



الجمهورية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة زيان عاشور-الجلفة
Université Ziane Achour – Djelfa
كلية علوم الطبيعة و الحياة
Faculté des Sciences de la Nature et de la vie
قسم البيولوجيا
Département de biologie

Mémoire de fin d'étude

En vue de l'obtention du Diplôme de Master
Spécialité : Microbiologie appliquée

Thème

**Identification des voies physiologiques conservées par les
bactéries lors de l'endosymbiose et de la pathogénie**

Présenté par : Sioussiou Saida et Si bachir Warda

Soutenue le : 11 juillet 2019

DEVANT LE JURY :

Président	M.Boutaiba Saad	Maitre de conférences A	U-Djelfa
Examineur	M.Ounissi Morad	Maitre-assistant	U-Djelfa
Examineur	M.Belaouni Hadj Ahmed	Maitre-assistant	U-Djelfa
Encadrant	M.Berrabah.Fathi	Maitre de conférences B	U-Djelfa

Année universitaire 2018/2019

Remerciements

On tient à remercier tout d'abord :

Allah pour nous avoir aidé à atteindre notre but

On tient à remercier vivement notre encadreur monsieur Berrabah Fathi pour le soutien et l'aide qu'il n'a jamais manqué de nous apporter, pour ses conseils et ses orientations durant l'élaboration du travail.

Les mots ne seront jamais suffisants pour remercier tous ceux qui ont été présent de près ou de loin pendant notre parcours, en particulier notre collègue Ibrahim et Khalil Nigrou pour son impact positive et sa capacité d'apporter toujours son aide quand il le faut

On remercie aussi les membres du jury qui ont accepté l'évaluation de cet ouvrage et de nous faire part de leurs remarques surement pertinentes, qui avec un peu de recul contribuerons sans doute au perfectionnement du présent travail.

Dédicace

A ma très chère mère

Qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureuse. Quoi que je fasse ou que je dise ; je ne saurai point te remercier comme il se doit. Ta présence à mes côtés a toujours été ma source de force.

A mon très cher père

Pour ta confiance ; ton soutien ; tes sacrifices et toutes les valeurs que tu as su m'inculquer. Tu as été toujours à mes côtés malgré les distances qui nous séparent.

' Veuillez trouver ici le témoignage de mon amour éternel.'

A mes chères frères Nadir et Youcef ; mes chères soeurs Wahiba ; Soumia ;

Rayhane et ma puce Ihsane.

Que Dieu vous procure santé ; prospérité et bonheur ...

A tous mes amies ; mes collègues et tous ceux ou celles qui ont cru en moi et m'ont chaleureusement supporté et encouragé tout au long de mon parcours. Pour plusieurs raisons ; je veux remercier toutes les personnes qui ont aidé du près ou du loin pour réaliser ce modeste travail.

Saida Sioussiou

Dédicace

*Du profond de mon cœur, je dédie ce travail à tous ceux qui me sont chers, à ma
famille cousins cousines et amis*

A la mémoire de Mon père.

*L'homme brave, honnête, courageux qui m'a toujours encouragé et donner une
raison d'avancer et aller plus loin.*

*J'espère que du monde qui est sien maintenant, il apprécie cet humble geste
comme preuve de reconnaissance de la part d'une fille qui a toujours prie pour le
salut de son âme*

A ma chère mère

*Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma
considération pour les sacrifices qu'elle a consenti pour mon instruction, mon
éducation et mon bien être*

A ma sœur Wafa

*Le symbole de courage et honnêteté, en témoignage de mon affection fraternelle,
de ma profonde tendresse et reconnaissance pour tout ce que tu as fait pour moi
depuis toujours.*

A mon amie Saida

*En hommage à notre amitié et au parcours qu'on a fait ensemble, pour ta
patience et ta sagesse.*

A mon cousin , frère et ami Khaled Sibachir

Pour ta gentillesse , ton aide précieuse tout au long de ce travail .

Warda sibachir

Sommaire

Partie 1 : Synthèse bibliographique

Introduction

Chapitre I : Généralités sur les interactions plantes-microorganismes

1.La rhizosphère : réservoir de diversité	1
2. Mécanismes de contrôle de la composition rhizosphérique	1
3. Les interactions plantes-pathogènes :.....	2
4.Mode d'infection des phytopathogènes	3
5. Mécanismes moléculaires et déterminants du pouvoir pathogène :.....	3
5.1 Les exopolysaccharides (EPS) :	4
5.2 Les toxines :	5
5.3 Les enzymes lytiques :	5
6. Exemple d'agents pathogènes	6
6.1. Agrobacterium tumefaciens	6
6.3. Pseudomonas syringea	9
7. Mécanismes moléculaires de résistance chez les plantes	9
7.1 La PAMP Triggered Immunity (PTI).....	9
7.2. L'Effector Triggered Immunity (ETI)	10
Introduction à la symbiose légumineuses-rhizobia	12
2.Physiologie moléculaire de la symbiose fixatrice d'azote	12
2.1. La phase de reconnaissance plante-bactérie	12
2.2. L'infection des racines	12
2.3. NCRs & différenciation des bacteroides	13
2.4. Survie intracellulaire des bacteroides.....	13
2.4. La fixation d'azote point final de la symbiose	14
3.Les nodosités de type déterminé & indéterminé	15
4. L'immunité au cours de la symbiose	16
4.1. L'immunité dans les étapes précoces	16
4.2. L'immunité dans les étapes tardives	17
1. La génomique.....	17
1.1. La génomique structurale	18

2. La génomique fonctionnelle.....	22
3. La génomique comparative	22
3.1. Outils de génomique comparative	23
4. La transcriptomique	25
4.1. Les puces à ADN.....	25
4.2. Séquençage et analyse d'ARN (RNA seq technology).....	26
1. Contexte générale de l'étude	28
1.1. La symbiose et l'immunité d'un point de vue procaryotique.....	28
1.2. L'adaptation des endosymbiotes à la vie intracellulaire	28
2. Questions et problèmes abordés.....	29
Partie 2: Matériels et méthodes	
1. Sélection des souches 30	
2. Annotation RAST (Rapid annotation subsystem technology) :.....	31
3. Le serveur Microscope (Microbial Genome Annotation & Analysis Platform)	32
3.1. Comparaison génomique.....	32
3.2. Analyse de sythénie.....	34
4. La base de données Omnibus (GEO)	
5. Analyse transcriptomique.....	36
6. Phylogénie moléculaire	36
7. Prédiction de la localisation.....	37
6.1. Prédiction des interactions protéine-protéine	37
Partie 3: Résultats et discussion	
1. Description génomique des souches employées	39
2. Résultats des annotations des génomes par RAST	40
3. Peu de différence en quantité de gènes entre les souches	41
4. Les souches présentent des différences qualitatives en composition génique.....	43
5. Les plasmides à la base du style de vie	45
6. Les rhizobia présentent des divergences intragroupes d'ampleur équivalente à celle des pathogènes	47
7. Les pathogènes et les symbiotes présentent des gènes conservées.....	48

8. L'analyse de synténie montre une conservation des génomes entre pathogènes et symbiontes.....	49
9. Les gènes conservés entre <i>S. meliloti</i> et les pathogènes sont exprimés au cours de la symbiose	50
11. Le flagelle impliqué dans la survie intracellulaire ?	54
12.. L'hémolysine D (HlyD) un candidat pour l'étude de la survie intracellulaire.....	56
13 NolX, un gène important pour la symbiose et la pathogénie	57
Conclusion	59
Références bibliographiques	

Liste des figures

Figure 1: La coiffe racinaire	2
Figure 2 : Représentation schématique des différents facteurs impliqués dans le pouvoir pathogène des bactéries	4
Figure 3 : le plasmide <i>ti</i> d' <i>A.tumefaciens</i>	7
Figure 4: Processus d'infection par <i>A.tumefaciens</i>	8
Figure 5 : ETI et PTI suite à la perception des PAMPs	11
Figure 6 : L'infection et la différenciation	14
Figure 7 : Les nodosités indéterminées et déterminées.....	15
Figure 8 : Les étapes relatives à la construction d'une banque d'ADN génomique	19
Figure 9 : Les instruments NGS	20
Figure 10 : La construction et l'analyse de réseaux de synthèse.....	24
Figure 11 : Principe d'analyse transcriptomique par puce à ADN.....	26
Figure 12 : Vue d'ensemble de l'ARN-seq.....	27
Figure 13 : Interface de base de données NCBI.....	30
Figure 14 : les étapes d'obtention d'annotation par RAST server.....	31
Figure 15 : interface microscope server	32
Figure 16 : Résultat de comparaison des souches.....	33
Figure 17 : Evaluation de la synthèse.....	34
Figure 18 : Interface de la base de données Geo datasets.....	35
Figure 19 : Interface du logiciel MEGA 7.....	37
Figure 20 : Interface de la base de données string.....	38
Figure 21 : Recouvrement de génomes annotés par l'outil RAST.....	41
Figure 22 : Représentation du pourcentage des gènes de chaque voie métabolique pour les différentes souches étudiées.....	42
Figure 23 : Représentation du pourcentage des gènes plasmidiques de chaque voie métabolique pour les différentes souches étudiées.....	46
Figure 24 : Nombre de familles de gènes communes et souche spécifique au sein des rhizobia ou des pathogènes.....	47
Figure 25 : Nombre de famille de gènes communs et divergents entre les pathogènes et les symbiotes.....	48
Figure 26 : Conservation et divergence génomique entre pathogène et symbiote.....	49
Figure 27 : Des orthologues de <i>b. Abortus</i> et <i>a. Tumefaciens</i> sont exprimés chez <i>s. Meliloti</i> 1021 au cours de la symbiose.....	51

Figure 28 : Le Profile d'expression des orthologues d' <i>A. tumefasciens</i>	52
Figure 29 : Expression des orthologues de <i>B. abortus</i> chez <i>S. meliloti</i> 1021 lors de la symbiose.....	54
Figure 30 :Expression relative des gènes du flagelle.....	55
Figure 31 :HlyD et survie intracellulaire.....	57
Figure 32 : Arbre phylogénique des orthologues de la protéine.....	58

Listes des abréviations :

BCL	Border Cell like
EPS	Exopolysaccharid
TI	Tumor inducer
PAMP	Pathogen associated Molecular Pattern
PTI	Pamp Triggered immunity
ETI	Effector Triggered immunity
PRR	Pathogens recognition receptor
FLS	Flagelin sensing
Bak 1	Brassinosteroid associated
ETS	Effector triggered sensitivity
NFS	Nod factors
RLK	Receptor like kinase
Lys M	Lysine motif
ENOD	Early Noduline
CI	infection cord
MPB	Membrane peri- bacteroidienne
EPB	Espace peri-bacteoidien
NCR	Nodule Cystein Reach
DNF 1	Do not fixe nitrogen
ND	Determined Nod
IN	Indeterminate Nod
ZI	First zone
ZII	Infection zone
ZIII	Nitrogen fixation zone
ZII- III	Inter zone
ZII-D	Distal infection zone
ZII-P	Proximal infection zone

MAMP	Microbial associated Molecular pattern
NGS	Next generation sequencing
EST	Expression sequence tag
RAST	Rapid Annotation using subsystem technology
GEO	Gene expression Omnibus
MEGA	Molecular evolutionary Genetics Analysis
ORF	open reading fram
COG	cluster of orthologus genes

Introduction

Les plantes constituent une niche écologique favorable pour les différents microorganismes. Ces interactions peuvent être sans effet sur leurs développement et leurs croissance, bénéfique dans le cas des symbioses ou au contraire peuvent avoir un effet négatif lors d'interactions avec des pathogènes. (Gastroenterology, 2017)

La symbiose a souvent été considérée comme un processus radicalement différent de la pathogénicité. De nombreuses études montrent que ces deux phénomènes font intervenir des mécanismes moléculaires similaires, ainsi la perception des facteurs nod des rhizobia implique des récepteurs de l'immunité et inversement, des récepteurs symbiotiques sont requis pour la réponse à des pathogènes. (Gastroenterology, 2017)

Avec l'émergence de la génomique d'autres approches ont été développées notamment par les évolutionnistes et consiste à comparer les génomes de plusieurs organismes afin d'identifier les points communs et divergents.

Afin de comprendre les connexions entre immunité et pathogénicité, Notre projet vise à étudier les différences entre des souches bactériennes pathogènes, symbiotiques, extracellulaires ou intracellulaires afin d'identifier les points communs et divergents entre ces différentes bactéries et de comprendre l'évolution des différentes associations ainsi que de dégager des gènes candidats afin de réaliser des analyses fonctionnelles.

Pour cela plusieurs approches ont été utilisées, la première, génomique a pour but de mettre en évidence les différences au niveau des gènes ainsi que de l'organisation des génomes. La seconde, transcriptomique vise à étudier ces différences au niveau de l'expression génique. A cela s'ajoute des analyses d'interaction protéine-protéine et d'évolution afin d'identifier d'émettre et de valider des hypothèses concernant les associations symbiotiques et pathogènes.

Partie I :

Synthèse

Bibliographique

Chapitre I. Généralités sur les interactions plantes-microorganismes

Les plantes interagissent tout au long de leur vie avec des microorganismes environnementaux. Ces interactions déterminent le développement, la nutrition et la forme physique des plantes dans un environnement dynamique et stressant, formant la base du concept d'holobionte dans lequel les plantes et les microbes associés aux plantes ne sont pas considérés comme des entités indépendantes, mais comme une unité évolutive unique. L'une des principales questions en suspens concerne le point de savoir si la structure de l'holobionte est formée par ses membres microbiens ou uniquement par la plante. Les connaissances actuelles sur les interactions plante-microbe soutiennent que l'établissement de la symbiose conditionne directement et indirectement le microbiome associé à la plante.

La rhizosphère : réservoir de diversité

La rhizosphère est un microenvironnement de sol particulier dans lequel les propriétés du sol, les caractéristiques des plantes-racines et des microorganismes et leurs activités interagissent de manière coordonnée. Toutes les fonctions biologiques des racines des plantes, telles que l'absorption, la respiration et l'exsudation, exercent leurs effets dans la rhizosphère en modifiant les paramètres biogéochimiques du sol (concentrations en éléments nutritifs, polluants, concentrations en composés complexants ou chélateurs, pH et potentiel d'oxydoréduction, pressions partielles d'O₂ et de CO₂), etc.). (Gianfreda, 2015)

Au sein de ce microenvironnement, il existe des microorganismes bénéfiques tels que les PGPR) mais aussi néfastes tels que les pathogènes (*Ralstonia solanacearum*, *Agrobacterium tumefaciens*...etc). Il est donc important pour les plantes de contrôler leur microbiote afin de favoriser le développement de microorganismes bénéfiques et de défavoriser les néfastes.

2. Mécanismes de contrôle de la composition rhizosphérique

Le contrôle de la composition rhizosphérique par les plantes est essentiel afin de survivre au sein de milieux souvent hostiles (Mur et al. 2008). Les végétaux utilisent différents mécanismes afin de favoriser la multiplication de microorganismes bénéfiques dans la zone environnant la racine, par exemple en sécrétant des molécules dérivées du métabolisme secondaire agissant comme chimio-attractant ou par la sécrétion de polysaccharides et de molécules azotées via par exemple des cellules spécialisées de la coiffe racinaire, les BCL (Border Cell-Like) (Broghammer et al. 2012) qui après détachement de la racine, restent vivants jusqu'à 14 jours (Figure 1). Les plantes peuvent allouer jusqu'à 11% de carbone issus de la photosynthèse et 6 à 10 % d'azote organique, ce qui est une dépense considérable et qui sous-entend l'importance accordée par les plantes à ce phénomène (Bourcy et al, 2013).

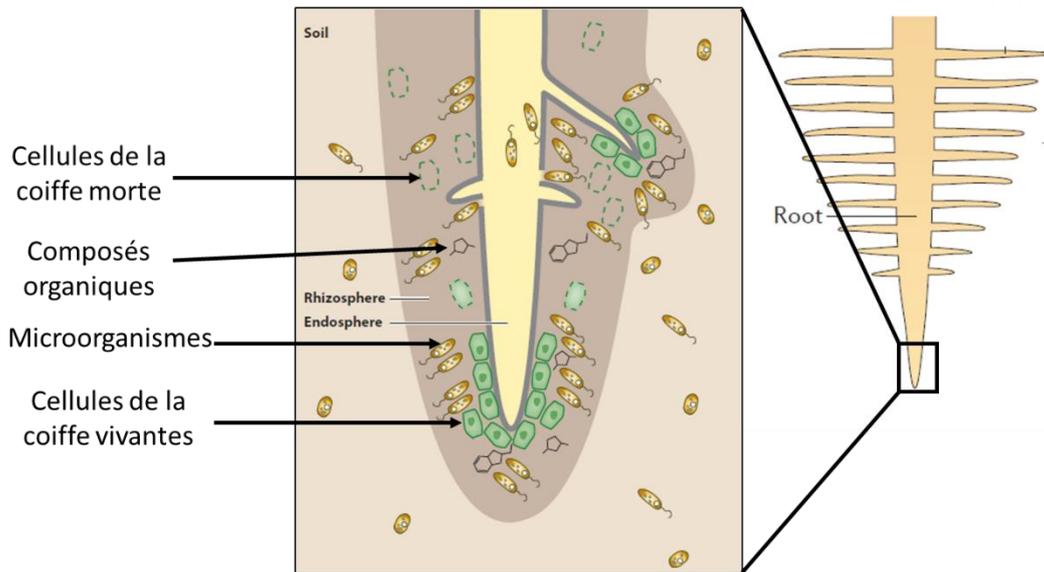


Figure 1 :La coiffe racinaire représente le premier tissu colonisé par les microorganismes ainsi que l'un des tissus les plus sensibles. Les plantes présentent une forte activité de sélection à ce site afin de favoriser les microorganismes bénéfiques, notamment par la sécrétion de composés chimio-attractants et de nutriments. Ce phénomène est en partie assuré par des cellules de la coiffe qui se desquament lors du processus de pénétration dans le sol. (Bulgarelli et al, Annual Review 2013).

3. Les interactions plantes-pathogènes :

Les plantes sont les cibles de différents types de pathogènes : virus, bactéries, champignons et oomycètes. Ces pathogènes attaquent, dans certains cas, des espèces d'intérêt agricole et peuvent infliger de fortes pertes pour le secteur agricole (FAO, 1983).

De par la nature pauvre des sols, les microorganismes colonisant ces milieux présentent une forte compétition, entre-autre pour l'acquisition des nutriments d'origine végétale

La plupart des agents pathogènes présents dans le sol doivent se développer de manière saprophyte dans la rhizosphère pour atteindre leur hôte ou être suffisamment nombreux sur celui-ci a

vant de pouvoir infecter les tissus de ce dernier (Berendsen, Pieterse, & Bakker, 2012). Le succès d'un agent pathogène dépend de la communauté microbienne du sol dans lequel se produit l'infection et donc à fortiori du degré de compétition présente au sein du biotope (Berendsen et al., 2012)

4. Mode d'infection des phytopathogènes :

Les plantes présentent des lignes de défenses passives telles que la cuticule (Serrano et al. 2014). Les pathogènes utilisent différentes stratégies afin de les surmonter ou de les contourner et d'infecter ainsi les plantes. Ils peuvent pénétrer par les stomates ou les hydathodes, qui sont des orifices naturels situés sur les feuilles dont la fonction est de permettre les échanges gazeux avec l'environnement et l'exsudation des gouttelettes d'eau lors du phénomène de guttation (Berrabah 2016). Le pathogène bactérien *Pseudomonas syringae* utilise les stomates pour infecter ces hôtes (Melotto et al. 2006) alors que certains *Xanthomonas campestris* empruntent préférentiellement les hydathodes (Hugouvieux et al. 1998). Les blessures représentent aussi une voie d'entrée. Les craquelures situées à la base des émergences des racines latérales sont par exemple utilisées par le phytopathogène racinaire *Ralstonia solanacearum* (Genin 2010). Enfin les pathogènes peuvent pénétrer via des régions sensibles tels que les fleurs ou les méristèmes apicaux. A titre d'exemple, les méristèmes racinaires sont également un des sites d'infection pour *R. solanacearum* (Genin 2010). Certains pathogènes sécrètent des enzymes lytiques telles que les pectinases et les cellulases qui digèrent les parois cellulaires et facilitent le processus infectieux (Kubicek et al. 2014).

Peu d'informations expérimentales sont actuellement disponibles sur les premières phases de l'infection et les données disponibles concernant le rôle de la mobilité bactérienne (flagelles) et de la chemotaxie font encore l'objet de controverses (Moens et al. 2014). Cependant il est clair que la capacité d'attachement d'une bactérie sur les tissus végétaux est un facteur qui influe fortement sur le développement ultérieur de l'interaction pathogène. Ainsi il a été démontré qu'un mutant *attR* d'*Agrobacterium tumefaciens* qui a perdu la capacité à synthétiser des fibrilles de cellulose impliquées dans l'attachement aux cellules végétales, est fortement affecté dans sa capacité à coloniser le système racinaire des plantes. Un tel mutant a également pratiquement totalement perdu la capacité d'induire le développement de tumeurs sur des plantes de tomate (Matthysse et al. 2014).

5. Mécanismes moléculaires et déterminants du pouvoir pathogène :

Il existe dans la nature une grande diversité de mécanismes moléculaires permettant aux pathogènes d'exercer leurs effets nocifs sur les plantes. Ils peuvent notamment perturber l'immunité de l'hôte à travers l'utilisation de composés de surfaces tel que les exopolysaccharides et de petites protéines appelées effecteurs, dans certain cas ils sécrètent des enzymes et des toxines permettant la destruction des tissus de l'hôte, enfin une perturbation de

l'absorption de certains ions peut être observée, ce qui affecte le métabolisme et l'immunité cellulaire (Figure 2) (Boucher, Genin, & Arlat, 2002)

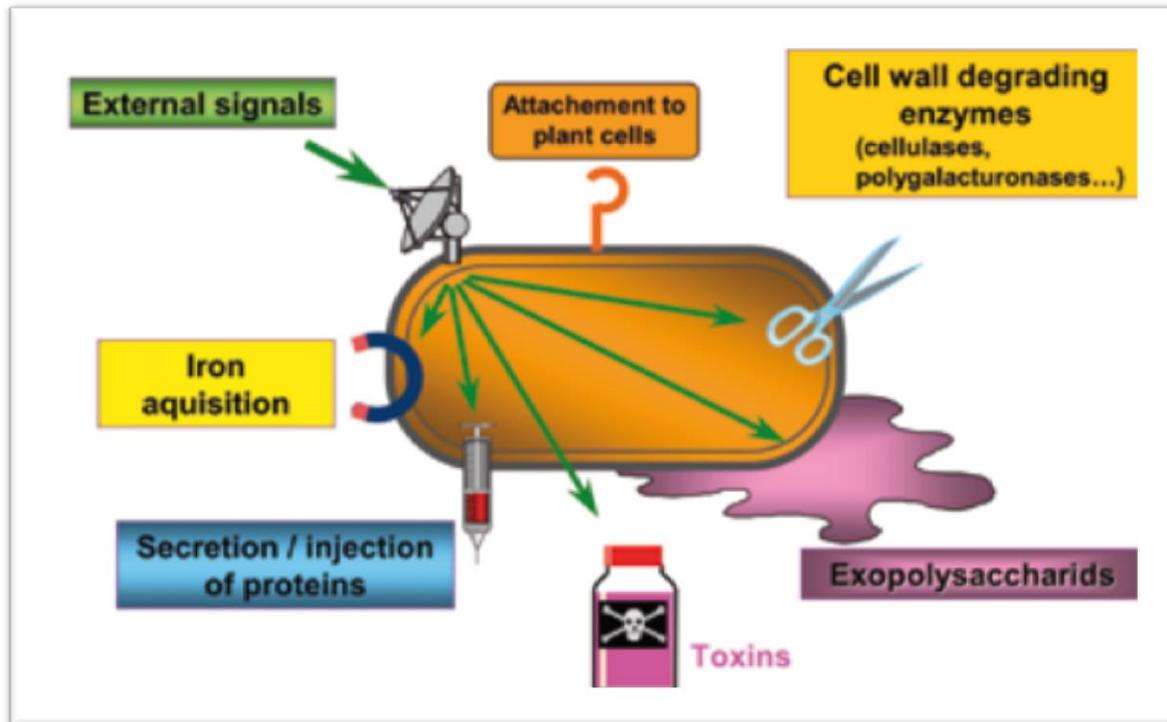


Figure 2 : Représentation schématique des différents facteurs impliqués dans le pouvoir pathogène des bactéries (Humaine, 2011)

5.1 Les exopolysaccharides (EPS) :

Les exopolysaccharides (EPS) bactériens constituent des systèmes modèles importants pour l'étude de l'assemblage et de la sécrétion moléculaire, de la régulation des gènes, des interactions cellule-cellule, de la symbiose et de la pathogénicité (John A. Leigh, 1992). Les EPS sont des polymères glucidiques sécrétés par une grande variété de bactéries, ils restent associés à la paroi cellulaire pour former une capsule liée à la paroi bactérienne, où ils peuvent être libérés dans le milieu extracellulaire (John Leigh, 1992).

Plusieurs espèces de bactéries phytopathogènes appartenant notamment aux genres *Xanthomonas*, *Erwinia*, *Pseudomonas* et *Ralstonia* produisent des quantités importantes d'exopolysaccharide en culture pure ou lors de leur multiplication dans les plantes (Cangelosi et al., 1987). Ils ont souvent été proposés comme déterminant important de la pathogénie, ces molécules agissent essentiellement en tant que stimulateur de l'infection (Van Sluys et al., 2002).

Chez *R. solanacearum* plusieurs études ont montré que ce pathogène produit un EPS à caractère acide (Boucher, Genin, & Arlat, 2001). Des mutants de cette bactérie spécifiquement affectés dans la production de ce composé sont capables d'infecter et de se multiplier dans la

plante, mais ils n'induisent pas les symptômes de flétrissement caractéristiques de la maladie (Boucher et al., 2001). Ceci suggère que les EPS interviennent uniquement dans les stades tardifs de la maladie, en induisant un flétrissement mortel de l'hôte (Denny, 1995).

Les EPS permettent aussi à la bactérie de se développer sur les tissus de l'hôte en décomposition et de repasser à une phase saprophytique favorable à la dissémination bactérienne. (Boucher et al., 2001). Ils jouent un rôle protecteur pour la bactérie vis-à-vis des réactions de défense de la plante (Pritchard & Birch, 2011) en intervenant notamment en tant que chélateur de calcium et en bloquant la pénétration au sein de la cellule du Ca^{+2} extracellulaire essentiel à l'activation de la réponse immunitaire de l'hôte (Denny, 1995)

5.2 Les toxines :

De nombreuses souches toxigènes ont été décrites chez les phytopathogènes, notamment dans le genre *Pseudomonas*. Celles-ci diffèrent cependant quant à la nature chimique des toxines produites et par l'activité anti-métabolique qui leur est associée (Uppalapati et al., 2008).

