



HAL
open science

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:**

Florent Ailloud. Le pouvoir pathogène chez *Ralstonia solanacearum* phylotype II génomique intégrative et paysages transcriptomiques en relation avec l'adaptation à l'hôte. Biologie moléculaire. Université de la Réunion, 2015. Français. NNT : 2015LARE0009 . tel-01211458

HAL Id: tel-01211458

<https://theses.hal.science/tel-01211458v1>

Submitted on 5 Oct 2015

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

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

UNIVERSITÉ DE LA RÉUNION
Faculté des Sciences et Technologies
Ecole doctorale Sciences Technologies et Santé

THÈSE
Pour obtenir le diplôme de
DOCTORAT

Discipline : Microbiologie, spécialité phytopathologie
UMR Peuplements Végétaux et Bioagresseurs en milieu Tropical

**Le pouvoir pathogène chez *Ralstonia solanacearum* phylotype II :
génomique intégrative et paysages transcriptomiques en relation
avec l'adaptation à l'hôte.**

par

Florent AILLOUD

Soutenu le 3 avril 2015 devant le jury composé de :

Stéphane POUSSIER, <i>Professeur, Université de la Réunion</i>	Président
Sylvain BRISSE, <i>Directeur de recherche, Institut Pasteur</i>	Rapporteur
Stéphane GENIN, <i>Directeur de recherche, CNRS</i>	Rapporteur
Charles MANCEAU, <i>Directeur de la Santé des Végétaux, ANSES</i>	Examineur
Valérie VERDIER, <i>Directrice de recherche, IRD</i>	Examinatrice
Philippe PRIOR, <i>Directeur de recherche, INRA</i>	Directeur de thèse

Titre : Le pouvoir pathogène chez *Ralstonia solanacearum* phylotype II : génomique intégrative et paysages transcriptomiques en relation avec l'adaptation à l'hôte.

Ralstonia solanacearum est une bactérie phytopathogène à la gamme d'hôte exceptionnellement large et à la répartition mondiale. Cet organisme présente une biologie à facettes multiples et s'est adapté à quasiment tous les types de sols, à la vie planctonique, et à de nombreux hôtes et plantes réservoirs. Cette capacité d'adaptation est attestée par une très forte hétérogénéité des souches qui unifient ce complexe d'espèces, aussi bien au plan de la diversité génétique, phénotypique, que de la gamme d'hôte. Des approches phylogénétiques ont montré une structuration de la population mondiale en quatre phylotypes qui correspondent globalement à l'origine géographique des souches. Les travaux de thèse portent sur des souches du phylotype II qui ont valeur de modèle expérimental car épidémiologiquement inféodées à un hôte particulier : souches **Moko** pathogènes du bananier, souches '**Brown rot**' adaptées à la pomme de terre et souches émergentes **NPB**, un variant du pouvoir pathogène. La question de recherche centrale porte sur la compréhension des **mécanismes d'adaptation à l'hôte**. Pour cela, une dizaine de génomes ont été séquencés dans une perspective (i) de revisiter la taxonomie de ce complexe d'espèce, (ii) d'en faire une analyse génomique comparative et (iii) d'analyser les paysages transcriptomiques produits lors de l'infection (*in planta*). L'ensemble de ces approches complémentaires permettent ainsi d'intégrer la complexité génétique et phénotypique de l'organisme de manière plus **systémique**.

Title : Pathogenicity of *Ralstonia solanacearum* phylotype II : integrative genomics and transcriptomic landscapes associated with host specificity.

Ralstonia solanacearum is a plant pathogenic bacterium globally distributed with a particularly broad host range. This organism is biologically diverse and is adapted to all types of soil, to planktonic lifestyle and to many plant hosts and natural reservoirs. This bacterium is a species complex and its genetic, phenotypic and host range diversity is a direct consequence of adaptation mechanisms. Phylogenetic analyses have divided this species complex into four distinct phylotypes correlating mostly with strains' geographical origin. This thesis focuses on using phylotype II strains as an experimental model due to their adaptation to specific hosts: **Moko** strains pathogenic to banana, '**Brown rot**' strains adapted to potatoes and emergent pathological variant **NPB** strains. Our main research topic is the understanding of **host adaptation processes**. In order to tackle this problematic we sequenced about ten genomes as a starting point of (i) a taxonomic revision of the species complex (ii) a comparative genomic analysis and (iii) an *in planta* transcriptomic analysis. Together, these complementary approaches allow a more **systemic** view of this organism's genetic and phenotypic complexity.

Remerciements

Je souhaite en premier lieu dédier ce manuscrit à mes parents, qui m'ont encouragé à poursuivre cette thèse, puis m'ont soutenu tout du long.

Ces travaux sont le résultat de plus de 3 années effectuées au Cirad (Pôle de Protection des Plantes) à la Réunion, entrecoupées de séjours au département de pathologies des plantes à l'Université du Wisconsin-Madison. A ce titre, je suis reconnaissant envers Philippe Prior, mon directeur de thèse, pour m'avoir procuré tant d'opportunités et de libertés dans la conduite de mes thématiques de recherche ; et Caitilyn Allen pour m'avoir accueilli dans son laboratoire et ainsi permis de poursuivre mes hypothèses. Je remercie également Gilles Cellier, mon encadrant du côté de l'Anses, compagnon de bureau et guide gastronomique.

Je remercie les membres de mon comité de thèse, Stéphane Poussier, Isabelle Robène, Bruno Hostachy, Frank Bertolla et Claudine Médigue pour leur soutien et leurs conseils avisés, ainsi que les membres de mon jury Sylvain Brisse, Stéphane Genin, Charles Manceau et Valérie Verdier pour avoir accepté de juger mes travaux.

Je tiens aussi à remercier nos collaborateurs du Genoscope : David Roche, David Vallenet et Valérie Barbe.

Du côté du 3P, merci à Jean-Jacques Cheron, Sylvain Lebon et Jean Michel Baptiste pour leur aide durant les nombreux essais en chambres climatiques. Je remercie également Pierre Lefeuvre pour son introduction sur R et Python. Du côté de Madison, merci aussi aux thésard(e)s d'UW Jonathan Jacobs, Tiffany Lowe, Beth Dalsing, Tuan Minh Tran et Alejandra Huerta pour leur soutien, leurs conseils et leurs projets de collaborations ; et à tous mes colocataires d'ICH et particulièrement Kyle, Leif, Luke et Lucy pour m'avoir fait découvrir les multiples facettes du Wisconsin.

Sommaire

INTRODUCTION	1
CONTEXTE	1
QUESTIONS DE RECHERCHE	4
CHAPITRE 1 – SYNTHÈSE BIBLIOGRAPHIQUE	7
1. RALSTONIA SOLANACEARUM, UNE ESPÈCE COMPLEXE	7
A. GENERALITES	7
B. TAXONOMIE	8
C. SPECTRE D'HÔTE, PATHOTYPES ET ECOTYPES.	11
2. PATHOGENÈSE	15
A. MECANISMES GÉNÉRAUX DU FLETRISSEMENT BACTÉRIEN	16
TACTISME ET RECONNAISSANCE DE L'HÔTE	16
FLAGELLES, PILI ET FIMBRIÆ	16
ENZYMES DE DÉGRADATION DE LA PAROI CELLULAIRE (CWDE)	17
SYSTÈME DE SÉCRETION TYPE III (T3SS) ET EFFECTEURS DE TYPE III (T3E)	18
EFFLUX	19
VOIES DE RÉGULATION	19
EXOPOLYSACCHARIDE (EPS)	21
B. SPECTRE D'HÔTE, ADAPTATION ET MECANISMES DE SPÉCIATION	21
3. GÉNOMIQUE COMPARATIVE	24
A. LE SÉQUENÇAGE HAUT DÉBIT	24
B. GÉNOMIQUE ET BIOINFORMATIQUE	29
C. GÉNOMIQUE ET SPECTRE D'HÔTE CHEZ LES PHYTOPATHOGENES	34
4. TRANSCRIPTION ET TRANSCRIPTOMIQUE	35
A. LES MÉTHODES D'ÉTUDE DE L'EXPRESSION GÉNÉRIQUE	35
AMPLIFICATION PAR PCR	37
HYBRIDATION	37
SÉQUENÇAGE	39
B. TRANSCRIPTOMIQUE ET SPECTRE D'HÔTE CHEZ LES PHYTOPATHOGENES	42
CHAPITRE 2 – RÉVISION TAXONOMIQUE DU COMPLEXE D'ESPÈCES	44
CHAPITRE 3 – GÉNOMIQUE COMPARATIVE CHEZ LES ECOTYPES DU PHYLOTYPE II	65
CHAPITRE 3 – ADDENDUM DÉVELOPPEMENT D'UN OUTIL DIAGNOSTIC	82
CHAPITRE 4 – TRANSCRIPTOMIQUE <i>IN PLANTA</i>	87
CONCLUSION GÉNÉRALE ET PERSPECTIVES	106
REFERENCES BIBLIOGRAPHIQUES	118

Liste des figures et tableaux

Figure 1 – Classification par phylotype et diversité de spectre d’hôte. _____	7
Figure 2 – Correspondance entre les classifications par race, biovar et division. _____	9
Figure 3 – Arbre phylogénétique réalisé à partir gène de l’endoglucanase (<i>egl</i>). _____	10
Figure 4 – Arbre phylogénétique du RSSC réalisé à partir d’une puce pangénomique. _____	10
Figure 5 – Pourriture brune chez la pomme de terre provoquée par les souches brown rot IIB-1. _____	11
Figure 6 – Flétrissement du bananier par les souches Moko. _____	12
Figure 7 – Symptômes tardifs de l’infection sur Anthurium (cv. Fire) par une souche IIB-4 NPB. _____	14
Figure 8 – Principales voies de régulations de la virulence chez <i>R. solanacearum</i> . _____	20
Figure 9 – Caractéristiques des plateformes de séquençage SGS et TGS. _____	26
Figure 10 – Poster décrivant l’intégralité des librairies disponibles sur plateforme Illumina. _____	27
Figure 11 – Construction d’un graphe dans le cadre d’une approche de type « OLC ». _____	30
Figure 12 – Construction d’un graphe dans le cadre d’une approche de type « K-mer ». _____	31
Figure 13 – Propriétés des méthodes de prédictions d’orthologues les plus répandues. _____	33
Figure 14 – Principe d’une puce à ADN basée sur des ADNc. _____	38
Figure 15 – Diversité nucléotidique des informations obtenues par séquençage dans le cadre d’analyses transcriptionnelles. _____	39
Figure 16 – Principe du SAGE. _____	40
Figure 17 – Evolution des méthodes de classification taxonomique au cours du temps. _____	44
Figure 18 – Méthodologie utilisée pour produire un arbre à partir de données protéomiques. _____	45
Figure 19 – Matrices de distances génomiques produites avec les algorithmes ANI, MUMi et GGDC. _____	63
Figure 20 – Matrice de présence/absence des facteurs de virulence au sein du RSSC réalisée à partir des génomes séquencés. _____	65
Figure 21 – Phénotypage (inoculation par le pseudotrunc) des mutants de délétion du T3E RipAA. _____	79
Figure 22 – Exemple d’un polymorphisme associé à la variation du spectre d’hôte (HAP) chez le gène d’un effecteur de type III (<i>ripAN</i>). _____	80
Figure 23 – La technologie ArrayTube (Alere Technologies). _____	82
Figure 24 – Pucés sur lame de verre. _____	84
Figure 25 – Concept du projet transcriptomique. _____	88
Figure 26 – Concept du Minimal Wilt Gene Set. _____	115
Table 1 – Liste d’effecteurs dont le rôle <i>in planta</i> est caractérisé. _____	22

Liste des figures associées aux publications :

Chapitre 2

<i>Figure 1 - Denitrification in RSSC.</i>	51
<i>Figure 2 - Phylogenetic network derived from the distance matrices generated using all three methods and the SplitsTree4 software.</i>	54
<i>Figure 3 - Majority-rule consensus tree based on the presence/absence of a consensus MALDI-TOF peak list obtained using MrBayes software.</i>	55

Chapitre 3

<i>Figure 1 - Phylogenetic tree of the R. solanacearum species complex inferred from whole genome comparisons.</i>	70
<i>Figure 2 - Host-range of sequenced R. solanacearum strains.</i>	71
<i>Figure 3 - Comparative diagram of the specific gene content in representative groups of RSSC.</i>	71

Chapitre 4

<i>Figure 1 - Differential expression across all tested comparison.</i>	94
<i>Figure 2 - Venn diagram of differentially expressed genes in planta compared to rich medium.</i>	96

Liste des annexes

Annexe 1 – Hydroxycinnamic acid degradation, a broadly conserved trait, protects Ralstonia solanacearum from chemical plant defenses and contributes to root colonization and virulence. Tiffany M. Lowe, Florent Ailloud, and Caitilyn Allen. MPMI. 2015

Liste des abréviations

ADNc ADN codant	MWGS Minimal Wilt Gene Set
ADNr ADN ribosomique	NGS Next Generation Sequencing
AHL AcylHomoserine Lactone	NO Nitrous Oxyde
AnGST Analyzer of Gene and Species Tree	NPB Not Pathogenic to Banana
ANI Average Nucleotidic Identity	NS-SNP Non-Synonymous Polymorphism
ARNm ARN messagers	OLC Overlap Layout Consensus
BAC Bacterial Artificial Chromosome	OMA Orthologous Matrix
BDB Blood Disease Bacterium	PAC P1 Artificial Chromosome
BLAST Basic Local Alignment Search Tool	PCA Principal Component Analysis
BMM Boucher's Minimal Medium	PCR Polymerase Chain Reaction
CCD Charge-Coupled Devices	PG PolyGalacturonase
CDS CoDing Sequence	qPCR quantitative PCR
CFU Colony Forming Unit	RFLP Random Fragment Length Polymorphism
CGH Comparative Genome Hybridization	RND Resistance Nodulation and cell Division
CPG Casamino acid-Peptone-Glucose	RSSC <i>Ralstonia solanacearum</i> Species Complex
CWDE Cell Wall Degrading Enzyme	SAGE Serial Analysis of Gene Expression
DD Differential Display	SGS Second Generation Sequencing
DDH DNA-DNA Hybridization	SMS Single Molecule Sequencing
DGE Differential Gene Expression	T1SS Type One Secretion System
DI Disease Index	T2SS Type Two Secretion System
EPS ExoPolySaccharide	T3E Type Three Effector
EST Expressed Sequenced Tags	T3SS Type Three Secretion System
FC Fold Change	TGS Third Generation Sequencing
FISH Fluorescence In Situ Hybridization	WGS Whole Genome Shotgun
GFP Green Fluorescent Protein	YAC Yeast Artificial Chromosome
GGDC Genome to Genome Distance Calculator	
HAP Host Associated Polymorphism	
HCA HydroxyCinnamic Acid	
HDF HrpB Dependent Factor	
HGP Human Genome Project	
HGT Horizontal Gene Transfer	
HTH Helix Turn HeJix	
INDEL INsertion DELetion	
IVET In Vivo Expression Technology	
Kb KiloBase	
MALDI-TOF Matrix-Assisted Laser Desorption/Ionization Time Of Flight	
MATE Multidrug And Toxic compound Extrusion	
MDR MultiDrug Resistant	
MLST Multi Locus Sequence Typing	
MPSS Massively Parallel Signature Sequencing	
MRCA Most Recent Common Ancestor	
MUMi Maximum Unique Matches index	

Introduction

Contexte

Les cultures maraîchères et fruitières occupent une place importante dans l'agriculture mondiale et certaines représentent des enjeux stratégiques aussi bien pour l'industrie agroalimentaire que pour l'économie de subsistance (tomate, pomme de terre, banane...). Ces cultures sont cependant mises régulièrement en échec par des maladies causées par un vaste panel de microorganismes (bactéries, champignons, oomycètes, virus, phytoplasmes et viroïdes) et d'insectes nuisibles. Historiquement, ces maladies étaient considérées comme endémiques de différentes régions du globe. La globalisation des échanges et les changements climatiques remettent cependant en question le statut épidémiologique de ces agents pathogènes et favorisent notamment leur distribution rapide (émergence ou réémergence) entre zones tropicales et tempérées.

En tant qu'agents phytopathogènes, les bactéries exercent des contraintes considérables aux cultures avec un impact socio-économique majeur (Elphinstone, 2005). Elles sont capables d'infecter diverses parties de la plante comme les racines, la tige les feuilles ou les fruits et provoquent des symptômes variés sous forme de pourriture, chancre, taches, tumeurs ou flétrissement pouvant entraîner la mort de la plante. Comme dans les domaines médicaux et vétérinaires, la lutte contre les bactéries nuisibles est rendue difficile par leur grande diversité et leur capacité d'adaptation aux hôtes ou aux conditions environnementales. Afin de pouvoir comprendre, limiter et prédire les processus d'adaptation, il est donc impératif d'obtenir des connaissances précises sur la relation entre la diversité bactérienne, les compartiments et niches écologiques occupées et pour ce qui va nous concerner dans ces travaux, les hôtes infectés.

Compte tenu de ses caractéristiques biologiques (qui seront évoquées dans la partie bibliographique), le flétrissement bactérien (Hayward, 1991) fournit un ensemble de modèles d'études académiques de choix, que l'on s'oriente vers l'étude

des interactions plantes bactéries sur des petits génomes modèles (Arabette, Medicago) ou sur des modèles à résonance plus agronomique (Solanées en général). Cet organisme est très étudié dans le monde¹. Pour ce qui nous concerne, les critères retenus pour élever ces pathosystèmes au rang de modèles ont été (i) l'émergence régulière de variant du pouvoir pathogène et (ii) l'existence de souches adaptées soit à des environnements originaux pour une bactériose tropicale (le cas des souches pomme de terre dites « froides »), soit à un hôte particulier comme les souches à l'épidémiologie clairement inféodée aux bananiers (plantain et Cavendish) et ornementales natives (Héliconia). L'ensemble fournissant un faisceau de ressources biologiques originales que nous avons exploité pour aller vers des questions de recherche innovantes sur la spéciation bactérienne, en relation avec la gamme d'hôte.

R. solanacearum est une bactérie phytopathogène rhizosphérique distribuée mondialement et qui infecte les racines et colonise le tissu vasculaire des plantes (xylème). Contrairement à la plupart des pathogènes de végétaux, cette bactérie a comme particularité une gamme d'hôtes exceptionnellement large. Par ailleurs, *R. solanacearum* est une espèce bactérienne très diverse d'un point de vu génétique et phénotypique comme l'atteste sa présence dans des environnements très différents. Cette diversité génétique a été souvent associée à la capacité naturelle de *R. solanacearum* pour la transformation bactérienne qui permet d'expliquer la structure mosaïque du génome (Boucher et al., 1992). De même, l'architecture du génome en deux réplicons est sans doute à associer à l'évolution rapide de ce génome, chaque réplicon ayant un rôle important dans la gestion des modifications génomiques: flux de gènes, 'hotspots' de recombinaisons (Lefeuvre et al., 2013). Une même souche pourra ainsi infecter plusieurs hôtes différents, tandis qu'une autre présentera un spectre d'hôte plus restreint et ne sera alors capable d'infecter qu'une seule espèce végétale. La forte capacité d'adaptation chez *R. solanacearum* est attestée par la diversité géographique, génétique et phénotypique des souches

¹ 44 publications concernant *R. solanacearum* par an depuis 1975 (1745 publications dont 29 reviews). Les contributeurs les plus importants sont l'Université du Wisconsin (8%) et Inra (8%) puis l'Université de Géorgie (4%).

(Cellier and Prior, 2010). Assez curieusement, aucune souche de *R. solanacearum* n'a été décrite à ce jour comme étant pathogène de l'ensemble des hôtes, ce qui laisse à penser que les mécanismes d'adaptations (et de « trade-off ») ont exercé leurs contraintes assez tôt dans le passé évolutif de l'organisme.

A plusieurs reprises, des associations entre des phénotypes et génotypes précis ont été démontrées. Quelques hôtes ne sont sensibles qu'à des souches assignées à une poignée de lignées phylogénétique parmi plus d'une centaine de lignées connues. Au contraire, certaines plantes peuvent flétrir en conditions tropicales suite à l'infection par des souches génétiquement très diverses, mais très peu de lignées conservent néanmoins leur pouvoir pathogène en conditions tempérées. Des phénomènes d'émergence de variants pathogènes ont également été observés. Des souches isolées dans une zone géographique restreinte suite à une épidémie présentaient ainsi une gamme d'hôte qui n'avait encore jamais été observée dans la lignée phylogénétique auxquelles elles ont été assignées (Wicker et al., 2005 ; Wicker et al., 2007). Cette interrelation entre génotype et pathotype peut ainsi être mise à profit comme modèle d'étude des interactions hôte-pathogène et, plus spécifiquement, pour aborder l'adaptation à l'hôte. En effet, la structure phylogénétique de ces populations, qualifiées d'écotypes, permet de supposer qu'elles sont le fruit d'un processus d'adaptation résultant d'une isolation géographique ou de pressions de sélection causées par les hôtes végétaux environnants. Le concept d'écotype n'est pas reconnu officiellement en taxonomie bactérienne (Brenner et al., 2000a) mais il est néanmoins utile pour décrire la diversité écologique au sein d'une espèce bactérienne (Schloter et al., 2000; Cohan, 2001). Un écotype est défini comme une population bactérienne occupant la même niche écologique et dont la diversité génétique est limitée par des pressions de sélection périodique et/ou la dérive génétique (Cohan and Perry, 2007). Un écotype conserve généralement la totalité (ou presque) des caractéristiques phénotypiques et écologiques de l'espèce et exploite une niche écologique peu différente en comparaison de son espèce (Konstantinidis et al., 2006).

En dehors de la sphère académique, décrire et appréhender le fonctionnement du spectre d'hôte a aussi une importance en agronomie. En effet, les épidémies de flétrissement bactérien sont souvent dramatiques aux plans économiques et sociaux. Une meilleure compréhension de l'adaptation à l'hôte permet de développer des outils diagnostics plus efficaces et d'améliorer ainsi le suivi épidémiologique. De la même façon, une meilleure maîtrise de la diversité génotypique et phénotypique autorise une meilleure gestion des schémas de sélection en amélioration variétale pour la résistance à la maladie, fournissant des critères de sélection plus efficaces.

Grâce à la disponibilité accrue des plateformes de séquençage et aux réductions des coûts des méthodes à haut débit, il est aujourd'hui possible d'obtenir rapidement une grande quantité de données génomiques. Le séquençage massif de souches de *R. solanacearum* représente ainsi une opportunité inédite d'identifier de nouveaux mécanismes d'adaptation à l'hôte et de revisiter les schémas précédemment établis.

Questions de recherche

L'axe thématique central de ma thèse concerne les divergences existantes à différents niveaux au sein du complexe d'espèces *R. solanacearum* (RSSC) (Gillings and Fahy, 1994; Genin and Denny, 2012) et la problématique principale consiste à identifier ces divergences, puis à évaluer leur implication dans la variation du pouvoir pathogène et du spectre d'hôte, voire l'émergence de variants du pouvoir pathogène.

D'un point de vu académique, le chapitre bibliographique introduira le contexte scientifique dans lequel la problématique présentée a été initialement abordée et décrira l'état des connaissances chez l'organisme étudié, puis présentera également les méthodologies retenues afin de tenter de répondre aux hypothèses de travail posées. D'un point de vu technique, le développement de la thématique repose essentiellement sur l'analyse de nombreux génomes de *R. solanacearum*, soit séquencés en préambule de ces travaux, soit issus de bases de données publiques.

Plus particulièrement, nous avons concentrés nos questions de recherche en fonction d'un modèle expérimentale constitué des écotypes du phylotype II (Moko, 'Brown rot' et NPB).

En fonction des objectifs que nous nous sommes assignés, l'interprétation et la restitution des données expérimentales progressivement accumulées sont structurées en trois chapitres complémentaires :

- **Hypothèse 1 : Les différences génomiques, protéomiques et phénotypiques au sein du RSSC justifient sa réorganisation taxonomique.** Ce projet se propose de diviser le complexe d'espèces actuel en trois espèces distinctes à l'aide de (i) distances génomiques obtenues par comparaison de génomes complets ; (ii) profils protéomiques obtenus par MALDI-TOF ; (iii) traits phénotypiques partiellement déduits de la littérature. L'intérêt de cette nouvelle classification au niveau académique et agronomique est discuté.

PUBLICATION 1 : ADDITIONAL EVIDENCE FOR THE DIVISION OF THE PLANT PATHOGEN *RALSTONIA SOLANACEARUM* INTO THREE SPECIES INFERRED THROUGH GENOMIC AND PROTEOMIC ANALYSES. PRIOR P., AILLOUD F. ET AL. *SYSTEMATIC AND APPLIED MICROBIOLOGY*. (SUBMITTED)

- **Hypothèse 2 : La comparaison génomique des écotypes Moko, Brown rot et NPB du phylotype II permet d'identifier des mécanismes associés à la spécificité d'hôte.** Les génomes de chacun des écotypes ont été comparés pair-à-pair ainsi qu'à l'ensemble du RSSC. De nouveaux core-génome, pan-génome et core-effectome ont été décrits. L'analyse du contenu en gènes, des polymorphismes et des transferts horizontaux ont permis l'identification de plusieurs gènes candidats. Les résultats d'une analyse fonctionnelle préliminaire par mutagenèse de certains gènes candidats sont discutés.

PUBLICATION 2 : COMPARATIVE GENOMIC ANALYSIS OF *RALSTONIA SOLANACEARUM* REVEALS CANDIDATE GENES FOR HOST SPECIFICITY. AILLOUD F. ET AL. *BMC GENOMICS*. (ACCEPTED)

- **Hypothèse 3 : La comparaison transcriptomique *in vitro* et *in planta* des écotypes Moko et NPB permet d'identifier des profils d'expression souche- ou plante-spécifiques associés à la spécificité d'hôte.** Les transcriptomes de chaque écotype ont été obtenus par RNAseq en milieu riche, minimum, sur tomate et sur bananier (Moko uniquement) ou melon (NPB uniquement). Plusieurs méthodes de comptage ont permis de déterminer l'expression différentielle entre des paires de conditions pertinentes d'un point de vue biologique. Le génome respectif de chaque souche a été utilisé comme référence afin d'inférer le rôle des gènes différentiellement exprimés dans la spécificité d'hôte.

PUBLICATION 3 : *IN PLANTA* COMPARATIVE TRANSCRIPTOMICS OF HOST-ADAPTED STRAINS OF *RALSTONIA SOLANACEARUM*. AILLOUD F. ET AL. (NOT SUBMITTED)

Enfin, les différents résultats obtenus sont regroupés sous la forme d'une conclusion générale et leurs contributions vis-à-vis de la problématique de l'adaptation à l'hôte sont discutées. Les perspectives envisagées concernant la poursuite de ces travaux ainsi que des hypothèses de travail ou des modèles expérimentaux alternatifs sont décrits.

Cette thèse s'intègre dans la thématique « Génomique et épidémiologie des agents pathogènes émergents » de l'UMR PVBMT. L'ensemble de mes travaux a bénéficié du financement du CIRAD, de l'Anses et de l'Université du Wisconsin.

- **Hypothèse 3 : La comparaison transcriptomique *in vitro* et *in planta* des écotypes Moko et NPB permet d'identifier des profils d'expression souche- ou plante-spécifiques associés à la spécificité d'hôte.** Les transcriptomes de chaque écotype ont été obtenus par RNAseq en milieu riche, minimum, sur tomate et sur bananier (Moko uniquement) ou melon (NPB uniquement). Plusieurs méthodes de comptage ont permis de déterminer l'expression différentielle entre des paires de conditions pertinentes d'un point de vue biologique. Le génome respectif de chaque souche a été utilisé comme référence afin d'inférer le rôle des gènes différentiellement exprimés dans la spécificité d'hôte.

PUBLICATION 3 : *IN PLANTA* COMPARATIVE TRANSCRIPTOMICS OF HOST-ADAPTED STRAINS OF *RALSTONIA SOLANACEARUM*. AILLOUD F. ET AL. (NOT SUBMITTED)

Enfin, les différents résultats obtenus sont regroupés sous la forme d'une conclusion générale et leurs contributions vis-à-vis de la problématique de l'adaptation à l'hôte sont discutées. Les perspectives envisagées concernant la poursuite de ces travaux ainsi que des hypothèses de travail ou des modèles expérimentaux alternatifs sont décrits.

Cette thèse s'intègre dans la thématique « Génomique et épidémiologie des agents pathogènes émergents » de l'UMR PVBMT. L'ensemble de mes travaux a bénéficié du financement du CIRAD, de l'Anses et de l'Université du Wisconsin.

Chapitre 1 – Synthèse bibliographique

1. *Ralstonia solanacearum*, une espèce complexe

a. Généralités

R. solanacearum est un bacille à Gram négatif, mobile, aérobic strict de la classe des bêta-protéobactéries. Cette bactérie est l'agent du flétrissement bactérien. Elle peut être isolée dans le sol, l'eau et les plantes. *R. solanacearum* est un phytopathogène vasculaire d'origine tellurique et rhizosphérique capable de transformation naturelle. Son génome, d'une taille totale d'environ 5,5 Mb est organisé en deux réplicons : un chromosome (~3,5 Mb) et un mégaplasmide (~2 Mb). *R. solanacearum* présente une forte diversité génétique associée à une forte diversité phénotypique qui se traduit par un très large spectre d'hôte comprenant environ 250 espèces végétales réparties dans 50 familles botaniques (Hayward, 1991) (**Figure 1**). Ces caractéristiques en font un modèle intéressant pour la compréhension des mécanismes moléculaires mis en jeu dans l'adaptation à l'hôte et pour l'identification des processus génétiques responsables des événements de spéciation chez les bactéries. La diversité de cette bactérie est également attestée par la vaste origine géographique des isolats. Bien que principalement isolé dans les régions tropicales et subtropicales, la répartition mondiale de *R. solanacearum* est en constante évolution, ce qui nécessite un effort permanent dans le développement de méthodes de diagnostic innovantes pour assurer une veille épidémiologique efficace. Les souches responsables de la pourriture brune ('Brown rot') chez la pomme de terre (French et al., 1977; Martin et al., 1982) sont notamment considérées comme agent potentiel de bioterrorisme aux États-Unis (Lambert, 2002) et les souches provoquant la maladie de Moko du bananier sont des organismes de quarantaine en Europe². La pomme de terre et le bananier ainsi que d'autres espèces comme la tomate, le tabac ou encore le géranium font partie des hôtes de *R. solanacearum* ayant une importance agronomique et économique

² www.eppo.org/QUARANTINE/listA2.htm

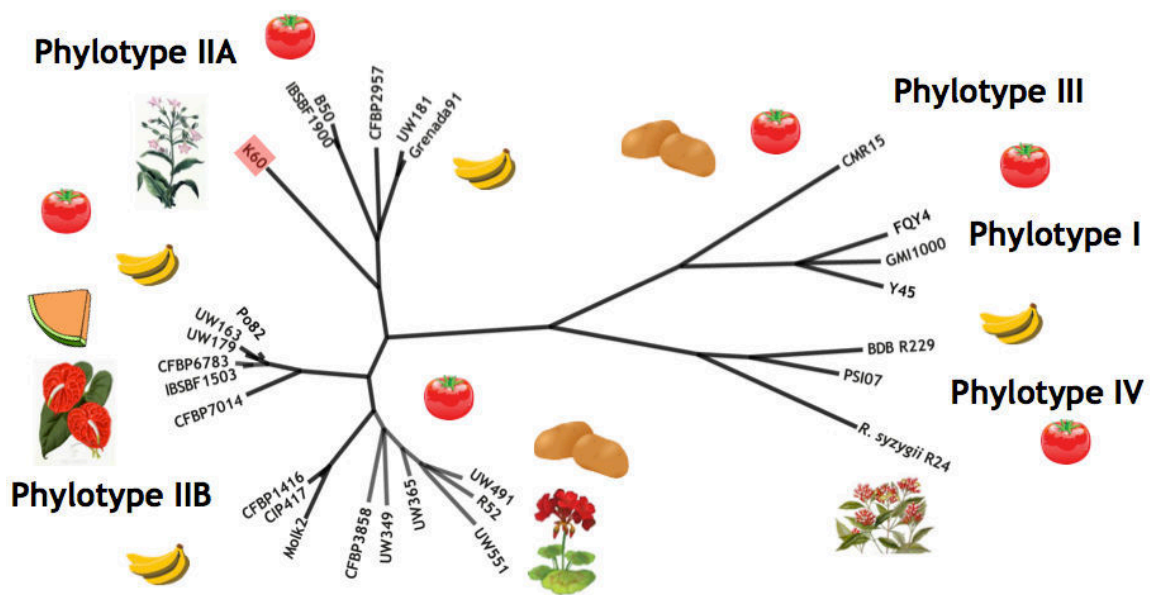


Figure 1 - Classification par phylotype et diversité de spectre d'hôte. L'arbre phylogénétique a été créé en fonction de distances génomiques calculées à partir de génomes complets. La souche type de l'espèce *R. solanacearum* est encadrée. Hôtes illustrés : banane, tomate, pomme de terre, melon, anthurium, tabac, géranium, giroflier.

majeure (Hayward, 1991). En effet les pertes pour l'industrie agroalimentaire dues au flétrissement bactérien chez la pomme de terre sont estimées à près d'un milliard US\$ par an (Grimault et al., 1994b). La plupart des hôtes sensibles sont des espèces vivrières stratégiques en matière de développement durable, ceci à l'échelle locale dans les pays en voie de développement où une seule de ces cultures peut représenter la majorité de la production et l'alimentation de base de la région. C'est le cas de nombreuses brèdes, terme qui regroupe un ensemble très divers de feuilles comestibles de nombreuses plantes (*Solanum nigrum*, *S. americanum*...) qui sont cuisinées avant d'être consommées, la cuisson inhibant sans doute les toxines de certains de ces végétaux.

L'étude de cet organisme présente donc aussi bien un intérêt dans le champ de la recherche fondamentale, que dans le monde agronomique. *R. solanacearum* est ainsi considéré par la communauté internationale comme la seconde plus importante bactérie phytopathogène (après *Pseudomonas syringae*) (Mansfield et al., 2012).

b. Taxonomie

Le flétrissement bactérien causé par *R. solanacearum* a été décrit pour la première fois par Erwin Fink Smith à la fin du 19ème siècle (Smith, 1896). Originellement dénommé *Bacillus solanacearum*, l'agent du flétrissement bactérien a ensuite été successivement connu sous le nom de *Pseudomonas solanacearum* (Smith, 1896) puis *Burkholderia solanacearum* (Yabuuchi et al., 1992) avant d'être finalement renommé *R. solanacearum* en 1995 après une caractérisation phylogénétique par l'intermédiaire entre autres du séquençage de l'ADNr 16S (Yabuuchi et al., 1995).

La classification intraspécifique de *R. solanacearum* a elle aussi fait débat bien avant que l'espèce ne soit intégrée dans le genre *Ralstonia*. Historiquement, un système de classification à trois races, basé sur le spectre d'hôte, a été utilisé pour décrire la diversité de *R. solanacearum* (Buddenhagen et al., 1962). La race 1 ubiquiste infecte de nombreuses solanées comme la tomate, le tabac ou le poivron, quelques mauvaises herbes (plantes adventices) ainsi que les bananiers

ornementaux diploïdes (*Musa Ensete ventricosum*). La race 2, au spectre d'hôte restreint, provoque le flétrissement chez les bananiers triploïdes (*Musa spp.*) et les *Heliconia*. Enfin, la race 3 est adaptée aux températures tempérées et est pathogène de la pomme de terre et de la tomate, mais apparaît généralement peu virulente chez les autres solanées. Plus tard, les races 4 et 5 furent décrites comme infectant respectivement le giroflier aux Philippines (Aragaki and Quinon, 1965) et le mûrier en Chine (He et al., 1983).

Indépendamment du système de race, celui du biovar fut développé sur la base de la capacité de la bactérie à métaboliser ou à oxyder divers disaccharides et hexoses alcool (Hayward, 1964). Quatre biovars ont été décrit en premier lieu par Hayward, puis deux nouveaux biovars ont ensuite été rapportés, les biovars 5 (He et al., 1983) et 2T (syn. N2) (Hayward et al., 1992). Aucun lien direct ne peut être réellement établi entre les deux systèmes de classification, suggérant qu'ils ne reflètent pas vraiment l'histoire évolutive de *R. solanacearum* (**Figure 2**).

L'évolution des techniques de typage et la transition des méthodes phénotypiques vers des méthodes génotypiques entraîneront un remaniement important de la classification chez *R. solanacearum*. L'analyse de la diversité génétique par le polymorphisme de longueur des fragments de restriction (RFLP) de plusieurs *loci* a ainsi permis de distinguer 40 groupes distincts (Cook et al., 1989; Cook and Sequeira, 1994). Parmi ces groupes deux sous-divisions corrélant avec l'origine géographique des souches peuvent être clairement identifiées : la Division 1 « Asiaticum » regroupe des isolats provenant d'Asie et d'Australie, la Division 2 « Americanum » héberge des souches originaires d'Amérique Centrale et d'Amérique du Sud. En étendant l'analyse à une autre collection de souches, une 3^{ème} Division « Africanum » regroupant uniquement des souches d'origine Africaines fut aussi identifiée (Poussier et al., 2000). Enfin, l'analyse de la séquence de l'ADNr 16S confirmera ces résultats et identifiera une sous-division de la division 2 comprenant des isolats d'Indonésie ainsi que des organismes proches : l'agent de maladie du sang du bananier (BDB) et *Pseudomonas syzygii*, l'agent de la maladie de Sumatra du giroflier (Li et al., 1993; Taghavi et al., 1996). Phylogénétiquement, ces deux espèces

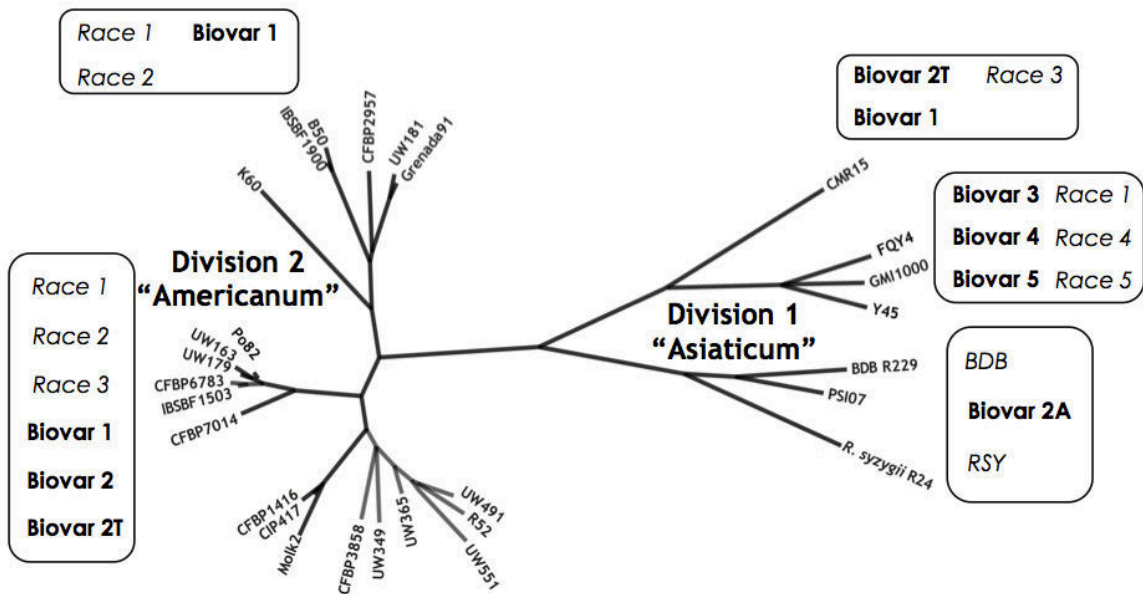


Figure 2 - Correspondance entre les classifications par race, biovar et division. L'arbre phylogénétique a été créé en fonction de distances génomiques calculées à partir de génomes complets.

sont très proches des souches de *R. solanacearum* ce qui justifie l'utilisation du terme de « complexe d'espèces » pour décrire cet organisme (Gillings and Fahy, 1994). Cette relation entre groupe génétique et région d'origine des souches suggère fortement que l'isolement géographique a joué un rôle important dans l'évolution de *R. solanacearum*.

Ce système de division fut uniformisé grâce à une analyse phylogénétique supplémentaire basée sur la région intergénique des ADNr 16S et 23S ainsi que sur les séquences des gènes *hrpB* et *egl* (Fegan and Prior, 2005). Le complexe d'espèce est alors divisé en quatre groupes appelés phylotypes (I, II, III et IV). Les phylotypes correspondant respectivement aux divisions Asiaticum, Americanum, Africanum et Indonésia. Chaque phylotype est également sous-divisé en lignées (les feuilles de l'arbre phylogénétique) dénommées sequevar, en fonction des variations (moins de 1%) de la séquence nucléotidique partielle du gène de l'endoglucanase (*egl*) (**Figure 3**). Sur cette base, deux branches majeures peuvent être distinguées au sein du phylotype II : IIA et IIB. Le phylotype IV regroupe des souches appartenant aux espèces *R. solanacearum*, BDB et *R. syzygii*. Dans un contexte de diagnostic, il est possible de déterminer le phylotype d'une souche à l'aide d'une PCR multiplexe (Fegan and Prior, 2005).

La classification en quatre groupes a été ensuite revisitée par des méthodes de plus en plus moderne utilisant toujours plus d'informations génétiques, notamment un schéma de Typage Moléculaire Multi Locus (MLST) (Castillo and Greenberg, 2007) et une puce à ADN d'Hybridation Génomique Comparative (CGH) (Guidot et al., 2007) développée suite au séquençage d'une souche du phylotype I, GMI1000 (Salanoubat et al., 2002) (**Figure 4**). Finalement, le séquençage d'autres souches dans chaque phylotype mais aussi de souches de BDB et de *R. syzygii* a permis d'établir un schéma global de la variabilité génétique entre les différentes classes du complexe d'espèces (Remenant et al., 2010; Remenant et al., 2011). Les distances génomiques établies entre ces génomes suggèrent alors que les phylotypes I+III, II et IV pourraient former trois espèces à part entière.

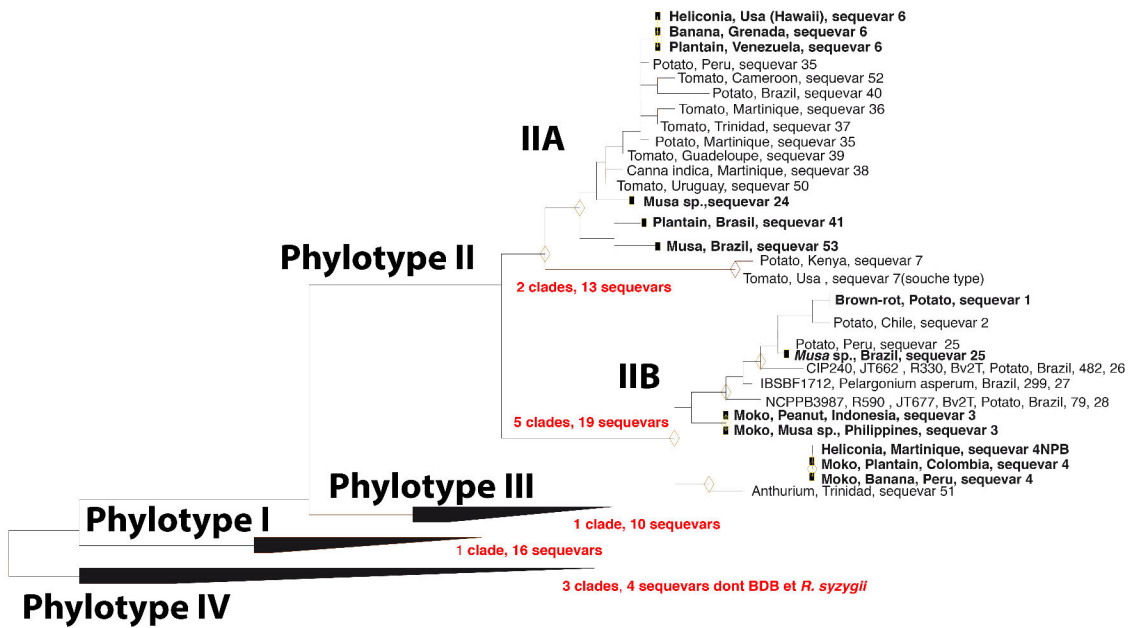


Figure 3 - Arbre phylogénétique réalisé à partir gène de l'endoglucanase (*egl*) (Fegan et Prior 2005).

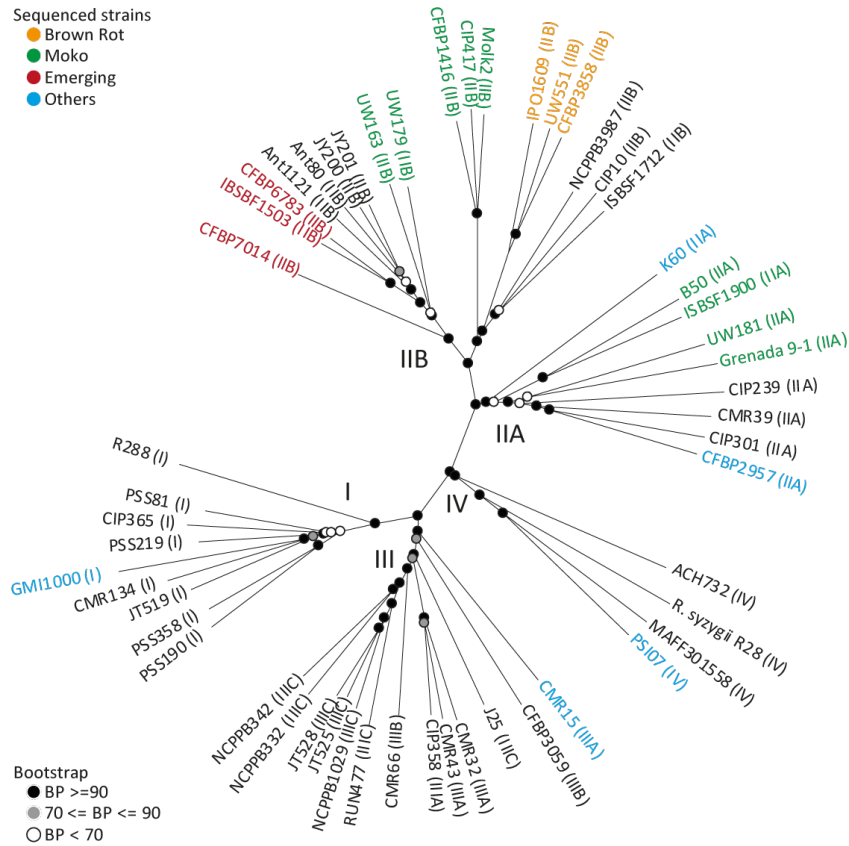


Figure 4 - Arbre phylogénétique du RSSC réalisé à partir d'une puce pangénomique (Cellier et al. 2012)

La connaissance exhaustive et l'assignation réfléchie des espèces et des sous-espèces n'ont pas uniquement un intérêt dans le monde de la taxonomie. La classification du complexe d'espèces chez *R. solanacearum* est particulièrement importante si l'on considère l'incidence pratique et décisive dans des domaines stratégiques clés tels que la sélection variétale en agronomie, la mise au point de méthodes de diagnostic et le maintien en quarantaine en épidémiologie.

c. Spectre d'hôte, pathotypes et écotypes.

La gamme d'hôte a toujours été un des principaux centres d'intérêt dans l'étude de *R. solanacearum* et cette propriété a tout d'abord été utilisée comme un outil de classification avec le système des races. Notons que contrairement à la classification des champignons, la classification en race chez *R. solanacearum* n'a aucune valeur de taxon. Aussi, au fur et à mesure que la connaissance du complexe d'espèce a progressé, il a été démontré que la classification en race était devenue obsolète et surtout, n'était plus un modèle approprié pour décrire l'histoire évolutive du complexe d'espèces *R. solanacearum*. Il a cependant été montré que certaines lignées phylogénétiques (sequevar) pouvaient être associées à des pathotypes distincts (Cellier and Prior, 2010).

De nombreuses lignées chez *R. solanacearum* sont capables d'infecter la pomme de terre et entraînent ce qui est communément appelé la pourriture brune ('Brown rot' des anglo-saxons). En dehors des signes caractéristiques de flétrissement et de jaunissement du feuillage, en cas d'infection, des symptômes peuvent également apparaître sur le tubercule et se présentent sous la forme d'exsudats bactériens caractéristiques et d'une décoloration brune à l'origine du nom de la maladie (**Figure 5**). Ces symptômes peuvent apparaître avant le flétrissement et provoquer l'arrêt de la croissance de la plante. Les tubercules infectés participent ensuite à l'infection de nouveaux hôtes et à la dispersion de la souche à une grande échelle. Parmi les lignées responsables de la pourriture brune, les souches de phylotype IIB-1 (historiquement Race 3 biovar 2) ont la particularité d'être adaptées au froid (<24°C) et donc de produire du flétrissement quand les autres lignées ne sont pas



Figure 5 - Pourriture brune chez la pomme de terre provoquée par les souches brown rot IIB-1. (Source : G. Cellier, P. Champoiseau)

capables de survivre (French, 1986; Cellier and Prior, 2010). Ces conditions sont aussi bien retrouvées en région tempérée en Europe et aux Etats-Unis qu'en région tropicale dans les hauts-plateaux andins et africains (Mahbou Somo Toukam et al., 2009). Cette capacité n'est cependant pas reliée à une adaptation directe au froid dans la mesure où il a été montré que cette tolérance est dépendante de la présence de la bactérie *in planta*. En effet, aucune différence significative au niveau de la survie et de la croissance à différentes températures n'a été observée *in vitro* entre les souches tempérées IIB-1 et les autres souches tropicales (Milling et al., 2009). La résistance au froid de cette lignée serait donc complètement induite par l'interaction entre la bactérie et le tubercule. A basse température, la pomme de terre constitue une niche écologique spécifique de la lignée IIB-1 qui peut alors être considérée comme un écotype de *R. solanacearum*.

Le bananier Cavendish (banane fruit, AAA, AAB) et le plantain (banane à cuire, BBB, ABB) sont également des hôtes qui sont infectés par différentes lignées chez *R. solanacearum* (**Figure 6**). Cependant, les symptômes et le mode de transmission peuvent varier en fonction de différents paramètres. Les souches capables d'infecter le genre *Musa* provoquent la maladie de Moko. Cette dernière est la plus connue et la plus dévastatrice des formes de flétrissement bactérien du bananier. Historiquement, la première référence de cette maladie précède la description d'E. F. Smith et eut lieu au milieu du 19^{ème} siècle lors des voyages de Schomburgk en Guyane Britannique (Martyn, 1931). Attribué plus tard, le nom de « Moko » vient d'une espèce de plantain (Bluggoe) particulièrement touchée à Trinidad en 1890 (Kelman, 1953). Initialement deux groupes de souches Moko ont été définis par Sequeira (Buddenhagen, 1961) : les souches « B » (pour 'Banana' en Amérique Centrale) qui s'établissent dans les tissus vasculaires et sont directement transmissibles par contact des racines ou avec des outils contaminés lors du bouturage ou de la récolte ; les souches « SFR » ('Small, Fluidal, Round') en Amérique du Sud et « A » (pour 'Amazon') qui se transmettent rapidement via insectes par l'intermédiaire des bourgeons avant d'envahir les vaisseaux des plantes infectées. De par leur mode de transmission, le 1^{er} groupe (B) entraîne souvent une



Figure 6 - Flétrissement du bananier par les souches Moko. Ici, la souche RUN62 a été inoculée soit par injection dans le pseudo-tronc (I) soit par inondation du substrat (C). (Source : G. Cellier)

décoloration vasculaire, un flétrissement et un jaunissement classique des feuilles tandis que les souches du 2nd groupe (SFR et A), transmises par insectes, tendent à provoquer des symptômes au niveau du fruit.

En plus de la maladie de Moko probablement introduite via des boutures de bananiers infectées, les îles des Philippines sont aussi touchées par la maladie de Bugtok sur plantain. Bien que les symptômes soient proches de ceux provoqués par les souches Moko SFR et A sur bananier, le faciès de la maladie dans cette interaction *R. solanacearum*-plantain a fait croire pendant des années à l'existence d'une autre bactérie phytopathogène. *R. solanacearum* sera formellement reconnue comme agent de la maladie de Bugtok dans les années 90 (Soguilon et al., 1995). Les symptômes consistent en un durcissement et une décoloration du fruit immature ainsi qu'un rougissement limité aux vaisseaux du rachis. La distinction faite entre les faciès de maladie Bugtok et Moko à cette époque a pour origine les différents génotypes de plantains cultivés dans les régions concernées par ces maladies, respectivement BBB (Saba et Cadaba) aux Philippines et ABB (Bluggoe) en Amérique.

La maladie du sang du bananier (BDB pour 'Blood Disease of Banana') est encore une autre manifestation du flétrissement bactérien sévissant uniquement en Indonésie et principalement sur les bananiers de génotype ABB (Pisang Kapok). Elle est décrite par Ernst Gäumann dans les années 20 et reconnue alors comme étant causée par *Pseudomonas celebensis* (Gäumann, 1921, 1923). Elle entraîne des symptômes proches de la maladie de Moko sur bananier Cavendish mais a la particularité de provoquer une décoloration rouge foncée du fruit à l'origine de son nom. Le rapprochement entre la maladie du sang et *R. solanacearum* sera fait par Eden-Green dans les années 90 (Eden-Green, 1994). Notons que les souches BDB sont inféodées au bananier et, à ce jour, on ne leur connaît aucun autre hôte sensible ou alternatif.

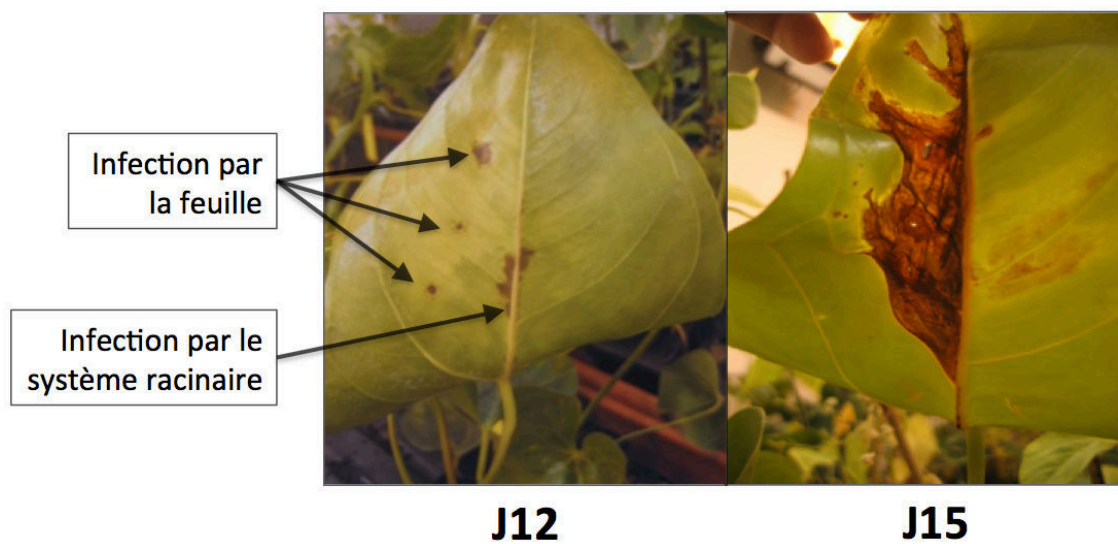
Les analyses phylogénétiques décrites ultérieurement démontrèrent rapidement que BDB appartient à une division distincte des autres maladies de Moko du

bananier, tombant dans le phylotype IV. Il a été aussi montré que les maladies Bugtok et Moko sont provoquées par des souches appartenant à la même lignée phylogénétique désignée aujourd'hui IIB-3. Les lignées IIB-4, IIA-6 et IIA-24 ont également été spécifiquement associées à la maladie de Moko (Fegan, 2005; Fegan and Prior, 2006).

Récemment, des souches initialement assignées à la lignée Moko IIB-4 ont été isolées en Martinique (Wicker et al., 2007; Wicker et al., 2009). Ces souches sont cependant des variants pathogènes qui ne présentent pas de virulence sur bananier et sont ainsi désignées NPB pour «Non Pathogène du Bananier». Ces dernières ont par ailleurs élargi leur gamme d'hôte et sont capables d'infecter des cultures de la famille des *Cucurbitaceae* ou des plantes ornementales (**Figure 7**) pour lesquelles les souches Moko (toutes lignées confondues) ne sont pas pathogènes. De plus, les souches NPB sont très agressives sur la tomate et sont capables d'infecter l'ensemble des principales sources de résistance au flétrissement utilisées par les sélectionneurs au plan international, tels qu'Hawaii 7996 (Lebeau et al., 2011).

La maladie de Sumatra correspond au flétrissement du giroflier par les souches de *R. syzygii*, espèce représentée phylogénétiquement par une lignée unique au sein du phylotype IV (Waller and Sitepu, 1975). Certaines souches de *R. solanacearum* sont capables de coloniser partiellement cet hôte, mais seule *R. syzygii* provoque un flétrissement complet. Cependant, le giroflier ne présente pas les caractéristiques d'un pathosystème expérimentale simple compte tenu de l'accomplissement très lent du cycle infectieux. En effet, les symptômes observables n'apparaissent qu'environ 200 jours après l'infection (Eden-Green and Adhi, 1986). De plus, la transmission naturelle ne se fait pas au niveau des racines mais activement, par l'intermédiaire d'un insecte foreur, *Hindola spp.*

Plusieurs écotypes ont donc été distingués au niveau du complexe d'espèces. En dehors du pouvoir pathogène, pouvant, dans une certaine mesure, être déduit par association phylogénétique, la connaissance de la gamme d'hôte de la plupart des souches isolées est très faible et se limite souvent à la plante sur laquelle la souche a



**Figure 7 - Symptômes tardifs de l'infection sur Anthurium (cv. Fire) par une souche IIB-4 NPB.
(Source : F. Ailloud)**

été isolée. De plus, la différence inévitable entre le processus infectieux naturel et les méthodes d'inoculations expérimentales peuvent conduire à surestimer la gamme d'hôte réelle d'une souche.

2. Pathogénèse

Initialement, l'infection débute au niveau du sol où *R. solanacearum* est capable de survivre de manière prolongée dans la rhizosphère de nombreuses espèces végétales (hôte et non-hôte). La bactérie colonise activement l'environnement racinaire aux niveaux des extrémités (coiffe racinaire), des zones d'émergences de racines secondaires ou bien des blessures provoquées par des insectes, des nématodes ou l'Homme (Vasse et al., 1995). Après infection de la plante, la colonisation se poursuit par l'invasion des espaces interstitielles du cortex racinaire et du parenchyme vasculaire sous forme de micro colonies. A ce stade, la bactérie peut alors pénétrer dans le cylindre vasculaire notamment grâce à la dégradation des parois cellulaires. *R. solanacearum* se multiplie ensuite exponentiellement au sein des vaisseaux de xylème et migre dans la tige jusqu'à atteindre des densités cellulaires de l'ordre de 10^{10} CFU/ mg de tissu (Grimault et al., 1994a). Finalement, la réduction de la circulation dans les tissus vasculaires est accentuée par la surproduction d'exopolysaccharides par la bactérie. Le flétrissement bactérien précédant la mort de plante hôte est donc provoqué par le blocage des vaisseaux de xylème empêchant ainsi la libre circulation du flux hydrique. Cette obstruction est le fait de la production massive d'exopolysaccharides bactériens, mais est aussi due à la réponse de la plante sensible à l'infection qui émet dans l'ensemble des vaisseaux du xylème (colonisés ou non) du matériel amorphe de type dépôt de callose (Grimault et al., 1994b). En fonction de l'interaction espèce hôte-souche bactérienne, plusieurs symptômes peuvent se développer avant l'apparition du flétrissement : retard de croissance de la plante, jaunissement des feuilles et/ou de la tige, croissance désordonnée des racines. Bien avant le début de la décomposition de la plante, *R. solanacearum* est également capable de recoloniser le sol en étant

directement relarguée dans le milieu par les racines et pouvant ainsi infecter les plantes alentours (Kelman and Sequeira, 1965; Swanson et al., 2005).

a. Mécanismes généraux du flétrissement bactérien

La versatilité du pouvoir pathogène chez *R. solanacearum* repose sur une large gamme de facteurs de virulence contrôlés par des voies de régulations en cascade.

Tactisme et reconnaissance de l'hôte

Afin de coloniser une plante hôte, *R. solanacearum* doit dans un premier temps l'identifier. Le chimiotactisme permet à la bactérie de répondre à des signaux chimiques émis par la plante. Il a été montré que la reconnaissance de différentes sources de carbones et d'exsudats racinaires (tomate et riz) varie en fonction des souches suggérant que la gamme de signaux reconnus pourrait être associée au spectre d'hôte et aux voies métaboliques disponibles. Un mutant *cheW* n'est plus capable de chimiotactisme et il présente une virulence moindre par rapport à la souche sauvage, une perte de motilité et une distribution désordonnée le long des racines (Yao and Allen, 2006). Un autre type de tactisme, l'aérotactisme, participe également à la virulence. Ce mécanisme consiste en la détection des niveaux d'oxygène dans l'environnement, afin d'optimiser la chaîne respiratoire. Un mutant *aer2* présente ainsi une déficience dans les étapes précoces de l'infection.

Flagelles, pili et fimbriae

R. solanacearum possède plusieurs types d'appendices de surface pour accomplir des tâches spécifiques à différentes étapes du cycle infectieux. La mobilité à l'extérieur de la plante, aussi appelée 'swimming motility', est assurée par un flagelle polaire (*flg*) dont la synthèse est assurée par le régulateur *flhDC*. Ce flagelle permet notamment à la bactérie de se diriger vers les racines après détection d'exsudats. Une fois dans la plante, la bactérie devient principalement non-motile. Un mutant non-motile n'est plus capable d'infecter une plante dans des conditions naturelles mais le phénotype virulent est entièrement restauré si la bactérie est directement

injectée dans le xylème suggérant que la motilité flagellaire est uniquement nécessaire dans les étapes précoces de l'infection (Tans-Kersten et al., 2001). La régulation négative de la motilité par l'intermédiaire de *flhDC* est en partie assurée par *motN*. Un mutant *motN* est hypermotile et hyperflagellé mais produit des biofilms plus fragiles (Meng et al., 2011). Ce mutant est également moins virulent, illustrant ainsi l'importance d'une régulation fine de la motilité chez *R. solanacearum*.

