydrogenase	Energy production and conversion
BRPE64_RS07975	1	NC_021287.1	1737536	1738837	*sucB*	2-oxoglutarate dehydrogenase E2 subunit dihydrolipoamide succinyltransferase	Energy production and conversion
BRPE64_RS10555	1	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS11045	1	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane/envelope biogenesis
BRPE64_RS12010	1	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12025	1	NC_021287.1	2606315	2606677	-	adenosine 5'-monophosphoramidase / Guanosine 5'-monophosphoramidase	Nucleotide transport and metabolism
BRPE64_RS12805	1	NC_021287.1	2752012	2752797	*garL*	2-dehydro-3-deoxyglucarate aldolase	Carbohydrate transport and metabolism

BRPE64_RS03955	2	NC_021287.1	858398	860509	*ptrB*	prolyl endopeptidase	Amino acid transport and metabolism
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS02145	3	NC_021287.1	462950	464950	*mutL*	DNA mismatch repair protein MutL	Replication, recombination and repair
BRPE64_RS02785	3	NC_021287.1	614758	615954	-	fatty acid desaturase	Lipid transport and metabolism
BRPE64_RS02795	3	NC_021287.1	617201	618067	*nadC*	nicotinate-nucleotide pyrophosphorylase	Coenzyme transport and metabolism
BRPE64_RS02975	3	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS03960	3	NC_021287.1	860619	861737	-	ABC transporter related	Carbohydrate transport and metabolism
BRPE64_RS03985	3	NC_021287.1	867246	867965	-	6-phosphogluconolactonase	Carbohydrate transport and metabolism
BRPE64_RS05685	3	NC_021287.1	1228362	1231877	-	chromosome partition protein Smc	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS05715	3	NC_021287.1	1238962	1241547	*glnD*	uridylyltransferase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS07555	3	NC_021287.1	1639396	1640298	*hflC*	band 7 protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS08850	3	NC_021287.1	1918274	1920037	-	acetolactate synthase large subunit biosynthetic type	Amino acid transport and metabolism
BRPE64_RS10125	3	NC_021287.1	2187950	2190004	-	NAD synthetase/Glutamine amidotransferase chain of NAD synthetase	Coenzyme transport and metabolism
BRPE64_RS11770	3	NC_021287.1	2554305	2555090	*trpC*	indole-3-glycerol phosphate synthase	Amino acid transport and metabolism
BRPE64_RS00270	4	NC_021287.1	59268	59876	-	methionine biosynthesis protein MetW	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS00275	4	NC_021287.1	59873	61018	-	homoserine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS01390	4	NC_021287.1	288271	289794	*ilvA*	L-threonine ammonia-lyase	Amino acid transport and metabolism
BRPE64_RS02135	4	NC_021287.1	460885	461940	*purM*	phosphoribosylformylglycinamidine cyclo-ligase	Nucleotide transport and metabolism
BRPE64_RS02140	4	NC_021287.1	462006	462953	*miaA*	tRNA dimethylallyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS02790	4	NC_021287.1	616077	617204	*nadA*	quinolinate synthase A	Coenzyme transport and metabolism

BRPE64_RS02980	4	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS02985	4	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS02990	4	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS02995	4	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS03000	4	NC_021287.1	655612	656925	*cysN*	sulfate adenylyltransferase large subunit	Inorganic ion transport and metabolism
BRPE64_RS03005	4	NC_021287.1	656943	657698	-	uroporphyrin-III C-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS03010	4	NC_021287.1	657835	658215	-	cobalamin (Vitamin B12) biosynthesis CbiX protein	Function unknown
BRPE64_RS03145	4	NC_021287.1	682640	683026	*panD*	aspartate 1-decarboxylase	Coenzyme transport and metabolism
BRPE64_RS03965	4	NC_021287.1	861787	862632	-	binding-protein-dependent transport systems inner membrane component	Carbohydrate transport and metabolism
BRPE64_RS03970	4	NC_021287.1	862634	863572	-	carbohydrate ABC transporter membrane protein 1 CUT1 family	Carbohydrate transport and metabolism
BRPE64_RS03975	4	NC_021287.1	863764	865017	-	extracellular solute-binding protein family 1	Carbohydrate transport and metabolism
BRPE64_RS03980	4	NC_021287.1	865641	867101	-	glucose-6-phosphate 1-dehydrogenase	Carbohydrate transport and metabolism
BRPE64_RS04240	4	NC_021287.1	916210	918189	*parE*	DNA topoisomerase	Replication, recombination and repair
BRPE64_RS04245	4	NC_021287.1	918228	920546	*parC*	DNA topoisomerase IV A subunit	Replication, recombination and repair
BRPE64_RS05675	4	NC_021287.1	1225755	1227020	-	succinyldiaminopimelate aminotransferase apoenzyme	Amino acid transport and metabolism
BRPE64_RS05680	4	NC_021287.1	1227225	1228205	-	hypothetical protein	Amino acid transport and metabolism
BRPE64_RS07490	4	NC_021287.1	1621985	1624027	-	poly(R)-hydroxyalkanoic acid synthase class I	Lipid transport and metabolism
BRPE64_RS07550	4	NC_021287.1	1639153	1639344	-	hypothetical protein	Function unknown
BRPE64_RS11755	4	NC_021287.1	2551158	2552654	*trpE*	anthranilate synthase component I	Amino acid transport and metabolism
BRPE64_RS11760	4	NC_021287.1	2552667	2553254	-	glutamine amidotransferase of anthranilate synthase	Amino acid transport and metabolism
BRPE64_RS11765	4	NC_021287.1	2553259	2554293	*trpD*	anthranilate phosphoribosyltransferase	Amino acid transport and metabolism
BRPE64_RS12165	4	NC_021287.1	2629958	2634661	*gltB*	glutamate synthase	Amino acid transport and metabolism

Annexe 9: Lists of fitness genes involved in CCR179 peptide resistance in *B. insecticola* identified by Con-ARTIST.

Annexe 9.1: Genes identified for the lowest concentration (100 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS04485	1	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS09940	1	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS10275	1	NC_021287.1	2223364	2223789	-	hypothetical protein	-
BRPE64_RS10555	1	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10580	1	NC_021287.1	2301632	2302519	-	dTDP-4-dehydrorhamnose reductase	Cell wall/membrane/envelope biogenesis
BRPE64_RS12280	1	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS02300	2	NC_021287.1	495263	496360	-	lipopolysaccharide heptosyltransferase II	Cell wall/membrane/envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS09930	2	NC_021287.1	2141048	2141443	-	helix-hairpin-helix motif protein	Replication, recombination and repair
BRPE64_RS09935	2	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10585	2	NC_021287.1	2302528	2303079	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10590	2	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10595	2	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS12010	2	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport

BRPE64_RS00270	3	NC_021287.1	59268	59876	-	methionine biosynthesis protein MetW	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS02975	3	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS02985	3	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS02995	3	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS11750	3	NC_021287.1	2550034	2550774	-	phosphoglycolate phosphatase	General function prediction only
BRPE64_RS11760	3	NC_021287.1	2552667	2553254	-	glutamine amidotransferase of anthranilate synthase	Amino acid transport and metabolism
BRPE64_RS12165	3	NC_021287.1	2629958	2634661	*gltB*	glutamate synthase	Amino acid transport and metabolism
BRPE64_RS12170	3	NC_021287.1	2634989	2635696	-	transposase IS200-family protein	Replication, recombination and repair
BRPE64_RS00275	4	NC_021287.1	59873	61018	-	homoserine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS02980	4	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS02990	4	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS03000	4	NC_021287.1	655612	656925	*cysN*	sulfate adenylyltransferase large subunit	Inorganic ion transport and metabolism
BRPE64_RS03005	4	NC_021287.1	656943	657698	-	uroporphyrin-III C-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS11755	4	NC_021287.1	2551158	2552654	*trpE*	anthranilate synthase component I	Amino acid transport and metabolism

Annexe 9.2: Genes identified for the highest concentration (200 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS04500	1	NC_021287.1	977887	979164	-	O-antigen polymerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10515	1	NC_021287.1	2282513	2283994	-	hypothetical protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS12010	1	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12280	1	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12285	1	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification,

							protein turnover, chaperones
BRPE64_RS02300	2	NC_021287.1	495263	496360	-	lipopolysaccharide heptosyltransferase II	Cell wall/membrane/envelope biogenesis
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS09930	2	NC_021287.1	2141048	2141443	-	helix-hairpin-helix motif protein	Replication, recombination and repair
BRPE64_RS09935	2	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10565	2	NC_021287.1	2295122	2296966	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS10570	2	NC_021287.1	2296988	2300452	-	hypothetical protein	-
BRPE64_RS10575	2	NC_021287.1	2300667	2301611	-	rhamnosyltransferase	General function prediction only
BRPE64_RS10580	2	NC_021287.1	2301632	2302519	-	dTDP-4-dehydrorhamnose reductase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10585	2	NC_021287.1	2302528	2303079	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10590	2	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10595	2	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS01390	3	NC_021287.1	288271	289794	*ilvA*	L-threonine ammonia-lyase	Amino acid transport and metabolism
BRPE64_RS01395	3	NC_021287.1	290326	294432	-	FAD linked oxidase domain protein	Energy production and conversion
BRPE64_RS02975	3	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS11750	3	NC_021287.1	2550034	2550774	-	phosphoglycolate phosphatase	General function prediction only
BRPE64_RS02980	4	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism

BRPE64_RS02985	4	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS02990	4	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS02995	4	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS03000	4	NC_021287.1	655612	656925	*cysN*	sulfate adenylyltransferase large subunit	Inorganic ion transport and metabolism
BRPE64_RS03005	4	NC_021287.1	656943	657698	-	uroporphyrin-III C-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS03010	4	NC_021287.1	657835	658215	-	cobalamin (Vitamin B12) biosynthesis CbiX protein	Function unknown
BRPE64_RS03975	4	NC_021287.1	863764	865017	-	extracellular solute-binding protein family 1	Carbohydrate transport and metabolism
BRPE64_RS07970	4	NC_021287.1	1735999	1737438	-	dihydrolipoyl dehydrogenase	Energy production and conversion
BRPE64_RS10125	4	NC_021287.1	2187950	2190004	-	NAD synthetase/Glutamine amidotransferase chain of NAD synthetase	Coenzyme transport and metabolism
BRPE64_RS11755	4	NC_021287.1	2551158	2552654	*trpE*	anthranilate synthase component I	Amino acid transport and metabolism
BRPE64_RS11760	4	NC_021287.1	2552667	2553254	-	glutamine amidotransferase of anthranilate synthase	Amino acid transport and metabolism
BRPE64_RS14010	4	NC_021287.1	3008347	3009735	-	tRNA modification GTPase MnmE	General function prediction only

Annexe 10: Lists of fitness genes involved in CCR480 peptide resistance in *B. insecticola* identified by Con-ARTIST.

Annexe 10.1: Genes identified for the lowest concentration (25 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS21955	1	NC_021288.1	254495	255832	-	acyl-CoA dehydrogenase domain protein	Lipid transport and metabolism
BRPE64_RS21945	2	NC_021288.1	252353	253501	-	outer membrane porin OmpC family	Cell wall/membrane/envelope biogenesis
BRPE64_RS21950	2	NC_021288.1	253610	254251	-	putative transcriptional regulator	Transcription
BRPE64_RS22155	2	NC_021288.1	291742	292509	-	NAD-dependent epimerase/dehydratase	Lipid transport and metabolism
BRPE64_RS22160	2	NC_021288.1	292855	293760	-	N-acetylneuraminate lyase	Amino acid transport and metabolism
BRPE64_RS22165	2	NC_021288.1	293807	294868	-	putative Glu/Leu/Phe/Val dehydrogenase	Amino acid transport and metabolism
BRPE64_RS22170	2	NC_021288.1	294922	295749	-	transcriptional regulator AraC family	Transcription
BRPE64_RS22175	2	NC_021288.1	295941	297044	-	putative transcriptional regulator Fis family	Transcription
BRPE64_RS22180	2	NC_021288.1	297041	297610	-	TetR family transcriptional regulator	Transcription
BRPE64_RS23075	2	NC_021288.1	512714	514834	-	glycogen debranching enzyme GlgX	Carbohydrate transport and metabolism
BRPE64_RS23080	2	NC_021288.1	514803	517304	-	phosphorylase	Carbohydrate transport and metabolism
BRPE64_RS23085	2	NC_021288.1	517492	518175	-	putative signal-transduction protein with CBS domains	General function prediction only
BRPE64_RS23090	2	NC_021288.1	518414	520756	-	small conductance mechanosensitive channel ion channel	Cell wall/membrane/envelope biogenesis
BRPE64_RS23525	2	NC_021288.1	637314	638576	-	efflux transporter RND family MFP subunit	Cell wall/membrane/envelope biogenesis
BRPE64_RS23530	2	NC_021288.1	639363	639701	-	transport-associated protein	General function prediction only
BRPE64_RS31990	2	NC_021288.1	727705	729990	-	hypothetical protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS23940	2	NC_021288.1	730076	730309	-	hypothetical protein	-
BRPE64_RS23945	2	NC_021288.1	730330	731085	-	hypothetical protein	-

Annexe 10.2: Genes identified for the highest concentration (100 µg.mL^{-1}).

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS12010	1	NC_021287.1	2604542	2605318	*tatC*	Sec-independent protein translocase TatC subunit	Intracellular trafficking, secretion, and vesicular transport

BRPE64_RS12280	1	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS21925	1	NC_021288.1	248477	249229	-	hypothetical protein	-
BRPE64_RS22190	1	NC_021288.1	298625	299338	-	NAD(P)H dehydrogenase quinone family	General function prediction only
BRPE64_RS22645	1	NC_021288.1	395257	396297	-	hypothetical protein	-
BRPE64_RS23075	1	NC_021288.1	512714	514834	-	glycogen debranching enzyme GlgX	Carbohydrate transport and metabolism
BRPE64_RS23100	1	NC_021288.1	522364	524328	-	hypothetical protein	-
BRPE64_RS12015	2	NC_021287.1	2605361	2605888	*tatB*	Sec-independent protein translocase protein TatB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12020	2	NC_021287.1	2605925	2606161	*tatA*	Sec-independent protein translocase protein TatA	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS12285	2	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS21855	2	NC_021288.1	234708	235418	-	ABC transporter related protein	General function prediction only
BRPE64_RS21860	2	NC_021288.1	235415	236122	-	ABC transporter related protein	Amino acid transport and metabolism
BRPE64_RS21865	2	NC_021288.1	236136	237122	-	acetamidase/Formamidase	Energy production and conversion
BRPE64_RS21870	2	NC_021288.1	237182	238336	-	ABC branched chain amino acid family transporter periplasmic ligand binding protein	Amino acid transport and metabolism
BRPE64_RS21875	2	NC_021288.1	238338	238988	-	response regulator receiver and ANTAR domain protein	Signal transduction mechanisms
BRPE64_RS21880	2	NC_021288.1	239019	239783	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS21885	2	NC_021288.1	239944	240759	-	transcriptional regulator DeoR family	Carbohydrate transport and metabolism
BRPE64_RS21890	2	NC_021288.1	240831	241796	-	PfkB domain protein	Carbohydrate transport and metabolism
BRPE64_RS21895	2	NC_021288.1	241789	243072	*kbaZ*	putative tagatose 6-phosphate kinase	Carbohydrate transport and metabolism
BRPE64_RS21900	2	NC_021288.1	243379	244284	-	probable sugar ABC transporter permease protein	Carbohydrate transport and metabolism
BRPE64_RS21905	2	NC_021288.1	244359	245306	-	putative sugar (D-ribose) ABC transporter	Carbohydrate transport and metabolism
BRPE64_RS21910	2	NC_021288.1	245354	246895	-	ABC transporter related	Carbohydrate transport and metabolism
BRPE64_RS21915	2	NC_021288.1	246960	247832	-	xylose isomerase domain-containing protein TIM barrel	Carbohydrate transport and metabolism
BRPE64_RS21920	2	NC_021288.1	247829	248458	-	NUDIX hydrolase	Nucleotide transport and metabolism

