Ecological specificities of Agrobacterium fabrum : role of its specific genes in the interaction with the plant

Rosa Maria Padilla Aguilar

N°d'ordre NNT : xxx

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 341
Ecosystèmes Evolution Modélisation Microbiologie

Spécialité de doctorat : Biologie
Discipline : Ecologie Microbienne

Soutenue publiquement le 29/06/2020, par :

Rosa Padilla

Spécificités écologiques d'*Agrobacterium fabrum* : rôle des gènes spécifiques dans l'interaction avec la plante

Devant le jury composé de :

CHIAPUSIO Geneviève, Maître de Conférences, Université de Savoie Mont-Blanc	Rapporteur
RHOUMA Ali, Professeur, PRIMA Foundation	Rapporteur
SANCHEZ Lisa, Ingénieur de Recherche, Université de Reims	Rapporteur
COMTE Gilles, Professeur, Université Lyon 1	Examinateur
UROZ Stephane, Directeur de Recherche, Université de Lorraine	Examinateur
WISNIEWSKI-DYÉ Florence, Professeure, Université Lyon 1	Examinateur
KERZAON Isabelle, Maître de Conférences, Université Lyon 1	Co-directrice de Thèse
NESME Xavier, Ingénieur de Recherche, Université Lyon 1	Directeur de Thèse

Université Claude Bernard – LYON 1

Président de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Damien VERHAEGHE

COMPOSANTES SANTE

Faculté de Médecine Lyon-Est – Claude Bernard	Doyen : M. Gilles RODE
Faculté de Médecine et Maïeutique Lyon Sud Charles. Mérieux	Doyenne : Mme Carole BURILLON
UFR d'Odontologie	Doyenne : Mme Dominique SEUX
Institut des Sciences Pharmaceutiques et Biologiques	Directrice : Mme Christine VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. Xavier PERROT
Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

UFR Biosciences	Directrice : Mme Kathrin GIESELER
Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
UFR - Faculté des Sciences	Administrateur provisoire : M. Bruno ANDRIOLETTI
UFR (STAPS)	Directeur : M. Yannick VANPOULLE
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Ecole Polytechnique Universitaire Lyon 1	Directeur : Emmanuel PERRIN
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Institut de Science Financière et d'Assurances	Directeur : M. Nicolas LEBOISNE
ESPE	Administrateur Provisoire : M. Pierre CHAREYRON

If we knew what it was we were doing, it would not be called research

Albert Einstein

Remerciements

Je tiens tout d'abord à remercier les membres du jury, Geneviève Chiapusio, Ali Rhouma Lisa Sanchez, Gilles Comte, Stephane Uroz et Florence Wisniewski-Dyé d'avoir accepté d'évaluer mes travaux de thèse.

Je remercie Yvan Moënne-loccoz pour m'avoir accueilli au sein de son laboratoire. J'adresse mes remerciements à Xavier Nesme, Céline Lavire et Ludovic Vial pour m'avoir accueilli dans leur équipe.

Je remercie aussi Xavier, mon directeur de thèse, avec qui les rigolades mélangées à la science n'ont jamais manqué et Isabelle, ma co-directrice, qui a toujours un sourire dans n'importe quelle situation. Je tiens à remercier tout particulièrement Ludo, mon co-encadrant, pour tous ses conseils, son aide, son temps, ces blagues et tous les bons moments passés ensemble, mais surtout pour avoir toujours cru en moi, même quand il ne croyait pas en moi. Tu sais que je ne serais pas arrivée jusque-là sans toi. Merci infiniment pour tout ce que tu m'as appris et apporté. J'espère ne pas t'avoir déçu. Et aussi merci à Céline, qui, dans cette thèse déjà bien surpeuplée de chefs, a tellement apporté avec toujours cette tendresse qui la caractérise si bien. Merci de ton aide, ton temps, ta gentillesse et de tous tes conseils.

Je remercie tous les membres du CESN pour toute leur aide. Tout particulièrement Gilles et Guillaume, merci pour votre patience, pour tout ce que vous m'aviez appris et tout le temps que vous m'avez consacré. Mais aussi à Laurent, Serge, Vincent, Marie-Geneviève et Anne-Emmanuelle, vous êtes des gens fantastiques et tellement sympa.