Ralstonia est également capable de mobilité sur une surface solide de type « twitching motility » produit par un pili de type IV composé principalement d'un seul type de piline, PilA. Un mutant *pilA* est avirulent quelle que soit la méthode d'infection et perd aussi sa compétence naturelle (Kang et al., 2002). Ces pilis jouent probablement plusieurs rôles au cours du processus infectieux : adhésion, migration au sein de l'hôte et formation de biofilms.

Enzymes de dégradation de la paroi cellulaire (CWDE)

Chez *R. solanacearum* ainsi que chez la plupart des bactéries à Gram négatif, le système de sécrétion de type II (T2SS) est utilisé pour exporter des protéines dans l'espace extracellulaire. Parmi ces protéines on peut trouver de multiples enzymes hydrolytiques qui pourraient permettre de dégrader les parois cellulaires afin d'obtenir des nutriments, mais aussi de progresser au sein de la plante hôte. Deux enzymes cellulolytiques sont secrétées : une endoglucanase (*egl*) et une exoglucanase (*cbhA*). Quatre enzymes pectinolytiques sont secrétées : une pectine méthylesterase (*pme*) dé-méthyle la pectine afin de faciliter sa dégradation par des polygalacturonases (PG), une endoPG (*pehA*) et deux exoPG (*pehB* et *pehC*). La contribution de ces enzymes à la pathogénèse semble cependant limitée. En effet, un mutant *egl* (Roberts et al., 1988), un triple mutant *pehABC* (Gonzalez and Allen, 2003) ou même un mutant pour les six CWDE (Liu et al., 2005) sont toujours capables d'entraîner du flétrissement bactérien et ne présentent qu'une faible réduction de virulence par rapport à la souche sauvage. Néanmoins, il existe très probablement d'autres exo-protéines secrétées par le T2SS encore non identifiées;

en effet un mutant T2SS est moins virulent que le mutant pour les six CWDE (Liu et al., 2005).

Système de sécrétion type III (T3SS) et effecteurs de type III (T3E)

Le système de sécrétion de type III (T3SS) est une autre structure de surface très répandue chez les bactéries pathogènes des végétaux ou des animaux, et également présente chez *R. solanacearum*, qui permet d'injecter des protéines à l'intérieur des cellules hôtes. Le T3SS est nécessaire pour établir une infection et il est recruté dans la plante ou en milieu pauvre en nutriment (Arlat et al., 1992). Il est aussi responsable de la réponse hypersensible (HR) chez les plantes non hôtes (Boucher et al., 1985). Il est codé par les îlots de gènes *hrp* ('hypersensitive response and pathogenicity') et *hrc* ('hrp conserved') comportant au total plus de 20 gènes. Les gènes *hrc* codent principalement pour des protéines conservées de la partie basale du T3SS, tandis que les gènes *hrp* codent pour les protéines plus spécifiques du pili constitué principalement par la piline HrpY. L'expression du T3SS ainsi que des T3E est contrôlée par un régulateur de type AraC, HrpB qui est lui-même activé par un système à deux composants, HrpG (Vasse et al., 2000). La régulation du T3SS est cependant complexe et implique vraisemblablement de nombreux facteurs et voies de régulation encore inconnues (Zuluaga et al., 2013). Par exemple, il a été montré qu'une protéine chaperonne HpaP est capable d'activer ou de réprimer la sécrétion de certains effecteurs spécifiques (AvrA et PopP1) et de la piline HrpY (Lohou et al., 2014).

Une des spécificités de *R. solanacearum* est son large répertoire d'effecteurs qui a notamment été décrit grâce au séquençage de la souche GMI1000 (Salanoubat et al., 2002). La recherche du motif de régulation d'HrpB (*hrp_{II}* box), similaire à la PIP-box présente chez les *Xanthomonas* (Cunnac et al., 2004), des analyses transcriptomiques (Occhialini et al., 2005) ainsi que des expériences de translocation (Cunnac et al., 2004; Tamura et al., 2005; Mukaihara et al., 2010) ont permis l'identification et la validation d'un grand nombre d'effecteurs candidats. 72 effecteurs seraient présents au total chez GMI1000 (Poueymiro and Genin, 2009)

et peuvent être aussi bien localisés sur le mégaplasme que sur le chromosome. Certains sont directement situés à côté du T3SS dont les effecteurs *popA*, *popB* et *popC*.

La plupart des effecteurs sont très peu caractérisés et leur rôle ou leur cible sont majoritairement inconnus. PopP1, PopP2, GALA7, AvrA ainsi que les T3E de la famille AWR ont été associés à des variations de la gamme d'hôte et au pouvoir pathogènes à différents niveaux (Coll and Valls, 2013).

Efflux

Chez de nombreuses bactéries pathogènes et environnementales, la tolérance à des composés toxiques passe par l'intermédiaire de pompes à efflux. Par exemple, les résistances multiples aux antibiotiques chez les agents pathogènes humains sont dues en partie à ce type de transporteur. Chez les plantes, les composés antimicrobiens sont des substances de faible poids moléculaire regroupées sous le nom de phytoalexines et sont synthétisés après exposition à des micro-organismes. Parmi les dizaines d'opérons codant pour des pompes à efflux putatives chez *R. solanacearum*, deux pompes à efflux ont été identifiées comme étant exprimées chez la plante au cours de l'infection : AcrA de la famille RND (Resistance/Nodulation/cell Division) et DinF de la famille MATE (Multidrug And Toxic compound Extrusion). Ces deux systèmes sont activés par deux régulateurs différents, RpoS et HrpB respectivement, et utilisent deux sources d'énergie distinctes, un gradient d'H⁺ et de Na⁺ respectivement. Les substrats exportés par ces transporteurs et leur rôle précis ne sont pas encore connus, des mutants pour chacune de ces pompes présentent cependant une sensibilité accrue à certaines toxines et leur virulence est atténuée (Brown et al., 2007).

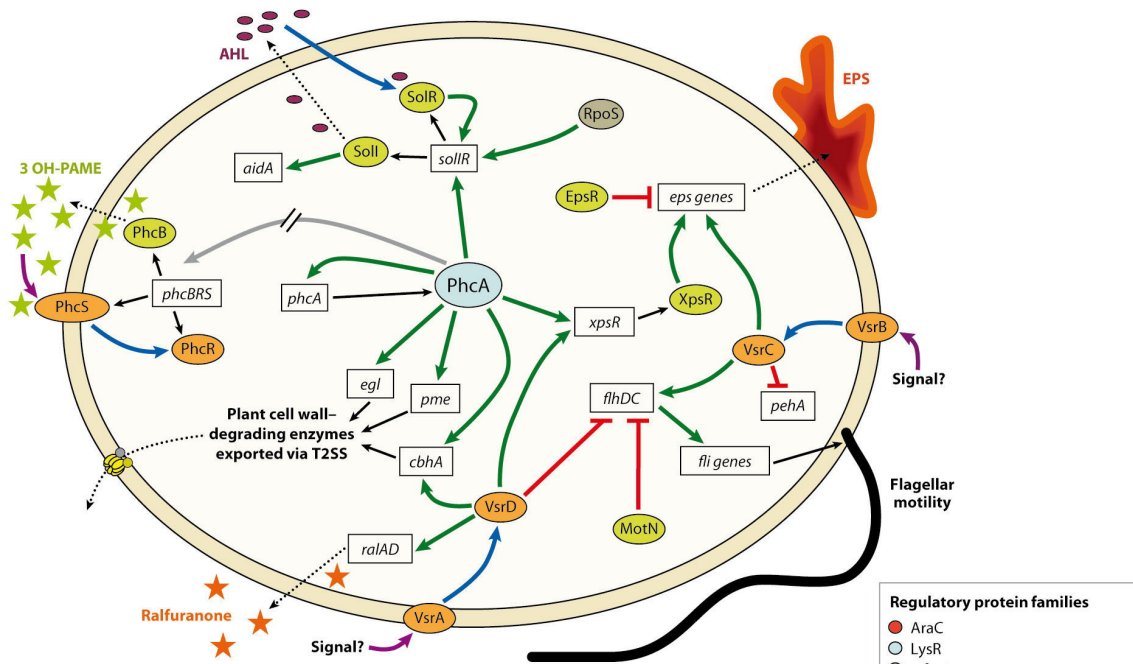
Voies de régulation

La pathogénicité chez *R. solanacearum* est contrôlée de façon distincte aux stades précoces et tardifs de l'infection en fonction de conditions environnementales telles que le contact avec les cellules de l'hôte ou la densité bactérienne.

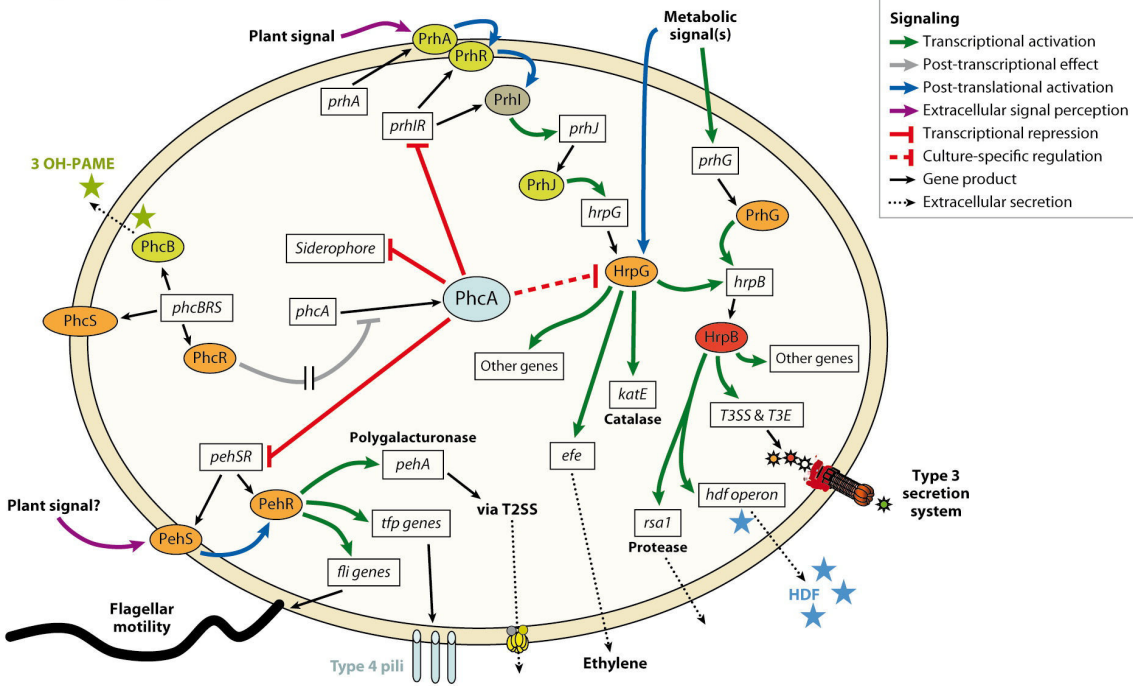
Au début de l'infection - stade précoce - le contact avec la paroi végétale est détecté par un récepteur membranaire PrhA qui transmet ce signal à une cascade de régulations composée de PrhR, PrhI, PrhJ, HrpG et enfin HrpB. HrpG contrôle deux voies de régulation indépendantes : la 1^{ère} concerne des gènes impliqués dans l'adaptation et la survie au sein de l'hôte tandis que la 2^{nde} module l'expression du T2SS, de PehC, du T3SS (genes *hrp*) via HrpB. Simultanément, HrpB est responsable de la production de l'HDF ('HrpB Dependant Factor'), un autoinducteur homologue de l'AHL (AcylHomoserine Lactone) impliqué dans la perception de la densité cellulaire ('quorum sensing'). Cette molécule pourrait également être impliquée dans la perturbation du quorum sensing d'autres bactéries (Delaspre et al., 2007). Néanmoins, le composé reconnu initialement par PrhA n'est pas encore caractérisé. En parallèle, HrpB est co-régulé par PrhG en réponse à divers signaux métaboliques (Plener et al., 2010). Plusieurs travaux suggèrent également qu'HrpG est également co-régulé par des signaux indépendants du contact avec la plante (Zuluaga et al., 2013).

Durant la colonisation - stade tardif -, le régulateur transcriptionnel central PhcA est responsable entre autres de la production de l'exopolysaccharide, d'enzymes cellulolytiques et pectinolytiques ainsi que de la répression des gènes *hrp*, de PG, de sidérophores et de la motilité par l'intermédiaire de nombreux mécanismes de régulations, notamment les systèmes à deux composants VsrA-VsrD et VsrB-VsrC (**Figure 8**). L'activité de PhcA est auto-induite via un 1^{er} système de quorum sensing qui repose sur l'accumulation de 3-OH PAME (3-hydroxypalmitic acid methyl ester) produit par PhcB au sein du xylème et reconnu par un système à deux composants PhcS-PhcR (Flavier et al., 1997). La répression des gènes *hrp* au stade tardif de l'infection et à haute densité cellulaire est cependant remise en cause par des travaux récents (Jacobs et al., 2012; Monteiro et al., 2012; Zuluaga et al., 2013). A son tour, PhcA est responsable d'un 2nd système de quorum sensing via la production d'AHL qui active d'autres facteurs de virulence sous la dépendance d'un système à deux composants SolI-SolR.

a Positively regulated traits



b Negatively regulated traits



Genin S, Denny TP. 2012. Annu. Rev. Phytopathol. 50:67–89

Figure 8 – Principales voies de régulations de la virulence chez *R. solanacearum* (Genin et Denny, 2012).

Les facteurs de virulence de *R. solanacearum* sont donc très souvent successivement ou simultanément contrôlés par plusieurs cascades de régulations intégrant à leur tour de nombreux signaux endogènes ou exogènes.

Exopolysaccharide (EPS)

R. solanacearum sécrète principalement un hétéropolysaccharide acide, de haut poids moléculaire et riche en azote composé de trimères de N-acétylé galactosamine, d'acide 2-Nacetyl-2-deoxy-L-galacturonic, et de 2-N-acetyl-4-N-(3-hydroxybutanoyl)-2,4,6 trideoxy-D-glucose, communément appelé EPS I (Orgambide et al., 1991) et dont la synthèse et l'exportation sont assurées par l'opéron *eps* de 18kb (Huang and Schell, 1995). L'EPS I représente 90% des exopolysaccharides produits par *R. solanacearum* ; 85% sont sécrétés, le reste étant associé à la membrane sous forme de capsule (McGarvey et al., 1999).

Des mutants déficients pour la production d'EPS sont quasiment avirulents mais sont toujours capables de croissance *in planta* (Denny and Baek, 1991; Kao et al., 1992). Il est généralement accepté que l'EPS joue un rôle dans l'occlusion des vaisseaux de xylème durant les stades tardifs de l'infection, et il est aussi suggéré qu'il pourrait servir à protéger la bactérie du système de défense de la plante en masquant les structures de surface qui pourraient être reconnues. Cependant, l'EPS I pourrait lui-même être la source d'une réponse immunitaire de la part de l'hôte, notamment dans le cas de plantes résistantes (Milling et al., 2011).

b. Spectre d'hôte, adaptation et mécanismes de spéciation

De par sa diversité génétique et phénotypique, *R. solanacearum* est souvent considérée comme un modèle pour l'identification et la compréhension des mécanismes d'adaptation à l'hôte.

Jusqu'à présent, un nombre limité de souches a fait l'objet de la majorité des études. La souche GMI1000 du phylotype I est de loin la plus étudiée car son génome a notamment été le premier à être complètement séquencé et

manuellement annoté (Salanoubat et al., 2002). La tolérance au froid des souches 'Brown rot' a été étudiée chez la souche UW551, dont le génome est disponible. Enfin, la souche K60^T pathogène de la tomate du phylotype IIA-7 est aussi souvent utilisée comme modèle compte tenu de son statut de souche type pour l'espèce (Remenant et al., 2012).

Certaines espèces végétales sont aussi préférentiellement utilisées pour les analyses *in planta*. *Arabidopsis thaliana* et *Medicago truncatula* sont des modèles très largement utilisés en pathologie végétale. Ces espèces modèles sont facilement et rapidement cultivables et sont adaptées aux études de microscopies. Leurs mécanismes de défense sont fortement caractérisés au niveau moléculaire. Leurs « petits » génomes sont séquencés et la simplicité de leur transformation a permis l'établissement d'une vaste collection de lignées mutantes. La tomate (*Solanum lycopersicum*) est également un hôte très répandu possédant des caractéristiques proches d'*Arabidopsis* et de *Medicago* mais qui a l'avantage d'être une espèce pertinente pour le monde agricole.

Comparé au nombre d'hôtes pouvant être infectés par *R. solanacearum*, seulement une poignée d'hôtes est en réalité étudiée. Jusqu'à présent, les variations du spectre d'hôte ont été très majoritairement associées à des effecteurs de type III (**Tableau 1**). PopP1, un membre de la famille YopJ/AvrRxv, restreint la virulence sur *Pétunia* chez la souche GMI1000 (Lavie et al., 2002). PopP2, également membre de la famille YopJ, est reconnu par la protéine de résistance RSS1-R et provoque une réponse immunitaire chez *A. thaliana* (Deslandes et al., 2003). PopP2 est une cystéine protéase qui interagit avec RSS1-R au sein du noyau et pourrait stabiliser cette protéine en empêchant son ubiquitinylation, évitant ainsi sa dégradation par le protéasome (Bernoux et al., 2008; Tasset et al., 2010). GALA7 permet la virulence sur *M. truncatula* chez GMI1000 (Angot et al., 2006). Les GALA sont une famille de protéines contenant des domaines F-box homologues des plantes qui pourraient également intervenir dans le détournement de la voie ubiquitine/protéasome. AvrA, en association avec PopP1, limitent la virulence de GMI1000 sur plusieurs espèces de tabac, *Nicotina tabacum*, *benthamiana* et *glutinosa* (Poueymiro et al.,

Table 1 – Liste d'effecteurs dont le rôle *in planta* est caractérisé (adapte de Coll et Valls, 2013).

Protein name	Family	Predicted domains	Role <i>in planta</i>	Hosts tested	Mode of action	References
AvrA	-	-	Avirulence Promotes growth	Nicotiana spp., Tomato	-	Carney and Denny (1990); Robertson <i>et al.</i> (2004); Turner <i>et al.</i> (2009); Macho <i>et al.</i> (2009)
AvrPphF	HopF2	-	Promotes growth	Tomato, Eggplant, Bean	-	Macho <i>et al.</i> (2010)
Awr1	AWR	-	Promotes growth	Tomato, Eggplant	-	Solé <i>et al.</i> (2012)
Awr2		-	Avirulence Promotes growth	Nicotiana spp., Tomato, Eggplant, Arabidopsis	-	
Awr4		-	Restricts growth	Arabidopsis	-	
Awr5		-	Avirulence Restricts growth	Nicotiana spp., Tomato, Eggplant, Arabidopsis	-	
Gala1	GALA	LRR repeats, F-box	-	-	Interaction with SKP1-like proteins	Angot <i>et al.</i> (2006)
Gala3			-	-		
Gala5			-	-		
Gala6			-	-		
Gala7			Host specificity factor	Medicago truncatula		
PopA	-	Harpin	-	Nicotiana	Formation of ion-conducting pores	Racapé <i>et al.</i> (2005)
-	PopC	LRR	Promotes growth	Eggplant, Bean	-	Macho <i>et al.</i> (2010)
PopP1	YopJ	Ser/Thr acetyl-transferase, functional NLS	Avirulence	Petunia	-	Lavie <i>et al.</i> (2002)
PopP2			Avirulence Promotes growth	Arabidopsis, Tomato, Eggplant, Bean	Nuclear relocalization of RRS1-R and RD19, binds RRS1-R	Deslandes <i>et al.</i> (2002); Deslandes <i>et al.</i> (2003); Bernoux <i>et al.</i> (2008); Macho <i>et al.</i> (2010)
Skwp4	SKWP	Heat/armadillo-related repeats	Promotes growth	Eggplant	-	Macho <i>et al.</i> (2010)
Rip19	AvrBs3	Central repeats	Promotes growth	Eggplant	-	
Rip34	HopD1	-	Promotes growth	Tomato, Eggplant, Bean	-	
Rip39	HopAV1	Coiled-coil	Promotes growth	Eggplant	-	
Rip64	HopR1	-	Promotes growth	Eggplant	-	
Rip3	-	Ankyrin repeat	Promotes growth	Eggplant	-	
Rip55	-	-	Promotes growth	Eggplant	-	
Rip23	-	-	Promotes growth	Eggplant	-	

2009). Certains membres de la famille AWR, caractérisée par le tripeptide alanine-tryptophane-arginine, restreignent la virulence chez *A. thaliana* (Sole et al., 2012).

Une seule protéine n'étant pas un T3E a pour l'instant été associée à la gamme d'hôte. Rsa 1, une aspartate protéase identifiée chez une souche du phylotype IV non pathogène du piment, entraîne la perte de la virulence sur piment une fois introduite chez une souche du phylotype I naturellement pathogène du piment (Jeong et al., 2011). Rsa1 est aussi requise pour la virulence sur pomme de terre mais ne confère cependant pas le phénotype sur la souche du phylotype I. Bien que n'étant pas un T3E, Rsa1 est régulée par HrpB via une *hrpII* box et serait potentiellement sécrétée par l'intermédiaire du T2SS.

La résistance au froid des souches responsables de la pourriture brune chez la pomme de terre peut également être considérée comme une forme d'adaptation à l'hôte compte tenu que ce phénotype est directement lié à l'interaction hôte-bactérie (Milling et al., 2009). En effet, *in vitro*, la survie et la croissance des souches R3bv2 n'est pas supérieure à celle des autres *R. solanacearum* à basse température. Les mécanismes moléculaires gouvernant cette adaptation n'ont cependant pas été découverts.

Hormis GALA7, la totalité des gènes identifiés pouvant influencer la gamme d'hôte sont des facteurs d'avirulence. L'adaptation à l'hôte et en particulier l'évolution du spectre d'hôte reste donc une question ouverte. L'ancêtre commun de *R. solanacearum* était potentiellement un organisme polyphagique/ubiquitaire pouvant exister au sein d'une large gamme d'espèces végétales à l'état latent ou pathogène à divers degrés. La gamme d'hôte aurait ensuite évolué via une combinaison de gain de facteurs entraînant l'avirulence chez certaines espèces végétales, mais fournissant probablement un avantage compétitif chez d'autres espèces ou conditions environnementales et, dans certains cas, un gain de facteurs de virulence élargissant le spectre d'hôte et entraînant l'émergence de nouveaux pathotypes.

3. Génomique comparative

La génomique est une discipline consistant en l'analyse de la structure et de la fonction des génomes qui a donné naissance à de nombreux champs de recherche, dont la génomique comparative. La génomique comparative consiste à comparer les génomes de différents organismes afin de mieux appréhender leurs fonctions biologiques à travers leurs points communs et leurs différences, et permet aussi d'étudier l'histoire évolutive de ces organismes.

a. Le séquençage haut débit

Le champ de la génomique a vu le jour grâce au développement de méthodes de séquençage de l'ADN et notamment la méthode de terminaison de chaîne mise au point par Frederick Sanger en 1977. La fragmentation aléatoire de l'ADN total, le clonage puis le séquençage individuel et aléatoire de chaque clone par la méthode de Sanger permet ainsi le séquençage de génomes entiers ('whole genome shotgun' ou WGS) (Staden, 1979). La qualité du génome peut être décrite par la profondeur de couverture qui se réfère au ratio du nombre total de bases séquencées sur la longueur totale du génome. La restriction de la longueur maximale de séquençage/lecture (< 1kb) signifiait cependant que la durée, le coût et le personnel requis étaient proportionnels à la taille du génome et à la couverture souhaitée.

La stratégie 'shotgun' a été adaptée au cours du temps en fonction des avancées technologiques. De nombreuses variantes se basant sur une cartographie préalable du génome ont été créées, entre autres, afin de limiter la composante aléatoire de cette stratégie: l'ADN total est dans un premier temps divisé en fragments de grandes tailles et clonés dans des vecteurs acceptant des inserts de grande taille (YAC, BAC, PAC, cosmid, fosmid); ces clones sont organisés hiérarchiquement à l'aide de méthodes de cartographie génétique (carte de restriction, fluorescence *in situ*) puis ensuite re-fragmentés, sous-clonés et finalement séquencés. Ces variantes sont tombées en désuétude dans les années 90 avec l'augmentation des capacités de

calcul des ordinateurs permettant un assemblage rapide des génomes à partir de large quantité de données de séquences aléatoires.

Le séquençage 'paired-end' représenta une évolution majeure dans le domaine du séquençage de génome. Cette méthode consiste à séquencer les extrémités de fragments de tailles prédéfinies (de 2kb jusqu'à plus de 100kb) : deux lectures, ne se chevauchant généralement pas, sont alors obtenues pour chaque clone (« mate-pair »). La taille des inserts étant contrôlée, la distance séparant chaque paire est connue et facilite alors le processus d'assemblage (Edwards et al., 1991; Roach et al., 1995).

Enfin, la méthode de terminaison de chaîne Sanger sera aussi fortement optimisée au cours du temps. La technique originale utilise la radioactivité pour détecter l'ADN et nécessite par conséquent 4 réactions et 4 migrations distinctes pour différencier les 4 nucléotides, l'utilisation de marqueurs fluorescents de couleurs différentes permet d'unifier le processus en une seule réaction puis migration (Smith et al., 1985; Smith et al., 1986; Prober et al., 1987). La migration sur gel est aussi remplacée par une électrophorèse en capillaire ce qui accélère grandement la séparation des fragments marqués (Cohen et al., 1988; Drossman et al., 1990; Luckey et al., 1990). Ces avancements ouvriront la voie au développement du séquençage automatique à haut débit, le premier séquenceur automatique utilisant des terminateurs fluorescents et des capillaires sera commercialisé par Applied Biosystems (Perkin-Elmer) en 1995.

Malgré ces évolutions technologiques majeures, le séquençage du génome humain débuté en 1990 (« Human Genome Project » ou HGP) prendra tout de même 13 années et nécessitera plusieurs milliards de dollars avant d'être complété par des dizaines de laboratoires répartis dans plus d'une dizaine de pays (Collins et al., 2003). L'achèvement du HGP contribua à l'augmentation de la demande en séquençage et au développement de méthodes plus rapides, précises et moins coûteuses dites de « nouvelle/seconde génération » (SGS). Associées à la bulle financière du secteur des biotechnologies, plusieurs solutions ont été

commercialisées et mises en concurrence, mais partagent cependant une philosophie commune dérivée du concept de Sanger et du WGS. La première plateforme a été déployée en 2005 par 454 Life Sciences/Roche (GS20) suivi notamment par Solexa/Illumina (Genome Analyzer) en 2006, Applied Biosystems en 2007 (SOLiD) et Ion Torrent en 2010 (PGM). Ces technologies reposent toujours sur la fragmentation de l'ADN total mais elles remplacent l'étape de clonage par une amplification par PCR des fragments en un point donné (fixe) à l'aide d'adaptateurs (librairies). La PCR peut être réalisée sur billes (émulsion PCR (Dressman et al., 2003); 454, SOLiD et PGM) ou sur une surface solide (bridge PCR (Adessi et al., 2000) ; Solexa) et a pour but de produire en parallèle des régions de haute densité à partir de chaque fragment (clusters) afin de faciliter la détection de la réaction de séquençage. La méthode utilisée pour déterminer chaque base diffère entre chaque plateforme, mais consiste toujours en des cycles d'incorporation des nucléotides (séquençage par synthèse ou par ligation), de lavage et de détection du signal (**Figure 9**).

Malgré les évolutions technologiques de chaque nouvelle version, les plateformes de 2nde génération ont souvent plusieurs défauts récurrents. L'amplification par PCR peut introduire des erreurs et entraîner des biais de couverture dans des régions répétées ou à taux extrêmes de GC. Les lectures sont en général plus courtes que par séquençage Sanger et ne permettent pas de résoudre efficacement la plupart des régions répétées. Sur les séquenceurs Illumina, cette différence de taille est due à la désynchronisation progressive des cycles entre les fragments au sein d'un même cluster, le nombre de cycle possible avant que ce phénomène n'entraîne une baisse trop importante de la qualité du signal détermine ainsi la longueur maximale des lectures. La grande quantité de données générée nécessite également des capacités de calculs accrues afin de réaliser un assemblage *de novo*. Par conséquent, les génomes séquencés par ces plateformes sont souvent incomplets ('Draft genome') et ne permettent pas certaines analyses approfondies comme, par exemple, la caractérisation des variations structurelles. Un recours au séquençage Sanger est souvent nécessaire pour parfaire l'assemblage et obtenir un

Platform	Library/ template preparation	NGS chemistry	Read length (bases)	Run time (days)	Gb per run	Machine cost (US\$)	Pros	Cons	Biological applications	Refs
Roche/454's GS FLX Titanium	Frag, MP/ emPCR	PS	330*	0.35	0.45	500,000	Longer reads improve mapping in repetitive regions; fast run times	High reagent cost; high error rates in homo- polymer repeats	Bacterial and insect genome <i>de novo</i> assemblies; medium scale (<3 Mb) exome capture; 16S in metagenomics	D. Muzny, pers. comm.
Illumina/ Solexa's GA ₁	Frag, MP/ solid-phase	RTs	75 or 100	4 [‡] , 9 [§]	18 [‡] , 35 [§]	540,000	Currently the most widely used platform in the field	Low multiplexing capability of samples	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Life/APG's SOLiD 3	Frag, MP/ emPCR	Cleavable probe SBL	50	7 [‡] , 14 [§]	30 [‡] , 50 [§]	595,000	Two-base encoding provides inherent error correction	Long run times	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Polonator G.007	MP only/ emPCR	Non- cleavable probe SBL	26	5 [§]	12 [§]	170,000	Least expensive platform; open source to adapt alternative NGS chemistries	Users are required to maintain and quality control reagents; shortest NGS read lengths	Bacterial genome resequencing for variant discovery	J. Edwards, pers. comm.
Helicos BioSciences HeliScope	Frag, MP/ single molecule	RTs	32*	8 [‡]	37 [‡]	999,000	Non-bias representation of templates for genome and seq-based applications	High error rates compared with other reversible terminator chemistries	Seq-based methods	91
Pacific Biosciences (target release: 2010)	Frag only/ single molecule	Real-time	964*	N/A	N/A	N/A	Has the greatest potential for reads exceeding 1 kb	Highest error rates compared with other NGS chemistries	Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks	S. Turner, pers. comm.

*Average read-lengths. [‡]Fragment run. [§]Mate-pair run. Frag, fragment; GA, Genome Analyzer; GS, Genome Sequencer; MP, mate-pair; N/A, not available; NGS, next-generation sequencing; PS, pyrosequencing; RT, reversible terminator; SBL, sequencing by ligation; SOLiD, support oligonucleotide ligation detection.

Figure 9 – Caractéristiques des plateformes de séquençage SGS et TGS (Metzker et al. 2010).

génomome complet. Malgré ces limitations, le SGS représente une évolution majeure par rapport au Sanger, un génome humain draft peut être obtenu en une journée pour environ 1000\$ (sous certaines conditions). Débuté en 2008, un consortium visant à séquencer plus de 1000 génomes humains a atteint la 2^{nde} phase de son projet en 2012 avec 1092 génomes disponibles (Abecasis et al., 2010). Cette génération a également participé à l'explosion des « omiques » dû aux champs d'applications possibles en modifiant le protocole de préparation des librairies. En effet, l'expression des gènes, la structure des ARN, la méthylation, la structure de la chromatine, les modifications des histones ou les interactions protéines-ADN peuvent être étudiés grâce aux plateformes de 2^{nde} génération (**Figure 10**).

Déjà commercialisés, les séquenceurs dits de 3^{ème} génération (TGS) se développent rapidement. Cette génération est caractérisée par la capacité de séquencer individuellement chaque molécule d'ADN ('single molecule sequencing' ou SMS) permettant de se passer d'une étape d'amplification. La détection se fait de manière continue éliminant les problèmes de désynchronisation. Deux types de technologies sont actuellement disponibles. Lancé en 2011, le PacBio (Pacific Biosciences) repose toujours sur le séquençage par synthèse et l'incorporation de nucléotides fluorescents détectés par une caméra. La polymérase est fixée au fond d'un puits de 10 nanomètres de diamètre reposant sur une lame de verre. Fixé sur le phosphate, chaque fluorophore est libéré après incorporation du nucléotide, puis diffuse à la surface du puits. Le très faible diamètre du puits empêche le laser de le traverser totalement et permet d'illuminer uniquement les fluorophores présents au niveau de la polymérase. Le temps d'incorporation étant de trois fois l'ordre du temps de diffusion, un nucléotide incorporé peut être différencié d'un nucléotide non incorporé en fonction de l'intensité du signal. Cette technologie est capable de produire des lectures jusqu'à 10kb, mais elle présente cependant un taux d'erreur généralement supérieur et un débit inférieur aux méthodes de 2^{nde} génération. L'observation en temps réel de l'incorporation permet également d'obtenir des informations supplémentaires sur le nucléotide incorporé comme son état de méthylation par exemple.

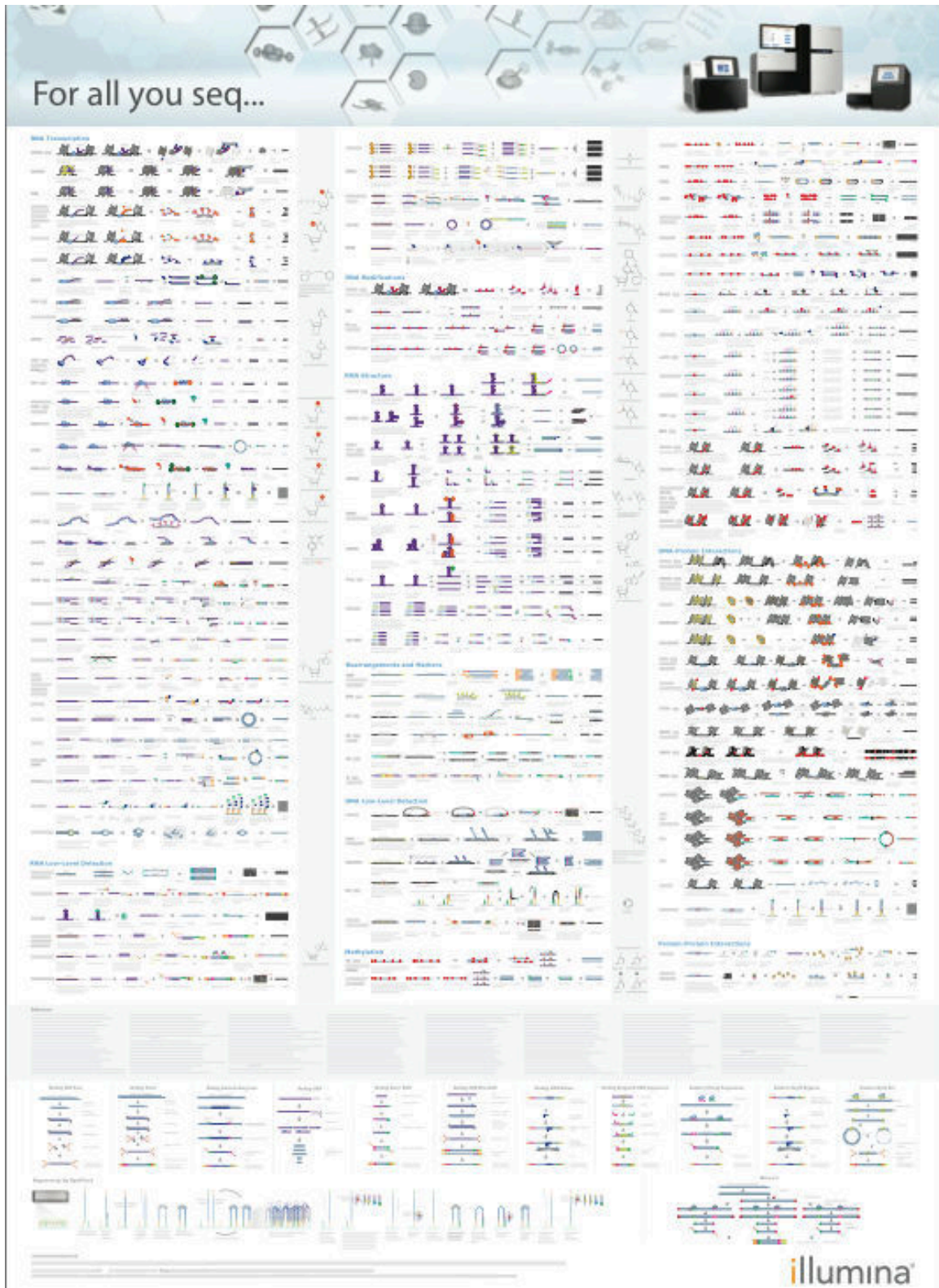


Figure 10 - Miniature d'un poster décrivant l'intégralité des librairies disponibles sur plateforme Illumina. (Source : www.illumina.com/LibraryPrepMethods)

Disponible de façon limitée, la technologie du Nanopore (Oxford Nanopore Technologies) représente une approche du séquençage radicalement différente. Intégré à une double membrane lipidique synthétique, un pore modifié d'alpha-hémolysine est couplé à une exonucléase et un senseur à base de cyclodextrine. Un courant est établi à travers la membrane à l'aide d'un gradient ionique. L'exonucléase dégrade l'ADN et chaque nucléotide passant par le nanopore est individuellement détecté par le changement spécifique de la force ionique.

La 3^{ème} génération n'est donc pas encore arrivée à maturité et plusieurs autres alternatives sont envisagées comme la visualisation directe des molécules d'ADN par microscopie électronique. La longueur des lectures, supérieure au SGS et au Sanger, ainsi que le taux d'erreur inhérent au séquençage de molécules individuelles seront aussi des paramètres à prendre en compte dans l'adaptation des modèles mathématiques permettant de traiter ces données. L'évolution des technologies de séquençage doit donc être accompagnée de l'évolution des méthodes bioinformatiques afin d'améliorer la qualité des génomes produits.

b. Génomique et bioinformatique

Contrairement à l'assemblage basé sur un génome de référence, l'assemblage *de novo* n'utilise que les données des lectures pour produire un génome et représente donc un processus complexe (Miller et al., 2010).

L'évolution des algorithmes d'assemblage *de novo* a naturellement suivi celle des techniques de séquençage, mais aussi celle des capacités informatiques. En effet, de la première à la dernière génération de séquenceurs, l'assemblage a toujours reposé sur le même principe de base : combiner des lectures en un minimum de séquences (contigs) en se basant sur les régions chevauchantes, puis organiser ces séquences en super-séquences (scaffolds) à partir des lectures en pair. La taille des lectures, les taux et les types d'erreurs, le débit de données sont autant de paramètres ayant influencés le développement des logiciels d'assemblage. La 2nde génération a entraîné des changements majeurs dans la philosophie de l'assemblage dus à une forte diminution de la taille des lectures associée à une forte augmentation de la quantité de données par rapport au Sanger (El-Metwally et al., 2013).

Les algorithmes utilisés par les assembleurs sont, pour la grande majorité, basés sur la théorie des graphes. Un graphe est un ensemble de nœuds connectés par des liens pouvant être orientés au sein duquel plusieurs liens forment un chemin. L'assemblage consiste en la réduction progressive par concaténation du nombre de nœuds formant un chemin. Trouver le chemin optimal est un problème de type NP-dur (i.e. intrinsèquement difficile) qui ne peut pas être résolu de manière efficace et nécessite donc l'utilisation d'algorithmes heuristiques et d'approximations.

Les répétitions représentent un problème redondant dans le processus d'assemblage (Zhi et al., 2006). Si la partie chevauchante de deux lectures appartient à une région répétée, il n'est pas correct d'assembler les lectures car il n'est pas possible de conclure avec certitude que ces lectures proviennent de la même région. L'augmentation de la taille des lectures et de la couverture accroît les chances d'obtenir une lecture couvrant la totalité de la répétition et possédant des séquences

uniques à ses extrémités. Les paires de lecture recouvrant la région répétée ou possédant un membre au sein des répétitions sont également utilisées afin de résoudre cette problématique, car elles permettent d'estimer la taille de cette région.

La précision des séquenceurs est également une composante importante et peut varier fortement en fonction des différentes machines (Harismendy et al., 2009). En effet, deux régions génomiques possédant moins de différences que le taux d'erreurs du séquenceur ne peuvent pas être facilement différenciées du point de vue de l'assembleur.

Les approches de types OLC ('Overlap-Layout-Consensus') sont surtout utilisées pour les données de lectures longues et de précisions moyennes (Sanger et 454) (Myers, 1995). Cette méthode organise les données en graphe contenant un nœud pour chaque lecture et un lien pour chaque chevauchement. Ces chevauchements ('Overlap') sont prédéterminés par des alignements pairs-à-pairs nécessitant beaucoup de puissance de calcul. Les lectures sont assemblées en identifiant des chemins dits Hamiltoniens ne passant qu'une seule fois par chaque nœud en passant par un minimum de lien ('Layout'). Enfin les contigs sont formés en combinant les régions chevauchantes de chaque lecture ('Consensus'). La taille des régions chevauchantes joue un rôle important dans l'efficacité de l'algorithme : une faible taille va augmenter le nombre de liens mais augmenter le nombre de faux-positifs, tandis qu'une grande taille va augmenter le nombre d'impasses dans le graphe avec des lectures ne contenant pas de chevauchement (**Figure 11**).

Les approches de type gloutonne ('greedy') ont été principalement utilisées chez les premiers assembleurs de lectures courtes (phrap, TIGR) et consistent à toujours joindre les lectures possédant le plus grand chevauchement (Pop and Salzberg, 2008). Cette approche est donc très locale et ne prend pas en compte les relations entre les lectures en pair, elle peut de ce fait se retrouver bloquée au niveau de maxima locaux et ne pas produire un assemblage optimal.

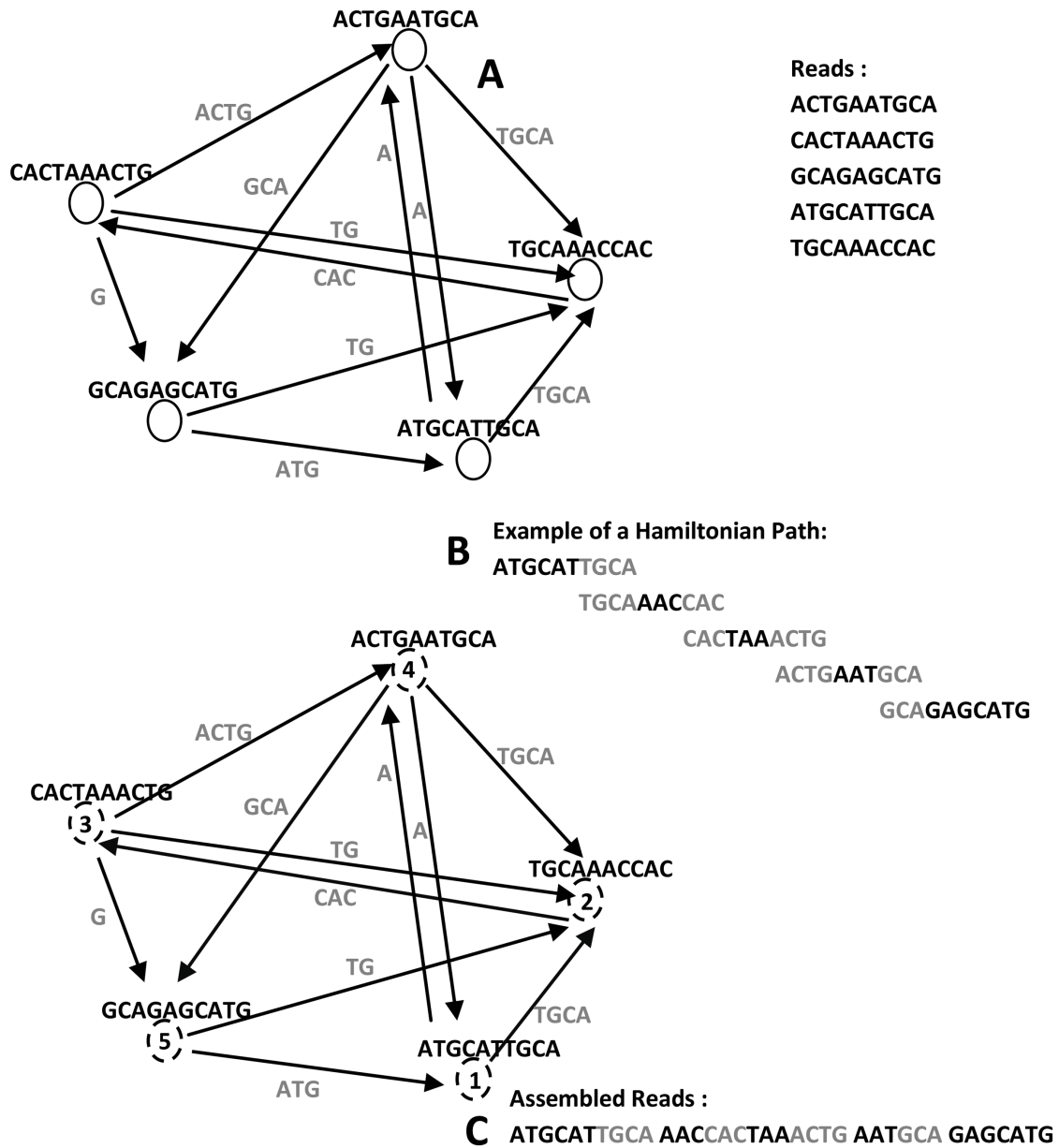
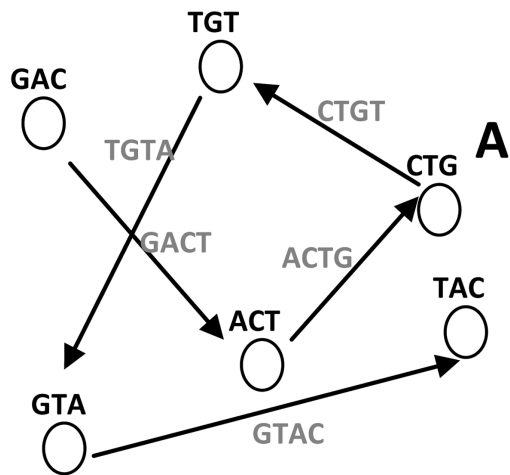


Figure 11 – Construction d'un graphe dans le cadre d'une approche de type « OLC ». (Metwally et al. 2013).

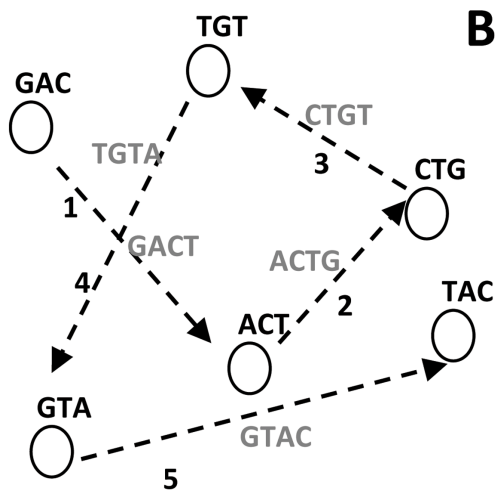
Les assembleurs basés sur les graphes de De Bruijn ont connu un développement accru avec l'émergence des lectures courtes et très précises (Illumina) mais sont néanmoins fondés sur des principes datant de l'assemblage de lectures Sanger (Idury and Waterman, 1995). Les nœuds correspondent à des K-mer (chaîne de K caractères) et les liens indiquent que ces K-mer se chevauchent par K-1 bases. Chaque lecture est représentée par un chemin au sein du graphe, deux lectures chevauchantes ont les extrémités de leur chemin en commun. Les lectures sont ensuite assemblées en identifiant des chemins dits Eulériens, ne passant qu'une seule fois par chaque lien (Pevzner et al., 2001). L'utilisation de K-mer pour identifier les régions chevauchantes entre des lectures a été un moyen universellement adopté par les assembleurs pour réduire la charge de calcul due au traitement de grande quantité de données. Deux lectures partageant un nombre important de K-mer peuvent-être considérées comme chevauchantes. Le K doit cependant être assez grand pour éviter les faux positifs et assez petit pour éviter les faux négatifs. Ce procédé réduit grandement la charge de calcul comparé notamment à des alignements pair-a-pair. Comparé aux assembleurs OLC, cette technique est néanmoins plus sensible aux répétitions et aux erreurs de séquençage, car chaque base erronée peut induire K nœuds artificiels et provoquer la formation de contigs chimériques (**Figure 12**).

Plusieurs structures peuvent être détectées dans les graphes formés par les assembleurs et correspondent à différents problèmes. Les 'spurs' sont des impasses généralement provoquées par une erreur de séquençage à la fin d'une lecture. Les nœuds contenant la base erronée étant artificiels, ils ont peu de chance d'être reliés à d'autres nœuds. Ces impasses peuvent aussi être dues à une absence totale de couverture dans une région particulière. Les 'bulles' sont formées par des chemins qui divergent puis convergent successivement. Ces dernières sont provoquées par des erreurs de séquençage au milieu des lectures formant un chemin alternatif. Les 'enchevêtrements', deux chemins convergeant puis divergeant ; et les 'cycles', des chemins convergeant sur eux-mêmes sont dus à des séquences répétées.



$R_1 = \text{GACTGTA}$ $R_2 = \text{ACTGTAC}$

Set of 3-Kmers of $R_1 = \text{GAC, ACT, CTG, TGT, GTA}$
 Set of 3-Kmers of $R_2 = \text{ACT, CTG, TGT, GTA, TAC}$



Example of an Eulerian path :

GACT
 ACTG
 CTGT
 TGTA
 GTAC

C Assembled Reads :
 GACTGTAC

Figure 12 – Construction d'un graphe dans le cadre d'une approche de type « K-mer ». (Metwally et al. 2013).

Les impasses sont résolues en supprimant les nœuds aboutissant à des chemins trop courts. Les bulles peuvent être résolues en se basant sur la fréquence des K-mer. Un K-mer contenant une ou plusieurs bases erronées devrait logiquement être retrouvé sur un très faible nombre de lectures. Il faut cependant veiller à prendre en compte les régions de basse couverture qui contiennent aussi naturellement des K-mer peu fréquents. Les enchevêtrements sont séparés en deux chemins distincts en réalignant les lectures sur le graphe ('read threading') ou bien en utilisant les informations des lectures en paire ('mate threading').

L'obtention d'un génome correctement assemblé (peu de contigs, contigs de grande taille) nécessite donc de choisir une méthode d'assemblage appropriée à la stratégie de séquençage choisie initialement. Dans le cadre de la génomique comparative et une fois plusieurs génomes obtenus, il est ensuite important d'établir les relations d'orthologies ou d'homologie de chacun des gènes identifiés.

Les relations d'homologie permettent de déterminer quels sont les gènes partagés par chacun des génomes afin de définir par exemple le core-génome, le pan-génome ou encore le génome accessoire. Les relations d'orthologie apportent entre autres des informations quant à la fonction du gène, s'il a été caractérisé chez une autre espèce. De nombreuses bases de données d'orthologues existent aujourd'hui : COG (Tatusov et al., 1997), eggNOG (Jensen et al., 2008), InParanoid (O'Brien et al., 2005), OrthoDB (Kriventseva et al., 2008), OrthoMCL (Li et al., 2003), OMA (Altenhoff et al., 2011). Chacune de ces bases se distingue par les organismes inclus (eucaryotes/procaryotes), le niveau de curation (manuel/automatique) et les méthodologies employées.

Par définition, les orthologues sont des homologues séparés par un évènement de spéciation. Les orthologues sont donc, généralement, plus divergents génétiquement que les homologues. Dans la littérature, les méthodes permettant de déterminer les relations ancestrales entre gènes sont communément désignées comme des méthodes de prédiction des orthologues, mais sont adaptées et couramment utilisées pour la prédiction des homologues.

Au plan méthodologique, plusieurs options existent et certaines sont adaptées à des situations particulières (**Figure 13**). Malgré la multiplication des méthodes publiées et des bases de données parfois associées, elles sont alternativement basées sur quelques principes centraux : la similarité des séquences ('graph based'), la phylogénie ('tree based') ou parfois la synténie. Les approches basées sur la phylogénie sont généralement plus adaptées aux familles de gènes complexes incluant des pertes et des duplications, que les approches basées sur la similarité (Trachana et al., 2011).

A partir de génomes assemblés dont les relations d'orthologies sont connues, il est alors possible de réaliser des comparaisons extensives à différents niveaux : gains et pertes de gènes, duplications, polymorphismes, phylogénies, transferts horizontaux, pressions de sélection... Quand la validation de certaines hypothèses nécessite des analyses peu ambiguës à mettre en place et à interpréter ; d'autres requièrent une réflexion minutieuse sur les méthodes à employer afin de garantir qu'elles soient appropriées aux données disponibles et à la réalité biologique. Par exemple, il existe des méthodes de détection des transferts horizontaux basées sur le GC%, la composition en nucléotides, la fréquence d'oligonucléotides, le contexte génomique, les profils phylogénétiques ou la congruence entre phylogénie du gène et de l'espèce... Un gène effectivement soumis à des transferts horizontaux ne sera pas nécessairement détecté par toutes ces méthodes ; certaines ne peuvent détecter qu'une signature de transfert au sein d'une famille tandis que d'autres sont capables d'inférer le donneur et le receveur.

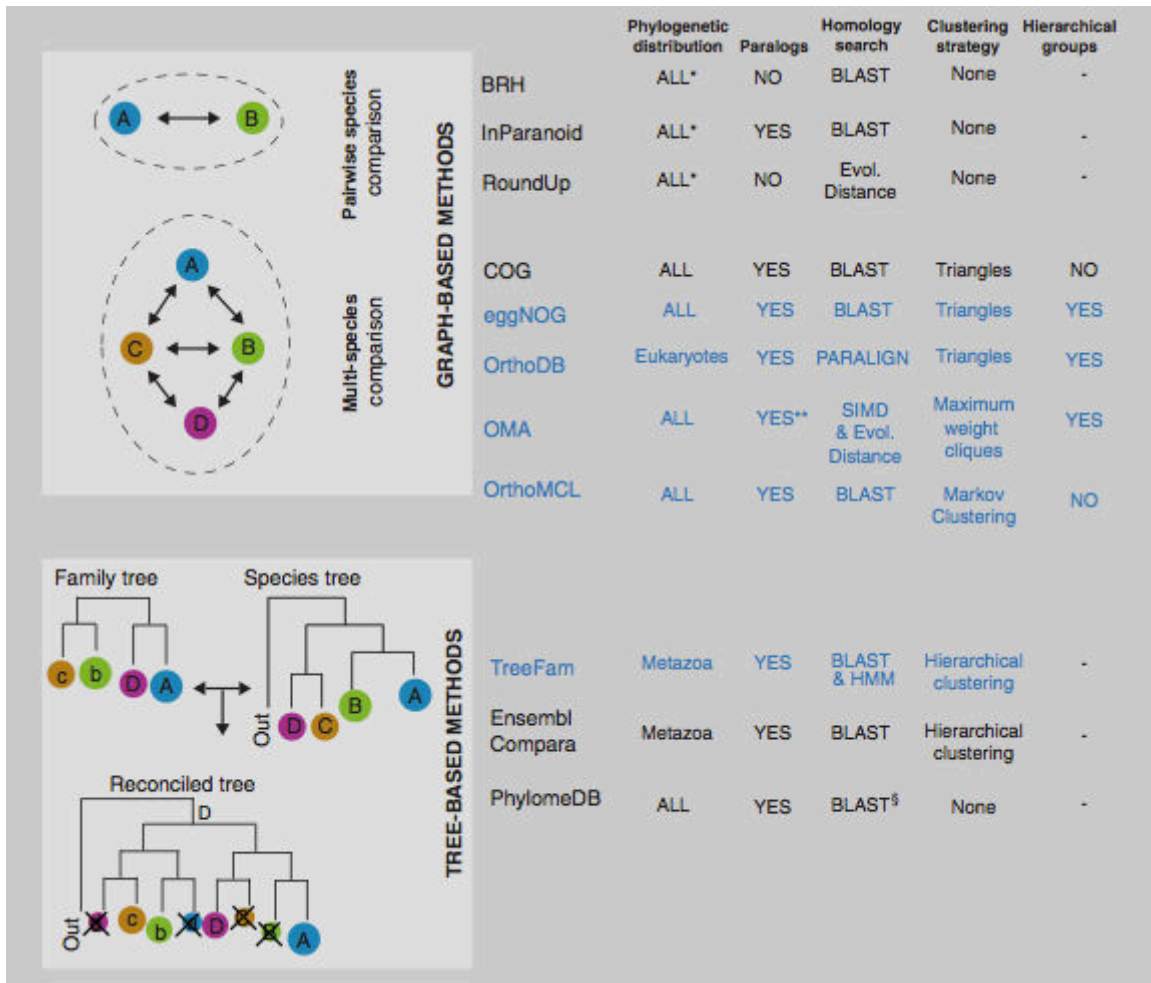


Figure 13 - Propriétés des méthodes de prédictions d'orthologues les plus répandues (Trachana et al. 2011).

c. Génomique et spectre d'hôte chez les phytopathogènes

La génomique comparative est donc un champ permettant de caractériser des organismes à divers niveaux à l'aide d'un vaste panel de méthodes bioinformatiques, elles-mêmes en constante évolution. Chez les organismes pathogènes en particulier, les phénomènes d'adaptation et les déterminants du spectre d'hôte peuvent être étudiés par la génomique comparative.

La comparaison de deux oomycètes proches, *Phytophthora infestans*, responsable du mildiou de la pomme de terre, et *P. mirabilis* infectant *Mirabilis jalapa*, une plante herbacée, a conduit à l'identification de plusieurs gènes sous sélection positive parmi lesquels un effecteur de type III inhibiteur de protéase et spécifiquement adapté à chaque hôte via une unique substitution d'acide aminé (Dong et al., 2014).

La caractérisation génomique d'une collection de souche de *Xanthomonas axonopodis*, une bactérie pathogène de nombreuses plantes à l'échelle de l'espèce, a décrit une évolution en deux étapes : une 1^{ère} étape de diversification généraliste, et une 2nde étape de formation d'écotypes avec des gammes d'hôtes distinctes. Les traces d'échanges génétiques récents entre écotypes suggèrent qu'ils pourraient favoriser l'émergence de nouveaux pathotypes (Mhedbi-Hajri et al., 2013).

Chez *R. solanacearum*, une dizaine de génomes répartis entre les quatre phylotypes ont été séquencés jusqu'à présent (Genin and Denny, 2012). La confrontation des génomes de deux souches Brown rot (IIB-1) présentant des différences de virulence importantes sur tomate et pomme de terre a mis à jour une délétion de 77kb hébergeant des facteurs contribuant à la pathogénèse (Gonzalez et al., 2011). L'utilisation de puces d'hybridation génomique comparative (CGH) (Guidot et al., 2007) (Cellier et al., 2012) (Lefeuvre et al., 2013) et la comparaison de génomes de chaque phylotype (Remenant et al., 2010) rapportent des différences de dynamiques dans l'évolution du chromosome et du mégaplasme. En effet les gènes situés sur le mégaplasme sont peu conservés et possèdent une synténie plus faible que les gènes portés par le chromosome. La taille du mégaplasme est également

plus variable que celle du chromosome, suggérant qu'il pourrait contenir préférentiellement des facteurs d'adaptation à l'hôte : il est notamment plus petit chez des pathotypes ayant une gamme d'hôte très réduite comme BDB et *R. syzygii* (Remenant et al., 2011).

4. Transcription et transcriptomique

a. Les méthodes d'étude de l'expression génique

La transcriptomique est, par définition, l'étude de toutes les molécules d'ARN produites par une cellule (i.e. le transcriptome). 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 se focalise, en pratique, sur l'expression des gènes et donc des ARNm. Cela s'explique en partie par le fait que l'implication des ARNm dans la biosynthèse des protéines est connue depuis les années 40 tandis que les mécanismes de régulation post-transcriptionnelle via des ARN interférants n'ont été découverts qu'à partir années 90 (Fire et al., 1998).

Historiquement, l'étude de l'expression des gènes se faisait principalement par Northern blot, une technique similaire au Southern blot consistant à extraire les ARN totaux, les séparer par électrophorèse puis détecter directement un transcrit d'intérêt à l'aide d'une sonde marquée après un transfert du gel sur membrane (Alwine et al., 1977). Plus tard, les sondes marquées par fluorescence pourront être directement hybridées sur des cellules ou des tissus perméabilisés et observées par microscopie afin d'étudier notamment la localisation des transcrits (FISH : Fluorescence In Situ Hybridization) (Langer-Safer et al., 1982). Du au caractère transitoire *in vivo* et à l'instabilité *in vitro* (RNAses présentes dans l'environnement) de la molécule d'ARN, la transcriptomique a rapidement évolué grâce au développement de multiples procédés ingénieux permettant de caractériser l'expression de manière indirecte. Globalement, deux types de méthodologie ont été universellement adoptés par la communauté scientifique :

Le premier type de méthodologie repose sur l'observation de l'expression des gènes au sein de la cellule (donc sans risque de dégradation accidentelle des ARN) par l'intermédiaire de gènes dits « rapporteurs » (marqueurs) transcrits par la machinerie cellulaire (IVET : In Vivo Expression Technology) (Mahan et al., 1993; Mahan et al., 2000; Rediers et al., 2005). Les marqueurs et les méthodes de détection associées sont très variés et autorisent une grande diversité de modèles expérimentaux : résistance antibiotique, auxotrophie, fluorescence (GFP), luminescence (luciférase), absorbance (B-galactosidase). Ces techniques ne détectent pas en tant que tels les transcrits, mais dépendent de l'utilisation du promoteur situé en amont du gène d'intérêt. Le gène rapporteur est placé dans un vecteur sans son promoteur d'origine, puis des séquences promotrices d'intérêt ou des morceaux d'ADN aléatoires provenant d'une librairie génomique sont insérés en amont de ce gène. La librairie de vecteurs d'expression ainsi créée est ensuite transfectée dans des cellules sauvages de l'organisme étudié. Enfin, l'expression du marqueur est criblée chez chacun des transformants dans des conditions expérimentales prédéterminées, afin de déterminer si l'insert contient un promoteur actif dans une condition spécifique. Par conséquent, les avantages majeurs de cette méthodologie sont la possibilité d'étudier la transcription, sans connaissance préalable du génome, *in vitro* ou *in vivo*, grâce à la diversité des marqueurs utilisables. Par exemple, une bactérie pathogène peut être étudiée dans son hôte naturel en identifiant une auxotrophie la rendant avirulente et un marqueur complétant cette auxotrophie; les transformants contenant un promoteur actif sont alors les seuls dont la virulence est restaurée. Alternativement et notamment dans le cas de la GFP, le gène marqueur peut être fusionné au gène dont on veut suivre l'expression qui produit alors une protéine de fusion fluorescente ; la transcription est alors détectée via la traduction de cette protéine. Cette méthode requiert cependant une connaissance plus approfondie du génome.