À titre d'exemple la phaseolotoxine produite par *P. syringae Pv. Phaseolicola*, de nature peptidique agit comme inhibiteur de l'ornithine carbamoyl transférase qui est une enzyme impliquée dans la biosynthèse de l'arginine. La toxine, qui diffuse librement au sein des tissus végétaux est directement responsable des symptômes de chlorose systémique caractéristique de la maladie dite de « graisse du haricot » (Mansfield et al., 2012).

La coronatine est une toxine largement décrite dans la littérature, cet analogue du jasmonate (une hormone immunitaire) est produit par certaines souches de *Pseudomonas* afin de stimuler la voie du jasmonate qui est antagoniste de l'acide salicylique, hormone essentielle pour la réponse à *Pseudomonas* (Urtuvia, 2017) .

Il convient de noter ici que l'aptitude d'une souche bactérienne à produire une toxine est toujours associée à la présence, chez cette bactérie, de fonctions qui lui permettent d'échapper aux effets délétères de sa propre toxine. Très souvent ces mécanismes de résistance font intervenir des formes résistantes de la cible moléculaire de la toxine ou l'absence de cette cible chez la souche productrice (Hatziloukas et al., 1992).

5.3 Les enzymes lytiques :

Certains microorganismes secrètent des enzymes lytiques tel que les pectinases qui dégradent les parois des cellules végétales et facilite ainsi la pénétration des pathogènes. D'autre enzyme tel que les chitinases dégradent la chitine extracellulaire libérée par les parois des champignons pathogènes et permet de réduire la réponse immunitaire de l'hôte déclenché par ces molécules (Boccaro et al., 1988).

5.4 Les effecteurs :

Les effecteurs sont des molécules de nature protéique injectées directement dans la cellule de l'hôte qui ont pour but de perturber son homéostasie cellulaire (Popa, Tabuchi, & Valls, 2016). Ils interviennent notamment dans le court-circuit de la signalisation immunitaire (Galán & Collmer, 1999). Les effecteurs sont délivrés par les bactéries via des systèmes de sécrétions, dénommé « seringue moléculaire », ils transpercent la membrane de l'hôte pour connecter le cytoplasme de la bactérie à celui de l'hôte. (Boucher et al., 2001). La pénétration des effecteurs fongiques au sein des cellules végétales est actuellement mal caractérisée, une hypothèse suggère qu'ils sont endocytés par l'hôte .

6. Exemple d'agents pathogènes

Un grand nombre de bactéries, de champignons et d'oomycètes sont capables d'infecter les plantes, causant des symptômes allant de l'altération de la croissance à la mort de l'individu infecté (Berrabah F, 2016).

6.1. *Agrobacterium tumefaciens*

a/Description du pathogène

Agrobacterium tumefaciens est une bactérie de type Gram (-) de la classe des α protéobactéries. Ce pathogène est responsable de la maladie de la tumeur du collet caractérisée par la production d'une masse résultant de l'hyper prolifération des cellules du cortex au niveau de la jonction racine-tige des hôtes infectés (Boivin, Fonouni-farde, & Frugier, 2016). Ce pathogène vie à l'état saprophyte sur les racines de l'hôte et n'enclenche le processus de pathogénicité qu'après blessure des racines. (Gohlke & Deeken, 2014).

b/ Plasmide Ti à la base la virulence

La virulence de ce pathogène est contenu dans un plasmide appelé Ti (Tumor inducer) contenant les gènes de virulence ou gène Vir, un gène de catabolisme d'opines qui sont des composés polycarbonnés utilisés pour la croissance bactérienne, enfin un cluster composé de gènes oncogènes et de biosynthèse d'opine est délimité au sein du plasmide par deux séquences appelées bordures droite et gauche. La partie d'ADN contenue entre les bordures est appelé région T du plasmide Ti (Figure. 3) (Gohlke & Deeken, 2014).

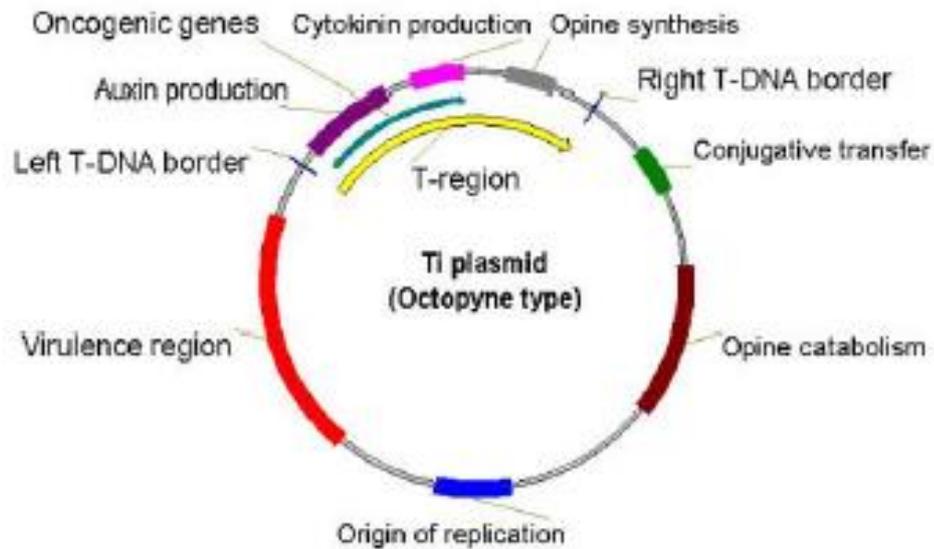
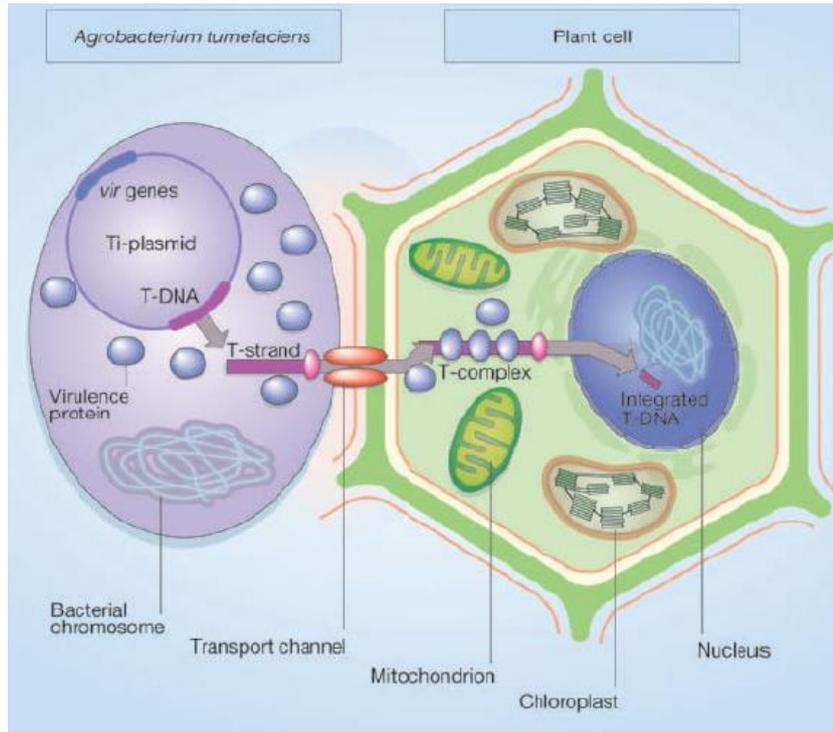


Figure 3. Le plasmide Ti d'*A. tumefaciens*. Le plasmide Ti est composé d'une origine de réplication bactérienne, d'une région de virulence, de gènes du catabolisme des opines, d'une région de conjugaison. Entre les séquences bordantes gauches et droite, un cluster composé de gènes oncogènes (production d'auxine, production de cytokinine) et de biosynthèse d'opine, constitue la région T du plasmide (Gohlke & Deeken, 2014).

c/. Mode d'infection par *A. tumefaciens*

Après blessure de la racine le processus de pathogénicité est enclenché par *A. tumefaciens*, qui est caractérisé par l'expression des gènes *Vir*, qui vont procéder à la réplication de l'ADN entouré par les bordures gauche et droite (ADNT), puis à son transfert vers les cellules de la plante. Une fois arrivé, l'ADNT est conduit vers le noyau cellulaire, aussi duquel il est intégré à l'ADN génomique végétal, suite à cela les gènes composant l'ADNT sont exprimés (Gohlke & Deeken, 2014).

Les gènes de biosynthèse d'auxine et cytokinine vont permettre de libérer des hormones végétales de même nom, responsables de la multiplication cellulaire et de la production de la tumeur, alors que le gène de biosynthèse des opines vas permettre la production de ce composé qui est utilisé comme source d'énergie par le pathogène (Figure 4).



1

Figure 4 : Processus d'infection par *A.tumefaciens*. Lors du processus infectieux le pathogène transfère l'ADN T sous l'action des protéines de virulence issues des gènes *Vir*. L'ADNT est conduit vers le noyau cellulaire de l'hôte au sein duquel le cluster de gènes est exprimé. Issue de (Gohlke & Deeken, 2014).

6.2 *Ralstonia solanacearum*

a/ Présentation du pathogène

Ralstonia solanacearum est une bactérie de la classe des α -protéobactéries, de gram (-) ce pathogène est un agent de la pourriture bactérienne chez les plantes. Largement répartie dans les régions subtropicales, tropicales ainsi que les régions chaudes des climats tempérés. (Genin, 2010). ce pathogène peut infecter une large gamme de plantes d'intérêt agronomique et constitue un agent de mise en quarantaine (Genin, 2010).

b/ Mode d'infection

Ce pathogène pénètre les racines de l'hôte via le méristème et les jonctions entre la racine principale et les racines secondaires. Une fois dans l'hôte ; la bactérie se multiplie au sein des tissus vasculaires et se propage dans l'individu conduisant ainsi à sa mort (Genin, 2010). Le pouvoir pathogène d'une souche diffère d'un hôte à l'autre, par exemple, une souche infectant un hôte ne sera pas obligatoirement pathogène chez une autre plante (Sevellec, 2018).

6.3. *Pseudomonas syringae*

a/ Présentation du pathogène

Les bactéries appartenant au groupe *Pseudomonas* sont des habitants habituels du sol et de l'eau et peuvent également se trouver à la surface des plantes et des animaux. (Uppalapati et al., 2008). Les bactéries *Pseudomonas* se trouvent dans la nature sous forme de biofilm ou de plancton. Dans cette dernière forme, la cellule unique peut afficher une très grande mobilité en raison de ses flagelles polaires. Les bactéries *Pseudomonas* sont réputées pour leur polyvalence métabolique car elles peuvent se développer dans diverses conditions de croissance et ne nécessitent aucun facteur de croissance organique. (Uppalapati et al., 2008) Ce sont généralement des agents phytopathogènes, bien que plusieurs souches soient pathogènes pour l'homme. *Pseudomonas syringae* comprend de nombreux agents phytopathogènes de cultures importantes, ce qui en fait un organisme modèle en phytopathologie. (Monteil, 2011) .

b/. Mode d'infection

Une fois que *P. syringae* a colonisé la surface extérieure de la plante (feuille en particulier), elle pénètre les espaces intercellulaires par des lésions ou des ouvertures naturelles. Le pouvoir pathogène de *P. syringae* s'exprime alors suivant deux processus : (i) sécrétion de protéines effectrices à l'intérieur des cellules des plantes hôtes pour moduler les fonctions de la plante par des systèmes de sécrétion transmembranaires ou (ii) la production d'exopolysaccharides, de phytotoxines, de phytohormones et d'enzymes dégradant les parois cellulaires. (Monteil, 2011)

7. Mécanismes moléculaires de résistance chez les plantes

Au cours de leurs évolutions les plantes ont développées une multitude de mécanismes afin de reconnaître les pathogènes et d'apporter une réponse adéquate (Bigéard et al. 2015). On peut distinguer deux niveaux de reconnaissance associés à deux amplitudes de réponse : la première d'amplitude faible à lieu suite à la connaissance de motif moléculaire conservée entre les microorganismes est nommée PTI (PAMP Triggered Immunity). La deuxième nommé ETI (Effector Triggered Immunity) d'amplitude forte se met en place suite à la reconnaissance d'effecteurs (Bigéard et al. 2015).

7.1 La PAMP Triggered Immunity (PTI)

Lors de ce processus les microorganismes sont perçus par les plantes via des récepteurs immunitaires appelés PRR (Pathogens Recognition Receptor) qui reconnaissent des motifs

conservés entre les pathogènes nommés PAMP (Pathogen Associated Molecular Pattern). Par exemple le récepteur FLS2 (FLagelin Sening 2) se lie avec le co-récepteur BAK1 (Brassinosteroid Associted Kinase 1) au peptide flg22 de la flageline bactérienne (Berrabah et al, 2018). Suite à l'activation des PRR plusieurs phénomènes cellulaires ont lieu notamment : une dépolarisation des membranes plasmique, un influx de calcium, une modification du statut redox cellulaire dû à la production de radicaux libre de l'oxygène, des phénomènes de phosphorylation de protéines immunitaires (Zipfel 2013). Ces phénomènes aboutissent à l'expression de gènes immunitaires et la libération de molécule antimicrobiens (Bigéard et al. 2015).

7.2. L'Effector Triggered Immunity (ETI)

Les pathogènes emploient des effecteurs afin de court-circuiter la réponse immunitaire de l'hôte. Si le pathogène arrive à bloquer l'immunité, on parlera d'ETS (Effector Triggered Sensitivity) (Bigéard et al. 2015). Les plantes emploient des récepteurs intracellulaires appelés NBS-LRR qui peuvent reconnaître l'effecteur (Meyers et al. 2003). Dans le cas où le récepteur reconnaît l'effecteur une ETI a lieu qui consiste à la réactivation et l'amplification de la réponse immunitaire aboutissant entre-autre à la mort cellulaire programmée des cellules infectées (Meyers et al. 2003). Lors de l'ETI, le gène codant l'effecteur sera appelé gène d'avirulence et celui codant le récepteur gène de résistance (Bigéard et al. 2015).

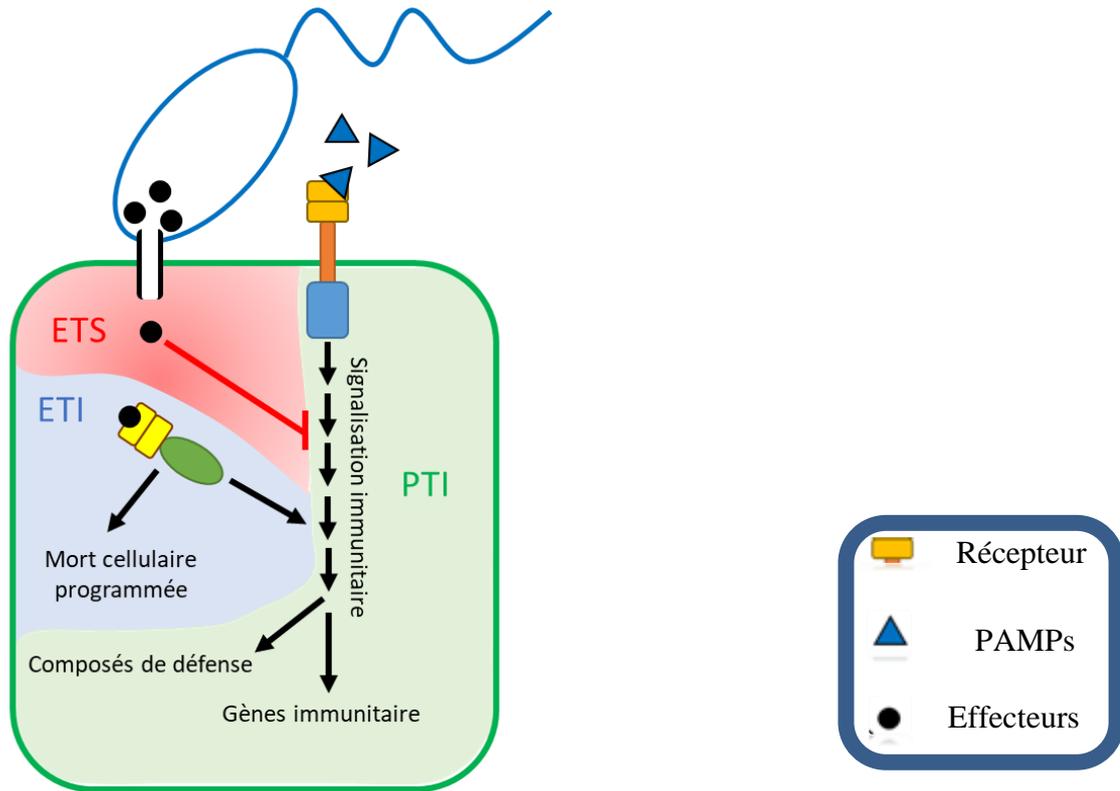


Figure 5 : ETI et PTI suite à la perception des PAMPs immunitaire qui aboutit à la libération de composés de défense et l'expression de gènes immunitaires, on parlera de PTI. Les pathogènes peuvent injecter des effecteurs qui bloquent la signalisation immunitaire conduisant à une ETS. Si l'hôte présente les récepteurs NBS-LRR adéquat, la perception de l'effecteur par ce dernier conduit à la réactivation de l'immunité et à l'ETI. (Berrabah et al., 2014).

Chapitre II. La symbiose fixatrice d'azote

Introduction à la symbiose légumineuses-rhizobia

Les légumineuses sont des plantes de la famille des fabacées, en condition de carence azoté elles ont la capacité d'établir une interaction symbiotique avec des bactéries du sol fixatrices d'azote, les rhizobia (Berrabah et al., 2014). Au cours de cette interaction un nouvel organe racinaire est formé, la nodosité fixatrice d'azote au sein de laquelle les bactéries fixent l'azote atmosphérique pour le bénéfice de leur hôte (Vasse et al., 2014).

Au cours de l'interaction deux évènements sont indispensables pour la mise en place de la nodosité ; **i**) l'infection racinaire, au cours de laquelle les rhizobia pénètrent dans les racines de la plante via les poils absorbants en formant des cordons d'infections, **ii**) parallèlement à cela, un mécanisme d'organogénèse est enclenché, il correspond à une réactivation de la division des cellules du cortex (Xiao *et al.*, 2014). Au cours de la symbiose, une infection massive et chronique des tissus de l'hôte par les bactéries est observée, malgré cela, la plante ne produit aucune réponse immunitaire, cette situation reste exceptionnelle dans la nature (Kereszt et al., 2014).

2. Physiologie moléculaire de la symbiose fixatrice d'azote

2.1. La phase de reconnaissance plante-bactérie

Quand l'azote organique est absent ou quand il se présente en concentration réduite dans le sol, les légumineuses produisent certains flavonoïdes (Van et al., 2006), composés issus du métabolisme secondaire perçus de manière spécifique par les rhizobia. En retour ces dernières produisent des facteurs nods (NFs) qui sont secrétés dans la rhizosphère, au sein de laquelle ils sont perçus par les plantes au niveau de l'épiderme racinaire.

Les NFs sont perçus au niveau des membranes plasmiques des cellules racinaires par des Recepteur-Like Kinase (RLK) possédant un domaine extracellulaire de liaison au LCOs, appelé domaine Lysine Motif (LysM) (Timmers et al., 2000). Suite à la liaison des NFs aux récepteurs correspondant, une voie de signalisation cellulaire est activée et conduit à des modifications de la physiologie des cellules hôtes aboutissant à l'activation de l'expression de gènes spécifiques et indispensables pour la mise en place de la nodosité, ces gènes sont nommés nodulines précoces ou Early NODuline (ENOD) (Timmers et al., 2000).

2.2. L'infection des racines

Suite au phénomène de reconnaissance, les rhizobia infectent les racines de l'hôte en formant un cordon d'infection (CI) qui passe via les poils absorbants et croit en direction du cortex

(Xiao *et al.*, 2014). Un fois que le cordon d'infection atteint les cellules cibles du cortex, les bactéries sont reléguées au sein des cellules (Timmers *et al.*, 2000).

Après leur endocytose au sein des cellules de l'hôte, le bacteroïde qui est la forme intracellulaires des rhizobia, est entourés par la membrane plasmique de l'hôte, appelée membrane peri-bacteroïdienne (MPB) et un espace peri-bacteroïdien (EPB) sépare le bacteroïde de la MPB, l'ensemble est nommé symbiosome (Ane *et al.*, 2004).

Chez les légumineuses du groupe des IRLC telle que *Medicago truncatula*, après l'infection des tissus, les bacteroïdes enclenchent un phénomène de différenciation terminal indispensable pour la fixation d'azote (Ane *et al.*, 2004). Au cours de ce processus le matériel génétique de la bactérie subit une endoreduplication, cela est associé à une augmentation de la taille du bacteroïdes. (Popp *et al.*, 2011). A la fin du processus, la bactérie perd toute capacité de survivre en dehors de son hôte, cela est aussi qualifié par certains scientifiques de « processus d'esclavagisme » (Mitra *et al.*, 2004).

2.3. NCRs & différenciation des bacteroïdes

La différenciation est médiée par des petits peptides antimicrobiens ; les Nodule Cystein Rich (NCRs) (Popp *et al.*, 2011). *M. truncatula* présente près de 600 gènes codant des NCRs au sein de son génome (Smit, 2005), la majorité d'entre eux sont exprimés au cours de la symbiose et ils présentent une expression spatio-temporelle différentielle (Kalo, 2005).

In planta, les NCRs sont adressés au symbiosome au sein duquel l'endopeptidase ; DNF1 est requise pour leur internalisation au sein de l'EPB (Vernié *et al.*, 2015). Les NCRs présentes un effet *in vitro* bactéricide (Cerri *et al.*, 2012). L'ABC transporteur BacA du symbionte de *M. truncatula*, *Sinorhizobium spp* est requis pour la survie intracellulaire des bacteroïdes (Cerri *et al.*, 2016). De plus le mutant *bacA* présente *in vitro*, une augmentation de sensibilité au NCRs par rapport à la souche sauvage (Cerri *et al.*, 2012).

Il est postulé que BacA est requis pour l'internalisation des NCRs au sein du bacteroïde, dans lequel leur pouvoir toxique est annulé (Cerri *et al.*, 2012). Au sein du microorganisme les NCRs peuvent cibler certaines fonctions biologiques afin d'induire la différenciation, par exemple le NCR247 interagit avec la protéine FtsZ requise pour la formation du septum bactérien et potentiellement inhibe par ce mécanisme la division cellulaire (Reid *et al.*, 2016) (Figure 6).

2.4. Survie intracellulaire des bacteroïdes :

Actuellement peu d'information sont disponibles quant aux différents mécanismes permettant au bacteroïdes de s'adapter à la cellule de la plante, mais l'aspect lié au NCR suggèrent que la bactérie subit un stress considérable conduisant à sa transformation. Par ailleurs certains

pathogènes intracellulaires, telle que *Brucella abortus* fait intervenir un homologue de BacA afin de réduire la toxicité des défensines intracellulaires, qui sont des peptides antimicrobiens immunitaires et ainsi d'échapper à l'hôte (Gibson, Kobayashi, & Walker, 2008). Ceci indique qu'il existe potentiellement des mécanismes partagés entre des bactéries très variées et permettant leurs adaptation aux cellules eucaryotiques (Gibson et al., 2008).

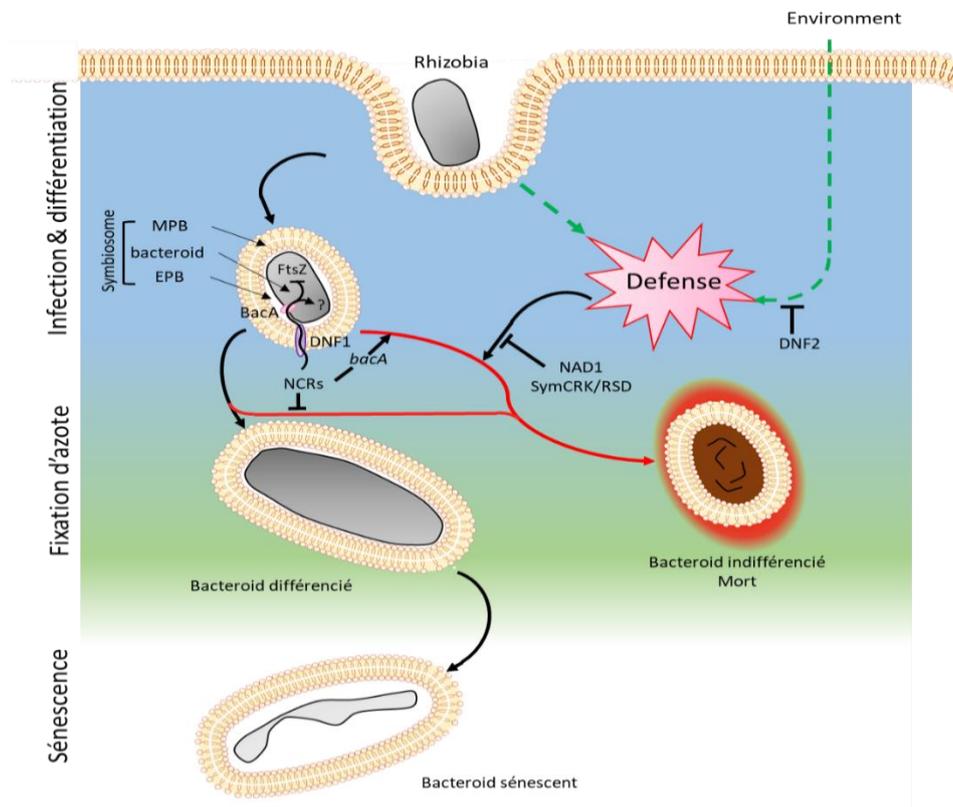


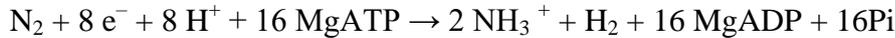
Figure 6 : L'infection & la différenciation. Lorsque la bactérie pénètre dans la cellule on parlera de bacteroid, ce dernier est entouré par une membrane peri-bacteroidienne (MPB) et séparé d'elle par un espace peri-bacteroidien (EPB), l'ensemble forme un symbiosome. Au cours de la différenciation les NCRs sont adressés au symbiosome via des peptidases telle que DNF1 chez *M. truncatula*. Les NCR pénètrent dans la bactérie via le transporteur BacA. Une fois dans le bacteroid ils interagissent avec des protéines telles que FtsZ et provoque la différenciation des bacteroids en bacteroids fixateur d'azote. Chez le mutant *bacA* les NCR s'accumulent dans l'EPB et provoque la mort des bacteroids indifférenciés. Chez *M. truncatula* plusieurs gènes interviennent dans la répression de l'immunité telle que *DNF2*, *NAD1*, *SymCRK* et *RSD*. A la fin du processus les bacteroids deviennent sénescents et ils sont recyclés par l'hôte. (Berrabah et al., 2014).