BRPE64_RS22155	2	NC_021288.1	291742	292509	-	NAD-dependent epimerase/dehydratase	Lipid transport and metabolism
BRPE64_RS22160	2	NC_021288.1	292855	293760	-	N-acetylneuraminate lyase	Amino acid transport and metabolism
BRPE64_RS22165	2	NC_021288.1	293807	294868	-	putative Glu/Leu/Phe/Val dehydrogenase	Amino acid transport and metabolism
BRPE64_RS22170	2	NC_021288.1	294922	295749	-	transcriptional regulator AraC family	Transcription
BRPE64_RS22175	2	NC_021288.1	295941	297044	-	putative transcriptional regulator Fis family	Transcription
BRPE64_RS22180	2	NC_021288.1	297041	297610	-	TetR family transcriptional regulator	Transcription
BRPE64_RS22185	2	NC_021288.1	297811	298542	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS22650	2	NC_021288.1	396290	397363	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS22655	2	NC_021288.1	397620	397844	-	hypothetical protein	-
BRPE64_RS22660	2	NC_021288.1	398148	398390	-	hypothetical protein	-
BRPE64_RS22665	2	NC_021288.1	398558	398986	-	Bll4598 protein	Function unknown
BRPE64_RS23080	2	NC_021288.1	514803	517304	-	phosphorylase	Carbohydrate transport and metabolism
BRPE64_RS23085	2	NC_021288.1	517492	518175	-	putative signal-transduction protein with CBS domains	General function prediction only
BRPE64_RS23090	2	NC_021288.1	518414	520756	-	small conductance mechanosensitive channel ion channel	Cell wall/membrane/envelope biogenesis
BRPE64_RS23095	2	NC_021288.1	521707	522336	-	hypothetical protein	General function prediction only
BRPE64_RS02785	3	NC_021287.1	614758	615954	-	fatty acid desaturase	Lipid transport and metabolism
BRPE64_RS02795	3	NC_021287.1	617201	618067	*nadC*	nicotinate-nucleotide pyrophosphorylase	Coenzyme transport and metabolism
BRPE64_RS02975	3	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS02980	3	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS02985	3	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS03000	3	NC_021287.1	655612	656925	*cysN*	sulfate adenylyltransferase large subunit	Inorganic ion transport and metabolism
BRPE64_RS05715	3	NC_021287.1	1238962	1241547	*glnD*	uridylyltransferase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS00270	4	NC_021287.1	59268	59876	-	methionine biosynthesis protein MetW	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS02790	4	NC_021287.1	616077	617204	*nadA*	quinolinate synthase A	Coenzyme transport and metabolism

BRPE64_RS02990	4	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS02995	4	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS03965	4	NC_021287.1	861787	862632	-	binding-protein-dependent transport systems inner membrane component	Carbohydrate transport and metabolism
BRPE64_RS03970	4	NC_021287.1	862634	863572	-	carbohydrate ABC transporter membrane protein 1 CUT1 family	Carbohydrate transport and metabolism
BRPE64_RS03975	4	NC_021287.1	863764	865017	-	extracellular solute-binding protein family 1	Carbohydrate transport and metabolism
BRPE64_RS07490	4	NC_021287.1	1621985	1624027	-	poly(R)-hydroxyalkanoic acid synthase class I	Lipid transport and metabolism

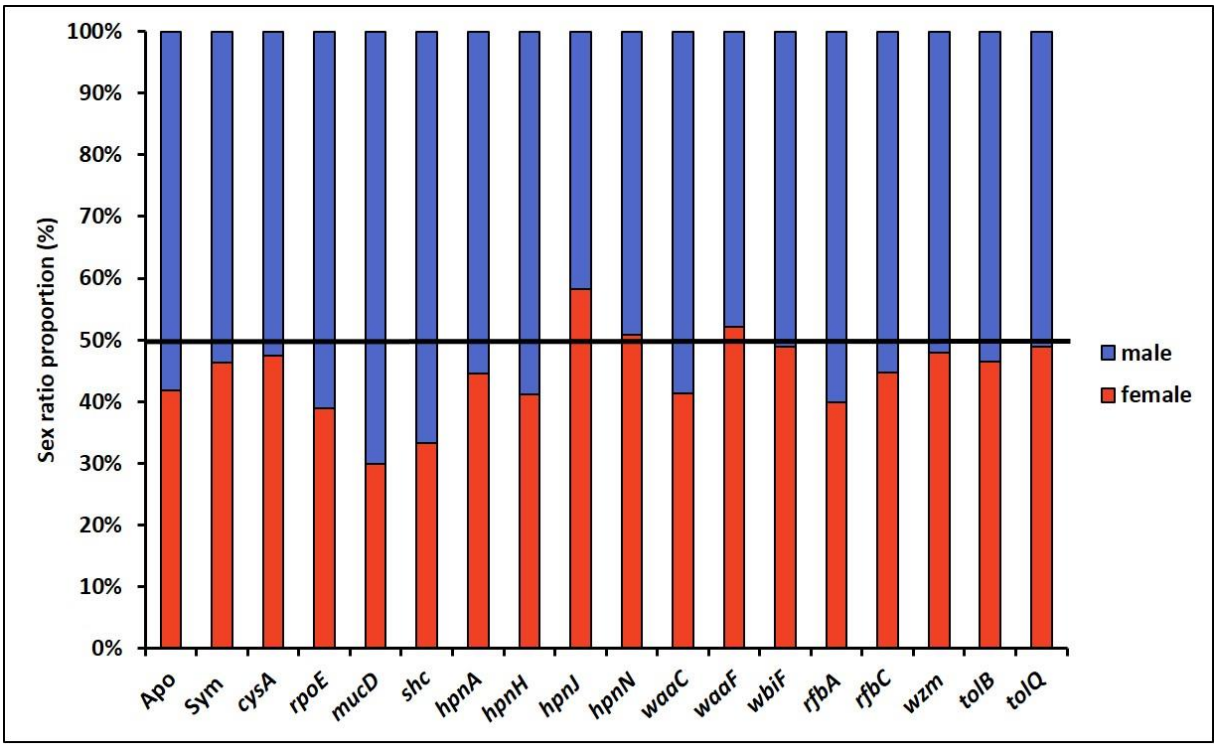
Annexe 11: List of putative Tat substrates in *B. insecticola* identified by the TATFIND 1.4 server (http://signalfind.org/tatfind.html).

Uniprot accession number	Gene tag	Replicon	Gene name	Gene product	Twin-arginine pattern	Hydrophobicity score
R4WES8	BRPE64_RS01050	NC_021287.1	-	lipoprotein	QRRNLL	3,13
R4WFP4	BRPE64_RS02880	NC_021287.1	-	alpha/beta hydrolase fold-3 domain protein	ARRRFI	4,89
R4WNH3	BRPE64_RS03205	NC_021287.1	*metG*	methionine--tRNA ligase	GRRQIL	1,87
R4WNI6	BRPE64_RS03275	NC_021287.1	-	hypothetical protein	ERRATI	0,71
R4WQB4	BRPE64_RS06490	NC_021287.1	-	hypothetical protein	NRRQFL	1,23
R4WQG6	BRPE64_RS06735	NC_021287.1	-	putative lipoprotein	QRRNFM	2,43
R4WWH5	BRPE64_RS07365	NC_021287.1	-	TonB-dependent receptor	ARRSAI	3,22
R4WH46	BRPE64_RS07430	NC_021287.1	-	hypothetical protein	NRRSIV	3,01
R4WYN6	BRPE64_RS07670	NC_021287.1	-	NMT1/THI5 like domain protein	KRRTFI	1,2
R4WR96	BRPE64_RS08145	NC_021287.1	-	aldehyde oxidase and xanthine dehydrogenase molybdopterin binding	SRRGFL	3,13
R4WS42	BRPE64_RS09535	NC_021287.1	*fabF*	3-oxoacyl-[acyl-carrier-protein] synthase 2	SRRRVV	3,7
R4WSE2	BRPE64_RS10165	NC_021287.1	-	urea ABC transporter urea binding protein	KRRSLL	-0,75
R4WI23	BRPE64_RS10210	NC_021287.1	-	amino acid ABC transporter substrate-binding protein PAAT family	ARRTTL	3,59
R4WI44	BRPE64_RS10785	NC_021287.1	*lpxK*	tetraacyldisaccharide 4'-kinase	RRRGPV	6,22
R4WZR9	BRPE64_RS10880	NC_021287.1	-	N-acetylmuramoyl-L-alanine amidase	RRRQVL	3,67
R4WY16	BRPE64_RS11095	NC_021287.1	-	LPS-assembly protein LptD	RRRRLV	7,02
R4WIG2	BRPE64_RS11155	NC_021287.1	-	rare lipoprotein B	SRRSFL	2,45
R4WIS9	BRPE64_RS11865	NC_021287.1	-	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	ARRRLI	4,92
R4WJ02	BRPE64_RS12235	NC_021287.1	-	hypothetical protein	RRRTAL	3,27
R4X0B1	BRPE64_RS12275	NC_021287.1	*yedY*	sulfoxide reductase catalytic subunit YedY	NRRRVL	0,93
R4WJ60	BRPE64_RS12720	NC_021287.1	-	hypothetical protein	SRRTFL	3,34
R4X112	BRPE64_RS14270	NC_021294.1	-	alkaline phosphatase	SRRALL	-1,12
R4WV33	BRPE64_RS15055	NC_021294.1	-	putative beta-lactamase	GRRRFL	4,52
R4X1B2	BRPE64_RS15095	NC_021294.1	-	alkaline phosphatase	DRRRFI	5,26
R4WK66	BRPE64_RS15325	NC_021294.1	*kgtP*	metabolite/H+ symporter major facilitator superfamily	TRRRVF	3,42
R4WK79	BRPE64_RS15390	NC_021294.1	-	hypothetical protein	TRRQFL	1,81
R4WZL9	BRPE64_RS15550	NC_021294.1	-	(2Fe-2S)-binding domain protein	SRRRFL	0,92
R4WVQ9	BRPE64_RS16245	NC_021294.1	-	hypothetical protein	ARRRVI	7,8
R4X261	BRPE64_RS17610	NC_021294.1	-	hypothetical protein	SRRGAM	7,01
R4X271	BRPE64_RS17705	NC_021294.1	-	methionine-R-sulfoxide reductase	TRRRFL	3,87
R4WWP9	BRPE64_RS18140	NC_021294.1	-	glyoxalase/bleomycin resistance protein/dioxygenase	DRRGVI	4,35
R4X0Z8	BRPE64_RS19530	NC_021294.1	-	phospholipase C phosphocholine-specific	SRRRFL	1,7
R4WMY2	BRPE64_RS19770	NC_021294.1	-	carboxymethylenebutenolidase	NRRTFI	1,82

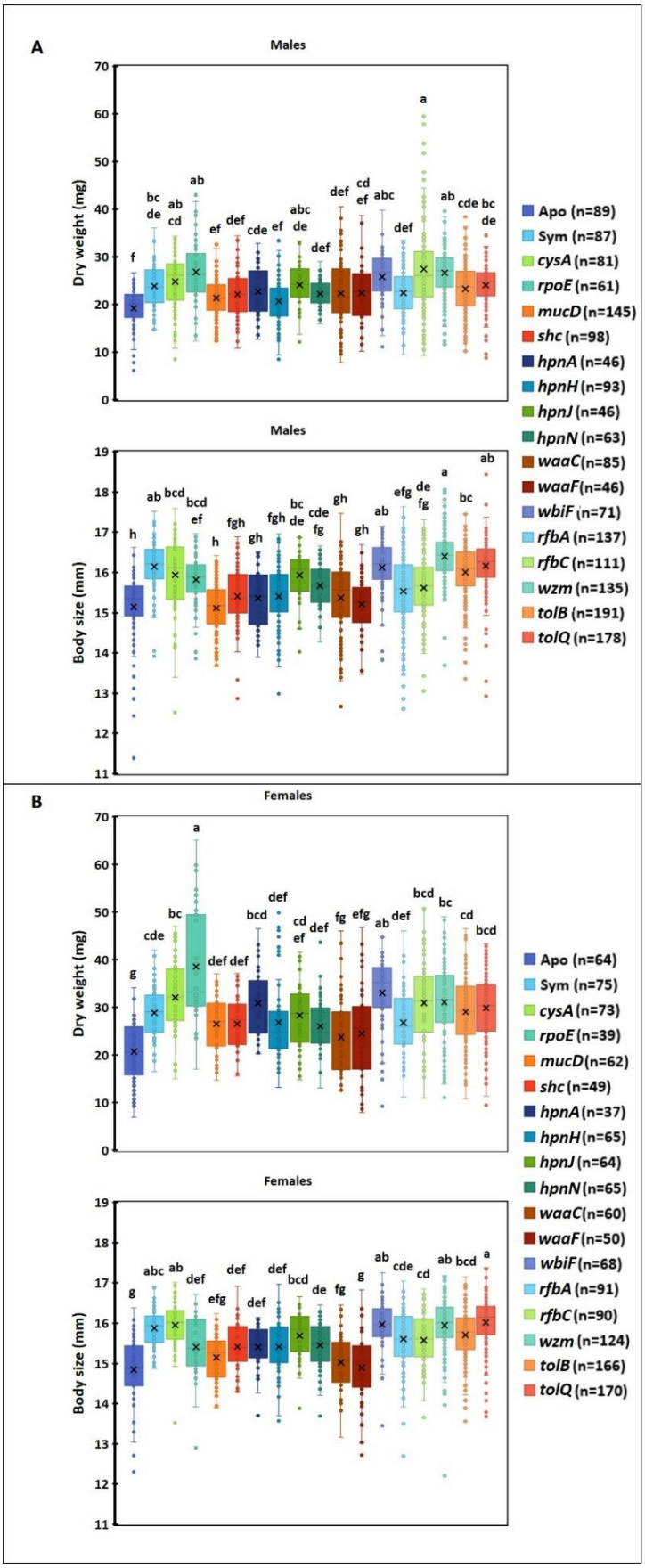
R4WNG3	BRPE64_RS20735	NC_021288.1	-	major facilitator superfamily MFS_1	DRRQAL	2,78
R4WXT4	BRPE64_RS20795	NC_021288.1	-	fumarate reductase/succinate dehydrogenase flavoprotein domain protein	SRRNFI	-1,66
R4X374	BRPE64_RS20855	NC_021288.1	-	ABC-type sugar transport system periplasmic component-like protein	TRRGLM	0,32
R4X3K0	BRPE64_RS22405	NC_021288.1	-	dioxygenase	RRRDFL	5,62
R4WP83	BRPE64_RS22555	NC_021288.1	-	amidohydrolase 2	KRREAL	1,86
R4X257	BRPE64_RS22820	NC_021288.1	-	TRAP dicarboxylate transporter DctP subunit	DRRTFL	1,36
R4X3N3	BRPE64_RS22850	NC_021288.1	-	hypothetical protein	SRRTFL	2,81
R4WPK7	BRPE64_RS23170	NC_021288.1	-	NAD(P) transhydrogenase subunit beta	ARRGNL	7,11
R4WYQ1	BRPE64_RS23420	NC_021288.1	-	extracellular solute-binding protein family 1	QRRRIV	5,49
R4WQ27	BRPE64_RS23505	NC_021288.1	-	beta-lactamase	GRRQFL	0,74
R4WYU0	BRPE64_RS23700	NC_021288.1	-	extracellular solute-binding protein family 5	SRRNVL	0,08
R4WYV4	BRPE64_RS23725	NC_021288.1	-	dipeptide ABC transporter periplasmic component	GRRKAM	4,65
R4WYX1	BRPE64_RS23840	NC_021288.1	-	extracellular solute-binding protein family 1	ARRRIL	3,18
R4WQ51	BRPE64_RS23960	NC_021288.1	-	secretion protein HlyD family protein	RRRRTL	6,98
R4WQ55	BRPE64_RS23985	NC_021288.1	-	aldehyde oxidase and xanthine dehydrogenase molybdopterin binding	ARRRFI	0,97
R4X3Y0	BRPE64_RS24150	NC_021288.1	-	ABC spermidine/putrescine transporter periplasmic ligand binding protein	SRRTFI	3,01
R4X3Y8	BRPE64_RS24240	NC_021288.1	*ytfQ*	periplasmic binding protein/LacI transcriptional regulator	KRRNVL	3,98
R4WZ57	BRPE64_RS24685	NC_021288.1	-	phospholipase C phosphocholine-specific	NRRDFL	2,87
R4X445	BRPE64_RS24875	NC_021289.1	-	TRAP dicarboxylate transporter-DctP subunit	SRRRFI	1,72
R4WRM2	BRPE64_RS25990	NC_021289.1	-	transcriptional regulator LacI family	QRRPTM	0,46
R4WZT7	BRPE64_RS26230	NC_021289.1	-	putative ABC transporter substrate-binding protein	TRRDVM	1,69
R4X3C1	BRPE64_RS26395	NC_021289.1	-	isoquinoline 1-oxidoreductase beta subunit	SRRAFL	-0,34
R4X4M1	BRPE64_RS26945	NC_021289.1	-	isocitrate lyase and phosphorylmutase	ARRALL	3,22
R4X3K3	BRPE64_RS27260	NC_021289.1	-	aliphatic sulfonates family ABC transporter periplasmic ligand-binding protein	SRRRAL	1,19
R4X0G5	BRPE64_RS28110	NC_021289.1	-	glycoside hydrolase family 28	TRRTFV	1,43
R4X3Z5	BRPE64_RS28740	NC_021289.1	-	hypothetical protein	RRRRLF	1,43
R4X524	BRPE64_RS28860	NC_021289.1	-	epoxide hydrolase domain protein	SRRRFI	3,52
R4X460	BRPE64_RS29520	NC_021289.1	-	hypothetical protein	SRRKAL	4,82
R4X184	BRPE64_RS30170	NC_021289.1	*napA*	periplasmic nitrate reductase	TRRAFI	-2,46
R4WU66	BRPE64_RS30380	NC_021289.1	-	putative sensor with GAF domain	QRRALI	6,42
A0A060PJ94	BRPE64_RS30900	NC_021295.1	-	hypothetical protein	GRRQAI	0,9