Le second type de méthodologie repose sur la détection des ARNm par amplification, hybridation et/ou séquençage via une conversion initiale en ADN codants (ADNc), plus stables, par rétrotranscription.

Amplification par PCR

La PCR quantitative en temps réel (RT-qPCR) est basée sur l'amplification d'un ADNc à l'aide d'amorces dirigées contre un gène spécifique et couplées ('quenched') à un fluorochrome (Ferre, 1992; Volkenandt et al., 1992). Lors de l'élongation, le fluorochrome est relâché par l'activité exonucléase 5' de la Taq polymérase. En parallèle, la fluorescence est détectée en temps réel par le thermocycleur (spécifiquement adapté à la qPCR) et est directement corrélée à la quantité d'ADN présent dans l'échantillon (i.e. donc au nombre de transcrits initiaux). La quantification peut se faire de manière relative en comparant la fluorescence de deux échantillons, ou absolue en la comparant à un standard. Cette méthode est très sensible (elle nécessite peu d'ARN au départ), et précise mais son débit est néanmoins limité par la nécessité de designer des amorces pour chaque gène et donc de connaître le génome *a priori*; il est cependant possible de multiplexer les réactions comme dans une PCR classique. La qPCR est encore très utilisée aujourd'hui, notamment pour valider individuellement les résultats obtenus par des méthodes haut-débit.

La méthode du 'Differential Display' (DD) (Liang and Pardee, 1992; Welsh et al., 1992) consiste à amplifier des ADNc avec des amorces aléatoires puis à séparer les amplicons par électrophorèse. L'intensité des bandes correspond alors à l'abondance des transcrits et peut être comparée entre différentes conditions. Les bandes peuvent également être extraites et séquencées individuellement. Comparées aux autres méthodes décrivant le transcriptome dans sa globalité, cette dernière a l'avantage d'être peu coûteuse et simple à mettre en place.

Hybridation

Ces techniques sont basées sur le principe des puces à ADN et possèdent donc les mêmes avantages et inconvénients. Elles nécessitent notamment de nombreuses étapes de normalisation dues aux variations des conditions expérimentales (pH, température, force ionique...), ainsi qu'une connaissance préalable du génome afin de dessiner la puce, et par conséquent ne permettent pas de découvrir de nouveaux

gènes chez des génotypes variants. Par ailleurs, le bruit de fond et la saturation du signal empêchent une quantification fiable des gènes très faiblement ou fortement exprimés (Okoniewski and Miller, 2006). Les puces à ADN sont néanmoins peu coûteuses et simples à utiliser une fois le design initial réalisé. Il existe deux types de design qui diffèrent par la quantité d'informations qu'ils fournissent. Le 1^{er} type correspond aux puces pangénomiques pour lesquelles chaque sonde correspond à un gène cible en particulier et dépend entièrement de la qualité des annotations du génome utilisé pour le design (Schena et al., 1995). Le 2nd type est communément appelé 'tiling array' et utilise des sondes dont les séquences, généralement chevauchantes, représentent partiellement ou complètement le génome à haute densité. Le design de ces dernières est donc indépendant de l'annotation du génome et des possibles erreurs informatiques ou humaines. En fonction du protocole utilisé, ces puces permettent d'étudier des phénomènes complémentaires de la transcription, contribuant à une meilleure interprétation des résultats d'expression obtenus (sites de fixation des régulateurs, méthylation...).

L'hybridation des ADNc avec les sondes est détectée grâce à un marqueur fluorescent ou luminescent et l'expression est quantifiée à partir de l'intensité du signal. Disposant de marqueurs différents, deux échantillons peuvent alors être mélangés et comparés sur une même puce ; leurs ADNc sont alors en compétition pour les mêmes sondes et le différentiel d'expression entre les échantillons est calculé à partir du ratio de l'intensité du signal produit par les deux marqueurs (Shalon et al., 1996) (**Figure 14**). Alternativement, il est possible de comparer les résultats provenant de plusieurs puces après une normalisation des intensités se basant sur un gène de référence, typiquement un gène de ménage dont l'expression est constante dans toutes les conditions et qui ne présente pas de dégradation du signal significativement différente des cibles étudiées. La transcription étant un phénomène complexe, il est néanmoins fréquent que l'expression des gènes de référence choisis soit en réalité variable (Thellin et al., 1999).

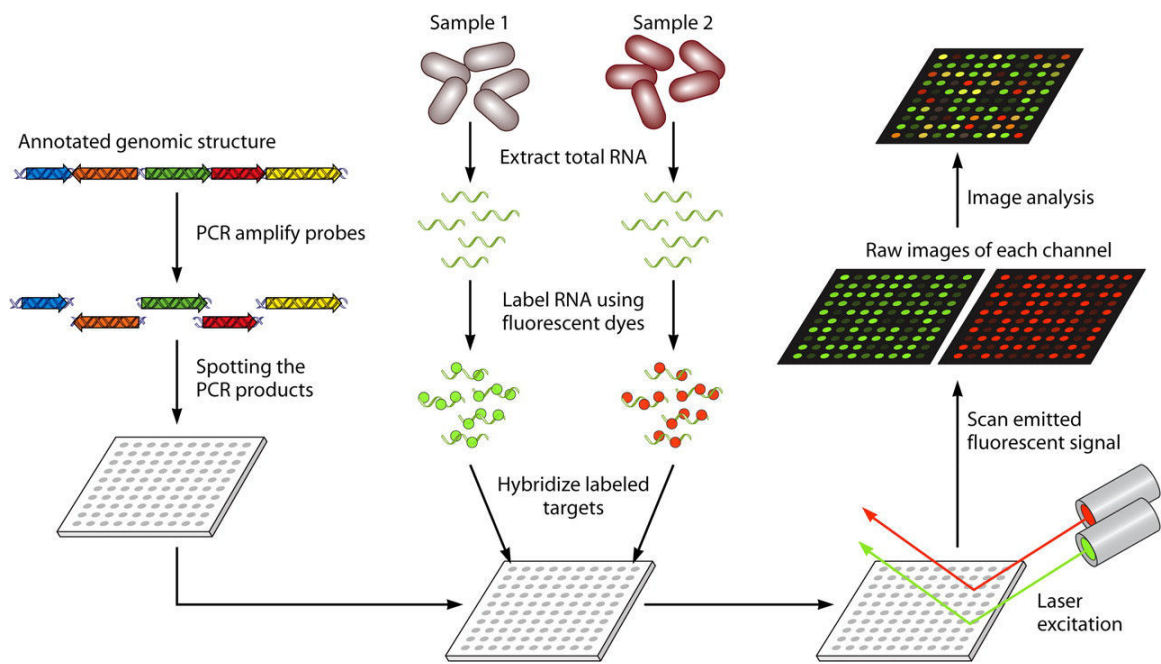


Figure 14 – Principe d'une puce à ADN basée sur des ADNc (Miller et al. 2009).

Séquençage

Les méthodes de quantification de l'expression des gènes basées sur le séquençage ne nécessitent pas de connaissance préalable du génome et elles ont donc l'avantage, entre autres, de permettre l'identification de nouveaux gènes. Initialement, ces méthodes ne reposent que sur une séquence partielle des transcrits ('tags'). Des bibliothèques d'ADNc sont ainsi séquencées aux deux extrémités (300-500pb) par Sanger et donnent lieu à la création de bases de données d'EST ('Expressed Sequenced Tags') (Adams et al., 1991). Ces fragments de séquences sont suffisamment uniques pour distinguer les transcrits, et la fréquence de chaque EST au sein de chaque bibliothèque est utilisée pour définir leur niveau d'expression. Puisqu'elle nécessite des étapes de clonage et de séquençage par Sanger, cette méthode est coûteuse et relativement laborieuse. (**Figure 15**)

Plusieurs autres méthodes basées sur les tags furent développées par la suite. La méthode SAGE ('Serial Analysis of Gene Expression') (Velculescu et al., 1995), développée à l'origine dans le cadre de l'étude des cellules cancéreuses (John Hopkins University), est relativement complexe dans son principe comme dans sa mise en place, et constitue les prémices du séquençage en parallèle. Les ADNc sont fixés à des billes (Streptavidine) via leur queue polyT (complément du polyA des ARNm, la méthode est donc en général destinée à l'étude des eucaryotes du à la rareté des ARNm polyadénylés chez les procaryotes), puis clivés par une 1^{ère} enzyme de restriction afin de fixer un adaptateur (NlaIII - 'Anchoring Enzyme'). Cet adaptateur contient lui-même un site de restriction d'une enzyme de type IIS clivant exactement 13pb en aval de ce site (BsmFI - 'Tagging Enzyme'). Le fragment détaché de la bille contient ainsi un tag de taille constante représentant la signature du transcrit. L'adaptateur est ensuite séparé du tag en réutilisant la 1^{ère} enzyme de restriction (NlaIII). Finalement, les tags provenant de plusieurs ADNc sont concaténés par ligation, clonés, amplifiés et séquencés par Sanger. Bien que cette méthode repose encore sur le clonage et le séquençage Sanger, comme les ESTs, l'étape de concaténation équivaut à une parallélisation du séquençage et permet d'obtenir bien plus de tags en une seule réaction (**Figure 16**).

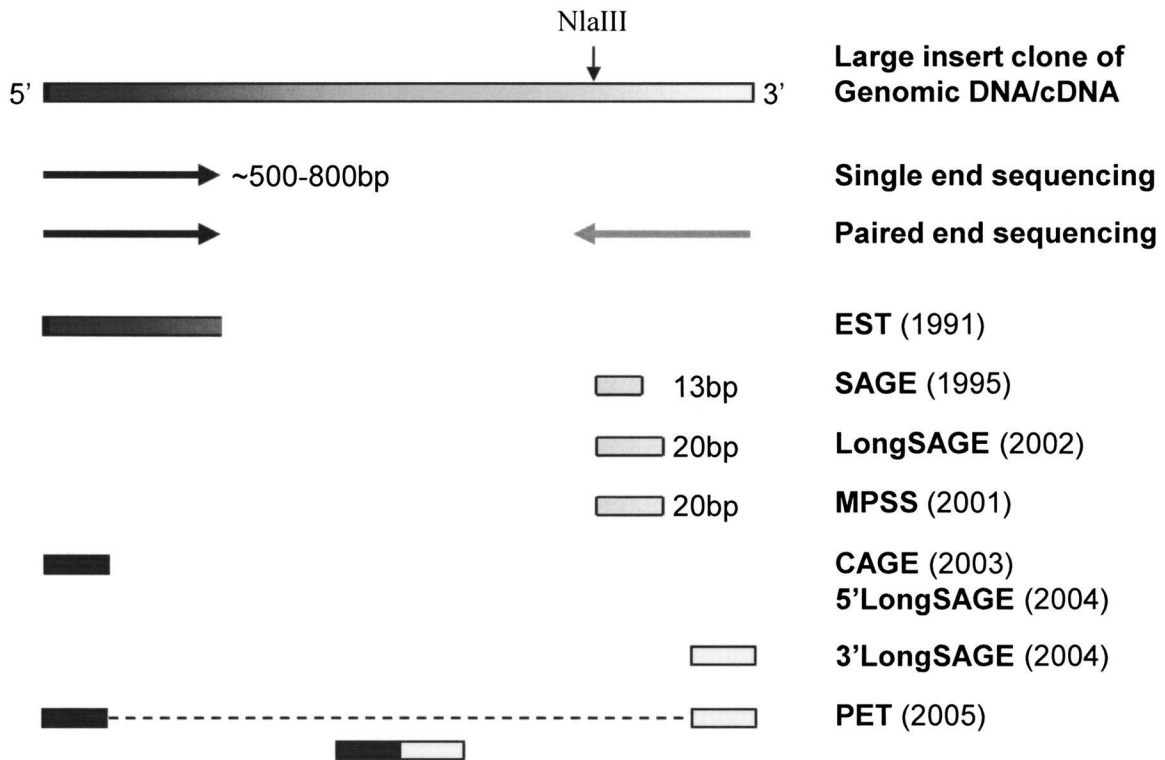


Figure 15 - Diversité nucléotidique des informations obtenues par séquençage dans le cadre d'analyses transcriptionnelles (Fullwood et al. 2009).

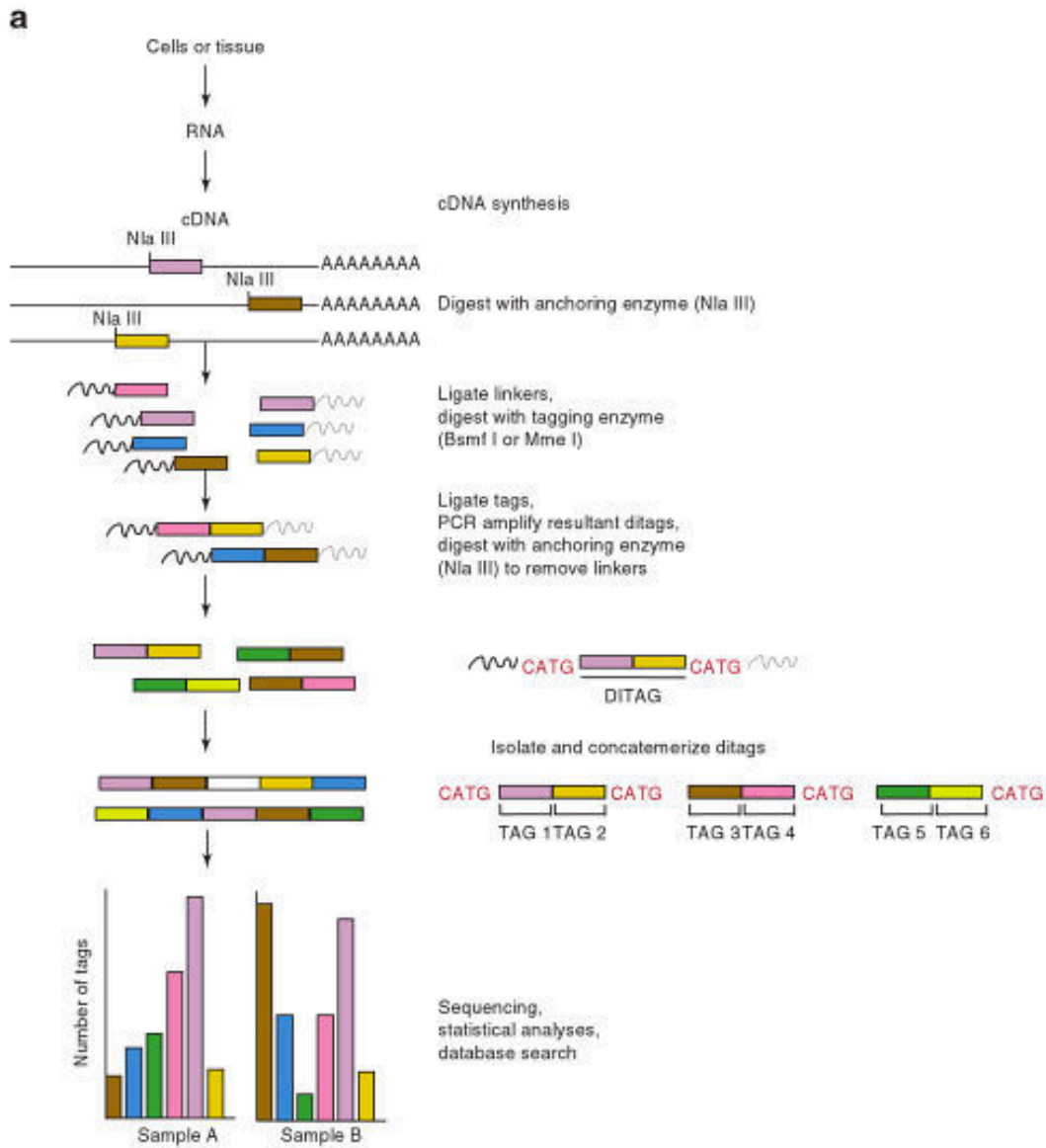


Figure 16 - Principe du SAGE (Modern Molecular Biology: Approaches for Unbiased Discovery in Cancer Research , Srinivasan Yegnasubramanian 2010).

Par la suite, la méthode SAGE a été améliorée fournissant la méthode MPSS ('massively parallel signature sequencing') (Brenner et al., 2000b). Les tags s'agrandissent grâce à l'utilisation d'une 'Tagging Enzyme' produisant des fragments de 20pb (MmeI). Après clonage, les concaténats de tags ne sont maintenant plus séquencés par Sanger, mais fixés de nouveau sur des billes. Les étapes de clonage et d'amplification rajoutent des tags « d'adressage » complémentaires d'« anti-tags » distincts fixés sur chaque bille afin de s'assurer qu'elles contiennent toujours les tags d'un même clone. Les billes sont ensuite triées sur une cellule microfluidique et séquencées avec une méthode de séquençage par ligation. Dans son ensemble, ce procédé permet un séquençage parallèle massif et produit ~20x plus de tags que par SAGE. La séquence est obtenue successivement par groupe de 4 nucléotides à l'aide d'adaptateurs fluorescents clivés après chaque cycle de ligation. Cette technique constitue l'origine des méthodes de 2nde génération ; la compagnie Lynxgen Therapeutics l'ayant commercialisé fusionnera d'ailleurs avec Solexa qui deviendra plus tard Illumina.

Cependant, les méthodes SAGE et MPSS sont dépendantes d'enzymes de restriction pour générer les tags, aussi les transcrits qui ne posséderaient pas certains sites de restriction ne sont donc pas pris en compte dans l'analyse. De plus, la taille limitée des tags rend parfois difficile l'identification des gènes dont ils proviennent, si leur séquence est présente au sein de plusieurs gènes ou bien en présence d'épissage alternatif.

Avec le développement des méthodes de 2nde génération, des méthodes de fragmentation (physiques ou chimiques) moins dépendantes de la séquence des ADNc sont utilisées. Plus important encore, le SGS permet de séquencer l'intégralité des ADNc et ainsi d'obtenir une quantité d'information bien plus importante (variations de séquences et structurales) (Wang et al., 2009). Par exemple, en l'absence de génome de référence sur lequel aligner les lectures de séquençage, le transcriptome peut être directement assemblé *de novo*. L'amplification des ADNc n'est également pas obligatoire, mais son absence nécessite cependant de plus grandes quantités d'ARN de départ. Sans amplification, la quantification de

l'expression est directement basée sur le nombre de transcrits présents dans l'échantillon et séquencé pour chaque gène (signal numérique). Le SGS a ainsi une plage de détection, bien plus large que les puces à ADN (signal analogue) et peut distinguer des gènes très fortement ou très faiblement exprimés. La détection des transcrits rares (faiblement exprimés) nécessite néanmoins d'avoir une profondeur de séquençage (couverture) suffisante pouvant augmenter significativement le coût de l'analyse. Le protocole le plus fréquemment utilisé est celui du RNAseq sur la plateforme Illumina. Avec le séquençage de 3^{ème} génération, il est maintenant possible de séquencer directement les ADNc en intégralité sans passer par une étape de fragmentation.

Les données produites par les différentes méthodes d'analyses décrites ci-dessus requièrent des approches statistiques fondamentalement différentes. Les puces à ADN fournissent des intensités lumineuses suivant théoriquement une distribution continue (loi Normale), tandis que le séquençage produit un compte de tags ou de lectures suivant une distribution discrète (loi de Poisson, binomiale négative). Par conséquent, une méthode statistique développée pour les puces à ADN ne peut être directement utilisée pour analyser des données de comptage (Rodríguez-Ezpeleta et al., 2012). De façon similaire à l'évolution des méthodes d'assemblage lors de la transition du Sanger vers le SGS, les modèles statistiques au cœur de l'analyse des comptages, dérivées du séquençage, sont devenus significativement plus nombreux et complexes que ceux utilisés pour les puces à ADN du aux quantités de données produites. Ces modèles sont encore en développement aujourd'hui et de nouveaux modèles sont fréquemment proposés (approche bayésienne) sans qu'un consensus définitif n'existe sur la méthode la plus efficace (Soneson and Delorenzi, 2013).

b. Transcriptomique et spectre d'hôte chez les phytopathogènes

L'expression des facteurs de virulence et leurs régulations ont été vastement étudiés chez *R. solanacearum*, notamment à l'aide de gènes rapporteurs (Allen et al., 1997; Flavier et al., 1997; Huang et al., 1998; Genin et al., 2005; Monteiro et al., 2012; Zuluaga et al., 2013). En revanche, les travaux concernant le profil d'expression global du génome de *R. solanacearum* sont fragmentaires.

Une analyse de type IVET, basée sur une souche mutante dont l'auxotrophie pour le tryptophane était associée à la perte de virulence chez la tomate a permis d'identifier plus de 150 gènes exprimés durant l'infection (Brown and Allen, 2004). Par la suite, des travaux similaires ont été réalisés avec une puce à ADN et les auteurs ont comparé les profils d'expression chez deux souches appartenant à des lignées distinctes (phylotype I et IIB-1) colonisant l'hôte sensible (la tomate) ou cultivant un milieu de culture riche (CPG) (Jacobs et al., 2012). Cette comparaison a révélé que les gènes partagés par les deux souches ont un profil d'expression similaire chez la tomate ; parmi ces gènes, 2898 sont exprimés chez la tomate et 402 sont différenciellement exprimés par rapport au milieu riche (sur- ou sous-expression). Spécifiquement, ces travaux ont démontré que le T3SS était fortement actif *in planta* (à haute densité cellulaire) contrairement au modèle de régulation précédemment établi (Genin et al., 2005; Yoshimochi et al., 2009). Ces résultats furent validés indépendamment à l'aide d'un marqueur luminescent (Monteiro et al., 2012). Aucune analyse transcriptomique par séquençage à haut-débit n'a été effectuée à ce jour chez *R. solanacearum*, néanmoins les changements induits chez un hôte (l'arachide) au cours de l'infection ont été analysés par RNAseq (Chen et al., 2014).

Au delà des aspects fondamentaux sur la connaissance des mécanismes de la virulence, la transcriptomique peut également être mise à profit dans l'identification des mécanismes d'adaptation expliquant les variations de la gamme d'hôte des phytopathogènes. Par exemple, chez *Xanthomonas citri*, une bactérie monomorphe pathogène des agrumes, la comparaison de deux génotypes par RNAseq a révélé la

surexpression de régulateurs centraux de la virulence (T3SS, T2SS, flagelles, chimiotactisme) chez la souche possédant une gamme d'hôte restreinte d'une part et, d'autre part, la surexpression de facteurs d'adhésion ou de résistance au stress chez la souche possédant une gamme d'hôte large (Jalan et al., 2013). De plus, ces travaux ont identifié un nombre plus important de gènes différentiellement exprimés par rapport à une étude précédente basée sur l'analyse par puce à ADN dans les mêmes conditions expérimentales (Astua-Monge et al., 2005); illustrant ainsi la sensibilité accrue des méthodes SGS.

Chapitre 2 – Révision taxonomique du complexe d'espèces

Depuis la caractérisation de l'agent du flétrissement bactérien par E. F. Smith en 1896, sa classification n'a cessé d'évoluer en fonction des avancées techniques et méthodologiques. Aujourd'hui, la définition standard des espèces en bactériologie repose sur une combinaison de traits phénotypiques et d'analyses d'hybridation ADN-ADN (DDH). Depuis l'avènement du séquençage, la validité de ces méthodes en tant que mètre-étalons pour la classification des espèces est fréquemment remise en cause (**Figure 17**). La DDH présente notamment des problèmes de répétabilité dus à la complexité de la méthode. C'est une technique coûteuse et laborieuse à mettre en place du à la rareté des laboratoires la pratiquant encore. La DDH ne permet également pas la mise en place de bases de données incrémentales. A une échelle plus globale, la DDH ne mesure pas directement le degré de similarité mais la capacité de l'ADN à s'hybrider et ne fournit pas d'autres informations supplémentaires. Plusieurs problèmes inhérents à la DDH, furent d'ailleurs spécifiquement discutés lors d'analyses chez *R. solanacearum* (Palleroni and Doudoroff, 1971) et *R. syzygii* (Roberts et al., 1990), encore incluses à l'époque dans le genre *Pseudomonas*. Par ailleurs, ces deux études faisaient déjà mention de la diversité génétique inhabituelle de *R. solanacearum* et suggéraient l'urgence d'approfondir les connaissances des relations phylogénétiques entre les différentes souches de l'agent du flétrissement bactérien.

La révision proposée dans cette étude fait suite à une publication précédente de l'équipe qui suggérait la séparation du complexe d'espèce en trois espèces distinctes sur des bases génomiques (Remenant et al., 2011) : le phylotype I et III regroupés en une seule espèce, le phylotype II et IV formant chacun une espèce distincte. Cette proposition s'est vue concrétisée récemment par Safni et al. (Safni et al., 2014) avec plusieurs modifications par rapport a la proposition originale, en particulier s'agissant des souches types proposées mais surtout pour ce qui concerne les noms d'espèces proposés afin de respecter les règles taxonomiques internationales

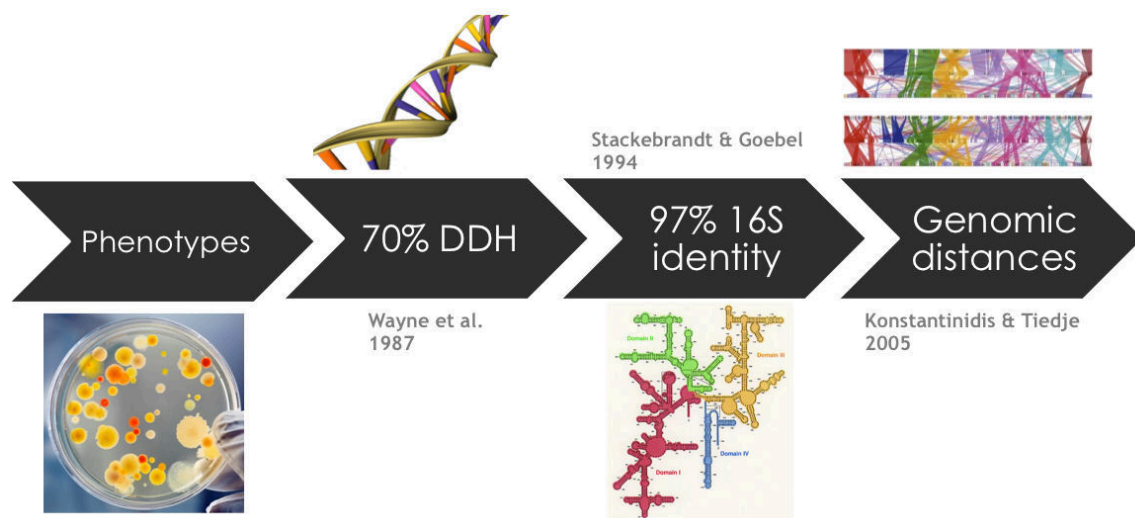


Figure 17 - Evolution des méthodes de classification taxonomique au cours du temps.

(Garrity et al., 2005). La classification des phylotypes en trois nouvelles espèces est identique, le phylotype I et III deviennent *R. pseudosolanacearum* sp. nov.; le phylotype II reste *R. solanacearum* car c'est dans ce groupe qu'est assigné la souche type; le phylotype IV devient *R. syzygii* – car le nom d'espèce est toujours valide d'un point de vue taxonomique – et comprend trois sous-espèces afin de décrire la diversité phénotypique de ce groupe particulier : les souches précédemment désignées comme *R. solanacearum* deviennent la sous-espèce *indonesiensis* subsp. nov., les souches BDB sont désignées comme la sous-espèce *celebensis* subsp. nov. et les *R. syzygii* originales deviennent la sous-espèce *syzygii* subsp. nov.. Safni et al. ont démontré la validité de cette séparation de manière polyphasique, avec des méthodes historiques regroupant des analyses DDH et des profils phénotypiques à haut-débit. Notre étude vient supporter cette révision via des techniques modernes reconnues par la communauté internationale se basant sur le séquençage à haut débit et une technique plus exploratoire reposant sur des profils protéomiques obtenus par spectrométrie de masse (**Figure 18**). Trois algorithmes distincts ont été utilisés pour obtenir ces valeurs : ANI (Average Nucleotidic Identity) qui a été une des premières méthodes répandues suggérant que les distances génomiques pouvaient remplacer la DDH pour la délimitation des espèces (Konstantinidis and Tiedje, 2005) ; MUMi (Maximum Unique Matches index) qui repose sur les mêmes principe que l'ANI mais améliore entre autre le temps de calcul (Deloger et al., 2009) ; GGDC (Genome to Genome Distance Calculator) qui est une méthode encore mise à jour récemment et qui propose des valeurs de DDH prédites *in silico* (Auch et al., 2010).

CES TRAVAUX ONT FAIT L'OBJET D'UNE PUBLICATION INTITULEE : « FURTHER EVIDENCES FOR THE DIVISION OF THE PLANT PATHOGEN *RALSTONIA SOLANACEARUM* INTO THREE SPECIES INFERRED BY GENOMIC AND PROTEOMIC » SOUMISE EN FEVRIER 2015 A LA REVUE SYSTEMATIC APPLIED MICROBIOLOGY [IF 3,51].

Matrix-Assisted Laser Desorption/Ionization-Time Of Flight

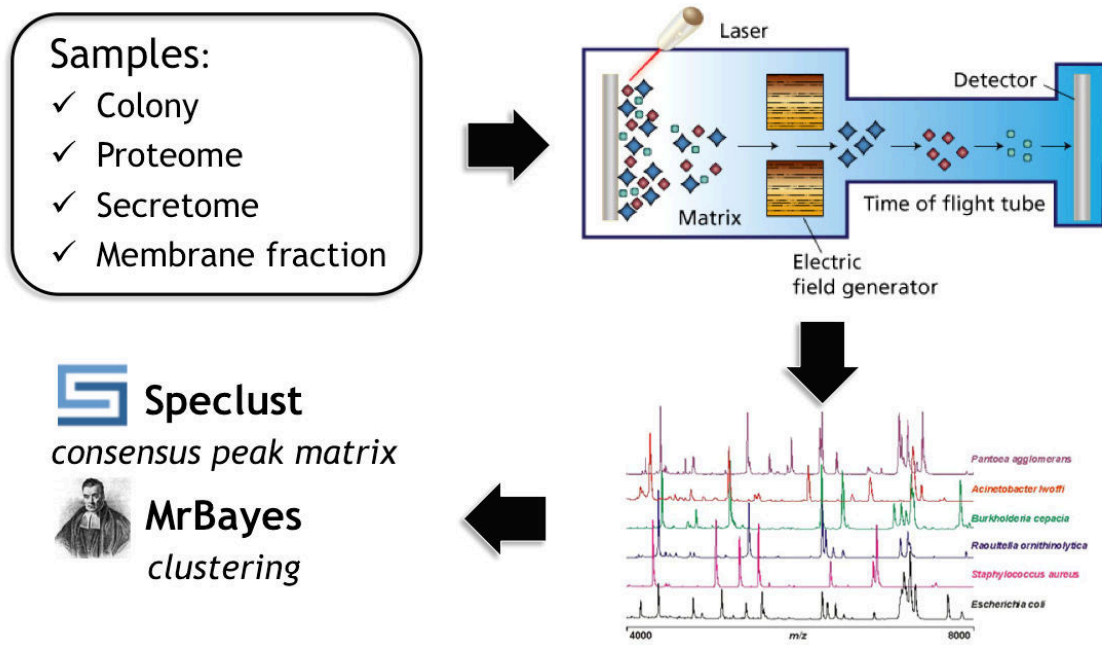


Figure 18 – Méthodologie utilisée pour produire un arbre à partir de données protéomiques.

1 **Additional evidence for the division of the plant pathogen *Ralstonia solanacearum***
2 **into three species inferred through genomic and proteomic analyses.**

3 Philippe Prior ^{a, d,*, 1}, Florent Ailloud ^{a, b, 1}, Beth L. Dalsing ^c, Benoit Remenant ^a, Borja
4 Sanchez ^{d, e}, and Caitilyn Allen ^c

5

6 ^a UMR PVBMT Peuplements Végétaux et Bioagresseurs en Milieu Tropical, CIRAD,
7 Saint Pierre, La Réunion, France

8 ^b Anses – Plant Health Laboratory (LSV), 7 chemin de l'IRAT, Saint-Pierre, La Réunion,
9 France

10 ^c Department of Plant Pathology, University of Wisconsin-Madison, 1630 Linden Drive,
11 Madison, WI 53706, USA

12 ^d Department of Plant Health and Environment (SPE), INRA, France.

13

14 ^e Department of Analytical and Food Chemistry, University of Vigo, Ourense, Spain

15

16

17 *Corresponding author: UMR PVBMT Peuplements Végétaux et Bioagresseurs en
18 Milieu Tropical, CIRAD, Saint Pierre, La Réunion, France. E-mail address:
19 philippe.prior@cirad.fr (P. Prior)

20 ¹ Both authors contributed equally to this study.

21 **Abstract** (214 words)

22 The increased availability of genome sequences has advanced the development of
23 genomic distance methods to describe bacterial diversity. Results of these fast-evolving
24 methods are highly correlated with those of the historically standard DNA-DNA
25 hybridization technique. They are thus a technically accessible replacement for species
26 delineation. Members of the *Ralstonia solanacearum* group cause bacterial wilt diseases
27 of many different plants. These diverse and widespread pathogens form a highly
28 heterogeneous species complex consisting of four genospecies, or phylotypes. We used
29 three different methods to compare the complete genomes of 29 strains from the *R.*
30 *solanacearum* species complex. In parallel we profiled the proteomes of 74 strains using
31 Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry
32 (MALDI-TOF-MS). Proteomic profiles together with genomic sequence comparisons
33 consistently and comprehensively described the diversity of *R. solanacearum*. In
34 addition, genome-driven functional phenotypic assays demonstrated that *R.*
35 *solanacearum* strains in phylotypes II and IV, which lack the NosZ nitrous oxide
36 reductase, do not produce dinitrogen gas from nitrate. Additionally, strains in phylotype I
37 and III can grow anaerobically on nitrate, indicating that they reduce nitrate to nitrite.
38 Together, the results of these studies support the proposed division of the *R.*
39 *solanacearum* species complex into three species, consistent with recent literature, and
40 demonstrate the utility of proteomic and genomic approaches to delineate bacterial
41 species.

42
43 *Keywords:* *Ralstonia solanacearum*, bacterial wilt, plant pathogen, taxonomy, genomics,
44 proteomics¹

ANI: Average Nucleotide Identity

GGDC: Genome-to-Genome Distance Calculator

BDB: Blood Disease Bacterium

DDH: DNA–DNA hybridization

MALDI-TOF: Matrix-assisted laser desorption/ionization-Time of flight

MLST: Multi Locus Sequence Typing

MUM: Maximum Unique Matches

RSSC: *Ralstonia solanacearum* species complex

45 **Introduction**

46 Thousands of genetically distinct strains within the *Ralstonia solanacearum* species
47 complex (RSSC) cause bacterial wilt diseases in plants. These bacteria colonize the
48 xylem tissue of host plant vascular systems, causing stunting, wilting, yield reduction,
49 and death. This pathogen group has major economic and social impact worldwide [1, 2].
50 Members of the RSSC can collectively infect over 250 hosts in 54 botanical families and
51 include: *R. solanacearum* strains, which collectively infect a broad host range and are
52 typically soil-borne; *R. syzygii*, a spittlebug-transmitted pathogen that causes Sumatra
53 disease in cloves; and the Blood Disease Bacterium (BDB), an unclassified organism
54 responsible for the pollinator-transmitted Blood Disease of bananas and plantains in the
55 Philippines.

56 Smith first described the morphological and chemotaxonomic characteristics of the
57 bacterial wilt pathogen as *Bacterium solanacearum*, and this species has most recently
58 been placed in the genus *Ralstonia* [3, 4]. The BDB was described and named
59 *Pseudomonas celebensis* in 1921 [5, 6]. However this name lost its standing in
60 nomenclature when the original strain got misplaced and could thus not serve as an
61 authentic type matching the description of the pathogen. The Sumatra disease pathogen,
62 originally described as *Pseudomonas syzygii*, was placed in a separate species in the
63 genus *Ralstonia* based on 16S sequences and DNA–DNA hybridization (DDH) data
64 showing substantial divergence from *R. solanacearum* [7]. However, the DDH study that
65 concluded *R. syzygii* should be placed in a separate species was based on a comparison
66 with *R. solanacearum* K60^T, a phylotype II strain that is quite divergent from phylotype
67 IV at the phenotypic and genotypic levels.

68 DNA–DNA hybridization has been used to distinguish species since the 1960s, and this
69 technique has contributed to the modern bacterial species concept [6, 8]. However,
70 because complete sequenced genomes contain significantly more information than can be
71 inferred from the results of DDH, this technique can now be replaced with bioinformatics
72 methods that take advantage of the full genetic features of different microbes [9, 10].
73 Early analyses based on the single-gene phylogeny of the conserved *egl*, *mutS*, *hrpB* or
74 ITS sequences divided the RSSC into four distinct genospecies, known as phylotypes,
75 corresponding to strain geographic origin: phylotype I (Asia), phylotype II (Americas),
76 phylotype III (Africa), and phylotype IV (Indonesia and Japan) [11-14]. The phylogenetic
77 structure of the RSSC was subsequently confirmed in an extensive series of genomic
78 studies involving a large array of analytical methods from microsatellites and MLST to
79 microarrays [12, 15, 16].

80 The analysis of the complete genome sequences of several strains in the RSSC provides
81 strong evidence supporting the phylotype structure [8, 11, 17-19]. These data reveal a
82 degree of evolutionary divergence among the phylotypes that warrants the division of the
83 RSSC into three species, as previously suggested [18]. Recently, Safni et al. (2014) [20]
84 supported this taxonomic revision, suggesting an amendment of the descriptions of the
85 RSSC based on a polyphasic approach: *R. pseudosolanacearum* sp. nov., corresponding
86 to phylotypes I and III, and *Ralstonia syzygii*, corresponding to phylotype IV. Based on
87 differences in pathological phenotype, *R. syzygii* is further divided into three subspecies.

88 The broad host-range soil-borne strains are renamed *R. syzygii* subspecies *indonesiensis*
89 subsp. nov. The unclassified banana Blood Disease Bacterium is named *R. syzygii*
90 subspecies *celebesensis* subsp. nov. and *R. syzygii*, which causes Sumatra disease in
91 cloves, is renamed *R. syzygii* subspecies *syzygii* subsp. nov. Finally, phylotype II strains
92 (from the Americas), which include the species type strain K60^T (=ATCC11696^T
93 =LMG2299^T), remain in *R. solanacearum*.

94 In the present study, we used a combination of genomic and proteomic methods to
95 unambiguously delineate the three species within the RSSC. We formally examined these
96 methods for correspondence to the recent reclassification of this taxonomically disputed
97 organism into three distinct species, consistent with Safni *et al.* (2014). This work
98 extends former research by providing tools for the rapid identification and classification
99 of new isolates into species and subspecies without the need to rely on archaic methods
100 including DDH.

101 **Materials and methods**

102 The sequenced strains used in the present study are listed in Table 1. The strains used for
103 the proteomic analyses are listed in Supplemental Table 2. The phylotype placement of
104 all strains was confirmed using the multiplex PCR method [13].

105 *Genomics.*

106 The complete and assembled genome sequence data used here are publicly available via
107 the MicroScope web interface at www.genoscope.cns.fr/agg/microscope/ralstoniascope.
108 The Average Nucleotide Identity (ANIb) between genomes was calculated according to
109 Konstantinidis and Tiedje [21], and the genomic distances were obtained after subtracting
110 the ANIb values from 1. The Maximal Unique Matches index (MUMi) distances between
111 genomes were calculated using the Perl script developed by Deloger *et al.* [22] using
112 MUMmer genome alignment software [23]. The Genome-to-Genome Distance
113 Calculator (GGDC) was used as previously described [24]. The DDH values were
114 derived from the GGDC distances using formula 2 [24]. A phylogenetic network derived
115 from the distance matrices produced with all three methods was created using SplitsTree4
116 software [25]. Distances matrices are available in **Supplemental Table 1**.

117 *Anaerobic inorganic nitrogen metabolism assessments.*

118 Nitrate respiration and complete denitrification were assessed using slightly modified
119 VDM medium [26, 27]. To decrease nitrate-independent anaerobic growth, we used
120 casamino acids instead of yeast extract [28]. Additionally, we omitted nitrate from the
121 base medium. This was done to allow assays to be conducted with and without nitrate
122 under otherwise similar conditions. Where specified, 30 mM NO₃⁻ (the concentration
123 found in host plant xylem sap) was added in the form of filter sterilized KNO₃ [28].
124 1.5mL of this modified VDM (+/- NO₃⁻) was inoculated with a specified bacterial strain
125 to a starting O.D.₆₀₀ of ~0.001 (~1x10⁶ CFU/mL). Tubes were incubated without
126 agitation at 28 °C under anaerobic conditions in a BD GasPak anaerobic system. 72 hours
127 post inoculation, O.D.₆₀₀ measurements were taken from each culture. Two to four

128 biological replicates were conducted per strain, per treatment (+/- NO₃⁻). To determine if
129 nitrate respiration contributed to anaerobic growth, O.D.₆₀₀ data were compared between
130 + and - NO₃⁻ treatments for each strain and depicted as a ratio. A ratio above 1 indicates
131 that the strain grew better anaerobically when provided with NO₃⁻. A ratio of 1 or below
132 indicates that the addition of NO₃⁻ did not enhance anaerobic growth, and that the strain
133 did not respire with NO₃⁻ under the conditions tested. Additionally, all cultures were
134 visually assessed (± bubbles) over the course of 96 hours for production of dinitrogen gas,
135 the end product of complete denitrification [26, 29].

136 *Analysis of denitrification genes.*

137 Presence or absence of homologs involved in denitrification were determined in all
138 sequenced strains (**Table 1**) using the MicroScope web interface and BLAST [30] to look
139 for loci identified in the GMI1000 strain: *fnr* (RSc1283); *narX* (RSp0979); *norR*
140 (RSp0959); *rpoNI* (RSc0408); *nosR* (RSp1369); a predicted nitric oxide metabolism-
141 related gene labeled *hypregI* (RSp1504); *nsrR* (RSc3397); *nnrS* (RSc3399); *norA*
142 (RSp0958); *narG* (RSp0974); *nosZ* (RSp1368); *aniA* (RSp1503); and *norB* (RSp1505).
143 Identity values were computed with the R package seqinr [31] after aligning amino-acid
144 sequences with MUSCLE [32].
145

146 *Bacterial typing using Matrix-Assisted Laser Desorption Ionization Time-Of-Flight Mass* 147 *Spectrometry.*

148 Seventy-three strains belonging to different phylotypes of the *Ralstonia solanacearum*
149 complex species were characterized at the proteomic level using Matrix-Assisted Laser
150 Desorption Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS). *Ralstonia*
151 *eutropha* LMG 1199 was included in the analysis as an outgroup (**Suppl. Table 2**).
152 Bacterial strains were grown on Kelman broth supplemented with agar for 48 h at 28°C.
153 For whole-cell protein extraction, 1 µL of the bacterial biomass was collected and
154 resuspended in a solution containing 50% (v/v) acetonitrile (Acros Organics, Fair Lawn,
155 NJ, USA) and 1% (v/v) trifluoroacetic acid (Sigma-Aldrich, Saint Louis, MO, USA) in
156 Milli-Q® ultrapure water (EMD Millipore Corporation, Billerica, MA, USA). The
157 suspensions were vortexed twice for 10 s and centrifuged at 20,000 g for 10 min at RT.
158 The supernatants were transferred and aliquoted into new tubes and stored at -20°C until
159 further analysis.

160 One microliter of the bacterial extracts was mixed with 1 µL of a saturated solution of α-
161 cyano-4-hydroxycinnamic acid (Sigma-Aldrich), which was used as a matrix. The
162 resulting sample/matrix mixture was deposited onto a stainless plate, dried at room
163 temperature, and introduced into the MALDI-TOF MS instrument for analysis. The mass
164 spectra profiles were obtained using a bench-top Microflex™ MALDI-TOF from Bruker
165 Daltonics, including the Flex Control and Flex Analysis v3.3 software, at the
166 Bacteriology Division of the CHU of St. Pierre, La Réunion. All spectra were obtained in
167 linear positive-ion mode with an m/z range of 2,000-20,000 Da. Each spectrum was
168 calculated as the sum of 320 accumulated laser shots obtained after a spiral trajectory of
169 the laser. For each sample, two bacterial extracts were obtained and measured in

170 duplicate, and all the spectra were calibrated using a standard preparation of *Escherichia*
171 *coli* DH5 α , according to Bruker Daltonics.

172 All bacterial spectra were analyzed using FlexAnalysis software (Bruker Daltonics) to
173 generate peak lists for each species, and only peaks with a relative intensity greater than
174 2% were considered for cluster analysis. The peak lists were exported to a CSV file,
175 exported to single files using a custom macro and loaded onto the SPECLUST web-
176 service (<http://bioinfo.thep.lu.se/speclust.html>) to obtain a consensus peak list for all
177 strains considered. For the consensus peaks, a peak match score (σ) width of ± 3 Da was
178 considered.

179 *Phylogenetic reconstruction using the MALDI-TOF data.*

180 The consensus peak list was formatted into a sequential Nexus binary file and loaded into
181 MrBayes 3.2.2 software (<http://mrbayes.sourceforge.net/>). Phylogeny was obtained
182 through Bayesian inference using the restriction data type (two states: absence or
183 presence of a peptide denoted by a 0 or a 1, respectively), assuming that the frequencies
184 of the two possible states had a Dirichlet (1.00, 1.00) prior. Bayesian analysis was
185 performed in two runs using 8 Markov chains and 3,000,000 generations. The potential
186 scale reduction factor implemented in MrBayes 3.2.2 was used as a convergence
187 diagnostic. A majority-rule consensus tree (50%) was obtained after discarding 25% of
188 the initial trees (burn-in=0.25) generated before the stabilization of the log likelihood
189 values of the data plotted against the number of generations. The trees were subsequently
190 edited using FigTree v1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).

191 **Results and Discussion**

192 *RSSC exhibits phenotypic diversity*

193 Safni et al. [20] analyzed the RSSC using phenotype microarrays and identified major
194 differences in the core metabolisms of each phylotype, supporting the idea that *R.*
195 *solanacearum* can be divided into multiple species based on biological data. In the
196 present study, we focused on the denitrification metabolic pathway, as it is a biologically
197 relevant phenotypic trait [28] related to known genetic features that differentiate
198 phlotypes I and III from the other *R. solanacearum* strains.

199 All tested strains in phlotypes I and III underwent complete denitrification and could
200 respire on nitrate, as demonstrated by the production of nitrogen gas and significant
201 growth when incubated anaerobically in nitrate-containing medium (**Fig. 1A**). In 1990,
202 Hayward recognized and examined the variability of anaerobic nitrogen metabolism
203 between *R. solanacearum* strains [25]. However, the biovar sub-classification system, in
204 use at the time, did not correspond to the phylogenetic relationships among RSSC strains.
205 No clear patterns differentiated biovars by anaerobic nitrogen metabolism and, thus
206 denitrification was not considered a useful trait for strain typing. The reclassification of
207 strains into phlotypes based on phylogenetic analyses revealed that this metabolic trait is
208 a defining characteristic of RSSC subgroups. Denitrification is an anaerobic respiration
209 process that allows strains to use nitrate as a terminal electron acceptor in order to grow
210 under anaerobic conditions. Nitrate (NO₃⁻) is converted successively to nitrite (NO₂⁻),

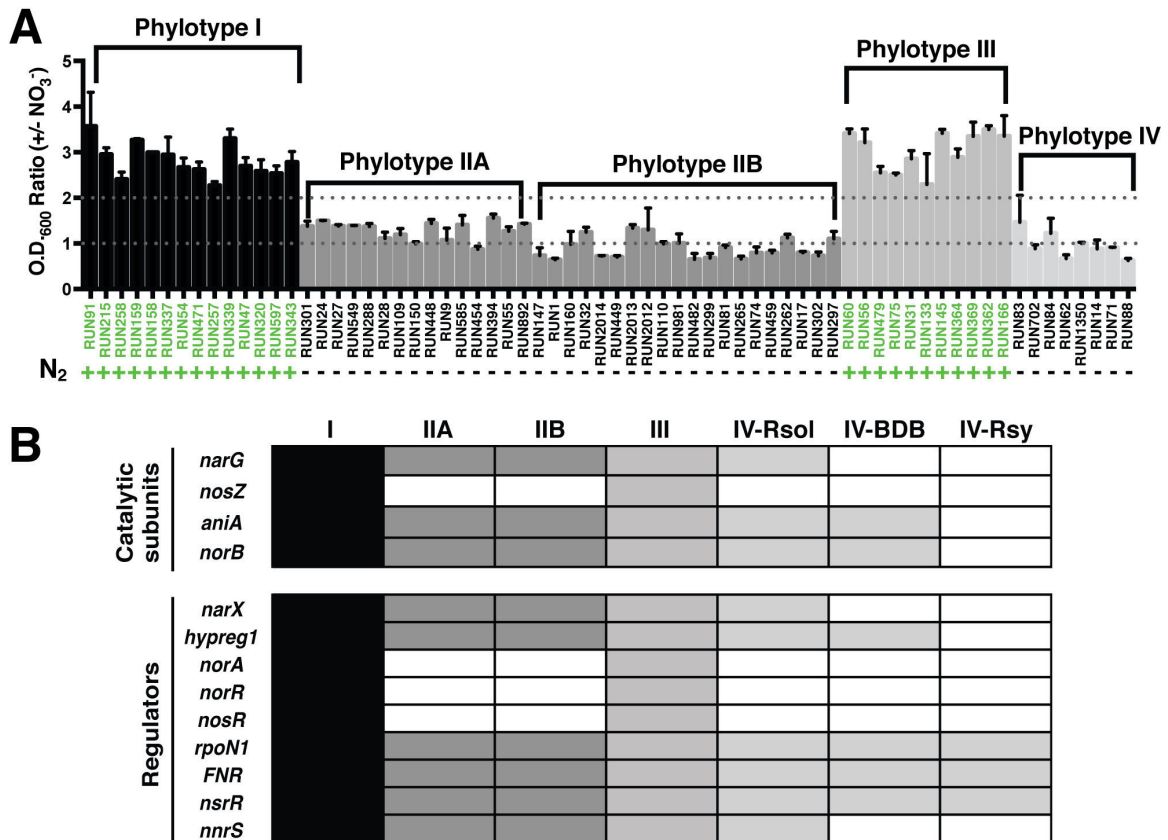


Figure 1 - Denitrification in RSSC A. Growth and production of nitrogen gas under anaerobic condition. Values represent the ratio of O.D.₆₀₀ readings following 72 hours of anaerobic incubation in VDM + NO₃⁻ vs - NO₃⁻. A value above 1 indicates that in the presence of NO₃⁻ a strain reached higher optical densities than in the absence of NO₃⁻, indicating NO₃⁻ respiration enhanced growth. A value above the arbitrary threshold of 2 meets our cut-off for proposed biological significance. Strain names in green and green '+' s indicate N₂ gas was produced within 96 hours of anaerobic inoculation in VDM + 30 mM NO₃⁻. Production of N₂ indicates that the strain completed the full denitrification pathway. Bars indicate standard error. Data represent the means of 4-6 biological replicates. B. Summary of the presence/absence of denitrification genes. Black cells indicate the presence of a gene in all the sequenced strains of this group while white cells indicate its absence.

211 nitric oxide (NO), nitrous oxide (N₂O), and finally nitrogen gas (N₂) in a series of four
212 reactions catalyzed by the products of the *narG*, *aniA*, *norB* and *nosZ* genes, respectively.

213 The production of N₂ reflects the nitrous oxide reductase activity encoded by *nosZ*.
214 Notably, this gene and nitrous oxide reductase activity are both absent from all phylotype
215 II and IV strains evaluated (**Fig. 1B**). Moreover, respiration on nitrate seems to be
216 dependent on *nosZ* since it is the only core catalytic subunit encoding gene missing from
217 the denitrification pathway in *R. solanacearum* strains evaluated from phylotypes II and
218 IV. However, the same is not necessarily true of phylotypes I and III, because a *nosZ*
219 deletion mutant can respire and grow on nitrate under anaerobic conditions [28]. Thus,
220 denitrification appears to be a complex process not yet fully understood. Indeed, up to
221 nine regulators and predicted inorganic nitrogen metabolism-related proteins are known
222 to be involved in denitrification in other organisms are found in RSSC. Some, like NosZ,
223 are present only in phylotypes I and III (NorA, NosR, NorR), and could be involved in
224 the expression of additional genes required for anaerobic respiration. For example, NorA
225 and NorR are both NO binding proteins in *R. eutropha* [33] and NO is an important
226 intermediate molecule produced prior to N₂ production in the denitrification pathway.
227 Furthermore, regulators present in all the RSSC do not exhibit a similar degree of
228 conservation across phylotypes. While the amino acid identities of RpoN1, Fnr, and NsrR
229 are greater than 90% across all phylotypes, the NnrS, NarX and Hypreg1 homologs only
230 share ~70% amino acid identities between phylotype I/III and phylotype II/IV. Further
231 characterization of these regulators would be required to better understand the observed
232 phenotypes.

233 Our functional studies of nitrogen metabolism conflict with those of Safni *et al.* in two
234 important respects. Safni *et al.* reported that most of the phylotype II strains and many
235 strains from phylotype IV produced dinitrogen gas. We did not detect N₂ gas production
236 from any tested phylotype II or IV strains. Considering that this function depends on
237 *nosZ*, a gene that is absent from all sequenced strains of these two phylotypes, the results
238 from the nitrate-to-gas experiment presented by Safni *et al.* likely reflect experimental
239 errors. Additionally, the authors stated that phylotype I strains are incapable of anaerobic
240 nitrate reduction to nitrite even though the same strains are described as producing N₂ gas
241 from nitrate; this is physiologically impossible because the production of dinitrogen gas
242 cannot occur without the initial conversion of nitrate to nitrite under the conditions tested
243 (nitrate as the sole terminal electron acceptor and inorganic nitrogen species). Our finding
244 that phylotype I and III strains can respire on nitrate anaerobically offers functional
245 evidence that they do reduce nitrate to nitrite. This experimental evidence is consistent
246 with the genomic data indicating the presence of NarG nitrate reductase in all sequenced
247 phylotype I strains. The formal descriptions of the species proposed by Safni *et al.* should
248 be revised to indicate that phylotype II and IV strains do not produce N₂ gas from nitrate,
249 and that phylotype I strains do reduce nitrate to nitrite.

250 *RSSC exhibits genomic diversity*

251 The 16S rRNA gene sequences of *R. solanacearum* strains are more than 97% identical,
252 suggesting that this group forms a single species that is distinct from its close relative *R.*
253 *eutropha* [4]. However, 16S rRNA sequences do not always accurately reflect similarities

254 at the whole-genome level and cannot distinguish between recently diverged species [34,
255 35]. Moreover, the identity threshold has not been universally accepted, and distinct
256 species with 98% identical 16S rRNA sequences have been described [36]. DDH has
257 been historically used for species delineation, and a 70% DNA–DNA similarity is
258 traditionally used to define species. Recently, Safni *et al.* provided evidence that *R.*
259 *solanacearum* can be divided into 3 species based on DDH values. In a complementary
260 approach, we evaluated the taxonomy of *R. solanacearum* using genomic and proteomic
261 data. As a method, DDH has significant drawbacks: it is technically difficult, is
262 performed only in a few specialized laboratories, and is prone to experimental errors [37].
263 DDH assays can only measure the potential for hybridization between purified DNA
264 from two organisms, without regard to biological function. Thus, the 70% DDH criterion
265 does not correspond to 70% shared orthologous genes or even 70% sequence identity
266 [38]. Strains showing more than 70% DDH can possess up to 21% divergent gene content,
267 equivalent to around 1000 genes in a typical 5.3 Mb *R. solanacearum* genome [39]. With
268 the recent development of *in silico* comparative methods using complete genome
269 sequences, DDH is no longer the best, most reliable method for determining sequence
270 identity.

271
272 In a previous study [18], we proposed division of the RSSC into 3 genomospecies based
273 on a genome-to-genome comparison using Average Nucleotide Identity (ANI) analyses
274 and 8 strains. In the present study, we included 15 additional genomes in the ANI
275 analysis and compared it to two recent methods showing a better correlation with DDH:
276 the Maximum Unique Matches index (MUMi) and the Genome-to-Genome Distance
277 Calculator (GGDC) [10, 21, 22, 24] (**Supplemental Table 1**).

278 Briefly, ANI detects the level of conservation or similarity of the total genomic sequences
279 shared between two strains based on the identification of homologous fragments of fixed
280 length using the BLAST algorithm. Strains with ANI >95% are considered as belonging
281 to the same species, consistent with the 70% DDH criterion [10, 21, 39-42]. Similar to
282 DDH, ANI accounts for the variability in conserved gene content and does not always
283 reflect differences between closely related strains. Thus, strains with similar ANI values
284 can have similar or dissimilar gene content during intraspecies comparison [43]. The
285 MUMi algorithm overcomes this issue and accounts for both the variability of
286 homologous gene content and the gain and loss of DNA. MUMi distances are derived
287 from a list of maximum unique matches (MUMs) of a given minimal length shared
288 between two genomes and the average length between genomes. Because this technique
289 uses a fast algorithm to detect MUMs, MUMi is significantly faster than ANI. A MUMi
290 value of 0.33 ± 0.03 corresponds to an ANI value of 95%. Moreover, the GGDC method
291 has been recently revised, showing the highest correlation with wet-lab DDH [24, 44].
292 The GGDC also infers *in silico* DDH values from genomic distances; therefore, a similar
293 70% threshold can be used. Although this method is based on principles similar to ANI
294 and MUMi, GGDC uses a different set of formulas to estimate genomic distances.

295 The ANI values obtained from pairwise comparisons between all genomes are presented
296 in Supplemental Table 1. The 29 strains analyzed in the present study fall into three
297 distinct groups. The first group includes strains from phylotypes I and III. The second

298 group comprises phylotype II strains, divided into subgroups IIA (containing the current
299 *R. solanacearum* type strain, K60^T) and IIB. The last group includes phylotype IV strains
300 (PSI07, BDB R229 and *R. syzygii* R24), originally described as a separate species before
301 the establishment of the species complex. These ANI results are consistent with a
302 previous analysis of a smaller group of genomes [18].

303 The genomic distances calculated using the MUMi algorithm are presented in
304 **Supplemental Table 1**. This method separates the RSSC into 3 or more species
305 depending on how strictly the 0.33 ± 0.03 criterion is applied. Consistent with the ANI
306 analysis, this method identifies two distinct species: one containing phylotype IV, and
307 one containing phylotypes I and III. However, the delineation of phylotype II as a single
308 species was not definitive, and the genomic distances were $< 0.33 \pm 0.03$ in 100% of the
309 strains within subgroups IIA and IIB. Notably, the distances between IIB strain UW551
310 and 5 of the 6 IIA strains were above the threshold (0.37-0.40). However, MUMi values
311 above 0.33 were also observed in other species. For example, although *Pseudomonas*
312 *syzygii* has an average intraspecies MUMi of 0.7, this species is considered distinct
313 from other *Pseudomonas*, with an average intragenus MUMi of 0.9 [22].

314 *In silico* DDH values inferred using the GGDC algorithm are presented in Supplemental
315 Table 1. According to the traditional 70% DDH criterion, GGDC distinguishes 5 species
316 within RSSC with phylotypes I, III and IV assigned to a single species and phylotype II
317 divided into 2 species corresponding to the IIA and IIB subgroups. Thus, GGDC more
318 clearly differentiates between closely related strains. Although GGDC divides the RSSC
319 into more species than ANI and MUMi, the raw result patterns are consistent between all
320 3 methods. As previously observed with the MUMi distances, IIA and IIB strains are
321 definitely divergent at the whole-genome level. The ANI values between phylotypes I
322 and III, or subtypes IIA and IIB, were approximately 96%, while the ANI values within
323 the species predicted using GGDC ranged from 97 to 99%. Moreover, a phylogenetic
324 network derived from the ANI, MUMi and GGDC distance matrices was built using
325 SplitsTree software. The results showed no obvious ambiguities, confirming that all three
326 genome sequence-based methods give broadly consistent results (**Fig. 2**).

327 The ANI, MUMi and GGDC methods are all based on whole-genome comparisons and
328 have been shown to correlate well with the traditional standard method, DDH.
329 Nonetheless, the RSSC can be divided into three to five species depending on how the
330 genomic distances are calculated and the criteria used. Taken together, the outputs of
331 these techniques illustrate the difficulty of consistently delineating species among closely
332 related strains. Based on phenotypic data, ANI and MUMi distances adequately reflect
333 the level of biological variability within the RSSC, with a three species division in which
334 phylotypes I and III cluster together, and phylotypes II and IV are further apart.