2.4. La fixation d'azote point final de la symbiose :

Après le processus de différenciation, les bacteroides débutent la fixation d'azote, en convertissant l'azote atmosphérique (N_2) en NH_3^+ , ce dernier est converti à son tour en NH_4^+ avant d'être transmis à la cellule hôte (Udvardi & Poole, 2013). La plante intègre le NH_4^+

dans la biosynthèse des acides aminées en synthétisant de la glutamine à partir du glutamate suite à l'action de la glutamate synthase (Udvardi & Poole, 2013).

La fixation d'azote est assurée par la nitrogénase, cette enzyme convertie le N_2 en NH_3^+ selon l'équation suivante :



La nitrogénase est composée de deux enzymes ; une nitrogénase réductase, qui fournit les électrons à la seconde enzyme qui est la nitrogénase, cette dernière est un tétramère contenant un cluster fer-molybdate-cobalt (Rubio and Ludden .2016). Le complexe enzymatique est codé par le cluster de gène *NifHDK* ; *NifH* code la nitrate réductase et *NifDK* code la nitrogénase (Rubio and Ludden .2016). En plus des cluster nif, d'autres gènes tels que le cluster Fix sont impliqués dans la régulation de la fixation d'azote (Dixon & Kahn, 2004).

La nitrogénase est une enzyme sensible à l' O_2 , afin d'éviter la dénaturation de cette enzyme par l'oxygène, les légumineuses synthétisent la légghémoglobine (Lbs) qui est une protéine régulant la diffusion d' O_2 au sein de la nodosité et qui maintient une faible concentration d'oxygène intracellulaire (Ott et al., 2005). La Lbs confère une couleur rose à la nodosité, cette couleur est un signe de fonctionnalité de l'organe (Bourcy, Berrabah, et al., 2013).

3. Les nodosités de type déterminé & indéterminé

Les nodosités déterminées (ND ; telles que celles formées par *Lotus japonicus*) ont la caractéristique d'avoir un méristème diffus, d'activité réduite, localisée au tour de l'organe en dessous de l'épiderme racinaire. La localisation, ainsi que l'activité réduite du méristème des nodosités déterminées (ND) a pour principal effet de limiter la croissance dans le temps et l'espace de l'organe, aboutissant ainsi à une forme arrondie (Vasse, De Billy, Camut, & Truchet, 1990). A l'inverse, le méristème des nodosités indéterminé (NI ; telle que celle formé par *M. truncatula*) est localisé en dessous de l'épiderme dans la partie apicale de l'organe et présente une forte activité tout au long de la vie de la nodosité, cela aboutit à un allongement du nodule (Xiao et al., 2014) (Figure 7).

L'organisation tissulaire des ND est très différente de celles des NI ; on trouve au sein des nodosités déterminées un tissu homogène composé de cellules infectées et non infectées organisées à la périphérie et une zone central, dite médullaire (Vasse et al., 1990). Les nodosités indéterminées quant à elles présentent une zonation de leurs tissus : une première zone ou zone I (ZI) au sein de laquelle on trouve le méristème est observée, elle est suivie par la zone d'infection ou zone II (ZII), au sein de laquelle les rhizobies infectent les cellules de l'hôte, puis une zone de fixation d'azote ou zone III (ZIII) dans laquelle les bactéroïdes

(forme intracellulaire des rhizobia) fixent l'azote atmosphérique (N₂). La ZIII est séparée de la ZII par une inter-zone ou ZII-III dans laquelle les bacteroids subissent un processus de différenciation terminal (Kereszt, Mergaert, Maróti, & Kondorosi, 2011). Enfin quand la nodosité devient trop âgée, quand les plantes subissent un stress ou que de l'azote organique est ajouté à des plantes nodulées, une zone de sénescence ou zone IV (ZIV) apparaît, au sein de laquelle les bacteroids et les composants des cellules de l'hôte sont dégradés (Van de Velde et al., 2006), par ailleurs cette zone peut être recolonisée par des rhizobia saprophytes (Timmers et al., 2000) (Figure 7).

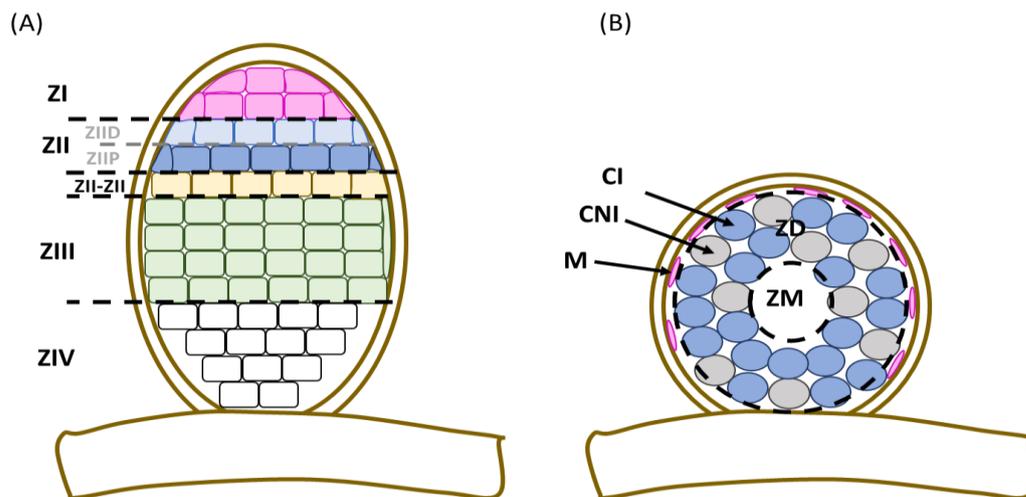


Figure 7 : Les nodosités indéterminées et déterminées. (A) nodosités de type indéterminé de forme allongée, elle est composée d'un méristème apical ou ZI, d'une zone d'infection ou ZII divisé en zone d'infection distale (ZIID) et proximal (ZIIP). La zone de fixation ou ZIII et une zone de sénescence ou ZIV. (B) les nodosités déterminées sont organisées en deux régions, une région distale (ZD) au sein de laquelle on trouve des cellules infectées (CI) et non infectées (CNI) et une région médullaire (ZM). Le méristème est diffus sous l'épiderme

4. L'immunité au cours de la symbiose

4.1. L'immunité dans les étapes précoces

Les rhizobia ont la capacité de stimuler des réponses immunitaires chez des plantes non-hôtes (non légumineuses) (Baier, 1999), mais aussi chez des plantes hôtes (Gourion, et al., 2010). Une induction de l'expression des gènes de défense a lieu aux premières étapes de l'interaction symbiotique (Maunoury et al., 2010). Une répression immunitaire est requise pour permettre la mise en place de la symbiose, elle est représentée par une répression massive de l'expression des gènes de défense (Maunoury et al., 2010). Différentes stratégies sont élaborées par les rhizobia pour éviter, réduire ou bloquer la stimulation de la réponse

immunitaire de l'hôte (Gourion et al., 2015) ; les exopolysaccharides contribuent à masquer les PAMPs/MAMPs (Microbial Associated Molecular Pattern) et réduisent la signalisation immunitaire en bloquant l'absorption du Ca^{+2} extracellulaire (Aslam et al., 2008). Les rhizobia au cours de leurs évolution ont modifiées certains MAMPs afin d'éviter une stimulation de la réponse chez l'hôte, par exemple *Sinorhizobium meliloti* présente un peptide flg22 avec une structure différente de la structure canonique de ce type de peptide (Lopez-Gomez, 2012). Des effecteurs symbiotiques dits Nodule Outer Protein (NOPs) peuvent être utilisés pour réprimer la réponse immunitaire de l'hôte par exemple : la souche NGR243 bloque la signalisation immunitaire en produisant des effecteurs symbiotiques qui sont injectés via le système de sécrétion de type III au sein des cellules hôte (Deakin and Broughton .2009). La plante de son côté réprime son immunité via la signalisation symbiotique ou la mise en place de gènes spécifiques (Gourion et al., 2015).

4.2. L'immunité dans les étapes tardives

Une fois la nodosité formée, plusieurs gènes impliqués dans la répression des défenses ont été décrits ; *NAD1*, *DNF2*, *SymCRK* et *RSD* code respectivement une protéine de fonction inconnue, une phospholipase C phosphatidyl inositol dépendante putative, un RLK de la famille des Domain Unknown Function 26 (DUF26) et un facteur de transcription de type C_2H_2 respectivement (Berrabah et al, 2016). Les mutants correspondants présentent une réactivation de la réponse immunitaire au sein de leurs nodosités caractérisés par une induction de l'expression de gènes de défense, l'accumulation de composés phénoliques et une mort des bacteroid(Gourion et al., 2015). Le mécanisme par lequel fonctionnent ces gènes est actuellement inconnu, mais certains facteurs influencent le phénotype des mutant ; quand *dnf2* est cultivé dans un milieu culture sans éliciteur de défense tel que l' ulvan (une molécule potentialisant les réponses immunitaires), le mutant produit des nodosités fixatrices d'azote, ce qui indique que l'environnement influence le phénotype du mutant (Berrabah et al. 2014). Les mutants *symCRK* et *rsd* présentent une forte réduction de la réponse immunitaire au sein de nodosité induite par le mutant de *S. meliloti bacA* qui incapable de survivre dans des cellules symbiotiques (Berrabah, Ratet, & Gourion, 2015), ce qui laisse à penser que l'infection correcte des tissus et/ou la différenciation des bacteroides sont requises pour la stimulation des défenses chez ces mutants (Figure6).

Chapitre III : Génomique et transcriptomique

1. La génomique :

La génomique est une discipline de la biologie qui s'attache à étudier la structure des génomes, leurs compositions en gènes et leurs évolution (Tambourin, 2005). Elle vise

notamment à mettre en place des bases de données disponible répertoriant les données génomiques des organismes, et microorganisme. Parmi les informations disponible : la composition en nucléotides de le l'ensemble de l'ADN pour un organisme, l'identification des zones du génome exprimées, l'annotation des gènes et l'identification de leurs fonction (Pierre Tambourin, 2005). De plus, les scientifiques développent des outils afin d'annoter et de comparer les génomes. On peut distinguer trois disciplines dans la génomique : la génomique structurale, la génomique fonctionnelle et la génomique comparative. (Pierre Tambourin, 2005).

1.1. La génomique structurale :

La génomique structurale consiste à identifier la composition en acide nucléique (séquence d'ADN) pour le génome d'un organisme donnée (Bidanel, Boichard, & Chevalet, 2008).

Parmi les premières approches en génomique, on peut citer, la cartographie génétique qui consiste à étudier de marqueurs génétiques évoluant au sein d'un groupe de liaison par analyse comparative entre différents individus (Tambourin, 2005). Par la suite les banques d'ADN génomique ont remplacé cette approche, couplée au séquençage bas débit telle que SANGER, elles ont permis de connaître des séquences d'ADN (Shendure & Ji, 2008). Actuellement une nouvelle génération de technologies est employée, il s'agit du séquençage haut débit (NGS : Next Generation Sequencing), à moindre coût, elles permettent d'accéder à l'information génomique de manière plus rapide, ce qui a participé à la démocratisation de la génomique (Besser, Carleton, Gerner-smidt, Lindsey, & Trees, 2018). Ces nouvelles technologies ouvrent des possibilités encore inexplorées, notamment pour l'analyse de la variabilité génétique au sein des populations (Bidanel et al., 2008).

1.1.1. Les banques d'ADN :

Une banque d'ADN est une collection complète de fragments d'ADN clonés provenant d'une cellule, d'un tissu ou d'un organisme. Les banques d'ADN peuvent être utilisées pour isoler un gène d'intérêt spécifique, car elles incluent généralement au moins un fragment contenant le gène. Elles peuvent être construites à l'aide de divers vecteurs (plasmides, phages, cosmides ou chromosomes artificiels) et sont généralement hébergées dans une population de cellules procaryotiques ou eucaryotiques (Carter & Shieh, 2010)

L'ADN à cloner peut être isolé de plusieurs manières. Il peut être synthétisé par PCR ou il peut être localisé sur le chromosome par Southern blot, ces deux méthodes nécessitent une certaine connaissance de la séquence nucléotidique de l'ADN à cloner (Figure 1).

L'objectif de la construction d'une banque d'ADN génomique est d'avoir le génome d'un organisme cloné sous forme de petits fragments dans des vecteurs distincts. Idéalement, le

génomique est représenté. C'est-à-dire que la somme des différents fragments équivaut à l'ensemble du génome. De cette manière, des groupes spécifiques de gènes peuvent être analysés et isolés.

La construction d'une bibliothèque génomique commence par le clivage du génome en petits morceaux par une endonucléase de restriction (Willey, Sherwood, & Woolverton, 2007).

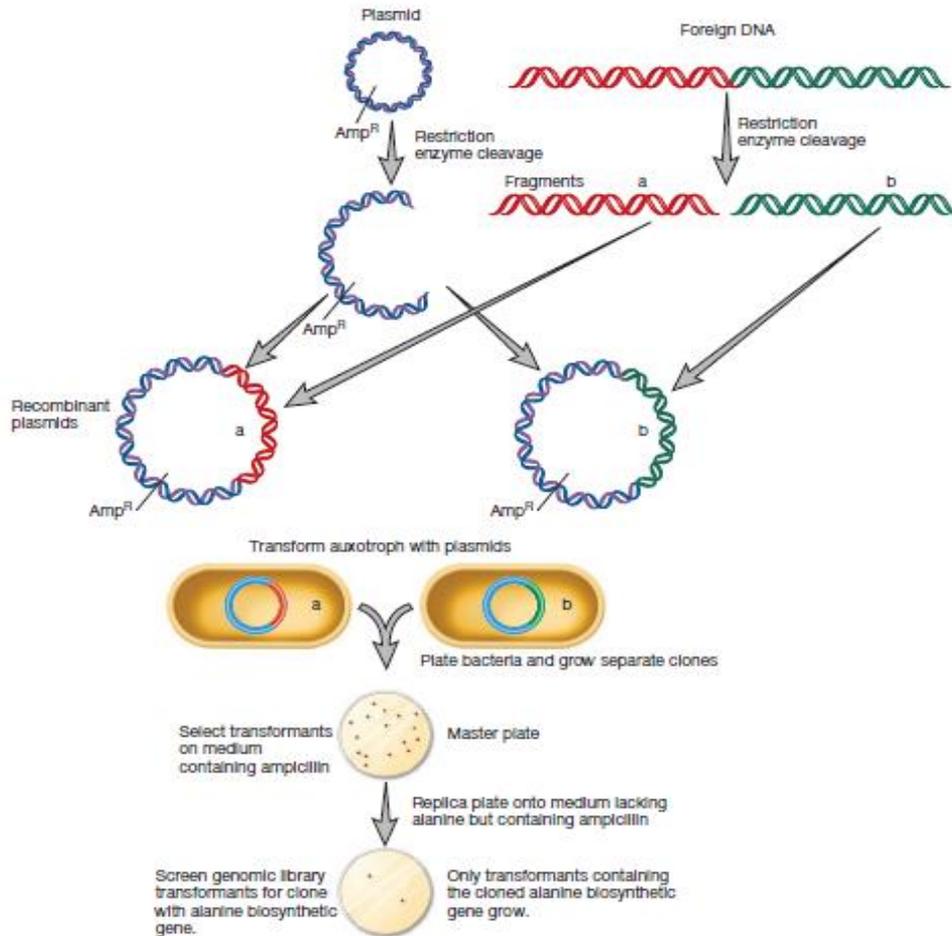


Figure 8 : Etapes relatives à la construction d'une banque génomique est réalisée par clonage des fragments du génome entier d'un organisme en un vecteur (Willey et al., 2007).

1.1.2. Le séquençage de l'ADN

Le séquençage de l'ADN est la détermination de la succession des nucléotides. Cette technique utilise les connaissances qui ont été acquises depuis une trentaine d'années sur les mécanismes de la réplication de l'ADN, il est maintenant possible de capturer des milliers de variables moléculaires. Cependant, ces techniques ont surtout été appliquées en laboratoire et dans des environnements contrôlés. La génération de tels ensembles de données volumineux appelées données «Omiques» - nécessite des considérations partiellement nouvelles pour les

montages expérimentaux, l'échantillonnage, l'analyse des données et la visualisation. (Alexandersson, 2014)

Au cours des dernières années, les technologies de séquençage d'ADN de nouvelle génération ont pris une importance croissante, avec l'adoption de plus en plus répandue de plates-formes qui implément individuellement différentes variantes de séquençage. Les caractéristiques communes s'étendent au-delà des technologies elles-mêmes, à la quantité et à la qualité des données générés, de sorte que toutes soulèvent un ensemble similaire de nouveaux défis pour la conception expérimentale, l'analyse et l'interprétation des données (Shendure & Ji, 2008) (Figure 9).

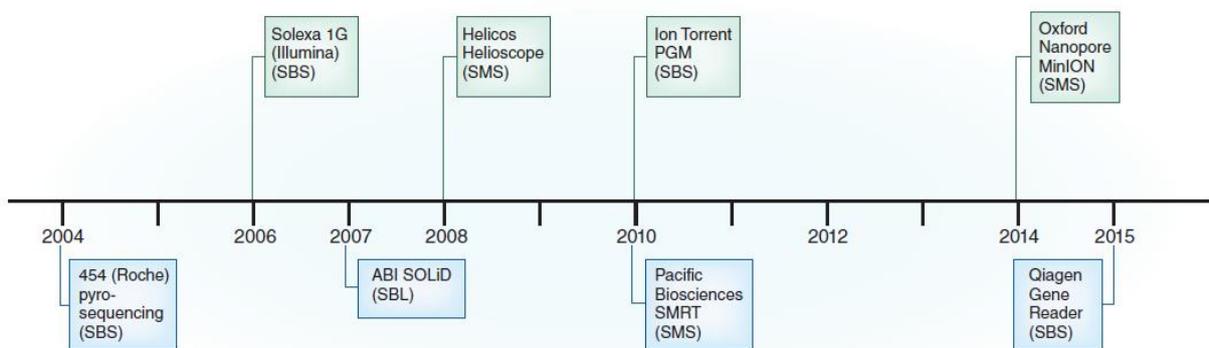


Figure 9 : Les instruments NGS introduits au cours des dernières années.

1.1.3. Les NGS: Next Generation Sequencing techniques

Le séquençage nouvelle génération (NGS) ou haut débit, est un terme commun utilisé pour décrire différentes technologies de séquençage moderne comme : l'Illumina, le Roche 454, l'ion torrent ou le SOLiD. Ces technologies récentes permettent de séquencer l'ADN beaucoup plus rapidement que des méthodes plus anciennes telle que le Sanger (Tambourin, 2005).

SYNTHESE BIBLIOGRAPHIQUE

Entreprise	Longueur de fragment	Applications
454/ roche	400 paires de base (extrémité unique)	1/Génomés bactériens et viraux 2/ validation de mutations ponctuelles, 3/détection ciblée de mutations somatiques
Illumina	150–300 paires de bases (extrémité appariée)	1/ Génomés complexes (humains, souris et plantes) et applications NGS à l'échelle du génome, produits d'ARN-seq, 2/ détection de mutation somatique, médecine légale
ABI SOLID	75 pb (extrémité unique) ou 50 pb (extrémité appariée)	1/Génomés complexes (humain, souris, plantes) et des applications NGS à l'échelle du génome, RNA-seq, 2/ produits de capture hybride ou PCR multiplex, détection de mutation somatique
Pacific Biosciences	Jusqu'à 40 kb (extrémité unique ou consensus circulaire)	1/ Génomés complexes (humain, souris et plantes), microbiologie et génomes de maladies infectieuses, 2/ détection par fusion-transcription, méthylation détection
Ion Torrent	200–400 pb (extrémité unique)	1/Produits de PCR multiplex, microbiologie et maladies infectieuses, mutation somatique 2/Détection, validation de mutations ponctuelles
Qiagen GeneReader	107 pb(extrémité unique)	1/Détection ciblée de mutations, biopsie liquide dans le cancer

Tableau 1 : Comparaison des techniques NGS disponibles

Le point commun entre les différentes technique NGS (Tableau 1), est qu'elles effectuent toutes le séquençage de millions de petits fragments d'ADN en parallèle. Les analyses bioinformatiques permettent de reconstituer ces fragments en superposant les lectures individuelles sur des génome de référence ou par détection de séquences chevauchantes(Steele et al., 2019)

Ces techniques ont considérablement augmenté la quantité de données de séquençage produites par instrument et ont considérablement réduit les coûts de production des données de séquençage. Couplées à des développements méthodologiques et informatiques innovants (Mardis, 2017).

Les défis informatiques consistent notamment à développer des outils pour l'intégration de tels ensembles de données, conduisant ainsi à une compréhension et à une prédiction plus poussées de la biologie à l'échelle du génome (Mardis, 2017).

2. La génomique fonctionnelle

La génomique fonctionnelle cherche à détecter les gènes et a leurs attribuer une fonction (Bidanel et al., 2008). Cette discipline emploie plusieurs approches afin de réaliser l'annotation, elles se basent toutes sur la comparaison de séquence (Bidanel et al., 2008).

L'identification d'EST (Expression Séquence Tags) permet de localiser les séquences exprimées au sein d'un génome, elle consiste à comparer la séquence de l'ADN génomique à celle d'ADN complémentaires issus de la reverse transcription des ARN messagers totaux de l'organisme étudié (Zhou et al., 2003). La comparaison entre des séquences de gènes dont la fonction biochimique est connue à celle du génome étudié permet quant à elle d'attribuer une fonction putative aux différentes séquences du génome (Bidanel et al., 2008)

La confirmation de la fonction de gènes annotés de manière bio-informatique peut se faire par la mise en évidence du rôle du gène dans un processus physiologique donné et cela par exemple en utilisant des approches de génétique directe ou reverse au cours desquels des mutants incapables de produire la protéine des gènes étudiés sont réalisés (Chandler et al. 2004) .L'analyse du phénotype des mutants permet de confirmer la fonction biologique des gènes annotés. Une autre approche consiste à purifier la protéine issue du gène et de tester son activité biochimique (May et al., 2008).

3. La génomique comparative

La génomique comparative consiste à comparer les séquence génomique de différents organismes afin de comprendre leurs fonctions et les relations évolutives à travers leurs points communs et leurs différences (Ailloud, 2015). L'une des premières études comparative de

génomés bactériens consiste à l'analyse de sept séquences du génome de bactéries associées à des plantes (d'*Agrobacterium tumefaciens*, de *Mesorhizobium loti*, de *Sinorhizobium meliloti*, de *Xanthomonas campestris pv campestris*, de *Xanthomonas axonopodis pv citri*, de *Xylella fastidiosa* et de *Ralstonia solanacearum*). L'étude a révélé que malgré l'importance reconnue du système de sécrétion de type III dans le contrôle de la compatibilité des bactéries à leurs hôtes, sa présence n'est pas universelle chez tous les agents pathogènes. Les hémolysines, les hémagglutinines et certaines adhésines, précédemment rapportées comme importantes pour les agents pathogènes de mammifères, sont présents dans la plupart des souches discutés. Différents nombres et combinaisons d'enzymes et de gènes dégradant la paroi cellulaire pour surmonter le stress oxydatif généralement induit par l'hôte végétal sont caractérisés dans ces génomes. Au total, 19 gènes ont été trouvés communs à toutes ces bactéries (Van Sluys et al., 2002).

3.1. Outils de génomique comparative

L'analyse computationnelle d'une séquence génomique implique essentiellement l'identification de gènes et l'attribution fonctionnelle putative (annotation de génome). Ces attributions reposent largement sur le concept de similarité de séquence. L'outil de détection de similarité le plus largement utilisé est le programme BLAST, généralement utilisé conjointement avec des bases de données accessibles au public, telles que GenBank (Médigue, Bocs, Labarre, Mathé, & Vallenet, 2002).

3.1.1. Le Blast

BLAST (outil de recherche d'alignement local de base) est l'outil bioinformatique clé pour la comparaison et l'extraction de séquences à partir de bases de données. il est souvent la première étape dans l'utilisation d'informations basées sur les séquences pour concevoir des expériences et contextualiser les résultats expérimentaux (Kerfeld & Scott, 2011).

Les résultats BLAST ont une mesure associée de signification statistique appelée e-valeur. Lorsqu'un nouveau gène correspond à une séquence de base de données sans fonction assignée, il est défini comme une hypothèse conservée. Lorsqu'aucune correspondance significative n'est trouvée, le nouveau gène est qualifié d'hypothétique. Tous les génomes procaryotes séquencés à ce jour possèdent une fraction assez élevée (entre 20 et 40%) de gènes pour lesquels aucune fonction n'a été attribuée (Boratyn et al., 2013).

3.1.2. Line plot (analyse en blocs synthénique)

L'essence de la génomique comparative réside dans la façon dont nous comparons les génomes pour révéler les relations évolutives des espèces. Les progrès des technologies de

séquençage ont permis l'exploration de nouveaux génomes dans tous les domaines de la vie. Contenu de la complexité des génomes qui généralement des millions ou des milliards de nucléotides et diffère du génome d'une espèce étroitement apparentée, cela en raison de processus évolutifs tels que des mutations séquentielles, des réarrangements chromosomiques et l'expansion ou la perte d'une famille de gènes. Il peut être difficile d'aligner correctement deux génomes (McCouch, 2001)

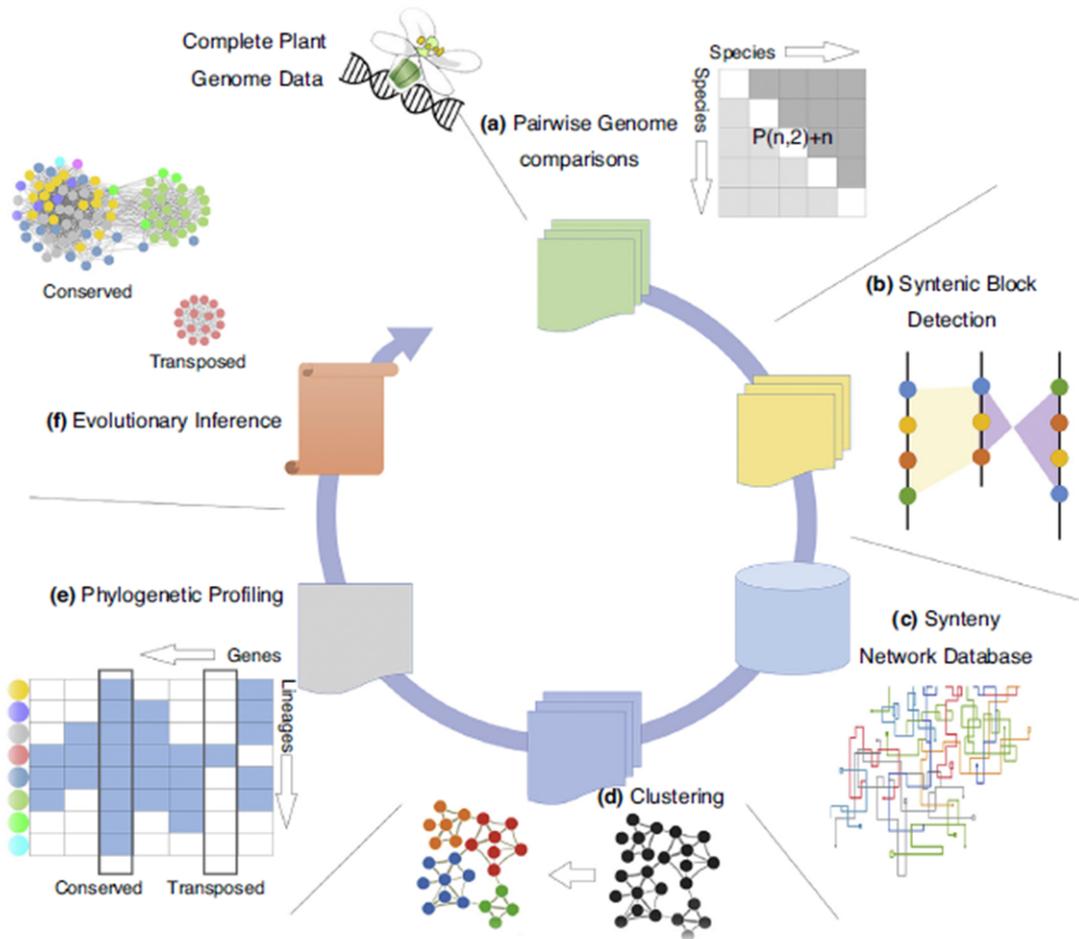


Figure 10 :La construction et l'analyse de réseaux de synténie comportent plusieurs étapes: (a) comparaisons du génome par paires, suivies de (b) détection de bloc synténique et l'intégration des données, afin de construire (c) une base de données de réseau de syntaxe permettant (d) le regroupement des données de réseau et (e) le profilage phylogénétique. Les étapes permettent (f) des inférences évolutives améliorées.(Zhao & Schranz, 2017).