Annexe 12: Sex ratio of adult insects' cohorts infected with each *Burkholderia* mutant strain.

The sex ratio proportion was calculated according to the gender of each adult insect mono-infected by each *B. insecticola* mutant strain, with males indicated in blue and females indicated in red. The black line indicates equivalent theoretical proportions of males (50%) and females (50%). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.

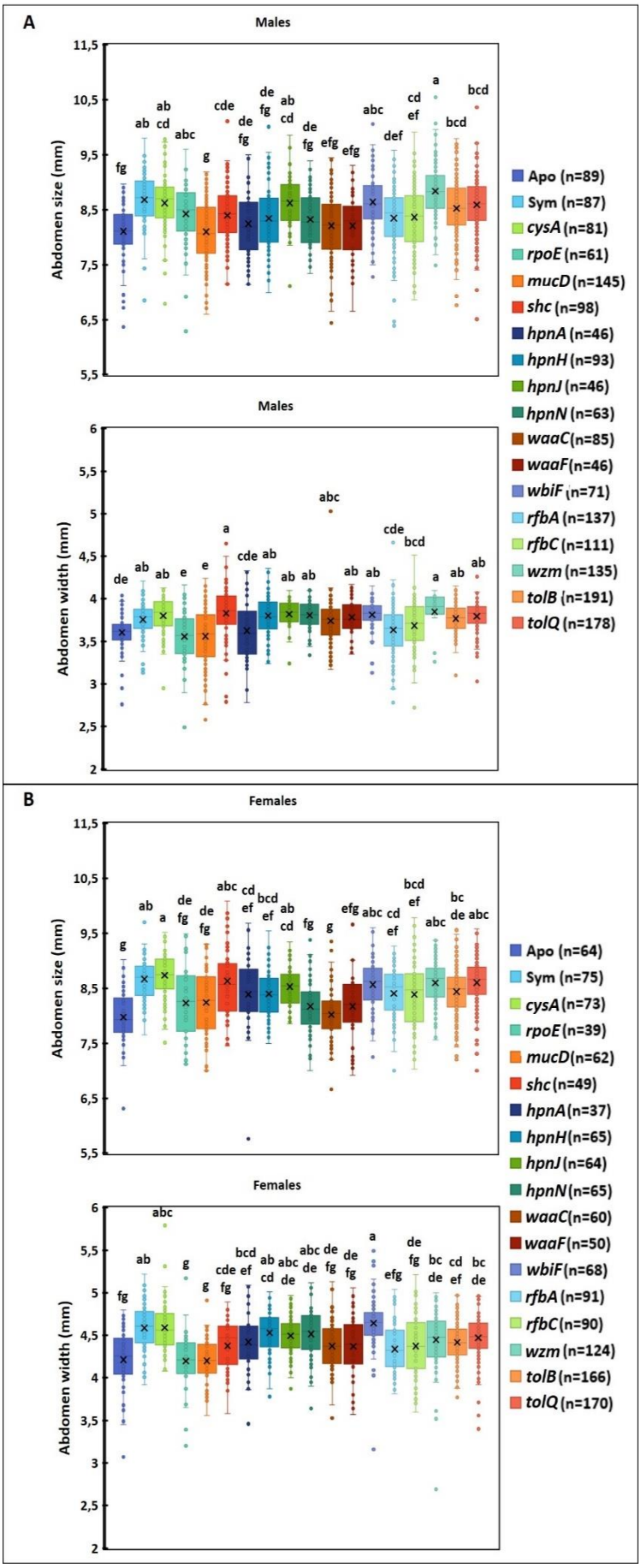


Annexe 13: Fitness parameters of *R. pedestris* males and females



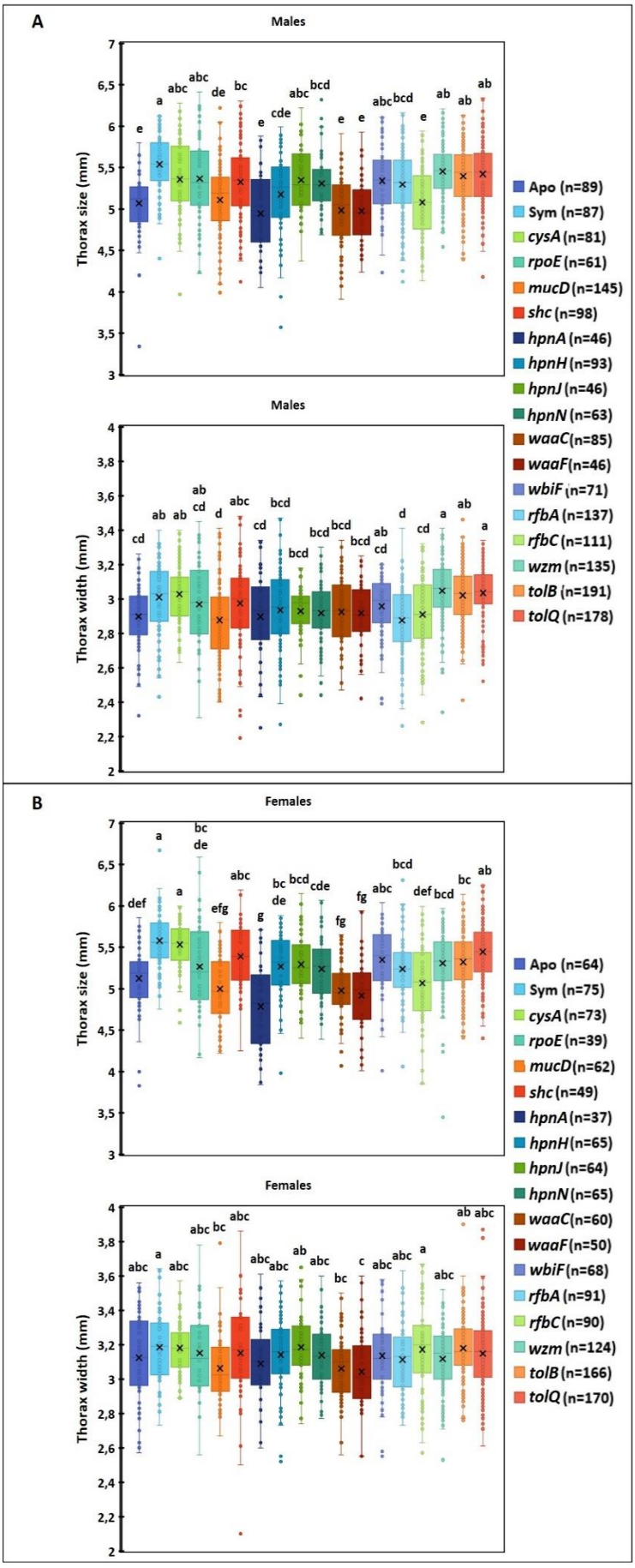
Annexe 13.1: Effects of the *Burkholderia* mutant strains on the body weight and body size of *R. pedestris* male and female adult insects.

A) Males. B) Females. Dry weight and body size were measured for each adult insect mono-infected by each *Burkholderia* strain. The mean values are indicated by a black cross on each boxplot. The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters on the top of each boxplot indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey correction). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.



Annexe 13.2: Effects of the *Burkholderia* mutant strains on the abdomen size and width of *R. pedestris* male and female adult insects.

A) Males. B) Females. Abdomen size and width were measured for each adult insect mono-infected by each *Burkholderia* strain. The mean values are indicated by a black cross on each boxplot. The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters on the top of each boxplot indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey correction). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.



Annexe 13.3: Effects of the *Burkholderia* mutant strains on the thorax size and width of *R. pedestris* male and female adult insects.

A) Males. B) Females. Thorax size and width were measured for each adult insect mono-infected by each *Burkholderia* strain. The mean values are indicated by a black cross on each boxplot. The number of insects indicated in parentheses for each condition (n) represents the pooled number of insects used in the three independent experiments. Different letters on the top of each boxplot indicate statistically significant differences (p-value < 0.05, one-way ANOVA with Tukey correction). Abbreviations: Apo: aposymbiotic insects, Sym: symbiotic insects.

Annexe 14: List of fitness genes in *B. insecticola* involved in the colonization of the M1 organ identified by Con-ARTIST.

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS01640	1	NC_021287.1	346795	347853	*ruvB*	Holliday junction ATP-dependent DNA helicase RuvB	Replication, recombination and repair
BRPE64_RS01665	1	NC_021287.1	351036	352130	*dusB*	tRNA-dihydrouridine synthase	Translation, ribosomal structure and biogenesis
BRPE64_RS02140	1	NC_021287.1	462006	462953	*miaA*	tRNA dimethylallyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS02300	1	NC_021287.1	495263	496360	-	lipopolysaccharide heptosyltransferase II	Cell wall/membrane/envelope biogenesis
BRPE64_RS02340	1	NC_021287.1	504058	504948	*purC*	phosphoribosylaminoimidazole-succinocarboxamidesynthase	Nucleotide transport and metabolism
BRPE64_RS04650	1	NC_021287.1	1020007	1022082	*recD*	exodeoxyribonuclease V alpha subunit	Replication, recombination and repair
BRPE64_RS04905	1	NC_021287.1	1075733	1076368	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms
BRPE64_RS06390	1	NC_021287.1	1369578	1370900	-	homoserine dehydrogenase	Amino acid transport and metabolism
BRPE64_RS07495	1	NC_021287.1	1624348	1625181	*yfiH*	hypothetical protein	Function unknown
BRPE64_RS09935	1	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS09970	1	NC_021287.1	2149002	2149682	*cmk*	cytidylate kinase	Nucleotide transport and metabolism
BRPE64_RS11035	1	NC_021287.1	2389995	2391107	-	protein TolA	Cell wall/membrane/envelope biogenesis
BRPE64_RS11215	1	NC_021287.1	2430339	2431349	-	glyceraldehyde-3-phosphate dehydrogenase type I	Carbohydrate transport and metabolism
BRPE64_RS12120	1	NC_021287.1	2620349	2620900	-	putative signal peptide protein toluene tolerance Ttg2C-like protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS01645	2	NC_021287.1	347911	348492	*ruvA*	Holliday junction ATP-dependent DNA helicase RuvA	Replication, recombination and repair
BRPE64_RS01650	2	NC_021287.1	348517	349059	*ruvC*	crossover junction endodeoxyribonuclease RuvC	Replication, recombination and repair
BRPE64_RS01655	2	NC_021287.1	349176	350741	*purH*	bifunctional purine biosynthesis protein PurH	Nucleotide transport and metabolism
BRPE64_RS01660	2	NC_021287.1	350806	351039	*fis*	DNA-binding protein Fis	Transcription
BRPE64_RS02145	2	NC_021287.1	462950	464950	*mutL*	DNA mismatch repair protein MutL	Replication, recombination and repair

BRPE64_RS02150	2	NC_021287.1	465159	465821	*dedA*	membrane-associated protein	Function unknown
BRPE64_RS02345	2	NC_021287.1	505002	505523	*purE*	N5-carboxyaminoimidazole ribonucleotide mutase	Nucleotide transport and metabolism
BRPE64_RS02350	2	NC_021287.1	505612	506799	*purK*	phosphoribosylaminoimidazole carboxylase ATPase subunit	Nucleotide transport and metabolism
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane/envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane/envelope biogenesis
BRPE64_RS04500	2	NC_021287.1	977887	979164	-	O-antigen polymerase	Cell wall/membrane/envelope biogenesis
BRPE64_RS04640	2	NC_021287.1	1012890	1016297	*recC*	exodeoxyribonuclease V gamma subunit	Replication, recombination and repair
BRPE64_RS04645	2	NC_021287.1	1016294	1020010	*recB*	exodeoxyribonuclease V beta subunit	Replication, recombination and repair
BRPE64_RS04910	2	NC_021287.1	1076365	1078893	-	multi-sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS06385	2	NC_021287.1	1368315	1369553	-	aminotransferase AlaT	Amino acid transport and metabolism
BRPE64_RS06595	2	NC_021287.1	1416832	1420869	*purL*	phosphoribosylformylglycinamidine synthase	Nucleotide transport and metabolism
BRPE64_RS07500	2	NC_021287.1	1625178	1626218	*rluD*	pseudouridine synthase	Translation, ribosomal structure and biogenesis
BRPE64_RS07700	2	NC_021287.1	1671421	1671747	-	thioredoxin	Energy production and conversion
BRPE64_RS09020	2	NC_021287.1	1954275	1956629	*uvrD*	UvrD/REP helicase	Replication, recombination and repair
BRPE64_RS09940	2	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS09975	2	NC_021287.1	2149758	2151062	*aroA*	3-phosphoshikimate 1-carboxyvinyltransferase	Amino acid transport and metabolism
BRPE64_RS10555	2	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10590	2	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS10595	2	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion,

							and vesicular transport
BRPE64_RS11045	2	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane/envelope biogenesis
BRPE64_RS11050	2	NC_021287.1	2393115	2393864	-	Tol-pal system protein YbgF	Function unknown
BRPE64_RS11220	2	NC_021287.1	2431398	2433449	-	transketolase 1	Carbohydrate transport and metabolism
BRPE64_RS11745	2	NC_021287.1	2549348	2550037	*rpe*	ribulose-phosphate 3-epimerase	Carbohydrate transport and metabolism
BRPE64_RS12125	2	NC_021287.1	2620994	2621761	-	hypothetical protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12130	2	NC_021287.1	2621758	2622573	-	ABC transporter related	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12280	2	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12285	2	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS13660	2	NC_021287.1	2940847	2941263	*dksA*	transcriptional regulator TraR/DksA family	Signal transduction mechanisms
BRPE64_RS14010	2	NC_021287.1	3008347	3009735	-	tRNA modification GTPase MnmE	General function prediction only

Annexe 15: List of fitness genes in *B. insecticola* involved in the colonization of the M3 organ identified by Con-ARTIST.