335 *RSSC exhibits proteome diversity*

336 A total of 73 bacterial strains representing the four phylotypes were subjected to
337 comparative proteomic analysis as a complementary method to the genome-based
338 analyses presented above. Protein mass spectra corresponding to each strain were
339 obtained using MALDI-TOF and clustered using SPECLUST software [45]. This

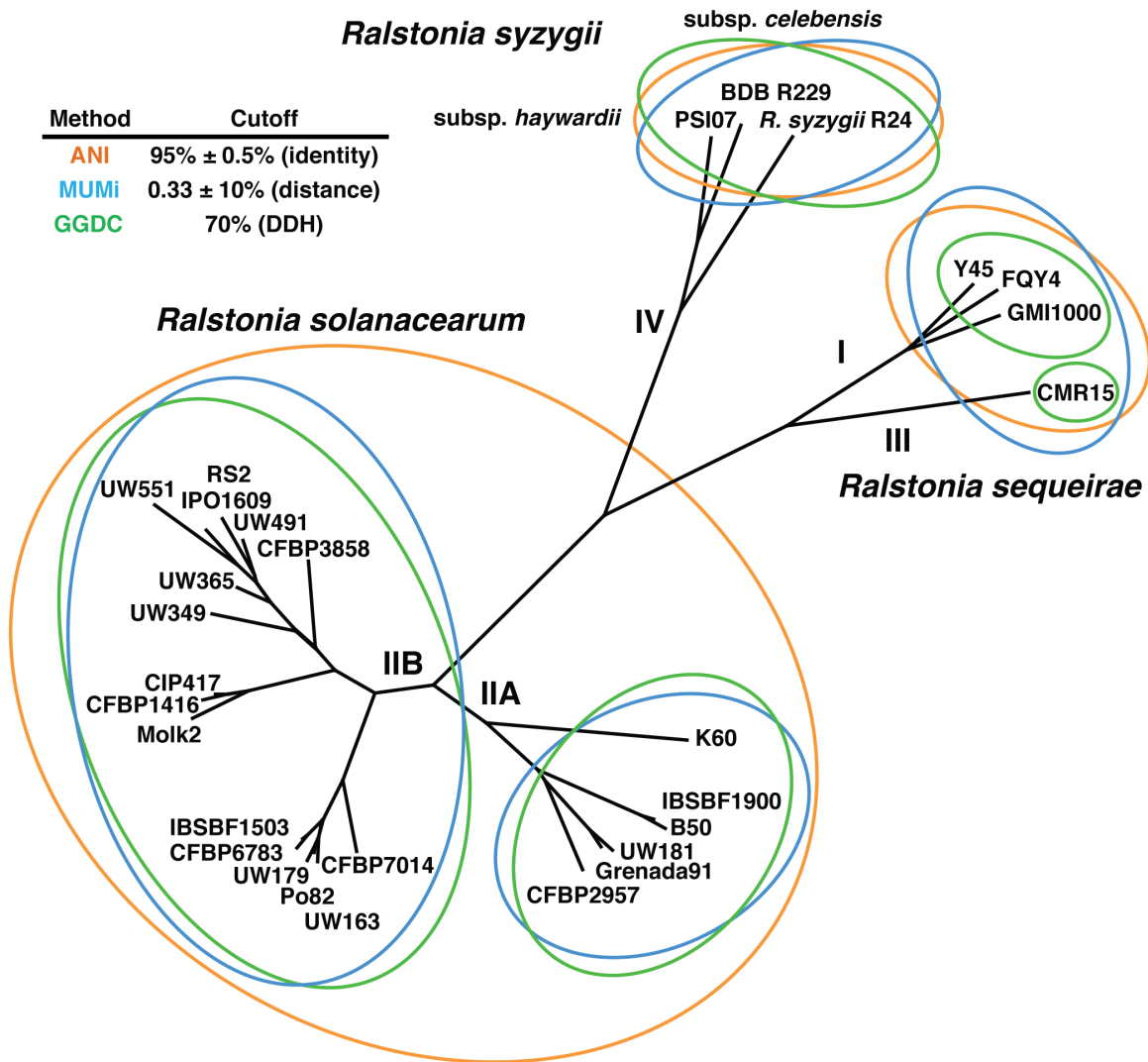


Figure 2 - Phylogenetic network derived from the distance matrices generated using all three methods and the SplitsTree4 software. The orange, blue and green cells represent strains clustered into species using the criteria specific to the ANI, MUMi and GGDC methods, respectively.

340 generated a list of common peaks represented as inter-sample consensus m/z values. The
341 best results were achieved using a “within peak match score (σ)” of 3 Da, as defined in
342 the SPECLUST documentation. The consensus spectra matrix was translated to a binary
343 matrix in which the absence/presence of a consensus peak in all strain profiles was
344 represented as 0 or 1, respectively. This binary matrix was used to infer the phylogenetic
345 relationships among the strains with the MALDI-TOF data and Bayesian analysis using
346 MrBayes v3.2.2 software [46].

347 The MALDI-TOF approach has previously been used for bacterial identification [47].
348 Mass fingerprinting has been used as a simple, quick and reproducible method for
349 bacterial identification through the generation of large spectral databases [48].

350
351 Taxonomically, molecular typing using protein profiles has been useful for bacterial
352 classification at the species and subspecies levels [49, 50] and at the strain level,
353 depending on the type and class of bacterial group considered [51]. In the present study, a
354 combination of MALDI-TOF profiling, consensus mass peak lists, and Bayesian
355 inference was used to cluster the 73 *Ralstonia* strains into three groups with strong branch
356 support. The first cluster contained phylotypes I and III, whereas the second cluster
357 contained phylotypes IIA and IIB, and the third cluster contained phylotype IV (**Fig. 3**).
358 These results were consistent with the findings of the genetic analysis, supporting the
359 division of *R. solanacearum* into three species.

360

361 *Phenotypic, genomic and proteomic data converge on a three-species model*

362 The taxonomic classification of *R. solanacearum* has changed repeatedly over the last 50
363 years, grouping strains with divergent ecological, geographical, genetic and phenotypic
364 profiles, including many pathological variants. Safni *et al.* recently used DDH on the
365 taxonomic revision of the RSSC, proposing the division of this complex into three
366 distinct species. Among other emerging classification methods, modern techniques, such
367 as 2nd generation sequencing, have provided great opportunities to understand the
368 phylogenetic relationships between strains and to question the relevance of the current
369 taxonomy of this species complex. Moreover, phenotypic data can be directly correlated
370 with genomic content in order to better understand the traits used to delineate species.
371 Using a combination of phenotypic analyses, whole-genome comparisons and proteomic
372 profiling, we provide additional information on the relationships between *R.*
373 *solanacearum* strains, supporting the proposed three-species delineation of the current
374 RSSC. The first species includes phylotypes I and III. The strains from these two
375 phylotypes undergo denitrification, among other unique phenotypic properties, and are
376 genetically closely related. They are, however, divergently adapted to temperatures
377 because phylotype III strains are primarily observed in temperate highlands, while
378 phylotype I strains are often isolated in tropical lowlands.

379 The second species corresponds to phylotype II. The strains in this phylotype present a
380 certain degree of genomic divergence, resulting in the vague delineation of species based
381 on genomic distances. Nonetheless, every strain belonging to phylotype II evaluated to-
382 date exhibits similar phenotypic properties and could therefore be considered a single
383 species. The third species comprises the phylotype IV strains currently classified as *R.*

384 *solanacearum*; *R. syzygii*, which is transmitted through tube-building *Hindola* spp.
385 cercopoid insects, with a host range limited to clove trees (Sumatra disease); and BDB,
386 the causative agent of banana wilt diseases in Indonesia. Despite their strikingly different
387 biological lifestyles [52], these strains are genetically related and display common core
388 metabolic activities. *R. solanacearum*, *R. syzygii* and BDB present different geographical
389 distributions and pathogenic potential. Moreover, because these groups are easily
390 genetically distinguishable, the members of this third group could be considered
391 subspecies.

392 **Conclusions**

393 Extensive biological, phenotypic, and genetic data demonstrate that *R. solanacearum* is
394 too diverse to be considered a single species. The modification of the taxonomy of this
395 organism is necessary to recognize three phylogenetically distinct groups with different
396 biological properties and evolutionary relationships. Newly isolated bacterial wilt strains
397 can readily be assigned to the proposed scheme using existing molecular methods [13].
398 These changes will benefit many different applications, including breeding plant
399 resistance to bacterial wilt, the identification of new pathological variants, management
400 of quarantine containment and the development of diagnostic tests.

401 **Acknowledgments**

402 We gratefully acknowledge Boris Vitnazer for providing some of the IIB-1 genomes and
403 Carolee Bull for valuable discussions. This research was supported through INRA,
404 CIRAD and the University of Wisconsin-Madison College of Agricultural and Life
405 Sciences. The authors would like to thank Aurélie Lajus and Claudine Médigue
406 (Genoscope, France) for technical support and A. Michault from CHU de La Réunion
407 Groupe Hospitalier Sud Réunion for generously providing the MALDI-TOF-MS
408 instrument. Borja Sánchez received funding from the INRA and a European Union
409 AgreenSkills postdoctoral fellowship (grant agreement n° 267196).

410 **References**

411

412 [1] T. Denny, Plant pathogenic *Ralstonia* species, in: S.S. Gnanamanickam (Ed.) Plant-
413 Associated Bacteria, Springer Netherlands, Dordrecht, The Netherlands, 2006, pp. 573-
414 644.

415 [2] A.C. Hayward, Biology and epidemiology of bacterial wilt caused by *Pseudomonas*
416 *solanacearum*, Annu Rev Phytopathol, 29 (1991) 65-87.

417 [3] E.F. Smith, A bacterial disease of tomato, eggplant, and Irish potato (*Bacterium*
418 *solanacearum* nov. sp.). USDA Div. Veg. Physiol. Pathol. Bull., 12:191 (1896).

419 [4] E. Yabuuchi, Y. Kosako, I. Yano, H. Hotta, Y. Nishiuchi, Transfer of two
420 *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. Nov.: Proposal of *Ralstonia*
421 *pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. Nov., *Ralstonia solanacearum*
422 (Smith 1896) comb. Nov. and *Ralstonia eutropha* (Davis 1969) comb. Nov, Microbiol
423 Immunol, 39 (1995) 897-904.

424 [5] S. Eden-Green, Banana Blood Disease, INIBAP Musa Disease Fact Sheet No.3,
425 (1994).

426 [6] E. Gäumann, Onderzoekingen over de bloedziekte der bananen op Celebes I,
427 Mededelingen van het Instituut voor Plantenziekten, 50 (1921) 47.

428 [7] M. Vaneechoutte, P. Kampfer, T. De Baere, E. Falsen, G. Verschraegen, Wautersia
429 gen. nov., a novel genus accommodating the phylogenetic lineage including *Ralstonia*
430 *eutropha* and related species, and proposal of *Ralstonia* [*Pseudomonas*] *syzygii* (Roberts
431 et al. 1990) comb. nov, Int J Syst Evol Microbiol, 54 (2004) 317-327.

432 [8] Z. Li, S. Wu, X. Bai, Y. Liu, J. Lu, Y. Liu, B. Xiao, X. Lu, L. Fan, Genome sequence
433 of the tobacco bacterial wilt pathogen *Ralstonia solanacearum*, J Bacteriol, 193 (2011)
434 6088-6089.

435 [9] J. Chun, F.A. Rainey, Integrating genomics into the taxonomy and systematics of the
436 Bacteria and Archaea, Int J Syst Evol Microbiol, 64 (2014) 316-324.

437 [10] M. Richter, R. Rossello-Mora, Shifting the genomic gold standard for the
438 prokaryotic species definition, Proc Natl Acad Sci U S A, 106 (2009) 19126-19131.

439 [11] G. Cellier, P. Prior, Deciphering phenotypic diversity of *Ralstonia solanacearum*
440 strains pathogenic to potato, Phytopathology, 100 (2010) 1250-1261.

441 [12] A. Guidot, P. Prior, J. Schoenfeld, S. Carrere, S. Genin, C. Boucher, Genomic
442 structure and phylogeny of the plant pathogen *Ralstonia solanacearum* inferred from
443 gene distribution analysis, J Bacteriol, 189 (2007) 377-387.

444 [13] M. Fegan, P. Prior, How complex is the "*Ralstonia solanacearum* species complex",
445 in: C. Allen, P. Prior, A.C. Hayward (Eds.) Bacterial wilt disease and the *Ralstonia*
446 *solanacearum* species complex, APS Press, St Paul, MN, 2005, pp. 449-461.

447 [14] P. Prior, M. Fegan, Recent developments in the phylogeny and classification of
448 *Ralstonia solanacearum*, Acta Horticulturae, 695 (2005) 127-136.

449 [15] J.A. Castillo, J.T. Greenberg, Evolutionary dynamics of *Ralstonia solanacearum*,
450 Appl Environ Microbiol, 73 (2007) 1225-1238.

451 [16] A. N'Guessan C, S. Brisse, A.C. Le Roux-Nio, S. Poussier, D. Kone, E. Wicker,
452 Development of variable number of tandem repeats typing schemes for *Ralstonia*
453 *solanacearum*, the agent of bacterial wilt, banana Moko disease and potato brown rot,
454 Journal of Microbiological Methods, 92 (2013) 366-374.

455 [17] D.W. Gabriel, C. Allen, M. Schell, T.P. Denny, J.T. Greenberg, Y.P. Duan, Z.
456 Flores-Cruz, Q. Huang, J.M. Clifford, G. Presting, E.T. Gonzalez, J. Reddy, J.
457 Elphinstone, J. Swanson, J. Yao, V. Mulholland, L. Liu, W. Farmerie, M. Patnaikuni, B.
458 Balogh, D. Norman, A. Alvarez, J.A. Castillo, J. Jones, G. Saddler, T. Walunas, A.
459 Zhukov, N. Mikhailova, Identification of open reading frames unique to a select agent:
460 *Ralstonia solanacearum* race 3 biovar 2, Mol Plant Microbe Interact, 19 (2006) 69-79.
461 [18] B. Remenant, J.C. de Cambiaire, G. Cellier, J.M. Jacobs, S. Mangenot, V. Barbe, A.
462 Lajus, D. Vallenet, C. Medigue, M. Fegan, C. Allen, P. Prior, *Ralstonia syzygii*, the
463 Blood Disease Bacterium and some Asian *R. solanacearum* strains form a single genomic
464 species despite divergent lifestyles, PLoS One, 6 (2011) e24356.
465 [19] M. Salanoubat, S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A.
466 Billault, P. Brottier, J.C. Camus, L. Cattolico, M. Chandler, N. Choisne, C. Claudel-
467 Renard, S. Cunnac, N. Demange, C. Gaspin, M. Lavie, A. Moisan, C. Robert, W. Saurin,
468 T. Schiex, P. Siguier, P. Thebault, M. Whalen, P. Wincker, M. Levy, J. Weissenbach,
469 C.A. Boucher, Genome sequence of the plant pathogen *Ralstonia solanacearum*, Nature,
470 415 (2002) 497-502.
471 [20] I. Safni, I. Cleenwerck, P. De Vos, M. Fegan, L. Sly, U. Kappler, Polyphasic
472 taxonomic revision of the *Ralstonia solanacearum* species complex: proposal to emend
473 the descriptions of *Ralstonia solanacearum* and *Ralstonia syzygii* and reclassify current *R.*
474 *syzygii* strains as *Ralstonia syzygii* subsp. *syzygii* subsp. nov., *R. solanacearum* phylotype
475 IV strains as *Ralstonia syzygii* subsp. *indonesiensis* subsp. nov., banana blood disease
476 bacterium strains as *Ralstonia syzygii* subsp. *celebesensis* subsp. nov. and *R.*
477 *solanacearum* phylotype I and III strains as *Ralstonia pseudosolanacearum* sp. nov, Int J
478 Syst Evol Microbiol, 64 (2014) 3087-3103.
479 [21] K.T. Konstantinidis, J.M. Tiedje, Genomic insights that advance the species
480 definition for prokaryotes, Proc Natl Acad Sci U S A, 102 (2005) 2567-2572.
481 [22] M. Deloger, M. El Karoui, M.A. Petit, A genomic distance based on MUM indicates
482 discontinuity between most bacterial species and genera, J Bacteriol, 191 (2009) 91-99.
483 [23] S. Kurtz, A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S.L.
484 Salzberg, Versatile and open software for comparing large genomes, Genome Biol, 5
485 (2004) R12.
486 [24] J.P. Meier-Kolthoff, A.F. Auch, H.P. Klenk, M. Goker, Genome sequence-based
487 species delimitation with confidence intervals and improved distance functions, BMC
488 Bioinformatics, 14 (2013) 60.
489 [25] D.H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary
490 studies, Mol Biol Evol, 23 (2006) 254-267.
491 [26] A.C. Hayward, H.M. El-Nashaar, U. Nydegger, L. De Lindo, Variation in nitrate
492 metabolism in biovars of *Pseudomonas solanacearum*, Journal of Applied Bacteriology,
493 69 (1990) 269-280.
494 [27] M. Van den Mooter, M. Steenackers, C. Maertens, F. Gossele, P. de Vos, J. Swings,
495 K. Kersters, J. de Ley, Differentiation between *Xanthomonas campestris* pv. *graminis*
496 (ISPP List 1980), pv. *phleipratensis* (ISPP List 1980) emend., pv. *poae* Egli and Schmidt
497 1982 and pv. *arrhenatheri* Egli and Schmidt 1982, by Numerical Analysis of Phenotypic
498 Features and Protein Gel Electrophoregrams, J Phytopathol, 118: 135–156 (1987).

499 [28] B. Dalsing, A. Truchon, E. Gonzalez-Orta, A. Milling, C. Allen, *Ralstonia*
500 *solanacearum* uses inorganic nitrogen metabolism for virulence, ATP production, and
501 detoxification in the oxygen-limited host xylem environment, *mBio*, (2015).
502 [29] W.G. Zumft, Cell biology and molecular basis of denitrification, *Microbiol Mol Biol*
503 *Rev*, 61 (1997) 533-616.
504 [30] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L.
505 Madden, BLAST+: architecture and applications, *BMC Bioinformatics*, 10 (2009) 421.
506 [31] D. Charif, J.R. Lobry, SeqinR 1.0-2: a contributed package to the R project for
507 statistical computing devoted to biological sequences retrieval and analysis., in:
508 U.B.a.M.P.a.H.E.R.a.M. Vendruscolo (Ed.) Structural approaches to sequence evolution:
509 Molecules, networks, populations, Springer Verlag, 2007, pp. 207-232.
510 [32] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high
511 throughput, *Nucleic Acids Res*, 32 (2004) 1792-1797.
512 [33] R. Cramm, A. Busch, K. Strube, NO-dependent transcriptional activation of gene
513 expression in *Ralstonia eutropha* H16, *Biochem Soc Trans*, 34 (2006) 182-184.
514 [34] R. Rossello-Mora, R. Amann, The species concept for prokaryotes, *FEMS Microbiol*
515 *Rev*, 25 (2001) 39-67.
516 [35] J.M. Janda, S.L. Abbott, 16S rRNA gene sequencing for bacterial identification in
517 the diagnostic laboratory: pluses, perils, and pitfalls, *J Clin Microbiol*, 45 (2007) 2761-
518 2764.
519 [36] M. Achtman, M. Wagner, Microbial diversity and the genetic nature of microbial
520 species, *Nat Rev Microbiol*, 6 (2008) 431-440.
521 [37] A.F. Auch, M. von Jan, H.P. Klenk, M. Goker, Digital DNA-DNA hybridization for
522 microbial species delineation by means of genome-to-genome sequence comparison,
523 *Stand Genomic Sci*, 2 (2010) 117-134.
524 [38] R. Rossello-Mora, Updating prokaryotic taxonomy, *J Bacteriol*, 187 (2005) 6255-
525 6257.
526 [39] J. Goris, K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme, J.M.
527 Tiedje, DNA-DNA hybridization values and their relationship to whole-genome sequence
528 similarities, *Int J Syst Evol Microbiol*, 57 (2007) 81-91.
529 [40] B.J. Haley, C.J. Grim, N.A. Hasan, S.Y. Choi, J. Chun, T.S. Brettin, D.C. Bruce, J.F.
530 Challacombe, J.C. Detter, C.S. Han, A. Huq, R.R. Colwell, Comparative genomic
531 analysis reveals evidence of two novel *Vibrio* species closely related to *V. cholerae*,
532 *BMC Microbiol*, 10 (2010) 154.
533 [41] M. Scortichini, S. Marcelletti, P. Ferrante, G. Firrao, A Genomic redefinition of
534 *Pseudomonas avellanae* species, *PLoS One*, 8 (2013) e75794.
535 [42] M. Kim, H.S. Oh, S.C. Park, J. Chun, Towards a taxonomic coherence between
536 average nucleotide identity and 16S rRNA gene sequence similarity for species
537 demarcation of prokaryotes, *Int J Syst Evol Microbiol*, 64 (2014) 346-351.
538 [43] K.T. Konstantinidis, A. Ramette, J.M. Tiedje, The bacterial species definition in the
539 genomic era, *Philos Trans R Soc Lond B Biol Sci*, 361 (2006) 1929-1940.
540 [44] S.R. Henz, D.H. Huson, A.F. Auch, K. Nieselt-Struwe, S.C. Schuster, Whole-
541 genome prokaryotic phylogeny, *Bioinformatics*, 21 (2005) 2329-2335.
542 [45] R. Alm, P. Johansson, K. Hjerno, C. Emanuelsson, M. Ringner, J. Hakkinen,
543 Detection and identification of protein isoforms using cluster analysis of MALDI-MS
544 mass spectra, *J Proteome Res*, 5 (2006) 785-792.

- 545 [46] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under
546 mixed models, *Bioinformatics*, 19 (2003) 1572-1574.
- 547 [47] K. Böhme, I.C. Fernandez-No, J. Barros-Velazquez, J. Gallardo, B. Cañas, P. Calo-
548 Mata, Comparative analysis of protein extraction methods for the identification of
549 seafood-borne pathogenic and spoilage bacteria by MALDI-TOF mass spectrometry.,
550 *Anal. Methods*, 2 (2010) 1941–1947.
- 551 [48] N. Degand, E. Carbonnelle, B. Dauphin, J.L. Beretti, M. Le Bourgeois, I. Sermet-
552 Gaudelus, C. Segonds, P. Berche, X. Nassif, A. Ferroni, Matrix-assisted laser desorption
553 ionization-time of flight mass spectrometry for identification of nonfermenting gram-
554 negative bacilli isolated from cystic fibrosis patients, *J Clin Microbiol*, 46 (2008) 3361-
555 3367.
- 556 [49] P.R. Murray, Matrix-assisted laser desorption ionization time-of-flight mass
557 spectrometry: usefulness for taxonomy and epidemiology, *Clin Microbiol Infect*, 16
558 (2010) 1626-1630.
- 559 [50] H.J. Ng, H.K. Webb, R.J. Crawford, F. Malherbe, H. Butt, R. Knight, V.V.
560 Mikhailov, E.P. Ivanova, Updating the taxonomic toolbox: classification of *Alteromonas*
561 spp. using multilocus phylogenetic analysis and MALDI-TOF mass spectrometry,
562 *Antonie Van Leeuwenhoek*, 103 (2013) 265-275.
- 563 [51] T.R. Sandrin, J.E. Goldstein, S. Schumaker, MALDI TOF MS profiling of bacteria
564 at the strain level: a review, *Mass Spectrom Rev*, 32 (2013) 188-217.
- 565 [52] B. Remenant, B. Coupat-Goutaland, A. Guidot, G. Cellier, E. Wicker, C. Allen, M.
566 Fegan, O. Pruvost, M. Elbaz, A. Calteau, G. Salvignol, D. Mornico, S. Mangenot, V.
567 Barbe, C. Medigue, P. Prior, Genomes of three tomato pathogens within the *Ralstonia*
568 *solanacearum* species complex reveal significant evolutionary divergence, *BMC*
569 *Genomics*, 11 (2010) 379.

570

571

572

573 **Appendices**574 **Table 1.** *Ralstonia* spp. strains used in whole-genome analyses.

Strain	Phy.-Seq.	Isolated from	Geographic origin	Acc. #
GMI1000	I	Tomato	Guyana	NC_003295 NC_003296
FQY_4	I	Soil	China	CP004012 CP004013
Y45	I	Tobacco	China	AFWL00000000
IPO1609	IIB-1	Potato	Netherlands	CU914168, CU914166
UW551	IIB-1	Geranium	Kenya	AAKL00000000
UW349	IIB-1	Potato	Brazil	JQOI00000000.1
UW365	IIB-1	Potato	China	JQSI00000000.1
UW491	IIB-1	Potato	Colombia	JQSH00000000.1
RS2	IIB-1	Potato	N/D	EMBL: PRJEB8309
CFBP3858	IIB-1	Potato	Netherlands	EMBL: PRJEB8309
MolK2	IIB-3	Banana	Philippines	CAHW01000040
CFBP1416	IIB-3	Plantain	Costa Rica	EMBL: PRJEB7434
CIP417	IIB-3	Banana	Philippines	EMBL: PRJEB7427
UW179	IIB-4	Banana	Colombia	EMBL: PRJEB7426
UW163	IIB-4	Plantain	Peru	EMBL: PRJEB7430
CFBP6783	IIB-4	Heliconia	French West Indies	EMBL: PRJEB7432
Po82	IIB-4	Potato	Mexico	CP002819 CP002820
IBSBF1503	IIB-4	Cucumber	Brazil	EMBL: PRJEB7433
CFBP7014	IIB-59	Anthurium	Trinidad	EMBL: PRJEB8309
CFBP2957	IIA-36	Tomato	French West Indies	EMBL: FP885897 EMBL: FP885907
K60 ^T	IIA-7	Tomato	United States	EMBL: CAGT01000001
Grenada 9-1	IIA-6	Banana	Grenada	EMBL: PRJEB7428
UW181	IIA-6	Plantain	Venezuela	EMBL: PRJEB8309
B50	IIA-24	Banana	Brazil	EMBL: PRJEB7421
IBSBF1900	IIA-24	Banana	Brazil	EMBL: PRJEB8309
CMR15	III	Tomato	Cameroon	EMBL: FP885895 EMBL: FP885896
PSI07	IV	Tomato	Indonesia	EMBL: FP885906 EMBL: FP885891
BDB R229	IV	Banana	Indonesia	EMBL: FR854059 to FR854085
<i>R. syzygii</i> R24	IV	Clove	Indonesia	EMBL: FR854086 to FR854092

575

576

577 **Figure 1. A.** Growth and production of nitrogen gas under anaerobic condition. Values
578 represent the ratio of O.D.₆₀₀ readings following 72 hours of anaerobic incubation in
579 VDM plus 30 mM NO₃⁻ vs. without added NO₃⁻. A value above 1 indicates that in the
580 presence of NO₃⁻ a strain reached higher optical densities than in the absence of NO₃⁻,
581 indicating NO₃⁻ respiration enhanced growth. A value above the arbitrary threshold of 2
582 meets our cut-off for biological significance. Strain names in green and green '+' s
583 indicate N₂ gas was produced within 96 hours of anaerobic inoculation in VDM + 30 mM
584 NO₃⁻. Production of N₂ indicates that the strain completed the full denitrification pathway
585 as indicated by production of visible N₂ gas bubbles. Bars indicate standard error. Data
586 represent the means of 4-6 biological replicates. **B.** Summary of the presence/absence of
587 denitrification genes. Black/gray cells indicate the presence of a gene in all the sequenced
588 strains of this group and white cells indicate its absence.

589 **Figure 2.** Phylogenetic network derived from the distance matrices generated with ANI,
590 MUMi and GGDC methods using the SplitsTree4 software. The orange, blue and green
591 cells represent strains clustered into species using the criteria specific to the ANI, MUMi
592 and GGDC methods, respectively.

593 **Figure 3.** Majority-rule consensus tree based on the presence/absence of a consensus
594 MALDI-TOF peak list obtained using the MrBayes software. The probability values are
595 indicated along the main branches. Red colored strains indicate the sequenced strains.
596 Black lines delineate strain clusters.

597 **Supplemental Table 1.** Genomes used in the present study. Pairwise comparisons of 28
598 sequenced genomes from the *R. solanacearum* species complex using the ANI, MUMi
599 and GGDC methods. The orange, blue and green cells represent strains clustered into
600 species using the criteria specific to each method.

601 **Supplemental Table 2.** Bacterial strains used for proteomic analysis. Phylotype and
602 sequevar classifications for the strains in the *R. solanacearum* species complex were
603 determined as previously described [13]. The sequenced strains are highlighted in green.

604

Conclusion

Ce travail soutient donc fortement la révision taxonomique du complexe d'espèces et complémente les méthodes historiques utilisées par Safni *et al.* en s'appuyant sur des techniques plus modernes, tirant notamment partie de la disponibilité croissante des séquences génomiques de *R. solanacearum*.

Les conclusions de cette étude reposent principalement sur les analyses de distances génomiques. Les résultats ont démontré l'élévation de *R. solanacearum* en plusieurs espèces à l'aide de valeurs seuils corrélant avec la DDH, en définitive. Bien que compatibles, les délimitations proposées par chaque algorithme sont néanmoins différentes (**Figure 19**). De manière intéressante, la technique la plus récente se montre la plus discriminante tandis que la plus ancienne est très conservatrice. Effectivement, GGDC classe, par exemple, le phylotype I et III comme des espèces distinctes alors qu'ANI les regroupe en une seule. Les données phénotypiques disponibles dans la littérature ainsi que celles produites par Safni *et al.* sont en faveur d'un schéma à trois espèces tel que celui proposé par ANI. Dans le cas de *R. solanacearum*, ANI semble donc être la distance génomique reflétant le mieux la réelle diversité biologique du complexe d'espèce. Les résultats fournis par GGDC n'en sont pas moins valides et illustrent quant à eux la forte diversité génétique de *R. solanacearum*, et plus généralement la complexité de la notion d'espèce chez les procaryotes.

La révision taxonomique chez *R. solanacearum* présente un intérêt multidisciplinaire. En agronomie, dans le domaine de la sélection variétale, une séparation distincte des espèces permettra de mieux définir le spectre de résistance des cultivars commerciaux. En effet, la plupart des obtentions variétales pour une résistance forte vis-à-vis de *R. solanacearum* n'expriment quasiment jamais cette propriété de résistance en réponse à l'ensemble des souches dans chacun des phylotypes. Par exemple, des variétés données comme résistantes au flétrissement bactérien se révèlent à l'usage être parfaitement sensibles à des souches endémiques autres que celles ayant contribué à sélectionner pour ce caractère.

De la même manière, cette résistance peut être contournée lors de l'introduction de matériel végétal contaminé par un autre phylotype. Ce flou concernant l'étendue de la résistance des variétés commerciales peut favoriser la survenue d'épidémie et est en partie due à la méconnaissance de la diversité génétique des souches de *R. solanacearum* dans les régions pour lesquelles ces variétés sont développées.

La classification proposée permettra également une clarification dans les textes légiférant sur les organismes de quarantaines. En dehors de l'écotype Brown rot, peu de distinction est faite au plan officiel entre les différents sous-groupes de *R. solanacearum*. Une division en plusieurs espèces facilitera également le suivi épidémiologique des souches et contribuera à éclaircir les routes d'invasion, les sources d'introduction et la dynamique de leur circulation.

Chapitre 3 – Génomique comparative chez les écotypes du phylotype II

Les mécanismes déterminants la gamme d'hôte ont toujours été au centre des questions de recherche dans le domaine de la pathologie végétale en général, et chez l'agent du flétrissement bactérien en particulier. De part sa capacité à infecter un nombre important d'espèces végétales par rapport aux phytopathogènes bactériens, fongiques ou viraux, *R. solanacearum* constitue un modèle expérimental de choix pour étudier l'adaptation à l'hôte. Ces dernières années, la communauté a cependant rencontré plusieurs barrières freinant la progression des connaissances sur le spectre d'hôte, malgré la disponibilité de collections de souches renfermant une importante diversité aussi bien génétique, phénotypique ou géographique. Les mécanismes fondamentaux du flétrissement bactérien ont été extensivement caractérisés au cours des avancées en biologie moléculaire, mais de nombreux points restent encore à approfondir. Le cycle infectieux chez *R. solanacearum* est présenté en deux étapes, une étape précoce et une autre tardive, étapes pendant lesquelles différents facteurs de virulences sont exprimés, la transition entre ces étapes étant déterminée par des systèmes de 'quorum sensing'. Des études récentes sur l'expression des régulateurs centraux de la virulence ont cependant remis en question le schéma précédemment établi (Jacobs et al., 2012 ; Zuluaga et al., 2013). La découverte des effecteurs de type III et du système de sécrétion de type III chez de nombreux agents phytopathogènes a permis d'identifier diverses stratégies impliquées dans la pathogénicité, la subversion des défenses de la plante et la modification du spectre d'hôte. Le répertoire d'effecteur chez *R. solanacearum* est en moyenne plus important que chez d'autres espèces bactériennes et un faible nombre d'effecteurs est conservé chez toutes les souches (**Figure 20**). De plus, les effecteurs possèdent souvent des effets redondants ou cumulatifs compliquant alors la caractérisation du rôle précis d'un effecteur unique (Genin and Denny, 2012).



Figure 20 - Matrice illustrant la présence/absence des facteurs de virulence au sein du RSSC réalisée à partir des génomes séquencés. Chaque ligne correspond à un génome et chaque colonne correspond à un facteur de virulence (la matrice détaillée est disponible dans données supplémentaires de la publication associée a ce chapitre).

Bien qu'il soit avéré que *R. solanacearum* possède une vaste gamme d'hôte à l'échelle de l'espèce, le potentiel infectieux de souches individuelles a rarement été évalué chez plus de trois ou quatre hôtes simultanément. Les travaux de Lebeau *et al.* (2011) ont démontré que les différents niveaux de virulence et d'agressivité entre souches, combinés à divers degrés de résistances de l'hôte, peuvent compliquer la désignation d'une souche donnée comme pathogène sur tel ou tel hôte (Lebeau *et al.*, 2011). Au contraire, les travaux de Cellier *et al.* (2010) ont contribué à prouver la forte association entre des groupes génétiques du complexe d'espèces et certains hôtes (Cellier and Prior, 2010). Sur les bases de ces études, plusieurs souches représentatives de ces groupes génétiques ont été séquencées avec des technologies de 2nde génération. Les groupes concernés sont :

- i.** Brown rot : souches IIB-1 adaptées à la pommes de terre en climat tempéré. Ce groupe est particulièrement étudié par des équipes Nord Américaines compte tenu de son statut d'agent de quarantaine et d'agent potentiel de bioterrorisme. Un génome complet préexistait (UW551) et un génome draft nous a été fourni par Boris Vinatzer.
- ii.** Moko : groupe polyphylétique composé des 4 lignées IIB-3, IIB-4, IIA-6 et IIA-24 pathogènes du bananier. Un génome draft d'une souche IIB-3 (Molk2) était préexistant et des génomes drafts pour les trois autres lignées ont été produit dans cette étude.
- iii.** NPB (Non Pathogène du Bananier) : groupe phylogénétiquement proche de la lignée Moko IIB-4 sans aucun pouvoir pathogène sur le bananier, mais variant particulièrement virulent sur certaines Cucurbitacées et Solanacées. Aucune souche de ce groupe n'avait été séquencée auparavant.

La production de génomes de souches provenant de groupes et de lignées encore non séquencés représentait donc une étape indispensable pour étudier l'adaptation à l'hôte. En effet, les groupes Brown rot et NPB sont tout deux phylogénétiquement proches de certaines lignées Moko. Notre hypothèse de départ stipulait que la proximité génétique entre ces différents groupes de souches combinées aux différences majeures entre leur spectre d'hôtes nous permettrait d'identifier des

gènes directement impliqués dans la spécificité d'hôte, en comparant leurs génomes *in silico*. La difficulté principale rencontrée lors cette étude a été l'analyse des génomes drafts. Les méthodes de séquençage de 2nde génération fournissent rapidement et à faible coût une séquence représentant la quasi-intégralité du génome d'une souche avec un très faible taux d'erreur, mais cette dernière demeure sous la forme de plusieurs contigs distincts et ne permet pas, par exemple, de distinguer les deux réplicons (chromosome et mégaplasmide) dans le cas de *R. solanacearum*. La séparation en contigs signifie également que certains gènes peuvent être incomplets car localisés en bordure de ces contigs.

L'ENSEMBLE DE CES TRAVAUX DE GENOMIQUE COMPARATIVE ONT FAIT L'OBJET D'UNE PUBLICATION ACCEPTEE LE 14 FEVRIER 2015 DANS LA REVUE INTERNATIONALE BMC GENOMICS [IF 4,04] : « COMPARATIVE GENOMIC ANALYSIS OF *RALSTONIA SOLANACEARUM* REVEALS CANDIDATE GENES FOR HOST SPECIFICITY».

RESEARCH ARTICLE

Open Access

Comparative genomic analysis of *Ralstonia solanacearum* reveals candidate genes for host specificity

Florent Ailloud^{1,3}, Tiffany Lowe², Gilles Cellier³, David Roche⁴, Caitilyn Allen² and Philippe Prior^{1,5*}

Abstract

Background: *Ralstonia solanacearum* is a vascular soil-borne plant pathogen with an unusually broad host range. This economically destructive and globally distributed bacterium has thousands of distinct lineages within a heterogeneous and taxonomically disputed species complex. Some lineages include highly host-adapted strains (ecotypes), such as the banana Moko disease-causing strains, the cold-tolerant potato brown rot strains (also known as R3bv2) and the recently emerged Not Pathogenic to Banana (NPB) strains.

Results: These distinct ecotypes offer a robust model to study host adaptation and the emergence of ecotypes because the polyphyletic Moko strains include lineages that are phylogenetically close to the monophyletic brown rot and NPB strains. Draft genomes of eight new strains belonging to these three model ecotypes were produced to complement the eleven publicly available *R. solanacearum* genomes. Using a suite of bioinformatics methods, we searched for genetic and evolutionary features that distinguish ecotypes and propose specific hypotheses concerning mechanisms of host adaptation in the *R. solanacearum* species complex. Genome-wide, few differences were identified, but gene loss events, non-synonymous polymorphisms, and horizontal gene transfer were identified among type III effectors and were associated with host range differences.

Conclusions: This extensive comparative genomics analysis uncovered relatively few divergent features among closely related strains with contrasting biological characteristics; however, several virulence factors were associated with the emergence of Moko, NPB and brown rot and could explain host adaptation.

Keywords: *Ralstonia solanacearum*, Comparative genomics, Host adaptation

Background

Although many plant pathogens are narrowly adapted to one or a few related plant hosts, *Ralstonia solanacearum* has an unusually broad host range that includes monocotyledonous and dicotyledonous hosts [1]. *R. solanacearum* infects plant vascular systems, causing diverse bacterial wilt diseases. Its extensive host range, together with a wide geographic distribution, makes it one of the world's most destructive crop pathogens [2]. *R. solanacearum* has many distinct lineages within a heterogeneous and taxonomically disputed species complex. Extensive phylogenetic analysis demonstrates that the species complex contains four

phylotypes that correspond to geographic origin; phylotype I strains originated in Asia, phylotype II strains came from the Americas, phylotype III strains originated in Africa, and phylotype IV strains are from the Indonesian archipelago [3]. Within each phylotype, strains can be further subclassified into sequevars based on the similarity of a 750-bp fragment of the endoglucases (*egl*) gene [3]. *R. solanacearum* strains share a conserved core genome that is presumably essential for their common biology: colonizing plant xylem vessels and causing wilt symptoms [4,5]. However, the group's pan-genome contains over 16,000 coding sequences (CDSs), and individual strains in the species complex vary enormously with respect to both epidemiology and host range. For example, the *R. solanacearum* species complex includes strains adapted to eucalyptus, mulberry, clove, ginger, banana, peanut, and solanaceous plants [1].

* Correspondence: philippe.prior@cirad.fr

¹CIRAD, UMR PVBMT, F-97410 Saint-Pierre, La Réunion, France

⁵Département de Santé des Plantes et Environnement, (SPE) Inra, Paris, France

Full list of author information is available at the end of the article

Functional genetic analyses have identified many of *R. solanacearum*'s broadly conserved common virulence mechanisms [6,7]. However, the molecular processes that are responsible for strains' adaptation to particular hosts are not well understood. A few studies have identified specific type III effectors and metabolic traits associated with host range [8-11]. However, these findings were validated in only a few model strains using plants that fail to represent the large diversity of the species complex's host range. Previous comparative analyses did not identify strong host specificity factors but were based on strains with a limited genetic diversity or failed to consider the entire host spectrum of each strain [12,13]. A recent study that characterized the bacterium's genetic and phenotypic heterogeneity allowed us to select new model strains and to further explore the underlying mechanisms that determine host range [14].

Microarray and multilocus sequence typing (MLST) analyses and surveys of biological diversity demonstrated the existence of several clonal lineages adapted to specific host plants inside the American phylotype II group of the *R. solanacearum* species complex [12,14-16]. Lineages consisting of strains with similar host ranges are grouped into ecotypes, as they represent genetically distinct populations adapted to particular ecological niches within the *R. solanacearum* species complex (RSSC). These lineages include the banana Moko disease-causing strains, the cold-tolerant potato brown rot strains (historically and for regulatory purposes known as Race 3, biovar 2 or R3bv2) and the Not Pathogenic to Banana (NPB) strains, a recently emerged group that does not wilt banana despite its phylogenetic location in sequevar 4; all other sequevar 4 strains cause Moko disease. We hypothesized that host range determinants could be identified by comparing the genomes of closely related strains that have highly divergent host ranges. The polyphyletic nature of the Moko ecotype and the unexpectedly close relationship of some Moko lineages to the monophyletic brown rot and NPB ecotypes make these highly adapted strains a robust model for the study of host adaptation.

In this study, we sequenced a representative group of strains from each of these ecotypes and analyzed their genomes using multiple comparative genomics methods. Genomes were compared using phyletic profiling to determine ecotype-specific gene content. We identified genetic variations associated with NPB and brown rot strains emergence. To gain a better understanding of the polyphyly of Moko strains, horizontal gene transfers (HGTs) were also investigated using an explicit phylogenetic method. Each of these analyses has proven successful in several other plant pathogenic bacteria, including *Pseudomonas syringae* [17] and *Xanthomonas* sp. [18,19], and has provided evidence that specific gene content, genetic variations and HGT can explain host adaptation.

These comparative genomic analyses identified several virulence factors associated with NPB and brown rot strain emergence, along with Moko strain polyphyly. However, there were surprisingly few differences in lineage-specific gene content that could explain the host adaptation of the various lineages despite the high plasticity of the pan-genome.

Results

Host-adapted strains from phylotype II offer a model to study host adaptation

The general features of the *R. solanacearum* genomes analyzed in this study are presented in Table 1. We selected eight phylotype II strains for sequencing based on their sequevar classifications and on phenotypic data that indicated host adaptation. Eleven additional RSSC genomes were obtained from public databases. The three brown rot strains form a monophyletic group adapted to temperate climate; these sequevar 1 strains are typically isolated from potato in the highland tropics. The Moko disease strains form a polyphyletic group adapted to banana and are represented in this analysis by seven strains from sequevars 3, 4, 6 and 24, all isolated from plants in the genus *Musa*. The NPB strains form an emerging monophyletic group that appears to have diverged recently from *Musa*-infecting strains in sequevar 4. This group is represented by two strains, one isolated from *Cucumis* and one from *Heliconia* [15].

We determined draft sequences of a total of eight new phylotype II draft genomes. The resulting whole-genome data were used to confirm the phylogenetic relationships previously inferred from the phylotype-sequevar classification. We used Maximum Unique Matches index (MUMi) genomic distances [20] to generate a new, more complete phylogenetic tree of the *R. solanacearum* species complex (Figure 1). The tree exhibits a similar topology as the previously inferred phylogeny, which was based on microarray or MLST data. Although the strains in phylotype II are closely related to the relative context of the species complex, the Moko group is clearly polyphyletic and includes four distinct lineages that correspond to sequevars 3, 4, 6 and 24. Interestingly, Moko sequevar 3 is closely related to the monophyletic potato brown rot lineage, and Moko sequevar 4 is even more closely related to the monophyletic NPB lineage.

To confirm the host-adapted nature of the selected phylotype II strains, we initially conducted pathogenicity assays under tropical conditions using tomato, potato, banana (*Musa*), melon (*Cucumis*) and *Anthurium* as representative hosts. The pathogenicity profiles obtained from these assays clearly demonstrate the host adaptation of each group (Figure 2). Moko strains from sequevars 3, 4, 6 and 24 wilted banana plants, but the sequevar 4-NPB strains did not. The NPB strains were

Table 1 General characteristics of the strains and genomes used in this study

Phylotype - sequevar	Strain	Genome length (Mb)	GC %	#Contig	#CDS	Isolated from	Geographic origin
I	GMI1000	5.81	66.9%	2	5635	Tomato	Guyana
IIB-1	IPO1609	5.24	66.7%	102	5203	Potato	Netherlands
	UW551	5.22	64.8%	561	5301	Geranium	Kenya
	UW491	5.27	66.7%	222	5035	Potato	Colombia
IIB-3	MoIk2	5.48	66.7%	30	5438	Banana	Philippines
	CFBP1416*	5.68	66.6%	653	5722	Plantain	Costa Rica
IIB-4	CIP417*	5.47	66.8%	609	5398	Banana	Philippines
	UW179*	5.37	66.7%	590	5354	Banana	N/A
IIB-4 NPB	UW163*	5.48	66.6%	572	5467	Plantain	Peru
	Po82	5.43	66.6%	115	5019	Potato	Mexico
	CFBP6783*	5.54	66.7%	655	5505	<i>Heliconia</i>	French West Indies
IIA-7	IBSBF1503*	5.45	66.7%	633	5452	Cucumber	Peru
	K60	5.33	66.7%	23	5102	Tomato	United States
IIA-6	Grenada91*	5.41	66.6%	670	5365	Banana	Grenada
IIA-24	B50*	5.49	66.4%	1088	5648	Banana	Peru
III	CMR15	5.61	65.0%	3	5149	Tomato	Cameroon
IV	PSI07	5.62	64.5%	3	5247	Tomato	Indonesia
	BDB R229	5.23	66.3%	13	5051	Banana	Indonesia
	<i>R. syzygii</i> R24	5.45	65.9%	2	5239	Clove	Indonesia

* Genomes sequenced during this study.

the only group able to infect melon and *Anthurium* plants. Finally, Moko, NPB, and brown rot strains can all infect potato plants at warm temperatures, but only brown rot strains are highly virulent under cooler temperate conditions [14]. It is important to note that not all strains in phylotype II exhibit this degree of host adaptation. For example, and as expected from the literature, the broad host range of *R. solanacearum* type strain K60 (which also causes disease in tobacco,

eggplant, and pepper) did not extend to banana, melon or anthurium under our experimental conditions.

Genomic diversity of phylotype II and the *R. solanacearum* species complex

A total of 16,757 distinct homolog families were identified across the 19 *R. solanacearum* genomes; we consider these homologs to be the current pan-genome of the species complex. Compared with a previous genomic

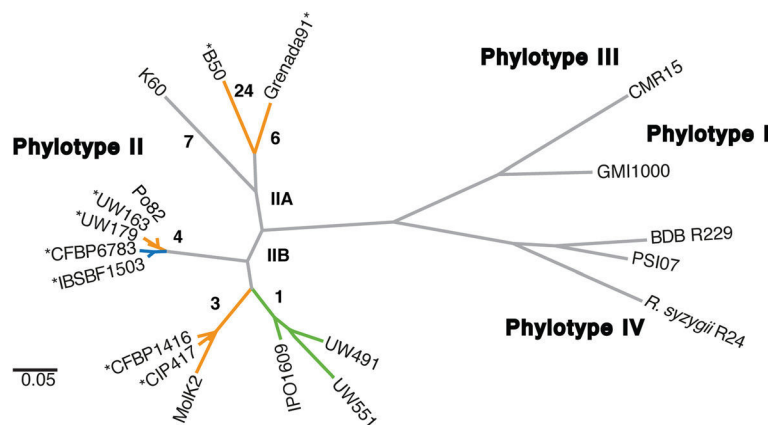


Figure 1 Phylogenetic tree of the *R. solanacearum* species complex inferred from whole-genome comparisons. The tree was computed from MUMi genomic distances. Stars next to strain names indicate genomes sequenced during this study. Phylotype and sequevars are given along branches. Colors indicate ecotypes. Orange: Moko (banana) strains; Blue: NPB strains (Not Pathogenic to Banana), Green: Potato brown rot strains. The scale bar corresponds to MUMi distances.

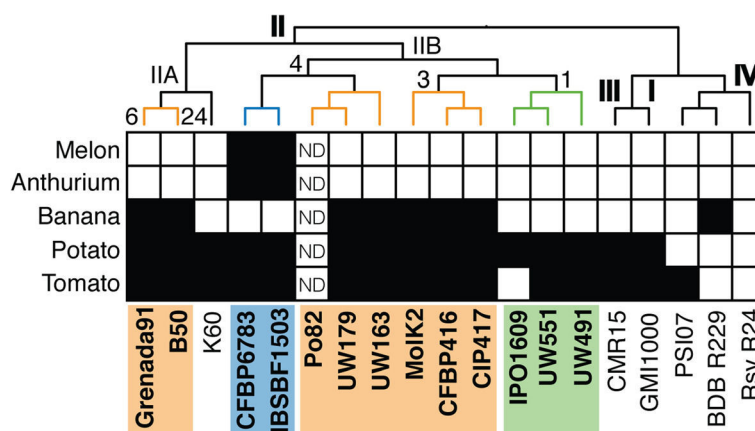


Figure 2 Host range of sequenced *R. solanacearum* strains. Black squares indicate compatible interactions, and white squares indicate incompatible interactions. The tree located alongside the matrix is the same as the one depicted in Figure 1. Colors indicate ecotypes: Orange: Moko (banana) strains; Blue: NPB strains (Not Pathogenic to Banana), Green: Potato brown rot strains.

analysis of eight *R. solanacearum* genomes [5], our results expand the pan-genome size by ~70%. The inclusion of previously unsequenced and phylogenetically distinct lineages, notably sequevars 4, 6 and 24, explains the steep increase in the pan-genome size. The 1,940 loci conserved in all strains constitute the current *R. solanacearum* core genome, which represents 17% of the pan-genome and 35% of an average *R. solanacearum* genome, which contains ~5,500 CDSs. Ecotype-specific gene content was determined by individually comparing each ecotype to every other sequenced genome. The NPB and brown rot strains each exhibited only a few unique genes specific to their phylogenetic groups: 99 and 109, respectively (Figure 3). The Moko strains are unlikely to exhibit specific genes because this polyphyletic group is composed of distant lineages. Three hypotheses can explain the emergence of a polyphyletic Moko ecotype. First, the most common recent ancestor (MRCA) of phylotype II may have been adapted to bananas. During the subsequent clonal expansion, variations in environmental selection pressures may have promoted maintenance of the ancient phenotype in certain lineages (which still cause Moko disease) and promoted its loss in others. Alternatively, because *R. solanacearum* is capable of natural transformation [21], pathogenicity to banana may have originated in an isolated lineage after divergence from the MRCA. Subsequently, the banana-infecting trait(s) may have spread to multiple phylotype II lineages via HGT, resulting in the current polyphyletic Moko group. A third possibility is that multiple lineages underwent convergent evolution towards the banana-infecting trait, which is consistent with the absence of ecotype-specific genes.

We selected a subset of 227 known or likely virulence factors from the pan-genome based on the literature and

manually annotated them (Additional file 1). These virulence factors have a broad range of functions during pathogenesis, including the secretion and synthesis of type III effectors (T3es), motility, chemotaxis, synthesis of exopolysaccharide (EPS), and degradation of plant cell walls. Compared with the other virulence factors, the T3e repertoire of *R. solanacearum* exhibits high plasticity. The pan-effectome of the species complex contains 113 T3es, but of those, only 14 effectors are present in every sequenced strain. These 14 effectors form the core effectome: RipG5, RipB, RipW, RipAC, RipAB, RipR,

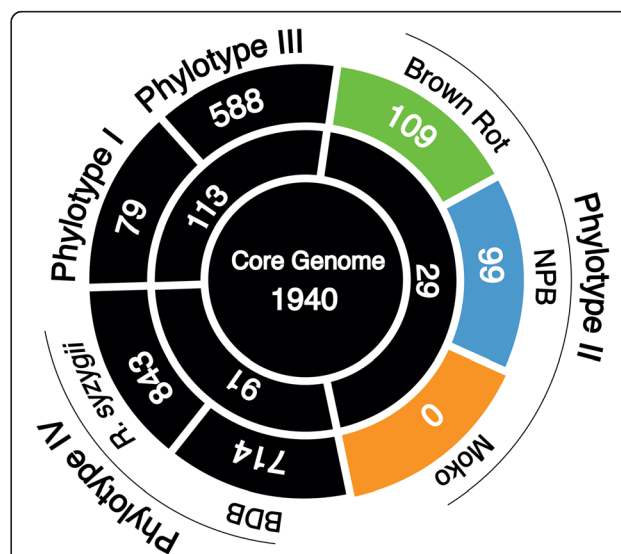


Figure 3 Comparative diagram of the specific gene contents in representative groups of the species complex. The number of genes unique to each group was determined using homolog prediction inferred with the OMA algorithm. Colors indicate ecotypes: Orange: Moko (banana) strains; Blue: NPB strains (Not Pathogenic to Banana), Green: Potato brown rot strains.

RipE1, RipAM, RipAN, RipAY, RipAJ, RipF1, RipAI and a PopC-like effector. The core effectome represents a ~36% size decrease relative to the previously inferred core effectome based on 11 strains [13]. On average, 64 T3es were present per strain. A total of 40 T3es were common to all Moko strains, whereas 76 were common to NPB strains, and 52 were common to brown rot strains.

Due to the large pan-genome, each strain appeared to exhibit highly diverse genetic content even within the Moko, NPB and brown rot groups. This variability can be partially explained by the estimation of gene gain and loss rates along lineages. Indeed, each phylogenetic node of the species complex appears to have undergone hundreds of gain and loss events, thus creating mosaic genomes (Additional file 2). This gain and loss explains why some homolog families are shared by distant lineages but not by closer ones. Nonetheless, it is important to consider that the new genomes are in draft form and are divided into contigs. Thus, these genomes contain fragmented CDSs that can artificially increase the number of predicted gene families and introduce some bias into subsequent analysis.

Pairwise comparison of the genomic content of Moko with the NPB and brown rot strains

To overcome the problem of these mosaic genomes and to specifically target gene content associated with host adaptation, we hypothesized that the brown rot clade emerged when the IIB-1 brown rot strains diverged from the IIB-3 Moko lineage. Similarly, we hypothesized that the NPB clade emerged when the IIB-4 NPB lineage diverged from the IIB-4 Moko strains. This methodology thus focuses on the evolution of genomic content during emergence events. To test these hypotheses, we directly compared the co-occurrence patterns of genes of the IIB-1 brown rot genomes to those of the Moko genomes and the IIB-4 NPB genomes to the Moko genomes. The first step was to establish which genes were unique to the brown rot and NPB strains compared with the Moko strains. Subsequently, we determined which genes were conserved in every Moko lineage but were absent in either the NPB or brown rot strains (Additional file 3).

The NPB strains exhibited 102 unique genes missing from all Moko strains. These genes fell into three putative genomic islands. The first one encoded a type I secretion system (T1SS), including an outer membrane protein (RALP6v1_3180004), a membrane fusion protein (RALP6v1_3180005), and an ATP-binding protein (RALP6v1_3180006). This secretion system is located next to a transposase homolog (RALP6v1_3180003), suggesting that the genes were most likely acquired by HGT. The second genomic island contained genes related to the degradation of the plant alkaloid isoquinoline (*iorA* and *iorB*) located next to a gene encoding an AraC-type transcriptional

regulator. The third genomic island contained a gene with a phosphatase 2C domain, a motif usually found in eukaryotic organisms. Phosphatase 2C can be involved in various cell functions, including the regulation of plant defense in *Arabidopsis* [22], and is sometimes secreted into host cells by pathogens [23]. The genomic island also contained a gene encoding a putative helix-turn-helix (HTH)-type transcriptional regulator as well as transposition-related genes, suggesting that it was also acquired by HGT. Outside of these clearly defined putative genomic islands, CDSs encoding another HTH-type transcriptional regulator and a GGDEF-domain signal transduction protein were unique to the NPB genomes.

Only six genes were conserved in Moko strains but absent from NPB strains. Interestingly, this gene-set included *ripAA*, a well-characterized T3e formerly known as *avrA* [8], which was either absent or pseudogenized in NPB strains. This effector was highly conserved in Moko strains (~85% nucleotide identity) and in the species complex (present in 15 out of the 19 sequenced strains), thus suggesting that *ripAA* was lost during the emergence of NPB. Another GGDEF-domain signal transduction protein was also absent from the NPB strains.

Brown rot strains exhibited 134 conserved genes that were absent from all Moko strains. One hundred three genes coded for conserved proteins of unknown function. However, a BLAST search on NCBI's nr database revealed that several of these hypothetical proteins exhibited homologs in other genera of soil-dwelling and plant pathogenic bacteria (e.g., *Xanthomonas*, *Pseudomonas*, *Cupriavidus* and others). We identified a single putative functional genomic island related to iron metabolism: a FecR homolog (UW551v3_mp40033) located next to an iron transporter (UW551v3_mp40034) and a sigma 70 factor (UW551v3_mp40032). Another gene block coding for hypothetical proteins without significant BLAST hits was also located next to an AraC-type transcriptional regulator. Finally, a member of the RTX exotoxin family (UW551v4_570022) was unique to the brown rot strains.

We identified 63 genes that were conserved in Moko strains and missing from brown rot strains. A large gene block included several genes of interest: the T3e *ripAU*, a two-component system involved in the stress response of *cpxR* and *cpxA*, multiple drug efflux pumps from the Acr and RND families, and an AraC-type transcriptional regulator. Several isolated transcriptional regulators, including one each from the TetR and MarR families, were also identified.

Our analyses identified similar numbers of NPB- and brown rot-specific genes relative to Moko strains. Although NPB strains exhibited three specific genomic islands that may be related to adaptation (potentially acquired by HGT), brown rot strains exhibited more than one hundred specific genes encoding proteins of unknown function that

appeared to be conserved in other bacteria. Although gene loss was globally more predominant in brown rot strains than in NPB strains, both ecotypes only lost a single distinct T3e each (*ripAU* and *ripAA*, respectively), possibly while diverging from their respective Moko lineages. These contrasting rates of gene loss between effectors relative to the rest of the genome suggest that the loss of these two effectors could have been key events in the emergence of these ecotypes.

HGT events between Moko lineages

To test the hypothesis that adaptation to banana spread through Moko lineages of phylotype II via HGT, we searched for signs of transfer between Moko lineages using the AnGST algorithm. This algorithm compares gene trees with their associated species tree and infers both HGT events and the direction of transfer by identifying conflicts between the two. We only considered genes with HGT patterns involving all four Moko lineages and directions of transfers compatible with a unique original donor. According to our hypothesis, this unique donor would be pathogenic to banana (Additional file 2). This strategy identified nine genes. Five of these genes exhibited the IIA Moko lineages MRCA as the likely donor with the IIB Moko lineages as acceptors. Of these five genes, two encode the T3es *ripAD* and *ripG4* from the GALA family. The others encode two hypothetical proteins and a bacteriophage-related protein. Three genes exhibited the IIB-3 lineage as the donor and putatively encode a protein belonging to the beta-lactamase superfamily, the translation initiation factor *infA*, and *hrcT*, a key component of the T3SS. The only gene that apparently exhibited the IIB-4 lineage as a donor encodes a putative adenine-specific methyl-transferase.

This analysis suggests that a small minority of genes seems to have spread through every Moko lineage via a single lineage (the hypothetical original donor). Although the original donors were not fully consistent, more than half of the HGT events originally started from the IIA Moko's MRCA and affected T3es that are likely to be directly involved in pathogenesis.

Association of sequence polymorphisms with host-adaptation

Due to the phylogenetic proximity of NPB and brown rot strains to individual Moko lineages, we analyzed the diversity of their shared genomic content. We analyzed multiple alignments of protein sequences to search for non-synonymous single-nucleotide polymorphisms (NS-SNPs) and insertions and deletions (INDELS) associated with the host-adapted groups. We will refer to these features as host-adapted polymorphisms (HAPs). More precisely, we selected polymorphisms that exhibited the same form in Moko strains but a different form in either the NPB or brown rot strains (Additional file 4).

Of the 2,855 genes common to the NPB and Moko strains, 96% were polymorphic. However, only 24 genes (<1%) contained HAPs, including 16 within conserved domains in protein sequences. This result can be explained by the very low divergence of the IIB4-NPB and IIB4 Moko lineages, as only 14% of their common genes are polymorphic. Notable NS-SNPs were identified in the T3e *ripAN* and in a PadR-like transcription regulator involved in the response to phenolic compounds; notable INDELS were identified in *motB*, a flagellar motor protein involved in the response to chemotaxis signals, and *epsF*, a membrane protein involved in the secretion of EPS. NS-SNPs were also identified in two multidrug resistance proteins (MDRs) and an HTH-type transcriptional regulator.

Ninety-nine percent of the 2459 common genes present in both the brown rot and Moko strains were polymorphic. Of these, 1024 genes (42%) contained HAPs, including 742 within protein domains. Twenty T3es contained the majority of NS-SNPs, including five T3es with more than ten NS-SNPs (*ripV1*, *ripG6*, *ripC1*, *ripAO* and *ripAD*), three members of the GALA T3e family and *ripAB* and *ripAC*. Two regulators and subunits of the T3SS were polymorphic: *prhJ* and *hrpAFGHJ*. Several HAPs were found in genes encoding members of diverse metabolic pathways, notably the NorB and NasF enzymes of inorganic nitrogen metabolism. Finally, 22 putative transcriptional regulators of various families contained HAPs in their substrate-binding domains.

Discussion

Using comparative genomic analysis, we compared the genomes of three host-adapted groups of strains from the phylotype II of *R. solanacearum*: brown rot strains adapted to potatoes and temperate climates, Moko strains adapted to bananas, and NPB strains adapted to melon or *Anthurium*. Together, our analyses demonstrated that the potato brown rot, Moko, and NPB strains constitute phylogenetically related populations that have adapted to biologically distinct host environmental conditions. Thus, these phylotype II groups, called ecotypes, provided a good model to explore host adaptation in *R. solanacearum*. We hypothesized that comparing the genomes of these closely related but biologically distinct groups could identify specific mechanisms of host adaptation.

Overall, the species complex appeared to be very dynamic, with a large pan-genome and a limited core genome. Specifically, the *R. solanacearum* T3e repertoire was unusually large compared with those of other well-known plant pathogens such as *Pseudomonas* spp. or *Xanthomonas* spp. [24,25]. The variability of T3es suggests that there may be extensive functional redundancy between effectors. Paralog families, such as the

GALA effectors, are known to exhibit functional redundancies, and single effector mutations rarely produce a virulence defect [10]. These functional redundancies drastically reduce the analytic power of comparative genomics; therefore, more data about the specific virulence functions of individual T3es are needed to better understand the biological mechanisms underlying host adaptation. Such data are especially important considering that among the few functionally characterized T3es [26], only *ripR* (*popS*) is a member of the *R. solanacearum* core effectome.

Three competing hypotheses could explain the origin of polyphyly of the Moko strains. The first hypothesis proposes that the phylotype II MRCA was a banana pathogen and that the capacity to infect banana was subsequently lost in multiple clades after clonal expansion due to variations in environmental selection pressures. Our analyses did not identify any genes present in all Moko strains but absent in all other *R. solanacearum* strains. Although this observation seems to contradict our hypothesis, it does not invalidate the hypothesis but rather suggests that different parts of the set of hypothetical genes responsible for the banana virulent phenotype present in phylotype II's MRCA could have been lost in non-Moko lineages. This phenomenon would have contributed to the mosaic nature of *R. solanacearum* genomes and would have been facilitated by functional redundancies.

Our second hypothesis states that banana pathogenicity emerged in a single clade and spread among phylotype II strains by HGT facilitated by natural transformation. Despite the large gene pool shared by Moko strains, only nine genes exhibited HGT patterns consistent with our hypothesis. We determined that the T3e *ripAD* that was transferred from the common ancestor of the IIA-6/24 Moko lineages to the IIB-3/4 Moko lineages. This effector does not possess the regulatory *hrp_{II}* box motif but is expressed in an *hrpB*-dependent manner and is translocated into plant cells [27]. Moreover, although its function is not characterized in *R. solanacearum*, it is also a member of the *hpx8* family found in plant pathogenic *Xanthomonas*. Another T3e, *ripG4*, exhibited similar HGT patterns originating from the IIA Moko lineages. This effector is a divergent member of the GALA family that interferes with PAMP-triggered immunity in *Arabidopsis* and contributes to host adaptation [10].