Une approche alternative à l'alignement de génome complet consiste à l'identification de blocs de synténie (figure10). Décrits pour la première fois comme des loci génétiques homologues qui co-apparaissent sur le même chromosome Les blocs de synténie sont plus

formellement définis comme des régions de chromosomes entre génomes qui partagent un ordre commun de gènes homologues dérivés d'un ancêtre commun. Les comparaisons de la syntenie génomique entre et au sein des espèces ont fourni une opportunité d'étudier les processus évolutifs conduisant à la diversité du nombre et de la structure des chromosomes. (D. Liu, Hunt, & Tsai, 2018).

L'analyse basée sur la syntenie est largement reconnue comme une approche efficace et fiable pour la génomique comparative. Cependant, l'organisation et le traitement efficaces à grande échelle des connexions microsynthétiques restent limitées. Avec l'augmentation rapide des génomes entiers de plantes et une multitude de nouveaux algorithmes et d'outils pour l'inférence de réseau, le plan de construction de réseaux de syntenie permettra de mieux comprendre l'histoire évolutive des génomes ainsi que celle des familles de gènes (Zhao & Schranz, 2017).

4. La transcriptomique

La Transcriptomique est par définition, l'étude de toutes les molécules d'ARN produites par une cellule, bien que théoriquement, la transcriptomique concerne aussi bien les ARN messagers (ARNm) que les ARN ribosomiaux, de transfert et les petits ARN. La caractérisation du transcriptome (ensemble des transcrit) se focalise, en pratique, sur l'expression des gènes et donc des ARNm, (Ailloud, 2015)

Cette méthode permet d'identifier les gènes différentiellement régulés entre deux ou plusieurs conditions ainsi que de caractériser le profil d'expression d'un gène dans plusieurs conditions (Bidanel et al., 2008). La fonction d'un gène n'étant pas limitée uniquement à sa fonction biochimique, la caractérisation de son expression spatio-temporelle permet de se rapprocher du rôle réel du gène dans des processus physiologiques.

4.1. Les puces à ADN

Cette technique de transcriptomique permet d'évaluer la quantité de transcrit pour plusieurs milliers de gènes dans différentes conditions. Suite à l'extraction des ARN totaux, des ADN complémentaire sont produits à partir de la reverse transcription des ARNm. Les ADNc mesurés par des fluorochromes, sont hybridés sur une puce à ADN contenant des sondes (ou probset) qui consiste en des oligonucléotides complémentaires au ADNc et disposés de manière ordonnée sur la puce de telle sorte qu'un type d'ADNc s'hybride sur une position bien déterminée sur la puce (Gallopain, Gallopain, & Statistiques, 2017). Les sondes sont

produites en se basant sur les FSTs identifié au sein d'un génome, ceci est l'une des plus grande limite de cette approche car l'analyse du transcriptome sous-entend une connaissance de l'ensemble des gènes et des régions transcrits pour un organisme qui souvent ne sont pas complètement caractérisé. Après hybridation des ADNc sur la puce, le niveau d'hybridation est détecté par mesure de l'intensité de fluorescence qui reflète la quantité d'ADNc hybridé par sondes. L'expression mesurée correspond donc à un niveau de fluorescence qui sera évalué entre les différentes conditions (Figure 11) (Cooper GM,2000).

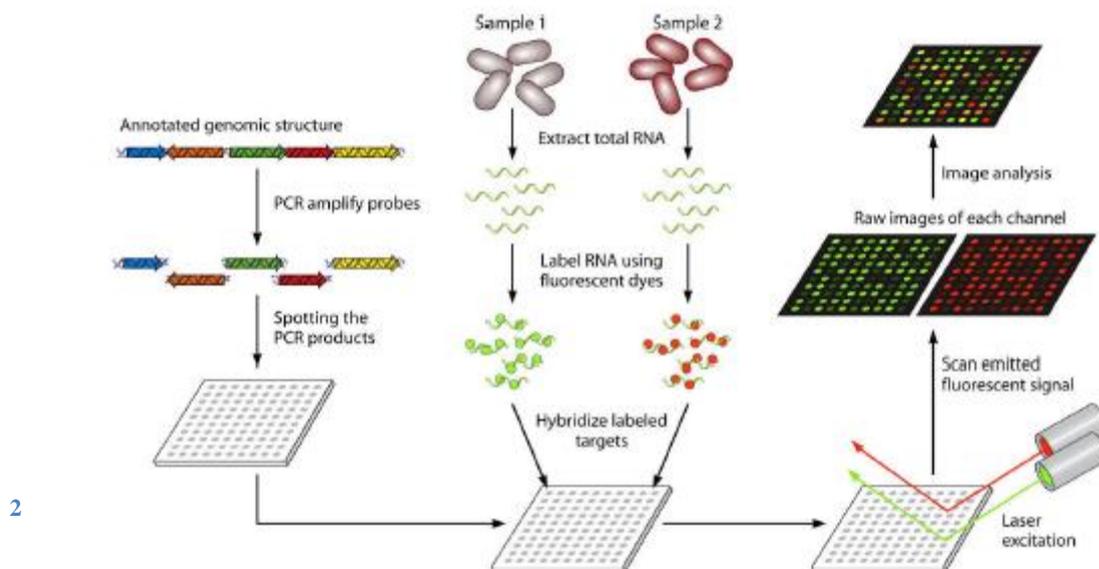


Figure 11. Principe d'une analyse transcriptomique par puce à ADN. Les ARNs sont extraits pour les différentes conditions, reverse transcrit pour produire des ADNc marqué par des fluorochromes de telle sorte que chaque condition possèdent sa propre fluorescence (exemple rouge ou vert). Après hybridation sur une puce contenant de manière ordonnée des oligonucléotides ou sonde correspondant aux FSTs, la fluorescence est évaluée à l'aide d'une excitation laser. La comparaison d'intensité de fluorescence entre les fluorochrome permet d'évaluer les variations d'expression relative. (Cooper GM .2000).

4.2. Séquençage et analyse d'ARN (RNA seq technology)

Le séquençage d'ARN (RNA-Seq) utilise les capacités des méthodes de séquençage à haut débit pour fournir des informations sur le transcriptome d'une cellule. Ainsi après l'extraction des ARNs, les ADNc sont séquencés et les séquences obtenues appelées « reads » sont confrontée à celle des gènes composant le génome. L'évaluation de la diversité et de la quantité de reads être plusieurs conditions permet d'identifié les régions exprimées de manière différentielle (Figure 12) (Kukurba & Montgomery,2005).

Par rapport à la puce à ADN, le RNA-Seq offre une couverture bien supérieure et une résolution plus grande de la nature dynamique du transcriptome. Au-delà de la quantification de l'expression des gènes, les données générées par RNA-Seq facilitent la découverte de nouveaux transcrits, l'identification de gènes épissés alternativement et la détection de l'expression spécifique d'allèles (Kukurba & Montgomery, 2005).

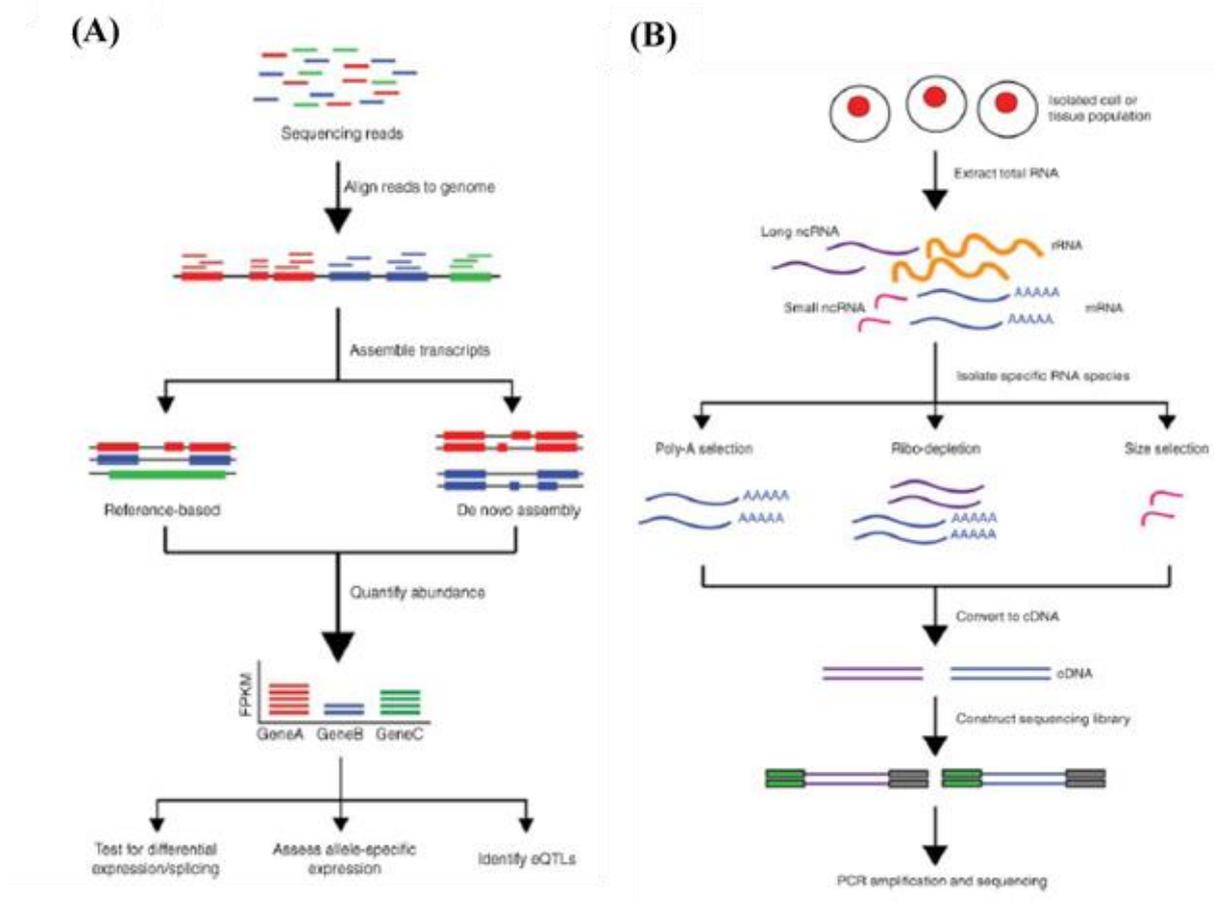


Figure 12 :Vue d'ensemble de l'ARN-Seq. A/ l'ARN est extrait du matériel biologique puis des sous-ensembles de molécules d'ARN sont isolés à l'aide d'un protocole spécifique puis converti en ADN complémentaire (ADNc) par transcription inverse et les adaptateurs de séquençage sont ligaturés aux extrémités des fragments d'ADNc. B/Analyses des données ARN-seq, les lectures peuvent être assemblées en transcriptions à l'aide d'annotations de transcriptions de référence ou d'approches d'assemblage de novo. (Kukurba & Montgomery, 2015)

Chapitre IV. Introduction à la problématique étudiée

1. Contexte générale de l'étude

1.1. La symbiose et l'immunité d'un point de vue procaryotique :

La symbiose a souvent été considérée comme un processus radicalement différents de la pathogénicité. De nombreuses études montrent que ces deux phénomènes font intervenir des mécanismes moléculaires similaires, ainsi la perception des facteurs nod des rhizobia implique des récepteurs de l'immunité et inversement, des récepteurs symbiotique sont requis pour la réponse à des pathogènes (N, Geiger, & Giorgio, 2016)

Actuellement, la majeure partie des études se base sur une approches similaire qui consiste à se focaliser sur une espèce donnée (bactéries ou hôtes) et à travers l'utilisation de différents outils, de montrer l'implication de gènes dans les processus de symbiose et/ou d'immunité. Avec l'émergence de la génomique une autre méthode de réflexions est développée notamment par les évolutionnistes et consiste à comparer les génomes de plusieurs organismes afin d'identifier les points commun et divergent.

Lors de notre projet, nous avons cherché à comprendre quels mécanismes peuvent être divergents ou conservés entre trois souches de rhizobia (*Sinorhizobium meliloti*, *Bradirhizobium japonicum* et *Cupriavidus Taiwanensis*), ainsi que trois souches de pathogènes (*Agrobacterium tumefaciens*, *Ralstonia solanacearum* et *Pseudomonas syringae*). L'analyse s'est focalisée sur la comparaison des grandes voies métaboliques entre les différentes souches, l'analyse de la synthénie ainsi que l'analyse de l'expression de gènes par RNA seq. L'objectif principale été de caractérisé le mode de vie symbiotique et pathogène ainsi que d'identifier des gènes pouvant intervenir au cours de la symbiose chez les rhizobia et du processus d'interaction avec l'hôte chez les pathogènes.

1.2. L'adaptation des endosymbiotes à la vie intracellulaire :

Le milieu intracellulaire des eucaryotes peut constituer un environnement hostile aux bactéries, qu'elle soit pathogènes ou symbiotiques. Les NCRs et les défensines constituent un bon exemple de réactions néfastes produites par l'hôte à l'encontre des endocellulaires.

Les bactéries doivent ainsi mettre en place des mécanismes d'adaptation à l'hôte afin de bloquer ou de contourner la réaction de l'hôte. Par exemple les rhizobia, au même titre que les pathogènes endocellulaires telle que *Brucella abortus*, utilisent des systèmes de sécrétion et des effecteurs afin de bloquer la signalisation immunitaire (Krachler & Orth, 2013).

Dans le but de comprendre les mécanismes d'adaptation bactéries au milieu endocellulaire. Nous avons comparé des modèles bactériens évoluant dans des milieux extrêmement

différents, mais présentant des points communs afin d'avoir une phase endocellulaire. Cette approche nouvelle permet théoriquement de se concentrer sur les gènes pouvant être réellement impliqués dans la survie endocellulaire et d'éviter les gènes impliqués dans l'adaptation à l'environnement. Nous avons ainsi comparé au niveau génomique et transcriptomique l'endosymbionte *S. meliloti* et le pathogène animale *B. abortus*.

2. Questions et problèmes abordés

Au cours de notre projet nous avons cherché à comprendre les divergences ainsi que les convergences entre le mode de vie pathogènes et symbiotique en s'appuyant sur la comparaison génomique et transcriptomique.

De plus nous avons testé une nouvelle méthode de prospection basée sur la comparaison génomique et transcriptomique entre des bactéries très différentes afin d'identifier des gènes potentiellement requis pour les processus de survie intracellulaire.

1. Sélection des souches

Nous nous sommes intéressés à trois souches dont les génomes sont complètement séquencés, elles ont été sélectionnées sur la base de leurs virulences, *Ralstonia Solanacearum* GMI1000, *Pseudomonas syringae* pv *syringae* B728a et *Agrobacterium tumefaciens* Ach5. En plus du pathogène intracellulaire *Brucella abortus* 2308 qui a été introduit dans nos analyses afin de comparer la conservation et la divergence de mode de vie intracellulaire avec la souche endosymbiotique *Sinorhizobium meliloti* 1021. Cette dernière avec *Bradyrhizobium japonicum* USDA6 et *Cupriavidus taiwanensis* LMG19424 correspondant à des symbiotes des légumineuses ont été utilisées comme modèle symbiotique lors de l'étude.

Les séquences des génomes sont obtenues à partir de la base de données (genome database) intégré au serveur NCBI (<https://www.ncbi.nlm.nih.gov/genome/genomes/>).

Figure 3

Figure 13. Interface de la base de données Genome dans l'NCBI.

Lors du téléchargement chaque fichier est complété par l'annotation lui correspondant (fichier Excel contenant un tableau des gènes) qui sert à affecter chaque fonction à un gène et montrer les différentes séquences composants chaque génome, les gènes, les pseudogènes, les séquences codant des protéines et d'autres unités génomiques fonctionnelles telles que les ARNt, les ARNr le nombre de gènes plasmidique.

2. Annotation RAST (Rapid annotation subsystem technology) :

RAST est un serveur (<http://rast.nmpdr.org/seedviewer.cgi>) entièrement automatisé servant à annoter les génomes bactériens et d'archèes. Le service identifie les gènes codant les protéines, les ARNr et les ARNt, attribue des fonctions aux gènes, prédit quels sous-systèmes sont représentés dans le génome. Il utilise ces informations pour reconstruire le réseau métabolique et permet de télécharger le résultat facilement pour une visualisation claire. De plus, le génome annoté peut être parcouru dans un environnement prenant en charge l'analyse comparative avec les génomes annotés.

Les outils d'annotation intégrée au serveur RAST ont été utilisés car ils sont conçus pour produire de manière cohérente des annotations comparables et une vision simple des résultats. Les étapes de base utilisées pour annoter un génome à l'aide de RAST sont décrites dans la figure ci-dessous. Le processus consiste en un génome procaryote sous la forme d'un ensemble de contigs au format FASTA, le serveur RAST permet à un utilisateur de spécifier un ensemble d'appels de gènes, mais dans le cas habituel, RAST effectuera ses propres appels. Les étapes suivantes décrivent en détail le reste de la méthode.

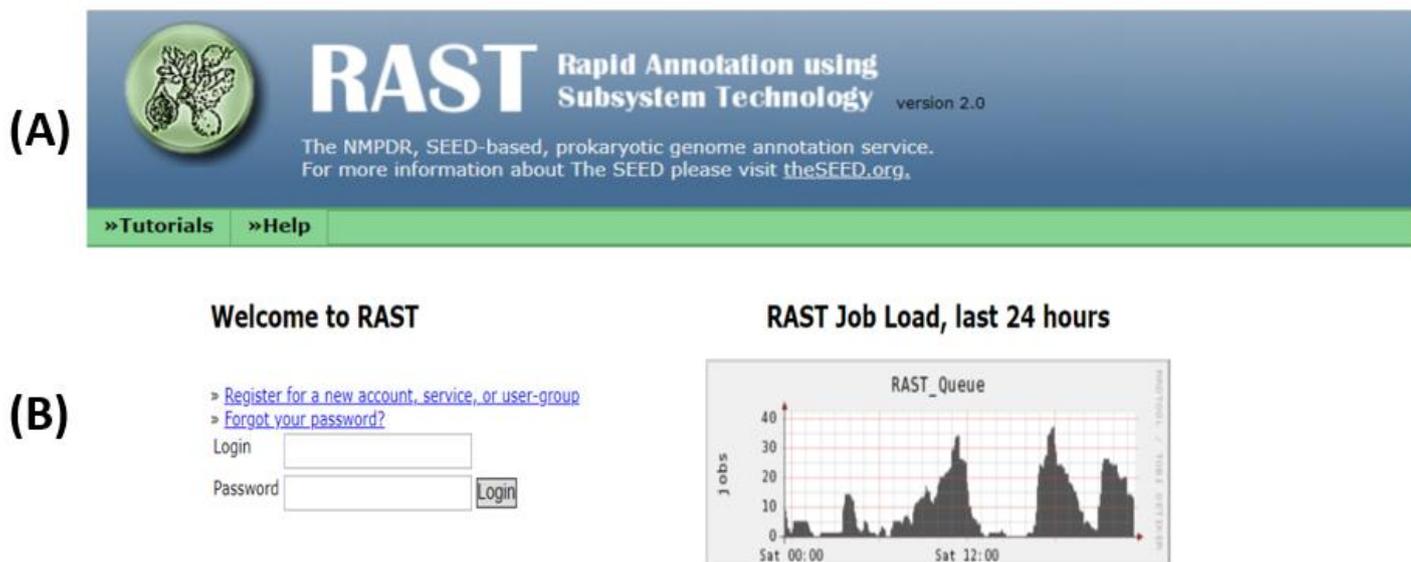


Figure 14. Les étapes d'obtention de l'annotation par RAST server, (A) Interface de serveur (B) insertion du fichier FASTA du génome.

Une fois que les fonctions ont été assignées, une reconstruction métabolique initiale est formée. Dans ce cas, cela revient à connecter des gènes à des rôles fonctionnels dans des sous-systèmes, ce qui permet de déterminer quand un ensemble de connexions à un sous-système spécifique est suffisant pour prendre en charge une variante active (Aziz et al., 2008).

Il convient de mentionner que les sous-systèmes couvrent tous les modules de la machinerie cellulaire et non pas seulement les voies métaboliques, ce qui permet de se focaliser sur l'étude ou la recherche effectuée sur les gènes intervenant lors de l'activation ou la régulation d'une voie caractéristique.

Par ailleurs le taux de recouvrement des génomes par l'annotation peut varier d'une souche à une autre, ce biais est à prendre en considération lors de l'interprétation des résultats.

3. Le serveur Microscope (Microbial Genome Annotation & Analysis Platform)

3.1. Comparaison génomique

Une grande variété de logiciels sont à la disposition de la communauté scientifique et peuvent être utilisés pour identifier des objets génomiques avant de prédire leurs fonctions biologiques. Cependant, seul un nombre limité de caractéristiques biologiquement intéressantes peut être révélé à partir d'une séquence isolée.

La plate-forme MicroScope (Figure 15) a été développée afin de proposer un cadre basé sur le Web pour la révision systématique et efficace de l'annotation et de l'analyse comparative du génome microbien.

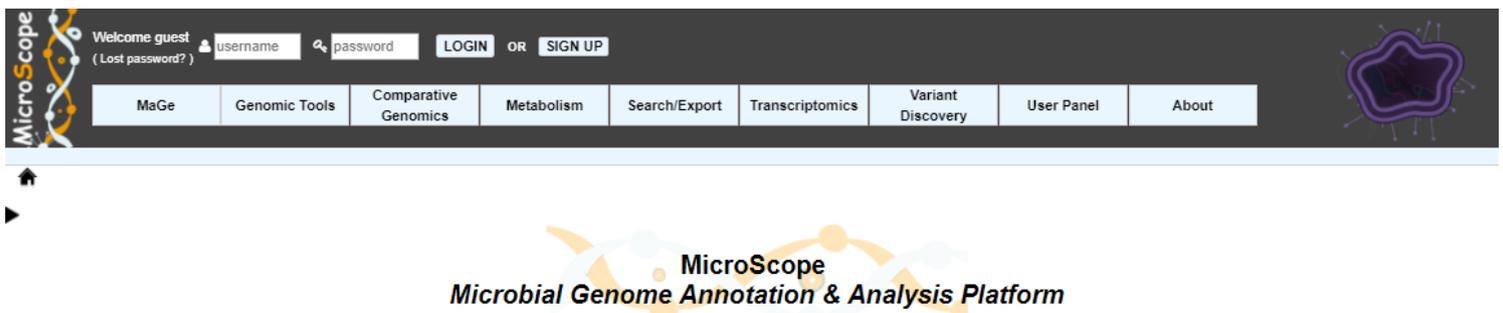


Figure 15. Interface microscope server

Une comparaison des séquences génomiques entre le groupe des rhizobias ou le groupe de pathogènes a été réalisée en utilisant l'outil en ligne disponible sur la base de données MicroScope (<http://www.genoscope.cns.fr/agc/microscope/>).

Il permet de calculer les tailles de pan et core génome, leurs évolutions pour un ensemble de génomes, tout en déterminant l'ensemble de variantes en commun entre eux.

Pour cela ;les souches sont sélectionnées pour avoir un diagramme de Venn rassemblant le nombre de gènes en commun(Figure16) , puis extraire les séquences et les annotations du core génome.

3. Browse results

Analysis summary

- Analyzed genomes: 2
- Excluded genome(s): 0
- MICFAM parameter:
 - 80% aa identity
 - 80% alignment coverage

Main results

Component	Families	Genes
Pan-genome	12074	12587
Core-genome	41	98
Variable-genome	12033	12489

Download and export

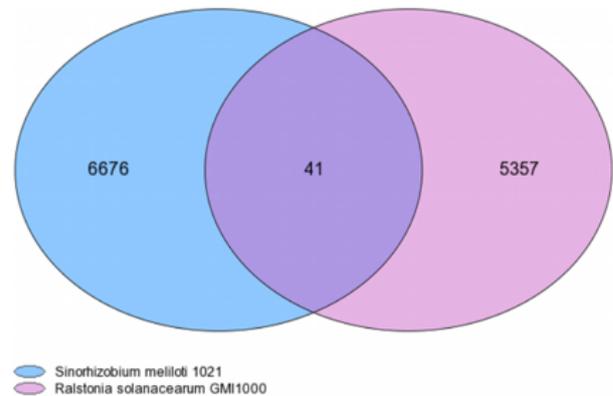
Organism Filter

	Fasta		Annotation	Export
Pan-genome	nuc	prot	csv	Gene Cart
Core-genome	nuc	prot	csv	Gene Cart
Variable-genome	nuc	prot	csv	Gene Cart
Strain-specific-genome	nuc	prot	csv	Gene Cart

(A)

Venn Diagram (family number)

Switch to Selected Genomes view



(B)

Figure 4

Figure 16. Résultat de comparaison des souches, (A) Nombre de gènes en commun et différents entre les souches étudiées. (B) liste des gènes en communs et souche spécifique format excel. Figure 5 resultats de comparaison des souches

Le résultat présente le nombre de familles de gènes communs ou divergents obtenus, une famille peut contenir un ou plusieurs gènes homologues entre les différentes souches.