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS09940	1	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane/envelope biogenesis
BRPE64_RS10590	1	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11045	1	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane/envelope biogenesis
BRPE64_RS11745	1	NC_021287.1	2549348	2550037	*rpe*	ribulose-phosphate 3-epimerase	Carbohydrate transport and metabolism
BRPE64_RS00910	2	NC_021287.1	193232	193723	-	PTS system fructose subfamily IIA component	Carbohydrate transport and metabolism
BRPE64_RS01090	2	NC_021287.1	227228	227761	-	hypothetical protein	-
BRPE64_RS01640	2	NC_021287.1	346795	347853	*ruvB*	Holliday junction ATP-dependent DNA helicase RuvB	Replication, recombination and repair
BRPE64_RS01645	2	NC_021287.1	347911	348492	*ruvA*	Holliday junction ATP-dependent DNA helicase RuvA	Replication, recombination and repair
BRPE64_RS01650	2	NC_021287.1	348517	349059	*ruvC*	crossover junction endodeoxyribonuclease RuvC	Replication, recombination and repair
BRPE64_RS01655	2	NC_021287.1	349176	350741	*purH*	bifunctional purine biosynthesis protein PurH	Nucleotide transport and metabolism
BRPE64_RS01660	2	NC_021287.1	350806	351039	*fis*	DNA-binding protein Fis	Transcription
BRPE64_RS09020	2	NC_021287.1	1954275	1956629	*uvrD*	UvrD/REP helicase	Replication, recombination and repair
BRPE64_RS09935	2	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS10555	2	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10595	2	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11220	2	NC_021287.1	2431398	2433449	-	transketolase 1	Carbohydrate transport and metabolism
BRPE64_RS12280	2	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones

BRPE64_RS12285	2	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS13660	2	NC_021287.1	2940847	2941263	*dksA*	transcriptional regulator TraR/DksA family	Signal transduction mechanisms
BRPE64_RS20665	2	NC_021294.1	1449187	1450284	-	N-acylglucosamine 2-epimerase	Carbohydrate transport and metabolism
BRPE64_RS17885	3	NC_021294.1	809675	812734	-	cyclic nucleotide-regulated ABC bacteriocin/lantibiotic exporter	Defense mechanisms
BRPE64_RS18525	3	NC_021294.1	949221	950699	-	aldehyde Dehydrogenase	Energy production and conversion
BRPE64_RS26860	3	NC_021289.1	496843	497250	-	cupin 2 conserved barrel domain protein	Function unknown
BRPE64_RS17890	4	NC_021294.1	812752	813504	-	PpiC-type peptidyl-prolyl cis-trans isomerase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS17895	4	NC_021294.1	813577	814950	-	ABC efflux pump membrane fusion protein HlyD subfamily	Cell wall/membrane/envelope biogenesis
BRPE64_RS17900	4	NC_021294.1	814928	817453	-	putative forkhead-associated protein	Signal transduction mechanisms
BRPE64_RS18530	4	NC_021294.1	950850	951329	-	transcriptional regulator AsnC family	Transcription
BRPE64_RS18535	4	NC_021294.1	951329	952354	-	ectoine utilization protein EutE	General function prediction only
BRPE64_RS18540	4	NC_021294.1	952359	953576	-	ectoine utilization protein EutD	Amino acid transport and metabolism
BRPE64_RS18545	4	NC_021294.1	953588	954598	-	ectoine utilization protein EutC	Amino acid transport and metabolism
BRPE64_RS18550	4	NC_021294.1	954603	955568	-	ectoine utilization protein EutB	Amino acid transport and metabolism
BRPE64_RS18555	4	NC_021294.1	955649	957046	-	transcriptional regulator GntR family with aminotransferase domain	Amino acid transport and metabolism
BRPE64_RS18560	4	NC_021294.1	957281	958135	-	ectoine/hydroxyectoine ABC transporter solute-binding protein	Amino acid transport and metabolism
BRPE64_RS18565	4	NC_021294.1	958216	958869	-	beta tubulin autoregulation binding site	Amino acid transport and metabolism
BRPE64_RS18570	4	NC_021294.1	958866	959534	-	amino acid ABC transporter permease protein 3-TM region His/Glu/Gln/Arg/opine	Amino acid transport and metabolism
BRPE64_RS18575	4	NC_021294.1	959587	960432	-	ectoine/hydroxyectoine ABC transporter ATP-binding protein	Amino acid transport and metabolism
BRPE64_RS26865	4	NC_021289.1	497349	498362	-	transcriptional regulator LysR family	Transcription

BRPE64_RS26870	4	NC_021289.1	498603	499121	-	OsmC family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS26875	4	NC_021289.1	499134	500522	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS26880	4	NC_021289.1	500586	501356	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26885	4	NC_021289.1	501370	501708	-	hypothetical protein	-
BRPE64_RS26890	4	NC_021289.1	501705	503450	-	fumarate reductase/succinate dehydrogenase flavoprotein domain protein	Energy production and conversion
BRPE64_RS26895	4	NC_021289.1	503505	504344	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS26900	4	NC_021289.1	504480	505373	-	transcriptional regulator AraC family	Transcription
BRPE64_RS26905	4	NC_021289.1	505496	507145	*treA*	alpha alpha-trehalase	Carbohydrate transport and metabolism
BRPE64_RS26910	4	NC_021289.1	507160	507696	-	cytochrome c class I	Energy production and conversion
BRPE64_RS26915	4	NC_021289.1	507689	508900	-	oxidoreductase molybdopterin binding protein	General function prediction only
BRPE64_RS26920	4	NC_021289.1	509169	509405	-	4-oxalocrotonate tautomerase family enzyme	General function prediction only

Annexe 16: List of fitness genes in *B. insecticola* involved in the colonization of the M4 organ at the second instar stage identified by Con-ARTIST.

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS00105	1	NC_021287.1	20466	21482	-	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding protein	Amino acid transport and metabolism
BRPE64_RS00515	1	NC_021287.1	113499	113798	-	hypothetical protein	-
BRPE64_RS01090	1	NC_021287.1	227228	227761	-	hypothetical protein	-
BRPE64_RS02350	1	NC_021287.1	505612	506799	*purK*	phosphoribosylaminoimidazole carboxylase ATPase subunit	Nucleotide transport and metabolism
BRPE64_RS03250	1	NC_021287.1	708925	709740	-	transcriptional regulator XRE family	Transcription
BRPE64_RS03750	1	NC_021287.1	812442	813461	-	fructose-1,6-bisphosphatase class 1,1	Carbohydrate transport and metabolism
BRPE64_RS05530	1	NC_021287.1	1195784	1197016	-	putative exported lipoprotein	Cell wall/membrane /envelope biogenesis
BRPE64_RS05810	1	NC_021287.1	1260186	1262603	-	phosphoenolpyruvate synthase	Carbohydrate transport and metabolism
BRPE64_RS06035	1	NC_021287.1	1308835	1311414	-	putative penicillin-binding (Peptidoglycan synthetase) transmembrane protein mrcA	Cell wall/membrane /envelope biogenesis
BRPE64_RS06685	1	NC_021287.1	1439287	1440297	-	aminodeoxychorismate lyase	General function prediction only
BRPE64_RS08840	1	NC_021287.1	1916590	1917606	*ilvC*	ketol-acid reductoisomerase	Amino acid transport and metabolism
BRPE64_RS09400	1	NC_021287.1	2038614	2039366	*otsB*	trehalose 6-phosphatase	Carbohydrate transport and metabolism
BRPE64_RS10555	1	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10585	1	NC_021287.1	2302528	2303079	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	Cell wall/membrane /envelope biogenesis
BRPE64_RS10595	1	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane /envelope biogenesis
BRPE64_RS11025	1	NC_021287.1	2388863	2389540	*tolQ*	protein TolQ	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11050	1	NC_021287.1	2393115	2393864	-	Tol-pal system protein YbgF	Function unknown
BRPE64_RS11090	1	NC_021287.1	2402511	2403872	*surA*	chaperone SurA	Posttranslational modification, protein turnover, chaperones
BRPE64_RS11215	1	NC_021287.1	2430339	2431349	-	glyceraldehyde-3-phosphate dehydrogenase type I	Carbohydrate transport and metabolism

BRPE64_RS12120	1	NC_021287.1	2620349	2620900	-	putative signal peptide protein toluene tolerance Ttg2C-like protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12160	1	NC_021287.1	2628391	2629857	*gltD*	glutamate synthase (NADH) small subunit	Amino acid transport and metabolism
BRPE64_RS13045	1	NC_021287.1	2801341	2801760	*fliS*	flagellar protein FliS	Cell motility
BRPE64_RS13115	1	NC_021287.1	2814264	2815058	*fliR*	flagellar biosynthetic protein fliR	Cell motility
BRPE64_RS13140	1	NC_021287.1	2822193	2823242	*flgJ*	flagellar rod assembly protein/muramidase FlgJ	Cell motility
BRPE64_RS13515	1	NC_021287.1	2908014	2908658	-	transcriptional regulator MarR family	Transcription
BRPE64_RS17380	1	NC_021294.1	711888	712550	-	hypothetical protein	-
BRPE64_RS20025	1	NC_021294.1	1307649	1308182	*cvpA*	putative bacteriocin production related protein	General function prediction only
BRPE64_RS20045	1	NC_021294.1	1311424	1312224	*trpA*	tryptophan synthase alpha chain	Amino acid transport and metabolism
BRPE64_RS20075	1	NC_021294.1	1318091	1319212	*asd*	aspartate-semialdehyde dehydrogenase	Amino acid transport and metabolism
BRPE64_RS24345	1	NC_021288.1	818932	820779	-	hypothetical protein	-
BRPE64_RS00275	2	NC_021287.1	59873	61018	-	homoserine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS00285	2	NC_021287.1	62059	62958	-	acetylglutamate kinase	Amino acid transport and metabolism
BRPE64_RS00520	2	NC_021287.1	113829	115370	*fliD*	flagellar hook-associated 2 domain protein	Cell motility
BRPE64_RS00525	2	NC_021287.1	115565	117082	*fliC*	flagellin	Cell motility
BRPE64_RS00600	2	NC_021287.1	130402	131262	-	chemotaxis protein MotA	Cell motility
BRPE64_RS00605	2	NC_021287.1	131312	132295	-	chemotaxis protein MotB	Cell motility
BRPE64_RS00610	2	NC_021287.1	132377	132754	-	response regulator receiver protein	Signal transduction mechanisms
BRPE64_RS00615	2	NC_021287.1	132805	135117	-	CheA Signal Transduction Histidine Kinases	Cell motility
BRPE64_RS00620	2	NC_021287.1	135163	135690	-	CheW protein	Cell motility
BRPE64_RS00625	2	NC_021287.1	135709	137559	*tsr*	methyl-accepting chemotaxis sensory transducer	Cell motility
BRPE64_RS00630	2	NC_021287.1	137713	138636	-	MCP methyltransferase CheR-type	Cell motility
BRPE64_RS00635	2	NC_021287.1	138633	139376	*cheD*	probable chemoreceptor glutamine deamidase CheD	Cell motility
BRPE64_RS00640	2	NC_021287.1	139373	140470	-	chemotaxis response regulator protein-glutamate methylesterase	Cell motility
BRPE64_RS00645	2	NC_021287.1	140520	140915	-	response regulator receiver protein	Signal transduction mechanisms
BRPE64_RS00650	2	NC_021287.1	140918	141646	*cheZ*	protein phosphatase CheZ	Cell motility
BRPE64_RS00670	2	NC_021287.1	144624	145271	*dsbA*	thiol disulfide interchange protein DsbA	Energy production and conversion
BRPE64_RS00675	2	NC_021287.1	145458	146231	-	sporulation domain protein	Cell cycle control, cell division,

							chromosome partitioning
BRPE64_RS00755	2	NC_021287.1	163567	164049	-	RfaE bifunctional protein	Cell wall/membrane /envelope biogenesis
BRPE64_RS00760	2	NC_021287.1	164063	164302	-	hypothetical protein	-
BRPE64_RS01395	2	NC_021287.1	290326	294432	-	FAD linked oxidase domain protein	Energy production and conversion
BRPE64_RS01725	2	NC_021287.1	364448	365170	-	hypothetical protein	Function unknown
BRPE64_RS01730	2	NC_021287.1	365167	366759	-	putative paraquat-inducible protein	General function prediction only
BRPE64_RS01735	2	NC_021287.1	366842	367525	-	paraquat-inducible protein A	Function unknown
BRPE64_RS01740	2	NC_021287.1	367522	368199	-	paraquat-inducible protein A	Function unknown
BRPE64_RS02135	2	NC_021287.1	460885	461940	*purM*	phosphoribosylformylglycinamidine cyclo-ligase	Nucleotide transport and metabolism
BRPE64_RS02140	2	NC_021287.1	462006	462953	*miaA*	tRNA dimethylallyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS02145	2	NC_021287.1	462950	464950	*mutL*	DNA mismatch repair protein MutL	Replication, recombination and repair
BRPE64_RS02300	2	NC_021287.1	495263	496360	-	lipopolysaccharide heptosyltransferase II	Cell wall/membrane /envelope biogenesis
BRPE64_RS02325	2	NC_021287.1	499591	500787	*pgk*	phosphoglycerate kinase	Carbohydrate transport and metabolism
BRPE64_RS02405	2	NC_021287.1	517090	518598	-	protease Do	Posttranslational modification, protein turnover, chaperones
BRPE64_RS02505	2	NC_021287.1	545166	545954	*znuB*	cation ABC transporter permease	Inorganic ion transport and metabolism
BRPE64_RS02510	2	NC_021287.1	545947	546855	*znuC*	ABC Mn2+/Zn2+ transporter ATPase subunit	Inorganic ion transport and metabolism
BRPE64_RS02515	2	NC_021287.1	546852	547733	*znuA*	periplasmic solute binding protein	Inorganic ion transport and metabolism
BRPE64_RS02520	2	NC_021287.1	547781	548239	-	transcriptional regulator Fur family	Inorganic ion transport and metabolism
BRPE64_RS03245	2	NC_021287.1	707442	708845	*argH*	argininosuccinate lyase	Amino acid transport and metabolism
BRPE64_RS03460	2	NC_021287.1	756139	757398	-	probable multidrug resistance protein	Amino acid transport and metabolism
BRPE64_RS03465	2	NC_021287.1	757460	758707	-	hypothetical protein	Lipid transport and metabolism
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane /envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane

							/envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane /envelope biogenesis
BRPE64_RS04500	2	NC_021287.1	977887	979164	-	O-antigen polymerase	Cell wall/membrane /envelope biogenesis
BRPE64_RS06475	2	NC_021287.1	1390679	1392052	-	integral membrane sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS06485	2	NC_021287.1	1393110	1393475	-	hypothetical protein	-
BRPE64_RS06530	2	NC_021287.1	1401612	1402265	*clpP*	ATP-dependent Clp protease proteolytic subunit	Posttranslational modification, protein turnover, chaperones
BRPE64_RS06535	2	NC_021287.1	1402436	1403707	*clpX*	ATP-dependent Clp protease ATP-binding subunit ClpX	Posttranslational modification, protein turnover, chaperones
BRPE64_RS06540	2	NC_021287.1	1403894	1406317	*lon*	Lon protease	Posttranslational modification, protein turnover, chaperones
BRPE64_RS06595	2	NC_021287.1	1416832	1420869	*purL*	phosphoribosylformylglycinamidine synthase	Nucleotide transport and metabolism
BRPE64_RS08845	2	NC_021287.1	1917676	1918167	-	acetolactate synthase small subunit	Amino acid transport and metabolism
BRPE64_RS08850	2	NC_021287.1	1918274	1920037	-	acetolactate synthase large subunit biosynthetic type	Amino acid transport and metabolism
BRPE64_RS09205	2	NC_021287.1	1996687	1997517	-	undecaprenyl-diphosphatase	Defense mechanisms
BRPE64_RS09935	2	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS09940	2	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane /envelope biogenesis
BRPE64_RS10075	2	NC_021287.1	2175984	2177834	-	hypothetical protein	General function prediction only
BRPE64_RS10300	2	NC_021287.1	2228407	2229402	*waaC*	lipopolysaccharide heptosyltransferase I	Cell wall/membrane /envelope biogenesis
BRPE64_RS10475	2	NC_021287.1	2273911	2275809	-	polysaccharide biosynthesis protein CapD	Carbohydrate transport and metabolism
BRPE64_RS10480	2	NC_021287.1	2275822	2276847	-	glycosyl transferase family 4	Cell wall/membrane /envelope biogenesis
BRPE64_RS10485	2	NC_021287.1	2277059	2278042	-	NAD-dependent epimerase/dehydratase	Carbohydrate transport and metabolism
BRPE64_RS10490	2	NC_021287.1	2278039	2278902	-	putative glycosyl transferase	General function prediction only

BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10565	2	NC_021287.1	2295122	2296966	-	glycosyl transferase family 2	Cell wall/membrane /envelope biogenesis
BRPE64_RS10570	2	NC_021287.1	2296988	2300452	-	hypothetical protein	-
BRPE64_RS10575	2	NC_021287.1	2300667	2301611	-	rhamnosyltransferase	General function prediction only
BRPE64_RS10580	2	NC_021287.1	2301632	2302519	-	dTDP-4-dehydrorhamnose reductase	Cell wall/membrane /envelope biogenesis
BRPE64_RS10590	2	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane /envelope biogenesis
BRPE64_RS10830	2	NC_021287.1	2355573	2356502	-	ornithine carbamoyltransferase	Amino acid transport and metabolism
BRPE64_RS10835	2	NC_021287.1	2356655	2357884	*argG*	argininosuccinate synthase	Amino acid transport and metabolism
BRPE64_RS10880	2	NC_021287.1	2364253	2365806	-	N-acetylmuramoyl-L-alanine amidase	Cell wall/membrane /envelope biogenesis
BRPE64_RS11030	2	NC_021287.1	2389555	2389998	-	TolR protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11035	2	NC_021287.1	2389995	2391107	-	protein TolA	Cell wall/membrane /envelope biogenesis
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11045	2	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane /envelope biogenesis
BRPE64_RS11085	2	NC_021287.1	2401523	2402506	-	4-hydroxythreonine-4-phosphate dehydrogenase	Coenzyme transport and metabolism
BRPE64_RS11220	2	NC_021287.1	2431398	2433449	-	transketolase 1	Carbohydrate transport and metabolism
BRPE64_RS12125	2	NC_021287.1	2620994	2621761	-	hypothetical protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12130	2	NC_021287.1	2621758	2622573	-	ABC transporter related	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12165	2	NC_021287.1	2629958	2634661	*gltB*	glutamate synthase	Amino acid transport and metabolism
BRPE64_RS12260	2	NC_021287.1	2652894	2653202	-	hypothetical protein	-

BRPE64_RS12265	2	NC_021287.1	2653248	2654510	*lysA*	diaminopimelate decarboxylase	Amino acid transport and metabolism
BRPE64_RS12985	2	NC_021287.1	2789444	2790283	*metF*	methylenetetrahydrofolate reductase	Amino acid transport and metabolism
BRPE64_RS12990	2	NC_021287.1	2790324	2790677	-	hypothetical protein	Function unknown
BRPE64_RS12995	2	NC_021287.1	2790753	2792174	-	adenosylhomocysteinase	Coenzyme transport and metabolism
BRPE64_RS13050	2	NC_021287.1	2802057	2802422	*fliE*	flagellar hook-basal body complex protein FliE	Cell motility
BRPE64_RS13055	2	NC_021287.1	2802721	2804499	*fliF*	flagellar FliF M-ring protein	Cell motility
BRPE64_RS13060	2	NC_021287.1	2804489	2805484	*fliG*	flagellar motor switch protein FliG	Cell motility
BRPE64_RS13065	2	NC_021287.1	2805477	2806154	*fliH*	flagellar assembly protein FliH	Cell motility
BRPE64_RS13070	2	NC_021287.1	2806157	2807707	*fliI*	flagellar protein export ATPase FliI	Cell motility
BRPE64_RS13075	2	NC_021287.1	2807773	2808228	*fliJ*	flagellar export protein FliJ	Cell motility
BRPE64_RS13080	2	NC_021287.1	2808268	2809680	*fliK*	putative flagellar hook-length control protein	Cell motility
BRPE64_RS13085	2	NC_021287.1	2810544	2811032	*fliL*	flagellar protein FliL	Cell motility
BRPE64_RS13090	2	NC_021287.1	2811057	2812055	*fliM*	flagellar motor switch protein FliM	Cell motility
BRPE64_RS13100	2	NC_021287.1	2812473	2812973	*fliO*	flagellar biosynthetic protein FliO	Cell motility
BRPE64_RS13105	2	NC_021287.1	2813190	2813960	*fliP*	flagellar biosynthetic protein FliP	Cell motility
BRPE64_RS13110	2	NC_021287.1	2813984	2814253	*fliQ*	flagellar biosynthetic protein FliQ	Cell motility
BRPE64_RS13125	2	NC_021287.1	2817036	2818271	*flgL*	flagellar hook-associated protein 3	Cell motility
BRPE64_RS13130	2	NC_021287.1	2818282	2820225	*flgK*	flagellar hook-associated protein FlgK	Cell motility
BRPE64_RS13145	2	NC_021287.1	2823255	2824382	*flgI*	flagellar P-ring protein	Cell motility
BRPE64_RS13150	2	NC_021287.1	2824385	2825068	*flgH*	flagellar L-ring protein	Cell motility
BRPE64_RS13155	2	NC_021287.1	2825099	2825887	*flgG*	flagellar basal-body rod protein FlgG	Cell motility
BRPE64_RS13160	2	NC_021287.1	2825919	2826671	*flgF*	FlgF	Cell motility
BRPE64_RS13165	2	NC_021287.1	2826707	2827960	*flgE*	flagellar basal body FlaE domain protein	Cell motility
BRPE64_RS13175	2	NC_021287.1	2828753	2829178	*flgC*	flagellar basal-body rod protein FlgC	Cell motility
BRPE64_RS13180	2	NC_021287.1	2829188	2829676	*flgB*	flagellar basal body rod protein FlgB	Cell motility
BRPE64_RS13185	2	NC_021287.1	2829939	2831282	*flgA*	flagella basal body P-ring formation protein FlgA	Cell motility
BRPE64_RS13195	2	NC_021287.1	2831889	2832323	-	hypothetical protein	-
BRPE64_RS13200	2	NC_021287.1	2832533	2833741	*flhB*	flagellar biosynthetic protein FlhB	Cell motility
BRPE64_RS13205	2	NC_021287.1	2833738	2835843	*flhA*	flagellar biosynthesis protein FlhA	Cell motility
BRPE64_RS13275	2	NC_021287.1	2852379	2853110	*fliA*	RNA polymerase sigma factor	Transcription
BRPE64_RS13280	2	NC_021287.1	2853129	2854046	-	flagellar biosynthesis protein FlhG	Cell cycle control, cell division, chromosome partitioning

BRPE64_RS13285	2	NC_021287.1	2854039	2855826	-	GTP-binding signal recognition particle SRP54 G-domain	Cell motility
BRPE64_RS14410	2	NC_021294.1	76678	77526	-	putative squalene/phytoene synthase	Lipid transport and metabolism
BRPE64_RS17375	2	NC_021294.1	711552	711869	-	hypothetical protein	-
BRPE64_RS17825	2	NC_021294.1	800846	801592	-	putative transmembrane transcriptional regulator	Transcription
BRPE64_RS18245	2	NC_021294.1	888588	889997	-	cytochrome bd ubiquinol oxidase subunit I	Energy production and conversion
BRPE64_RS18250	2	NC_021294.1	890002	891003	-	cytochrome d ubiquinol oxidase subunit II	Energy production and conversion
BRPE64_RS19345	2	NC_021294.1	1134818	1135984	-	outer membrane porin OmpC family	Cell wall/membrane /envelope biogenesis
BRPE64_RS19700	2	NC_021294.1	1219620	1221668	-	RNA polymerase sigma factor	Transcription
BRPE64_RS20020	2	NC_021294.1	1305948	1307495	*purF*	amidophosphoribosyltransferase	Nucleotide transport and metabolism
BRPE64_RS20050	2	NC_021294.1	1312310	1313188	-	DNA methylase N-4/N-6 domain protein	Replication, recombination and repair
BRPE64_RS20055	2	NC_021294.1	1313200	1314393	*trpB*	tryptophan synthase beta chain	Amino acid transport and metabolism
BRPE64_RS20060	2	NC_021294.1	1314443	1315111	-	N-(5'-phosphoribosyl)anthranilate isomerase	Amino acid transport and metabolism
BRPE64_RS20065	2	NC_021294.1	1315108	1315911	*truA*	tRNA pseudouridine synthase A	Translation, ribosomal structure and biogenesis
BRPE64_RS20070	2	NC_021294.1	1315913	1317862	-	hypothetical protein	Cell motility
BRPE64_RS20090	2	NC_021294.1	1321440	1322849	-	3-isopropylmalate dehydratase large subunit	Amino acid transport and metabolism
BRPE64_RS20095	2	NC_021294.1	1323212	1323460	-	hypothetical protein	-
BRPE64_RS20100	2	NC_021294.1	1323454	1323774	-	hypothetical protein	-
BRPE64_RS20665	2	NC_021294.1	1449187	1450284	-	N-acylglucosamine 2-epimerase	Carbohydrate transport and metabolism
BRPE64_RS20670	2	NC_021294.1	1450289	1451830	-	methyl-accepting chemotaxis sensory transducer	Cell motility
BRPE64_RS28705	2	NC_021289.1	888916	889830	*metR*	transcriptional regulator LysR family	Transcription
BRPE64_RS28710	2	NC_021289.1	889942	892233	*metE*	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	Amino acid transport and metabolism
BRPE64_RS31205	3	NC_021295.1	167106	168587	-	hypothetical protein	-
BRPE64_RS31220	3	NC_021295.1	170991	171959	-	hypothetical protein	-
BRPE64_RS31275	3	NC_021295.1	188371	189696	-	hypothetical protein	-
BRPE64_RS31210	4	NC_021295.1	168600	169685	-	hypothetical protein	-
BRPE64_RS31215	4	NC_021295.1	169691	170989	-	hypothetical protein	-
BRPE64_RS31225	4	NC_021295.1	172034	172987	-	hypothetical protein	-
BRPE64_RS32060	4	NC_021295.1	173113	174546	-	hypothetical protein	-
BRPE64_RS31235	4	NC_021295.1	174548	179116	-	hypothetical protein	-
BRPE64_RS32065	4	NC_021295.1	181546	184029	-	hypothetical protein	-

BRPE64_RS31265	4	NC_021295.1	185855	186658	-	hypothetical protein	-
BRPE64_RS31270	4	NC_021295.1	186814	187971	-	hypothetical protein	-
BRPE64_RS31280	4	NC_021295.1	189735	190688	-	hypothetical protein	-
BRPE64_RS31285	4	NC_021295.1	190739	198283	-	hypothetical protein	-
BRPE64_RS31290	4	NC_021295.1	198286	199815	-	hypothetical protein	-

Annexe 17: List of fitness genes in *B. insecticola* involved in the colonization of the M4 organ at the third instar stage identified by Con-ARTIST.

Gene tag	Essentiality score	Replicon	Start	End	Gene name	Gene product	Class description COG
BRPE64_RS31795	1	NC_021287.1	7627	8037	-	hypothetical protein	-
BRPE64_RS00520	1	NC_021287.1	113829	115370	*fliD*	flagellar hook-associated 2 domain protein	Cell motility
BRPE64_RS00750	1	NC_021287.1	162538	163416	-	hypothetical protein	-
BRPE64_RS00935	1	NC_021287.1	198756	199502	*gpmA*	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	Carbohydrate transport and metabolism
BRPE64_RS01290	1	NC_021287.1	265367	266569	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS01870	1	NC_021287.1	398708	403045	-	hypothetical protein	Function unknown
BRPE64_RS02150	1	NC_021287.1	465159	465821	*dedA*	membrane-associated protein	Function unknown
BRPE64_RS02250	1	NC_021287.1	484365	485531	-	succinyl-CoA ligase [ADP-forming] subunit beta	Energy production and conversion
BRPE64_RS02350	1	NC_021287.1	505612	506799	*purK*	phosphoribosylaminoimidazole carboxylase ATPase subunit	Nucleotide transport and metabolism
BRPE64_RS02370	1	NC_021287.1	512116	512778	-	DNA-binding response regulator	Transcription
BRPE64_RS02375	1	NC_021287.1	512778	514079	-	integral membrane sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS02505	1	NC_021287.1	545166	545954	*znuB*	cation ABC transporter permease	Inorganic ion transport and metabolism
BRPE64_RS02725	1	NC_021287.1	598292	598606	*clpS*	ATP-dependent Clp protease adapter protein ClpS	Function unknown
BRPE64_RS02975	1	NC_021287.1	650378	651319	*cysB*	transcriptional regulator LysR family	Transcription
BRPE64_RS03175	1	NC_021287.1	687273	691421	-	hypothetical protein	Function unknown
BRPE64_RS03455	1	NC_021287.1	754386	755552	-	hypothetical protein	-
BRPE64_RS03470	1	NC_021287.1	758704	759771	-	hypothetical protein	-
BRPE64_RS03475	1	NC_021287.1	759787	762261	-	putative uncharacterized protein XOO3672	Amino acid transport and metabolism
BRPE64_RS03750	1	NC_021287.1	812442	813461	-	fructose-1,6-bisphosphatase class 1,1	Carbohydrate transport and metabolism
BRPE64_RS03990	1	NC_021287.1	867946	869868	*glk*	bifunctional protein glk	Carbohydrate transport and metabolism
BRPE64_RS04515	1	NC_021287.1	982428	983897	-	RNAse G	Translation, ribosomal structure and biogenesis
BRPE64_RS04685	1	NC_021287.1	1027668	1029599	*thiC*	phosphomethylpyrimidine synthase	Coenzyme transport and metabolism