A genomic content analysis was designed to characterize the evolution of genomic content during the emergence of the NPB and brown rot clades and to identify the specific genes that could be responsible for host range differences. Our analysis revealed that few genes separate the Moko strains from the NPB and brown rot strains despite their biologically distinct phenotypes. One of these, *ripAA*, was lost by NPB strains' MRCA during its divergence from the IIB-4 Moko lineage. This effector is recognized by tobacco, so it prevents phylotype I strain GMI1000 from

infecting tobacco plants [8]. We speculate that *ripAA* could either be an effector that enables Moko strains to wilt banana or an avirulence factor that prevents pathogenicity in plants that are not part of the Moko strains' host range but are part of the NPB strains' host range. *ripAA* is a common effector in the species complex, but none of the strains with this effector are melon or *Anthurium* pathogens. Therefore, we propose that the loss of this effector by the NPB strains could have been a key event in this clade's host change. Similarly, *ripAU* was lost by the brown rot strains during the divergence from the IIB3 Moko lineage. This effector is also present in all of the other strains of the species complex, and its loss might have been necessary for brown rot strains to adapt to temperate climates. The cold tolerance of brown rot strains is strictly dependent on an interaction with the host [28], and T3es may play key roles in this interaction. Although the function of *ripAU* in *R. solanacearum* is unknown, this effector belongs to the *hpx8* family. Hpx families are T3e families sharing protein sequence similarities defined by Mukaihara, and the *hpx8* family shares similarities with the effector XopV, which is found in various plant pathogenic *Xanthomonas* [27]. However, the function of XopV in *Xanthomonas* is not well characterized.

A complete T1SS was acquired by NPB strains during their divergence from banana-infecting sequevar 4 Moko strains, most likely via HGT. T1SSs are ubiquitous and versatile systems often involved in the secretion of various virulence factors into the extracellular medium, such as proteases, toxins, or quorum-sensing molecules that can promote plant invasion [29]. T1SSs can also play a defensive role to counteract plant defense responses by exporting antimicrobial compounds out of the cell. Numerous genes encoding conserved proteins of unknown function were acquired by brown rot strains during their divergence from sequevar 3 Moko strains. Homologs of most of these genes are present in other soil-borne and plant pathogenic bacteria, suggesting that they may play a role in adaptation to either environmental conditions or unique hosts. Further characterization of these proteins of unknown function is required to determine if they participate in brown rot strain adaptation to temperate climate. Homologs of the transporter and the sensor of the iron-sensing *fec* system and a sigma70 factor were also encoded by brown rot-specific genes. The *fec* system is responsible for iron uptake and is essential in many human pathogens and a few enteric plant pathogens [30,31]. However, all *R. solanacearum* strains sequenced to date have the *fur* system, which depends on another siderophore. It is possible that this *fec* system has a ligand specificity that is more efficient at sequestering iron in potato tubers or at low temperatures. Functional analyses could determine if this additional iron uptake system is involved in brown rot cold

tolerance. Finally, we looked for polymorphisms (i.e., HAPs) that could affect the function of the genes shared between Moko and either NPB or brown rot strains. Polymorphisms are introduced by replication errors or intragenic recombination and can have a major impact on gene functionality. In plant pathogens, a single amino acid modification in TAL effectors can alter the host gene targeted by the effector [32]. Despite the genetic distances between Moko lineages and their proximity to the brown rot and NPB ecotypes, a high number of genes contained sites that were coincidentally fixed in the Moko strains but polymorphic in either NPB or brown rot strains, suggesting that these sites are functionally important and may be related to host adaptation.

NPB and Moko strains contained several HAPs in known virulence factors. The T3e *ripAN* had a single NS-SNP. Although RipAN is translocated into host cells [33], this effector's function in *R. solanacearum* is unknown. Homologs of *ripAN* have not been identified in other plant pathogenic bacteria. One HAP was also found in the gene encoding the inner membrane protein EpsF, which is predicted to be responsible for the modification and/or export of the exopolysaccharide EPS I, a major bacterial wilt virulence factor. A six-amino acid insertion was identified in the C-terminal end of the NPB *epsF* gene. The extent of cross-talk between plants and EPS is not yet entirely understood, but it is known that different EPS proteins trigger different plant defense responses and that EPS modifications play a role in biofilm formation, immune evasion and virulence in EPS-producing bacteria [34,35].

Furthermore, we also found a HAP in a transcriptional regulator containing a PadR-like domain. PadR domains modulate the expression of virulence factors, MDR efflux pumps and responses to phenolic stress [36,37]. Interestingly, HAPs were also identified in two independent MDR transporters. Because plants produce antimicrobial phenolic compounds in response to pathogens, the PadR and MDR genes might constitute a network for stress response/resistance to plant defense mechanisms adapted to NPB-specific host range. Altogether, the 24 HAPs identified may be necessary for NPB strains to adapt to the environmental conditions encountered in new hosts.

The brown rot and Moko strains exhibited hundreds of genes with HAPs in conserved protein domains, suggesting that *R. solanacearum* metabolism and virulence may function differently in brown rot strains. Remarkably, the list of HAP sites included central virulence regulators, T3SS regulators, T3SS machinery, and 19 T3es. T3es with HAPs notably included an effector with a nuclear localization signal (*ripAB*) and several functionally redundant effectors known to affect host range (*ripG2*, *ripG3*, and *ripG6*) [10,38]. Recent transcriptomic studies

suggest that regulation of virulence factors during pathogenesis is far more complex than previous models proposed [39,40]. Given that brown rot's adaptation to cold is dependent on interactions with a host, a large reorganization of core mechanisms may be required for *R. solanacearum* to cause disease in temperate climates.

Conclusions

This extensive comparative genomics analysis identified relatively few differences in gene content between closely related *R. solanacearum* strains with contrasting biological characteristics. However, several T3e were associated with the Moko, NPB and brown rot ecotypes (Table 2). Most differences between strains involved HAPs of uncertain biological significance, although many HAPs were located in genes associated with bacterial wilt virulence. Our study did identify specific hypotheses concerning mechanisms of host adaptation in the *R. solanacearum* species complex. These hypotheses

Table 2 Candidate T3e for host specificity identified in Moko, NPB and brown rot ecotypes

Gene name	Alternative name	Characteristics
<i>ripAU</i>		Absent from brown rot, present in Moko
<i>ripAA</i>	<i>avrA</i>	Absent from NPB, present in Moko
<i>ripG4</i>	<i>GALA4</i>	HGT from IIA Moko to IIB Moko lineages
<i>ripAD</i>		HGT from IIA Moko to IIB Moko lineages; HAPs between brown rot and Moko
<i>ripAB</i>	<i>popB</i>	HAPs between brown rot and Moko
<i>ripAC</i>	<i>popC</i>	HAPs between brown rot and Moko
<i>ripAD</i>		HAPs between brown rot and Moko
<i>ripAE</i>		HAPs between brown rot and Moko
<i>ripAI</i>		HAPs between brown rot and Moko
<i>ripAO</i>		HAPs between brown rot and Moko
<i>ripAP</i>		HAPs between brown rot and Moko
<i>ripAY</i>		HAPs between brown rot and Moko
<i>ripB</i>		HAPs between brown rot and Moko
<i>ripC1</i>		HAPs between brown rot and Moko
<i>ripD</i>	<i>avrPphD</i>	HAPs between brown rot and Moko
<i>ripE2</i>		HAPs between brown rot and Moko
<i>ripF1</i>		HAPs between brown rot and Moko
<i>ripG2</i>	<i>GALA2</i>	HAPs between brown rot and Moko
<i>ripG3</i>	<i>GALA3</i>	HAPs between brown rot and Moko
<i>ripG6</i>	<i>GALA6</i>	HAPs between brown rot and Moko
<i>ripH1</i>	<i>HLK1</i>	HAPs between brown rot and Moko
<i>ripH2</i>	<i>HLK2</i>	HAPs between brown rot and Moko
<i>ripV1</i>		HAPs between brown rot and Moko
<i>ripW</i>	<i>popW</i>	HAPs between brown rot and Moko
<i>ripAN</i>		HAPs between NPB and Moko

will be tested using functional genomics experiments such as gene swaps and deletions to determine whether these mechanisms play roles in host specificity. However, an important proportion of these candidate genes are related to regulatory function, suggesting that host range could evolve through changes in regulation. Small genomic differences could lead to drastically different expression profiles when the bacterium infects different hosts. Due to their unexpected overall genomic similarity and their clear-cut differences in host range, IIB-4 Moko and NPB form an elegant model for transcriptomic studies designed to identify differentially expressed genes associated with host specificity.

Methods

Genome sequencing and assembly

Total DNA was extracted using the phenol-chloroform method. Libraries were constructed using Nextera technology and sequenced on Illumina's HiSeq-2000 using a 2 × 50-nt paired-end strategy. Reads were pre-processed using Trimmomatic [41]. First, adapter sequences and low-quality nucleotides occurring at 5' and 3' ends with a Phred quality score < 20 were trimmed. Second, reads shorter than half their initial lengths were discarded. The resulting reads were assembled using Velvet [42]. Manual editing of the annotations was performed for genes of interest using the MaGe web interface of the MicroScope platform [43]. The accession numbers are as follows: CFBP1416 [EMBL:PRJEB7434], CIP417 [EMBL:PRJEB7427], B50 [EMBL:PRJEB7421], Grenada 91 [EMBL:PRJEB7428], UW179 [EMBL:PRJEB7426], UW163 [EMBL:PRJEB7430], CFBP6783 [EMBL:PRJEB7432], and IBSBF1503 [EMBL:PRJEB7433]. Annotation files are provided in genbank format (Additional file 5).

Pathogenicity assays

Pathogenicity was assessed on 3–4 fully expanded leaves of tomato cv. L390 (T10) (30 plants), potato cv. Désirée (30 plants), Cavendish banana cv. 902 (8 plants), melon cv. Amish (30 plants) and the ornamental plant Anthurium cv. Fire (4 plants). Plants were placed in a full containment security level growth chamber with a 12-h photoperiod, 28 ± 2°C (day) and 24 ± 2°C (night), and 90% relative humidity. Bacterial suspensions were prepared as described in Cellier & Prior 2010. Using the soil-soak method, each plant was inoculated with 5 × 10⁹ colony-forming units (CFU) after lightly damaging the roots with a scalpel. Strains were considered pathogenic if more than 50% of plants presented wilt symptoms within one month after inoculation. Each assay was repeated 2 times. We could not analyze the phenotype of strain Po82 because the authors of this published genome would not share their strains [44].

Exploratory virulence dataset

The list of *R. solanacearum* virulence factors analyzed in this study was based on Remenant et al. [12]. Orthology relationships were determined for each strain by BLAST and synteny data using the web-based MaGe interface. Every gene annotation was then manually validated to ensure homogeneity of the start codon positions and to detect frameshifts and pseudogenization. T3es were annotated using the IANT “Ralstonia T3E” database [13]. T3es located on contig borders were considered as present to establish the core effectome.

Species complex phylogeny

The species complex phylogeny was inferred by neighbor-joining using MUMi genomic distances [20]. MUMi values were computed from whole-genome comparisons conducted with MUMMer 3.0 [45].

Phyletic profiles

Homology relationships across all genomes analyzed in this study were inferred using the Orthologous Matrix algorithm (OMA) [46] with the following default criteria: alignment length > 60% of the minimum gene length and alignment score > 181 in Gonnet PAM matrix units. Phyletic profiles [47] were subsequently determined using R to identify co-occurrence patterns for specific genes in a given group of strains. The specific gene content was locally blasted on the NCBI nr database to identify an eventual source organism.

Detection of gain, loss and HGT events

Amino acid sequences of each homolog family were aligned with Muscle [48], and a neighbor-joining phylogenetic tree supported by 1000 bootstrap replicates was computed with FastTree [49]. Gain, loss, and HGT events were inferred by reconciliation of the gene trees and the species complex tree topologies using the AnGST algorithm (Analyzer of Gene and Species Tree) [50]. This algorithm compares the topology of a gene tree with its associated species tree, which is generally defined as an explicit phylogenetic method within the scope of HGT detection. The algorithm identifies differences between the gene and the species trees and explains them (“reconciles”) according to a set of evolutionary events, including gain, loss, duplication or horizontal transfer, inferred with a parsimony-based model. Loci with > 95% nucleotide identity were discarded to avoid false positives caused by phylogenetic trees with insufficient resolution.

Allelic variation analysis

Nucleic and protein sequences of each homolog family were aligned with Muscle [48]. The resulting alignments were trimmed on the 5' ends and then screened for

host-associated amino-acid polymorphisms using R. The strain MolK2 was used as a reference to assign the position of each polymorphism in the sequence. Functional domains containing polymorphisms were identified using the CDD database from NCBI.

Availability of supporting data

The data sets supporting the results of this article are included within the article and its additional files. Sequence data are available at the EMBL nucleotide sequence database:

- CFBP1416 <http://www.ebi.ac.uk/ena/data/view/PRJEB7434>
- CIP417 <http://www.ebi.ac.uk/ena/data/view/PRJEB7427>
- B50 <http://www.ebi.ac.uk/ena/data/view/PRJEB7421>
- Grenada 91 <http://www.ebi.ac.uk/ena/data/view/PRJEB7428>
- UW179 <http://www.ebi.ac.uk/ena/data/view/PRJEB7426>
- UW163 <http://www.ebi.ac.uk/ena/data/view/PRJEB7430>
- CFBP6783 <http://www.ebi.ac.uk/ena/data/view/PRJEB7432>
- ISBSF1503 <http://www.ebi.ac.uk/ena/data/view/PRJEB7433>

Additional files

Additional file 1: Subset of 227 virulence factors from the pan-genome. These genes were manually selected and annotated. Red cells: absent gene. Orange cells: pseudogenized or frameshifted gene. Black: Fragmented genes due to sequencing gaps.

Additional file 2: Results from the analysis with the AnGST algorithm. Gene gain and loss events in the species complex and HGT patterns between Moko lineages. Black boxes indicate that the lineage is involved in the HGT.

Additional file 3: Results of phyletic profiling. List of genes present or absent in NPB or brown rot strains when compared to Moko strains. Results of BLASTp in the NCBI nr database are provided.

Additional file 4: Results from the analysis of HAPs. List of HAPs contained in genes shared by NPB or brown rot strains with Moko strains are provided. Protein domains predicted from NCBI CDD database and containing HAPs are indicated.

Additional file 5: Annotations files (.gb). Annotations are formatted in genbank format.

Abbreviations

AnGST: Analyzer of gene and species tree; CDS: Coding sequences; EPS: Exopolysaccharide; HAP: Host associated polymorphism; HGT: Horizontal gene transfer; HTH: Helix turn helix; INDEL: Insertion deletion; MDR: Multidrug resistant; MLST: Multilocus sequence typing; MRCA: Most recent common ancestor; MUMi: Maximum unique matches index; NPB: Not pathogenic to banana; NS-SNP: Non-synonymous polymorphism; OMA: Orthologous matrix; RSSC: *R. solanacearum* species complex; T1SS: Type I secretion system; T3e: Type III effector; T3SS: Type III secretion system.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The work presented resulted from collaboration between all authors and each author has participated sufficiently in the work to take public responsibility for the content. PP defined the research theme. FA, GC and PP performed the virulence bioassay. DR contributed to identify the repertoire of T3e. FA and PP realized the genes annotation. FA carried out all the bioinformatics analysis. FA, CA and PP drafted the article, CA and TL participated in revising it critically for intellectual content. Each author gave final approval of the version of the manuscript to be published.

Acknowledgements

The European Union (ERDF), Conseil Régional de La Réunion, the French Agence Nationale de la Recherche and CIRAD provided financial support. ANSES secured part of the doctoral fellowship to F. Ailloud (contract Number 11-237/BSL). We also thank B. Vinatzer (Virginia Tech, USA) for providing the early genomic sequence of strain UW491.

Author details

¹CIRAD, UMR PVBMT, F-97410 Saint-Pierre, La Réunion, France. ²Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI, USA. ³Anses - Plant Health Laboratory, F-97410 Saint-Pierre, La Réunion, France. ⁴Institut de Génétique, Genoscope, Commissariat à l'Energie Atomique (CEA), Evry, Paris, France. ⁵Département de Santé des Plantes et Environnement, (SPE) Inra, Paris, France.

Received: 26 November 2014 Accepted: 20 March 2015

Published online: 08 April 2015

References

1. Hayward AC. Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. *Annu Rev Phytopathol.* 1991;29:65–87.
2. Denny T. Plant pathogenic *Ralstonia* species. In: Plant-associated bacteria. Samuel S. Gnanamanickam, editor: Springer Netherlands; 2006. p. 573–644.
3. Fegan M, Taghavi M, Sly, LI, Hayward AC. Phylogeny, diversity and molecular diagnostics of *Ralstonia solanacearum*. In: Bacterial Wilt Disease: Molecular and Ecological Aspects. Prior P, editor INRA; 1998: 19–33.
4. Guidot A, Coupat B, Fall S, Prior P, Bertolla F. Horizontal gene transfer between *Ralstonia solanacearum* strains detected by comparative genomic hybridization on microarrays. *ISME J.* 2009;3(5):549–62.
5. Remenant B, de Cambiaire JC, Cellier G, Jacobs JM, Mangenot S, Barbe V, et al. *Ralstonia solanacearum*, the blood disease bacterium and some Asian *R. solanacearum* strains form a single genomic species despite divergent lifestyles. *PLoS One.* 2011;6(9):e24356.
6. Genin S, Denny TP. Pathogenomics of the *Ralstonia solanacearum* species complex. *Annu Rev Phytopathol.* 2012;50:67–89.
7. Peeters N, Guidot A, Vaillau F, Valls M. *Ralstonia solanacearum*, a widespread bacterial plant pathogen in the post-genomic era. *Mol Plant Pathol.* 2013;14(7):651–62.
8. Poueymiro M, Cunnac S, Barberis P, Deslandes L, Peeters N, Cazale-Noel AC, et al. Two type III secretion system effectors from *Ralstonia solanacearum* GMI1000 determine host-range specificity on tobacco. *Mol Plant Microbe Interact.* 2009;22(5):538–50.
9. Lavie M, Shillington E, Eguiluz C, Grimsley N, Boucher C. PopP1, a new member of the YopJ/AvrRxv family of type III effector proteins, acts as a host-specificity factor and modulates aggressiveness of *Ralstonia solanacearum*. *Mol Plant Microbe Interact.* 2002;15(10):1058–68.
10. Remigi P, Anisimova M, Guidot A, Genin S, Peeters N. Functional diversification of the GALA type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts. *New Phytol.* 2011;192(4):976–87.
11. Sole M, Popa C, Mith O, Sohn KH, Jones JD, Deslandes L, et al. The awr gene family encodes a novel class of *Ralstonia solanacearum* type III effectors displaying virulence and avirulence activities. *Mol Plant Microbe Interact.* 2012;25(7):941–53.
12. Remenant B, Coupat-Goutaland B, Guidot A, Cellier G, Wicker E, Allen C, et al. Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC Genomics.* 2010;11:379.

13. Peeters N, Carrere S, Anisimova M, Plener L, Cazale AC, Genin S. Repertoire, unified nomenclature and evolution of the Type III effector gene set in the *Ralstonia solanacearum* species complex. *BMC Genomics*. 2013;14:859.
14. Cellier G, Prior P. Deciphering phenotypic diversity of *Ralstonia solanacearum* strains pathogenic to potato. *Phytopathology*. 2010;100(11):1250–61.
15. Wicker E, Grassart L, Coranson-Beaudou R, Mian D, Guilbaud C, Fegan M, et al. *Ralstonia solanacearum* strains from Martinique (French West Indies) exhibiting a new pathogenic potential. *Appl Environ Microbiol*. 2007;73(21):6790–801.
16. Cellier G, Remenant B, Chiroleu F, Lefeuvre P, Prior P. Phylogeny and population structure of brown rot- and Moko disease-causing strains of *Ralstonia solanacearum* phylotype II. *Appl Environ Microbiol*. 2012;78(7):2367–75.
17. Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, et al. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathog*. 2011;7(7):e1002132.
18. Jalan N, Aritua V, Kumar D, Yu F, Jones JB, Graham JH, et al. Comparative genomic analysis of *Xanthomonas axonopodis* pv. citrumelo F1, which causes citrus bacterial spot disease, and related strains provides insights into virulence and host specificity. *J Bacteriol*. 2011;193(22):6342–57.
19. Bogdanove AJ, Koebnik R, Lu H, Furutani A, Angiuoli SV, Patil PB, et al. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *J Bacteriol*. 2011;193(19):5450–64.
20. Deloger M, El Karoui M, Petit MA. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J Bacteriol*. 2009;191(1):91–9.
21. Bertolla F, Van Gijsegem F, Nesme X, Simonet P. Conditions for natural transformation of *Ralstonia solanacearum*. *Appl Environ Microbiol*. 1997;63(12):4965–8.
22. Widjaja I, Lassowskat I, Bethke G, Eschen-Lippold L, Long HH, Naumann K, et al. A protein phosphatase 2C, responsive to the bacterial effector AvrRpm1 but not to the AvrB effector, regulates defense responses in *Arabidopsis*. *Plant J*. 2010;61(2):249–58.
23. Gilbert LA, Ravindran S, Turetzky JM, Boothroyd JC, Bradley PJ. Toxoplasma gondii targets a protein phosphatase 2C to the nuclei of infected host cells. *Eukaryot Cell*. 2007;6(1):73–83.
24. Hajri A, Pothier JF, Fischer-Le Saux M, Bonneau S, Poussier S, Boureau T, et al. Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Appl Environ Microbiol*. 2012;78(2):371–84.
25. Zumaquero A, Macho AP, Rufian JS, Beuzon CR. Analysis of the role of the type III effector inventory of *Pseudomonas syringae* pv. phaseolicola 1448a in interaction with the plant. *J Bacteriol*. 2010;192(17):4474–88.
26. Deslandes L, Genin S. Opening the *Ralstonia solanacearum* type III effector tool box: insights into host cell subversion mechanisms. *Curr Opin Plant Biol*. 2014;20C:110–7.
27. Mukaihara T, Tamura N, Iwabuchi M. Genome-wide identification of a large repertoire of *Ralstonia solanacearum* type III effector proteins by a new functional screen. *Mol Plant Microbe Interact*. 2010;23(3):251–62.
28. Milling A, Meng F, Denny TP, Allen C. Interactions with hosts at cool temperatures, not cold tolerance, explain the unique epidemiology of *Ralstonia solanacearum* race 3 biovar 2. *Phytopathology*. 2009;99(10):1127–34.
29. Tseng TT, Tyler BM, Setubal JC. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol*. 2009;9 Suppl 1:S2.
30. Expert D. Withholding and exchanging iron: Interactions Between *Erwinia* spp. and Their Plant Hosts. *Annu Rev Phytopathol*. 1999;37:307–34.
31. Ratledge C, Dover LG. Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol*. 2000;54:881–941.
32. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. 2009;326(5959):1509–12.
33. Mukaihara T, Tamura N. Identification of novel *Ralstonia solanacearum* type III effector proteins through translocation analysis of *hrpB*-regulated gene products. *Microbiology*. 2009;155(Pt 7):2235–44.
34. Vuong C, Kocianova S, Voyich JM, Yao Y, Fischer ER, DeLeo FR, et al. A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *J Biol Chem*. 2004;279(52):54881–6.
35. Milling A, Babujee L, Allen C. *Ralstonia solanacearum* extracellular polysaccharide is a specific elicitor of defense responses in wilt-resistant tomato plants. *PLoS One*. 2011;6(1):e15853.
36. Huillet E, Velge P, Vallaëys T, Pardon P. LadR, a new PadR-related transcriptional regulator from *Listeria monocytogenes*, negatively regulates the expression of the multidrug efflux pump MdrL. *FEMS Microbiol Lett*. 2006;254(1):87–94.
37. Li Y, Peng Q, Selimi D, Wang Q, Charkowski AO, Chen X, et al. The plant phenolic compound p-coumaric acid represses gene expression in the *Dickeya dadantii* type III secretion system. *Appl Environ Microbiol*. 2009;75(5):1223–8.
38. Gueneron M, Timmers AC, Boucher C, Arlat M. Two novel proteins, PopB, which has functional nuclear localization signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the hrp-secretion apparatus of *Ralstonia solanacearum*. *Mol Microbiol*. 2000;36(2):261–77.
39. Zuluaga AP, Puigvert M, Valls M. Novel plant inputs influencing *Ralstonia solanacearum* during infection. *Front Microbiol*. 2013;4:349.
40. Jacobs JM, Babujee L, Meng F, Milling A, Allen C. The in planta transcriptome of *Ralstonia solanacearum*: conserved physiological and virulence strategies during bacterial wilt of tomato. *MBio*. 2012;3(4):e00114–12.
41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
42. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18(5):821–9.
43. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, et al. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res*. 2006;34(1):53–65.
44. Xu J, Zheng HJ, Liu L, Pan ZC, Prior P, Tang B, et al. Complete genome sequence of the plant pathogen *Ralstonia solanacearum* strain Po82. *J Bacteriol*. 2011;193(16):4261–2.
45. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5(2):R12.
46. Roth AC, Gonnet GH, Dessimoz C. Algorithm of OMA for large-scale orthology inference. *BMC Bioinformatics*. 2008;9:518.
47. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*. 1999;96(8):4285–8.
48. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7.
49. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*. 2009;26(7):1641–50.
50. David LA, Alm EJ. Rapid evolutionary innovation during an Archaeal genetic expansion. *Nature*. 2011;469(7328):93–6.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Conclusion

Peu de différences ont donc été identifiées entre les écotypes Moko, Brown rot et NPB. De façon intéressante, les divergences observées concernent tout de même plusieurs facteurs de virulence dont des T3E qui ont été associés à la détermination du spectre d'hôte : AvrA (RipAA) module la virulence sur différentes espèces de tabac et GALA4 restreint la virulence sur *A. thaliana* en activant le PTI ('PAMP Triggered Immunity'). Des analyses fonctionnelles ont été effectuées mais n'ont cependant pas pu être intégrées dans la publication dans les délais impartis. Deux T3E présentent des différences entre les Moko et les NPB : RipAA et RipAN.

Le gène *ripAA* est présent chez toutes les souches Moko mais absent des souches NPB : une souche Moko mutant chez laquelle *ripAA* a été supprimé ainsi qu'une souche NPB dans laquelle *ripAA* a été rajouté ont donc été produites. La délétion de ce gène chez les Moko nous permet de déterminer si cet effecteur contribue à la virulence chez le bananier ou bien la tomate tandis que son ajout chez les NPB permet de savoir s'il restreint la virulence chez le bananier ; la caractérisation fonctionnelle permet donc de déterminer si RipAA est facteur de virulence ou d'avirulence dans le cadre de ce modèle. Dans le cadre d'analyses préliminaires, un changement phénotypique majeur a été observé : la souche Moko délétée pour *ripAA* perd complètement son pouvoir pathogène chez le bananier tandis que la souche NPB complétementée avec *ripAA* n'a pas acquis de virulence sur bananier (**Figure 21**).

Le gène *ripAN* est présent chez les souches Moko et NPB mais présente un polymorphisme non-synonyme qui distingue ces deux groupes (HAP) (**Figure 22**) : des souches mutantes Moko et NPB chez lesquelles *ripAN* est absent et des souches mutantes Moko et NPB complétementées avec le gène contenant le polymorphisme NPB et Moko, ont été construites respectivement. Aucun changement phénotypique n'a été observé suite à la délétion ou à la modification du polymorphisme. Ces facteurs de virulence étant des T3E, il est donc probable qu'ils jouent malgré tout un rôle dans le pouvoir pathogène, ce rôle ne semble cependant pas être assez



Figure 21 - Phénotypage (inoculation par le pseudotronc) des mutants de délétion du T3E RipAA. De droite à gauche : souche Moko IIB4 sauvage ; souche Moko IIB4 $\Delta ripAA$; souche Moko IIB4 $\Delta ripAA$; souche NPB sauvage. (Source : P. Prior)

important pour modifier le spectre d'hôte des souches Moko et NPB. Bien que *ripAA* apparaisse désormais comme un déterminant primaire du pouvoir pathogène sur le bananier, cette étude préliminaire va être reproduite en y incluant les mutants de délétion complémenté par *ripAA*.

De façon similaire aux NPB, un unique T3E, *ripAU*, semble avoir été perdu lors de la divergence des Brown rot. Le nombre de T3E polymorphiques par rapport aux souches Moko est cependant bien plus important. Ce résultat pouvait être en partie attendu en considérant la distance génétique plus importante entre les lignées IIB-1 et IIB-3 qu'entre les lignées IIB-4-Moko et IIB-4-NPB. Néanmoins, l'analyse des polymorphismes prend en compte toutes les lignées Moko qui sont également éloignées génétiquement au sein du phylotype II, ces résultats indiquent donc que des mutations de sites conservées dans de nombreux T3E ont probablement été sélectionnées chez les Brown rot. En sachant que l'adaptation de cet écotype aux conditions tempérées est entièrement dépendante de l'environnement de la pomme de terre, il est possible que l'évolution de facteurs de virulence ait joué un rôle dans l'apparition de ce phénotype.

A l'exception du cas de *ripAA*, ces travaux n'ont donc pas permis de fournir une réponse définitive à la question du spectre d'hôte. Le modèle d'étude, constitué notamment par les groupes Moko et NPB, reste un modèle valide et de choix. En effet, les divergences phénotypiques entre ces deux groupes sont flagrantes et l'analyse des génomes complets a définitivement démontré la forte proximité génétique de ces deux groupes. Plusieurs directions sont envisageables pour poursuivre les travaux sur ce modèle. La finition des génomes déjà séquencés permettrait d'augmenter la résolution des analyses comparatives et pourrait potentiellement permettre de caractériser de nouvelles différences plus subtiles. Le séquençage de souches additionnelles appartenant à de nouvelles lignées Moko récemment décrites au Brésil (Albuquerque et al., 2014) enrichirait également l'analyse. La caractérisation fonctionnelle approfondie des T3E représenterait également une avancée majeure dans la compréhension des mécanismes fins de la virulence. La détermination de leur cible et de leur localisation *in planta* ainsi que

les conditions et la cinétique de leur expression et sécrétion permettraient de mieux comprendre leur rôle précis lors de colonisation et de déchiffrer les phénomènes de redondance.

Finalement, nous avons identifié plusieurs régulateurs transcriptionnels qui laissent à penser que l'adaptation à l'hôte repose plus sur des modifications de l'expression plutôt que sur des modifications du génome. Mes travaux se sont donc orientés vers la caractérisation des différences d'expression entre les souches Moko et NPB sur différents hôtes et milieux de culture.

Chapitre 3 – Addendum

Développement d'un outil diagnostique

Mes travaux précédents ont porté sur la caractérisation génomique de groupes d'intérêt chez *R. solanacearum* afin d'identifier des gènes candidats associés à la détermination du spectre d'hôte. Bien qu'elles aient été entreprises dans une perspective fondamentale, les méthodes bioinformatiques mises en place dans le cadre de cette étude peuvent être utilisées à des fins de recherche appliquée. En effet, les techniques de génomique comparative sont également adaptées à l'identification de marqueurs spécifiques permettant de distinguer les groupes et lignées à valeur épidémiologique au sein du complexe d'espèce. Dans cette perspective, une collaboration entre l'Anses (G. Cellier) et le CIRAD (I. Robène, S. Arribat, J. Barthet, P. Grygiel) a permis le développement d'une puce à ADN présentant les caractéristiques requises pour une utilisation dans le domaine du diagnostic : faible coût d'achat et d'investissement, protocole rapide, résultats répétables.

Mon implication dans ces travaux a concerné principalement la sélection de certains répertoires de gènes spécifiques des lignées d'intérêt et la participation aux discussions relatives à l'évolution du protocole expérimental en fonction des résultats intermédiaires. La puce ainsi développée est basée sur la technologie ArrayTube (Alere Technologies) pouvant regrouper jusqu'à 196 sondes sur un support de 3x3mm placé au fond d'un tube d'une taille équivalente à celle d'un Eppendorf 2ml (**Figure 23**). Les échantillons à diagnostiquer doivent se présenter sous forme d'ADN, extrait par exemple de colonies isolées sur milieu solide à partir de tissu végétal infecté, et sont préparés avec des techniques de biologie moléculaire de routine pour une durée totale de 3h. Finalement, la lecture de la puce consiste en une détection colorimétrique (marquage par un conjugué HRP) qui s'effectue à l'aide d'un scanner doté d'une caméra CCD compact nécessitant peu de maintenance.

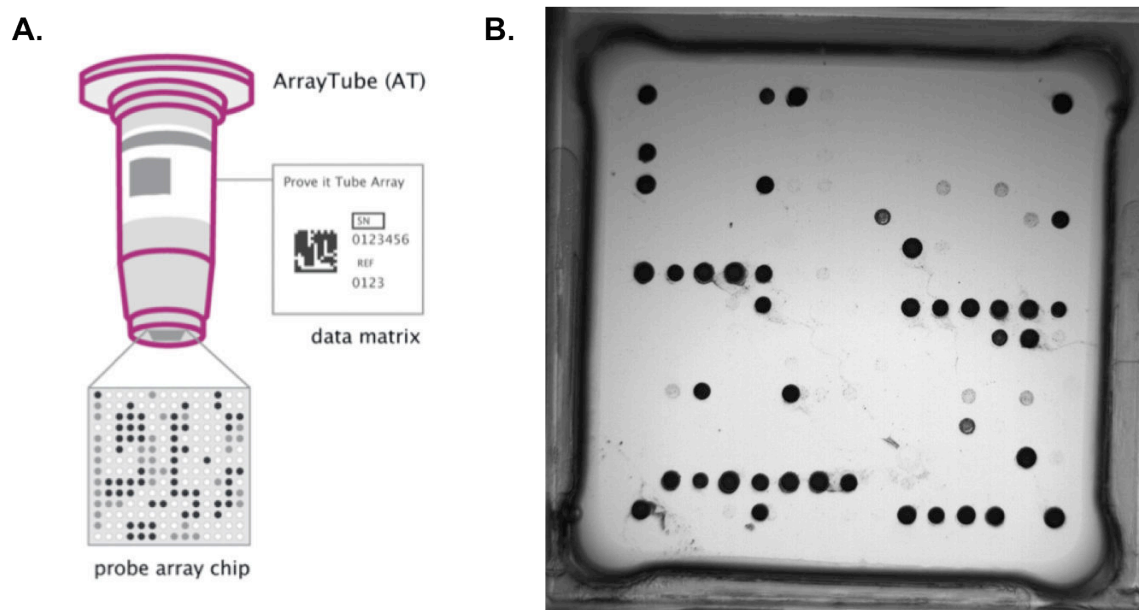


Figure 23 – A. Représentation schématique d'un ArrayTube (Alere Technologies) B. Exemple d'une image obtenue lors de la lecture de la puce.

La technologie ArrayTube est déjà utilisée dans le cas des pathogènes humains, comme par exemple pour la distinction des nombreux variants du virus de la grippe *Haemophilus influenzae*. Elle apparaît donc adaptée pour établir un diagnostic précis et capable d'appréhender la diversité génétique du complexe d'espèce.

Le développement de la puce a naturellement débuté par le design de sondes spécifiques à chacune des lignées génétiques du complexe d'espèce pouvant avoir un intérêt stratégique en terme de diagnostic et d'études épidémiologiques. Les génomes disponibles publiquement, ainsi que ceux séquencés dans le cadre des travaux de génomique comparative précédent (26 génomes au total) ont été utilisés comme point de départ pour le design des sondes par les logiciels ArrayOligoSelector (Bozdech et al., 2003) et OligoArray (Rouillard et al., 2003). Les résultats des deux logiciels ont été sélectionnés et triés afin d'obtenir un design homogène sur les gènes cibles. En effet, les algorithmes étant différents, le design de sondes sur certains gènes ne pouvaient se faire qu'avec un des deux logiciels. Certaines sondes ont été désignées manuellement, notamment pour cibler les amplicons produits par les PCR diagnostiques existantes : 759/760 (universelles *R. solanacearum*); 630/631 (spécifiques Brown rot); 93F/93R (spécifiques Moko) (Cellier et al. 2015 *in press*). Les sondes candidates ainsi obtenues sont longues de 50pb et ont une température de fusion proche de 80°C. Leur spécificité ainsi que celle des cibles contre lesquelles elles sont dirigées a ensuite été validée *in silico* par BLAST, puis *in vitro* par PCR, réduisant le nombre initial de 256 sondes candidates à 100. Au total, 17 groupes d'intérêts peuvent être distingués à l'aide de ces sondes, chaque groupe contient entre 2 et 5 gènes spécifiques et chaque gène est ciblé par 2 sondes distinctes:

- Un groupe « GC » correspondant au core génome de *R. solanacearum*
- Les phylotypes I, II, III et IV
- La lignée IIA7 de la souche type K60 de l'espèce *R. solanacearum*
- Les sous-groupes du phylotype II : IIA et IIB
- Les lignées correspondant aux écotypes du phylotype II : IIB-1 (Brown rot), IIB-3 (Moko), IIB-4 (Moko), IIB-4 (NPB), IIA-6 (Moko) et IIA-24 (Moko)
- Les lignées correspondant aux différentes espèces regroupées au sein du phylotype IV : IV-10 (*R. solanacearum*), IV-9 (*R. syzygii*), IV-10 (BDB)

Afin d'optimiser les coûts de développement, la performance des 100 sondes candidates a été évaluée premièrement à l'aide d'une puce sur lame de verre car le laboratoire était déjà en possession du matériel permettant leur hybridation et leur lecture. Contrairement aux puces haute densité de type Affymetrix, ces puces nécessitent une amplification préalable des cibles avant hybridation. Au total, 50 puces sur lame de verre ont ainsi été spottées par la Plateforme GeT Biopuces de Toulouse (**Figure 24**). Cette étape de criblage a permis notamment de vérifier que chaque sonde s'hybridait correctement et exclusivement avec son groupe cible respectif. En effet, l'hybridation dépend de nombreux paramètres expérimentaux (force ionique, concentration en ADN, fréquence et conditions de lavages, etc) et la prédiction des sondes *in silico* ne permet pas de garantir son efficacité malgré les étapes de validation par BLAST et PCR.

La validation des sondes candidates s'est ensuite poursuivie directement sur ArrayTube. Chaque sonde a été spottée en trois exemplaires repartis aléatoirement afin de tester la répétabilité, tout en évitant un biais éventuel dû à la position du spot sur la puce. De manière similaire aux puces sur lame de verre, les ArrayTube nécessitent une amplification préalable des cibles. Lors de cette étape de développement, l'amplification des cibles par PCR linéaire a été testée et validée afin d'adapter le protocole au contexte particulier du diagnostic. En effet, cette technique permet l'obtention d'un amplicon à l'aide d'une seule amorce par cible et le

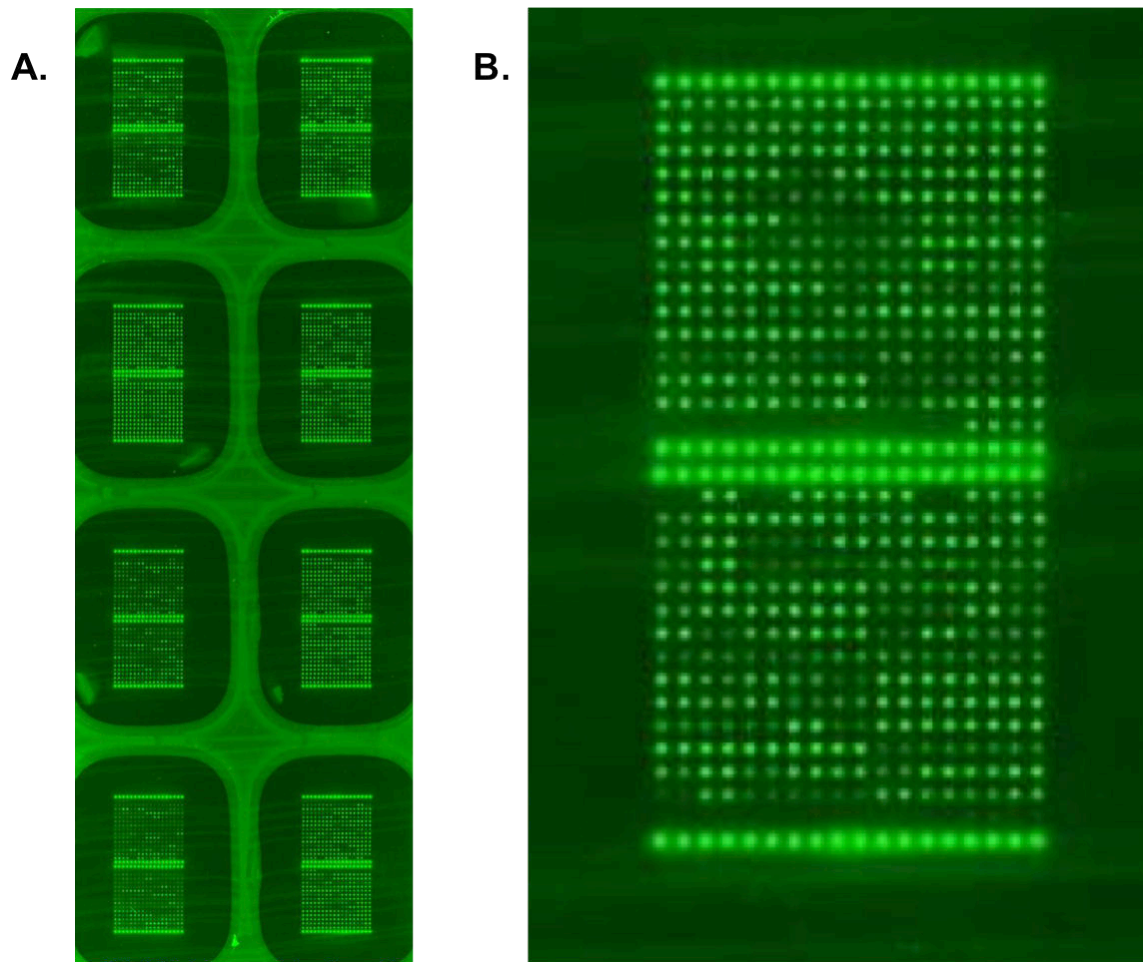


Figure 24 - A. Puces sur lame de verre. B. Agrandissement.

multiplexage de plusieurs centaines d'amorces en une seule réaction. Les volumes de réactifs, le temps de manipulation et les coûts s'en trouvent donc réduits comparés à une PCR conventionnelle équivalente. La validation sur ArrayTube permet non seulement de confirmer que les sondes candidates demeurent spécifiques dans les conditions expérimentales d'hybridation propres à cette technologie, mais également de déterminer le cutoff d'intensité du signal permettant de distinguer les spots positifs des négatifs à l'aide de méthodes statistiques (courbes Receiver Operating Characteristic, seuil de Youden, etc). A l'issue de cette étape, 35 sondes ont été validées.

Le prototype final de cette puce de type ArrayTube peut donc distinguer 17 groupes au sein de RSSC à l'aide de 35 sondes spottées en triplicat. La position phylogénétique d'une souche dans le RSSC est ainsi estimée à un niveau de précision s'étendant au minimum du phylotype et potentiellement, jusqu'au sequevar. Dans une certaine mesure, la détection de lignées associées à des écotypes permet également de prédire le potentiel infectieux d'une souche. En dehors de son application diagnostic direct, cette puce pourrait rendre possible un suivi épidémiologique de *R. solanacearum* bien plus précis que celui permis actuellement par les méthodes officielles. Bien que le prototype final de cet ArrayTube coûte environ 60 € à l'unité, ce qui représente un coût conséquent comparé à une PCR multiplexe (dans le contexte du diagnostic), 35x3 réactions sont réalisées simultanément sur ce prototype. Par ailleurs, 91 spots sont encore libres sur la puce et pourront, dans le futur, être occupés par des sondes dirigées contre des lignées émergentes et/ou encore non caractérisées. Par exemple, de nouvelles lignées Moko ont été récemment identifiées au Brésil (Albuquerque, et al. 2014). Eventuellement, ces spots pourraient aussi être occupés par des sondes détectant d'autres phytopathogènes ayant des hôtes communs avec *R. solanacearum*, afin de faire de cette étude un outil de diagnostic multi-organisme.

Comparé aux méthodes de diagnostic officielles basées sur des PCR, l'ArrayTube possède de nombreux avantages. La détection de chaque cible par deux sondes spottées chacune en trois exemplaires en fait une méthode robuste et le nombre

total de cibles détectées en fait une méthode haut débit. La simplicité du protocole et la standardisation des réactifs et du matériel de lecture permettent une grande répétabilité. Cette technologie possède cependant certaines limitations concernant notamment le matériel de départ. Quand les PCR diagnostic actuelles peuvent s'effectuer directement sur une colonie isolée sans réaliser d'extraction ADN, voire directement à partir d'une suspension de tissu végétal infecté, la puce nécessite obligatoirement un isolement, de grande quantité d'ADN et une amplification intermédiaire. Néanmoins, contrairement aux virus phytopathogènes qui ne nécessitent pas de disposer du virus purifié, l'isolement sur milieu solide est toujours requis pour l'établissement d'un diagnostic bactérien officiel. Par la suite, la sensibilité et la spécificité de la puce seront caractérisées à plus grande échelle par des procédures certifiées afin de pouvoir en publier les résultats et, éventuellement, les intégrer dans la démarche officielle du diagnostic de *R. solanacearum*.

Chapitre 4 – Transcriptomique *in planta*

Le paysage transcriptomique chez *R. solanacearum* a été défini principalement par des travaux sur l'expression des facteurs de virulence (Allen et al., 1997; Brito et al., 1999) et leur modulation par des stimuli environnementaux, comme le contact avec les cellules végétales (Aldon et al., 2000), la densité cellulaire (Clough et al., 1997) ou des études de type IVET (Brown and Allen, 2004). Des travaux plus récents, utilisant des méthodes plus modernes, ont cependant démontré que certains schémas bien établis sont bien plus complexes qu'imaginé auparavant. Ces travaux prouvent, par exemple, que les T3E ne sont pas exclusivement exprimés uniquement durant les étapes précoces, mais tout au long de l'infection (Monteiro et al., 2012), ou encore que le régulateur HrpG, responsable en partie de l'expression des T3E, n'est pas induit uniquement par le contact avec les plantes mais également par des signaux encore inconnus (Zuluaga et al., 2013).

A l'aide d'une puce à ADN, une image globale du transcriptome de la souche modèle GMI1000 a pu être obtenue au cours de l'infection d'un plant de tomate (Jacobs et al., 2012). De la même manière qu'en génomique, notre connaissance transcriptomique à l'échelle du RSSC reste donc très limitée. Suite aux conclusions des analyses de génomique comparatives entreprises précédemment, je me suis donc intéressé aux différences d'expression génique entre nos deux écotypes modèles (souches Moko vs. NPB). En effet, les faibles différences de séquence (polymorphismes) et de contenu en gènes entre les souches Moko et NPB nous ont conduit à émettre l'hypothèse que la variation du spectre d'hôte entre ces deux écotypes est probablement associée à des divergences transcriptomiques plutôt qu'à du contenu en gènes (Ailloud et al. 2015). Les méthodologies associées à l'étude du transcriptome étant encore fastidieuses et coûteuses, le modèle expérimental a été restreint à deux souches de la lignée IIB-4 : une souche Moko (UW163) et une souche NPB (RUN302). Au sein de cette lignée, les souches Moko et NPB partagent plus de 90% de leurs génomes et plus de 80% des loci communs sont identiques au niveau nucléotidique. Cette forte similarité permet donc de comparer directement

les niveaux d'expression de ces deux souches malgré leur spectre d'hôte divergent (**Figure 25**). Afin d'identifier les différences entre ces deux souches de la manière la plus exhaustive possible, l'expression génique de *R. solanacearum* a été caractérisée dans les conditions suivantes à des densités cellulaires similaires (10^8 à 10^9 CFU par ml de milieu ou par g de tissu végétal):

- **Milieu riche liquide CPG** : ce milieu est communément utilisé pour cultiver *R. solanacearum* et a notamment été utilisée comme point de comparaison dans l'étude du transcriptome *in planta* de GMI1000 chez la tomate.
- **Milieu minimum BMM** : ce milieu imite en partie l'environnement végétal et induit l'expression des T3E.
- **Tomate** : cet hôte est commun aux deux souches et permet de déterminer si les deux souches expriment les mêmes facteurs de virulence quand elles colonisent un environnement identique.
- **Bananier**, uniquement pour la souche Moko : cet hôte n'est infecté que par les souches Moko.
- **Melon**, uniquement pour la souche NPB : cet hôte n'est infecté que par les souches NPB.

Au total, chaque souche est donc testée dans quatre conditions. Chaque condition a été répétée trois fois (réplicats biologiques) donnant 24 échantillons d'ARN séquencés en double (réplicats techniques). Certaines étapes ont été limitantes. Premièrement, pour les conditions *in planta*, il est nécessaire de prélever des plants présentant théoriquement des symptômes précoces de flétrissement (D.I. = 1) afin d'obtenir des culots bactériens correspondant aux densités cellulaires requises de 10^8 à 10^9 CFU/g de tissu. En pratique, certains plants auront une densité bactérienne inférieure à celle requise en présence de symptômes ou bien supérieure sans symptômes apparent et cela malgré l'utilisation de plants homogènes et l'infection avec une dose d'inoculum contrôlée. Ce phénomène est d'autant plus amplifié que les souches sélectionnées sont plus agressives que les souches modèles généralement utilisées pour des tests de virulence. Sur tomate, les symptômes peuvent progresser de D.I. = 1 (25% des feuilles flétries) à 3 (75% des feuilles flétries) en moins d'une journée. De plus, l'indice de notation des symptômes n'est pas vraiment adapté au bananier. En effet, le flétrissement progresse de manière

Differential expression analysis between IIB4 Moko and NPB

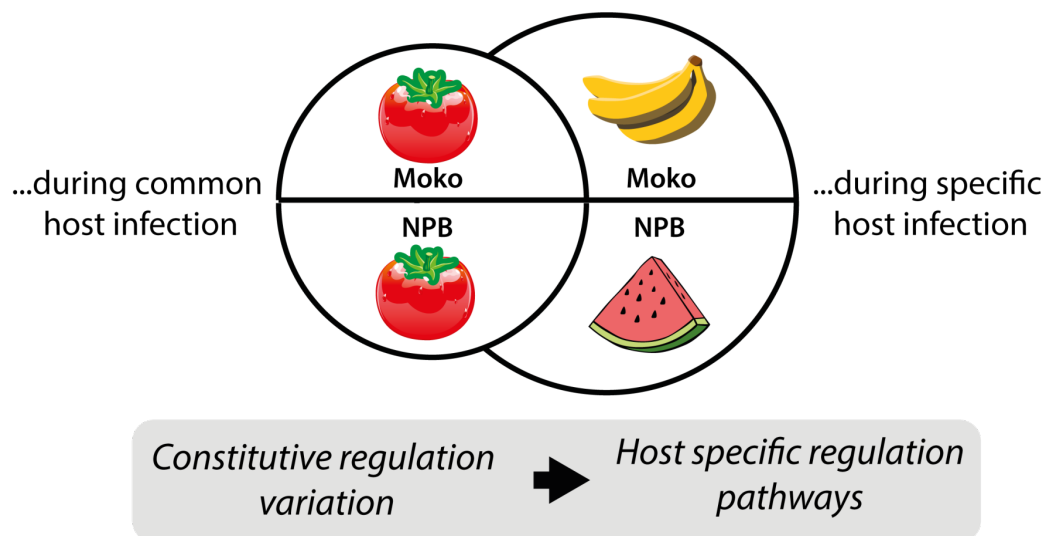


Figure 25 – Représentation schématique du projet transcriptomique.

bien moins radicale sur bananier que sur la tomate ou le melon, l'équivalent de D.I. = 1 correspond sommairement à une seule feuille commençant à flétrir et pouvant être confondu avec la senescence normale de la plante (en chambre climatique). Deuxièmement, l'extraction d'ARN et leur séquençage par RNAseq requièrent des conditions de stérilité strictes et l'utilisation de consommables adaptés. Certains échantillons obtenus lors d'une première expérience au Pole de Protection des Plantes (3P) ne possédaient pas le niveau de qualité nécessaire pour être séquencés. Les échantillons obtenus lors d'une seconde expérience à l'Université du Wisconsin (Laboratoire de Caitilyn Allen) se sont révélés d'une qualité bien plus adéquate. En effet, les conditions de laboratoire, mais surtout le transport international des échantillons sur carboglace du a l'absence d'un centre de séquençage sur l'île de la Réunion, sont probablement à l'origine de l'échec partiel de la première expérience.

A l'aide de méthodes statistiques se basant sur un modèle de distribution binomiale négative, les transcriptomes obtenus peuvent être comparés pair- à -pair afin d'identifier les gènes différentiellement exprimés entre chaque conditions. Avec quatre conditions par souche, cela revient à 28 comparaisons possibles. Toutes les comparaisons ne répondent cependant pas à des questions ayant une signification biologique. La comparaison du transcriptome d'une même souche dans des conditions différentes fournit surtout des indications sur l'influence de l'environnement, tandis que la comparaison des transcriptomes chez deux souches dans des conditions équivalentes indique plutôt les divergences entre écotypes. Par exemple, les niveaux d'expression en milieu riche peuvent être comparés avec ceux observés *in planta* afin d'identifier les gènes potentiellement induits sous le contrôle de la plante; et l'expression différentielle entre les deux souches au sein de la tomate reflète la conservation des stratégies de virulence au sein d'un même environnement.

CES DONNEES ONT ETE DECRITES ET INTERPRETEES SOUS LA FORME D'UN PROJET DE PUBLICATION QUI SERA SOUMIS A LA REVUE BMC GENOMICS ET DONT LE TITRE PRELIMINAIRE EST: « IN PLANTA COMPARATIVE TRANSCRIPTOMICS OF HOST-ADAPTED STRAINS OF *RALSTONIA SOLANACEARUM* ».

1 ***In planta* comparative transcriptomics of host-adapted strains of *Ralstonia***
2 ***solanacearum***

3 Florent Ailloud^{1,2}, Tiffany Lowe³, Caitilyn Allen³, and Philippe Prior^{1,4}

4 **Author's affiliations**

5 ¹ CIRAD, UMRC53 PVBMT, F-97410, Saint-Pierre, La Réunion, France

6 ²Anses - Plant Health Laboratory, F-97410, Saint-Pierre, La Réunion, France

7 ³Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI USA

8 ⁴Département de Santé des Plantes et Environnement (SPE), Inra, France

9 **Corresponding author and E-mail address:** Philippe Prior, philippe.prior@cirad.fr

10 **Abstract**

11 **Background**

12 *Ralstonia solanacearum* is an economically important plant pathogen with an unusually large
13 host range. Moko and NPB are groups of closely related strains that are adapted to distinct
14 hosts. Previous studies have uncovered very few genomic differences that could account for
15 the host range variation of these pathotypes. To better understand the basis of host
16 specificities, we used RNAseq to obtain *R. solanacearum* transcriptome profiles of a Moko
17 strain and an NPB strain under *in vitro* and *in planta* conditions.

18 **Results**

19 RNAs were sequenced from cells grown in rich and minimal media, as well as from bacteria
20 extracted from mid-stage infected tomato, banana and melon plants. Differential expression
21 between a pair of conditions was computed to represent gene expression differences between
22 Moko and NPB. We found that T3Es were globally up-regulated upon plant cell contact in the
23 NPB strain compared with the Moko strain. Siderophore biosynthesis and nitrogen
24 metabolism genes were highly differentially expressed during the colonization of distinct,
25 pathotype-specific hosts.

26 **Conclusions**

27 This study provides the first report of differential gene expression associated with host range
28 variation. Despite minimal genomic divergence, the pathogenicity of Moko and NPB strains
29 is characterized by striking differences in their expression of virulence- and metabolism-
30 related genes.

31 **Keywords:** *Ralstonia solanacearum*, Moko, NPB, host range, transcriptomics, differential
32 expression

33 **Background**

34 Although the virulence mechanisms of model strains such as GMI1000 have been extensively
35 investigated in model host plants such as *Solanum lycopersicum* (tomato) or *Arabidopsis*
36 *thaliana*, *R. solanacearum* is part of a genetically diverse species complex (RSSC) with a
37 cumulative host range that includes more than 250 vegetal species [1]. Very little is known
38 about the prevalence of virulence factors in the RSSC or their contribution to pathogenicity in
39 a large array of hosts. Moreover, recent *in planta* studies have tended to disprove models of
40 pathogenicity regulation that were previously validated *in vitro* [2, 3].

41 Within the RSSC, several groups of strains appear to have undergone adaptation to specific
42 hosts and are grouped in distinct phylogenetic lineages. The most studied group is the brown
43 rot pathotype (IIB-1), which corresponds to strains that are adapted to potatoes in temperate
44 climates [4]. Less studied groups include the Moko and NPB pathotypes. The Moko
45 pathotype is a polyphyletic group (IIB3, IIB4, IIA6, IIA24, etc.) of strains that are pathogenic
46 to banana [4]. The NPB pathotype is monophyletic group that is similar the IIB4 Moko
47 lineage but has lost the ability to infect banana; however, it is virulent towards several
48 members of the *Cucurbitaceae* family [5, 6]. Remarkably, most of the strains in these groups
49 are also pathogenic to tomato plants. The host spectrums of Moko and NPB strains thus
50 comprise both common and exclusive host plants. Because of their genetic closeness on the
51 scale of the RSSC, these strains represent an interesting experimental model that provides a
52 unique opportunity to pinpoint the mechanisms associated with host specificity.

53 We previously used a comparative genomics analysis to characterize the gene content and
54 sequence differences between Moko and NPB strains from the IIB4 lineage. The lack of
55 significant differences between these strains, particularly with respect to virulence factors, led
56 us to hypothesize that the host range variations between Moko and NPB strains are associated
57 with significant differences in gene expression (Ailloud et al. 2015, *submitted*).

58 This work represents the first report of transcriptomic differences associated with host range
59 variation. It is also the first gene expression analysis of *R. solanacearum* strains from the
60 Moko and NPB ecotypes assigned to the IIB4 lineage. RNA samples were extracted from the
61 Moko strain UW163 and the NPB strain RUN302 under biologically relevant conditions:
62 minimal medium, rich medium, tomato infection and either banana infection (UW163) or
63 melon infection (RUN302). Differential gene expression was inferred from pairwise
64 comparisons that were organized to determine either the influence of different environments
65 on individual pathotypes or differences in gene expression between pathotypes. Striking

66 differences in the induction levels of type III effector genes upon plant cell contact, as well as
67 the differential regulation of genes encoding proteins involved in siderophore biosynthesis
68 and nitrogen metabolism, were observed between the two pathotypes.

69 **Materials and methods**

70 **Bacterial strains and growth conditions**

71 The *R. solanacearum* strains UW163 and RUN302 were used in this study. UW163 was
72 isolated in 1967 from plantain in Peru; it belongs to the phylotype IIB4 and to the banana
73 Moko disease-causing ecotype. RUN302 was isolated in 1999 from cucumber in Brazil; it
74 belongs to the phylotype IIB4 and to the NPB ecotype, a pathological variant that is
75 pathogenic towards Cucurbits. Both strains were grown aerobically at 28°C in Boucher's
76 minimal medium (BMM) [7] and rich medium composed of casamino acid, peptone and
77 glucose (CPG) supplemented with yeast extract [8].

78 **RNA extraction for RNAseq**

79 To observe the transcriptomic landscape of the bacterial cells *in vitro* (i.e., outside plants),
80 both strains were grown independently in CPG and BMM to a density of $\sim 6 \times 10^8$ CFU/ml
81 (O.D. = 0.8). RNA was extracted from bacterial pellets as described in Jacobs *et al.* 2012 [3].
82 Three biological replicates were performed for each condition.

83 To observe the transcriptomic landscape of the bacterial cells during colonization in planta,
84 plants were infected with 5×10^8 CFU via soil-soak inoculation as described in Cellier *et al.*
85 2010 [4]. Tomato plants (cv. Bonny Best) were inoculated with either UW163 or RUN302.
86 Banana plants (Cavendish) were inoculated with UW163. Melon plants (cv. Amish) were
87 inoculated with RUN302. Bacteria were extracted from plants stems showing early wilt
88 symptoms at a disease index (D.I.) of 1. Bacteria and RNA were extracted as described in
89 Jacobs *et al.* 2012 [3]. Only stems containing between 10^8 and 10^9 CFU/g of tissue were used
90 for extraction. Three biological replicates were performed for each condition, and each
91 replicate consisted of a pool of RNA extracted from ~ 15 stems.

92 RNA sequencing was carried out by the University of Wisconsin Biotech Center (UWBC).
93 One hundred base pair single-end libraries were sequenced on an HiSeq 2000 (Illumina).
94 Each library was sequenced twice to provide technical replicates. Read quality was controlled
95 with FastQC (www.bioinformatics.bbsrc.ac.uk/projects/fastqc). Low-quality bases and
96 adapters were trimmed using Trimmomatic [9].

97 **Differential expression analysis**

98 Reads from each library were mapped onto their corresponding reference genome using
99 Bowtie2 [10]; the number of uniquely mapped reads for each coding sequences (CDS) was
100 then counted using Bedtools [11]. Differential gene expression between multiple pairs of
101 conditions was computed using edgeR [12] and DEseq2 [13]. Only CDS predicted by both
102 methods with an FDR (corrected p-value) < 0.01 and a $-2 > \logFC > 2$ were considered
103 differentially expressed. UW163 accession number EMBL: PRJEB7430; RUN302 accession
104 number EMBL: PRJEB7433. Genome annotations are available on the MicroScope platform
105 (www.genoscope.cns.fr/agc/microscope).

106 **Results and Discussion**

107 **Initial results**

108 The mapping results indicated efficient bacterial extraction and library preparation. For each
109 sample, rRNA or plant RNA contamination was limited, and 60 to 90% of the total sequenced
110 reads were mapped to CDS (> 10 M mapped reads per sample).

111 Principal component analysis (PCA) of the counts data revealed that the biological and
112 technical replicates were consistent. Differentially expressed genes were obtained by merging
113 the results of the edgeR and DEseq2 packages. Although there were some differences
114 between the two sets of results, as expected, the differences were marginal [14, 15]. An
115 inspection of the MA and volcano plots suggested no bias between the CPM and either the
116 logFC or the FDR adjusted p-value.

117 DGE was computed between multiple pairs of samples to characterize the differences
118 between the IIB4 Moko and NPB strains *in vitro* and *in planta* during the colonization of
119 several hosts (**Fig. 1**) (**Additional file 1**). Overall, many more differentially expressed genes
120 were observed between CPG and BMM media ($\sim 40\%$ of the genome) than in plant-to-plant
121 comparisons (10 to 20% of the genome).