Je remercie les permanents du Mendel. Hasna, Petar, Danis, Daniel, Florence WD, Franck, Zahar, Jeanne, Alessandro, Xavier et Sabine, vous êtes de gens formidables. Mais aussi les non-Mendel, Veronica, Patricia, Mylène, merci pour votre bonne humeur. Merci également à Elise pour son aide et pour toutes ces plantes que j'ai ensuite massacré. Et merci à toi ma Corinne, sans toute ton aide, je ne sais pas ce que j'aurais fait. T'es une femme en or !

Je ne peux pas oublier de remercier les anciens Pauline, Armelle et Florence. Mais aussi à tous les anciens thésards qui m'ont tellement appris. Merci infiniment à tous mes stagiaires, notamment Xan et Clém, vous êtes les meilleures les filles et vous allez arriver tellement loin ! Et finalement toi ma petite Flo, merci de m'avoir redonné le sourire, merci de tous ces moments passés ensembles à rire et à nous soutenir l'une et l'autre. T'es une trop géniale !

Mais je tiens tout particulièrement à remercier ma famille. Rémi, j'ai toujours su que la vie, il faut la partager avec quelqu'un meilleur que soi et essayer toute ta vie de lui ressembler. Merci pour être celui qui m'inspire à devenir une meilleure version de moi-même. Tu as toujours été mon soutien, tu as toujours su m'encourager et m'apaiser. Merci pour tous les sacrifices que tu as dû endurer depuis mon master et puis ma thèse. Je m'excuse pour toutes ces soirées et weekends en solitude en attendant que j'arrive à la maison. Merci pour tout ce que tu fais pour moi, merci pour ton aide dans tous les domaines, tu es tellement magnifique. Et finalement toi, mon Petit Chou, tu es un bébé si parfait, avec toujours un sourire, plein d'énergie et l'envie de gambader partout. Ta curiosité et bonne humeur m'émerveillent toujours. Ne grandit pas trop vite ! Je te demande pardon pour ne pas t'avoir dédié autant de temps que je l'aurais voulu pendant cette rédaction. Cette thèse vous est dédiée.

Spécificités écologiques d'*Agrobacterium fabrum* : rôle des gènes spécifiques dans l'interaction avec la plante

Résumé

Agrobacterium est un genre bactérien composé de nombreuses espèces capables d'établir des interactions commensales voire favorables dans la rhizosphère des plantes. Mais elles sont principalement connues pour être phytopathogènes en induisant des tumeurs (symptôme de la galle du collet), sur de nombreuses espèces végétales. Remarquablement, des études de terrain ont montré que plusieurs espèces d'agrobactéries coexistent généralement dans les mêmes biotopes. Pour expliquer l'occurrence et la persistance de cette biodiversité, nous émettons l'hypothèse que les différentes espèces doivent avoir des niches écologiques particulières leur permettant d'échapper aux compétitions avec les espèces les plus apparentées. Des études de génomique comparative ont montré l'existence de gènes spécifiques d'*A. fabrum* groupés en sept régions génomiques appelées « régions spécifiques » codant des unités fonctionnelles hypothétiques liées au métabolisme ou au transport de composés végétaux (sucres, composés phénoliques…). Cela nous a permis de prédire une écologie spécifique liée aux interactions bactéries-plantes et leur implication dans la construction de la niche écologique spécifique de cette bactérie.

Ces travaux visent à vérifier si les régions spécifiques à *A. fabrum* sont impliquées dans les adaptations à la plante dans ces deux styles de vie (commensal et pathogène). À cette fin, nous avons étudié l'effet des régions spécifiques sur la valeur adaptative bactérienne *in planta* et sur le métabolome végétal (racine et tumeur) en utilisant une approche métabolomique. Grâce à des tests de compétition, nous avons constaté que la plupart des régions spécifiques d'*A. fabrum* lui confèrent une meilleure valeur adaptative dans la plante. Certaines régions sont plus impliquées dans un style de vie plutôt que l'autre, certaines dans les deux à la fois. D'autre part, les analyses métabolomiques ont montré qu'*A. fabrum* est capable de moduler la teneur en composés phénoliques, notamment en flavonoïdes, dans les racines et tumeurs, et que cette modulation était essentiellement liée aux régions spécifiques.