BRPE64_RS04905	1	NC_021287.1	1075733	1076368	-	Two component transcriptional regulator LuxR family	Signal transduction mechanisms
BRPE64_RS06035	1	NC_021287.1	1308835	1311414	-	putative penicillin-binding (Peptidoglycan synthetase) transmembrane protein mrcA	Cell wall/membrane /envelope biogenesis
BRPE64_RS06390	1	NC_021287.1	1369578	1370900	-	homoserine dehydrogenase	Amino acid transport and metabolism
BRPE64_RS06685	1	NC_021287.1	1439287	1440297	-	aminodeoxychorismate lyase	General function prediction only
BRPE64_RS07500	1	NC_021287.1	1625178	1626218	*rluD*	pseudouridine synthase	Translation, ribosomal structure and biogenesis
BRPE64_RS09525	1	NC_021287.1	2061524	2062123	-	RNA polymerase sigma factor	Transcription
BRPE64_RS10550	1	NC_021287.1	2291177	2293348	-	hypothetical protein	-
BRPE64_RS10585	1	NC_021287.1	2302528	2303079	*rfbC*	dTDP-4-dehydrorhamnose 3,5-epimerase	Cell wall/membrane /envelope biogenesis
BRPE64_RS10880	1	NC_021287.1	2364253	2365806	-	N-acetylmuramoyl-L-alanine amidase	Cell wall/membrane /envelope biogenesis
BRPE64_RS11050	1	NC_021287.1	2393115	2393864	-	Tol-pal system protein YbgF	Function unknown
BRPE64_RS12120	1	NC_021287.1	2620349	2620900	-	putative signal peptide protein toluene tolerance Ttg2C-like protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12255	1	NC_021287.1	2652520	2652837	*cyaY*	protein CyaY	Inorganic ion transport and metabolism
BRPE64_RS13115	1	NC_021287.1	2814264	2815058	*fliR*	flagellar biosynthetic protein fliR	Cell motility
BRPE64_RS13140	1	NC_021287.1	2822193	2823242	*flgJ*	flagellar rod assembly protein/muramidase FlgJ	Cell motility
BRPE64_RS13290	1	NC_021287.1	2856138	2857844	-	thiamine pyrophosphate protein domain protein TPP-binding	Amino acid transport and metabolism
BRPE64_RS14410	1	NC_021294.1	76678	77526	-	putative squalene/phytoene synthase	Lipid transport and metabolism
BRPE64_RS17825	1	NC_021294.1	800846	801592	-	putative transmembrane transcriptional regulator	Transcription
BRPE64_RS17830	1	NC_021294.1	801582	802253	-	RNA polymerase sigma factor	Transcription
BRPE64_RS19075	1	NC_021294.1	1074002	1074937	-	serine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS20075	1	NC_021294.1	1318091	1319212	*asd*	aspartate-semialdehyde dehydrogenase	Amino acid transport and metabolism

BRPE64_RS30505	1	NC_021295.1	8302	9750	-	conserved hypothetical branched-chain amino acid ABC transporter	Amino acid transport and metabolism
BRPE64_RS30625	1	NC_021295.1	31557	32693	-	hypothetical protein	-
BRPE64_RS31725	1	NC_021295.1	293051	294433	-	hypothetical protein	-
BRPE64_RS31735	1	NC_021295.1	295019	295345	-	hypothetical protein	-
BRPE64_RS00270	2	NC_021287.1	59268	59876	-	methionine biosynthesis protein MetW	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS00275	2	NC_021287.1	59873	61018	-	homoserine O-acetyltransferase	Amino acid transport and metabolism
BRPE64_RS00285	2	NC_021287.1	62059	62958	-	acetylglutamate kinase	Amino acid transport and metabolism
BRPE64_RS00525	2	NC_021287.1	115565	117082	*fliC*	flagellin	Cell motility
BRPE64_RS00600	2	NC_021287.1	130402	131262	-	chemotaxis protein MotA	Cell motility
BRPE64_RS00605	2	NC_021287.1	131312	132295	-	chemotaxis protein MotB	Cell motility
BRPE64_RS00610	2	NC_021287.1	132377	132754	-	response regulator receiver protein	Signal transduction mechanisms
BRPE64_RS00615	2	NC_021287.1	132805	135117	-	CheA Signal Transduction Histidine Kinases	Cell motility
BRPE64_RS00620	2	NC_021287.1	135163	135690	-	CheW protein	Cell motility
BRPE64_RS00625	2	NC_021287.1	135709	137559	*tsr*	methyl-accepting chemotaxis sensory transducer	Cell motility
BRPE64_RS00630	2	NC_021287.1	137713	138636	-	MCP methyltransferase CheR-type	Cell motility
BRPE64_RS00635	2	NC_021287.1	138633	139376	*cheD*	probable chemoreceptor glutamine deamidase CheD	Cell motility
BRPE64_RS00640	2	NC_021287.1	139373	140470	-	chemotaxis response regulator protein-glutamate methylesterase	Cell motility
BRPE64_RS00645	2	NC_021287.1	140520	140915	-	response regulator receiver protein	Signal transduction mechanisms
BRPE64_RS00650	2	NC_021287.1	140918	141646	*cheZ*	protein phosphatase CheZ	Cell motility
BRPE64_RS00670	2	NC_021287.1	144624	145271	*dsbA*	thiol disulfide interchange protein DsbA	Energy production and conversion
BRPE64_RS00675	2	NC_021287.1	145458	146231	-	sporulation domain protein	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS00755	2	NC_021287.1	163567	164049	-	RfaE bifunctional protein	Cell wall/membrane /envelope biogenesis
BRPE64_RS00760	2	NC_021287.1	164063	164302	-	hypothetical protein	-
BRPE64_RS01090	2	NC_021287.1	227228	227761	-	hypothetical protein	-

BRPE64_RS01395	2	NC_021287.1	290326	294432	-	FAD linked oxidase domain protein	Energy production and conversion
BRPE64_RS01655	2	NC_021287.1	349176	350741	*purH*	bifunctional purine biosynthesis protein PurH	Nucleotide transport and metabolism
BRPE64_RS01725	2	NC_021287.1	364448	365170	-	hypothetical protein	Function unknown
BRPE64_RS01730	2	NC_021287.1	365167	366759	-	putative paraquat-inducible protein	General function prediction only
BRPE64_RS01735	2	NC_021287.1	366842	367525	-	paraquat-inducible protein A	Function unknown
BRPE64_RS01740	2	NC_021287.1	367522	368199	-	paraquat-inducible protein A	Function unknown
BRPE64_RS02135	2	NC_021287.1	460885	461940	*purM*	phosphoribosylformylglycinamidine cyclo-ligase	Nucleotide transport and metabolism
BRPE64_RS02140	2	NC_021287.1	462006	462953	*miaA*	tRNA dimethylallyltransferase	Translation, ribosomal structure and biogenesis
BRPE64_RS02145	2	NC_021287.1	462950	464950	*mutL*	DNA mismatch repair protein MutL	Replication, recombination and repair
BRPE64_RS02245	2	NC_021287.1	483602	484255	-	hypothetical protein	-
BRPE64_RS02300	2	NC_021287.1	495263	496360	-	lipopolysaccharide heptosyltransferase II	Cell wall/membrane /envelope biogenesis
BRPE64_RS02325	2	NC_021287.1	499591	500787	*pgk*	phosphoglycerate kinase	Carbohydrate transport and metabolism
BRPE64_RS02405	2	NC_021287.1	517090	518598	-	protease Do	Posttranslational modification, protein turnover, chaperones
BRPE64_RS02510	2	NC_021287.1	545947	546855	*znuC*	ABC Mn2+/Zn2+ transporter ATPase subunit	Inorganic ion transport and metabolism
BRPE64_RS02515	2	NC_021287.1	546852	547733	*znuA*	periplasmic solute binding protein	Inorganic ion transport and metabolism
BRPE64_RS02520	2	NC_021287.1	547781	548239	-	transcriptional regulator Fur family	Inorganic ion transport and metabolism
BRPE64_RS02730	2	NC_021287.1	598603	600906	-	putative ATP-dependent Clp protease ATP-binding subunit	Posttranslational modification, protein turnover, chaperones
BRPE64_RS02980	2	NC_021287.1	651558	653234	*cysI*	ferredoxin--nitrite reductase	Inorganic ion transport and metabolism
BRPE64_RS02985	2	NC_021287.1	653245	653784	-	uncharacterized conserved protein UCP030820	Function unknown
BRPE64_RS02990	2	NC_021287.1	653788	654510	*cysH*	adenylylsulfate reductase thioredoxin dependent	Amino acid transport and metabolism
BRPE64_RS02995	2	NC_021287.1	654622	655584	*cysD*	sulfate adenylyltransferase subunit 2	Amino acid transport and metabolism
BRPE64_RS03000	2	NC_021287.1	655612	656925	*cysN*	sulfate adenylyltransferase large subunit	Inorganic ion transport and metabolism

BRPE64_RS03005	2	NC_021287.1	656943	657698	-	uroporphyrin-III C-methyltransferase	Coenzyme transport and metabolism
BRPE64_RS03010	2	NC_021287.1	657835	658215	-	cobalamin (Vitamin B12) biosynthesis CbiX protein	Function unknown
BRPE64_RS03245	2	NC_021287.1	707442	708845	*argH*	argininosuccinate lyase	Amino acid transport and metabolism
BRPE64_RS03460	2	NC_021287.1	756139	757398	-	probable multidrug resistance protein	Amino acid transport and metabolism
BRPE64_RS03465	2	NC_021287.1	757460	758707	-	hypothetical protein	Lipid transport and metabolism
BRPE64_RS03985	2	NC_021287.1	867246	867965	-	6-phosphogluconolactonase	Carbohydrate transport and metabolism
BRPE64_RS04485	2	NC_021287.1	974821	975888	-	glycosyl transferase group 1	Cell wall/membrane /envelope biogenesis
BRPE64_RS04490	2	NC_021287.1	975892	976653	-	glycosyl transferase family 2	Cell wall/membrane /envelope biogenesis
BRPE64_RS04495	2	NC_021287.1	976721	977716	-	glycosyl transferase family 2	Cell wall/membrane /envelope biogenesis
BRPE64_RS04500	2	NC_021287.1	977887	979164	-	O-antigen polymerase	Cell wall/membrane /envelope biogenesis
BRPE64_RS04910	2	NC_021287.1	1076365	1078893	-	multi-sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS05240	2	NC_021287.1	1142691	1144043	*astB*	N-succinylarginine dihydrolase	Amino acid transport and metabolism
BRPE64_RS05570	2	NC_021287.1	1204190	1205836	-	DEAD/DEAH box helicase domain protein	Translation, ribosomal structure and biogenesis
BRPE64_RS05810	2	NC_021287.1	1260186	1262603	-	phosphoenolpyruvate synthase	Carbohydrate transport and metabolism
BRPE64_RS06385	2	NC_021287.1	1368315	1369553	-	aminotransferase AlaT	Amino acid transport and metabolism
BRPE64_RS06475	2	NC_021287.1	1390679	1392052	-	integral membrane sensor signal transduction histidine kinase	Signal transduction mechanisms
BRPE64_RS06485	2	NC_021287.1	1393110	1393475	-	hypothetical protein	-
BRPE64_RS06530	2	NC_021287.1	1401612	1402265	*clpP*	ATP-dependent Clp protease proteolytic subunit	Posttranslational modification, protein turnover, chaperones
BRPE64_RS06535	2	NC_021287.1	1402436	1403707	*clpX*	ATP-dependent Clp protease ATP-binding subunit ClpX	Posttranslational modification, protein turnover, chaperones
BRPE64_RS06540	2	NC_021287.1	1403894	1406317	*lon*	Lon protease	Posttranslational modification, protein

							turnover, chaperones
BRPE64_RS06595	2	NC_021287.1	1416832	1420869	*purL*	phosphoribosylformylglycinamidine synthase	Nucleotide transport and metabolism
BRPE64_RS08840	2	NC_021287.1	1916590	1917606	*ilvC*	ketol-acid reductoisomerase	Amino acid transport and metabolism
BRPE64_RS08845	2	NC_021287.1	1917676	1918167	-	acetolactate synthase small subunit	Amino acid transport and metabolism
BRPE64_RS08850	2	NC_021287.1	1918274	1920037	-	acetolactate synthase large subunit biosynthetic type	Amino acid transport and metabolism
BRPE64_RS09205	2	NC_021287.1	1996687	1997517	-	undecaprenyl-diphosphatase	Defense mechanisms
BRPE64_RS09400	2	NC_021287.1	2038614	2039366	*otsB*	trehalose 6-phosphatase	Carbohydrate transport and metabolism
BRPE64_RS09935	2	NC_021287.1	2141518	2142510	-	ADP-L-glycero-D-manno-heptose-6-epimerase	Carbohydrate transport and metabolism
BRPE64_RS09940	2	NC_021287.1	2142518	2143492	-	RfaE bifunctional protein	Cell wall/membrane /envelope biogenesis
BRPE64_RS10075	2	NC_021287.1	2175984	2177834	-	hypothetical protein	General function prediction only
BRPE64_RS10300	2	NC_021287.1	2228407	2229402	*waaC*	lipopolysaccharide heptosyltransferase I	Cell wall/membrane /envelope biogenesis
BRPE64_RS10475	2	NC_021287.1	2273911	2275809	-	polysaccharide biosynthesis protein CapD	Carbohydrate transport and metabolism
BRPE64_RS10485	2	NC_021287.1	2277059	2278042	-	NAD-dependent epimerase/dehydratase	Carbohydrate transport and metabolism
BRPE64_RS10490	2	NC_021287.1	2278039	2278902	-	putative glycosyl transferase	General function prediction only
BRPE64_RS10555	2	NC_021287.1	2293424	2294173	-	hypothetical protein	-
BRPE64_RS10560	2	NC_021287.1	2294175	2294954	-	ABC-2 type transporter	Carbohydrate transport and metabolism
BRPE64_RS10565	2	NC_021287.1	2295122	2296966	-	glycosyl transferase family 2	Cell wall/membrane /envelope biogenesis
BRPE64_RS10570	2	NC_021287.1	2296988	2300452	-	hypothetical protein	-
BRPE64_RS10575	2	NC_021287.1	2300667	2301611	-	rhamnosyltransferase	General function prediction only
BRPE64_RS10580	2	NC_021287.1	2301632	2302519	-	dTDP-4-dehydrorhamnose reductase	Cell wall/membrane /envelope biogenesis
BRPE64_RS10590	2	NC_021287.1	2303064	2303957	*rfbA*	glucose-1-phosphate thymidylyltransferase	Cell wall/membrane /envelope biogenesis
BRPE64_RS10595	2	NC_021287.1	2303970	2305031	*rfbB*	dTDP-glucose 4,6-dehydratase	Cell wall/membrane /envelope biogenesis