122 To validate our analysis pipeline, the transcriptomic profile in rich medium (CPG) was
123 compared with either minimal medium (BMM) or tomato stems. In both UW163 and
124 RUN302 samples, T3SS components and T3E were highly overexpressed in BMM-cultured
125 cells and in plants compared with CPG-cultured cells. These observations are consistent with
126 previous findings that T3SS and T3E expression is not repressed in minimal medium or in
127 tomato stems at high cell densities ($> 5 \times 10^8$ CFU/ml) in the phylotype I strain GMI1000 [3].
128 Similar observations were made in the banana and melon samples. Together, these results
129 tended to demonstrate that our bioinformatics pipeline is suitable to analyze our model given
130 our experimental conditions.

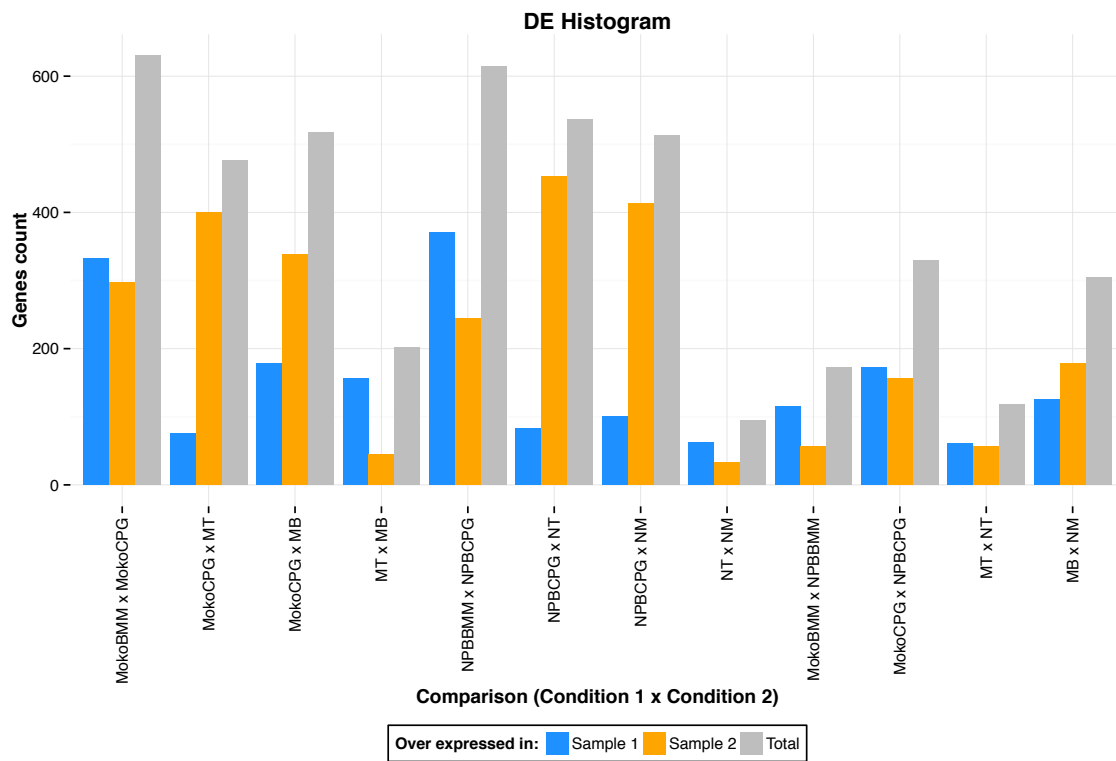


Figure 1 - Differential expression across all tested comparison. Grey bars indicate the total number of differentially expressed genes in the comparison. Blue bars indicate the number of genes more expressed in the first condition while orange bars indicate the number of genes more expressed in the second condition.

131 **Conserved regulatory pathways in similar environments: Moko vs. NPB in CPG, BMM**
132 **and tomato plants**

133 Moko and NPB IIB4 strains share more than 90% of their gene content, and 80% of these
134 common genes display perfect nucleotide identity (Ailloud et al. 2015, submitted). By
135 comparing UW163 and RUN302 gene expression in similar environments (CPG, BMM and
136 tomato stems), we were able to characterize the extent to which these small genomic
137 differences have constitutively modified the transcriptomic profiles of these strains. The
138 detailed results are available in **Additional file 2**.

139 A total of 4148 genes were never differentially expressed between the Moko and NPB strains
140 in CPG, BMM and tomato stems based on a logFC threshold of ± 2 . Conversely, 47 genes
141 were always differentially expressed in similar environments. These values represent ~90%
142 and only ~1% of the genes shared by the two strains. A major portion of the 1% of genes that
143 were differentially expressed followed the same regulatory patterns in CPG, BMM and
144 tomato plants, indicating that they were always more expressed in the same strain irrespective
145 of the conditions. Notably, genes encoding a complete putative amino acid ABC transporter
146 and the catalase KatB were overexpressed in the NPB strain under all three conditions. KatB
147 is involved in the oxidative stress response during plant infection [16]. Similarly, genes
148 encoding a putative hexuronate transporter ExuT2 and a lysozyme protein were
149 overexpressed in the Moko strain in all conditions. The role of ExuT2 has not been
150 characterized; the transporter ExuT is involved in the uptake of compounds released during
151 cell wall degradation, though it does not contribute directly to wild-type virulence [17].
152 However, ExuT was not differentially expressed in any of the three conditions.

153 An additional 407 genes, or ~9% of the genes shared between the strains exhibited
154 inconsistent expression patterns between the different conditions. These genes were
155 differentially expressed in only one or two of the conditions that were compared. This result
156 can be explained by regulatory pathways that are unique to each strain and are specifically
157 activated by compounds present in only one of the three conditions. Of these genes, 9% were
158 only differentially expressed in tomato stems, 50% only in CPG, and 20% only in BMM.
159 These results could also be due to experimental bias if some parameters were not exactly
160 identical between the Moko and NPB strains in one of the conditions.

161 It should be noted that some computational bias could arise when comparing the UW163 and
162 RUN302 strains due to the use of distinct reference genomes during the mapping process. In
163 this case, the estimation of differentially expressed genes is based on homolog families shared
164 by both strains. Even if homolog families are correctly predicted, a given locus can

165 sometimes be fragmented (because it is located on the border of a contig) or have a slightly
166 different START position in one strain compared with the other, which results in more reads
167 being mapped in one locus and eventually gives rise to a false positive.

168 Most of the genes shared by the Moko and NPB strains were thus similarly expressed under
169 similar conditions. A total of 10% of the shared genes were differentially expressed to some
170 degree, but only 1% of the differences were seen under all three conditions and could be
171 considered constitutive differences. Taken together, these results suggest that the Moko and
172 NPB strains use a highly similar virulence strategy to colonize tomato plants.

173 **Plant signal-dependent regulation across hosts and pathotypes: growth in CPG vs.** 174 **plants**

175 To pinpoint genes whose expression is modulated by the interaction with plants, samples
176 extracted from tomato, banana and melon plants were compared with corresponding samples
177 extracted from CPG medium. Because BMM medium appears to at least partially mimic the
178 plant environment [18, 19], CPG medium appeared best suited to serve as a neutral baseline
179 for comparison with plant samples. The detailed results are available in **Additional file 2**.

180 A total of 1071 genes were differentially regulated by plant signals in at least one comparison.
181 These genes were classified into 3 categories in an attempt to properly describe the influence
182 of each plant x strain combination on the transcriptomic landscape (**Fig. 2**).

183 **Category 1.** A total of 123 genes were differentially expressed in all the plant x strain
184 combinations, and the majority of these genes followed a similar pattern of expression.
185 Twenty-two genes exhibited lower expression in plants in both strains, and most were related
186 to amino acid transport and metabolism according to COG predictions, including the *gcv*
187 (glycine catabolism) and *dpp* (dipeptide uptake) operons. Eighty genes showed higher
188 expression in plants, and functions related to energy production or carbohydrate transport and
189 metabolism were highly represented, including the *cox* (cytochrome C oxidase) and *scr*
190 operons [sucrose uptake and catabolism, contributing to virulence [3]]. More importantly, this
191 category includes 10 of the 69 T3Es shared by these two strains (e.g., *ripAM*, *ripAT*, *ripAY*,
192 *ripG3*, *ripR*, *ripC1*, *ripM*, *ripA5*, *ripAN*, and *ripAU*). Interestingly, 4 of these T3Es are part of
193 the core effectome (Ailloud et al., 2015, submitted). Moreover, RipAN is one of the rare
194 genes with non-synonymous polymorphisms that is conserved in all sequenced Moko and
195 NPB lineages. Although these genes are differentially expressed in all conditions and do not
196 diverge across strains and hosts, they provide important information about which genes are
197 modulated by plant signals in these *R. solanacearum* lineages.

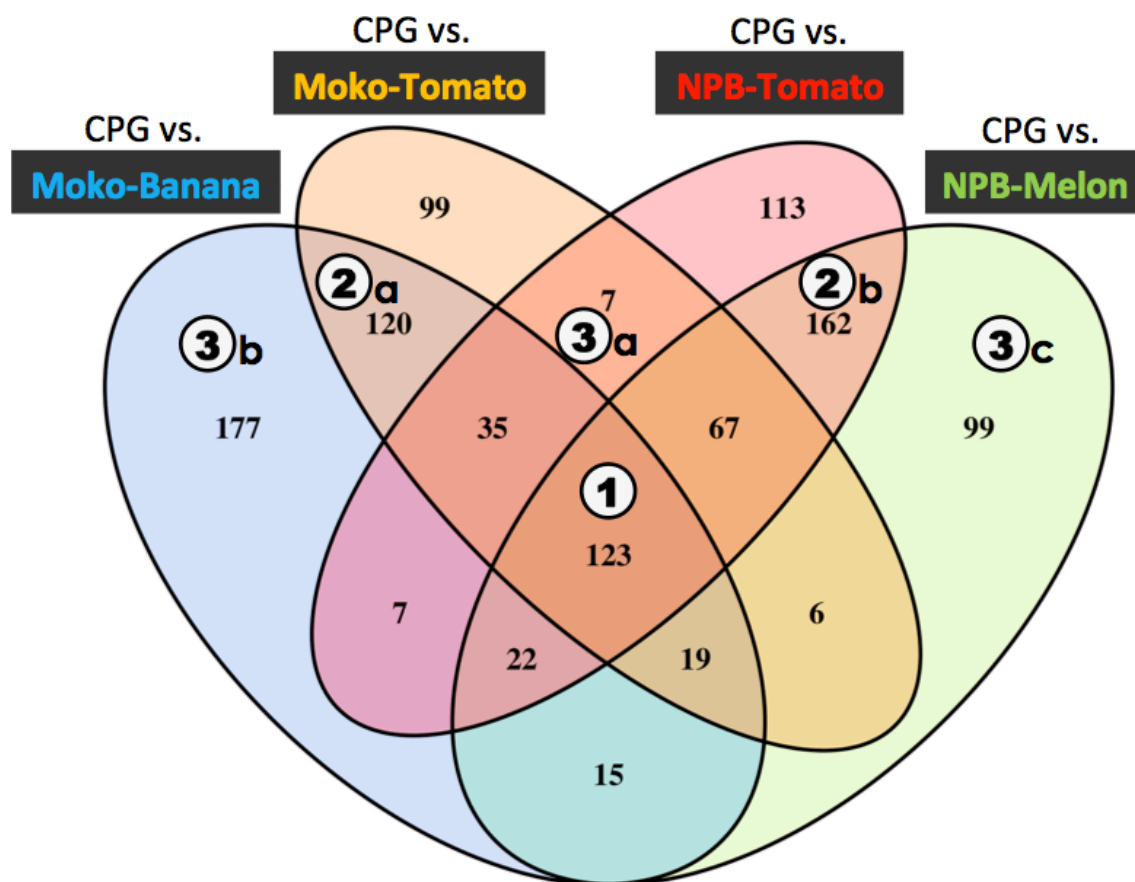


Figure 2 - Venn diagram of differentially expressed genes *in planta* compared to rich medium. Numbers displayed in circles correspond to biologically relevant categories defined in the text.

198 Nonetheless, few genes showed divergent expression patterns across strains and plants.
199 Compared with CPG, 17 genes were up-regulated in the NPB strain in melon plants and in
200 both strains in tomato plants but were down-regulated in the Moko strain in banana plants.
201 Notable among these 17 genes was the *cco* operon (cytochrome C oxidase *cbb-3* type), which
202 is involved in microaerobic energy metabolism [20] and virulence in tomato plants [21], and
203 the anaerobic ribonucleotide reductase genes (*nrdD* and *nrdG*), which are involved in strict
204 anaerobic growth. Structural differences between tomato or melon stems and banana stems,
205 commonly referred to as pseudostems, could explain the Moko-strain-specific expression
206 patterns observed during banana infection.

207 **Category 2a.** A total of 120 genes were differentially expressed only in the Moko strain in
208 both tomato and banana plants. The gene encoding the endoglucanase *egl*, which is related to
209 cell wall degradation, was only overexpressed in Moko strains. Only one T3E, *ripG5*, was up-
210 regulated in the Moko strain in both tomato and banana plants. This effector is part of the core
211 effectome and belongs to the GALA family, which contributes to adaptation to different hosts
212 [22].

213 **Category 2b.** A total of 162 genes were differentially expressed only in the NPB strain in
214 both tomato and melon plants. T3SS components, regulators and 31 T3Es were up-regulated
215 in the NPB strains in both tomato and melon plants (*ripZ*, *ripAC*, *ripY*, *ripF1*, *ripH1*, *ripG6*,
216 *ripAO*, *ripG7*, *ripD*, *ripAJ*, *ripQ*, *ripAR*, *ripP1*, *ripP2*, *ripE2*, *ripN*, *ripL*, *ripAP*, *ripBH*, *ripAE*,
217 *ripX*, *ripAB*, *ripH2*, *ripBC*, *ripH2_2*, *ripAX1*, *ripAQ*, *ripV2*, *ripBD*, *ripAW*, and
218 *RS_T3E_Hyp1*). Remarkably, some of these effectors (e.g., *ripV2*, *ripAQ*) were not present in
219 the IIB4 Moko strains but are widespread throughout the RSSC and in different Moko
220 lineages, indicating that they were most likely lost by the IIB4 Moko lineage.

221 These missing genes represent a major, yet somewhat unexpected, difference in type III
222 system regulation. Although experimental error cannot be excluded given the number of
223 parameters involved, these results nevertheless appear to be quite robust. The T3Es that were
224 found to be exclusively up-regulated *in planta* for the NPB strains (category 2b) were
225 classified as such according to a stringent logFC threshold of ± 2 , which corresponds to a 4-
226 fold change in expression between conditions. In Moko strains, these T3Es were also more
227 expressed in plants than in CPG medium, albeit with a lower fold change. Biologically, this
228 result should be interpreted as a much greater plant-induced expression of some type III genes
229 in the NPB strain rather than as a lack of induction in the Moko strain.

230 **Category 3a.** Surprisingly, there were only 7 genes that were specifically differentially
231 expressed in tomato plants for both the Moko and NPB strains, including the *czcAB* genes,
232 which are involved in heavy metal resistance, which were up-regulated *in planta*.

233 **Category 3b.** A total of 177 genes were differentially expressed exclusively in the Moko
234 strain during the colonization of banana plants. However, this interaction did not specifically
235 modulate any T3Es or major metabolic operons. Several neighbor genes with putative
236 functions related to siderophores and fur, the ferric uptake regulator, were up-regulated in
237 banana plants.

238 **Category 3c.** In total, 99 genes were differentially expressed only in the NPB strain during
239 the colonization of melon plants. Specifically, the T3E *ripTPS* was down-regulated *in planta*,
240 whereas the T3E *ripI* was up-regulated.

241 Overall, the differential expression arising from plant interactions appears to be more strongly
242 associated with strains rather than with hosts. These results suggest that differences in host
243 range between the Moko and NPB strains are not due to specific molecular adaptations to a
244 given host but rather a reshaping of the strain expression profile to be compatible with
245 different hosts. Among the numerous virulence factors characterized in *R. solanacearum*,
246 expression of T3SS and T3E genes appears to be differentially triggered in the Moko and
247 NPB strains upon plant cell contact. Recent studies [2, 3, 23] investigating T3SS regulation
248 have raised concerns regarding the established model; it is therefore difficult to draw
249 definitive conclusions as to the functional consequences of these results, particularly
250 considering that this analysis utilized pathotypes that are poorly characterized relative to the
251 model strain GMI1000.

252 **Host-specific differential gene expression: plant vs. plant comparisons**

253 Plant samples were compared across hosts and across strains. To assess intra-strain variations,
254 Moko and NPB tomato samples were compared with corresponding banana and melon
255 samples, respectively. For the Moko strain, 202 genes were differentially expressed between
256 tomato and banana. For the NPB strain, 95 genes were differentially expressed between
257 tomato and melon. Next, to assess inter-strain variation, the tomato samples were compared
258 with one another, and the banana sample was compared with the melon sample. In the tomato
259 samples, 118 genes were differentially expressed between the Moko and NPB strains. When
260 comparing banana and melon, 305 genes were found to be differentially expressed between
261 the Moko and NPB strains. These results were then cross-referenced to identify candidate
262 genes.

263 As expected, some genes that were previously identified as being differentially expressed
264 between the Moko and NPB strains under similar growth conditions (CPG, BMM and tomato
265 plants) were expressed in a similar fashion when comparing the expression profiles of these
266 strains in banana and melon plants. The hexuronate transporter gene ExuT2 was up-regulated
267 in the Moko strain in banana plants compared with the NPB strain in melon plants. In the
268 NPB strain, this gene was also down-regulated in melon plants compared with tomato plants.
269 These patterns suggest that the function of ExuT2 is beneficial to fitness in banana but could
270 be detrimental in melon. Similarly, the catalase KatB gene was up-regulated in melon plants
271 compared with banana plants. However, this gene was not down-regulated in the Moko strain
272 in banana plants compared with tomato plants. Although the roles of these two genes in
273 virulence are not clearly defined, they were among the most differentially expressed genes
274 between the Moko and NPB strains under all the conditions tested, with a logFC of up to 6,
275 corresponding to a ~65-fold change.

276 In the Moko strain, the ferric uptake regulator *fur*, as well as numerous genes related to
277 siderophore biosynthesis, were up-regulated in banana plants compared with tomato and
278 melon plants. Considering that these same genes were also up-regulated exclusively in banana
279 plants compared with rich medium (category 2a), it appears that they are specifically induced
280 by the banana environment and may contribute to the fitness of the bacteria in this condition.
281 Iron-scavenging by *R. solanacearum* has been investigated to some degree, and siderophores
282 have been shown to not contribute to pathogenicity in tomato plants [24]. However, tomato
283 xylem is not iron limiting, and transgenic tomato plants that display iron-binding activity
284 exhibit increased resistance to *R. solanacearum*, suggesting that iron availability could be a
285 determining factor for virulence. Moreover, some plant pathogenic species of the *Erwinia*
286 family do require siderophore production to exhibit a complete pathogenic phenotype [25,
287 26]. Further investigation of the iron content in banana xylem sap, as well as the regulation of
288 ferric uptake in Moko and NPB strains, would help determine whether siderophores
289 contribute to pathogenicity in banana plants and to the host range of these two strains.

290 In the NPB strain, genes encoding the T3SS regulators HrpX and HrpY and nine T3Es
291 (RipO1, **RipP1**, RipC2, **RipY**, RipS4, **RipE2**, **RipX**, **RipAB**, and **RipAC**) were up-regulated
292 in melon plants compared with banana plants. The expression profiles of RipP1, RipY,
293 RipE2, RipX, RipAB, and RipAC were expected, because these genes were also up-regulated
294 exclusively in this strain compared with rich medium (category 2b). However, these effectors
295 did not display any intra-strain variation. They are thus only induced by plant signals in the
296 NPB strain and are not influenced by differences between the tomato and melon

297 environments. In the Moko strain, RipO1 and RipAB were down-regulated in banana plants
298 compared with tomato plants; their expression in the Moko strain might thus be detrimental to
299 bacterial fitness in banana plants. RipC2 was the only effector that was also up-regulated in
300 the NPB strain compared with the Moko strain in tomato plants. RipAX, RipAB and RipAC
301 are encoded by an HrpB-activated operon (formerly referred as PopABC) and are not required
302 for pathogenicity towards tomato or HR on tobacco. Moreover, RipAB and RipAC belong to
303 the core effectome. PopP1 is a member of the YopJ/AvrBsT family and has been shown to act
304 as an avirulence factor by stopping the GMI1000 strain from being pathogenic toward
305 *Petunia* cultivars and *Arabidopsis thaliana* [27]. The functions of the other T3Es, as well as
306 their roles in virulence, remain to be determined.

307 Finally, the Moko and NPB strains appeared to favor distinct nitrogen metabolic pathways in
308 banana and melon plants. A nitrite reductase encoded by the *nirBD* gene and a nitrate uptake
309 transporter encoded by *narK3* are part of the nitrate assimilation pathway and are up-
310 regulated in banana plants compared with melon plants. *R. solanacearum* is known to
311 scavenge nitrate from the xylem sap, and nitrate assimilation has been shown to contribute to
312 the development of early-stage infection [28]. Conversely, the genes *narHIJGKL*, *aniA*, *norB*,
313 and *hmp* were up-regulated in melon plants compared with banana plants. In the Moko strain,
314 these genes were also down-regulated in banana plants compared with tomato plants. These
315 genes encode enzymatic subunits or regulators and can induce partial denitrification (up to
316 nitrous oxide) and partial nitric oxide detoxification. These functions contribute to both
317 virulence and growth in xylem sap (Dalsing et al. 2015, submitted). The advantages conferred
318 by favoring either of these pathways remain to be determined, but these results suggest that
319 the banana xylem environment might require metabolic adaptation from the bacteria to
320 achieve colonization.

321 **Conclusions**

322 Contrary to our previous comparative genomics results, Moko and NPB strains from the IIB
323 lineage appear to exhibit far more divergent transcriptomics profiles than are suggested by
324 their gene content. Gene expression is generally convergent between pathotypes in similar
325 environmental conditions, but divergences are evident in distinct hosts. In banana plants, the
326 Moko strain tended to favor siderophore biosynthesis and nitrate assimilation, whereas the
327 NPB strain appeared to favor T3E expression and partial denitrification in melon plants.
328 Nevertheless, the differences in gene expression upon plant cell contact appear to be more
329 pathotype-specific than host-specific, particularly with respect to virulence-related genes.

330 Future functional studies should address the degree to which the differential expression of
331 T3Es, siderophores and nitrogen metabolic pathways are involved in the host range variations
332 observed between Moko and NPB strains.

333 **Author contributions**

334 The above work resulted from collaboration between all authors, and each author participated
335 sufficiently in the work to take public responsibility for the content. PP and FA defined the
336 research theme. FA, TL and PP performed the experiments. FA carried out the bioinformatics
337 analysis. FA and PP drafted the article, and CA and TL participated in revising it critically for
338 intellectual content. Each author gave final approval of the version of the manuscript to be
339 published.

340 **Competing interest Section**

341 The authors declare that they have no competing interests.

342 **Acknowledgements**

343 The European Union (ERDF), Conseil Régional de La Réunion, the French Agence Nationale
344 de la Recherche and CIRAD provided financial support. ANSES secured part of the doctoral
345 fellowship awarded to F. Ailloud (contract Number 11-237/BSL).

346 **Figures**

347 **Figure 1 - Differential expression across all tested comparison.**

348 Grey bars indicate the total number of differentially expressed genes in the comparison. Blue
349 bars indicate the number of genes more expressed in the first condition while orange bars
350 indicate the number of genes more expressed in the second condition.

351 **Figure 2 - Venn diagram of differentially expressed genes in planta compared to rich**
352 **medium.** Numbers displayed in circles correspond to biologically relevant categories defined
353 in the text

354 **Additional files**

355 **Additional file 1** Differentially expressed genes in each tested comparison (merged from
356 edgeR and DEseq2 results and filtered with $-2 > \log_{2}FC > 2$ and $FDR < 0.01$)

357 **Additional file 2** Classification of differentially expressed genes according to criteria
358 described in text. In the first sheet, on each line, “CONV”, “DIV”, “ODD” describes the

359 expression pattern of the gene between Moko and NPB in similar conditions: conserved,
360 divergent and inconsistent, respectively. “Plant”, “Moko”, “NPB”, “Tom.”, “Mel.” and
361 “Ban.” corresponds to *in planta* expression categories 1, 2a, 2b, 3a, 3b, 3c, respectively.

362 **References**

- 363 1. Genin S, Denny TP: **Pathogenomics of the *Ralstonia solanacearum* species**
364 **complex**. *Annual review of phytopathology* 2012, **50**:67-89.
- 365 2. Monteiro F, Genin S, van Dijk I, Valls M: **A luminescent reporter evidences active**
366 **expression of *Ralstonia solanacearum* type III secretion system genes throughout**
367 **plant infection**. *Microbiology* 2012, **158**(Pt 8):2107-2116.
- 368 3. Jacobs JM, Babujee L, Meng F, Milling A, Allen C: **The in planta transcriptome of**
369 ***Ralstonia solanacearum*: conserved physiological and virulence strategies during**
370 **bacterial wilt of tomato**. *MBio* 2012, **3**(4).
- 371 4. Cellier G, Prior P: **Deciphering phenotypic diversity of *Ralstonia solanacearum***
372 **strains pathogenic to potato**. *Phytopathology* 2010, **100**(11):1250-1261.
- 373 5. Wicker E, Grassart L, Coranson-Beaudu R, Mian D, Guilbaud C, Fegan M, Prior P:
374 ***Ralstonia solanacearum* strains from Martinique (French West Indies) exhibiting**
375 **a new pathogenic potential**. *Appl Environ Microbiol* 2007, **73**(21):6790-6801.
- 376 6. Wicker E, Coranson-Beaudu R, Cadasse S, William MA: **Emerging strains of**
377 ***Ralstonia solanacearum* in the French West Indies raise new challenges to tomato**
378 **breeders**. *Acta Horticulturae* 2009(808):279-286.
- 379 7. Boucher CA, Barberis P, Trigalet AP, Demery DA: **Transposon mutagenesis of**
380 ***Pseudomonas solanacearum*: isolation of Tn5-induced avirulent mutants**. *Journal*
381 *of General Microbiology* 1985, **131**:2449-2457.
- 382 8. Hendrick CA, Sequeira L: **Lipopolysaccharide-Defective Mutants of the Wilt**
383 **Pathogen *Pseudomonas solanacearum***. *Appl Environ Microbiol* 1984, **48**(1):94-101.
- 384 9. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina**
385 **sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
- 386 10. Langdon WB: **Performance of genetic programming optimised Bowtie2 on**
387 **genome comparison and analytic testing (GCAT) benchmarks**. *BioData mining*
388 2015, **8**(1):1.
- 389 11. Quinlan AR: **BEDTools: The Swiss-Army Tool for Genome Feature Analysis**.
390 *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*
391 2014, **47**:11 12 11-11 12 34.
- 392 12. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for**
393 **differential expression analysis of digital gene expression data**. *Bioinformatics*
394 2010, **26**(1):139-140.
- 395 13. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion**
396 **for RNA-seq data with DESeq2**. *Genome biology* 2014, **15**(12):550.
- 397 14. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson
398 MD: **Count-based differential expression analysis of RNA sequencing data using**
399 **R and Bioconductor**. *Nature protocols* 2013, **8**(9):1765-1786.
- 400 15. Guo Y, Li CI, Ye F, Shyr Y: **Evaluation of read count based RNAseq analysis**
401 **methods**. *BMC genomics* 2013, **14 Suppl 8**:S2.
- 402 16. Flores-Cruz Z, Allen C: ***Ralstonia solanacearum* encounters an oxidative**
403 **environment during tomato infection**. *Molecular plant-microbe interactions :*
404 *MPMI* 2009, **22**(7):773-782.

- 405 17. Gonzalez ET, Allen C: **Characterization of a *Ralstonia solanacearum* operon**
406 **required for polygalacturonate degradation and uptake of galacturonic acid.**
407 *Molecular plant-microbe interactions : MPMI* 2003, **16(6):536-544.**
- 408 18. Rahme LG, Mindrinos MN, Panopoulos NJ: **Plant and environmental sensory**
409 **signals control the expression of hrp genes in *Pseudomonas syringae* pv.**
410 **phaseolicola.** *Journal of bacteriology* 1992, **174(11):3499-3507.**
- 411 19. Tang X, Xiao Y, Zhou JM: **Regulation of the type III secretion system in**
412 **phytopathogenic bacteria.** *Molecular plant-microbe interactions : MPMI* 2006,
413 **19(11):1159-1166.**
- 414 20. Pitcher RS, Watmough NJ: **The bacterial cytochrome cbb3 oxidases.** *Biochimica et*
415 *biophysica acta* 2004, **1655(1-3):388-399.**
- 416 21. Colburn-Clifford J, Allen C: **A cbb(3)-type cytochrome C oxidase contributes to**
417 ***Ralstonia solanacearum* R3bv2 growth in microaerobic environments and to**
418 **bacterial wilt disease development in tomato.** *Molecular plant-microbe interactions*
419 *: MPMI* 2010, **23(8):1042-1052.**
- 420 22. Remigi P, Anisimova M, Guidot A, Genin S, Peeters N: **Functional diversification of**
421 **the GALA type III effector family contributes to *Ralstonia solanacearum***
422 **adaptation on different plant hosts.** *New Phytol* 2011, **192(4):976-987.**
- 423 23. Zuluaga AP, Puigvert M, Valls M: **Novel plant inputs influencing *Ralstonia***
424 ***solanacearum* during infection.** *Frontiers in microbiology* 2013, **4:349.**
- 425 24. Bhatt G, Denny TP: ***Ralstonia solanacearum* iron scavenging by the siderophore**
426 **staphyloferrin B Is controlled by *PhcA*, the global virulence regulator.** *Journal of*
427 *bacteriology* 2004, **186(23):7896-7904.**
- 428 25. Dellagi A, Brisset MN, Paulin JP, Expert D: **Dual role of desferrioxamine in**
429 ***Erwinia amylovora* pathogenicity.** *Molecular plant-microbe interactions : MPMI*
430 1998, **11(8):734-742.**
- 431 26. Expert D: **WITHHOLDING AND EXCHANGING IRON: Interactions Between**
432 ***Erwinia* spp. and Their Plant Hosts.** *Annual review of phytopathology* 1999,
433 **37:307-334.**
- 434 27. Lavie M, Shillington E, Eguiluz C, Grimsley N, Boucher C: **PopP1, a new member**
435 **of the YopJ/AvrRxv family of type III effector proteins, acts as a host-specificity**
436 **factor and modulates aggressiveness of *Ralstonia solanacearum*.** *Molecular plant-*
437 *microbe interactions : MPMI* 2002, **15(10):1058-1068.**
- 438 28. Dalsing BL, Allen C: **Nitrate assimilation contributes to *Ralstonia solanacearum***
439 **root attachment, stem colonization, and virulence.** *Journal of bacteriology* 2014,
440 **196(5):949-960.**

441

442

Conclusion

Les données présentées et analysées dans cette étude représentent une approche originale abordant la question de la variation du spectre d'hôte chez *R. solanacearum* grâce des techniques modernes. Principalement exploratoires, ces travaux constituent la première analyse globale du transcriptome de *R. solanacearum* par séquençage (RNAseq), ainsi que la première analyse de l'expression génique de la lignée IIB-4, des écotypes Moko et NPB et de la colonisation du bananier ou du melon. Globalement, ces données permettent de décrire l'étendue des différences d'expression pouvant exister au plan de l'induction par le contact avec les plantes, de la colonisation d'un hôte commun ou d'hôtes distincts, au sein d'une même souche ou de souches d'écotypes distincts Moko et NPB.

Les différences d'expression les plus importantes ont été observées lors de la comparaison des conditions de culture *in vitro*, entre un milieu riche et un milieu minimum. Des différences bien plus faibles caractérisent les comparaisons entre hôtes végétaux, suggérant une similarité relative entre les environnements vasculaires des plantes. En moyenne, les comparaisons des échantillons plantes avec le milieu minimum pointent moins de gènes différentiellement exprimés que les comparaisons avec le milieu riche, ce qui est cohérent avec le fait que le xylème soit apparenté à un milieu de croissance pauvre pour les bactéries (Pegg, 1985).

Les transcriptomes des souches Moko et NPB de la lignée IIB-4 sont significativement divergents, notamment lors de l'infection d'hôtes exclusifs à leur gamme d'hôte respective : le bananier et le melon. En accord avec notre hypothèse de départ, les différences génomiques précédemment observées entre les deux écotypes semblent se traduire par des différences d'expressions de gènes associés au flétrissement bactérien. Certains effecteurs de type III présentent des différences constitutives d'expression entre les écotypes Moko et NPB notamment au niveau de l'induction par la plante mais aussi durant la colonisation du bananier et du melon. Le manque de données fonctionnelles quant aux rôles et aux cibles de ces effecteurs

ne permet cependant pas d'émettre d'hypothèse sur leur implication dans la variation du spectre d'hôte.

Il a été identifié parmi les gènes les plus différenciellement exprimés entre Moko et NPB et entre toutes les conditions : *exuT2*, un homologue du gène *exuT* codant pour un transporteur d'hexuronate (Gonzalez and Allen, 2003), un produit de la dégradation enzymatique de la paroi végétale ; et *katB*, un gène codant pour une catalase impliquée dans la réponse au stress oxydatif (Flores-Cruz and Allen, 2009). Le gène *exuT2* est 65 fois plus exprimé chez la souche Moko que chez la souche NPB et vice versa pour le gène *katB*. Bien qu'il soit difficile de prédire l'implication directe de ces gènes présentant une forte expression différentielle dans la virulence, il est tout de même très probable qu'ils contribuent au fitness de la bactérie lors de la colonisation. Des gènes présentant une expression différentielle plus faible sont probablement aussi impliqués dans la variation du spectre d'hôte et nécessiteront d'analyser ces données RNAseq de manière plus fine. En effet, les analyses présentées ont été réalisées avec une limite de 4-fold afin d'identifier les différences les plus flagrantes. Dans la littérature, une limite de 2-fold est souvent considérée comme acceptable pour observer des différences biologiquement significatives.

La description de ces données transcriptomiques n'est donc pas encore exhaustive et nécessitera des analyses plus fines afin d'identifier toutes les différences caractérisant les souches Moko et NPB. Les résultats les plus significatifs pourront être confirmés par des méthodes alternatives comme la PCR quantitative et les gènes candidats explorés fonctionnellement par mutagenèse, afin de déterminer si ils sont le moteur ou seulement la conséquence de la variation du spectre d'hôte.

Conclusion générale et perspectives

Les différents travaux entrepris au cours de ma thèse ont été conduits autour d'une thématique commune : la caractérisation des divergences au sein du complexe d'espèces à partir de données génomiques. En effet, malgré les problèmes inhérents au séquençage de 2nde génération, le séquençage d'une dizaine de génomes de souches du phylotype II a constitué le socle de la réorganisation taxonomique du RSSC, du développement d'un outil diagnostique et de l'analyse des caractères génomiques et transcriptomiques propres à certains écotypes.

La réévaluation taxonomique de *R. solanacearum* fut suggérée depuis les années 70 suite à la caractérisation du phylotype IV ayant amenée à la désignation de complexe d'espèces (RSSC). Paradoxalement, le phylotype IV fut le dernier à être séquencé parmi les quatre phylotypes existants, et l'étude des génomes de chaque espèce le composant – *R. solanacearum*, BDB et *R. syzygii* – démontra définitivement qu'il n'était composé que d'une seule et même espèce (Remenant et al., 2011) confirmant ainsi la légitimité d'une évolution taxonomique du RSSC vers un modèle à trois espèces à l'aide de distances génomiques obtenues *in silico* (ANI) : les phylotypes II et IV formeraient chacun une espèce distincte tandis que les phylotype I et III sont élevés au rang de la même espèce [Cette proposition fut par ailleurs poursuivie indépendamment par la communauté à l'aide de techniques complémentaires]. Une équipe Australienne valida indépendamment le schéma à trois espèces avec les méthodes utilisées historiquement en taxonomie consistant en des expériences d'hybridation ADN-ADN et des analyses biochimiques (Safni et al., 2014). Dans le cadre de ma thèse, j'ai exploré la taxonomie de *R. solanacearum* à l'aide de plusieurs algorithmes de distances génomiques (ANI, MUMi et GGDC) appliqués à 29 génomes. Cette approche a été renforcée par une analyse protéomique portant sur la caractérisation de 74 souches par spectrométrie de masse (MALDI-TOF), puis par des analyses phénotypiques portant sur la dénitrification. Bien que les données phénotypiques disponibles actuellement ne supportent qu'un schéma à trois espèces, les distances génétiques calculées par les

algorithmes les plus récents (MUMi et GGDC) suggèrent que la division du RSSC pourrait théoriquement aller jusqu'à cinq espèces distinctes en se basant sur des distances génomiques seuils, qui corrélerent avec la DDH. Ceci renforce l'idée d'une grande hétérogénéité dans le complexe d'espèce, et que la diversité génétique réelle chez *R. solanacearum* est bien plus élevée que celle suggérée par le schéma taxonomique en place actuellement. De plus, nos données concernant la dénitrification étaient en contradiction avec celles du laboratoire Australien. La répétition des analyses ainsi que l'association du phénotype aux voies métaboliques identifiées dans les génomes étudiés ont confirmé la validité de nos données, tout en illustrant l'importance d'intégrer des séquences génomiques dans la démarche taxonomique globale. La classification proposée est la suivante :

- *Ralstonia solanacearum* (phylotype II)
- *Ralstonia pseudosolanacearum* sp. nov. (phylotype I et III)
- *Ralstonia syzygii* sp. nov. (phylotype IV)
 - Sous-espèce *indonesiensis* subsp. nov. (*R. solanacearum* phylotype IV)
 - Sous-espèce *celebesensis* subsp. nov. (BDB)
 - Sous-espèce *syzygii* subsp. nov. (*R. syzygii*)

En décrivant formellement les différences fondamentales existant au sein du RSSC sous la forme de nouvelles espèces, cela permettra entre-autres de dynamiser la recherche fondamentale (choix des modèles d'études), de faciliter la recherche appliquée en diagnostic, en épidémiologie et surtout en sélection pour la résistance variétale.

Les pathotypes n'ont absolument aucune valeur de taxon même pour des germes monomorphes et donc ne sont d'aucune utilité pour la délimitation de nouvelles espèces. Néanmoins, certaines lignées phylogénétiques de *R. solanacearum* sont facilement distinguables grâce à leur spectre d'hôte. Adaptées à certaines espèces végétales, ces lignées sont ainsi considérées comme des écotypes par la communauté des phytobactériologistes. Ces écotypes sont retrouvés chez le phylotype IV incluant les souches BDB, restreintes au bananier, et les souches

R. syzygii, restreintes au girofler. Ils sont aussi présent chez le phylotype II incluant les souches Brown rot, infectant la pomme de terre à basse température, les souches Moko adaptées au bananier, et les souches NPB adaptées aux Cucurbitacée. Dans le cadre de mes travaux de génomique comparative, les écotypes du phylotype II ont été choisis comme modèle d'étude car ils sont proches d'un point de vue génétique et phénotypique tout en étant nettement différenciables. En effet, les trois écotypes Moko, Brown rot et NPB sont tous pathogènes de la tomate, mais sont chacun adaptés à un hôte spécifique. De plus, l'écotype Moko est polyphylétique, certaines lignées Moko du phylotype IIB sont ainsi génétiquement proches des lignées Brown rot et NPB, alors que d'autres sont relativement plus distantes (IIA). Par ailleurs, les souches du phylotype II sont généralement plus disponibles dans les collections locales et internationales et représentent de meilleurs pathosystèmes eu égard à la possibilité d'étudier leur pouvoir pathogène en conditions artificielles, facilitant la caractérisation fonctionnelle de gènes candidats *in planta*. Compte tenu de ces caractéristiques uniques au sein du RSSC, nous avons formulé l'hypothèse que la comparaison directe de ces écotypes permettrait d'identifier des caractères génomiques contribuant à expliquer la variation du spectre d'hôte. D'un point de vue quantitatif, peu de différences génétiques entre les écotypes étaient cependant attendues. En effet, les phylogénies *egl* ou MLST décrivaient déjà les écotypes du phylotype II comme très proches à l'échelle du RSSC ; la comparaison génomique d'une souche Moko IIB-3 et d'une souche Brown rot IIB-1 avait également permis d'observer une grande similarité. Durant cette thèse, la phylogénie basée sur la comparaison de génomes complets a confirmé la proximité génétique globale des écotypes du phylotype II, notamment des lignées Moko et NPB non séquencées auparavant.

Les nouveaux génomes que nous avons produits ont servi dans un premier temps à redéfinir le core-génome du RSSC et par conséquent à identifier les facteurs de virulence conservés chez toutes ces souches, notamment les T3E. Parmi le répertoire de 113 T3E observé au sein des 19 génomes analysés, seuls 14 effecteurs sont conservés dans tous les génomes. Comparé à d'autres bactéries

phytopathogènes possédant un T3SS, le core-effectome de *R. solanacearum* apparaît donc minimal et illustre ainsi la forte plasticité génétique de l'espèce. Ces effecteurs ont probablement un rôle central dans l'établissement du flétrissement bactérien et leur caractérisation fonctionnelle permettra de mieux concevoir les mécanismes fondamentaux mis en jeu lors de l'infection. L'étude des T3E non-conservés du répertoire sera néanmoins requise pour expliquer pourquoi autant d'effecteurs existent à l'échelle du RSSC, quand si peu semblent indispensables au pouvoir pathogène. Paradoxalement, la délétion d'un seul effecteur ne provoque que rarement une faible modification observable du phénotype virulent, mais la délétion du système de sécrétion en lui-même s'accompagne de la perte de la virulence.

La génomique comparative des écotypes retenus dans nos travaux a été effectuée à plusieurs niveaux en faisant appel à un large panel d'outils bioinformatiques. Les génomes des différentes lignées Moko ont été successivement comparés pair-à-pair aux génomes Brown rot, puis NPB. Cette comparaison paire-à-paire des écotypes permet de concentrer l'analyse sur des souches possédant des fonds génétiques adaptés à des hôtes très spécifiques à l'échelle du RSSC. Il est probable que les phénotypes observés chez les écotypes soient le résultat d'une combinaison de plusieurs facteurs génétiques. La distribution des gènes est mosaïque au sein du RSSC, dû à la forte diversité génétique du RSSC et à la capacité naturelle chez *R. solanacearum* d'acquérir des gènes par transfert horizontal. Certains gènes contribuant partiellement à la virulence chez ces écotypes pourraient ainsi avoir été conservés ou transférés chez d'autres lignées. Ces dernières, ne possédant pas le même fond génétique que les écotypes, ne sont néanmoins pas capables de restituer un phénotype identique. De plus, la structure phylogénétique du phylotype II suggère que les Brown rot et NPB ont émergé à partir des lignées IIB-3 et IIB-4 respectivement, et donc potentiellement à partir d'un fond génétique de « type » Moko. Les comparaisons paire-à-paire permettent ainsi d'identifier précisément l'évolution de ce fond génétique. Les gènes acquis ou perdus au sein de chaque écotpe ont été identifiés en observant le contenu en gène spécifique à chaque groupe et les polymorphismes protéiques conservés ont été criblés chez les

gènes partagés par plusieurs écotypes. Enfin, une analyse de réconciliation phylogénétique a déterminé d'une part les gains et pertes de gènes chez chaque lignée et, d'autre part, les transferts horizontaux au sein du RSSC, notamment entre lignées Moko.

Ces multiples analyses ont permis d'identifier principalement des traits spécifiques de chaque écotype et, par conséquent, potentiellement impliqués dans la variation du spectre d'hôte. La caractérisation des transferts horizontaux a servi également à tester plusieurs hypothèses concernant l'origine polyphylétique des souches Moko : i) dispersion verticale du phénotype à partir d'un ancêtre commun à toutes les lignées Moko actuelles ii) propagation du phénotype par transfert horizontal à partir d'une des lignées Moko actuelles. Les résultats obtenus n'ont cependant pas permis de conclure définitivement sur une hypothèse en particulier. L'utilisation d'algorithmes plus sensibles pour détecter les transferts horizontaux, mais principalement une meilleure compréhension des mécanismes mis en jeu dans la virulence sur bananier, permettraient d'appréhender plus efficacement l'apparition et l'évolution de ce phénotype au sein du RSSC, notamment vis-à-vis des souches BDB du phylotype IV qui sont relativement éloignées en terme de distance phylogénétique.

Bien que ces analyses n'aient révélées qu'un faible nombre global de gènes et de polymorphismes spécifiques des écotypes, elles ont permis d'identifier de nombreux effecteurs de type III polymorphiques ou spécifiques d'un écotype. Les analyses fonctionnelles préliminaires sont encourageantes et la délétion de *ripAA* (plus connu sous la dénomination *avrA*, un T3E perdu par les souches NPB mais présent dans toutes les lignées Moko) chez une souche IIB-4 Moko entraîne une perte totale de virulence chez le bananier. Néanmoins, l'effecteur *ripAA* est très répandu dans le RSSC et il est présent chez des souches non pathogènes du bananier. De plus, il est également impliqué dans la modulation de la virulence chez différentes espèces de tabac. La mutagénèse des autres effecteurs identifiés par ces travaux pourrait donc permettre d'identifier des facteurs de virulence indispensables à la pathogénicité sur certains hôtes en fonction du fond génétique.

La compréhension fine du mécanisme d'action de ces effecteurs nécessitera cependant une caractérisation plus approfondie impliquant la cible ou le substrat, la localisation, la cinétique d'expression en fonction de l'hôte, du tissu végétal et de la densité cellulaire. Ces approches seront nécessaires pour déterminer quels sont précisément les processus biologiques impliqués dans la variation du spectre d'hôte. En effet, les effecteurs de type III sont génétiquement et fonctionnellement très variés (et parfois redondant) et leur sécrétion permet une interaction extrême avec l'hôte. De plus, ils représentent, à ce jour, la seule famille de gènes ayant été associés à la gamme d'hôte chez *R. solanacearum*. Les différentes fonctions assurées par les effecteurs de type III décrites dans la littérature suggèrent cependant qu'ils représentent une véritable trousse à outil moléculaire coopérant avec de multiples systèmes n'étant pas eux-mêmes reliés immédiatement à la virulence.

Chez *Salmonella typhimurium*, une entérobactérie pathogène, le burst oxydatif produit par les leucocytes au sein de la muqueuse de l'hôte, et induit par le T3SS, est responsable de l'inflammation intestinale et entraîne en parallèle l'oxydation d'un composé soufré qui est ensuite utilisé comme accepteur terminal d'électron par le système respiratoire des bactéries présentes dans la lumière intestinale (Winter et al., 2010). La respiration permise par ce composé oxydé confère un avantage métabolique à *S. typhi* par rapport aux autres organismes de la flore intestinale et est ainsi directement dépendante de l'inflammation provoquée par le T3SS.

Chez *Pseudomonas aeruginosa*, un pathogène opportuniste, le T3SS induit la formation de pores dans la membrane des macrophages et provoque alors une fuite de nutriments ainsi qu'une molécule chimio-attractante ayant pour conséquence l'accumulation des bactéries autour des macrophages (Dacheux et al., 2001). En amont, le T3SS semble définitivement être relié à l'état nutritionnel de la bactérie, car sa production a été associée au métabolisme du pyruvate (Dacheux et al., 2002), de l'histidine (Rietsch et al., 2004) et du tryptophane (Shen et al., 2008).

Tout comme les facteurs de virulence, le métabolisme d'un agent pathogène a besoin de s'adapter pour survivre au sein d'un hôte dont les conditions environnementales sont souvent très différentes de celles du réservoir bactérien.

L'interdépendance du métabolisme avec un système aussi versatile que celui des effecteurs de type III apparaît donc comme un mécanisme permettant plus de flexibilité dans l'adaptation face à l'évolution des conditions environnementales. Dans le cas de *R. solanacearum*, la multiplicité des réservoirs (eau, sol, mauvaises herbes), des vecteurs (insectes dans le cas de *R. syzygii* subsp. *syzygii*) et surtout des hôtes sensibles a probablement entraîné une forte capacité d'adaptation du métabolisme de la bactérie. Bien que peu de données soient disponibles sur les différences de composition du xylème entre les végétaux concernés (mono et dicotylédones), il est possible qu'elles soient en partie responsables des variations du spectre d'hôte observées chez *R. solanacearum*, en sus des mécanismes de défense et des gènes de résistances de la plante. La compréhension du métabolisme de la plante en complément de celui la bactérie apparaît donc indispensable pour mieux appréhender l'adaptation à l'hôte.

Les résultats de l'analyse transcriptomique des souches IIB-4 Moko et NPB dans divers conditions environnementales, *in vitro* et *in planta*, ont mis en évidence des différences d'expression impliquant le T3SS, le métabolisme de l'azote et des systèmes putatifs d'acquisition du fer (sidérophores). En effet, cette approche a montré que la régulation transcriptionnelle de nombreux effecteurs diffère entre les écotypes Moko et NPB, ainsi qu'entre les différents fonds génétiques hôtes testés. Au contraire, les autres facteurs de virulence caractérisés chez *R. solanacearum* présentent très peu de différences d'expression dans les mêmes conditions, suggérant que les T3E sont potentiellement des déterminants majeurs de la spécificité d'hôte *in planta*. Outre leur complexité aux niveaux génomique et transcriptomique, les T3E pourraient également être modulés au niveau protéomique. En effet, il a été démontré que la sécrétion de certains effecteurs est contrôlée par une protéine chaperonne (incluant AvrA) (Lohou et al., 2014). En parallèle, des différences d'expression des voies métaboliques de l'azote et de l'acquisition du fer ont été observées entre des hôtes distincts (le bananier et le melon). Cependant, la corrélation entre les ressources nutritives existantes au sein de chaque hôte et les variations observées restent encore à prouver. De même, une

éventuelle coordination fonctionnelle entre les T3E et les voies métaboliques que nous avons pointées devra être approfondie. L'expression des voies métaboliques de l'azote est particulièrement remarquable de par son antinomie. En effet la voie d'assimilation et la voie de dénitrification ont une expression opposée chez les souches Moko et NPB lors de l'infection du bananier et du melon, respectivement. Ces deux voies utilisent le nitrate présent dans le milieu, la première le métabolise successivement en ammonium qui sera intégré par la suite dans des composés organiques azotés ; tandis que la dernière l'utilise comme accepteur terminal d'électrons. Bien qu'elles remplissent des fonctions métaboliques radicalement différentes, ces voies contribuent chacune à la virulence (Dalsing and Allen, 2014) (Dalsing et al. 2015, *in press*). L'assimilation du nitrate est associée à l'attachement aux racines et à la production de l'EPS ; la dénitrification permet également de détoxifier le monoxyde d'azote (NO) potentiellement produit par les défenses de la plante. Les variations d'expression pourraient donc aussi bien constituer une adaptation métabolique à l'environnement qu'une adaptation de la stratégie de virulence en réponses aux mécanismes de défenses de chaque hôte. Dans un contexte global, la dénitrification est aussi un processus qui nous a permis de définir plusieurs espèces au sein du RSSC, car seuls les phylotypes I et III sont capables d'effectuer complètement la voie aboutissant à la production de di-nitrogène. Par conséquent, la dénitrification a très probablement constitué un phénotype important dans l'évolution de *R. solanacearum*.

Mes travaux de thèse montrent que nos approches génomique et transcriptomique ont permis de caractériser des différences qualitatives et quantitatives associées à la variation du spectre d'hôte. Dans de futurs travaux, cette problématique pourra être analysée sous différents angles. Les analyses de protéomiques réalisées dans le cadre de cette thèse en vue de délimiter des espèces au sein des RSSC ont démontré que le protéome globale est assez discriminant pour distinguer chaque lignée par spectrométrie de masse. Plus particulièrement, les T3E réellement sécrétés par chaque écotype en fonction des hôtes infectés et le sécrétome en général pourraient

être identifiés à haut débit à l'aide de techniques de spectrométrie de masse qui ne nécessitent pas de déterminer *a priori* les protéines dont on veut étudier la sécrétion ; en comparaison des méthodes utilisant des protéines de fusions (Mukaihara and Tamura, 2009). Par exemple, les protéines surexprimées par les souches Brown rot exposées à de faibles températures ont déjà été caractérisées (Bocsanczy et al., 2014).

Les changements induits chez l'hôte lors de la colonisation sont aussi à prendre en compte et peuvent servir à déduire indirectement le fonctionnement métabolique de la bactérie et les enjeux majeurs de la virulence. Par exemple, la modification du contenu en acide aminés et en sucres au sein du xylème de la tomate au cours de l'infection a été déterminée par chromatographie à échange d'ions (Zuluaga et al., 2013).

Les hypothèses génomiques et transcriptomiques pourront également être approfondies au fur et à mesure des avancées technologiques. Le reséquencage avec des techniques de 3^{ème} génération (PacBio) améliorerait grandement la qualité des génomes par rapport à la 2nde génération et permettrait par exemple de distinguer le chromosome du mégaplasme, ouvrant la voie à l'étude de la dynamique évolutive des deux réplicons. Le séquençage des microARN et de la méthylation révélerait des surcouches de régulations de l'expression encore inexplorées chez *R. solanacearum*, tandis qu'une analyse haut débit des interactions protéines/ADN identifierait, à l'échelle du génome, les régions cibles des éléments trans-régulateurs déjà caractérisés.

De manière générale, l'explosion des « omics » permet aujourd'hui d'envisager une caractérisation approfondie et haut-débit des multiples facettes de la spécificité d'hôte chez *R. solanacearum*. La stratégie et le choix du modèle expérimental restent cependant les plus importants afin d'obtenir des résultats correspondant à une réalité biologique.

La thématique centrale de ma thèse et les stratégies associées ont visé à caractériser des différences au sein du RSSC en se basant sur des souches

présentant une forte diversité aussi bien génétique que phénotype. Inversement, la diversité de ces souches peut être mise à profit en caractérisant les similarités entre ces dernières conformément à l'hypothèse « classique » que les traits génétiques ayant subsistés au cours du temps sont essentiels (i.e. core génome) pour maintenir, entre autres, le phénotype virulent. Dans un 2nd temps il est possible de tirer parti de la disponibilité exponentielle de génomes dans les bases de données en comparant dans un premier temps le core génome i) aux génomes de bactéries telluriques non phytopathogènes afin d'y soustraire, par exemple, les orthologues des voies métaboliques centrales ; puis successivement ii) aux génomes de phytopathogènes colonisant le xylème afin de cerner les gènes potentiellement impliqués dans la colonisation de ce compartiment végétal. Les gènes restants constitueraient ainsi le « patho-core-genome du flétrissement bactérien » (ou MWGS pour 'Minimal Wilt Gene Set') (**Figure 26**).

En dehors du modèle étudié lors de cette thèse, d'autres modèles intéressants (car complémentaires) existent au sein du RSSC. Le phylotype IV héberge des souches dont la gamme d'hôtes est restreinte à un unique représentant (bananier, giroflier) et divergente en fonction des lignées. Elles présentent, par exemple, une opportunité de réaliser des expériences d'évolution expérimentale visant à étendre la gamme d'hôte d'une souche *via* le transfert de facteurs de spécificité d'hôte candidats. L'accroissement de nos connaissances de la diversité génotypique et phénotypique chez *R. solanacearum* permettra probablement de mettre en évidence de nouveaux modèles permettant des approches originales.

Conceptuellement, quel que soit le modèle choisi, les approches de type Omics tendent à établir les bases génétiques associées à un caractère phénotypiques ('forward genetics'). La caractérisation fonctionnelle d'un trait phénotypique particulier nécessite cependant de (re)démontrer le phenotype produit par un trait génétique spécifique ('reverse genetics'). La transition est d'autant plus complexe quand des différences génétiques subtiles sont à l'origine de modification majeures du phénotype. Dans le cas spécifique de l'adaptation à l'hôte chez les phytopathogènes, il a été démontré chez deux espèces voisines d'oomycètes,

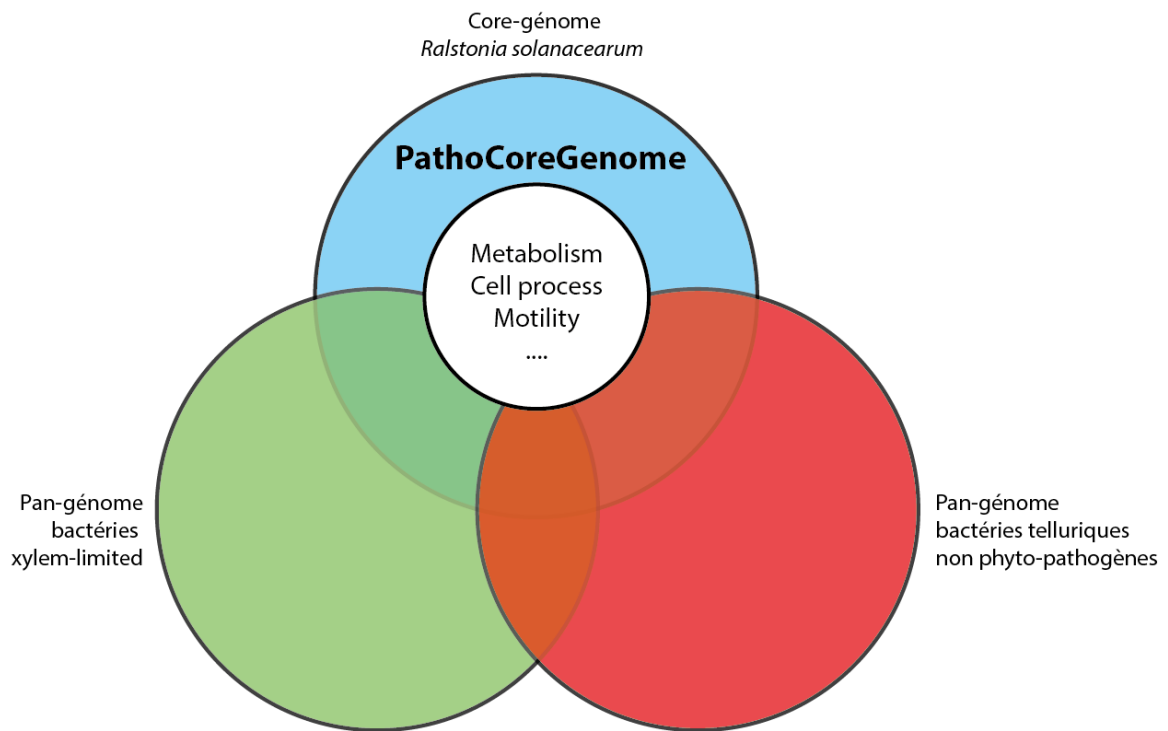


Figure 26 - Représentation schématique du Minimal Wilt Gene Set.

Phytophthora infestans et *P. mirabilis*, adaptées respectivement aux Solanées et aux plantes ornementales du genre *Mirabilis*, qu'un unique polymorphisme au sein d'un effecteur inhibiteur de protéase, identifié initialement par génomique comparative, est responsable de l'adaptation de son activité enzymatique aux protéases spécifiques de chaque espèce végétale (Dong et al., 2014). Chez *R. solanacearum*, des expériences d'évolution expérimentale couplées à du séquençage haut-débit ont permis d'identifier un nombre très limité de mutations acquises suite un gain de fitness chez un hôte distant (absence de flétrissement mais colonisation possible de la tige). La mutation de ces régions a néanmoins démontré par la suite qu'elles étaient effectivement associées au processus d'adaptation observé expérimentalement (Guidot et al., 2014). Dans le cadre de ma thèse, une collaboration avec T. Lowe (UW Madison, Allen's lab) a permis le rapprochement entre l'investigation fonctionnelle de la voie de dégradation des acides hydroxycinnamiques (HCAs) en relation avec la résistance aux défenses des plantes et la distribution génétique de cette voie au sein du complexe d'espèce ainsi que chez d'autres phytopathogènes (Lowe et al., 2015) (**Annexe 2**). Les évolutions futures de la rapidité de mise en œuvre, de l'efficacité et de la fiabilité des méthodes de mutagenèse permettront progressivement de coupler de plus en plus aisément des approches génétiques haut-débit forward et reverse a grande échelle et contribuera ainsi à une vue plus systématique (et intégrative) des questions de biologie que nous avons abordées.

La large gamme d'hôte de *R. solanacearum* représente une des caractéristiques distinguant cette bactérie des autres agents phytopathogènes. Paradoxalement, la plupart des souches étudiées sont rarement testées sur des hôtes autres que les plantes sur lesquelles elles ont été isolées. Un examen approfondie du spectre d'hôte du RSSC dans son ensemble pourrait pourtant permettre d'identifier d'autres écotypes et de déterminer s'il existe une corrélation entre tous les végétaux infectés par une même souche (famille végétale, composition ou structure des tissus, mécanismes de défenses). Un échantillonnage réfléchi de niches écologiques telles que les réservoirs naturels comme le sol, l'eau ou certains végétaux dans lesquels

l'organisme réside à l'état latent comme les innombrables mauvaises herbes (plantes adventices), contribuerait à la compréhension de l'évolution de ce pathogène (compartiments naturels vs. cultivés) grâce à l'identification de souches potentiellement moins aptes à produire du flétrissement. En effet, les collections de bactéries pathogènes sont fréquemment biaisées par le fait que les souches prélevées sont souvent celles ayant effectivement provoqué une infection et représentent par conséquent les génotypes les mieux adaptés à la virulence dans ces conditions environnementales. À leur tour, les souches les plus virulentes sont surreprésentées suite au « bottleneck » artificiel généré par la faible diversité des espèces végétales et des cultivars exploités par l'agriculture moderne.

Références bibliographiques

- Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061-1073.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., and et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651-1656.
- Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J.J., Mayer, P., and Kawashima, E. 2000. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic acids research* 28:E87.
- Albuquerque, G.M., Santos, L.A., Felix, K.C., Rollemberg, C.L., Silva, A.M., Souza, E.B., Cellier, G., Prior, P., and Mariano, R.L. 2014. Moko Disease-Causing Strains of *Ralstonia solanacearum* from Brazil Extend Known Diversity in Paraphyletic Phylotype II. *Phytopathology* 104:1175-1182.
- Aldon, D., Brito, B., Boucher, C., and Genin, S. 2000. A bacterial sensor of plant cell contact controls the transcriptional induction of *Ralstonia solanacearum* pathogenicity genes. *The EMBO journal* 19:2304-2314.
- Allen, C., Gay, J., and Simon-Buela, L. 1997. A regulatory locus, *pehSR*, controls polygalacturonase production and other virulence functions in *Ralstonia solanacearum*. *Molecular plant-microbe interactions : MPMI* 10:1054-1064.
- Altenhoff, A.M., Schneider, A., Gonnet, G.H., and Dessimoz, C. 2011. OMA 2011: orthology inference among 1000 complete genomes. *Nucleic acids research* 39:D289-294.
- Alwine, J.C., Kemp, D.J., and Stark, G.R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* 74:5350-5354.
- Angot, A., Peeters, N., Lechner, E., Vailliau, F., Baud, C., Gentzbittel, L., Sartorel, E., Genschik, P., Boucher, C., and Genin, S. 2006. *Ralstonia solanacearum* requires F-box-like domain-containing type III effectors to promote disease on several host plants. *Proceedings of the National Academy of Sciences of the United States of America* 103:14620-14625.
- Aragaki, M., and Quinon, V. 1965. Bacterial wilt of ornamental gingers (*Hedychium* spp.) caused by *Pseudomonas solanacearum*. *Plant Dis. Rptr.* 49:378-379.
- Arlat, M., Gough, C.L., Zischek, C., Barberis, P.A., Trigalet, A., and Boucher, C.A. 1992. Transcriptional organization and expression of the large *hrp* gene cluster of *Pseudomonas solanacearum*. *Molecular plant-microbe interactions : MPMI* 5:187-193.