De plus, nous avons tenté d'exclure du résultat final les gènes dits de ménage, en incluant dans la base de données le génome complet d'une bactérie de référence non-associée à des plantes. L'élimination de ces gènes nous permet de minimiser le nombre de gènes en commun.

3.2. Analyse de syntonie

Le deuxième outil utilisé lors de la comparaison génomique est le « Line plot », qui illustre un résultat sous-forme de représentation graphique globale de syntagmes conservés entre deux génomes sélectionnés.

En premier lieu l'analyse de la syntonie se base sur la sélection d'un organisme référence, qui se fait en haut de l'interface du server Microscope (figure 15), dont le but est de vérifier la validité des résultats des outils précédents et d'analyser la conservation et la divergence des régions génomiques entre les souches étudiées.

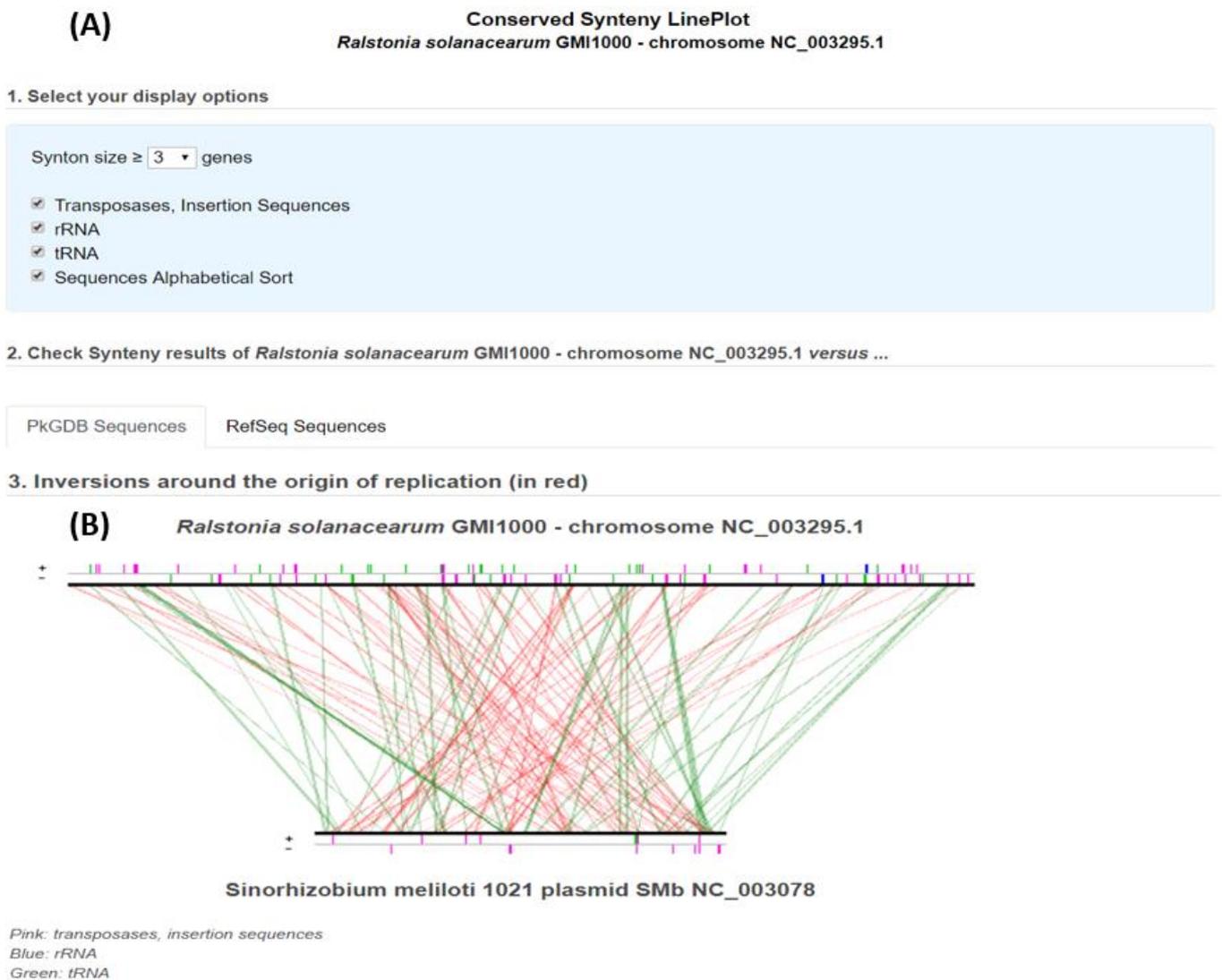


Figure 17. Evaluation de la syntonie, (A) comparaison entre une bactérie pathogène *R. solanacearum* et un symbionte *S. meliloti*. (B) résultat de la syntonie.

L'analyse de la conservation des régions peut même être établie entre chromosome/plasmide ce qui pourrait détecter de nouvelle région intégré dans le plasmide d'une souche tout en étant une région hautement conservé dans le chromosome de celle comparé.

4. La base de données Omnibus (GEO)

La base de donnée GEO (Gene Expression Omnibus)

(<https://www.ncbi.nlm.nih.gov/gds>) est un référentiel public international (figure 18) qui archive et distribue librement l'expression de gènes à haut débit et d'autres ensembles de données de génomique fonctionnelle (B. H. Liu, 2018).

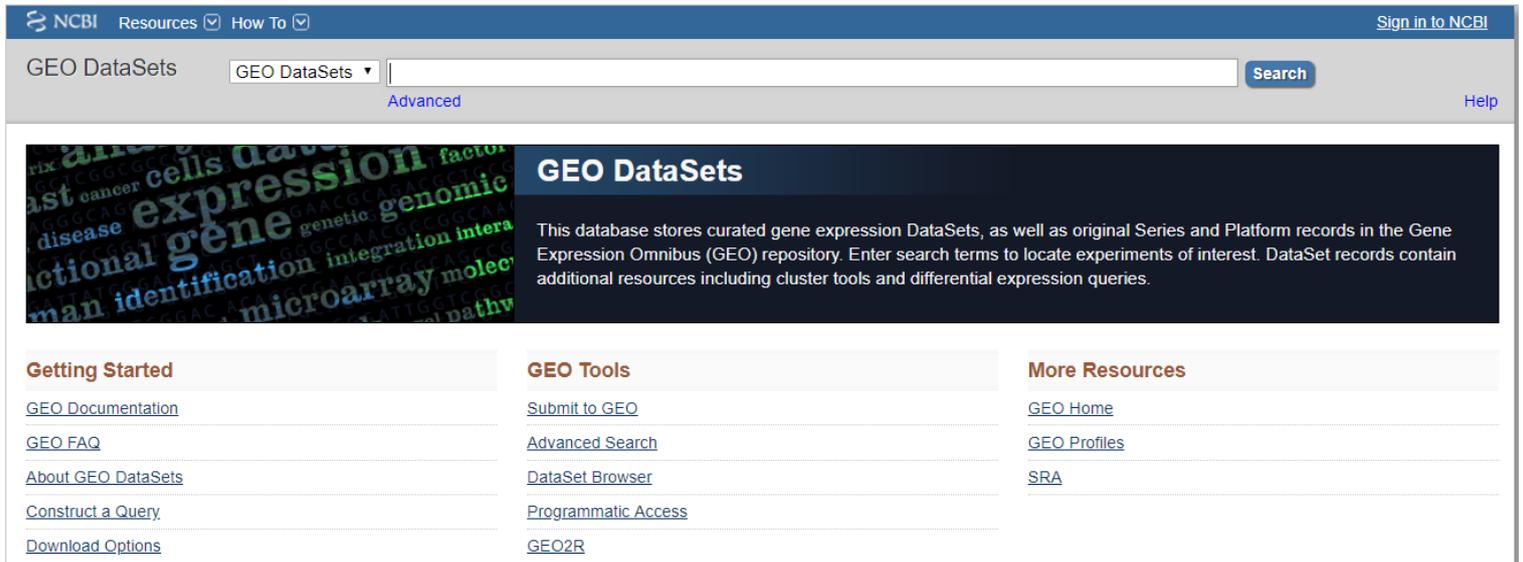


Figure 18. Interface de la base de données Geo datasets.

Créée en 2000 en tant que ressource mondiale pour les études d'expression génique, GEO a évolué avec les technologies en évolution rapide et accepte désormais les données à haut débit pour de nombreuses autres applications de données, y compris celles qui traitent de la méthylation du génome, de la structure de la chromatine et des interactions génome-protéine. GEO prend en charge les normes de dérivées de la communauté qui spécifient la fourniture de plusieurs éléments d'études critiques, notamment les données brutes, les données traitées et les métadonnées descriptives (B. H. Liu, 2018).

Lors de notre étude nous avons utilisé cette base de données afin de télécharger des transcriptomes permettant après traitement statistique de visualiser les profils d'expression adaptée à l'objectif de la thématique.

5. Analyse transcriptomique

Afin d'aller plus loin et d'identifier des mécanismes moléculaires conservés entre les pathogènes et les rhizobia au cours de l'interaction avec leurs hôtes nous avons analysé les profils d'expression des gènes orthologues de pathogènes chez *S. meliloti* 1021 au cours de la symbiose.

Le Profile d'expression est déterminé chez *S.meliloti* 1021 après avoir intégré les gènes en commun entre les deux souches *S. meliloti* 1021 et *A. tumefaciens* ou *S. meliloti* 1021 et *B. abortus* .

L'expression est visualisée sur l'ensemble des zones de la nodosité chez *Medicago truncatula* (ZI, ZII, ZIII). Les orthologues sont réparties en fonction de deux à quatre classes fonctionnelles et leurs expression est comparée à l'ensemble des gènes d'une même classe (orthologues ou non). Les voix biochimiques étudiées sont : la formation et le fonctionnement du flagelle, l'oxydo-réduction, le chimiotaxie et le métabolisme aminé.

6. Phylogénie moléculaire

L'analyse phylogénétique est établie en commençant par une comparaison des orthologues par l'outil BLAST, les séquences résultantes sont téléchargées dans un fichier Fasta multiple pour ensuite l'intégrer dans le logiciel MEGA 7 (Molecular Evolutionary Genetics Analysis) figure 19.

Par la suite le logiciel établie la construction en se basant sur une matrice de distance dite « p-distance » par la méthode neighbour-joining Bootstrap 1000.

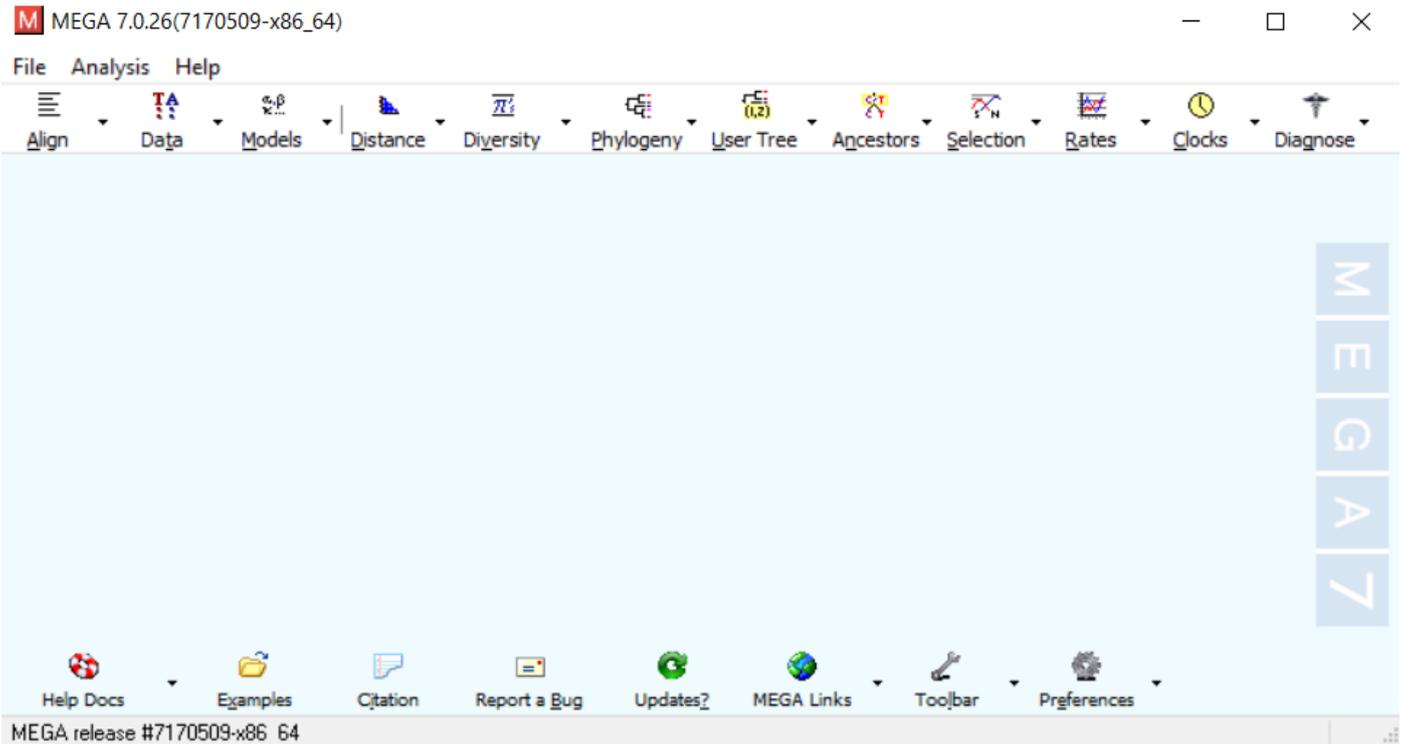


Figure 19. Interface du logiciel MEGA7 (Molecular Evolutionary Genetics Analysis).

7. Prédiction de la localisation

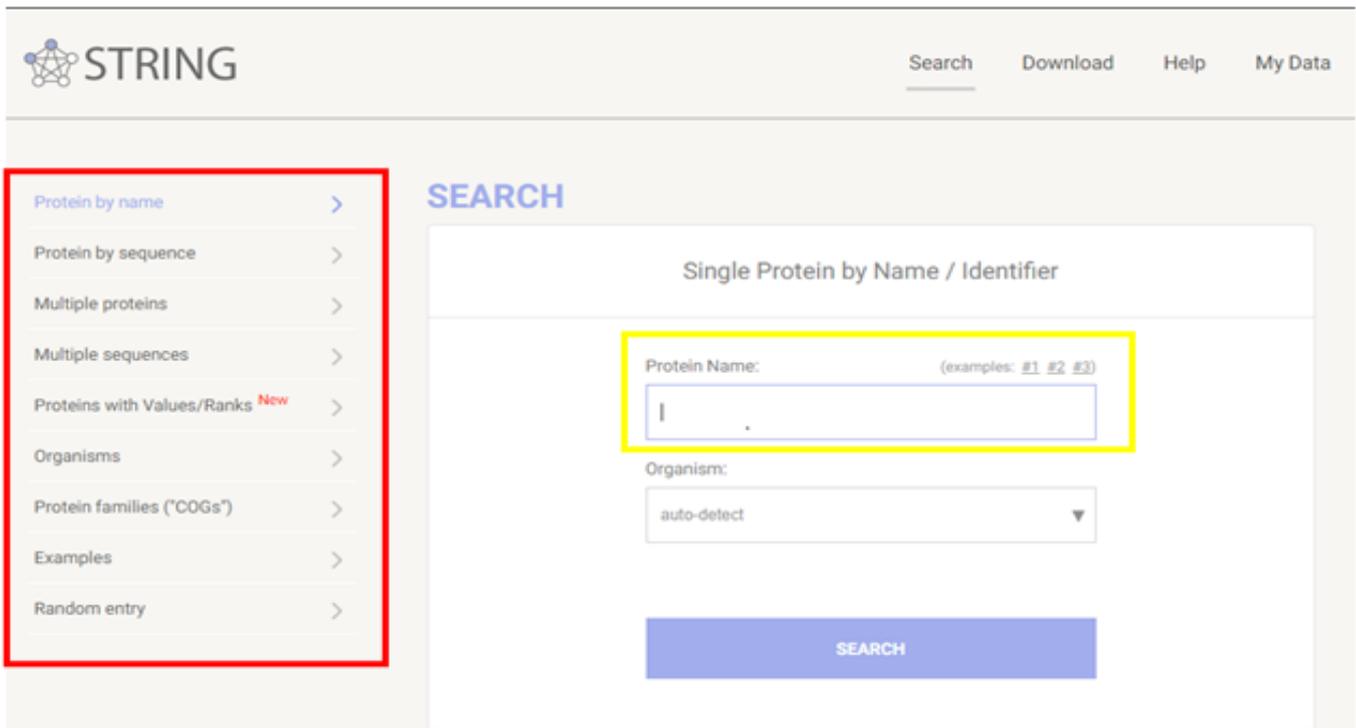
La première étape consiste à détecter la présence du peptide signal, pour cette prédiction il existe plusieurs outils, parmi eux le serveur SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP-4.1/>) qui prédit la présence et l'emplacement des sites de clivage du peptide signal dans les séquences d'acides.

Pour confirmer l'adressage et la localisation de la protéine, la deuxième étape de détection des domaines transmembranaire en utilisant DAS « Transmembrane Prediction server » <https://tmdas.bioinfo.se/DAS/index.html> est réalisée.

6.1. Prédiction des interactions protéine-protéine

Afin d'identifier les protéines interagissant avec la protéine candidate, une recherche sur la base de données STRING répertoriant les interactions protéine-protéine connues est réalisée. Les associations dans STRING incluent des interactions directes (physiques), ainsi que des interactions indirectes (fonctionnelles), pour autant que les deux soient spécifiques et biologiquement significatives.

Le réseau d'interactants résultant (Figure20) permet de connaître l'implication de la protéine et son rôle dans chaque fonction de la cellule et les différentes interactions qui existent au sein d'un réseau.



The image shows the STRING database search interface. On the left, a sidebar menu is highlighted with a red border, listing search options: Protein by name, Protein by sequence, Multiple proteins, Multiple sequences, Proteins with Values/Ranks (marked as 'New'), Organisms, Protein families ("COGs"), Examples, and Random entry. The main search area is titled "SEARCH" and "Single Protein by Name / Identifier". It contains a "Protein Name:" input field with a yellow border, a "Protein Name:" label, and a "(examples: #1 #2 #3)" note. Below it is an "Organism:" dropdown menu currently set to "auto-detect". A blue "SEARCH" button is located at the bottom of the search area.

Figure 20. Interface de la base de données SRTING, montrant le champ de l'insertion du nom de la protéine à étudier et l'organisme dans lequel elle est exprimée

Notre projet vise à étudier les différences entre des souches bactériennes pathogènes, symbiotiques, extracellulaires ou intracellulaires afin d'identifier les points communs et divergents entre ces différentes bactéries et de comprendre l'évolution des différentes associations ainsi que de dégager des gènes candidats afin de réaliser des analyses fonctionnelles.

Pour cela plusieurs approches ont été utilisées, la première, génomique a pour but de mettre en évidence les différences au niveau des gènes ainsi que de l'organisation des génomes. La seconde, transcriptomique vise à étudier ces différences au niveau de l'expression génique.

1. Description génomique des souches employées

Les souches pathogènes étudiées ont été sélectionnées sur la base de leurs virulences, ainsi *Agrobacterium tumefaciens* Ach5, *Ralstonia solanacearum* GMI1000 et *Pseudomonas syringae* pv *synringae* B728a font partie des pathogènes les plus virulents ainsi que les plus étudiés au sein de la communauté scientifique mondiale (Bidanel et al., 2008). Le pathogène intracellulaire *Brucella abortus* 2308 a été introduit dans nos analyses afin de comparer la conservation et la divergence des voies métaboliques entre un pathogène animale, des pathogènes végétaux et des endosymbiontes. Les bactéries symbiotiques sélectionnées sont des rhizobia dont deux α -protéobactéries : *Sinorhizobium meliloti* 1021, *Bradyrhizobium japonicum* USDA6 et une β -protéobactérie : *Cupriavidus taiwanensis* LMG19424.

L'analyse comparative présentée dans ce travail repose sur des génomes publiés et disponibles dans des bases de données accessibles ; Genome de l'NCBI <https://www.ncbi.nlm.nih.gov/genome/> et l'EMBL-EBI <https://www.ebi.ac.uk/>. Le tableau 2 présente les principales caractéristiques génomiques des différentes souches utilisées. Ces caractéristiques sont obtenues à partir de la base de données Genome dans l'NCBI.

Résultats et discussion

Tableau 2. Les souches utilisées lors de l'étude et leurs principales caractéristiques (K : chromosome, P : plasmide, Mb : longueur en Méga bases, Nbre : Nombre)

	<i>A. tumefasceins</i>	<i>R. solanacearum</i>	<i>P. syringae</i>	<i>S. meliloti</i>	<i>B. japonicum</i>	<i>C. taiwanansis</i>	<i>B. abortus</i>
Structure	2K 2P	1K 1P	1K	1K 2P	1K	2K 1P	2K
Mb	5.66866	5.8109	6.037	6.69169	9.20738	6.47652	3.28
GC %	58.47	66.96	59.2	62.15	63.7	66.9	57.2
Nbre de gènes	5404	5208	5220	6292	8743	5918	3360
Nbre de Protéines	5188	4984	5089	6218	8327	5719	3153
ARNr	15	12	15	9	6	15	9
ARNt	56	57	64	54	55	64	55

Certaines souches possèdent un chromosome (exemple *P. syringae*) ou deux (exemple : *C. taiwanensis*, *R. solanacearum*) et des plasmides reçus lors d'échanges horizontaux. La longueur donnée en Mb du génome montre que les souches symbiotiques possèdent un plus grand génome que celui des pathogènes, ce qui est corrélé avec le nombre de gènes et de protéines. Le nombre de protéines n'est pas systématiquement corrélé avec celui des gènes pour une même souche, ceci peut être expliqué par la présence de gènes non fonctionnels (pseudogènes) dans le génome. L'apparition des pseudogènes peut faire suite à une mutation ou une erreur affectant un gène fonctionnel ou pouvant être lié à des remaniements génomiques aboutissant à la modification d'ORF (Open Reading Frame).

2. Résultats des annotations des génomes par RAST

L'analyse de la composition en gène d'un génome permet de mettre en évidence les fonctions biochimiques d'un organisme. Dans le but de vérifier la conservation et la divergence des voies métaboliques entre les différentes souches, on a utilisé le serveur RAST (Rapid annotation using subsystem technology) afin d'obtenir une annotation rapide des génomes. Ceci nous permet de confirmer l'annotation des génomes en plus d'avoir une répartition des gènes en fonction de grande voie biochimique ou COG (Cluster of Orthologous Genes).

La figure 21 présente les pourcentages de recouvrement pour chaque génome avec l'outil RAST.

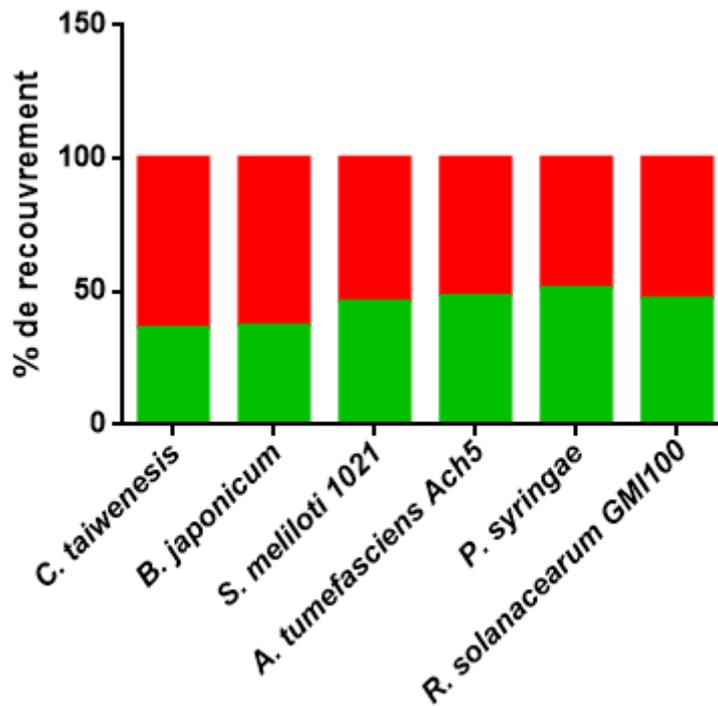


Figure 21. Recouvrement de génomes annotés par l’outil RAST. En vert pourcentage du génome annoté (couvert), En rouge pourcentage de régions génomiques non annotées (non couvertes).Figure 7 recouvrement de génomes annotés par l’outil RAST

En vert le pourcentage des régions génomiques pour lesquelles une annotation des gènes a pu être réalisé par l’outil RAST. Globalement près de 50 % des génomes pu être annotés par l’outil RAST. Pour le reste, le logiciel n’a pas pu déterminer la composition des régions restantes. Afin d’annoter l’ensemble du génome, d’autres logiciels peuvent être utilisés, mais ces derniers sont plus simple que RAST et ils nécessitent des ordinateurs d’une puissance supérieure à ceux dont nous disposons au laboratoire avec un débit de connexion plus important. Néanmoins les résultats obtenus avec l’annotation RAST mettent en évidence des différences intéressantes et exploitables pour la suite de l’étude.

3. Peu de différence en quantité de gènes entre les souches

L’analyse en composition de gènes des différentes souches est représentée dans la figure 22. Les gènes sont triés selon leur fonction biochimique en sous-système, ainsi chaque sous-système représente une voie métabolique contenant le nombre en pourcentage de gènes intervenants dans chaque voie/fonction.