BRPE64_RS10830	2	NC_021287.1	2355573	2356502	-	ornithine carbamoyltransferase	Amino acid transport and metabolism
BRPE64_RS10835	2	NC_021287.1	2356655	2357884	*argG*	argininosuccinate synthase	Amino acid transport and metabolism
BRPE64_RS11020	2	NC_021287.1	2388249	2388692	*ybgC*	4-hydroxybenzoyl-CoA thioesterase	General function prediction only
BRPE64_RS11025	2	NC_021287.1	2388863	2389540	*tolQ*	protein TolQ	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11030	2	NC_021287.1	2389555	2389998	-	TolR protein	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11035	2	NC_021287.1	2389995	2391107	-	protein TolA	Cell wall/membrane /envelope biogenesis
BRPE64_RS11040	2	NC_021287.1	2391222	2392514	*tolB*	protein TolB	Intracellular trafficking, secretion, and vesicular transport
BRPE64_RS11045	2	NC_021287.1	2392581	2393087	-	peptidoglycan-associated lipoprotein	Cell wall/membrane /envelope biogenesis
BRPE64_RS11205	2	NC_021287.1	2429035	2429730	-	GntR domain protein	Transcription
BRPE64_RS11210	2	NC_021287.1	2429845	2430234	-	hypothetical protein	-
BRPE64_RS11215	2	NC_021287.1	2430339	2431349	-	glyceraldehyde-3-phosphate dehydrogenase type I	Carbohydrate transport and metabolism
BRPE64_RS11220	2	NC_021287.1	2431398	2433449	-	transketolase 1	Carbohydrate transport and metabolism
BRPE64_RS12125	2	NC_021287.1	2620994	2621761	-	hypothetical protein	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12130	2	NC_021287.1	2621758	2622573	-	ABC transporter related	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS12135	2	NC_021287.1	2622670	2623314	-	hypothetical protein	-
BRPE64_RS12140	2	NC_021287.1	2623761	2624567	*thiG*	thiazole synthase	Coenzyme transport and metabolism
BRPE64_RS12160	2	NC_021287.1	2628391	2629857	*gltD*	glutamate synthase (NADH) small subunit	Amino acid transport and metabolism
BRPE64_RS12165	2	NC_021287.1	2629958	2634661	*gltB*	glutamate synthase	Amino acid transport and metabolism
BRPE64_RS12260	2	NC_021287.1	2652894	2653202	-	hypothetical protein	-
BRPE64_RS12265	2	NC_021287.1	2653248	2654510	*lysA*	diaminopimelate decarboxylase	Amino acid transport and metabolism

BRPE64_RS12280	2	NC_021287.1	2656361	2657563	*nrfE*	cytochrome c assembly protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12285	2	NC_021287.1	2657568	2659808	-	ResB family protein	Posttranslational modification, protein turnover, chaperones
BRPE64_RS12985	2	NC_021287.1	2789444	2790283	*metF*	methylenetetrahydrofolate reductase	Amino acid transport and metabolism
BRPE64_RS12990	2	NC_021287.1	2790324	2790677	-	hypothetical protein	Function unknown
BRPE64_RS12995	2	NC_021287.1	2790753	2792174	-	adenosylhomocysteinase	Coenzyme transport and metabolism
BRPE64_RS13045	2	NC_021287.1	2801341	2801760	*fliS*	flagellar protein FliS	Cell motility
BRPE64_RS13050	2	NC_021287.1	2802057	2802422	*fliE*	flagellar hook-basal body complex protein FliE	Cell motility
BRPE64_RS13055	2	NC_021287.1	2802721	2804499	*fliF*	flagellar FliF M-ring protein	Cell motility
BRPE64_RS13060	2	NC_021287.1	2804489	2805484	*fliG*	flagellar motor switch protein FliG	Cell motility
BRPE64_RS13065	2	NC_021287.1	2805477	2806154	*fliH*	flagellar assembly protein FliH	Cell motility
BRPE64_RS13070	2	NC_021287.1	2806157	2807707	*fliI*	flagellar protein export ATPase FliI	Cell motility
BRPE64_RS13075	2	NC_021287.1	2807773	2808228	*fliJ*	flagellar export protein FliJ	Cell motility
BRPE64_RS13080	2	NC_021287.1	2808268	2809680	*fliK*	putative flagellar hook-length control protein	Cell motility
BRPE64_RS13085	2	NC_021287.1	2810544	2811032	*fliL*	flagellar protein FliL	Cell motility
BRPE64_RS13090	2	NC_021287.1	2811057	2812055	*fliM*	flagellar motor switch protein FliM	Cell motility
BRPE64_RS13100	2	NC_021287.1	2812473	2812973	*fliO*	flagellar biosynthetic protein FliO	Cell motility
BRPE64_RS13105	2	NC_021287.1	2813190	2813960	*fliP*	flagellar biosynthetic protein FliP	Cell motility
BRPE64_RS13110	2	NC_021287.1	2813984	2814253	*fliQ*	flagellar biosynthetic protein FliQ	Cell motility
BRPE64_RS13125	2	NC_021287.1	2817036	2818271	*flgL*	flagellar hook-associated protein 3	Cell motility
BRPE64_RS13130	2	NC_021287.1	2818282	2820225	*flgK*	flagellar hook-associated protein FlgK	Cell motility
BRPE64_RS13145	2	NC_021287.1	2823255	2824382	*flgI*	flagellar P-ring protein	Cell motility
BRPE64_RS13150	2	NC_021287.1	2824385	2825068	*flgH*	flagellar L-ring protein	Cell motility
BRPE64_RS13155	2	NC_021287.1	2825099	2825887	*flgG*	flagellar basal-body rod protein FlgG	Cell motility
BRPE64_RS13160	2	NC_021287.1	2825919	2826671	*flgF*	FlgF	Cell motility
BRPE64_RS13165	2	NC_021287.1	2826707	2827960	*flgE*	flagellar basal body FlaE domain protein	Cell motility
BRPE64_RS13175	2	NC_021287.1	2828753	2829178	*flgC*	flagellar basal-body rod protein FlgC	Cell motility
BRPE64_RS13180	2	NC_021287.1	2829188	2829676	*flgB*	flagellar basal body rod protein FlgB	Cell motility

BRPE64_RS13185	2	NC_021287.1	2829939	2831282	*flgA*	flagella basal body P-ring formation protein FlgA	Cell motility
BRPE64_RS13195	2	NC_021287.1	2831889	2832323	-	hypothetical protein	-
BRPE64_RS13200	2	NC_021287.1	2832533	2833741	*flhB*	flagellar biosynthetic protein FlhB	Cell motility
BRPE64_RS13205	2	NC_021287.1	2833738	2835843	*flhA*	flagellar biosynthesis protein FlhA	Cell motility
BRPE64_RS13275	2	NC_021287.1	2852379	2853110	*fliA*	RNA polymerase sigma factor	Transcription
BRPE64_RS13280	2	NC_021287.1	2853129	2854046	-	flagellar biosynthesis protein FlhG	Cell cycle control, cell division, chromosome partitioning
BRPE64_RS13285	2	NC_021287.1	2854039	2855826	-	GTP-binding signal recognition particle SRP54 G-domain	Cell motility
BRPE64_RS20060	2	NC_021294.1	1314443	1315111	-	N-(5'-phosphoribosyl)anthranilate isomerase	Amino acid transport and metabolism
BRPE64_RS20065	2	NC_021294.1	1315108	1315911	*truA*	tRNA pseudouridine synthase A	Translation, ribosomal structure and biogenesis
BRPE64_RS20070	2	NC_021294.1	1315913	1317862	-	hypothetical protein	Cell motility
BRPE64_RS20655	2	NC_021294.1	1447264	1448247	-	hypothetical protein	-
BRPE64_RS20660	2	NC_021294.1	1448353	1449036	-	MgtC/SapB transporter	Function unknown
BRPE64_RS20665	2	NC_021294.1	1449187	1450284	-	N-acylglucosamine 2-epimerase	Carbohydrate transport and metabolism
BRPE64_RS28700	2	NC_021289.1	888489	888818	-	hypothetical protein	-
BRPE64_RS28705	2	NC_021289.1	888916	889830	*metR*	transcriptional regulator LysR family	Transcription
BRPE64_RS28710	2	NC_021289.1	889942	892233	*metE*	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	Amino acid transport and metabolism
BRPE64_RS30510	2	NC_021295.1	9867	10808	-	inner-membrane translocator	Amino acid transport and metabolism
BRPE64_RS30515	2	NC_021295.1	10811	12103	-	putative permease component of branched-chain amino acid transport system	Amino acid transport and metabolism
BRPE64_RS30520	2	NC_021295.1	12100	12858	-	ABC transporter related	Amino acid transport and metabolism
BRPE64_RS30525	2	NC_021295.1	12855	13574	-	ABC transporter related	Amino acid transport and metabolism
BRPE64_RS30530	2	NC_021295.1	13571	14395	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS30535	2	NC_021295.1	14451	16640	-	2-oxoisovalerate dehydrogenase beta subunit	Energy production and conversion
BRPE64_RS30540	2	NC_021295.1	16672	17172	-	lactoylglutathione lyase	Amino acid transport and metabolism

BRPE64_RS30545	2	NC_021295.1	17174	18424	-	dihydrolipoamide acetyltransferase	Energy production and conversion
BRPE64_RS30550	2	NC_021295.1	18435	19838	-	dihydrolipoyl dehydrogenase	Energy production and conversion
BRPE64_RS30555	2	NC_021295.1	19878	20618	-	short-chain dehydrogenase/reductase SDR	Lipid transport and metabolism
BRPE64_RS30560	2	NC_021295.1	20789	21610	-	shikimate/quinate 5-dehydrogenase family protein 2	Amino acid transport and metabolism
BRPE64_RS30565	2	NC_021295.1	21814	22995	-	purine efflux pump PbuE	Carbohydrate transport and metabolism
BRPE64_RS30570	2	NC_021295.1	23388	23897	-	transcriptional regulator MarR family	Transcription
BRPE64_RS30575	2	NC_021295.1	24246	24584	-	transport-associated	General function prediction only
BRPE64_RS30580	2	NC_021295.1	25122	25328	-	hypothetical protein	-
BRPE64_RS30585	2	NC_021295.1	25724	26674	-	transcriptional regulator AraC family with amidase-like domain	Transcription
BRPE64_RS30630	2	NC_021295.1	32745	32948	-	hypothetical protein	-
BRPE64_RS30635	2	NC_021295.1	32985	34418	-	NAD-dependent aldehyde dehydrogenase protein	Energy production and conversion
BRPE64_RS32035	2	NC_021295.1	34568	37150	-	transcriptional regulator winged helix family	Transcription
BRPE64_RS30645	2	NC_021295.1	37709	39088	-	transcriptional regulator winged helix family	Posttranslational modification, protein turnover, chaperones
BRPE64_RS30650	2	NC_021295.1	39749	42373	-	hypothetical protein	-
BRPE64_RS30655	2	NC_021295.1	42519	42899	-	response regulator receiver protein	Signal transduction mechanisms
BRPE64_RS30660	2	NC_021295.1	42928	43578	-	DNA-binding response regulator FixJ	Signal transduction mechanisms
BRPE64_RS30665	2	NC_021295.1	43691	44278	-	probable transcriptional regulator protein TetR family	Transcription
BRPE64_RS30670	2	NC_021295.1	44870	45253	-	hypothetical protein	-
BRPE64_RS30675	2	NC_021295.1	45392	45817	-	hypothetical protein	-
BRPE64_RS30680	2	NC_021295.1	46395	47273	-	aldo/keto reductase	Energy production and conversion
BRPE64_RS30685	2	NC_021295.1	47467	48279	-	hypothetical protein	Carbohydrate transport and metabolism
BRPE64_RS30690	2	NC_021295.1	48872	49231	-	hypothetical protein	-
BRPE64_RS30695	2	NC_021295.1	49465	50949	-	RND efflux system outer membrane lipoprotein NodT family	Cell wall/membrane /envelope biogenesis

BRPE64_RS30725	2	NC_021295.1	58984	60837	-	cyclic nucleotide-regulated FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Posttranslational modification, protein turnover, chaperones
BRPE64_RS30730	2	NC_021295.1	61024	61329	-	transcriptional regulator ArsR family	Transcription
BRPE64_RS30735	2	NC_021295.1	61596	62168	-	transcriptional regulator TetR family	Transcription
BRPE64_RS30740	2	NC_021295.1	62227	63429	-	major facilitator superfamily MFS_1	Carbohydrate transport and metabolism
BRPE64_RS30745	2	NC_021295.1	63587	64324	-	putative short-chain dehydrogenase	Lipid transport and metabolism
BRPE64_RS30750	2	NC_021295.1	64367	65071	-	isochorismatase hydrolase	Secondary metabolites biosynthesis, transport and catabolism
BRPE64_RS32040	2	NC_021295.1	65673	66659	-	FMN-dependent NADH-azoreductase	Lipid transport and metabolism
BRPE64_RS30760	2	NC_021295.1	66722	67012	-	hypothetical protein	-
BRPE64_RS30765	2	NC_021295.1	67095	67721	-	hypothetical protein	-
BRPE64_RS30770	2	NC_021295.1	68009	68512	-	hypothetical protein	-
BRPE64_RS30775	2	NC_021295.1	68780	70030	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS30780	2	NC_021295.1	70293	71234	-	alcohol dehydrogenase zinc-binding domain protein	Energy production and conversion
BRPE64_RS30785	2	NC_021295.1	71342	71791	-	4-oxalocrotonate tautomerase	General function prediction only
BRPE64_RS32045	2	NC_021295.1	71877	72335	-	hypothetical protein	-
BRPE64_RS30795	2	NC_021295.1	72732	73694	-	transcriptional regulator LysR family	Transcription
BRPE64_RS30800	2	NC_021295.1	73879	74562	-	hypothetical protein	-
BRPE64_RS30805	2	NC_021295.1	74574	75956	-	amidase	Translation, ribosomal structure and biogenesis
BRPE64_RS30810	2	NC_021295.1	76019	77179	-	extracellular ligand-binding receptor	Amino acid transport and metabolism
BRPE64_RS30815	2	NC_021295.1	77181	78374	-	peptidase	Amino acid transport and metabolism
BRPE64_RS30820	2	NC_021295.1	78444	79416	-	hypothetical protein	-
BRPE64_RS30825	2	NC_021295.1	79791	81029	-	major facilitator superfamily MFS_1	Amino acid transport and metabolism
BRPE64_RS30830	2	NC_021295.1	81753	87071	-	nodulation protein NodV	General function prediction only
BRPE64_RS30835	2	NC_021295.1	87174	88053	-	hypothetical protein	-
BRPE64_RS30840	2	NC_021295.1	88189	88740	-	hypothetical protein	-
BRPE64_RS30845	2	NC_021295.1	89130	90047	-	hypothetical protein	-
BRPE64_RS30850	2	NC_021295.1	90108	90854	-	hypothetical protein	-
BRPE64_RS30855	2	NC_021295.1	91822	92931	-	hypothetical protein	-