- Astua-Monge, G., Freitas-Astua, J., Bacocina, G., Roncoletta, J., Carvalho, S.A., and Machado, M.A. 2005. Expression profiling of virulence and pathogenicity genes of *Xanthomonas axonopodis* pv. citri. *J Bacteriol* 187:1201-1205.
- Auch, A.F., von Jan, M., Klenk, H.P., and Goker, M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in genomic sciences* 2:117-134.
- Bernoux, M., Timmers, T., Jauneau, A., Briere, C., de Wit, P.J., Marco, Y., and Deslandes, L. 2008. RD19, an Arabidopsis cysteine protease required for RRS1-R-mediated resistance, is relocalized to the nucleus by the *Ralstonia solanacearum* PopP2 effector. *Plant Cell* 20:2252-2264.
- Bocsanczy, A.M., Achenbach, U.C., Mangravita-Novo, A., Chow, M., and Norman, D.J. 2014. Proteomic comparison of *Ralstonia solanacearum* strains reveals temperature dependent virulence factors. *BMC genomics* 15:280.
- Boucher, C.A., Gough, C.L., and Arlat, M. 1992. Molecular genetics of pathogenicity determinants of *Pseudomonas solanacearum* with special emphasis on *HRP* genes. *Annual Review of Phytopathology* 30:433-461.
- Boucher, C.A., Barberis, P., Trigalet, A.P., and Demery, D.A. 1985. Transposon mutagenesis of *Pseudomonas solanacearum*: isolation of Tn5-induced avirulent mutants. *Journal of General Microbiology* 131:2449-2457.
- Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B., and DeRisi, J.L. 2003. Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome biology* 4:R9.
- Brenner, D., Staley, J., and Krieg, N. 2000a. Classification of prokaryotic organisms and the concept of Bacterial speciation. in: *Bergey's manual of systematic bacteriology.*, Springer, New York, NY.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. 2000b. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature biotechnology* 18:630-634.
- Brito, B., Marena, M., Barberis, P., Boucher, C., and Genin, S. 1999. prhJ and hrpG, two new components of the plant signal-dependent regulatory cascade controlled by PrhA in *Ralstonia solanacearum*. *Molecular microbiology* 31:237-251.
- Brown, D.G., and Allen, C. 2004. *Ralstonia solanacearum* genes induced during growth in tomato: an inside view of bacterial wilt. *Molecular microbiology* 53:1641-1660.
- Brown, D.G., Swanson, J.K., and Allen, C. 2007. Two host-induced *Ralstonia solanacearum* genes, *acrA* and *dinF*, encode multidrug efflux pumps and contribute to bacterial wilt virulence. *Applied and environmental microbiology* 73:2777-2786.
- Buddenhagen, I., Sequeira, L., and Kelman, A. 1962. Designation of races in *Pseudomonas solanacearum*. *Phytopathology* 52:726.

- Buddenhagen, I.W. 1961. Bacterial wilt of bananas: history and known distribution. *Tropical Agriculture* 38:107-121.
- Castillo, J.A., and Greenberg, J.T. 2007. Evolutionary dynamics of *Ralstonia solanacearum*. *Applied and environmental microbiology* 73:1225-1238.
- Cellier, G., and Prior, P. 2010. Deciphering phenotypic diversity of *Ralstonia solanacearum* strains pathogenic to potato. *Phytopathology* 100:1250-1261.
- Cellier, G., Remenant, B., Chiroleu, F., Lefeuvre, P., and Prior, P. 2012. Phylogeny and population structure of brown rot- and Moko disease-causing strains of *Ralstonia solanacearum* phylotype II. *Applied and environmental microbiology* 78:2367-2375.
- Chen, Y., Ren, X., Zhou, X., Huang, L., Yan, L., Lei, Y., Liao, B., Huang, J., Huang, S., Wei, W., and Jiang, H. 2014. Dynamics in the resistant and susceptible peanut (*Arachis hypogaea* L.) root transcriptome on infection with the *Ralstonia solanacearum*. *BMC genomics* 15:1078.
- Clough, S.J., Lee, K.E., Schell, M.A., and Denny, T.P. 1997. A two-component system in *Ralstonia (Pseudomonas) solanacearum* modulates production of *PhcA*-regulated virulence factors in response to 3-hydroxypalmitic acid methyl ester. *J. Bacteriol.* 179:3639-3648.
- Cohan, F.M. 2001. Bacterial species and speciation. *Systematic biology* 50:513-524.
- Cohan, F.M., and Perry, E.B. 2007. A systematics for discovering the fundamental units of bacterial diversity. *Current biology* : CB 17:R373-386.
- Cohen, A.S., Najarian, D., Smith, J.A., and Karger, B.L. 1988. Rapid separation of DNA restriction fragments using capillary electrophoresis. *J Chromatogr* 458:323-333.
- Coll, N.S., and Valls, M. 2013. Current knowledge on the *Ralstonia solanacearum* type III secretion system. *Microb Biotechnol* 6:614-620.
- Collins, F.S., Morgan, M., and Patrinos, A. 2003. The Human Genome Project: lessons from large-scale biology. *Science* 300:286-290.
- Cook, D., and Sequeira, L. 1994. Strain differentiation of *Pseudomonas solanacearum* by molecular genetic methods. Pages 77-93 in: *Bacterial wilt: the disease and its causative agent, Pseudomonas solanacearum, A.C. Hayward and G.L. Hartman, eds. CAB International, Wallingford, United Kingdom.*
- Cook, D., Barlow, E., and Sequeira, L. 1989. Genetic diversity of *Pseudomonas solanacearum*: detection of restriction fragment polymorphisms with DNA probes that specify virulence and hypersensitive response. *Molecular Plant-Microbe Interactions* 2:113-121.
- Cunnac, S., Boucher, C., and Genin, S. 2004. Characterization of the *cis*-acting regulatory element controlling *HrpB*-mediated activation of the type III secretion system and effector genes in *Ralstonia solanacearum*. *J. Bacteriol.* 186:2309-2318.
- Dacheux, D., Goure, J., Chabert, J., Usson, Y., and Attree, I. 2001. Pore-forming activity of type III system-secreted proteins leads to oncosis of *Pseudomonas aeruginosa*-infected macrophages. *Mol Microbiol* 40:76-85.

- Dacheux, D., Epaulard, O., de Groot, A., Guery, B., Leberre, R., Attree, I., Polack, B., and Toussaint, B. 2002. Activation of the *Pseudomonas aeruginosa* type III secretion system requires an intact pyruvate dehydrogenase aceAB operon. *Infection and immunity* 70:3973-3977.
- Dalsing, B.L., and Allen, C. 2014. Nitrate assimilation contributes to *Ralstonia solanacearum* root attachment, stem colonization, and virulence. *J Bacteriol* 196:949-960.
- Delaspre, F., Nieto Penalver, C.G., Saurel, O., Kiefer, P., Gras, E., Milon, A., Boucher, C., Genin, S., and Vorholt, J.A. 2007. The *Ralstonia solanacearum* pathogenicity regulator HrpB induces 3-hydroxy-oxindole synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 104:15870-15875.
- Deloger, M., El Karoui, M., and Petit, M.A. 2009. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J Bacteriol* 191:91-99.
- Denny, T.P., and Baek, S.R. 1991. Genetic evidence that extracellular polysaccharide is a virulence factor of *Pseudomonas solanacearum*. *Molecular Plant-Microbe Interactions* 4:198-206.
- Deslandes, L., Olivier, J., Peeters, N., Feng, D.X., Khounlotham, M., Boucher, C., Somssich, I., Genin, S., and Marco, Y. 2003. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 100:8024-8029.
- Dong, S., Stam, R., Cano, L.M., Song, J., Sklenar, J., Yoshida, K., Bozkurt, T.O., Oliva, R., Liu, Z., Tian, M., Win, J., Banfield, M.J., Jones, A.M., van der Hoorn, R.A., and Kamoun, S. 2014. Effector specialization in a lineage of the Irish potato famine pathogen. *Science* 343:552-555.
- Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., and Vogelstein, B. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America* 100:8817-8822.
- Drossman, H., Luckey, J.A., Kostichka, A.J., D'Cunha, J., and Smith, L.M. 1990. High-speed separations of DNA sequencing reactions by capillary electrophoresis. *Anal Chem* 62:900-903.
- Eden-Green, S. 1994. Banana Blood Disease. INIBAP Musa Disease Fact Sheet No.3.
- Eden-Green, S.J., and Adhi, E.M. 1986. Sumatra disease of cloves and *Pseudomonas solanacearum*. *Bacterial Wilt Newsletter* 1:2-3.
- Edwards, A., Civitello, A., Hammond, H.A., and Caskey, C.T. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* 49:746-756.
- El-Metwally, S., Hamza, T., Zakaria, M., and Helmy, M. 2013. Next-generation sequence assembly: four stages of data processing and computational challenges. *PLoS computational biology* 9:e1003345.

- Elphinstone, J.G. 2005. The current bacterial wilt situation: A global overview. Pages 9-28 in: Bacterial wilt disease and the *Ralstonia solanacearum* species complex, C. Allen, P. Prior, and A.C. Hayward, eds. APS Press, St Paul, MN.
- Fegan, M. 2005. Bacterial wilt disease of Banana: evolution and ecology. Pages 379-386 in: Bacterial wilt disease and the *Ralstonia solanacearum* species complex, C. Allen, P. Prior, and A.C. Hayward, eds. APS Press, St Paul, MN.
- Fegan, M., and Prior, P. 2005. How complex is the "*Ralstonia solanacearum* species complex". Pages 449-461 in: Bacterial wilt disease and the *Ralstonia solanacearum* species complex, C. Allen, P. Prior, and A.C. Hayward, eds. APS Press, St Paul, MN.
- Fegan, M., and Prior, P. 2006. Diverse members of the *Ralstonia solanacearum* species complex cause bacterial wilts of banana. *Australasian Plant Pathology* 35:93-101.
- Ferre, F. 1992. Quantitative or semi-quantitative PCR: reality versus myth. *PCR methods and applications* 2:1-9.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.
- Flavier, A.B., Clough, S.J., Schell, M.A., and Denny, T.P. 1997. Identification of 3-hydroxypalmitic acid methyl ester as a novel autoregulator controlling virulence in *Ralstonia solanacearum*. *Molecular microbiology* 26:251-259.
- Flores-Cruz, Z., and Allen, C. 2009. *Ralstonia solanacearum* encounters an oxidative environment during tomato infection. *Molecular plant-microbe interactions* : MPMI 22:773-782.
- French, E., Martin, C., and Nydegger, U. 1977. Races and biovars of *Pseudomonas solanacearum* that affect potatoes. *American Potato Journal* 54:10:479-480.
- French, E.R. (1986). Interaction between strains of *Pseudomonas solanacearum*, its hosts and the environment. In *Bacterial wilt disease in Asia and the South Pacific: proceedings n°13 of an international workshop held at PCARRD, Los Banos, Philippines, 8-10 October 1985*, G. Persley, ed (Canberra, Australia: Australian Centre for International Agricultural Research), pp. 99-104.
- Garrity, G.M., Bell, J.A., and Lilburn, T. 2005. *Bergey's Manual of Systematic Bacteriology*. Originally published by Williams & Wilkins, 1984, Baltimore.
- Gäumann, E. 1921. Onderzoekingen over de bloedziekte der bananen op Celebes I. *Mededelingen van het Instituut voor Plantenziekten* 50:47.
- Gäumann, E. 1923. Onderzoekingen over de bloedziekte der bananen op Celebes II. *Mededelingen van het Instituut voor Plantenziekten* 59:45.
- Genin, S., and Denny, T.P. 2012. Pathogenomics of the *Ralstonia solanacearum* species complex. *Annual Review of Phytopathology* 50:67-89.
- Genin, S., Brito, B., Denny, T.P., and Boucher, C. 2005. Control of the *Ralstonia solanacearum* Type III secretion system (Hrp) genes by the global virulence regulator PhcA. *FEBS letters* 579:2077-2081.
- Gillings, M.R., and Fahy, P. 1994. Genomic Fingerprinting: towards a unified view of the *Pseudomonas solanacearum* species complex. Pages 95-112 in: *Bacterial*

- wilt: the disease and its causative agent, *Pseudomonas solanacearum*, A.C. Hayward and G.L. Hartman, eds. CAB International, Wallingford, United Kingdom.
- Gonzalez, A., Plener, L., Restrepo, S., Boucher, C., and Genin, S. 2011. Detection and functional characterization of a large genomic deletion resulting in decreased pathogenicity in *Ralstonia solanacearum* race 3 biovar 2 strains. *Environmental microbiology* 13:3172-3185.
- Gonzalez, E.T., and Allen, C. 2003. Characterization of a *Ralstonia solanacearum* operon required for polygalacturonate degradation and uptake of galacturonic acid. *Molecular plant-microbe interactions : MPMI* 16:536-544.
- Grimault, V., Anais, G., and Prior, P. 1994a. Distribution of *Pseudomonas solanacearum* in the stem tissues of tomato plants with different levels of resistance to bacterial wilt. *Plant Pathol.* 43:663-668.
- Grimault, V., Gelie, B., Lemattre, M., Prior, P., and Schmit, J. 1994b. Comparative histology of resistant and susceptible tomato cultivars infected by *Pseudomonas solanacearum*. *Physiological and Molecular Plant Pathology* 44:105-123
- Guidot, A., Prior, P., Schoenfeld, J., Carrere, S., Genin, S., and Boucher, C. 2007. Genomic structure and phylogeny of the plant pathogen *Ralstonia solanacearum* inferred from gene distribution analysis. *J. Bacteriol.* 189:377-387.
- Guidot, A., Jiang, W., Ferdy, J.B., Thebaud, C., Barberis, P., Gouzy, J., and Genin, S. 2014. Multihost experimental evolution of the pathogen *Ralstonia solanacearum* unveils genes involved in adaptation to plants. *Molecular biology and evolution* 31:2913-2928.
- Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S., and Frazer, K.A. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome biology* 10:R32.
- Hayward, A., Sequeira, L., French, E., El-Nashar, H., and Nydegger, U. 1992. Tropical variant of biovar 2 of *Pseudomonas solanacearum*. *J. Phyto Pathol.* 82: 608. *J. Phyto Pathol:*82:608.
- Hayward, A.C. 1964. Characteristics of *Pseudomonas solanacearum*. *Journal of Applied Bacteriology* 27:265-277.
- Hayward, A.C. 1991. Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. *Annual Review of Phytopathology* 29:65-87.
- He, L.Y., Sequeira, L., and Kelman, A. 1983. Characteristics of strains of *Pseudomonas solanacearum* from China. *Plant Disease* 67:1357-1361.
- Huang, J., and Schell, M. 1995. Molecular characterization of the *eps* gene cluster of *Pseudomonas solanacearum* and its transcriptional regulation at a single promoter. *Molecular microbiology* 16:977-989.
- Huang, J., Yindeeyoungyeon, W., Garg, R.P., Denny, T.P., and Schell, M.A. 1998. Joint transcriptional control of *xpsR*, the universal signal integrator of the *Ralstonia solanacearum* virulence gene regulatory network, by a response

- regulator and a *LysR*-type transcriptional activator. *J. Bacteriol.* 180:2736-2743.
- Idury, R.M., and Waterman, M.S. 1995. A new algorithm for DNA sequence assembly. *Journal of computational biology : a journal of computational molecular cell biology* 2:291-306.
- Jacobs, J.M., Babujee, L., Meng, F., Milling, A., and Allen, C. 2012. The in planta transcriptome of *Ralstonia solanacearum*: conserved physiological and virulence strategies during bacterial wilt of tomato. *MBio* 3.
- Jalan, N., Kumar, D., Andrade, M.O., Yu, F., Jones, J.B., Graham, J.H., White, F.F., Setubal, J.C., and Wang, N. 2013. Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. *BMC genomics* 14:551.
- Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., and Bork, P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic acids research* 36:D250-254.
- Jeong, Y., Cheong, H., Choi, O., Kim, J.K., Kang, Y., Kim, J., Lee, S., Koh, S., Moon, J.S., and Hwang, I. 2011. An HrpB-dependent but type III-independent extracellular aspartic protease is a virulence factor of *Ralstonia solanacearum*. *Mol Plant Pathol* 12:373-380.
- Kang, Y., Liu, H., Genin, S., Schell, M.A., and Denny, T.P. 2002. *Ralstonia solanacearum* requires type 4 pili to adhere to multiple surfaces and for natural transformation and virulence. *Molecular microbiology* 46:427-437.
- Kao, C.C., Barlow, E., and Sequeira, L. 1992. Extracellular Polysaccharide Is Required for Wild-Type Virulence of *Pseudomonas solanacearum*. *J. Bacteriol.* 174:1068-1071.
- Kelman, A. 1953. The bacterial wilt caused by *Pseudomonas solanacearum*. A literature review and bibliography, Raleigh, NC, USA.
- Kelman, A., and Sequeira, L. 1965. Root-to-root spread of *Pseudomonas solanacearum*. *Phytopathology* 55:304-309.
- Konstantinidis, K.T., and Tiedje, J.M. 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 102:2567-2572.
- Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. 2006. The bacterial species definition in the genomic era. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 361:1929-1940.
- Kriventseva, E.V., Rahman, N., Espinosa, O., and Zdobnov, E.M. 2008. OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic acids research* 36:D271-275.
- Lambert, C.D. (2002). Agricultural Bioterrorism Protection Act of 2002; Possession, Use, and Transfer of Biological; Agents and Toxins; Interim Final Rule, A.a.P.H.I. Service, ed (Federal Register), pp. 76907-76938.
- Langer-Safer, P.R., Levine, M., and Ward, D.C. 1982. Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* 79:4381-4385.

- Lavie, M., Shillington, E., Eguiluz, C., Grimsley, N., and Boucher, C. 2002. *PopP1*, a new member of the YopJ/AvrRxv family of type III effector proteins, acts as a host-specificity factor and modulates aggressiveness of *Ralstonia solanacearum*. *Molecular Plant-Microbe Interactions* 15:1058-1068.
- Lebeau, A., Daunay, M.C., Frary, A., Palloix, A., Wang, J.F., Dintinger, J., Chiroleu, F., Wicker, E., and Prior, P. 2011. Bacterial wilt resistance in tomato, pepper, and eggplant: genetic resources respond to diverse strains in the *Ralstonia solanacearum* species complex. *Phytopathology* 101:154-165.
- Lefeuvre, P., Cellier, G., Remenant, B., Chiroleu, F., and Prior, P. 2013. Constraints on genome dynamics revealed from gene distribution among the *Ralstonia solanacearum* species. *PLoS one* 8:e63155.
- Li, L., Stoeckert, C.J., Jr., and Roos, D.S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* 13:2178-2189.
- Li, X., Dorsch, M., Del Dot, T., Sly, L., Stackebrandt, E., and Hayward, A.C. 1993. Phylogenetic studies of the rRNA group II pseudomonads based on 16S rRNA gene sequences. *Journal of Applied Bacteriology* 74:324-329.
- Liang, P., and Pardee, A.B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967-971.
- Liu, H., Zhang, S., Schell, M.A., and Denny, T.P. 2005. Pyramiding unmarked deletions in *Ralstonia solanacearum* shows that secreted proteins in addition to plant cell-wall-degrading enzymes contribute to virulence. *Molecular plant-microbe interactions : MPMI* 18:1296-1305.
- Lohou, D., Turner, M., Lonjon, F., Cazale, A.C., Peeters, N., Genin, S., and Vaillau, F. 2014. HpaP modulates type III effector secretion in *Ralstonia solanacearum* and harbours a substrate specificity switch domain essential for virulence. *Mol Plant Pathol* 15:601-614.
- Lowe, T.M., Ailloud, F., and Allen, C. 2015. Hydroxycinnamic Acid Degradation, a Broadly Conserved Trait, Protects *Ralstonia solanacearum* from Chemical Plant Defenses and Contributes to Root Colonization and Virulence. *Molecular plant-microbe interactions : MPMI* 28:286-297.
- Luckey, J.A., Drossman, H., Kostichka, A.J., Mead, D.A., D'Cunha, J., Norris, T.B., and Smith, L.M. 1990. High speed DNA sequencing by capillary electrophoresis. *Nucleic acids research* 18:4417-4421.
- Mahan, M.J., Slauch, J.M., and Mekalanos, J.J. 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* 259:686-688.
- Mahan, M.J., Heithoff, D.M., Sinsheimer, R.L., and Low, D.A. 2000. Assessment of bacterial pathogenesis by analysis of gene expression in the host. *Annual review of genetics* 34:139-164.
- Mahbou Somo Toukam, G., Cellier, G., Wicker, E., Guilbaud, C., Kahane, R., Allen, C., and Prior, P. 2009. Broad Diversity of *Ralstonia solanacearum* Strains in Cameroon. *Plant Disease* 93:1123-1130.
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S.V., Machado, M.A., Toth, I., Salmond, G., and Foster,

- G.D. 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology* 13:614-629.
- Martin, C., French, E., and Nydegger, U. 1982. Strains of *Pseudomonas solanacearum* affecting Solanaceae In the Americas. *Plant Disease* 66:6:458-460
- McGarvey, J.A., Denny, T.P., and Schell, M.A. 1999. Spatial-Temporal and Quantitative Analysis of Growth and EPS I Production by *Ralstonia solanacearum* in Resistant and Susceptible Tomato Cultivars. *Phytopathology* 89:1233-1239.
- Meng, F., Yao, J., and Allen, C. 2011. A MotN mutant of *Ralstonia solanacearum* is hypermotile and has reduced virulence. *J. Bacteriol.* 193:2477-2486.
- Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., Fischer-Le Saux, M., Manceau, C., Poussier, S., Pruvost, O., Lemaire, C., and Jacques, M.A. 2013. Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PloS one* 8:e58474.
- Miller, J.R., Koren, S., and Sutton, G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95:315-327.
- Milling, A., Babujee, L., and Allen, C. 2011. *Ralstonia solanacearum* extracellular polysaccharide is a specific elicitor of defense responses in wilt-resistant tomato plants. *PloS one* 6:e15853.
- Milling, A., Meng, F., Denny, T.P., and Allen, C. 2009. Interactions with hosts at cool temperatures, not cold tolerance, explain the unique epidemiology of *Ralstonia solanacearum* race 3 biovar 2. *Phytopathology* 99:1127-1134.
- Monteiro, F., Genin, S., van Dijk, I., and Valls, M. 2012. A luminescent reporter evidences active expression of *Ralstonia solanacearum* type III secretion system genes throughout plant infection. *Microbiology* 158:2107-2116.
- Mukaihara, T., and Tamura, N. 2009. Identification of novel *Ralstonia solanacearum* type III effector proteins through translocation analysis of *hrpB*-regulated gene products. *Microbiology* 155:2235-2244.
- Mukaihara, T., Tamura, N., and Iwabuchi, M. 2010. Genome-wide identification of a large repertoire of *Ralstonia solanacearum* type III effector proteins by a new functional screen. *Molecular plant-microbe interactions : MPMI* 23:251-262.
- Myers, E.W. 1995. Toward simplifying and accurately formulating fragment assembly. *Journal of computational biology : a journal of computational molecular cell biology* 2:275-290.
- O'Brien, K.P., Remm, M., and Sonnhammer, E.L. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic acids research* 33:D476-480.
- Occhialini, A., Cunnac, S., Reymond, N., Genin, S., and Boucher, C. 2005. Genome-wide analysis of gene expression in *Ralstonia solanacearum* reveals that the *hrpB* gene acts as a regulatory switch controlling multiple virulence pathways. *Molecular plant-microbe interactions : MPMI* 18:938-949.

- Okoniewski, M.J., and Miller, C.J. 2006. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC bioinformatics* 7:276.
- Orgambide, G., Montrozier, H., Servin, P., Roussel, J., Trigalet-Demery, D., and Trigalet, A. 1991. High heterogeneity of the exopolysaccharides of *Pseudomonas solanacearum* strain GMI 1000 and the complete structure of the major polysaccharide. *Journal of Biological Chemistry* 266:8312-8321.
- Palleroni, N.J., and Doudoroff, M. 1971. Phenotypic characterization and deoxyribonucleic acid homologies of *Pseudomonas solanacearum*. *J. Bacteriol.* 107:690-696.
- Pegg, G. 1985. Life in a black hole – the micro-environment of the vascular pathogen. . *Transactions of the British Mycological Society* 85:1-20.
- Pevzner, P.A., Tang, H., and Waterman, M.S. 2001. An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America* 98:9748-9753.
- Plener, L., Manfredi, P., Valls, M., and Genin, S. 2010. PrhG, a transcriptional regulator responding to growth conditions, is involved in the control of the type III secretion system regulon in *Ralstonia solanacearum*. *J Bacteriol* 192:1011-1019.
- Pop, M., and Salzberg, S.L. 2008. Bioinformatics challenges of new sequencing technology. *Trends in genetics : TIG* 24:142-149.
- Poueymiro, M., and Genin, S. 2009. Secreted proteins from *Ralstonia solanacearum*: a hundred tricks to kill a plant. *Current Opinion in Microbiology* 12:44-52.
- Poueymiro, M., Cunnac, S., Barberis, P., Deslandes, L., Peeters, N., Cazale-Noel, A.C., Boucher, C., and Genin, S. 2009. Two type III secretion system effectors from *Ralstonia solanacearum* GMI1000 determine host-range specificity on tobacco. *Molecular plant-microbe interactions : MPMI* 22:538-550.
- Poussier, S., Trigalet-Demery, D., Vandewalle, P., Goffinet, B., Luisetti, J., and Trigalet, A. 2000. Genetic diversity of *Ralstonia solanacearum* as assessed by PCR-RFLP of the *hrp* gene region, AFLP and 16S rRNA sequence analysis, and identification of an African subdivision. *Microbiology* 146 (Pt 7):1679-1692.
- Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., and Baumeister, K. 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238:336-341.
- Rediers, H., Rainey, P.B., Vanderleyden, J., and De Mot, R. 2005. Unraveling the secret lives of bacteria: use of in vivo expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. *Microbiology and molecular biology reviews : MMBR* 69:217-261.
- Remenant, B., Babujee, L., Lajus, A., Medigue, C., Prior, P., and Allen, C. 2012. Sequencing of K60, type strain of the major plant pathogen *Ralstonia solanacearum*. *J. Bacteriol.* 194:2742-2743.

- Remenant, B., de Cambiaire, J.C., Cellier, G., Jacobs, J.M., Mangenot, S., Barbe, V., Lajus, A., Vallenet, D., Medigue, C., Fegan, M., Allen, C., and Prior, P. 2011. *Ralstonia syzygii*, the Blood Disease Bacterium and some Asian *R. solanacearum* strains form a single genomic species despite divergent lifestyles. *PLoS one* 6:e24356.
- Remenant, B., Coupat-Goutaland, B., Guidot, A., Cellier, G., Wicker, E., Allen, C., Fegan, M., Pruvost, O., Elbaz, M., Calteau, A., Salvignol, G., Mornico, D., Mangenot, S., Barbe, V., Medigue, C., and Prior, P. 2010. Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC genomics* 11:379.
- Rietsch, A., Wolfgang, M.C., and Mekalanos, J.J. 2004. Effect of metabolic imbalance on expression of type III secretion genes in *Pseudomonas aeruginosa*. *Infection and immunity* 72:1383-1390.
- Roach, J.C., Boysen, C., Wang, K., and Hood, L. 1995. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* 26:345-353.
- Roberts, D.P., Denny, T.P., and Schell, M.A. 1988. Cloning of the *egl* gene of *Pseudomonas solanacearum* and analysis of its role in phytopathogenicity. *J. Bacteriol.* 170:1445-1451.
- Roberts, S.J., Eden-Green, S.J., Jones, P., and Ambler, D.J. 1990. *Pseudomonas syzygii*, sp. nov., the cause of Sumatra disease of cloves. *Syst. Appl. Microbiol.* 13:34-43.
- Rodríguez-Ezpeleta, N., Hackenberg, M., and Aransay, A. 2012. *Bioinformatics for High Throughput Sequencing*. Springer.
- Rouillard, J.M., Zuker, M., and Gulari, E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic acids research* 31:3057-3062.
- Safni, I., Cleenwerck, I., De Vos, P., Fegan, M., Sly, L., and Kappler, U. 2014. Polyphasic taxonomic revision of the *Ralstonia solanacearum* species complex: proposal to emend the descriptions of *Ralstonia solanacearum* and *Ralstonia syzygii* and reclassify current *R. syzygii* strains as *Ralstonia syzygii* subsp. *syzygii* subsp. nov., *R. solanacearum* phylotype IV strains as *Ralstonia syzygii* subsp. *indonesiensis* subsp. nov., banana blood disease bacterium strains as *Ralstonia syzygii* subsp. *celebesensis* subsp. nov. and *R. solanacearum* phylotype I and III strains as *Ralstonia pseudosolanacearum* sp. nov. *International journal of systematic and evolutionary microbiology* 64:3087-3103.
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., and Boucher, C.A. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415:497-502.

- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-470.
- Schlöter, M., Leubner, M., Heulin, T., and Hartmann, A. 2000. Ecology and evolution of bacterial microdiversity. *FEMS microbiology reviews* 24:647-660.
- Shalon, D., Smith, S.J., and Brown, P.O. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome research* 6:639-645.
- Shen, D.K., Filopon, D., Chaker, H., Boullanger, S., Derouazi, M., Polack, B., and Toussaint, B. 2008. High-cell-density regulation of the *Pseudomonas aeruginosa* type III secretion system: implications for tryptophan catabolites. *Microbiology* 154:2195-2208.
- Smith, E.F. 1896. A bacterial disease of tomato, eggplant, and Irish potato (*Bacterium solanacearum* nov. sp.). *USDA Div. Veg. Physiol. Pathol. Bull.* 12:191.
- Smith, L.M., Fung, S., Hunkapiller, M.W., Hunkapiller, T.J., and Hood, L.E. 1985. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic acids research* 13:2399-2412.
- Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., and Hood, L.E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674-679.
- Soguilon, C.E., Magnaye, L.V., and Natural, M.P. 1995. La maladie de Bugtok du bananier. INIBAP Maladie des Musa - Fiche technique No.6.
- Sole, M., Popa, C., Mith, O., Sohn, K.H., Jones, J.D., Deslandes, L., and Valls, M. 2012. The *awr* gene family encodes a novel class of *Ralstonia solanacearum* type III effectors displaying virulence and avirulence activities. *Molecular plant-microbe interactions : MPMI* 25:941-953.
- Soneson, C., and Delorenzi, M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC bioinformatics* 14:91.
- Staden, R. 1979. A strategy of DNA sequencing employing computer programs. *Nucleic acids research* 6:2601-2610.
- Swanson, J.K., Yao, J., Tans-Kersten, J., and Allen, C. 2005. Behavior of *Ralstonia solanacearum* Race 3 Biovar 2 During Latent and Active Infection of Geranium. *Phytopathology* 95:136-143.
- Taghavi, M., Hayward, C., Sly, L.I., and Fegan, M. 1996. Analysis of the phylogenetic relationships of strains of *Burkholderia solanacearum*, *Pseudomonas syzygii*, and the blood disease bacterium of banana based on 16S rRNA gene sequences. *International Journal of Systematic Bacteriology* 46:10-15.
- Tamura, N., Murata, Y., and Mukaihara, T. 2005. Isolation of *Ralstonia solanacearum* *hrpB* constitutive mutants and secretion analysis of *hrpB*-regulated gene products that share homology with known type III effectors and enzymes. *Microbiology* 151:2873-2884.

- Tans-Kersten, J., Huang, H., and Allen, C. 2001. *Ralstonia solanacearum* needs motility for invasive virulence on tomato. *J. Bacteriol.* 183:3597-3605.
- Tasset, C., Bernoux, M., Jauneau, A., Pouzet, C., Briere, C., Kieffer-Jacquiod, S., Rivas, S., Marco, Y., and Deslandes, L. 2010. Autoacetylation of the *Ralstonia solanacearum* effector PopP2 targets a lysine residue essential for RRS1-R-mediated immunity in Arabidopsis. *PLoS Pathog* 6:e1001202.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* 278:631-637.
- Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., and Heinen, E. 1999. Housekeeping genes as internal standards: use and limits. *Journal of biotechnology* 75:291-295.
- Trachana, K., Larsson, T.A., Powell, S., Chen, W.H., Doerks, T., Muller, J., and Bork, P. 2011. Orthology prediction methods: a quality assessment using curated protein families. *BioEssays : news and reviews in molecular, cellular and developmental biology* 33:769-780.
- Vasse, J., Frey, P., and Trigalet, A. 1995. Microscopic studies of intercellular infection and protoxylem invasion of tomato roots by *Pseudomonas solanacearum*. *Molecular Plant-Microbe Interactions* 8:241-251.
- Vasse, J., Genin, S., Frey, P., Boucher, C., and Brito, B. 2000. The *hrpB* and *hrpG* regulatory genes of *Ralstonia solanacearum* are required for different stages of the tomato root infection process. *Molecular plant-microbe interactions : MPMI* 13:259-267.
- Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. 1995. Serial analysis of gene expression. *Science* 270:484-487.
- Volkenandt, M., Dicker, A.P., Banerjee, D., Fanin, R., Schweitzer, B., Horikoshi, T., Danenberg, K., Danenberg, P., and Bertino, J.R. 1992. Quantitation of gene copy number and mRNA using the polymerase chain reaction. *Proc Soc Exp Biol Med* 200:1-6.
- Waller, J.M., and Sitepu, D. 1975. Sumatra disease of cloves in Indonesia. *International Journal of Pest Management* 21:141-147.
- Wang, Z., Gerstein, M., and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* 10:57-63.
- Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D., and McClelland, M. 1992. Arbitrarily primed PCR fingerprinting of RNA. *Nucleic acids research* 20:4965-4970.
- Wicker, E., Grassart, L., Coranson-Beaudu, R., Mian, D., and Prior, P. 2009. Epidemiological evidence for the emergence of a new pathogenic variant of *Ralstonia solanacearum* in Martinique (French West Indies). *Plant Pathol.* 58:853-851.
- Wicker, E., Grassart, L., Coranson Beaudu, R., Mian, D., Guilbaud, C., and Prior, P. 2005. Emerging strains of *Ralstonia solanacearum* in Martinique (French West Indies): a case study for epidemiology of bacterial wilt. *Acta Horticulturae*:145-152.

- Wicker, E., Grassart, L., Coranson-Beaudu, R., Mian, D., Guilbaud, C., Fegan, M., and Prior, P. 2007. *Ralstonia solanacearum* strains from Martinique (French West Indies) exhibiting a new pathogenic potential. *Applied and environmental microbiology* 73:6790-6801.
- Winter, S.E., Thiennimitr, P., Winter, M.G., Butler, B.P., Huseby, D.L., Crawford, R.W., Russell, J.M., Bevins, C.L., Adams, L.G., Tsolis, R.M., Roth, J.R., and Baumler, A.J. 2010. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* 467:426-429.
- Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H., and Nishiuchi, Y. 1995. Transfer of two *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. Nov.: Proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. Nov., *Ralstonia solanacearum* (Smith 1896) comb. Nov. and *Ralstonia eutropha* (Davis 1969) comb. Nov. *Microbiol Immunol* 39:11:897-904.
- Yabuuchi, E., Kosako, Y., Oyaizu, H., Yano, I., Hotta, H., Hashimoto, Y., Ezaki, T., and Arakawa, M. 1992. Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiology and Immunology* 36:1251-1275.
- Yao, J., and Allen, C. 2006. Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*. *J. Bacteriol.* 188:3697-3708.
- Yoshimochi, T., Hikichi, Y., Kiba, A., and Ohnishi, K. 2009. The global virulence regulator PhcA negatively controls the *Ralstonia solanacearum* hrp regulatory cascade by repressing expression of the PrhIR signaling proteins. *J. Bacteriol.* 191:3424-3428.
- Zhi, D., Raphael, B.J., Price, A.L., Tang, H., and Pevzner, P.A. 2006. Identifying repeat domains in large genomes. *Genome biology* 7:R7.
- Zuluaga, A.P., Puigvert, M., and Valls, M. 2013. Novel plant inputs influencing *Ralstonia solanacearum* during infection. *Front Microbiol* 4:349.

1 **Hydroxycinnamic acid degradation, a broadly conserved trait, protects *Ralstonia***
2 ***solanacearum* from chemical plant defenses and contributes to root colonization and**
3 **virulence**

4 Tiffany M. Lowe^{1,2}, Florent Ailloud³, and Caitilyn Allen^{2*}

5
6 ¹Microbiology Doctoral Training Program, University of Wisconsin-Madison,
7 Madison, WI, USA

8 ²Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI,
9 USA

10 ³Peuplements Végétaux et Bioagresseurs en Milieu Tropical (UMR PVBMT), INRA-
11 CIRAD, Saint Pierre, La Réunion, France

12
13 *Corresponding Author: Caitilyn Allen; cza@plantpath.wisc.edu; 608-262-9578
14
15

16 **Abstract**

17 Plants produce hydroxycinnamic acid defense compounds (HCAs) to combat
18 pathogens, such as the bacterium *Ralstonia solanacearum*. We showed that an HCA
19 degradation pathway is genetically and functionally conserved across diverse *R.*
20 *solanacearum* strains. Further, a Δfcs (feruloyl-CoA synthetase) mutant that cannot degrade
21 HCAs was less virulent on tomato plants. To understand the role of HCA degradation in
22 bacterial wilt disease, we tested the following hypotheses: HCA degradation helps the
23 pathogen (1) **grow**, as a carbon source; (2) **spread**, by reducing physical barriers HCA-
24 derived; and (3) **survive** plant antimicrobial compounds. Although HCA degradation
25 enabled *R. solanacearum* growth on HCAs *in vitro*, HCA degradation was dispensable for
26 growth in xylem sap and root exudate, suggesting that HCAs are not significant carbon
27 sources *in planta*. Acetyl-bromide quantification of lignin demonstrated that *R.*
28 *solanacearum* infections did not affect the gross quantity or distribution of stem lignin.
29 However, the Δfcs mutant was significantly more susceptible to inhibition by two HCAs:
30 caffeate and *p*-coumarate. Finally, plant colonization assays suggested that HCA
31 degradation facilitates early stages of infection and root colonization. Together, these
32 results indicated that ability to degrade HCAs contributes to bacterial wilt virulence by
33 facilitating root entry and by protecting the pathogen from HCA toxicity.

34

35 **Introduction**

36 Plants produce thousands of phenolic compounds, which play roles in plant
37 development and interactions with microbes (Mandal et al., 2010; Naoumkina et al., 2010).
38 Among these are hydroxycinnamic acids (HCAs), which are monocyclic phenylpropanoid
39 molecules. Roots exude HCAs and related phenolics to chelate metals, thereby facilitating
40 uptake and transport of metals in the xylem sap (Ishimaru et al., 2011). In response to root
41 pathogens, many plants release *de novo* synthesized HCAs into the rhizosphere, and
42 grapevines infected with *Xylella* accumulate HCAs and HCA-conjugates in their xylem sap
43 (Mandal and Mitra, 2008; Lanoue et al., 2010; Wallis and Chen, 2012). HCAs are broadly
44 antimicrobial; they disrupt membrane integrity and decouple the respiratory proton
45 gradient (Fitzgerald et al., 2004; Harris et al., 2010). Additionally, HCAs reinforce protective
46 physical barriers in plants by cross-linking primary cell wall polysaccharides and by
47 serving as precursors for the phenolic polymer lignin (Fry et al., 2000; Naoumkina et al.,
48 2010; Campos et al., 2014).

49 Plants defend their vascular systems with phenolic-storing cells stationed along the
50 xylem (Beckman, 2000). These phenolic storing cells decompartmentalize in response to
51 infection and release phenolics into the xylem lumen, in a process similar to neutrophil
52 degranulation in animal immunity. Exposing tomato roots to a xylem-dwelling fungal
53 vascular wilt pathogen, *Fusarium oxysporum* f. sp. *lycopersici*, leads to increased
54 accumulation of the HCAs ferulate and *p*-coumarate (Mandal and Mitra, 2008).
55 Ultrastructure studies of xylem infected with the vascular pathogen *Ralstonia*

56 *solanacearum* show phenomena consistent with phenolic release (Mueller and Beckman,
57 1984; Grimault et al., 1994; Rahman et al., 1999; Nakaho et al., 2000).

58 *R. solanacearum* causes bacterial wilt disease, which limits production of key crops
59 like potato, banana, peanut, and tomato (Elphinstone, 2005). This soil-dwelling pathogen
60 generally enters hosts through the roots and then colonizes the xylem elements throughout
61 the plant. Extensive colonization of the xylem ultimately blocks water transport, leading to
62 stunting and wilting. *R. solanacearum* strains form a large, heterogeneous species complex
63 that collectively infects hundreds of different plant species (Peeters et al., 2013).

64 Several lines of evidence suggest that hydroxycinnamic acids are involved in tomato
65 interactions with *R. solanacearum*. Quantitative resistance of tomato cultivars against *R.*
66 *solanacearum* is correlated with early expression of phenylalanine ammonia lyase (PAL),
67 which catalyzes the first step in phenylpropanoid biosynthesis (Vanitha et al., 2009).
68 Transcriptomic analysis showed that multiple phenylpropanoid biosynthesis genes are
69 upregulated in *R. solanacearum*-infected, resistant tomato plants compared to healthy
70 plants (Ishihara et al., 2012) (Milling and Allen, unpublished). We previously found that
71 drug efflux pumps protect *R. solanacearum* from the toxicity of many plant defense
72 chemicals, including the HCA caffeate (Brown et al., 2007). More specifically, the genomes
73 of many *R. solanacearum* strains encode an enzymatic pathway that is homologous to a
74 *Pseudomonas fluorescens* pathway that breaks down the HCAs ferulate, *p*-coumarate, and
75 caffeate to central carbon metabolites (Narbad and Gasson, 1998) (Fig. 1). These HCA
76 degradation pathway genes are expressed by *R. solanacearum* cells growing in tomato
77 xylem vessels at the onset of wilt symptoms (Salanoubat et al., 2002; Jacobs et al., 2012).

78 We explored the hypothesis that HCA degradation contributes to bacterial wilt
79 disease. We found that HCA degradation is widely conserved in the *R. solanacearum* species
80 complex. A feruloyl-CoA synthetase mutant (Δfcs) that cannot degrade HCAs had reduced
81 virulence on tomato, delayed colonization of tomato roots, and increased susceptibility to
82 the toxicity of the HCAs caffeate and *p*-coumarate.

83 **Results**

84 **Organization of HCA degradation genes in *R. solanacearum* GMI1000.**

85 HCA degradation enzymes encoded by the genes *fcs*, *fca*, *vdh*, *vanAB*, and *pobA*
86 convert the HCAs *p*-coumarate, caffeate, and ferulate to protocatechuate and acetyl-CoA
87 (Fig. 1A). The β -keto adipate enzymes encoded by the *pca* genes further metabolize
88 protocatechuate to the central carbon metabolites succinyl-CoA and a second acetyl-CoA. In
89 *R. solanacearum* strain GMI1000, the HCA degradation and β -keto adipate genes are
90 organized as five putative operons at three genomic loci (Fig. 1B): *fca*-*vdh*-*fcs* (RSp0225-
91 0227), *vanAB* (RSp0222-0223), *pobA* (RSc02242), *pcaGH* (RSc1141-1142), and *pcaIJFBDC*
92 (RSc2249-2255).

93 We used the Orthologous MAtrix (OMA) browser to investigate the conservation of
94 genes for HCA and protocatechuate degradation across bacteria (Supplementary Table S2).
95 OMA uses a strict algorithm to categorize orthologous proteins from complete publicly
96 available genome sequences (Altenhoff et al., 2011). The eukaryotic β -keto adipate genes
97 are not homologous to the bacterial genes, so we did not analyze eukaryotic genomes. OMA
98 analysis indicated that HCA degradation is a rare trait among the 1,281 bacterial strains
99 considered. Many plant pathogenic bacteria lacked HCA degradation genes, including the

100 necrotroph *Dickeya dadantii* 3937, which uses feruloyl-esterases to cleave HCAs from cell
101 wall polysaccharides (Hassan and Hugouvieux-Cotte-Pattat, 2011), and *Xylella fastidiosa*,
102 which encounters HCAs in grapevine xylem (Wallis and Chen, 2012). Although the OMA
103 database tends to yield false negatives, we gained insight on the prevalence and
104 distribution of this pathway in bacteria because HCA degradation has been functionally
105 characterized in several of the strains included in the OMA database (Parke and Ornston,
106 2003; Plaggenborg et al., 2003; Abdelkafi et al., 2006; Kim et al., 2008; Pérez-Pantoja et al.,
107 2008; Romero-Silva et al., 2013; Campillo et al., 2014). For example, although OMA analysis
108 indicated that *Cupriavidus pinatubonensis* JMP134 (formerly *C. necator* and *Ralstonia*
109 *eutropha*) lacks 4/14 HCA degradation genes, this strain is known to degrade HCAs.
110 Therefore, we hypothesized that the 33 strains containing more than 10/14 HCA
111 degradation genes likely degrade HCAs. These strains are predominantly in genera known
112 to spend part of their lifecycles in soil: *Burkholderia*, *Brucella*, and *Pseudomonas*. OMA
113 analysis identified several plant-associated genera that appear to have functional β -
114 ketoadipate pathways but lack the upstream HCA degrading enzymes: *Xanthomonas*
115 *campestris*, *Rhizobium* spp., and *Agrobacterium* spp. (although Campillo et al. (2014) show
116 that *Agrobacterium* strain C58 degrades HCAs).

117 **The hydroxycinnamic acid (HCA) degradation pathway is broadly conserved in the *R.***
118 ***solanacearum* species complex.**

119 Because hydroxycinnamic acids are common plant metabolites, the ability to
120 degrade these metabolites could benefit *R. solanacearum*. To explore the genetic
121 conservation of the HCA degradation pathway in the large and heterogeneous *R.*

122 *solanacearum* species complex, we searched for homologs of HCA degradation genes in the
123 genomes of 23 available *R. solanacearum* strains (Fig. 2A). Only two strains lacked multiple
124 HCA degradation genes: Phylotype IIA strain K60 and Phylotype IV Blood Disease
125 Bacterium (BDB) strain R229. We identified homologs of each HCA degradation gene in the
126 remaining 21 strains (91%), but in eight of these strains, one or more genes were located
127 on a contig border or were annotated as putative pseudogenes.

128 The number of potential pseudogenes and genes lying on contig borders made it
129 difficult to predict the HCA degradation ability of a third of the sequenced strains, so we
130 functionally characterized the HCA degradation ability of all available strains. We could not
131 analyze strains Po82, FQY-4, and Y45 since the authors of these published genomes would
132 not share their strains (Li et al., 2011; Xu et al., 2011; Cao et al., 2013). Each strain was
133 tested for its ability to grow on the HCAs ferulate (Fer) and *p*-coumarate (Cou) as well as on
134 the pathway intermediates vanillin (Van), vanillate (VA), *p*-hydroxybenzoate (HBA), and
135 protocatechuate (PCA) (see Fig. 1A). Most strains grew on all tested compounds, except
136 where genomic data indicated an incomplete pathway (Fig 2A). For example, because it
137 lacks *fcs*, *fca*, *vdh*, *vanB*, *pobA*, and *pcaH*, strain K60 did not grow on any tested substrate.
138 Similarly, BDB strain R229, lacking *fcs*, *fca*, *vdh*, and *vanB*, grew only on protocatechuate.

139 In several cases, bioinformatic data did not accurately predict biological function.
140 Surprisingly, although strains CFBP2957 and Molk2 grew on ferulate, they did not grow on
141 the ferulate degradation intermediates vanillin (in the case of CFBP2957) or vanillin and
142 vanillate (in the case of Molk2). It is possible that higher sensitivity to toxicity of vanillin
143 and vanillate prevented these strains from growing. Several strains with a putatively

144 pseudogenized *fcs* displayed contrasting growth phenotypes. While strain UW179 grew on
145 all compounds, the phylotype II sequevar 1 (Race 3 biovar 2) strains UW491 and UW551
146 unexpectedly did not grow on ferulate, *p*-coumarate, vanillin, or vanillate. Further, strains
147 CIP417 and CMR15 grew on intermediate metabolites but not ferulate or *p*-coumarate even
148 though these strains apparently possess complete sets of HCA degradation genes. Taken
149 together, these analyses indicate that most *R. solanacearum* strains can degrade at least
150 some HCAs; the unexpected positive and negative results for growth on HCAs also highlight
151 the importance of functional experiments to confirm genomic analyses.

152 To determine whether HCA degradation contributes to bacterial wilt disease, we
153 created an *fcs* deletion mutant in the background of phylotype I strain GMI1000. Hereafter,
154 strain GMI1000 is referred to as wildtype or WT and the GMI1000 deletion mutant lacking
155 the feruloyl-CoA synthetase open reading frame is referred to as the Δfcs mutant. While WT
156 used HCAs and intermediate phenolics as a carbon source (Fig. 2), the Δfcs mutant did not
157 grow on the HCAs *p*-coumarate, caffeate, or ferulate, as predicted (Fig. 2B). Additionally,
158 the mutant grew as well as wildtype on all pathway intermediates (data not shown). This
159 result confirmed the bioinformatic annotation of this gene, and also confirmed the deletion
160 of the *fcs* gene. Genetic complementation of the mutant with the cloned *fcs* operon under
161 control of the native promoter restored its growth on HCAs (Fig. 2B).

162 **HCA degradation contributes to *R. solanacearum* virulence on tomato.**

163 We used a naturalistic soil soak virulence assay to measure the contribution of HCA
164 degradation to *R. solanacearum* virulence on tomato. Bacterial suspensions were poured

165 into the soil of individually potted unwounded tomato plants, and symptom development
166 was measured over time.

167 The Δfcs mutant displayed a modest but significant reduction in virulence on plants
168 grown at the tropical temperature 28°C (Fig. 3A; repeated measure ANOVA; $P=0.0123$).
169 The reduced virulence of the Δfcs mutant did not result from an altered rate of symptom
170 progression. Once they became symptomatic, plants inoculated with either strain
171 progressed to end-stage disease at the same rate (average time was 1.2 and 1.1 days for
172 WT and Δfcs inoculated plants, respectively, between the first symptoms and the highest
173 disease index rating; $P=0.402$, unpaired t -test). To determine whether the virulence defect
174 was due to a delay in symptom onset, we used survival analysis. Although survival analysis
175 was originally developed to analyze patient outcome data in clinical trials, this statistical
176 tool can analyze any discrete biological events in a time course. This analysis revealed that
177 symptom onset was earlier in WT-inoculated plants than in Δfcs -inoculated plants (Fig. 3B;
178 log-rank Mantel-Cox test; $P = 0.0118$). The median time until symptom onset was 6 days
179 after WT inoculation and 7 days after Δfcs mutant inoculation. This result suggested that
180 the virulence defect of the Δfcs mutant affects an early stage of the infection process before
181 symptom onset.

182 Since HCA degradation genes were highly expressed when *R. solanacearum* infected
183 plants at cool temperatures (Meng, Jacobs and Allen, unpublished), we also quantified the
184 virulence of the Δfcs mutant in a growth chamber at 24°C day and 19°C night (Fig 3C).
185 Under these cooler conditions, the Δfcs mutant also displayed a virulence defect.

186 **HCA degradation is not required for *R. solanacearum* growth in plant-associated**
187 **environments.**

188 We hypothesized that HCA degradation contributes to *R. solanacearum* virulence by
189 providing the bacterium with a carbon source in the competitive and nutrient-limited
190 niches in and around plants. Plant roots exude HCAs into the rhizosphere, and HCAs
191 comprise up to 10% of the water-soluble carbon in soil (Smolander et al., 2005). Thus, the
192 ability to use HCAs as a carbon source could provide bacteria with a competitive edge in
193 the soil. We asked whether HCA degradation increased growth of *R. solanacearum* in water-
194 soluble potting soil extract and in root exudate from sterile tomato seedlings. The Δfcs
195 mutant grew as well as its wild type parent in both substrates (Fig. 4A-B). However, sterile
196 tomato seedlings may produce less HCAs than mature plants with a diverse microbiome
197 since pathogens induce production and release of HCAs into the rhizosphere (Neumann
198 and Römheld, 2007). HCAs chelate and transport metals in the xylem sap, and
199 concentrations of HCA conjugates increase in grapevines infected with *Xylella* (Ishimaru et
200 al., 2011; Wallis and Chen, 2012). Therefore we asked whether HCA degradation provides a
201 growth benefit to *R. solanacearum* in xylem sap. Xylem sap was harvested by detopping
202 healthy tomato plants and allowing root pressure to exude the sap. The sap was filter
203 sterilized and used as a growth substrate for WT, Δfcs , and $\Delta fcs+fcs$ (complemented)
204 bacteria. HCA degradation ability did not affect growth of any strain in healthy xylem sap
205 (Fig. 4C). Because *R. solanacearum* infections induce expression of tomato phenolic
206 biosynthesis genes (Ishihara et al., 2012; Mitra, Milling, and Allen unpublished), we
207 hypothesized that xylem sap from infected plants contains higher HCA concentrations that

208 would benefit growth of the WT strain. However, we detected no differences in growth
209 between WT and Δfcs when they were grown in sap harvested from plants infected with
210 WT or the Δfcs mutant (data not shown). Because the Δfcs mutant grew normally on these
211 substrates, we infer that HCAs were not significant sources of carbon or present in
212 sufficient concentrations to inhibit bacterial growth.

213 **HCA degradation contributes to colonization of tomato roots.**

214 Studies of *R. solanacearum* growth in potting soil extract, root exudate from sterile
215 seedlings, and xylem sap cannot reflect the complex process of invading and multiplying in
216 hosts. Therefore, we transitioned to whole-plant assays. To measure the overall fitness of
217 the Δfcs mutant in a naturalistic infection, we used a competition assay where plants were
218 co-inoculated with a 1:1 suspension of WT and Δfcs mutant bacteria using the soil-soaking
219 method. At the first sign of wilt symptoms, we quantified the population size of each strain
220 in the midstem (Fig. 5C). With a median competitive index (CI) of 0.46, the Δfcs mutant was
221 significantly less fit than WT bacteria (Wilcoxon signed rank test; $P < 0.0001$). The WT strain
222 outcompeted the Δfcs mutant by 2.2-fold. This assay requires strains to compete during
223 several stages of the *R. solanacearum* infection cycle: survival in bulk soil and colonization
224 of host rhizosphere, roots, and stems. The cumulative effects of these competitive
225 interactions are assessed by comparing population sizes of the two strains in tomato stems.
226 The observed reduced competitive fitness of the Δfcs mutant indicates that HCA
227 degradation contributes to at least one stage of the *R. solanacearum* infection cycle.

228 To more narrowly investigate the role of HCA degradation in early stages of
229 infection, we individually soil-soak inoculated tomato plants with gentamicin-marked WT

230 and Δfcs bacteria and quantified population sizes of the strains in surface-sterilized roots
231 (Fig. 5A). At 3 dpi, the population sizes of the Δfcs mutant were lower than those of WT
232 bacteria in roots ($P < 0.0094$, t -test), but by 6 dpi, the population size of the mutant caught
233 up to wildtype levels. These results indicate that HCA degradation contributed significantly
234 to *R. solanacearum*'s ability to enter and/or grow within the root.

235 To investigate the role of HCA degradation when the bacterium is in tomato stems,
236 we used a two-pronged approach. First we quantified population sizes of the WT and Δfcs
237 strains in the midstem stem after individual soil-soak inoculations (Fig 5B). At 3 dpi, few
238 stems were colonized with detectable levels of bacteria. At 6 dpi, the average stem
239 population size of WT bacteria was slightly, but insignificantly, higher than that of the Δfcs
240 mutant. We next used an *in planta* competition assay that can reveal subtle colonization
241 defects that are missed in individual colonization assays (Yao and Allen, 2006; Macho et al.,
242 2010). For this assay, 2,000 CFU of reciprocally-marked WT and Δfcs bacteria were co-
243 inoculated into tomato plants via a cut petiole. At symptom onset, the midstem population
244 size of each strain was determined by grinding stem tissue and dilution plating. The
245 population size of the WT strain was slightly larger than that of the Δfcs mutant in this in-
246 stem competition assay, but the two strains were not significantly different. Overall,
247 colonization assays indicated that HCA degradation may improve stem colonization and
248 showed that HCA degradation contributed significantly to the bacterium's ability to
249 colonize roots.

250 ***R. solanacearum* HCA degradation does not detectably affect quantity or distribution**
251 **of lignin in susceptible tomato stems.**

252 As phenylpropanoids, HCAs are precursors to many plant physical defenses, such as
253 diferulate cross-links in primary cell walls and lignin in secondary cell walls. Plant hosts
254 often respond to pathogens by increasing biosynthesis and deposition of phenylpropanoids
255 (Dixon and Paiva, 1995). Previous studies found that bacterial wilt-resistant tomato plants
256 express phenylpropanoid biosynthesis genes early during *R. solanacearum* infections
257 (Vanitha et al., 2009). We hypothesized that HCA degradation by the pathogen may reduce
258 or prevent lignin formation by decreasing the pool of lignin precursors (HCAs). To test this
259 prediction, we used the acetyl bromide assay to measure gross lignin amounts in whole
260 stems from healthy tomato plants and from plants infected with WT or Δfcs bacteria. The
261 lignin content in stems of the wilt-susceptible Bonny Best cultivar did not increase in
262 response to infection with either *R. solanacearum* strain (Fig. 6A). Analysis of whole stems
263 could overlook variation in lignin distribution between conditions, but a histopathological
264 analysis of stem cross sections using the lignin-specific stain phloroglucinol revealed that
265 lignin was similarly distributed in all samples, mainly around the xylem vessels in the
266 vascular bundles (Fig. 6B-D). This result suggested that *R. solanacearum*'s HCA degradation
267 pathway does not detectably alter the amount or the distribution of host lignin, at least
268 under the conditions tested.

269 These experiments do not determine if HCA degradation affects the amount of
270 diferulate cross-links in the primary cell wall. Several plant pathogens with large
271 repertoires of cell-wall degrading enzymes use feruloyl esterases to cleave ferulate from
272 cell wall sugars (DiGuistini et al., 2011; Hassan and Hugouvieux-Cotte-Pattat, 2011;
273 Balcerzak et al., 2012). It is possible that *R. solanacearum* encounters diferulate bridges

296 the WT strain at sub-MIC concentrations of both *p*-coumarate and caffeate with the Δfcs
297 mutant showing a significant growth defect in as little as 23 μ M of *p*-coumarate and
298 caffeate. The Δfcs mutant suffered near-complete growth inhibition at 375 μ M *p*-coumarate
299 even though WT bacteria were unaffected by this concentration (Fig. 7A and B). In contrast,
300 there was no difference in growth between WT and Δfcs bacteria at any ferulate
301 concentration (Fig. 7C). Adding a wild-type copy of the *fcs* operon to the mutant restored
302 full wild-type levels of *p*-coumarate and caffeate tolerance to the complemented strain.
303 Together, these results suggested that HCA degradation protects *R. solanacearum* from
304 toxicity of caffeate and *p*-coumarate but not ferulate.

305 Discussion

306 Plant-associated bacteria experience a complex cocktail of secondary metabolites
307 produced by their eukaryotic hosts. Some of these compounds may provide nutrition, while
308 many are inhibitory or toxic. We found that the ability to degrade a group of such
309 compounds, the HCAs, is a quantitative virulence factor for *R. solanacearum*. More specific
310 analyses revealed that HCA degradation contributes to bacterial wilt pathogenesis, possibly
311 by protecting the pathogen from inhibition by toxic HCAs during root colonization.

312 Several lines of evidence support our model that HCA degradation contributes to *R.*
313 *solanacearum* fitness at early stages of disease. First, we found that the HCA degradation
314 and β -keto adipate pathways are found predominantly in soil-inhabiting bacteria. In
315 contrast, we did not detect conservation of HCA degradation genes in the xylem-colonizing
316 bacteria *Clavibacter michiganensis*, *Xylella fastidiosa*, *Dickeya dadantii*, or *Erwinia*
317 *amylovora*. Second, the virulence defect of the Δfcs mutant includes a delay in the first

318 appearance of symptoms. The delay in symptom onset likely results from this mutant's
319 slower colonization of host roots. Surprisingly, the delay in root colonization by the Δfcs
320 mutant did not result in smaller eventual population sizes in the stems of infected plants.
321 This suggests that once the pathogen gains entry to xylem vessels, a strain unable to
322 degrade HCAs can grow to the same final density as its wild-type parent. Thus, HCA
323 degradation appears to be most useful to *R. solanacearum* in host roots and rhizospheres.

324 Our *in vitro* inhibition assay demonstrates that the mutant is more susceptible to
325 toxicity of certain HCAs. Although the HCA concentrations required for growth inhibition *in*
326 *vitro* are 10 to 100-fold higher than concentrations measured *in planta*, concentrations of
327 these compounds in the xylem of roots and stems may be locally high where phenolics are
328 released by sentinel phenolic-storing-cells (Beckman, 2000; Alvarez et al., 2008; Mandal
329 and Mitra, 2008; Wallis and Chen, 2012). Additionally, the metabolic state of *R.*
330 *solanacearum* cells affected their susceptibility to HCAs; HCAs were more inhibitory when
331 *R. solanacearum* was grown in glucose minimal media than when grown in succinate
332 minimal media (data not shown). Moreover, it takes less HCA to inhibit microbial growth
333 when HCAs are present in mixtures than when only one HCA is present (Harris et al., 2010).
334 *R. solanacearum* cells likely encounter mixtures of HCAs and other antimicrobial
335 compounds when infecting plants. It is therefore possible that the effective inhibitory
336 concentrations of caffeate and *p*-coumarate are lower in the complex chemical
337 environment of an infected plant than in a single-chemical *in vitro* MIC assay.