Résultats et discussion

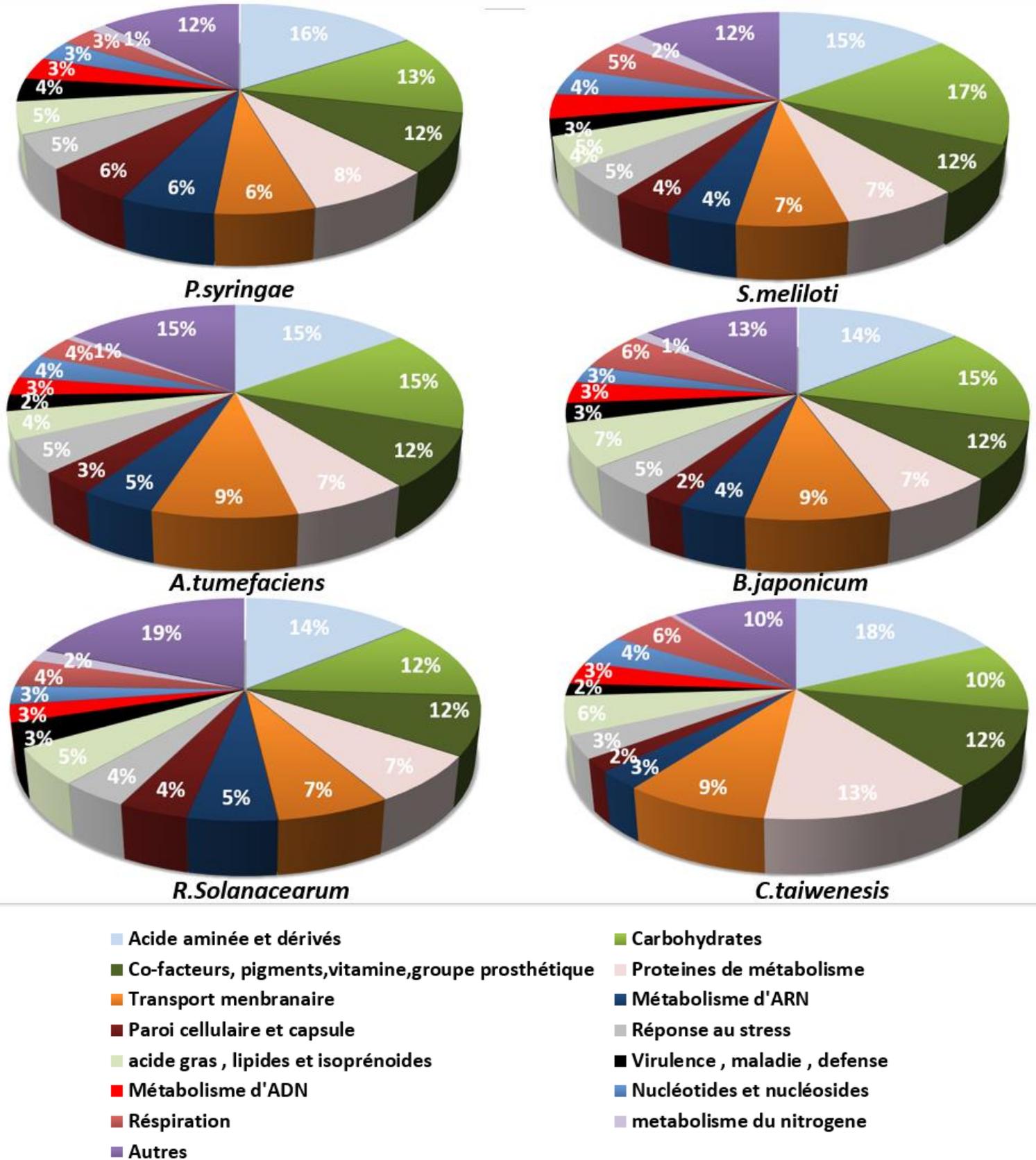


Figure 22 : Représentation du pourcentage des gènes plasmidiques de chaque voie métabolique pour les différentes souches étudiées.

Les résultats obtenus montrent qu'il n'y a pas de différences drastiques d'un point de vue nombre de gènes par voie métabolique. Néanmoins certaines différences peuvent être relevées. Le symbionte *C. taiwanensis* présente les plus faibles nombres de gènes pour le métabolisme lié à la paroi cellulaire (2%), la réponse au stress (3%), le transport (3%). Ceci est compensé par les gènes codant des protéines du métabolisme (13%) qui sont les plus nombreux chez le symbionte, alors qu'il est en moyenne de 7 à 8%. La souche *S. meliloti* présente le plus grand nombre de gènes liés au métabolisme des carbohydrates (17%).

La présence d'un nombre de gènes équivalent entre les différentes souches par rapport à la virulence, indique que cette fonction biochimique est importante pour les souches symbiotiques autant que pathogènes. Le métabolisme de l'azote est important pour le processus de pathogénicité chez certaines bactéries telle que *R. solanacearum* (Dalsing, Truchon, Gonzalez-Orta, Milling, & Allen, 2015). Ce nombre reste stable entre les différentes souches avec un pourcentage de 3 à 5 %. Ce qui indique qu'au niveau génomique un nombre équivalent de gènes liés à l'azote est retrouvé entre les souches pathogènes et non pathogènes.

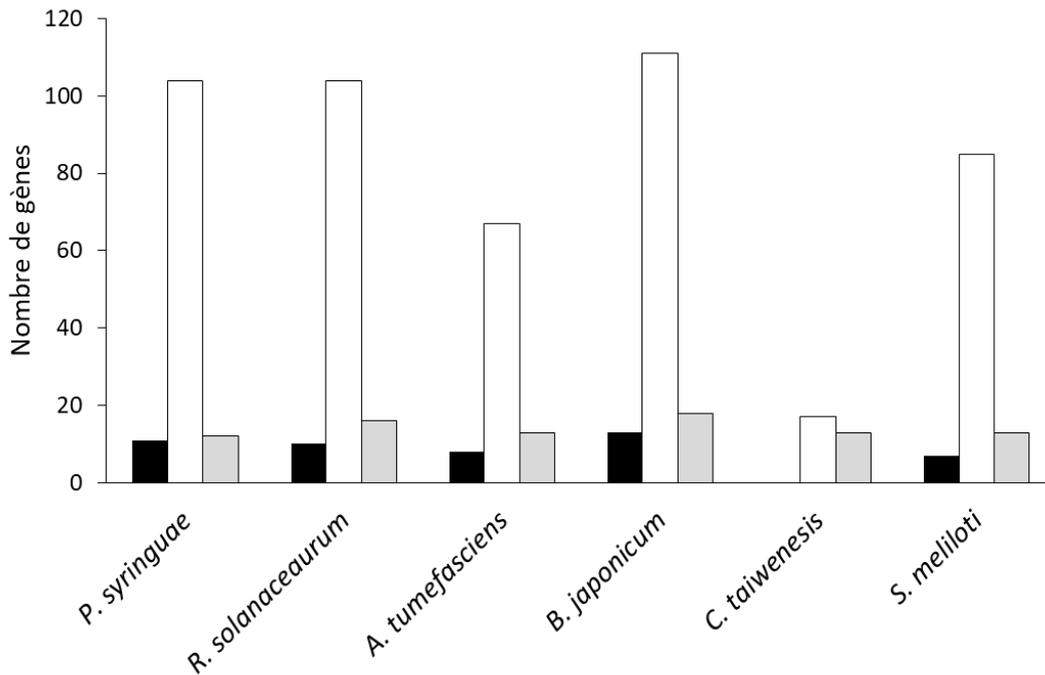
Le peu de différence au niveau génomique peut être expliqué par le fait que la plus part des gènes requis pour les processus d'interaction avec l'hôte sont situés sur le matériel extra chromosomique (plasmide). Nos résultats tant à indiquer que l'acquisition d'un plasmide peut être à la base d'un changement du style de vie (pathogène/symbionte). Ces résultats sont appuyés par les travaux de l'équipe de C. Maison-Boivin qui montre que l'ajout du plasmide symbiotique de *C. taiwanensis* à *R. solanacearum* suffit à la création d'une souche chimères nodulants *Mimosa pudica* l'hôte de *C. taiwanensis* (Gilbert et al., 2010).

4. Les souches présentent des différences qualitatives en composition génique

Afin d'aller plus loin dans nos analyses, nous nous sommes focalisés sur le groupe de gènes de virulence. Le graphe 1. Présente le nombre de gènes pour les différents types de gènes de virulence entre les souches étudiées.

Résultats et discussion

On observe que les gènes associés à la résistance aux antibiotiques et aux composés toxiques sont variables d'une souche à une autre, chez *A. tumefaciens* leur nombre est le plus faible avec 67 gènes, alors que pour *B. japonicum* il est de 111 gènes. Ces résultats contrastent avec ceux de la figure 2 et indique que malgré peu de variation au niveau de la quantité totale de gènes de virulence (2 à 4 %), il existe des différences qualitatives avec une répartition des gènes différentielle entre les catégories de gènes de virulence.



Graphe 1. Les gènes de virulences chez les pathogènes et symbiontes. Le graphe représente le nombre total de gènes pour trois catégories de gènes de virulence : bactériocines/peptides antimicrobiens (en noir). Résistance aux antibiotiques et composés toxiques (blanc). Invasion et résistance intracellulaire (gris).

De manière intéressante peu de variation existe pour les gènes d'invasion et de résistance intracellulaire. Ce résultat est surprenant étant données que les rhizobia étudiées sont des endosymbiontes et peuvent se retrouver au sein d'une cellule eucaryote lors de la symbiose. Cette observation est potentiellement expliquée par le fait que l'ensemble des bactéries étudiées présentent deux modes de vie : l'un lié à l'interaction avec leurs hôtes et l'autre épiphyte dans le sol. Il est possible donc que les gènes faiblement variables soient impliqués préférentiellement dans le deuxième mode de vie.

5. Les plasmides à la base du style de vie

S. meliloti et *A. tumefaciens* présentent deux plasmides alors que *C. taiwanensis* et *R. solanacearum* présentent un seul (Figure 23). L'analyse de la répartition des gènes en fonction de leur rôle montre une hétérogénéité plus importante que celle observé au niveau des génomes. Ainsi les pathogènes possèdent le nombre le plus important de gènes associés au métabolisme de l'ARN, alors que les symbiontes présentent des gènes impliqués dans le métabolisme des parois et de la capsid (Figure 23). *S. meliloti* présente des gènes associés à la réponse au stress au niveau du plasmide pSymB, *C. taiwanensis* est dépourvue au niveau du plasmide de gènes du métabolisme du carbone et de l'azote, à l'inverse chez cette dernière catégorie *S. meliloti* présente le plus grand nombre de gènes (Figure 23).

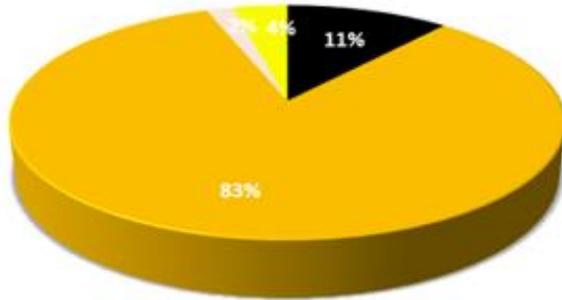
Chez le symbionte *S. meliloti* les gènes du métabolisme d'azote se situent au sein des 2 plasmides (pSymA et pSymB) avec une plus grande représentation sur pSymA. *S. meliloti* présente 9 gènes associés à la fixation d'azote (*nifA*, *nifB*, *nifH/D/K/E*, *nifX*, *nifN*, and *nifS*) qui sont tous portés sur pSymA, par ailleurs deux orthologue de *nifS* et *nifV* se retrouve sur le plasmide pSymB.

De manière surprenante les gènes de nodulation requis pour la synthèse et l'exportation des facteurs Nod sont localisés également sur pSymA avec des clusters nod dupliqués dans le génome. La présence d'une séquence d'insertion près du cluster *nodPQ* dans pSymA suggère que cette copie est acquise par transposition à partir de la copie ancestrale située sur le plasmide pSymB. Ces résultats indiquent que les gènes nod ont 2 origines distinctes : un transfert horizontal du plasmide pSymA à partir d'une souche inconnue et une duplication au sein du génome.

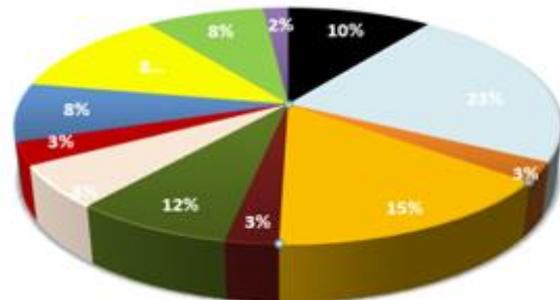
Les travaux de Catherine et ses collègues (Gilbert et al., 2010) sur l'évolution expérimentale montrent que le transfert du plasmide de *Cupriavidus taiwanensis* au pathogène *R. solanacearum* est suffisant pour rendre ce dernier capable de noduler *Mimosa pudica*. Or l'annotation du plasmide n'a révélé aucun gène lié au métabolisme d'azote ce qui est du potentiellement à la limite des logiciels de RAST (expliqué par un taux de recouvrement assez faible).

Chez la souche d'*A. tumefaciens* étudiée nous avons observé que les gènes de virulence sont préférentiellement associée aux plasmide At (Figure 23) alors que les gènes du métabolisme

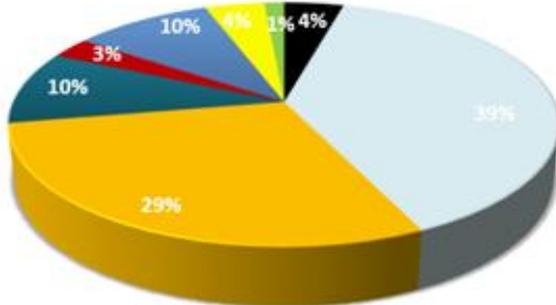
Résultats et discussion



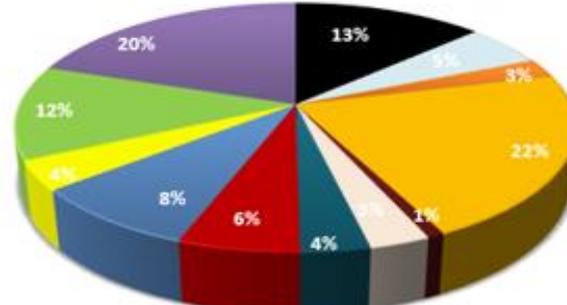
Plasmide Ti *A.tumefaciens*



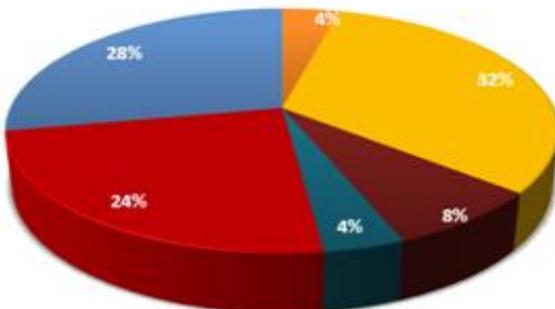
PsymA de *S.meliloti*



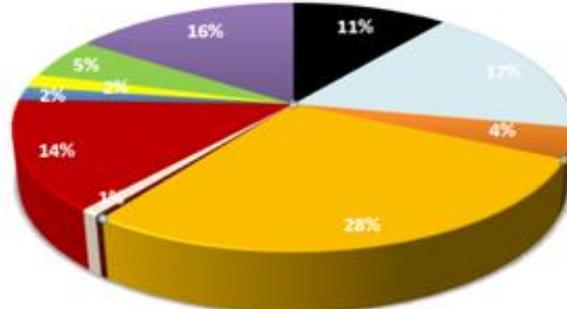
Plasmide At *A.tumefaciens*



PsymB de *S.meliloti*



C.taiwenensis



R.Solanacearum



Figure 23. Représentation du pourcentage des gènes plasmidiques de chaque voie métabolique pour les différentes souches étudiées sont localisé sur le plasmide Ti. Ces résultats indiquent que les gènes de virulences et de métabolisme peuvent être localisés sur deux plasmides

L'ensemble de ces observations, contraste avec les observations réalisées au niveau génomique, ainsi la majeure partie des différences semble être localisée au niveau des plasmides. Ceci est en accord avec le rôle du plasmide dans l'adaptation des bactéries à leurs environnements (Dziewit et al. 2015), ainsi une plus grande diversité génétique est attribuée à ce matériel extra chromosomique et reflète la complexité des mécanismes d'adaptation.

6. Les rhizobia présentent des divergences intragroupes d'ampleur équivalente à celle des pathogènes

Une comparaison des séquences génomiques entre le groupe des rhizobia ou le groupe de pathogènes a été réalisée en utilisant l'outil en ligne disponible sur la base de données MicroScope (<http://www.genoscope.cns.fr/agc/microscope/>).

Le nombre de gènes présentant un minimum d'homologie de 80% avec un recouvrement minimum de la séquence de 80% ont été isolés. La figure 24 présente le nombre de familles de gènes communes ou divergentes obtenues, une famille peut contenir un ou plusieurs gènes homologues entre les différentes souches.

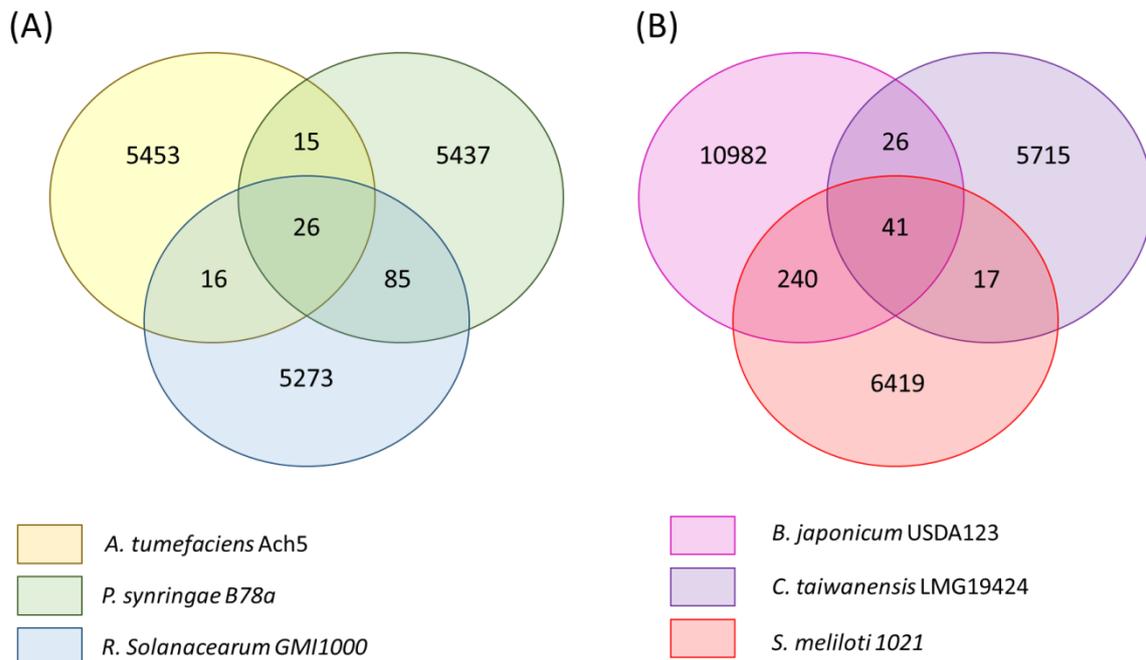


Figure 24 : Nombre de familles de gènes communes et souche spécifique au sein des rhizobia ou des pathogènes. (A) Nombre de gènes communs et divergents entre les souches pathogènes (*A. tumefaciens* Ach5, *P. syringae* B728 et *R. solanacearum* GMI1000). (B) Nombre de gènes communs et divergents entre les souches symbiotiques (*B. japonicum* USA123, *C. taiwanensis* LMG19424 et *S. meliloti* 1021).

Les pathogènes présentent un nombre de gènes souches spécifiques peu variables allant de 5273 pour *R. solanacearum* à 5453 pour *A. tumefasciens* (Figure 24A). Ce nombre est plus hétérogène dans le cas des souches symbiotiques avec 10982 gènes pour *B. japonicum* et 5715 gènes pour *S. meliloti* (Figure 24B). Par ailleurs le nombre de gènes conservés entre les pathogènes est proche de celui des rhizobia allant de 15 à 85 gènes, excepté pour *B. japonicum* et *S. meliloti* qui est de 240 gènes communs.

Ces résultats indiquent qu'appart certaines exceptions, le nombre de gènes communs et divergents est équivalent d'un groupe à l'autre. Ceci est en corrélation avec les observations précédentes qui indique une conservation des fonctions au niveau des génomes.

7. Les pathogènes et les symbiontes présentent des gènes conservés

Dans le but d'évaluer la conservation et la divergence génique entre les pathogènes et les symbiontes nous avons réalisé des comparaisons deux à deux sélectionnant les gènes qui présentent 80% d'identité avec 80% de recouvrement entre les différentes souches. La figure 25 présente les résultats pour les comparaisons entre *C. taiwanensis* Vs. *S. meliloti* ou *R. solanacearum* et *A. tumefasciens* Vs. *S. meliloti* et *R. solanacearum*.

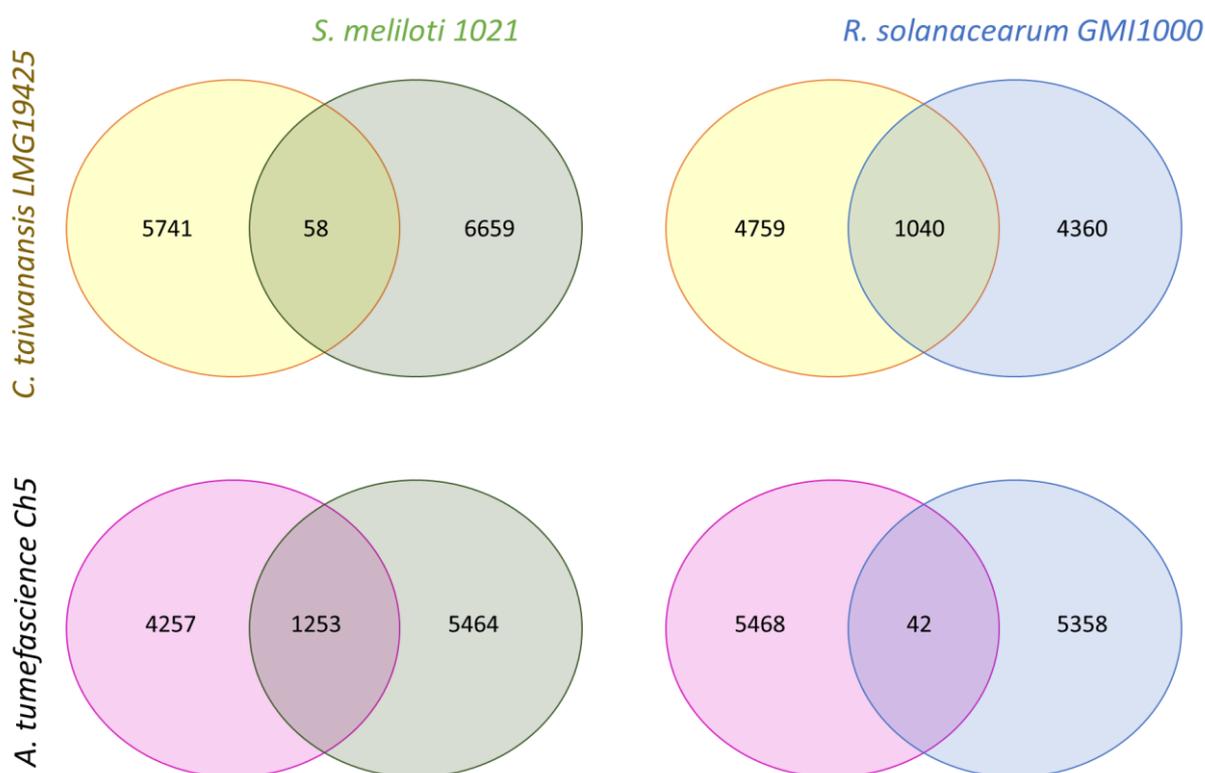


Figure 25 : Nombre de famille de gènes communs et divergents entre les pathogènes et les symbiontes.

Résultats et discussion

Les résultats indiquent que le symbionte *C. taiwanensis* ne partage que 58 familles de gènes communes avec *S. meliloti*, alors qu'il en partage 1040 avec le pathogène *R. solanacearum*. Inversement le pathogène *A. tumefaciens* partage 42 gènes en commun avec *R. solanacearum* et 1263 avec *S. meliloti*.

Ces résultats indiquent qu'il existe une conservation des génomes chromosomiques entre les pathogènes et les symbiontes et que cette conservation peut être plus importante entre les deux groupes (pathogène et rhizobia) qu'au sein d'un même groupe (Pathogène ou rhizobia).

8. L'analyse de synténie montre une conservation des génomes entre pathogènes et symbiontes

Afin de vérifier et de valider les observations précédentes nous avons réalisé une analyse de la synténie entre le chromosome de *R. solanacearum* Vs. *C. taiwanensis* d'une part et *S. meliloti* Vs. *A. tumefaciens* d'un autre part. Cette analyse a été réalisée en utilisant l'outil Line plot de la base de données MicroScope. La figure 18 présente les résultats obtenus.

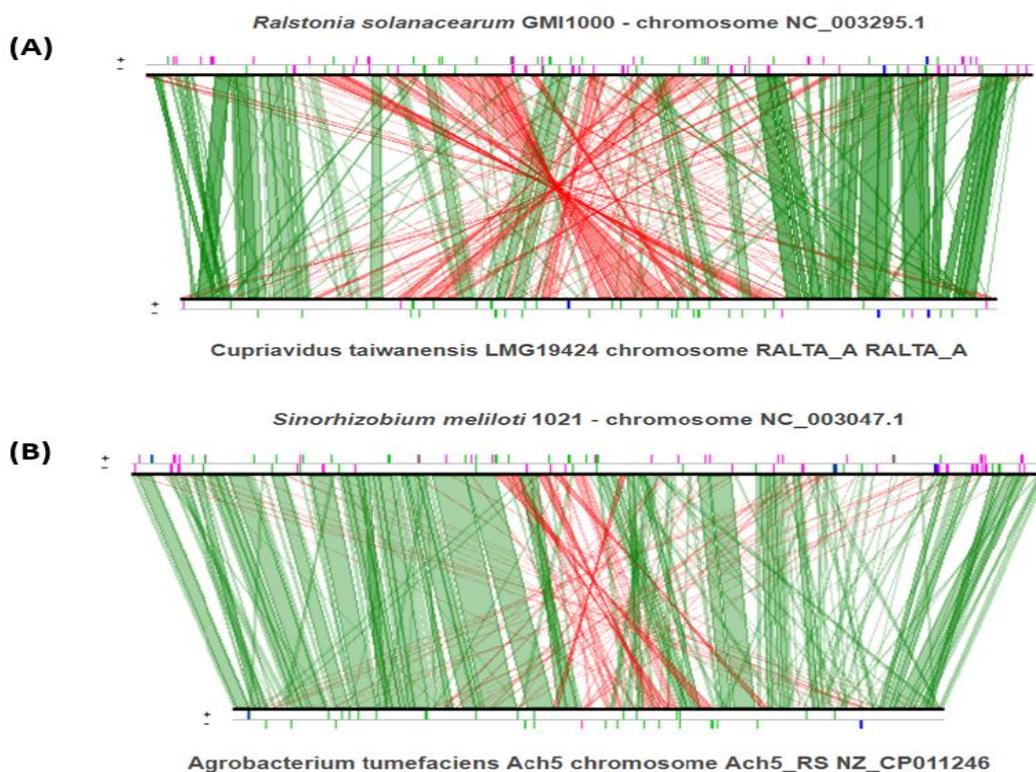


Figure 27 : Conservation et divergence génomique entre pathogène et symbionte. (A) analyse de synténie entre *R. solanacearum* GMI1000 et *C. taiwanensis* LMG19424 (B) Analyse de synténie entre *A. tumefaciens* Ach5 et *S. meliloti* 1021. En vert les régions conservés, en rouge les régions inversés.