BRPE64_RS30860	2	NC_021295.1	93610	94500	-	hypothetical protein	-
BRPE64_RS30865	2	NC_021295.1	94587	95312	-	hypothetical protein	-
BRPE64_RS30870	2	NC_021295.1	95356	96012	-	hypothetical protein	-
BRPE64_RS30875	2	NC_021295.1	96009	96773	-	hypothetical protein	-
BRPE64_RS30880	2	NC_021295.1	96817	97629	-	hypothetical protein	-
BRPE64_RS30885	2	NC_021295.1	97667	98758	-	hypothetical protein	-
BRPE64_RS30890	2	NC_021295.1	98772	99614	-	hypothetical protein	-
BRPE64_RS30895	2	NC_021295.1	99611	100531	-	hypothetical protein	-
BRPE64_RS30900	2	NC_021295.1	100537	101634	-	hypothetical protein	-
BRPE64_RS30905	2	NC_021295.1	101797	103272	-	hypothetical protein	-
BRPE64_RS30910	2	NC_021295.1	103269	104930	-	hypothetical protein	-
BRPE64_RS30915	2	NC_021295.1	105233	105832	-	hypothetical protein	-
BRPE64_RS30920	2	NC_021295.1	105829	107238	-	hypothetical protein	-
BRPE64_RS30925	2	NC_021295.1	107337	108152	-	hypothetical protein	-
BRPE64_RS30930	2	NC_021295.1	108169	109161	-	hypothetical protein	-
BRPE64_RS30935	2	NC_021295.1	109201	110409	-	hypothetical protein	-
BRPE64_RS30940	2	NC_021295.1	110442	111215	-	hypothetical protein	-
BRPE64_RS30945	2	NC_021295.1	111365	112249	-	hypothetical protein	-
BRPE64_RS30950	2	NC_021295.1	112595	113662	-	hypothetical protein	-
BRPE64_RS30955	2	NC_021295.1	113841	114644	-	hypothetical protein	-
BRPE64_RS30960	2	NC_021295.1	114658	115797	-	hypothetical protein	-
BRPE64_RS30965	2	NC_021295.1	115810	116553	-	hypothetical protein	-
BRPE64_RS30970	2	NC_021295.1	116830	117870	-	hypothetical protein	-
BRPE64_RS30975	2	NC_021295.1	117892	118218	-	hypothetical protein	-
BRPE64_RS30980	2	NC_021295.1	118352	120031	-	hypothetical protein	-
BRPE64_RS32050	2	NC_021295.1	120907	121407	-	hypothetical protein	-
BRPE64_RS30990	2	NC_021295.1	121804	122598	-	hypothetical protein	-
BRPE64_RS30995	2	NC_021295.1	123209	124093	-	hypothetical protein	-
BRPE64_RS31000	2	NC_021295.1	124527	125150	-	hypothetical protein	-
BRPE64_RS31005	2	NC_021295.1	125246	126421	-	hypothetical protein	-
BRPE64_RS31010	2	NC_021295.1	126418	126624	-	hypothetical protein	-
BRPE64_RS31015	2	NC_021295.1	126624	127769	-	hypothetical protein	-
BRPE64_RS31050	2	NC_021295.1	135067	135987	-	hypothetical protein	-
BRPE64_RS31055	2	NC_021295.1	136001	137194	-	hypothetical protein	-
BRPE64_RS31060	2	NC_021295.1	137500	137871	-	hypothetical protein	-
BRPE64_RS31065	2	NC_021295.1	138612	139517	-	hypothetical protein	-
BRPE64_RS31070	2	NC_021295.1	139849	141180	-	hypothetical protein	-
BRPE64_RS31100	2	NC_021295.1	146560	147528	-	hypothetical protein	-
BRPE64_RS31105	2	NC_021295.1	147562	148065	-	hypothetical protein	-
BRPE64_RS31110	2	NC_021295.1	148094	148837	-	hypothetical protein	-
BRPE64_RS31115	2	NC_021295.1	148857	149480	-	hypothetical protein	-
BRPE64_RS31120	2	NC_021295.1	149491	149826	-	hypothetical protein	-
BRPE64_RS31130	2	NC_021295.1	150076	150846	-	hypothetical protein	-
BRPE64_RS31135	2	NC_021295.1	150881	151867	-	hypothetical protein	-
BRPE64_RS31140	2	NC_021295.1	152187	152957	-	hypothetical protein	-
BRPE64_RS31145	2	NC_021295.1	153006	153803	-	hypothetical protein	-
BRPE64_RS31150	2	NC_021295.1	153796	154509	-	hypothetical protein	-
BRPE64_RS31155	2	NC_021295.1	154506	155201	-	hypothetical protein	-
BRPE64_RS31160	2	NC_021295.1	155416	156498	-	hypothetical protein	-
BRPE64_RS31165	2	NC_021295.1	156499	157347	-	hypothetical protein	-
BRPE64_RS31170	2	NC_021295.1	157670	158845	-	hypothetical protein	-
BRPE64_RS31175	2	NC_021295.1	158982	159881	-	hypothetical protein	-

BRPE64_RS31180	2	NC_021295.1	159827	160282	-	hypothetical protein	-
BRPE64_RS31185	2	NC_021295.1	160383	161420	-	hypothetical protein	-
BRPE64_RS31190	2	NC_021295.1	162835	165273	-	hypothetical protein	-
BRPE64_RS31195	2	NC_021295.1	165456	166256	-	hypothetical protein	-
BRPE64_RS31300	2	NC_021295.1	201640	202482	-	hypothetical protein	-
BRPE64_RS31305	2	NC_021295.1	202507	202812	-	hypothetical protein	-
BRPE64_RS31310	2	NC_021295.1	203409	203663	-	hypothetical protein	-
BRPE64_RS31315	2	NC_021295.1	203627	205234	-	hypothetical protein	-
BRPE64_RS31320	2	NC_021295.1	205334	205687	-	hypothetical protein	-
BRPE64_RS31325	2	NC_021295.1	205687	206160	-	hypothetical protein	-
BRPE64_RS31330	2	NC_021295.1	206166	206540	-	hypothetical protein	-
BRPE64_RS31335	2	NC_021295.1	207351	209084	-	hypothetical protein	-
BRPE64_RS31340	2	NC_021295.1	209303	210118	-	hypothetical protein	-
BRPE64_RS31345	2	NC_021295.1	210740	211516	-	hypothetical protein	-
BRPE64_RS31350	2	NC_021295.1	211762	212604	-	hypothetical protein	-
BRPE64_RS31355	2	NC_021295.1	212912	214051	-	hypothetical protein	-
BRPE64_RS31360	2	NC_021295.1	214255	215559	-	hypothetical protein	-
BRPE64_RS31365	2	NC_021295.1	215650	216714	-	hypothetical protein	-
BRPE64_RS31415	2	NC_021295.1	223687	225048	-	hypothetical protein	-
BRPE64_RS31420	2	NC_021295.1	225287	226261	-	hypothetical protein	-
BRPE64_RS31425	2	NC_021295.1	226266	227015	-	hypothetical protein	-
BRPE64_RS31430	2	NC_021295.1	227012	228223	-	hypothetical protein	-
BRPE64_RS31435	2	NC_021295.1	228243	228566	-	hypothetical protein	-
BRPE64_RS31440	2	NC_021295.1	228607	230043	-	hypothetical protein	-
BRPE64_RS31445	2	NC_021295.1	230104	231420	-	hypothetical protein	-
BRPE64_RS31450	2	NC_021295.1	231758	232489	-	hypothetical protein	-
BRPE64_RS31455	2	NC_021295.1	232690	233378	-	hypothetical protein	-
BRPE64_RS31460	2	NC_021295.1	233935	235119	-	hypothetical protein	-
BRPE64_RS31465	2	NC_021295.1	235535	237064	-	hypothetical protein	-
BRPE64_RS31470	2	NC_021295.1	237219	238169	-	hypothetical protein	-
BRPE64_RS31475	2	NC_021295.1	238206	239153	-	hypothetical protein	-
BRPE64_RS31480	2	NC_021295.1	239260	239994	-	hypothetical protein	-
BRPE64_RS31485	2	NC_021295.1	240098	240718	-	hypothetical protein	-
BRPE64_RS31490	2	NC_021295.1	240781	241296	-	hypothetical protein	-
BRPE64_RS31495	2	NC_021295.1	241990	242880	-	hypothetical protein	-
BRPE64_RS31500	2	NC_021295.1	242877	243521	-	hypothetical protein	-
BRPE64_RS31505	2	NC_021295.1	243535	244371	-	hypothetical protein	-
BRPE64_RS31510	2	NC_021295.1	244430	245188	-	hypothetical protein	-
BRPE64_RS31515	2	NC_021295.1	245179	246198	-	hypothetical protein	-
BRPE64_RS31520	2	NC_021295.1	246299	247327	-	hypothetical protein	-
BRPE64_RS31525	2	NC_021295.1	247558	248985	-	hypothetical protein	-
BRPE64_RS31530	2	NC_021295.1	249052	250404	-	hypothetical protein	-
BRPE64_RS31535	2	NC_021295.1	250429	251199	-	hypothetical protein	-
BRPE64_RS31540	2	NC_021295.1	251306	252007	-	hypothetical protein	-
BRPE64_RS31545	2	NC_021295.1	252004	252996	-	hypothetical protein	-
BRPE64_RS31550	2	NC_021295.1	253143	254042	-	hypothetical protein	-
BRPE64_RS31555	2	NC_021295.1	254322	254822	-	hypothetical protein	-
BRPE64_RS31560	2	NC_021295.1	254914	256197	-	hypothetical protein	-
BRPE64_RS31565	2	NC_021295.1	256656	257705	-	hypothetical protein	-
BRPE64_RS31570	2	NC_021295.1	258159	259061	-	hypothetical protein	-
BRPE64_RS31655	2	NC_021295.1	277234	277836	-	hypothetical protein	-
BRPE64_RS31660	2	NC_021295.1	278035	278655	-	hypothetical protein	-

BRPE64_RS31665	2	NC_021295.1	278892	280436	-	hypothetical protein	-
BRPE64_RS31670	2	NC_021295.1	280572	281984	-	hypothetical protein	-
BRPE64_RS31675	2	NC_021295.1	282087	283295	-	hypothetical protein	-
BRPE64_RS31680	2	NC_021295.1	283994	284341	-	hypothetical protein	-
BRPE64_RS31685	2	NC_021295.1	284507	284710	-	hypothetical protein	-
BRPE64_RS31695	2	NC_021295.1	287055	287390	-	hypothetical protein	-
BRPE64_RS32075	2	NC_021295.1	287772	288287	-	hypothetical protein	-
BRPE64_RS31705	2	NC_021295.1	288764	289216	-	hypothetical protein	-
BRPE64_RS31710	2	NC_021295.1	289345	290529	-	hypothetical protein	-
BRPE64_RS31715	2	NC_021295.1	290972	291652	-	hypothetical protein	-
BRPE64_RS31720	2	NC_021295.1	291752	293041	-	hypothetical protein	-
BRPE64_RS31740	2	NC_021295.1	295563	295826	-	hypothetical protein	-
BRPE64_RS31745	2	NC_021295.1	295877	296185	-	hypothetical protein	-
BRPE64_RS31750	2	NC_021295.1	296424	297023	-	hypothetical protein	-
BRPE64_RS31755	2	NC_021295.1	297743	299197	-	hypothetical protein	-
BRPE64_RS31760	2	NC_021295.1	299380	300027	-	hypothetical protein	-
BRPE64_RS31765	2	NC_021295.1	300657	301178	-	hypothetical protein	-
BRPE64_RS31770	2	NC_021295.1	301280	303670	-	hypothetical protein	-
BRPE64_RS31775	2	NC_021295.1	303740	304744	-	hypothetical protein	-
BRPE64_RS31780	2	NC_021295.1	305624	306268	-	hypothetical protein	-
BRPE64_RS31785	2	NC_021295.1	306466	307641	-	hypothetical protein	-
BRPE64_RS32080	2	NC_021295.1	308088	308813	-	hypothetical protein	-
BRPE64_RS31200	3	NC_021295.1	166343	167047	-	hypothetical protein	-
BRPE64_RS31205	4	NC_021295.1	167106	168587	-	hypothetical protein	-
BRPE64_RS31210	4	NC_021295.1	168600	169685	-	hypothetical protein	-
BRPE64_RS31215	4	NC_021295.1	169691	170989	-	hypothetical protein	-
BRPE64_RS31220	4	NC_021295.1	170991	171959	-	hypothetical protein	-
BRPE64_RS31225	4	NC_021295.1	172034	172987	-	hypothetical protein	-
BRPE64_RS31235	4	NC_021295.1	174548	179116	-	hypothetical protein	-
BRPE64_RS31240	4	NC_021295.1	179142	180404	-	hypothetical protein	-
BRPE64_RS31245	4	NC_021295.1	180421	181542	-	hypothetical protein	-
BRPE64_RS32065	4	NC_021295.1	181546	184029	-	hypothetical protein	-
BRPE64_RS31255	4	NC_021295.1	184049	185200	-	hypothetical protein	-
BRPE64_RS31260	4	NC_021295.1	185202	185858	-	hypothetical protein	-
BRPE64_RS31265	4	NC_021295.1	185855	186658	-	hypothetical protein	-
BRPE64_RS31270	4	NC_021295.1	186814	187971	-	hypothetical protein	-
BRPE64_RS31275	4	NC_021295.1	188371	189696	-	hypothetical protein	-
BRPE64_RS31280	4	NC_021295.1	189735	190688	-	hypothetical protein	-
BRPE64_RS31285	4	NC_021295.1	190739	198283	-	hypothetical protein	-
BRPE64_RS31290	4	NC_021295.1	198286	199815	-	hypothetical protein	-
BRPE64_RS31295	4	NC_021295.1	199817	200602	-	hypothetical protein	-

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences du végétal : du gène à l'écosystème

Titre : Identification des facteurs de résistance aux peptides antimicrobiens et de colonisation de l'insecte *Riptortus pedestris* chez la bactérie symbiotique *Burkholderia insecticola.*

Mots clés : Peptides antimicrobiens – Tn-seq – *Burkholderia* – symbiose – *Riptortus pedestris*

Résumé : L'insecte phytophage *Riptortus pedestris*, appartenant au sous-ordre des Hétéroptères, est un ravageur notoire de cultures agricoles en Asie du sud-est qui se nourrit préférentiellement de plants de soja. Cette punaise est associée à une bactérie symbiotique du genre *Burkholderia* nommée *Burkholderia insecticola*, localisée dans une région spécifique de l'intestin de l'insecte appelée la région M4. Cette région M4, organisée en cryptes, constitue l'organe symbiotique dans lequel le symbiote prolifère de manière extracellulaire. Cette interaction favorise la croissance et le développement de la punaise. Récemment, il a été montré que *Riptortus* produit des peptides antimicrobiens au sein des cryptes, appelés "crypt-specific cysteine-rich peptides" ou peptides CCR pour lesquels le symbiote est particulièrement résistant. Il a été proposé que les peptides antimicrobiens de l'hôte, incluant les peptides CCR, participent à la colonisation spécifique de l'organe symbiotique par *B. insecticola*. Dans ce travail, une approche Tn-seq a été utilisée pour identifier les gènes bactériens impliqués dans la résistance aux peptides antimicrobiens et dans la symbiose. Dans un premier temps, la robustesse de la méthode Tn-seq a été évaluée en identifiant le génome essentiel de *B. insecticola*. Puis dans un second temps, les facteurs bactériens impliqués dans la résistance aux peptides antimicrobiens ont été caractérisés via une approche gènes-candidats et l'approche Tn-seq. Dans une dernière partie, une expérience de Tn-seq *in vivo* a permis d'évaluer l'ampleur du goulot d'étranglement sur la population symbiotique lors de l'infection de l'organe symbiotique et d'identifier les facteurs symbiotiques impliqués dans la colonisation de *R. pedestris*.

Title : Identification in the bacterial symbiont *Burkholderia insecticola* of factors involved in antimicrobial peptide-resistance and colonization of the insect *Riptortus pedestris*.

Keywords : Antimicrobial peptides – Tn-seq – *Burkholderia* – symbiosis – *Riptortus pedestris*

Abstract : The phytophagous insect *Riptortus pedestris*, belonging to the *Heteroptera* suborder, is a notorious crop pest in South-Eastern Asia which feeds preferentially on soybean plants. This bean bug is associated with a bacterial symbiont, a specific *Burkholderia* species named *Burkholderia insecticola*, located in the M4 region of the insect's midgut. This M4 region is organized in crypts and constitutes the symbiotic organ where the symbiont proliferates extracellularly. This interaction promotes the growth and the development of the bean bug. Recently, it was demonstrated that *Riptortus* produces antimicrobial peptides in the midgut crypts called crypt-specific cysteine-rich peptides (CCR) for which the bacterial symbiont demonstrates a high resistance profile. It was proposed that host antimicrobial peptides, including the CCR peptides, contribute to the specific colonization of the symbiotic organ by *B. insecticola*. In this work, a Tn-seq approach was used to find bacterial fitness genes involved in antimicrobial peptide resistance and symbiosis. First, the robustness of the Tn-seq method was assessed by identifying the essential genome of *B. insecticola*. Second, the bacterial factors for antimicrobial peptide resistance were characterized, based on both a candidate-gene and the Tn-seq approach. Finally, a Tn-seq *in vivo* experiment was performed to reveal the infection bottleneck effect on the symbiotic population and to identify the bacterial symbiosis factors for the colonization of *R. pedestris*.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France