338 Surprisingly, HCA degradation ability did not affect the toxicity of ferulate. *R.*
339 *solanacearum* may have an *fcs*-independent pathway that specifically degrades ferulate,

340 although this seems unlikely because the Δfcs mutant could not grow on ferulate.
341 Alternatively, *R. solanacearum* could have a drug efflux pump that is highly effective at
342 removing ferulate but less active on caffeate and *p*-coumarate. The drug efflux pumps
343 encoded by *dinF* and *acrA* are very important for strain K60's virulence on tomato (Brown
344 et al., 2007). The *acrA* mutant had heightened susceptibility to caffeate toxicity, but other
345 HCAs were not tested. Strain K60 lacks the HCA degradation pathway, so it would be
346 interesting to determine if drug efflux pumps also protect other *R. solanacearum* strains
347 from toxicity of caffeate and other HCAs.

348 Our virulence studies used a single strain, GMI1000, but the capacity to degrade
349 HCAs is genetically well conserved across the *R. solanacearum* species complex. Although
350 bioinformatic analysis accurately predicted the ability of strains to grow on various HCA
351 compounds 83% of the time, several strains did not grow on all predicted HCA carbon
352 sources. These disparities demonstrate that predictions based on genomic analysis require
353 functional validation, especially to confirm enzyme substrate specificity (Airola et al., 2014).
354 The multi-strain screen for HCA degradation ability revealed several surprising results.
355 Although CFBP2957 and Molk2 grew on ferulate, they did not grow on ferulate degradation
356 intermediates. It is possible that higher sensitivity to toxicity of vanillin and vanillate
357 prevented these strains from growing. Alternatively, vanillin and vanillate may not induce
358 expression of *vdh* and *vanAB* genes as these compounds appear to do in most *R.*
359 *solanacearum* strains. Pseudogenization of the *fcs* gene in the Race 3 biovar 2 strains
360 (UW491 and UW551) appears to prevent this strain from growing on most metabolites

361 upstream of protocatechuate. This is in contrast to strain GMI1000 where deletion of *fcs*
362 did not affect growth on any compounds besides HCAs.

363 Although *R. solanacearum* expresses its HCA degradation genes in stem xylem
364 vessels during tomato pathogenesis, these genes are only expressed at moderate levels
365 (Jacobs et al., 2012). It is not surprising that HCA degradation genes were not identified in
366 our previous IVET screen for root exudate induced genes because that study used *R.*
367 *solanacearum* strain K60, which has lost HCA degradation ability (Colburn-Clifford and
368 Allen, 2010). Transcriptional analysis could be used to compare expression of HCA
369 degradation genes at different stages in the *R. solanacearum* life cycle, particularly in the
370 rhizosphere. Phenolics in root exudate are chemoattractants for many rhizosphere
371 bacteria, including *Agrobacterium tumefaciens* (Mandal et al., 2010). Chemotaxis allows *R.*
372 *solanacearum* to locate host plants, but it remains to be determined whether root-exuded
373 phenolics serve as chemoattractants (Yao and Allen, 2006).

374 Plant phenolics influence expression of virulence genes in many plant mutualists
375 and pathogens. Ferulate and other phenolics induce expression of *Agrobacterium vir* genes,
376 and flavonoids induce *Rhizobium nod* genes, both of which are required for association with
377 plants (Bhattacharya et al., 2010). Expression of the *Dickeya dadantii* type III secretion
378 genes is induced by the phenolics *o*-coumarate and *trans*-cinnamate and repressed by *p*-
379 coumarate (Li et al 2009 and Yang et al 2008). We cannot rule out the possibility that the
380 virulence defect of the Δfcs mutant is due to HCA-mediated repression of the pathogen's
381 type III secretion system, although such repression cannot be complete because the Δfcs
382 mutant still triggers a hypersensitive response in the non-host tobacco (data not shown).

383 In response to pathogen attack, plants can reinforce cell walls with lignin, which is
384 an HCA polymer. Although we didn't identify an obvious difference in stem lignin in
385 susceptible tomato after *R. solanacearum* infections, we cannot rule out subtle but
386 biologically important differences that would be undetectable in our gross analyses.
387 Moreover, this trait may play an important role in resistant tomato cultivars. A previous
388 study observed increased lignification in response to *R. solanacearum* infections in the
389 quantitatively wilt-resistant tomato cv. LS-89, but not in susceptible tomato cv. Ponderosa
390 (Ishihara et al., 2012). Histopathological studies of resistant and susceptible tomato found
391 that *R. solanacearum* colonized fewer xylem vessels in wilt-resistant varieties (Grimault et
392 al., 1994; Rahman et al., 1999). It is therefore possible that physical barriers to pathogen
393 spread may be a component of tomato resistance to bacterial wilt.

394 Taken together, our results indicate that *R. solanacearum*'s ability to enzymatically
395 disarm HCAs contributes to the success of this widespread pathogen. Pathogens have
396 adopted multiple strategies to evade plant defense compounds. *Pseudomonas syringae* uses
397 type III effectors to manipulate plant phenylpropanoids (Truman et al., 2006). Many plant
398 pathogens protect themselves with drug efflux pumps, while others enzymatically degrade
399 the plant defense compounds pisatin, tomatine, and HCAs (Tegtmeier and VanEtten, 1982;
400 Tegos et al., 2002; Brown et al., 2007; Seipke and Loria, 2008; Michielse et al., 2012). Our
401 results support a general model that root-infecting pathogens encounter toxic
402 concentrations of HCAs, and that degradation of these defenses is important for pathogenic
403 success.

404 **Materials and Methods**

405 Cultures and stock solutions

406 The bacterial strains and plasmids used in this study are described in Table 1. *E. coli*
407 was grown in Luria-Bertani (LB) medium at 37°C. *R. solanacearum* was grown in CPG broth
408 or TZC plates at 28 °C (Kelman, 1954). When appropriate, the antibiotics gentamicin (15
409 mg/L), kanamycin (25 mg/L), and ampicillin (50 m/L) were added. Boucher's minimal
410 medium (BMM) buffered with 10 mM MES (2-(*N*-morpholino) ethanesulfonic acid) pH 5.5
411 or 7.0 was used as a minimal medium (Boucher et al., 1985). For sampling from the soil or
412 roots, *R. solanacearum* was plated on modified SMSA semi-selective medium (10 g/L
413 peptone, 5 ml/L glycerol, 1 g/L casamino acids, 2.5 mg/L crystal violet, 2.5 mg/L TZC
414 [tetrazolium chloride], 13 mg/L bacitracin, 0.3 mg/L penicillin, 2.5 mg/L chloramphenicol,
415 25 mg/L cycloheximide) (Engelbrecht, 1994). Stock solutions of phenolic compounds were
416 prepared in DMSO. Chemicals were from Sigma-Aldrich, Fisher Scientific, or Difco
417 Laboratories.

418 A plate assay was used to detect HCA degradation ability in 20 *R. solanacearum*
419 isolates. BMM MES pH 7.0 plates were supplemented with 1 mM succinate, ferulate, *p*-
420 coumarate, vanillin, vanillate, *p*-hydroxybenzoate, or 5 mM protocatechuate. Compound
421 concentration was chosen empirically as there was a trade-off between compound toxicity
422 at high concentrations and minimal bacterial growth at low concentrations. To assess
423 growth, 2 µl of a dense overnight culture of each strain was spotted onto the plates. After
424 incubation at 28 °C for three to five days, growth of the strains on each substrate was
425 assessed relative to growth on BMM plates without supplemented carbon. In several cases,
426 plate growth phenotypes were indeterminate or contradicted predictions from genomic

427 data. So, growth was further tested by culturing these strains for 48 hr in liquid BMM with
428 the relevant carbon source and quantifying cell density by dilution plating.

429 **Plant growth conditions**

430 Wilt-susceptible tomato plants (cv. Bonny Best) were grown in Sunshine Redimix
431 professional growing mix at 28 °C in a climate controlled growth chamber with a 12 hr
432 day/ 12 hr night cycle. To test virulence during cool conditions, plants were grown in a
433 climate controlled chamber with a 24 day/19 °C night cycle. Plants were watered with
434 Hoagland solution.

435 **Genomic analysis of *R. solanacearum* species complex and identification of HCA** 436 **degradation gene homologs**

437 The phylogenetic tree was designed around a matrix of genomic distances obtained
438 using the Maximum Unique Match index (MUMi) algorithm (Deloger et al., 2009). MUMi
439 values were computed from pairwise genome comparisons made with MUMMer 3.0 (Kurtz
440 et al., 2004). The distances were then clustered together into a tree using the neighbor-
441 joining method (Saitou and Nei, 1987).

442 We identified homologs of the GMI1000 and UW551 HCA degradation pathway
443 using the OMA algorithm with translated coding sequences from the genomes of the other
444 *R. solanacearum* isolates (Altenhoff et al., 2011).

445 **Strain construction**

446 The Δfcs strain was created using a *sacB* suicide vector designed to precisely excise
447 the *fcs* ORF. Briefly, we amplified ~1 kb regions directly upstream (*fcs*KOupF: 5'-
448 CTCGACGATGCGGACCTG-3'; *fcs*KOupR: 5'-GACAGCGACCTCGCATCAG-3') and downstream

449 of the *fcs* ORF (*fcsK0dwnF*: 5'-ctcatgcgaggtcgctgtcGAGTGTTGAGCGGGGCC-3'; *fcsK0dwnR*:
450 5'-GGAAGGCGAATTCGAGCG-3') by PCR, fused the fragments by splice by overlap extension
451 PCR (SOE-PCR) (Heckman and Pease, 2007), and blunt-end ligated them into the pCR-blunt
452 subcloning vector (Life Technologies). This knockout construct was transferred by
453 restriction digestion and ligation to the *sacB* vector pUFR80 to create pUFR80-*fcsK0*
454 (Castañeda et al., 2005). *R. solanacearum* strain GMI1000 (WT) was transformed with
455 pUFR80-*fcsK0* by electroporation and plated on kanamycin media to select for
456 merodiploids that were sucrose sensitive and Kan^R. A clone was then counter-selected on
457 CPG+5% sucrose to select for excision of the *sacB*-containing vector backbone. This process
458 either restored the wildtype genotype or yielded a markerless deletion of the *fcs* ORF.
459 Colony PCR using *fcsK0upF* and *fcsK0dwnR* primers was used to screen for loss of the *fcs*
460 ORF.

461 A miniTn7 vector was used to complement the Δfcs strain (Choi et al., 2005). The
462 miniTn7 transposon integrates into the selectively neutral *att* site downstream of *glmS*. We
463 amplified the 5.6 kb putative operon encompassing 500 bp upstream of *fca*, the *fca* ORF, the
464 *vdh* ORF and the *fcs* ORF using primers *fcsoperonF* (5'-TGCACCAGGACCAATACCTC-3') and
465 *fcsoperonR* (5'-CTCAACGTGTTCCCATCCA-3'). The resulting PCR product was subcloned
466 into pCR-blunt and transferred by restriction enzyme digestion and ligation into pUC18t-
467 miniTn7-Gm to create pTn7*fcsComp*. The Δfcs *R. solanacearum* strain was transformed
468 with pMiniTn7*fcsComp* and the helper vector pTNS1 encoding the transposase TnsABC+D
469 and plated on gentamicin media. Complementation was confirmed by restoration of the
470 ability to grow on HCAs.

471 To create antibiotic marked strains for colonization and competition assays,
472 GMI1000 and the Δfcs strain were transformed with the chromosomal insertion vectors
473 pRCG-GWY or pRCK-GWY carrying gentamicin or kanamycin cassettes, respectively
474 (Monteiro et al., 2012).

475 **Virulence assay**

476 To assess virulence following soil soak inoculation, 17- to 21-day old plants with
477 unwounded roots were inoculated by pouring bacterial suspensions into the soil to a final
478 concentration of 1×10^8 CFU/g soil (Tans-Kersten et al., 1998). Symptoms on each plant
479 were rated daily using a disease index scale of 0-4 corresponding to wilt severity: 0,
480 asymptomatic plants; 1, less than 25%; 2, less than 50%; 3, less than 75%; and 4, up to
481 100% leaves wilted.

482 **Colonization and competition assays**

483 To assess colonization ability of individual strains, plants were soil soak inoculated
484 with either WT-Gm or Δfcs -Gm. At 3 and 6 days after inoculation, bacterial populations
485 were determined in surface-sterilized root and midstem stem tissue. To surface sterilize
486 the roots, soil was removed by gentle shaking and washing. Then roots were swirled in a
487 10% bleach solution for 15 sec and rinsed three times in successive water baths to remove
488 remaining bleach. Roots were sectioned into evenly distributed slices totaling 0.3 g. From
489 the same plants, a 0.1 g midstem stem slice was sampled. Tissue was ground in water with
490 0.28 mm metal beads using a homogenizer (MoBio). Stem grinding required 2 cycles, and
491 root grinding required 3 cycles of 2200 rpm for 1.5 min with a 4 min rest between cycles.

492 Homogenized root and stem tissue were dilution plated onto SMSA and CPG with
493 gentamicin, respectively.

494 A competition assay was used to investigate subtler differences in stem colonization
495 ability (Yao and Allen, 2006). For soil soak inoculations, plants were inoculated with a 1:1
496 mixture of antibiotic-marked WT: Δfcs bacteria totaling 1×10^8 CFU/g soil. Marker
497 swapping was used to ensure that competitive fitness differences were not caused by the
498 antibiotic resistance marker: thus a set of plants were inoculated with a WT-Gm+ Δfcs -Km
499 mixture and another set were inoculated with a WT-Km+ Δfcs -Gm mixture. Antibiotic
500 markers did not significantly impact the fitness of either strain (Fig. S1). At the first sign of
501 wilt symptoms (disease index = 1), plants were harvested and population sizes of each
502 strain in the stem were determined by grinding and dilution plating on selective media.
503 Population size was normalized to initial inoculum of each strain. Then, competitive index
504 (CI) was calculated by dividing the normalized Δfcs population size by the normalized WT
505 population size from the same plant. For cut-petiole inoculations, bacteria were directly
506 introduced into the stem by placing a 2 μ l drop of bacterial suspension onto a freshly cut
507 petiole. Each plant was inoculated with 4,000 cells in a 1:1 mixture of marked WT and Δfcs
508 bacteria.

509 **Growth in xylem sap, root exudate, and potting soil extract**

510 Xylem sap was collected from healthy and soil soak inoculated plants displaying the
511 first signs of symptoms, as previously described (Jacobs et al., 2012). Plants were detopped
512 with a sharp blade, and sap was allowed to pool on the stump by root pressure. The first
513 drop was discarded to avoid contamination by cell debris, and the stump was rinsed with

514 water and blotted dry. Sap was only collected for 30 minutes to avoid damage-response-
515 induced changes in sap composition. Samples were flash-frozen and kept at -80°C until use.
516 Growth of strains on 0.2 µM filter-sterilized xylem sap, root exudate was measured in a
517 plate reader (Bio-Tek). Overnight cultures were washed and adjusted to OD_{600nm} of 1.0. In a
518 half-area 96-well plate (Corning), 45 µl of each growth substrate was combined with 5 µl of
519 bacterial suspensions. Optical density was measured hourly until growth plateaued. Each
520 experiment was repeated twice.

521 Root exudate was collected as described (Yao and Allen, 2006). Briefly, seeds were
522 sterilized and germinated on 1% water agar plates in the dark for 3 days. Sterile roots were
523 transferred into a 50 ml conical tube containing 5 ml of BMM with 10 mM MES pH 7.0.
524 Tubes were incubated in the dark for 24 hr, and root exudate was used immediately as
525 previous studies reported loss of potency with time.

526 To collect water-soluble potting soil extract, 1 g of potting soil was suspended in 50
527 ml of distilled water in a 50 ml conical tube and incubated horizontally with shaking for 2
528 hr at room temperature (Smolander et al., 2005). Filtered potting soil extract was used
529 immediately. Dense overnight cultures were washed, and 0.5 µl of the cell suspension was
530 inoculated into 5 ml of potting soil extract. Cell density was determined periodically by
531 dilution plating.

532 **Lignin quantification and visualization**

533 Twenty-one day-old tomato plants were left healthy or inoculated by pouring WT or
534 *Δfcs* bacteria into the soil to a final concentration of 1 x 10⁸ CFU/g soil. Total stem was
535 harvested 3, 6, and 9 days after inoculation and desiccated, yielding approximately 30 mg

536 dry weight/plant. Total lignin was quantified by the spectroscopic acetyl bromide assay
537 (Fukushima and Hatfield, 2004). Wood pulp inulin was used as a lignin standard. Lignin in
538 cross-sections of tomato stems was stained by phloroglucinol:HCl (Nakano and Meshitsuka,
539 1992).

540 **Growth inhibition assay**

541 A growth inhibition assay modified from the standard minimum inhibitory
542 concentration (MIC) assay was used to test the toxicity of HCAs (Brown et al., 2007). BMM
543 MES pH 5.5 +10 mM succinate supplemented with 23-3000 μ M of ferulate, caffeate, or *p*-
544 coumarate or with no inhibiting compound was inoculated with bacterial strains to 1×10^5
545 CFU/ml final concentration. After incubating strains at 28 °C with shaking, cell density was
546 measured by optical density in a Bio-Tek plate reader. Growth of strains in each condition
547 was calculated relative to growth of the WT strain without inhibitory HCAs.

548 **Acknowledgements**

549 The authors gratefully acknowledge Paul Weimer for help with lignin analysis,
550 Mehdi Kabbage for help with microscopy, and Frederique van Gijsegem, Devanshi
551 Khokhani, and Philippe Prior for valuable discussions. We thank Brianna Fochs for
552 technical help. This research was supported by the USDA-ARS Floral and Nursery Crops
553 Research Initiative and the UW-Madison College of Agricultural and Life Sciences. TML was
554 supported by NIH National Research Service Award T32 GM07215 and by a Agriculture
555 and Food Research Initiative Competitive Grant from the USDA National Institute of Food
556 and Agriculture.

557 **Author Contributions**

558 T. Lowe designed most experiments and analyzed data. F. Ailloud constructed the
 559 whole-genome phylogenetic tree and analyzed the genetic conservation of the HCA
 560 degradation pathway. C. Allen advised T. Lowe in experimental design and analysis. All
 561 authors participated in drafting the manuscript and approved of the final version for
 562 submission.

563 Literature Cited

- 564 Abdelkafi, S., Sayadi, S., Gam, A., Ben, Z., Casalot, L., and Labat, M. 2006. Bioconversion of
 565 ferulic acid to vanillic acid by *Halomonas elongata* isolated from table-olive
 566 fermentation. *FEMS Microbiol. Lett.* 262:115-120.
- 567 Airola, M.V., Tumolo, J.M., Snider, J., and Hannun, Y.A. 2014. Identification and biochemical
 568 characterization of an acid sphingomyelinase-like protein from the bacterial plant
 569 pathogen *Ralstonia solanacearum* that hydrolyzes ATP to AMP but not
 570 sphingomyelin to ceramide. *PLoS ONE* 9:e105830.
- 571 Altenhoff, A.M., Schneider, A., Gonnet, G.H., and Dessimoz, C. 2011. OMA 2011: orthology
 572 inference among 1000 complete genomes. *Nucleic Acids Res.* 39:D289-D294.
- 573 Alvarez, S., Marsh, E.L., Schroeder, S.G., and Schachtman, D.P. 2008. Metabolomic and
 574 proteomic changes in the xylem sap of maize under drought. *Plant Cell Environ.*
 575 31:325-340.
- 576 Balcerzak, M., Harris, L.J., Subramaniam, R., and Ouellet, T. 2012. The feruloyl esterase gene
 577 family of *Fusarium graminearum* is differentially regulated by aromatic compounds
 578 and hosts. *Fungal Biol.* 116:478-488.
- 579 Beckman, C.H. 2000. Phenolic-storing cells: keys to programmed cell death and periderm
 580 formation in wilt disease resistance and in general defence responses in plants?
 581 *Physiol. Mol. Plant Pathol.* 57:101-110.
- 582 Bhattacharya, A., Sood, P., and Citovsky, V. 2010. The roles of plant phenolics in defence and
 583 communication during *Agrobacterium* and *Rhizobium* infection. *Mol. Plant Pathol.*
 584 11:705-719.
- 585 Boucher, C., Barberis, P., Trigalet, A., and Demery, D. 1985. Transposon mutagenesis of
 586 *Pseudomonas solanacearum*: isolation of Tn5-induced avirulent mutants. *J. Gen.*
 587 *Microbiol.* 131:2449-2457.
- 588 Brown, D.G., Swanson, J.K., and Allen, C. 2007. Two host-induced *Ralstonia solanacearum*
 589 genes, *acrA* and *dinF*, encode multidrug efflux pumps and contribute to bacterial wilt
 590 virulence. *Appl. Environ. Microbiol.* 73:2777-2786.
- 591 Campillo, T., Renoud, S., Kerzaon, I., Vial, L., Baude, J., Gaillard, V., Bellvert, F., Chamignon, C.,
 592 Comte, G., and Nesme, X. 2014. Analysis of hydroxycinnamic acid degradation in
 593 *Agrobacterium fabrum* reveals a coenzyme A-dependent, beta-oxidative
 594 deacetylation pathway. *Appl. Env. Microbiol.* 80:3341-3349.

- 595 Campos, L., Lisón, P., López-Gresa, M.P., Rodrigo, I., Zacarés, L., Conejero, V., and Bellés, J.M.
 596 2014. Transgenic tomato plants overexpressing tyramine N-
 597 hydroxycinnamoyltransferase exhibit elevated hydroxycinnamic acid amide levels
 598 and enhanced resistance to *Pseudomonas syringae*. *Mol. Plant-Microbe Interact.*
 599 27:1159-1169.
- 600 Cao, Y., Tian, B., Liu, Y., Cai, L., Wang, H., Lu, N., Wang, M., Shang, S., Luo, Z., and Shi, J. 2013.
 601 Genome sequencing of *Ralstonia solanacearum* FQY_4, isolated from a bacterial wilt
 602 nursery used for breeding crop resistance. *Genome Announc.* 1:e00125-00113.
- 603 Castañeda, A., Reddy, J.D., El-Yacoubi, B., and Gabriel, D.W. 2005. Mutagenesis of all eight
 604 *avr* genes in *Xanthomonas campestris* pv. *campestris* had no detected effect on
 605 pathogenicity, but one *avr* gene affected race specificity. *Mol. Plant-Microbe Interact.*
 606 18:1306-1317.
- 607 Choi, K.-H., Gaynor, J.B., White, K.G., Lopez, C., Bosio, C.M., Karkhoff-Schweizer, R.R., and
 608 Schweizer, H.P. 2005. A Tn7-based broad-range bacterial cloning and expression
 609 system. *Nat. Methods* 2:443-448.
- 610 Colburn-Clifford, J., and Allen, C. 2010. A *cbb3*-type cytochrome C oxidase contributes to
 611 *Ralstonia solanacearum* R3bv2 growth in microaerobic environments and to
 612 bacterial wilt disease development in tomato. *Mol. Plant-Microbe Interact.* 23:1042-
 613 1052.
- 614 Deloger, M., El Karoui, M., and Petit, M.A. 2009. A genomic distance based on MUM indicates
 615 discontinuity between most bacterial species and genera. *J. Bacteriol.* 191:91-99.
- 616 DiGuistini, S., Wang, Y., Liao, N.Y., Taylor, G., Tanguay, P., Feau, N., Henrissat, B., Chan, S.K.,
 617 Hesse-Orce, U., and Alamouti, S.M. 2011. Genome and transcriptome analyses of the
 618 mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine
 619 pathogen. *Proc. Natl. Acad. Sci. U.S.A.* 108:2504-2509.
- 620 Dixon, R.A., and Paiva, N.L. 1995. Stress-induced phenylpropanoid metabolism. *Plant Cell*
 621 7:1085-1097.
- 622 Elphinstone, J.G. 2005. The current bacterial wilt situation: A global overview. Pages 9-28
 623 in: *Bacterial Wilt: The Disease and the Ralstonia solanacearum* Species Complex, C.
 624 Allen, P. Prior, and A.C. Hayward, eds. American Phytopathological Society Press, St
 625 Paul.
- 626 Engelbrecht, M. 1994. Modification of a semi-selective medium for the isolation and
 627 quantification of *Pseudomonas solanacearum*. *Bacterial Wilt Newsletter*:3-5.
- 628 Fitzgerald, D., Stratford, M., Gasson, M., Ueckert, J., Bos, A., and Narbad, A. 2004. Mode of
 629 antimicrobial action of vanillin against *Escherichia coli*, *Lactobacillus plantarum* and
 630 *Listeria innocua*. *J. Appl. Microbiol.* 97:104-113.
- 631 Fry, S.C., Willis, S.C., and Paterson, A.E. 2000. Intraprotoplasmic and wall-localised
 632 formation of arabinoxylan-bound diferulates and larger ferulate coupling-products
 633 in maize cell-suspension cultures. *Planta* 211:679-692.
- 634 Fukushima, R.S., and Hatfield, R.D. 2004. Comparison of the acetyl bromide
 635 spectrophotometric method with other analytical lignin methods for determining
 636 lignin concentration in forage samples. *J. Agric. Food Chem.* 52:3713-3720.
- 637 Gabriel, D.W., Allen, C., Schell, M., Denny, T.P., Greenberg, J.T., Duan, Y.P., Flores-Cruz, Z.,
 638 Huang, Q., Clifford, J.M., Presting, G., González, E.T., Reddy, J., Elphinstone, J.,

- 639 Swanson, J., Yao, J., Mulholland, V., Liu, L., Farmerie, W., Patnaikuni, M., Balogh, B.,
640 Norman, D., Alvarez, A., Castillo, J.A., Jones, J., Saddler, G., Walunas, T., Zhukov, A., and
641 Mikhailova, N. 2006. Identification of open reading frames unique to a select agent:
642 *Ralstonia solanacearum* Race 3 biovar 2. Mol. Plant-Microbe Interact. 19:69-79.
- 643 Grimault, V., Gelie, B., Lemattre, M., Prior, P., and Schmit, J. 1994. Comparative histology of
644 resistant and susceptible tomato cultivars infected by *Pseudomonas solanacearum*.
645 Physiol. Mol. Plant Pathol. 44:105-123.
- 646 Harris, V., Jiranek, V., Ford, C.M., and Grbin, P.R. 2010. Inhibitory effect of hydroxycinnamic
647 acids on *Dekkera* spp. Appl. Microbiol. Biotechnol. 86:721-729.
- 648 Hassan, S., and Hugouvieux-Cotte-Pattat, N. 2011. Identification of two feruloyl esterases in
649 *Dickeya dadantii* 3937 and induction of the major feruloyl esterase and of pectate
650 lyases by ferulic acid. J. Bacteriol. 193:963-970.
- 651 Heckman, K.L., and Pease, L.R. 2007. Gene splicing and mutagenesis by PCR-driven overlap
652 extension. Nat. Protoc. 2:924-932.
- 653 Ishihara, T., Mitsuhara, I., Takahashi, H., and Nakaho, K. 2012. Transcriptome analysis of
654 quantitative resistance-specific response upon *Ralstonia solanacearum* infection in
655 tomato. PLoS ONE 7:e46763.
- 656 Ishimaru, Y., Kakei, Y., Shimo, H., Bashir, K., Sato, Y., Sato, Y., Uozumi, N., Nakanishi, H., and
657 Nishizawa, N.K. 2011. A rice phenolic efflux transporter is essential for solubilizing
658 precipitated apoplasmic iron in the plant stele. J. Biol. Chem. 286:24649-24655.
- 659 Jacobs, J.M., Babujee, L., Meng, F., Milling, A., and Allen, C. 2012. The *in planta* transcriptome
660 of *Ralstonia solanacearum*: conserved physiological and virulence strategies during
661 bacterial wilt of tomato. mBio 3:e00114-00112.
- 662 Kelman, A. 1954. The relationship of pathogenicity of *Pseudomonas solanacearum* to colony
663 appearance in tetrazolium medium. Phytopathol. 44:693-695.
- 664 Kim, D., Kim, S.W., Choi, K.Y., Lee, J.S., and Kim, E. 2008. Molecular cloning and functional
665 characterization of the genes encoding benzoate and *p*-hydroxybenzoate
666 degradation by the halophilic *Chromohalobacter* sp. strain HS-2. FEMS Microbiol.
667 Lett. 280:235-241.
- 668 Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg,
669 S.L. 2004. Versatile and open software for comparing large genomes. Genome Biol.
670 5:R12.
- 671 Lanoue, A., Burlat, V., Henkes, G.J., Koch, I., Schurr, U., and Röse, U.S. 2010. *De novo*
672 biosynthesis of defense root exudates in response to *Fusarium* attack in barley. New
673 Phytol. 185:577-588.
- 674 Li, L., and Steffens, J.C. 2002. Overexpression of polyphenol oxidase in transgenic tomato
675 plants results in enhanced bacterial disease resistance. Planta 215:239-247.
- 676 Li, Z., Wu, S., Bai, X., Liu, Y., Lu, J., Liu, Y., Xiao, B., Lu, X., and Fan, L. 2011. Genome sequence
677 of the tobacco bacterial wilt pathogen *Ralstonia solanacearum*. J. Bacteriol.
678 193:6088-6089.
- 679 Macho, A.P., Guidot, A., Barberis, P., Beuzón, C.R., and Genin, S. 2010. A competitive index
680 assay identifies several *Ralstonia solanacearum* Type III effector mutant strains with
681 reduced fitness in host plants. Mol. Plant-Microbe Interact. 23:1197-1205.

- 682 Mandal, S., and Mitra, A. 2008. Accumulation of cell wall-bound phenolic metabolites and
683 their upliftment in hairy root cultures of tomato (*Lycopersicon esculentum* Mill.).
684 *Biotechnol. Lett.* 30:1253-1258.
- 685 Mandal, S.M., Chakraborty, D., and Dey, S. 2010. Phenolic acids act as signaling molecules in
686 plant-microbe symbioses. *Plant Signal. Behav.* 5:359-368.
- 687 Michielse, C.B., Reijnen, L., Olivain, C., Alabouvette, C., and Rep, M. 2012. Degradation of
688 aromatic compounds through the β -keto adipate pathway is required for
689 pathogenicity of the tomato wilt pathogen *Fusarium oxysporum* f. sp. *lycopersici*. *Mol.*
690 *Plant Pathol.* 13:1089-1100.
- 691 Monteiro, F., Solé, M., Dijk, I.v., and Valls, M. 2012. A chromosomal insertion toolbox for
692 promoter probing, mutant complementation, and pathogenicity studies in *Ralstonia*
693 *solanacearum*. *Mol. Plant-Microbe Interact.* 25:557-568.
- 694 Mueller, W., and Beckman, C. 1984. Ultrastructure of the cell wall of vessel contact cells in
695 the xylem of tomato stems. *Ann. Botany* 53:107-114.
- 696 Nakaho, K., Hibino, H., and Miyagawa, H. 2000. Possible mechanisms limiting movement of
697 *Ralstonia solanacearum* in resistant tomato tissues. *J. Phytopathol.* 148:181-190.
- 698 Nakano, J., and Meshitsuka, G. 1992. The detection of lignin. Pages 23-32 in: *Methods Lignin*
699 *Chemistry*, S.Y. Lin and C.W. Dence, eds. Springer-Verlag, Berlin.
- 700 Naoumkina, M.A., Zhao, Q., GALLEGO-GIRALDO, L., Dai, X., Zhao, P.X., and Dixon, R.A. 2010.
701 Genome-wide analysis of phenylpropanoid defence pathways. *Mol. Plant Pathol.*
702 11:829-846.
- 703 Narbad, A., and Gasson, M.J. 1998. Metabolism of ferulic acid via vanillin using a novel CoA-
704 dependent pathway in a newly-isolated strain of *Pseudomonas fluorescens*.
705 *Microbiology* 144:1397-1405.
- 706 Neumann, G., and Römheld, V. 2007. The release of root exudates as affected by the plant
707 physiological status. Pages 23-72 in: *The Rhizosphere Biochemistry and Organic*
708 *Substances at the Soil-Plant Interface*, R. Pinton, Z. Veranini, and P. Nannipieri, eds.
709 CRC Press, New York.
- 710 Parke, D., and Ornston, L.N. 2003. Hydroxycinnamate (*hca*) catabolic genes from
711 *Acinetobacter* sp. strain ADP1 are repressed by HcaR and are induced by
712 hydroxycinnamoyl-coenzyme A thioesters. *Appl. Environ. Microbiol.* 69:5398-5409.
- 713 Peeters, N., Guidot, A., Vailleau, F., and Valls, M. 2013. *Ralstonia solanacearum*, a widespread
714 bacterial plant pathogen in the post-genomic era. *Mol. Plant Pathol.* 14:651-662.
- 715 Pérez-Pantoja, D., De la Iglesia, R., Pieper, D.H., and González, B. 2008. Metabolic
716 reconstruction of aromatic compounds degradation from the genome of the amazing
717 pollutant-degrading bacterium *Cupriavidus necator* JMP134. *FEMS Microbiol. Rev.*
718 32:736-794.
- 719 Plaggenborg, R., Overhage, J., Steinbüchel, A., and Priefert, H. 2003. Functional analyses of
720 genes involved in the metabolism of ferulic acid in *Pseudomonas putida* KT2440.
721 *Appl. Microbiol. Biotechnol.* 61:528-535.
- 722 Rahman, M.A., Abdullah, H., and Vanhaeke, M. 1999. Histopathology of susceptible and
723 resistant *Capsicum annuum* cultivars infected with *Ralstonia solanacearum*. *J.*
724 *Phytopathol.* 147:129-140.

- 725 Remenant, B., Babujee, L., Lajus, A., Médigue, C., Prior, P., and Allen, C. 2012. Sequencing of
726 K60, type strain of the major plant pathogen *Ralstonia solanacearum*. J. Bacteriol.
727 194:2742-2743.
- 728 Remenant, B., Cambiaire, J.-C.D., Cellier, G., Barbe, V., Medigue, C., Jacobs, J.M., Fegan, M.,
729 Allen, C., and Prior, P. 2011. Phylotype IV strains of *Ralstonia solanacearum*, *R.*
730 *syzygii* and the Blood Disease Bacterium form a single genomic species despite their
731 divergent life-styles. PLoS ONE 6:e24356.
- 732 Remenant, B., Coupat-Goutaland, B., Guidot, A., Cellier, G., Wicker, E., Allen, C., Fegan, M.,
733 Pruvost, O., Elbaz, M., Calteau, A., Salvignol, G., Mornico, D., Mangenot, S., Barbe, V.,
734 Médigue, C., and Prior, P. 2010. Genomes of three tomato pathogens within the
735 *Ralstonia solanacearum* species complex reveal significant evolutionary divergence.
736 BMC Genomics 11:379.
- 737 Romero-Silva, M.J., Méndez, V., Agullo, L., and Seeger, M. 2013. Genomic and functional
738 analyses of the gentisate and protocatechuate ring-cleavage pathways and related 3-
739 hydroxybenzoate and 4-hydroxybenzoate peripheral pathways in *Burkholderia*
740 *xenovorans* LB400. PloS ONE 8:e56038.
- 741 Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for
742 reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.
- 743 Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A.,
744 Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C.,
745 Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W.,
746 Schiex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J.,
747 and Boucher, C.A. 2002. Genome sequence of the plant pathogen *Ralstonia*
748 *solanacearum*. Nature 415:497-502.
- 749 Seipke, R.F., and Loria, R. 2008. *Streptomyces scabies* 87-22 possesses a functional
750 tomatinase. J. Bacteriol. 190:7684-7692.
- 751 Smolander, A., Loponen, J., Suominen, K., and Kitunen, V. 2005. Organic matter
752 characteristics and C and N transformations in the humus layer under two tree
753 species, *Betula pendula* and *Picea abies*. Soil Biol. Biochem. 37:1309-1318.
- 754 Tans-Kersten, J., Guan, Y., and Allen, C. 1998. *Ralstonia solanacearum* pectin methylesterase
755 is required for growth on methylated pectin but not for bacterial wilt virulence.
756 Appl. Environ. Microbiol. 64:4918-4923.
- 757 Tegos, G., Stermitz, F.R., Lomovskaya, O., and Lewis, K. 2002. Multidrug pump inhibitors
758 uncover remarkable activity of plant antimicrobials. Antimicrobial Agents and
759 Chemotherapy 46:3133-3141.
- 760 Tegtmeier, K., and VanEtten, H. 1982. The role of pisatin tolerance and degradation in the
761 virulence of *Nectria haematococca* on peas: a genetic analysis. Phytopathology
762 72:608-612.
- 763 Truman, W., Zabala, M.T., and Grant, M. 2006. Type III effectors orchestrate a complex
764 interplay between transcriptional networks to modify basal defence responses
765 during pathogenesis and resistance. Plant J. 46:14-33.
- 766 Vanitha, S.C., Niranjana, S.R., and Umesha, S. 2009. Role of phenylalanine ammonia lyase
767 and polyphenol oxidase in host resistance to bacterial wilt of tomato. J. Phytopathol.
768 157:552-557.

- 769 Wallis, C.M., and Chen, J. 2012. Grapevine phenolic compounds in xylem sap and tissues are
770 significantly altered during infection by *Xylella fastidiosa*. *Phytopathology* 102:816-
771 826.
- 772 Wicker, E., Grassart, L., Coranson-Beaudu, R., Mian, D., Guilbaud, C., Fegan, M., and Prior, P.
773 2007. *Ralstonia solanacearum* strains from Martinique (French West Indies)
774 exhibiting a new pathogenic potential. *Appl. Environ. Microbiol.* 73:6790-6801.
- 775 Xu, J., Zheng, H.-j., Liu, L., Pan, Z.-c., Prior, P., Tang, B., Xu, J.-s., Zhang, H., Tian, Q., and Zhang,
776 L.-q. 2011. The complete genome sequence of plant pathogen *Ralstonia*
777 *solanacearum* strain Po82. *J. Bacteriol.* 193:4261-4262.
- 778 Yao, J., and Allen, C. 2006. Chemotaxis is required for virulence and competitive fitness of
779 the bacterial wilt pathogen *Ralstonia solanacearum*. *J. Bacteriol.* 188:3697-3708.
- 780

781 **Table 1: Strains and plasmids used in this study.**

Strain or Plasmid ^a	Relevant characteristics ^b	Source or Reference
Strains		
	F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>)	
	ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i>	
	Δ (<i>ara-leu</i>)7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (Str ^r)	
<i>E. coli</i> TOP10	<i>endA1</i> <i>nupG</i>	Life Technologies
<i>R. solanacearum</i>		
GMI1000 (WT)	Wildtype, phylotype I	(Boucher et al., 1985)
GMI1000-Gm	GMI1000 transformed with pRCG-GWY, Gm ^r	This study
GMI1000-Km	GMI1000 transformed with pRCK-GWY, Kan ^r	This study
Δ <i>fcs</i>	GMI1000 with unmarked, precise deletion of the feruloyl-CoA synthetase (<i>fcs</i>) ORF	This study
Δ <i>fcs</i> -Gm	Δ <i>fcs</i> transformed with pRCG-GWY, Gm ^r	This study
Δ <i>fcs</i> -Km	Δ <i>fcs</i> transformed with pRCK-GWY, Kan ^r	This study
Δ <i>fcs</i> + <i>fcs</i>	Complemented Δ <i>fcs</i> with Tn7FcsComp integrated into chromosome at the selectively neutral <i>att</i> site, Gm ^r	This study

Plasmids

	Vector that integrates downstream of	
pRCG-GWY	<i>glmS</i> on the GMI1000 chromosome, Gm ^r	(Monteiro et al., 2012)
	Vector that integrates downstream of	
pRCK-GWY	<i>glmS</i> on the GMI1000 chromosome, Kan ^r	(Monteiro et al., 2012)
pCR-blunt	Cloning vector	Life Technologies
pUFR80	pUFR80, Suc ^s (<i>sacB</i>), Kan ^r	(Castañeda et al., 2005)
	Vector that integrates into selectively	
pUC18T- miniTn7T-Gm	neutral <i>att</i> site on <i>R. solanacearum</i> chromosome Gm ^r , Amp ^r	(Choi et al., 2005)
	Helper plasmid for pUC18T-miniTn7T-Gm encoding the site-specific TnsABCD Tn7 transposase, Amp ^r	(Choi et al., 2005)
pTNS1		
	pUFR80 + <i>fcs</i> markerless deletion construct inserted into <i>sacI/xbal</i> sites in	
pUFR80-fcsKO	MCS, Suc ^s (<i>sacB</i>), Kan ^r	This study
	<i>fcs</i> operon (<i>fca-vdh-fcs</i>) with native promoter cloned into <i>hindIII/speI</i> sites in	
pMiniTn7FcsComp	pUC18T-miniTn7T-Gm, Gm ^r , Amp ^r	This study

782 ^a*R. solanacearum* isolates characterized in Fig 2 are listed in Table S1.

- 783 ^bAmp^r, ampicillin resistance; Gm^r, gentamicin resistance; Kan^r, kanamycin resistance; Str^r,
784 streptomycin resistance; Suc^s, sucrose sensitivity

785 **Fig. 1.** Hydroxycinnamic acid (HCA) degradation pathway and genes in *R. solanacearum*
 786 GMI1000. **A**, The HCA degradation pathway is shown with enzyme names in boldface. Fcs,
 787 Fca, and Vdh convert the HCAs ferulate, *p*-coumarate, and caffeate to the phenolic acids
 788 vanillate, *p*-hydroxybenzoate, and protocatechuate, respectively. VanAB and PobA convert
 789 vanillate and *p*-hydroxybenzoate to protocatechuate, which is further metabolized by the
 790 β -ketoacid enzymes; **B**, A locus containing genes encoding multiple enzymes in the HCA
 791 degradation pathway. White arrows indicate ORFs encoding HCA degradation, and grey
 792 arrows indicate neighboring ORFs. RSp0221, RSp0224, and RSp0228 encode
 793 transcriptional regulators of the LysR family, MarR family, and Fis family, respectively. The
 794 dashed line above the genes indicates the region that was precisely excised to create the
 795 feruloyl-CoA synthetase deletion mutant (Δfcs). The solid line below the genes indicates the
 796 region used to genetically complement the Δfcs mutation in the $\Delta fcs+fcs$ strain.

797
 798 **Fig. 2.** HCA degradation is widely conserved in the *R. solanacearum* species complex. **A**,
 799 Genetic and functional conservation of HCA degradation. A whole genome comparison
 800 phylogenetic tree is shown on the left. Presence of HCA degradation genes and growth of *R.*
 801 *solanacearum* strains on ferulate (Fer), *p*-coumarate (Cou), vanillin (Van), vanillate (VA), *p*-
 802 hydroxybenzoate (HBA), and protocatechuate (PCA) are indicated. ^aGrowth phenotype
 803 differs from genotype prediction; **B**, *fcs* encodes a functional feruloyl-CoA synthetase in
 804 strain GMI1000. Strains were grown in minimal media supplemented with 0.2 mM
 805 succinate, *p*-coumarate, caffeate, ferulate, or no carbon (-) for 72 hr. Bars represent the
 806 mean of 3 biological replicates and error bars indicate standard error of the mean.

807

808 **Fig. 3.** HCA degradation is required for full virulence of *R. solanacearum*. **A**, Disease
809 progress of WT and Δfcs mutant strains on susceptible tomato plants grown at 28 °C.
810 Twenty-one-day-old unwounded plants (cv. Bonny Best) grown at constant 28 °C were
811 inoculated by pouring a bacterial suspension into the soil of each pot. Symptoms were
812 rated using a 0 to 4 disease index scale. Each point represents the mean disease index of a
813 total of 82 plants per strain, in 6 biological replicates. Bars indicate standard error of the
814 mean. Disease progress of the Δfcs mutant was significantly slower than that of wild-type
815 ($P=0.0123$, two-way repeated measures ANOVA); **B**, Survival analysis of the above dataset
816 showing the rate of symptom onset after inoculations with WT and the Δfcs mutant; **C**,
817 Disease progress of strains on tomato plants grown in a 24 °C day /19 °C night cycle (1
818 biological replicate with N=16 plants per strain).

819

820 **Fig. 4.** Hydroxycinnamic acid degradation does not enhance *R. solanacearum* growth in
821 plant associated environments. **A-C**, *ex vivo* bacterial growth in: **A**, water extract of potting
822 soil; **B**, tomato root exudate; and **C**, tomato xylem sap harvested from stems of un-
823 inoculated, healthy plants. Graphs show the mean of 3 replicates.

824

825 **Fig. 5.** HCA degradation contributes to root entry and competitive fitness following soil
826 soak inoculation of tomato. **A-B**, Plants grown at 28 °C were soil-soak inoculated with
827 suspensions of WT or Δfcs bacteria. At 3- and 6-days post inoculation (dpi); **A**, 300 mg root
828 tissue; or **B**, 100 mg of midstem tissue were harvested, ground, and dilution plated to

829 determine cell density of *R. solanacearum* (N=30 plants for root colonization at 3 and 6 dpi;
830 N=20 for stem colonization at 3 dpi and N=30 at 6 dpi); Solid lines represent the median
831 population sizes and the dashed lines represent the limit of detection. WT-gm colonized
832 roots better than Δfcs -gm at 3 days after inoculation ($P<0.0094$; t -test); **C**, Competitive
833 fitness of WT vs. Δfcs bacteria following soil-soak inoculation. Tomato plants were co-
834 inoculated with mixtures of reciprocally-marked WT and Δfcs strains. At the first stage of
835 disease (less than 25% leaves wilted), midstem tissue was harvested, ground, and dilution
836 plated. Population size of each strain was normalized by initial inoculum. Median
837 competitive index (CI) of the Δfcs mutant was 0.46 ($P<0.0001$, Wilcoxon Signed Rank Test;
838 N=13 plants per co-inoculation, 26 total); **D**, Competition of Δfcs and WT bacteria in tomato
839 stem following direct stem inoculation. Tomato plants were co-inoculated via a cut leaf
840 petiole with 4,000 CFU in a 1:1 suspension of reciprocally-marked WT and Δfcs strains.
841 Midstem tissue was harvested at the first sign of symptoms, ground, and dilution plated.
842 Population size of each strain was normalized by initial inoculum. Median CI of the Δfcs
843 mutant was 0.71 ($P=0.225$, Wilcoxon Signed Rank Test; N=14 plants per co-inoculation, 28
844 total).

845
846 **Fig. 6.** HCA degradation by *R. solanacearum* did not affect total lignin quantity or
847 distribution in tomato stems. **A**, Mean gross lignin content in tomato stems at 6 days post
848 soil-soak inoculation. Whole stems of healthy (mock-inoculated) or infected tomato plants
849 ($n=6$ per condition) were dried, ground, and analyzed by the acetyl bromide lignin
850 quantification assay using wood pulp inulin as a standard. Error bars indicate standard

851 error of the mean. Similar results were obtained at 3 and 9 days after inoculation; **B-D**,
852 Phloroglucinol HCl-stained cross-sections of stems from representative healthy or
853 symptomatic infected plants. Pink precipitate indicates lignin.

854

855 **Fig. 7.** HCA degradation protected *R. solanacearum* from HCA toxicity. Bacterial growth in
856 succinate minimal medium supplemented with increasing concentrations of: **A**, *p*-
857 coumarate (*p*-Cou); **B**, caffeate (Caf); or **C**, ferulate (Fer). Culture optical density was
858 measured by a plate reader 48 hr after inoculation with 10⁵ CFU/ml of bacteria. Growth of
859 each strain was calculated relative to that of wild-type bacteria growing without HCAs.
860 Error bars indicate standard error of the mean. The WT strain was less inhibited than the
861 Δfcs mutant by *p*-coumarate and caffeate (*t*-test; *P* < 0.005).

862

863 **e-Xtra Fig. S1.** Kanamycin and gentamicin resistance markers did not affect competitive
864 fitness of strains. **A**, Soil soak competition data of WT and Δfcs bacteria from Fig. 5; **B**,
865 Petiole competition data of WT and Δfcs bacteria from Fig. 5 Results from the WT-kan: Δfcs -
866 gm inoculation are shown in filled squares and results from the WT-gm: Δfcs -kan
867 inoculation are shown in open circles.

868 **e-Xtra: Table S1: *R. solanacearum* strains used for genomic and functional analyses**

Strain	Phylotype- sequevar	Host isolated from	Location	Genome Reference
GMI1000	I	Tomato	French Guiana	(Boucher et al., 1985; Salanoubat et al., 2002)
FQY4	I	Bacterial wilt nursery	China	(Cao et al., 2013)
Y45	I	Tobacco	China	(Li et al., 2011)
Grenada91	IIA-6	Banana	Grenada	Ailloud and Prior, unpublished
UW181	IIA-6	Plantain	Venezuela	Ailloud and Prior, unpublished
K60	IIA-7	Tomato	USA	(Remenant et al., 2012)
B50	IIA-24	Banana	Brazil	Ailloud and Prior, unpublished
IBSBF1900	IIA-24	Banana	Brazil	(Wicker et al., 2007)
CFBP2957	IIA-36	Tomato	French West Indies	(Remenant et al., 2010)

UW491	IIB-1	Potato	Colombia	Vinatzer, unpublished
UW551	IIB-1	Geranium	Kenya	(Gabriel et al., 2006)
CFBP1416	IIB-3	Plantain	Costa Rica	Ailloud and Prior, unpublished
CIP417	IIB-3	Banana	Philippines	Ailloud and Prior, unpublished
Molk2	IIB-3	Banana	Philippines	Boucher et al., unpublished
CFBP6783	IIB-4	Heliconia	West Indies	Ailloud and Prior, unpublished
IBSBF1503	IIB-4	Cucumber	Brazil	Ailloud and Prior, unpublished
Po82	IIB-4	Potato	Mexico	(Xu et al., 2011)
UW163	IIB-4	Plantain	Peru	Ailloud unpublished
UW179	IIB-4	Banana	Colombia	Ailloud and Prior, unpublished
CMR15	III	Tomato	Cameroon	(Remenant et al., 2010)
BDB R229	IV	Banana	Indonesia	(Remenant et al.,

Lowe, Tiffany - *MPMI*

				2011)
PSI07	IV	Tomato	Indonesia	(Remenant et al., 2010)
R. syzygii R24	IV	Clove	Indonesia	(Remenant et al., 2011)

869

870 **e-Xtra Table S2** can be downloaded at:871 http://www.scientificsocieties.org/MPMIXtras/2014/MPMI-09-14-0292-FI_TableS2.xlsx

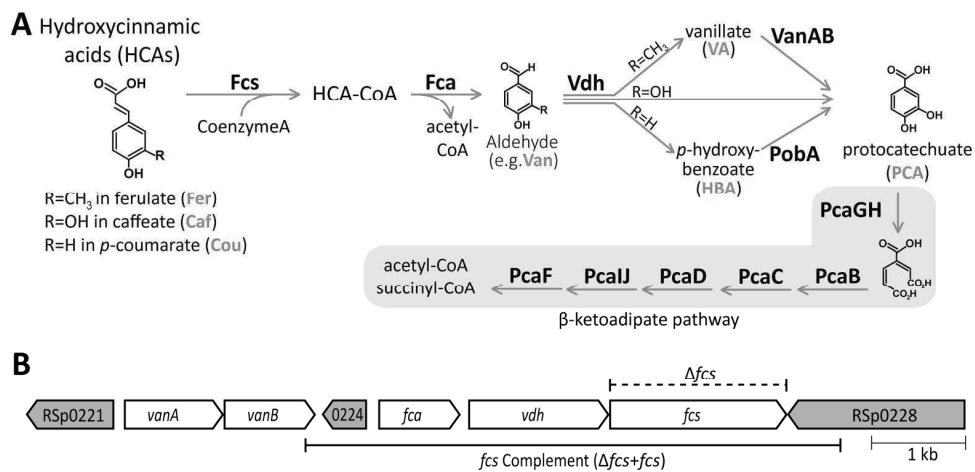


Figure 1
175x85mm (300 x 300 DPI)

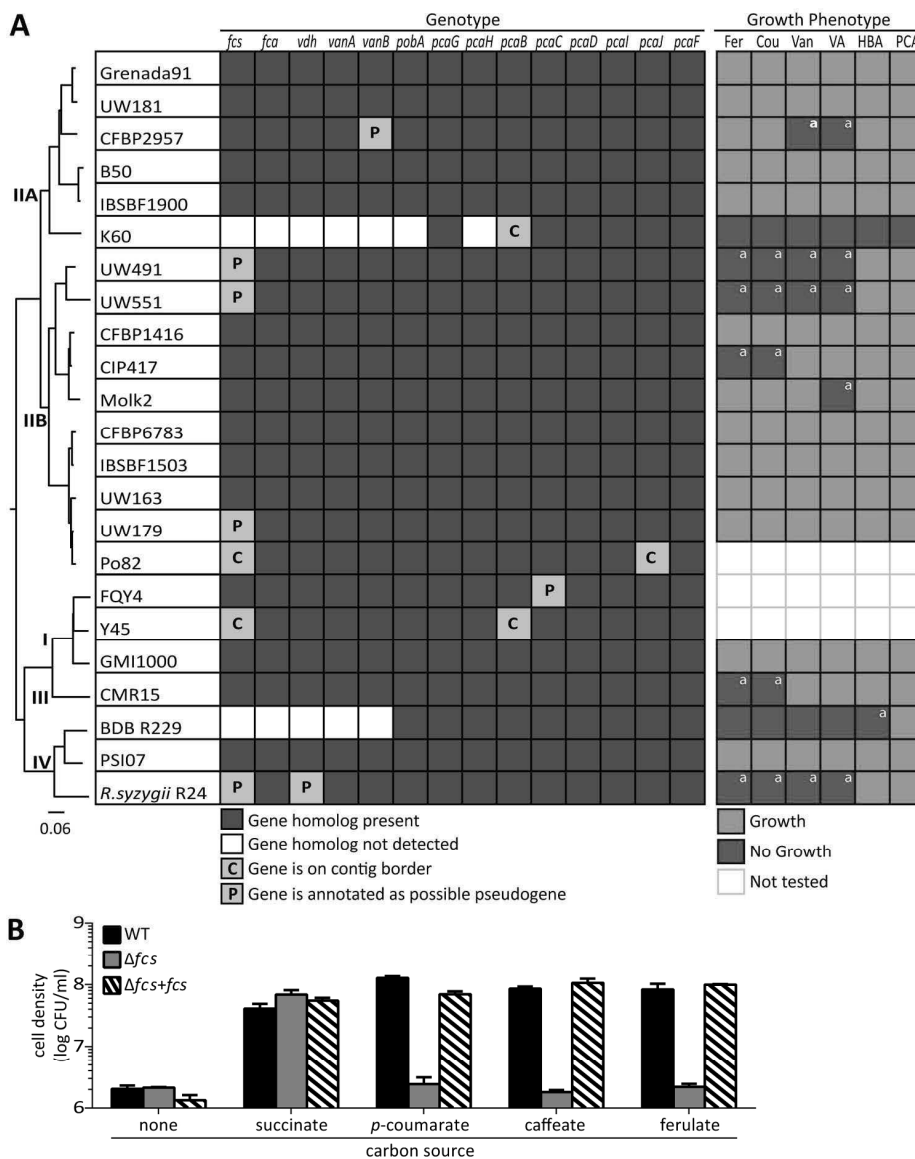


Figure 2
 175x220mm (300 x 300 DPI)

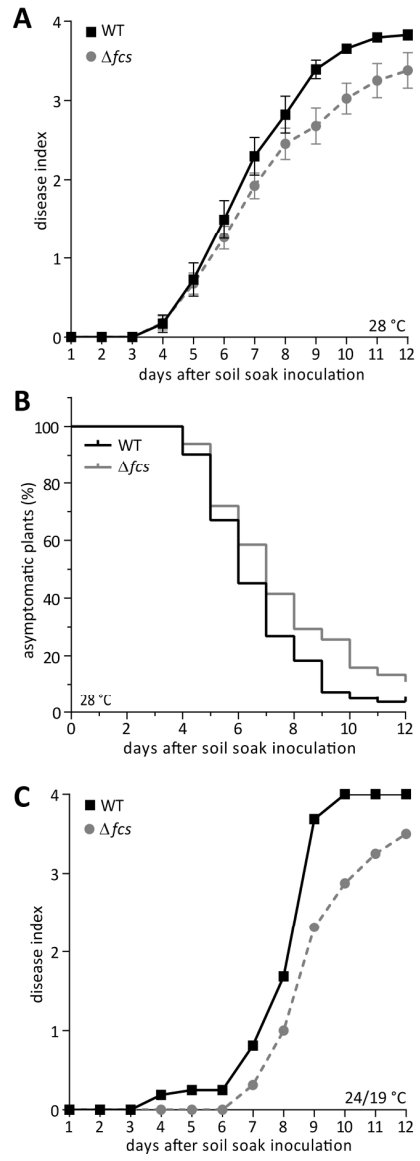


Figure 3.
81x232mm (300 x 300 DPI)

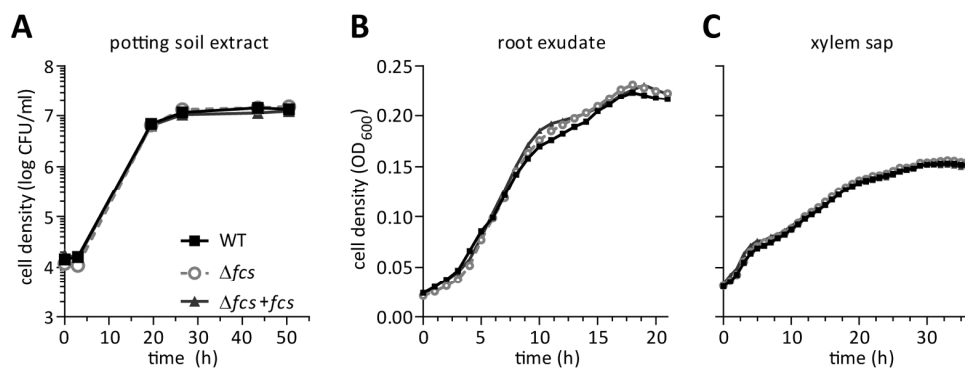


Figure 4.
169x64mm (300 x 300 DPI)

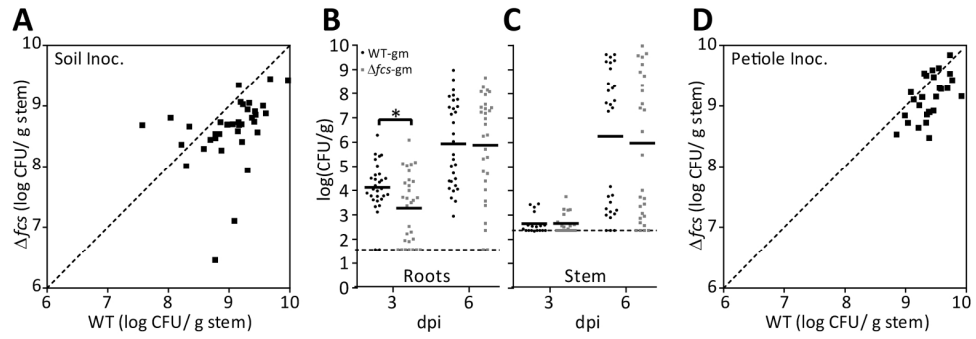


Figure 5.
176x60mm (300 x 300 DPI)

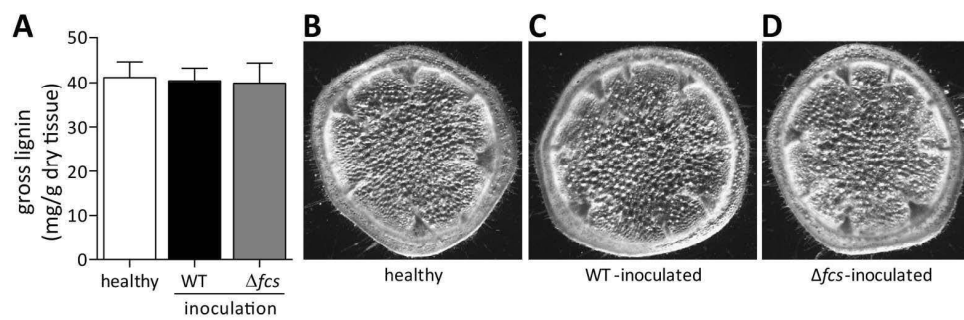


Figure 6.
175x57mm (300 x 300 DPI)

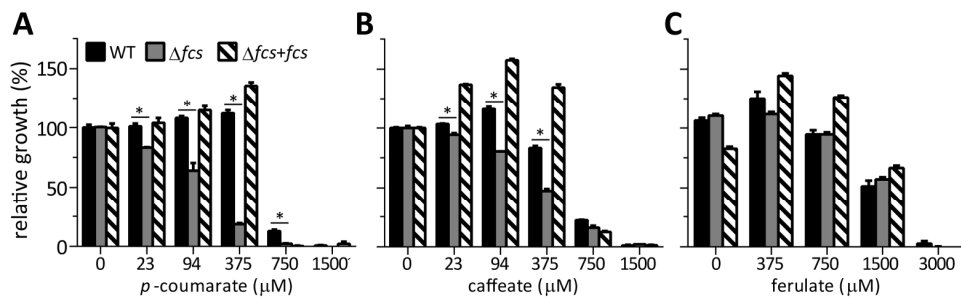
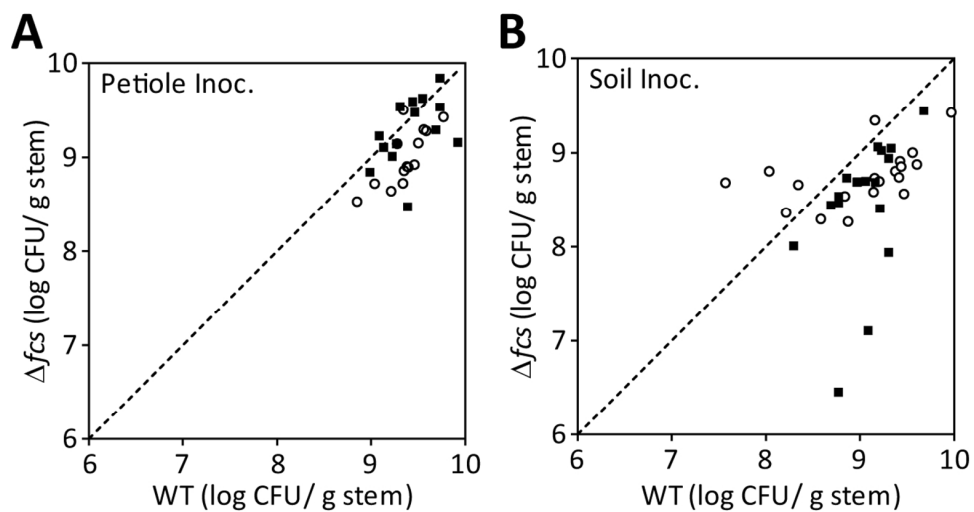


Figure 7.
 175x53mm (300 x 300 DPI)



Supplemental Figure 1.
113x60mm (300 x 300 DPI)

***e*-Xtra Table S2: Distribution of HCA degradation genes in bacteria with complete genomes**

Data table can be downloaded at the following location:

https://drive.google.com/file/d/0Byypeh_Akk8yMlpUOU1rS0xHS00/