On observe une grande conservation des régions génomiques entre *R. solanacearum* et *C. taiwanensis* ainsi qu'entre *S. meliloti* et *A. tumefasciens* avec des régions inversées plus importantes entre *R. solanacearum* et *C. taiwanensis* expliquées par la dynamique des génomes notamment via les transposants et les éléments de transition.

9. Les gènes conservés entre *S. meliloti* et les pathogènes sont exprimés au cours de la symbiose

Dans le but de valider les observations génomiques d'un point de vue transcriptionnel et d'identifier des mécanismes moléculaires conservés entre les pathogènes et les rhizobias au cours de l'interaction avec leurs hôtes. Une analyse des profils d'expression des orthologues d'*A. tumefasciens*, *B. abortus* chez *S. meliloti* 1021 est réalisé chez ce dernier au cours de la symbiose. Cette bactérie présente 354 familles d'orthologues avec le pathogène humain *Brucella abortus* (Figure 27A) et 1253 familles d'orthologues avec le pathogène *A. tumefasciens* (Figure 27C).

Nous avons identifié 369 orthologues d'*A. tumefasciens* et 230 orthologues de *B. abortus* chez *S. meliloti* 1021 pour lesquelles une expression est détectée à partir du transcriptome réalisé par Roux et al. Ce dernier a analysé l'expression des gènes de la bactérie symbiotique au sein des différentes zones de la nodosité.

Les gènes identifiés présentent une expression chez *S. meliloti* au sein de la nodosité ce qui indique que les orthologues identifiés peuvent être stimulés lors de la symbiose. L'analyse du profil d'expression des gènes indique qu'ils sont plus exprimés en ZI, IZ et ZIII pour les différents orthologues (Figure 27B et 27D), avec une plus forte stimulation en ZIII dans le cas des orthologues d'*A. tumefasciens* (Figure 27D).

L'ensemble de ces observations suggèrent que les gènes orthologues entre les pathogènes et les symbiontes sont stimulés lors de l'interaction symbiotique et donc jouent un rôle lors de ce processus.

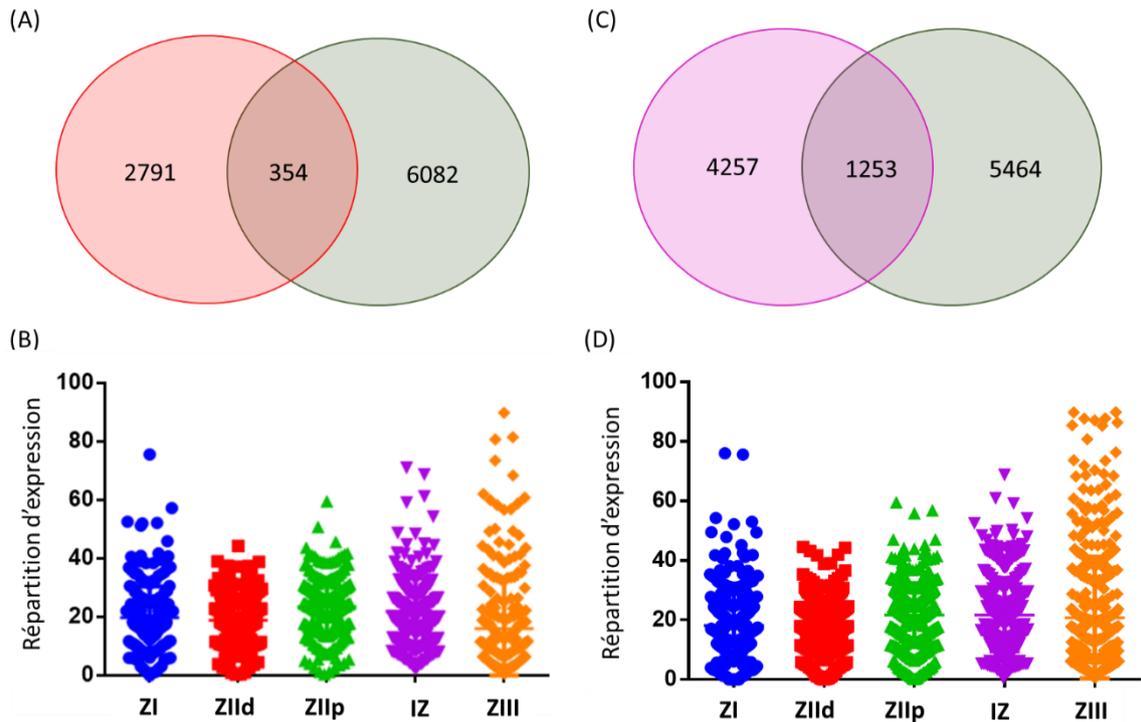


Figure 28 : Des orthologues de *B. abortus* et *A. tumefaciens* sont exprimés chez *S. meliloti* 1021 au cours de la symbiose. (A) nombre de famille de gène orthologues ou divergents entre *S. meliloti* (gris) et *B. abortus* (rouge). (B) Analyse de l'expression d'orthologues de *B. abortus* chez *S. meliloti* 1021 au de la zone I, II distal (ZII d), II proximal (ZII p), Inter zone (IZ) et de la zone III de la nodosité fixatrice d'azote. (C) nombre de famille de gènes orthologues ou divergents entre *S. meliloti* (gris) et *A. tumefaciens* (violet). (D) Analyse de l'expression d'orthologues de *B. abortus* chez *S. meliloti* 1021 au sein de la zone I, II distal (ZII d), II proximal (ZII p), Inter zone (IZ) et de la zone III de la nodosité fixatrice d'azote. Les résultats en B et D présente la répartition d'expression des gènes en pourcentage pour chaque zone.

10 Les orthologues d'*A. tumefaciens* chez *S. meliloti* présentent un profil d'expression spécifique

Dans le but d'analyser plus en détail le comportement transcriptionnel des orthologues d'*A. tumefaciens* chez *S. meliloti*. Leur expression est étudiée par catégorie fonctionnelle (COG) : l'assemblage et le fonctionnement du flagelle, l'oxydo-réduction, le chimiotactisme et le métabolisme des acides aminés. Le choix de ces catégories est lié au fait qu'elles sont requises lors du processus d'interaction hôte-symbionte.

Résultats et discussion

L'expression des orthologues est comparée à celle de l'ensemble des gènes de *S. meliloti* par catégorie afin de déterminer si les orthologues présentent un comportement équivalent à celui de l'ensemble des gènes de la bactérie, ou au contraire présentent un comportement différents. La figure 28 présente les résultats obtenus.

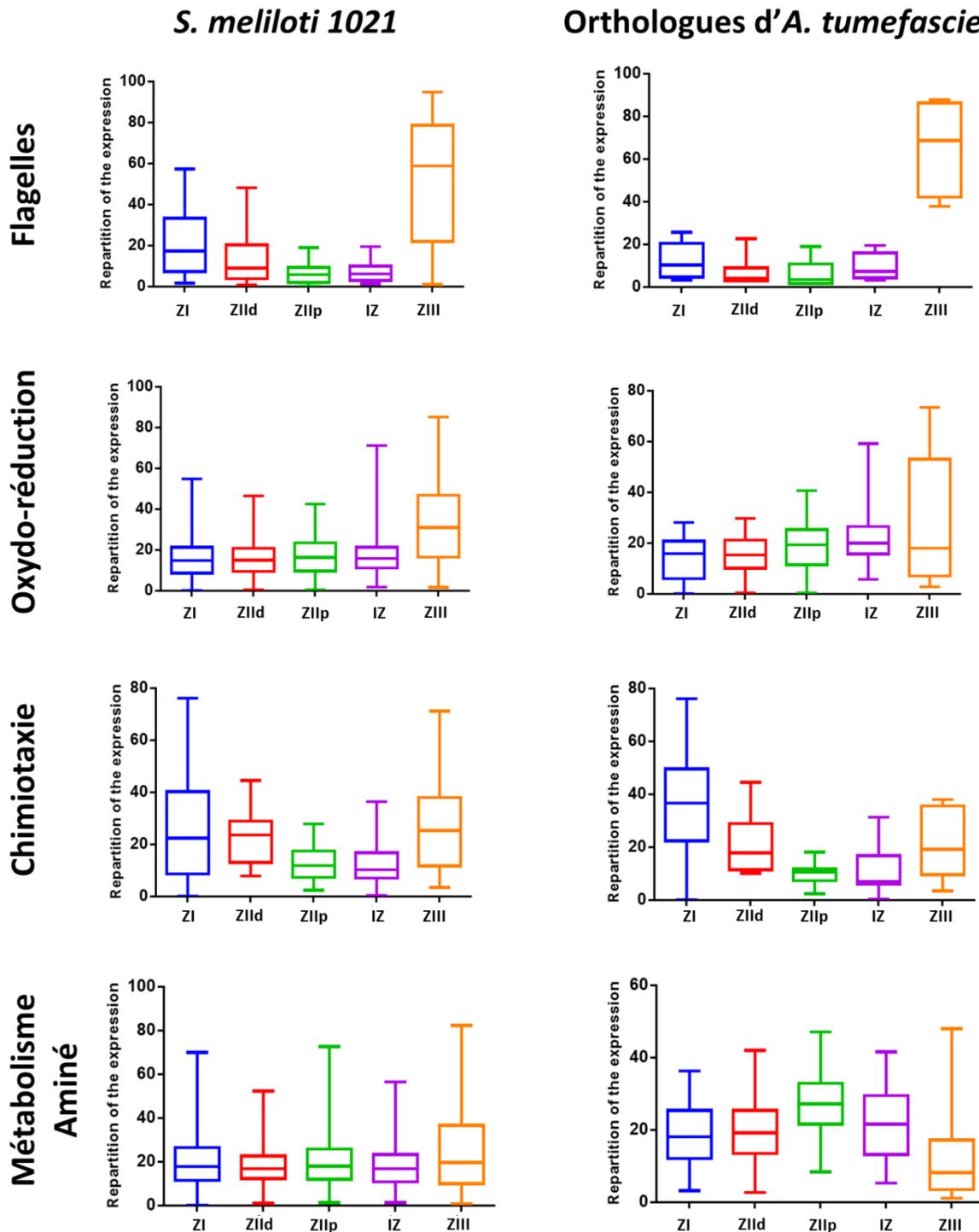


Figure 28. Le Profil d'expression des orthologues d'*A. tumefaciens* est divergent de l'ensemble des gènes de *S. meliloti*. Pour différentes voies biochimiques (formation et fonctionnement du flagelle, oxydo-réduction, chimiotaxie et métabolisme aminé) les gènes

de *S. meliloti* 1021 orthologues d'*A. tumefaciens* ont été isolés et leur expression a été comparé à celui de l'ensemble des gènes pour les différentes zones de la nodosité (ZI, ZII distale ; ZIIId, ZII proximale ; ZIIp, inter zone ; IZ, ZIII).

La détection de l'expression des gènes associés au flagelle dans les zones infectées par *S. meliloti* et plus précisément en zone de fixation d'azote (ZII) est surprenante car aucune étude n'a encore reporté un mouvement des bacteroides au sein des cellules symbiotiques. Par ailleurs les orthologues d'*A. tumefaciens* présentent un profil d'expression plus marqué en zone III comparé à celle de l'ensemble des gènes détectés (Figure 28).

Les gènes associés à l'oxydoréduction présentent un profil d'expression plus marqué dans les zones infectées dans le cas des orthologues par rapport à celui de l'ensemble des gènes de cette fonction physiologique qui intervient dans différents aspects du processus infectieux, notamment par l'activation des voies de signalisation cellulaire spécifiques et l'acclimatation métabolique. De manière inverse les orthologues impliqués dans le chimiotactisme présentent une plus forte expression en zone méristématique (ZI), alors que le profil de ce type de gènes est plus important dans les zones infectées.

Les orthologues des gènes associées aux métabolismes des acides aminés ont une expression plus marquée zone d'infection (ZIIp, ZIIId) et inter zone, alors que cette catégorie de gènes présentent une expression répartie de manière homogène entre les différentes zones.

Ces observations indiquent que le comportement des gènes orthologues d'*A. tumefaciens* est différent de celui de l'ensemble des gènes pour les différentes fonctions cellulaires étudiées. Nous pouvons ainsi définir deux groupes de gènes basés sur les profils d'expressions, le premier qui représente les gènes de *S. meliloti* et le second représente les orthologues d'*A. tumefaciens* qui ont une expression plus marquée dans les zones infectées.

Le même type d'observation est réalisé pour les orthologues de *B. abortus*, dont l'expression est représentée dans la figure 29. Comme dans le cas d'*A. tumefaciens* les gènes orthologues de *B. abortus* associés à l'oxydoréduction sont exprimés de manière croissante en zone d'infection pour atteindre un pique en zone III. Alors que les gènes associés au métabolisme des acides aminés sont les moins exprimés en zone de fixation d'azote (Figure 29). Ces résultats montrent qu'entre deux pathogènes il existe un comportement similaire des orthologues.

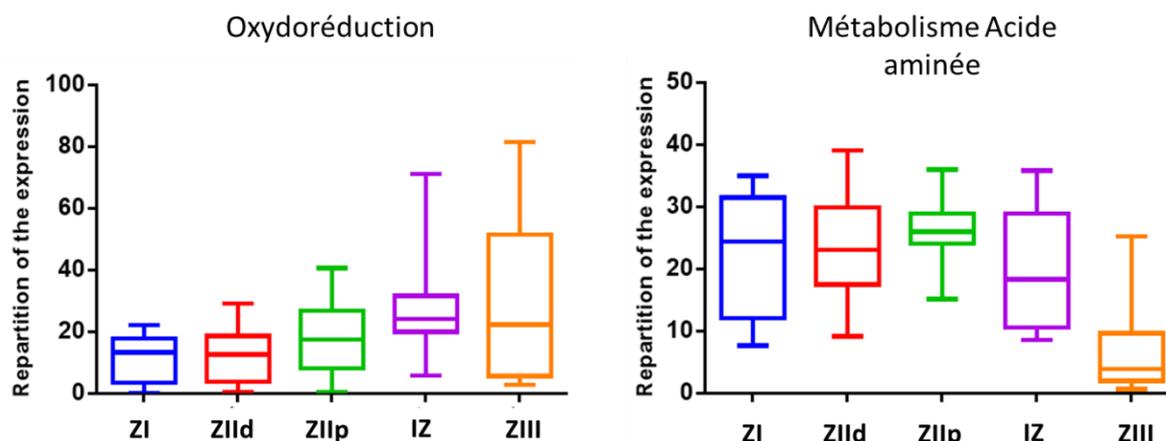


Figure 29. Expression des orthologues de *B. abortus* chez *S. meliloti* 1021 lors de la symbiose.

En postulant pour une conservation fonctionnelle entre les orthologues de la bactérie symbiotique et des pathogènes étudié d'une part et en se basant sur les profils d'expression des orthologues chez *S. meliloti* lors de la symbiose d'une autre part, on peut émettre l'hypothèse que les orthologues sont potentiellement impliqués de manière préférentielle dans l'infection et la survie intracellulaire chez *S. meliloti*.

11. Le flagelle impliqué dans la survie intracellulaire ?

La détection de l'expression de gènes associés à la biosynthèse et au fonctionnement du flagelle nous a questionné sur le rôle potentiel de ce composant cellulaire dans la survie intracellulaire des rhizobia. Par ailleurs aucune publication n'a encore mis en évidence le rôle de cet élément au sein du bacteroïde ou n'a même discuté de ce phénomène. Ces gènes sont exprimés avec des niveaux différents chez *S. meliloti*, ainsi les gènes *FlaB*, *FliQ*, *FliG*, *FlaF*, *FlgA*, *FlhB* et *FlgE* sont les gènes les plus exprimés dans la nodosité (Figure 30A).

L'analyse de la répartition d'expression globale des gènes du flagelle dans les différentes zones de la nodosité révèle que les gènes sont exprimés en zone I et zone II distale, puis une répression en zone II proximale et interzone a lieu. Leur expression est la plus forte en zone de fixation d'azote (Figure 30B). Ces résultats indiquent que les gènes du flagelle sont exprimés chez *S. meliloti* au sein des nodosités, ils sont potentiellement requis lors de la phase d'infection puis plus tard lors de la phase de fixation d'azote.

Ces résultats peuvent être expliqués par **i)** Un rôle réel du flagelle dans le processus d'infection et de survie, potentiellement en créant un système de sécrétion. **ii)** Lors de

Résultats et discussion

l'endosymbiose la plante contrôle la majeure partie des gènes bactériens néanmoins il est possible que dans le cas du flagelle, il y'a une fuite d'expression des gènes.

Dans le but de voir si l'expression des gènes du flagelle peut avoir lieu dans d'autres cas de survie intracellulaire. Nous avons vérifié l'expression des gènes chez *B. abortus* en condition contrôle et au sein de sa cellule hôte, le macrophage (Figure 30C). On observe que plusieurs gènes sont détectés avec une tendance d'expression supérieure en condition d'infection par rapport au contrôle. Les gènes les plus exprimés sont *FlgF*, *MotB*, un gène codant une protéine L-ring. Ces résultats indiquent que l'expression des gènes du flagelle semble être une caractéristique commune entre les bactéries infectant des cellules eucaryotes. Ces résultats sont à mettre en corrélation avec des observations réalisées sur des endosymbiontes d'insecte, *Buchnera aphidicola*, qui présente une conservation de l'ensemble des gènes du flagelle, alors que cette bactérie est incapable de se déplacer ou même de survivre en dehors de son hôte (Brinza et al., 2009).

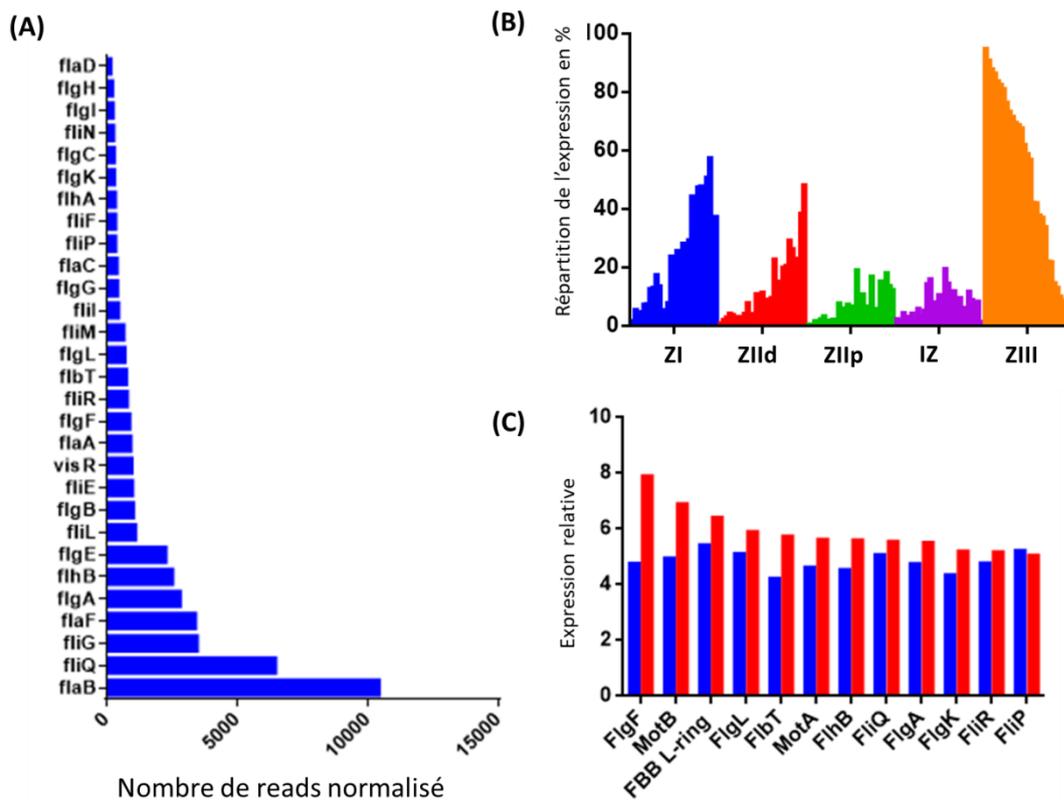


Figure 30 : Expression relative des gènes du flagelle exprimée en nombre de reads (séquence de lecture) normalisé par des contrôles internes. (B) Répartition en pourcentage de l'expression des gènes du flagelle entre les différentes zones de la nodosité. (C) Expression des gènes du flagelle chez *B. abortus* en condition de culture liquide (control) en bleu et au sein du macrophage en rouge.

12.. L'hémolysine D (HlyD) un candidat pour l'étude de la survie intracellulaire

Lors de l'analyse comparative génomique entre *B. abortus* et *S. meliloti* nous avons pu identifier un nombre de 354 familles de gènes communs entre les deux bactéries. Le tableau supplémentaire 1 présente la liste des gènes ainsi que leurs profils d'expression chez *S. meliloti* en condition symbiotique.

En se basant sur le profil d'expression ainsi que sur l'annotation génomique nous nous sommes focalisés sur le gène *HlyD* qui est un gène de virulence pour plusieurs pathogènes (López-fernández et al., 2015). Six homologues du gène sont exprimés chez *S. meliloti* en condition de symbiose avec des niveaux d'expressions différents (Figure 31A) avec HlyD_A le plus exprimé et HlyD_F le moins stimulé.

L'analyse de la répartition d'expression des six gènes montre qu'ils sont préférentiellement exprimés en zone d'infection, interzone et zone de fixation avec une absence d'expression en zone I pour HlyD_E (Figure 31B). Par ailleurs, un homologue HlyD est détecté au sein de nodosités de *M. truncatula* induite par la souche *Sinorhizobium medicae* WSM419. La protéine est spécifiquement détectée au niveau du symbiosome (Figure 31C) (Berrabah et al 2018).

Ces résultats montrent que l'hémolysine D est préférentiellement exprimée au niveau des zones infectées par *Sinorhizobium sp* et que la protéine est accumulée au niveau du symbiosome où elle joue probablement son rôle.

Afin de vérifier l'implication de HlyD dans la survie intracellulaire de *B. abortus*, nous avons vérifié l'expression de l'orthologue ainsi que des membres de sa famille chez la bactérie en condition d'infection de cellules macrophagiques. Quatre gènes codent l'hémolysine D sont détectés. HlyD_1 à 2 présente une augmentation d'expression du gène au cours de l'infection et HlyD_3 présente une légère réduction. HlyD_4 quant à lui ne présente pas de variation significative. Ceci est en corrélation avec les observations réalisées chez *S. meliloti* et suggèrent que HlyD4 est un facteur important pour le processus de survie intracellulaire.

L'HlyD est l'un des membres les mieux caractérisés de la famille des protéines de fusion membranaire (MFP) formant un système de sécrétion type I (Letoffe et al. 1990). Cette protéine est ancrée dans la membrane cytoplasmique par un seul domaine

transmembranaire et possède un grand domaine périplasmique. Les 100 acides aminés carboxy-terminaux forment avec HlyB un complexe stable impliqué dans le transport de l'hémolysine α chez *E. coli* uropathogène. Cette dernière est un facteur de virulence important en raison de son activité cytotyrique et cytotoxique contre un large éventail de types de cellules de mammifères (érythrocytes, granulocytes, monocytes et cellules endothéliales) (Letoffe et al. 1990).

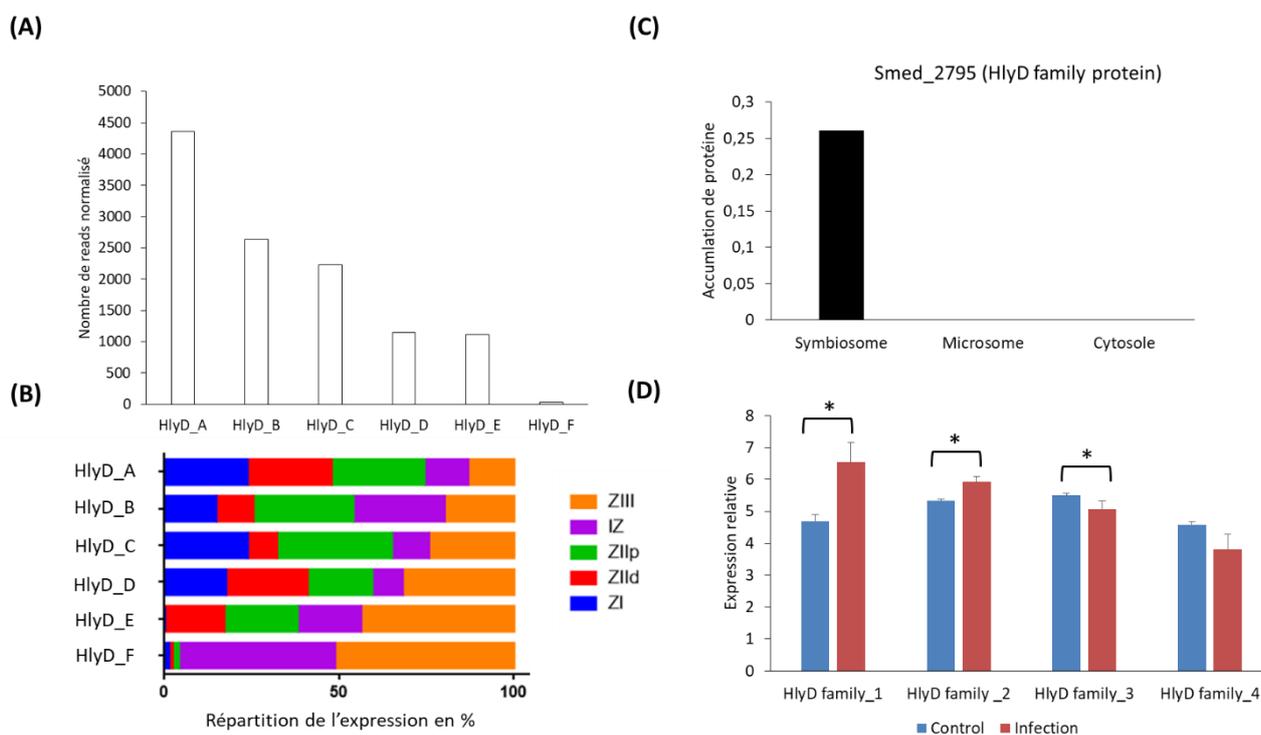


Figure 31 :HlyD et survie intracellulaire. (A) Niveaux d'expression globale en nombre de reads normalisés de six *HlyD* détectées au niveau du transcriptome de *S. meliloti* 1021. (B) Répartition en pourcentage de l'expression des *HlyD* de *S. meliloti* 1021 dans les différentes zones la nodosité. (C) Accumulation d'une HlyD (Smed_2795) chez la bactérie symbiotique *Sinorhizobium medicae* WSM419. (D) Expression relative de quatre *HlyD* de *B. abortus* en condition contrôle et d'infection de cellules macrophagiques. La barre d'erreur représente l'écartype. L'Astérisque indique la significativité du test statistique Man Whitney.

13 NolX, un gène important pour la symbiose et la pathogénie

Lors de nos prospections, nous avons identifié par analyse de transcriptome de *R. solanacearum* sur tomate un gène induit, conservé entre le pathogène et plusieurs rhizobia (Figure 32). La caractérisation *in silico* de la protéine a révélé la présence des régions transmembranaires et l'absence du peptide signal (Figure sup.1 et 2), indiquant ainsi que la protéine est de nature transmembranaire.

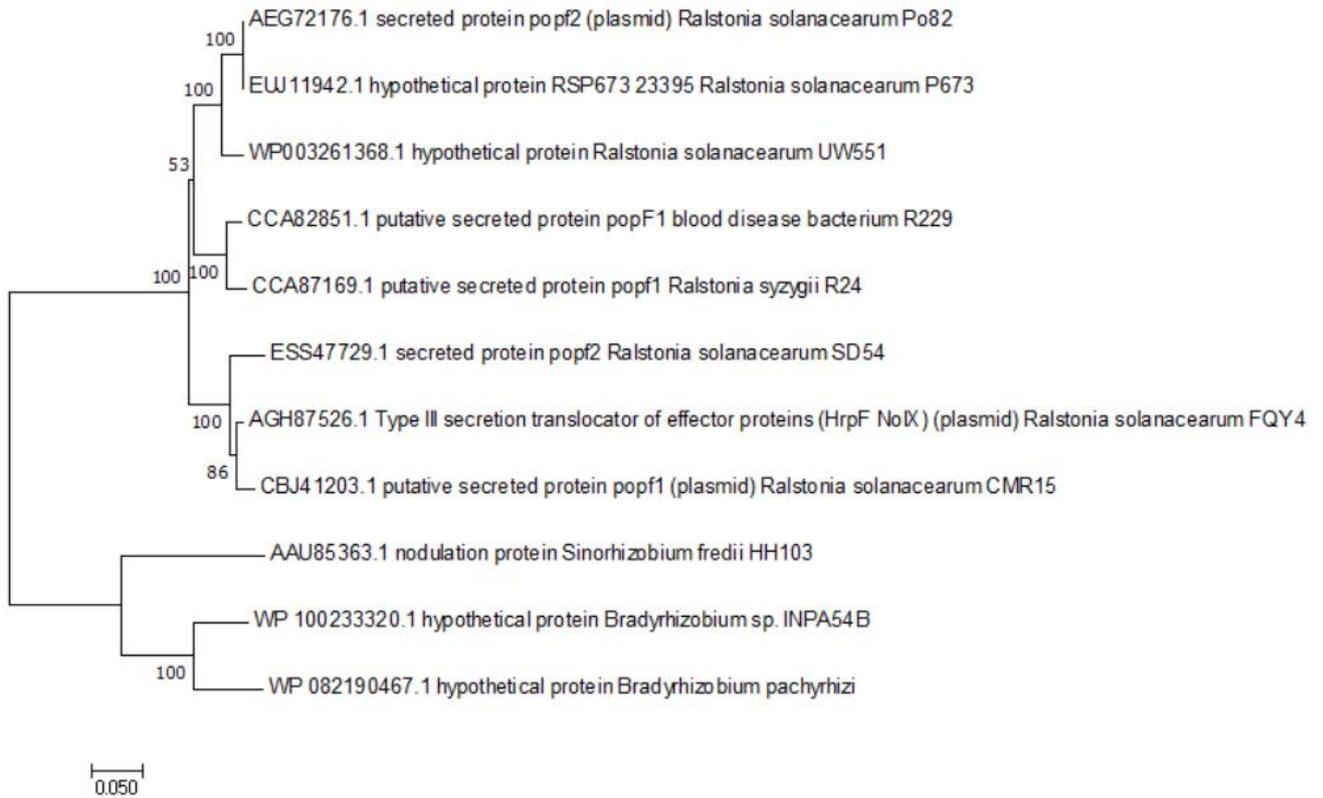


Figure 32 : Arbre phylogénique des orthologues de la protéine NoIX chez les différentes souches ayant une similarité supérieur à 60%. (Construit par la méthode Neighbors joining)

Nous avons cherché à savoir quel type de protéine interagissant avec NoIX, pour cela une recherche sur la base de données STRING répertoriant les interactions protéine-protéine connues est réalisée (Figure 33). La majeure partie des interactants retrouvés intervient dans la formation du système de sécrétion de type III bactérien.

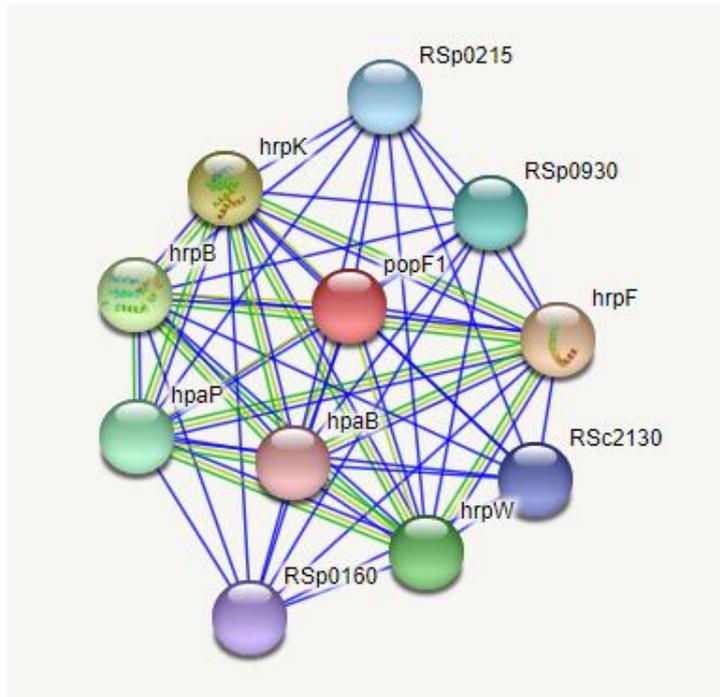


Figure 33. Réseau d'interactants de la protéine NolX. Les interactants sont des protéines composants le système de sécrétion type 3 (hrpK ;hrpB ;hrpW ;hrpF), des protéines chaperonnes et des effecteurs.

NolX est une protéine intervenant dans la formation de système de sécrétion requis lors du processus de pathogénie, mais aussi potentiellement dans la symbiose car présente chez différentes souches de rhizobia (*Sinorhizobium fredii* ;*Bradyrhizobium sp* et *Bradyrhizobium apachyrhizi*).

Cette protéine est apparentée à la protéine de la famille popF2, les travaux de Masson-Boivin et al sur l'évolution expérimentale ont montré qu'un double mutant des translocons *popF1* et son homologue *popF2* est altéré dans l'invasion intracellulaire des cellules hôtes chez les rhizobia (Gilbert et al., 2010).

Conclusion

Lors de notre projet nous nous sommes intéressés aux interfaces symbiose-immunité en utilisant une nouvelle approche dans l'identification de candidats moléculaires pouvant intervenir dans les différents processus mais aussi en réalisant une analyse plus globale afin de comparer les points communs et divergents entre des pathogènes et les rhizobias.

Nos résultats ont montré que la majeure partie des divergences quantitatives et qualitatives en gènes entre pathogènes et symbiontes se situe essentiellement sur les plasmides, au niveau des génomes ; peu de variations sont observées en terme de fonction biochimique et elle semble être essentiellement qualitative, néanmoins une analyse en utilisant des logiciels plus performants doit être réalisée afin de confirmer ces observations.

La comparaison génomique entre pathogènes ou entre symbiontes a révélé que l'amplitude des différences est équivalente au sein d'un même groupe (Pathogène ou symbionte), suggérant un même niveau de divergence intragroupe et appuyant l'hypothèse d'une spécialisation via le matériel extra chromosomique. Cette théorie se confirme lors de l'analyse comparative entre pathogène et symbionte, ainsi le rhizobium *C. taiwanensis* partage plus de gènes en commun avec le pathogène *R. solanacearum* qu'avec *S. meliloti*, inversement, ce dernier présente plus de gènes en commun avec *A. tumefaciens* qu'avec *C. taiwanensis*.

Au niveau du transcriptome, les gènes partagés entre *S. meliloti* et *A. tumefaciens* présentent un profil d'expression divergent de celui de l'ensemble des gènes exprimés chez *S. meliloti* lors de la symbiose. Ainsi on retrouve globalement une augmentation d'expression des orthologues d'*A. tumefaciens* au niveau des zones d'infection et au niveau des nodosités, ces résultats suggèrent qu'en fonction de l'origine du gène (orthologue ou non) le profil d'expression peut être divergent.

Afin d'analyser les points communs entre deux bactéries ne partageant que l'état intracellulaire, une comparaison génomique entre *S. meliloti* et *B. abortus* est réalisée, elle révèle la présence de plusieurs centaines de gènes, une grande partie d'entre-eux sont exprimés chez l'endosymbionte et sont stimulés au sein des zones infectées. Plusieurs gènes intervenant dans la virulence sont induits, notamment l'hémolysine D. Plusieurs membres de cette famille sont exprimés lors de la symbiose, dont sont accumulés au sein du symbiosome au niveau des régions infectées Zone II et Zone III. De manière intéressante les orthologues chez *B. abortus* sont induits lors de l'infection des cellules macrophagiques. Ces résultats font de HlyD un bon candidat pour réaliser des études fonctionnelles et de confirmer ainsi son implication dans les processus de symbiose et de pathogénie.

L'approche employée lors de notre étude bien que nouvelle a permis de mettre en évidence les différences et les points communs entre pathogènes et symbiontes ainsi que d'identifier des

Conclusion

gènes candidats permettant de connecter l'immunité à la symbiose. Ce travail pionnier doit être poursuivi afin de valider les hypothèses émises et les observations réalisées.

1. Ailloud, F. (2015). *Le pouvoir pathogène chez Ralstonia solanacearum phylotype II génomique intégrative et paysages transcriptomiques en relation avec l'adaptation à l'hôte* Florent Ailloud To cite this version : HAL Id : tel-01211458 UNIVERSITÉ * DE * LA * RÉUNION * ! TH.
2. Alexandersson, E., Jacobson, D., Vivier, M. A., Weckwerth, W., & Andreasson, E. (2014). Field-omics understanding large-scale molecular data from field crops. *Frontiers in Plant Science*, 5(June), 1–6. <https://doi.org/10.3389/fpls.2014.00286>
3. Aslam, S. N., Newman, M., Erbs, G., Morrissey, K. L., Chinchilla, D., Boller, T., ... Cooper, R. M. (2008). *Report Bacterial Polysaccharides Suppress Induced Innate Immunity by Calcium Chelation*. 1078–1083. <https://doi.org/10.1016/j.cub.2008.06.061>
4. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., ... Zagnitko, O. (2008). *The RAST Server : Rapid Annotations using Subsystems Technology*. 15, 1–15. <https://doi.org/10.1186/1471-2164-9-75>
5. Baier, R., Schiene, K., Kohring, B., Flaschel, E., & Niehaus, K. (1999). *Alfalfa and tobacco cells react differently to chitin oligosaccharides and Sinorhizobium meliloti nodulation factors*.
6. Berendsen, R. L., Pieterse, C. M. J., & Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. *Trends in Plant Science*, 17(8), 478–486. <https://doi.org/10.1016/j.tplants.2012.04.001>
7. Berrabah, F. (2016). *Contrôle symbiotique de l'immunité au cours des étapes tardives de la symbiose Medicago-Sinorhizobium*.
8. Berrabah, F., Bourcy, M., Cayrel, A., Eschstruth, A., Mondy, S., Ratet, P., & Gourion, B. (2014). Growth conditions determine the DNF2 requirement for symbiosis. *PLoS One*, 9(3), e91866. <https://doi.org/10.1371/journal.pone.0091866>
9. Berrabah, F., Bourcy, M., Eschstruth, A., Cayrel, A., Guefrachi, I., Mergaert, P., ... Ratet, P. (2014). A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. *New Phytologist*, 203(4), 1305–1314. <https://doi.org/10.1111/nph.12881>
10. Berrabah, F., Ratet, P., & Gourion, B. (2015). *Multiple steps control immunity during the intracellular accommodation of rhizobia*. 66(7), 1977–1985. <https://doi.org/10.1093/jxb/eru545>
11. Besser, J., Carleton, H. A., Gerner-smidt, P., Lindsey, R. L., & Trees, E. (2018). *to the*

- Study and Control of Bacterial Infections*. 24(4), 335–341.
<https://doi.org/10.1016/j.cmi.2017.10.013>.Next-Generation
12. Bidanel, J. P., Boichard, D., & Chevalet, C. (2008). From genetics to genomics TT - De la genetique a la genomique. *INRA Productions Animales*, 21(1), 15-31 ST-From genetics to genomics.
13. Boivin, S., Fonouni-farde, C., & Frugier, F. (2016). *How Auxin and Cytokinin Phytohormones Modulate Root Microbe Interactions*. 7(August), 1–12.
<https://doi.org/10.3389/fpls.2016.01240>
14. Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., ... Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, 41(Web Server issue), 29–33.
<https://doi.org/10.1093/nar/gkt282>
15. Boucher, C., Genin, S., & Arlat, M. (2001). *Concepts actuels sur la pathogénie chez les bactéries phytopathogènes*. 324, 915–922.
16. Boucher, C., Genin, S., & Arlat, M. (2002). Concepts actuels sur la pathogénie chez les bactéries phytopathogènes. *Comptes Rendus de l'Académie Des Sciences - Series III - Sciences de La Vie*, 324(10), 915–922. [https://doi.org/10.1016/s0764-4469\(01\)01375-0](https://doi.org/10.1016/s0764-4469(01)01375-0)
17. Bourcy, M., Berrabah, F., Ratet, P., & Gourion, B. (2013). Failure of self-control - Defense-like reactions during legume:rhizobia symbiosis. *Plant Signaling & Behavior*, 8(4), e23915. <https://doi.org/10.4161/psb.23915>
18. Bourcy, M., Brocard, L., Pislariu, C. I., Cosson, V., Mergaert, P., Tadege, M., ... Ratet, P. (2013). *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytologist*, 197(4), 1250–1261.
<https://doi.org/10.1111/nph.12091>
19. Brinza, L., Viñuelas, J., Cottret, L., Calevro, F., Rahbé, Y., Febvay, G., ... Charles, H. (2009). Systemic analysis of the symbiotic function of *Buchnera aphidicola*, the primary endosymbiont of the pea aphid *Acyrtosiphon pisum*. *Comptes Rendus - Biologies*, 332(11), 1034–1049. <https://doi.org/10.1016/j.crv.2009.09.007>
20. Cangelosi, G. A., Hung, L., Puvanesarajah, V., Stacey, G., Ozga, D. A., Leigh, J. A., & Nester, E. W. (1987). Common loci for *Agrobacterium tumefaciens* and *Rhizobium meliloti* exopolysaccharide synthesis and their roles in plant interactions. *Journal of Bacteriology*, 169(5), 2086–2091. <https://doi.org/10.1128/jb.169.5.2086-2091.1987>

21. Carter, M., & Shieh, J. C. (2010). Molecular Cloning and Recombinant DNA Technology. *Guide to Research Techniques in Neuroscience*, (1), 207–227.
<https://doi.org/10.1016/b978-0-12-374849-2.00009-4>
22. Chandler, M. (n.d.). *Les génomes et ses enjeux*.
23. Dalsing, B. L., Truchon, A. N., Gonzalez-Orta, E. T., Milling, A. S., & Allen, C. (2015). *Ralstonia solanacearum* uses inorganic nitrogen metabolism for virulence, ATP production, and detoxification in the oxygen-limited host xylem environment. *MBio*, 6(2), 1–13. <https://doi.org/10.1128/mBio.02471-14>
24. Deakin, W. J., & Broughton, W. J. (n.d.). *Symbiotic use of pathogenic strategies : rhizobial protein secretion systems*.
25. Dixon, R., & Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology*, 2(8), 621–631. <https://doi.org/10.1038/nrmicro954>
26. Gallopin, M., Gallopin, M., & Statistiques, R. (2017). *Classification et inférence de réseaux pour les données RNA-seq To cite this version : HAL Id : tel-01424124 Thèse de Doctorat de l ' Université Paris-Saclay préparée à l ' Université Paris-Sud 11 École doctorale n o 574 Spécialité : Mathématiques aux Int*.
27. Gastroenterology, E. (2017). *The ecological community of commensal , symbiotic , and pathogenic gastrointestinal microorganisms – an appraisal*. 91–103.
28. Genin, S. (2010). Molecular traits controlling host range and adaptation to plants in *Ralstonia solanacearum*. *New Phytologist*, (August), 920–928.
29. Gianfreda, L. (2015). *Enzymes of importance to rhizosphere processes*. 15(2), 283–306.
30. Gibson, K. E., Kobayashi, H., & Walker, G. C. (2008). Molecular Determinants of a Symbiotic Chronic Infection. *Annual Review of Genetics*, 42(1), 413–441.
<https://doi.org/10.1146/annurev.genet.42.110807.091427>
31. Gilbert, L. B., Heeb, P., Gris, C., Timmers, T., Batut, J., & Masson-boivin, C. (2010). *Experimental Evolution of a Plant Pathogen into a Legume Symbiont*. 8(1).
<https://doi.org/10.1371/journal.pbio.1000280>
32. Gourion, B., Berrabah, F., Ratet, P., & Stacey, G. (2015). Rhizobium – legume symbioses : the crucial role of plant immunity. *Trends in Plant Science*, 20(3), 186–194. <https://doi.org/10.1016/j.tplants.2014.11.008>
33. Humaine, A. (2011). *La bactérie et son hôte*.
34. John A. Leigh, D. L. C. (1992). EXOPOL YSACCHARIDES IN PLA1'JT - BACTERIAL INTERACTIONS. *Annu.Rev.Microbiol*, 46.

35. Kereszt, A., Mergaert, P., Maróti, G., & Kondorosi, É. (2011). Innate immunity effectors and virulence factors in symbiosis. *Current Opinion in Microbiology*, 14(1), 76–81. <https://doi.org/10.1016/j.mib.2010.12.002>
36. Kerfeld, C. A., & Scott, K. M. (2011). Using BLAST to Teach “E-value-tionary” Concepts. *PLoS Biology*, 9(2), <https://doi.org/10.1371/journal.pbio.1001014>
37. Krachler, A. M., & Orth, K. (2013). Targeting the bacteria-host interface strategies in anti-adhesion therapy. *Virulence*, 4(4), 284–294. <https://doi.org/10.4161/viru.24606>
38. Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis Kimberly. *Cold Spring Harbor Protocols*, 2015(11), 951–969. <https://doi.org/10.1101/pdb.top084970.RNA>
39. Liu, B. H. (2018). Statistical Genomics. *Statistical Genomics*, (301), 93–110. <https://doi.org/10.1201/9780203738658>
40. Liu, D., Hunt, M., & Tsai, I. J. (2018). Inferring synteny between genome assemblies: A systematic evaluation. *BMC Bioinformatics*, 19(1), 1–13. <https://doi.org/10.1186/s12859-018-2026-4>
41. Lòpez-fernàndez, S., Sonego, P., Moretto, M., Pancher, M., Engelen, K., Pertot, I., & Campisano, A. (2015). Whole-genome comparative analysis of virulence genes unveils similarities and differences between endophytes and other symbiotic bacteria. 6(May), 1–15. <https://doi.org/10.3389/fmicb.2015.00419>
42. Lopez-Gomez, M., Sandal, N., Stougaard, J., & Boller, T. (2012). Interplay of flg22-induced defence responses and nodulation in *Lotus japonicus*. *Journal of Experimental Botany*, 63(1), 393–401. <https://doi.org/10.1093/jxb/err291>
43. Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., ... Foster, G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13(6), 614–629. <https://doi.org/10.1111/j.1364-3703.2012.00804.x>
44. Mardis, E. R. (2017). DNA sequencing technologies: 2006-2016. *Nature Protocols*, 12(2), 213–218. <https://doi.org/10.1038/nprot.2016.182>
45. Maunoury, N., Redondo-nieto, M., Bourcy, M., Velde, W. Van De, Alunni, B., Ratet, P., ... Mergaert, P. (2010). Differentiation of Symbiotic Cells and Endosymbionts in *Medicago truncatula* Nodulation Are Coupled to Two. 5(3). <https://doi.org/10.1371/journal.pone.0009519>
46. May, P., Wienkoop, S., Kempa, S., Usadel, B., Christian, N., Rupprecht, J., ... Walther, D. (2008). Metabolomics- and proteomics-assisted genome annotation and

- analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. *Genetics*, *179*(1), 157–166. <https://doi.org/10.1534/genetics.108.088336>
47. McCouch, S. R. (2001). Genomics and Synteny. *Plant Physiology*, *125*(1), 152–155. <https://doi.org/10.1104/pp.125.1.152>
48. Médigue, C., Bocs, S., Labarre, L., Mathé, C., & Vallenet, D. (2002). Bio-informatique (1) l'annotation in silico des séquences génomiques. *Medecine/Sciences*, *18*(2), 237–250.
49. N, E. C. D. O., Geiger, M. E., & Giorgio, P. (2016). *De L ' U Niversite P Aris -S Aclay Preparee a L ' U Niversite P Aris -S Ud.*
50. Ott, T., Van Dongen, J. T., Günther, C., Krusell, L., Desbrosses, G., Vigeolas, H., ... Udvardi, M. K. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology*, *15*(6), 531–535. <https://doi.org/10.1016/j.cub.2005.01.042>
51. Pritchard, L., & Birch, P. (2011). A systems biology perspective on plant-microbe interactions: Biochemical and structural targets of pathogen effectors. *Plant Science*, *180*(4), 584–603. <https://doi.org/10.1016/j.plantsci.2010.12.008>
52. Rubio, L. M., & Ludden, P. W. (n.d.). *Biosynthesis of the Cofactor of Nitrogenase*. <https://doi.org/10.1146/annurev.micro.62.081307.162737>
53. Sevellec, M. (2018). *les differents microbiotes de l'holobionte grand coregone (Coregonus Clupeaformis) dans un contexte de spéciation.*
54. Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135–1145. <https://doi.org/10.1038/nbt1486>
55. Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C. I., Nakashima, J., ... Udvardi, M. K. (2013). The C2H2 transcription factor regulator of symbiosome differentiation represses transcription of the secretory pathway gene VAMP721a and promotes symbiosome development in *Medicago truncatula*. *Plant Cell*, *25*(9), 3584–3601. <https://doi.org/10.1105/tpc.113.114017>
56. Steele, C. D., Tarabichi, M., Oukrif, D., Webster, A. P., Ye, H., Fittall, M., ... Pillay, N. (2019). Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. *Cancer Cell*, *35*(3), 441–456.e8. <https://doi.org/10.1016/j.ccell.2019.02.002>
57. Tambourin, P. (2005). Les grands instruments de la biologie moléculaire, prémices de la médecine de demain. *La Revue Pour l'histoire Du CNRS*, (12). <https://doi.org/10.4000/histoire-cnrs.1308>
58. Timmers, A. C. J., Soupène, E., Auriac, M.-C., de Billy, F., Vasse, J., Boistard, P., &

- Truchet, G. (2000). Saprophytic Intracellular Rhizobia in Alfalfa Nodules. *Molecular Plant-Microbe Interactions*, 13(11), 1204–1213.
<https://doi.org/10.1094/MPMI.2000.13.11.1204>
59. Udvardi, M., & Poole, P. S. (2013). Transport and Metabolism in Legume-Rhizobia Symbioses. *Annual Review of Plant Biology*, 64(1), 781–805.
<https://doi.org/10.1146/annurev-arplant-050312-120235>
60. Uppalapati, S. R., Ishiga, Y., Wangdi, T., Urbanczyk-Wochniak, E., Ishiga, T., Mysore, K. S., & Bender, C. L. (2008). Pathogenicity of *Pseudomonas syringae* pv. tomato on Tomato Seedlings: Phenotypic and Gene Expression Analyses of the Virulence Function of Coronatine. *Molecular Plant-Microbe Interactions*, 21(4), 383–395. <https://doi.org/10.1094/mpmi-21-4-0383>
61. Urtuvia, V., Maturana, N., Acevedo, F., Peña, C., & Díaz-Barrera, A. (2017). Bacterial alginate production: an overview of its biosynthesis and potential industrial production. *World Journal of Microbiology and Biotechnology*, 33(11), 0.
<https://doi.org/10.1007/s11274-017-2363-x>
62. Van de Velde, W., Guerra, J. C. P., De Keyser, A., De Rycke, R., Rombauts, S., Maunoury, N., ... Goormachtig, S. (2006). Aging in legume symbiosis. A molecular view on nodule senescence in *Medicago truncatula*. *Plant Physiology*, 141(2), 711–720. <https://doi.org/10.1104/pp.106.078691>
63. Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E. A., Menck, C. F. M., da Silva, A. C. R., Ferro, J. A., ... Simpson, A. J. (2002). Comparative Genomic Analysis of Plant-Associated Bacteria. *Annual Review of Phytopathology*, 40(1), 169–189.
<https://doi.org/10.1146/annurev.phyto.40.030402.090559>
64. Vasse, J., De Billy, F., Camut, S., & Truchet, G. (1990). Correlation between ultrastructural differentiation of bacterioids and nitrogen fixation in alfalfa nodules. *Journal of Bacteriology*, 172(8), 4295–4306.
65. Wang, C., Yu, H., Luo, L., Duan, L., Cai, L., He, X., ... Zhang, Z. (2016). *NODULES WITH ACTIVATED DEFENSE 1* is required for maintenance of rhizobial endosymbiosis in *Medicago truncatula*.
66. Willey, J., Sherwood, L., & Woolverton, C. (2007). Prescott, Harley, and Klein's Microbiology, Seventh Edition. *Microbiology*, 1088.
67. Xiao, T. T., Schilderink, S., Moling, S., Deinum, E. E., Kondorosi, E., Franssen, H.,

- ... Bisseling, T. (2014). Fate map of *Medicago truncatula* root nodules. *Development (Cambridge, England)*, *141*(18), 3517–3528. <https://doi.org/10.1242/dev.110775>
68. Zhao, T., & Schranz, M. E. (2017). Network approaches for plant phylogenomic synteny analysis. *Current Opinion in Plant Biology*, *36*, 129–134. <https://doi.org/10.1016/j.pbi.2017.03.001>
69. Zhou, Y., Tang, J., Walker, M. G., Zhang, X., Wang, J., Hu, S., ... Wang, J. (2003). Gene identification and expression analysis of 86,136 Expressed Sequence Tags (EST) from the rice genome. *Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute*, *1*(1), 26–42. [https://doi.org/10.1016/S1672-0229\(03\)01005-2](https://doi.org/10.1016/S1672-0229(03)01005-2)