Ces résultats contribuent à une meilleure compréhension de la construction de niche écologique d'*A. fabrum* en mettant en évidence l'importance de ses gènes spécifiques dans l'établissement de cette interaction fine.

Mots-clés : *Agrobacterium fabrum*, interaction plantes-bactéries, gènes spécifiques, valeur adaptative, métabolomique, métabolites secondaires, flavonoïdes.

Ecological specificities of *Agrobacterium fabrum*: role of its specific genes in the interaction with the plant

Abstract

Agrobacterium is a bacterial genus composed of many species capable of establishing commensal or pathogenic (by inducing the crown-gall disease) interactions. Field investigations showed that several species of agrobacteria generally co-exist in the same biotopes. To explain this biodiversity, we hypothesize that different species must have particular ecological niches allowing them to evade from competitions. Comparative genomics showed the existence of *A. fabrum*-specific genes clustered into genomic regions called "specific regions" encoding putative functional units related to metabolism or transport of plant compounds. This allowed us to predict a specific ecology related to bacteria-plant interactions and their implication in the specific niche construction of this bacteria.

This study aims at verifying whether *A. fabrum*-specific regions are involved in adaptations to the plant in both lifestyles (commensal and pathogenic). For this purpose, we studied the effect of the specific regions on the bacterial fitness *in planta* and on the plant metabolome (rhizosphere and tumor). Competition tests reveal that almost all the *A. fabrum*-specific regions confer better fitness to bacteria. Some regions are more involved in one lifestyle more than the other, some in both. Metabolomic analyzes showed that *A. fabrum* is capable of modulating the content of phenolic compounds in roots and tumors, in particular flavonoids. This modulation was essentially linked to the specific regions.

These results contribute to a better understanding of the ecological niche construction of *A. fabrum* highlighting the importance of its specific genes in the establishment of this fine-tuned interaction.

Keywords: *Agrobacterium fabrum*, plant-bacteria interaction, specific genes, fitness, metabolomics, secondary metabolites, flavonoids.

Table de matières

Liste de figures et tableaux

Tableaux

Chapitre 2 : Implication des gènes spécifiques d'*A. fabrum* sur l'adaptation *in planta*

Figures

Chapter 3: Effects of *A. fabrum*-specific regions on plant metabolomic profile

I. Metabolomic study of *A. fabrum* commensal lifestyle on root secondary metabolites

Figures

II. Metabolomic study of *A. fabrum* pathogenic lifestyle in plant tumors

A. Secondary metabolites

B. Opines: A new method of detection and quantification

Figures

Tableaux

General Introduction

The rhizosphere of plants, defined as the area of the soil under the direct influence of roots and plant tissue, is in close contact with innumerable microorganisms (Bais et al. 2006). Plant-bacteria interactions can be commensal or even mutually beneficial, such as the root nodule symbiosis with nitrogen-fixing bacteria (Pitzschke 2013). But these interactions can also be pathogenic and damage the plant causing diseases. Agrobacteria, ubiquitous bacteria from the soil and the rhizosphere belonging to the family *Rhizobiaceae*, are an interesting case as they are capable of having two different lifestyles. One is a commensal lifestyle, where they can live in a saprophytic way in rhizospheres establishing even beneficial interactions with the plant in a PGPB type (Plant Growth-Promoting Bacteria) (Walker et al. 2013; Glaeser et al. 2016; Hao et al. 2012). The other is the pathogenic lifestyle, in which agrobacteria cause crow-gall disease, inducing the formation of tumors (Kado 2014). However, the real etiological agent of this plant disease is a plasmid (the pTi) which carries all the virulence genes and thus confers to bacteria the capacity to induce this bacteriosis (Hong et al. 1997). Some of these virulence genes will be transferred, integrated and expressed in the plant genome (**Fig. 1**). Thereafter, an anarchic growth of plant cells takes place, as well as the diversion of the plant metabolism towards the production of small molecules called opines that serves as carbon sources specifically to agrobacteria (Tempé and Petit 1982; Baek et al. 2003). Thus, agrobacteria genetically modify the plant cells host to create its specific ecological niche where they can be fed and housed. An ecological niche is defined, in a given environment, by the availability of nutritive resources, which can be specifically assimilated by certain living organisms to promote their proliferation (Lang et al. 2014).

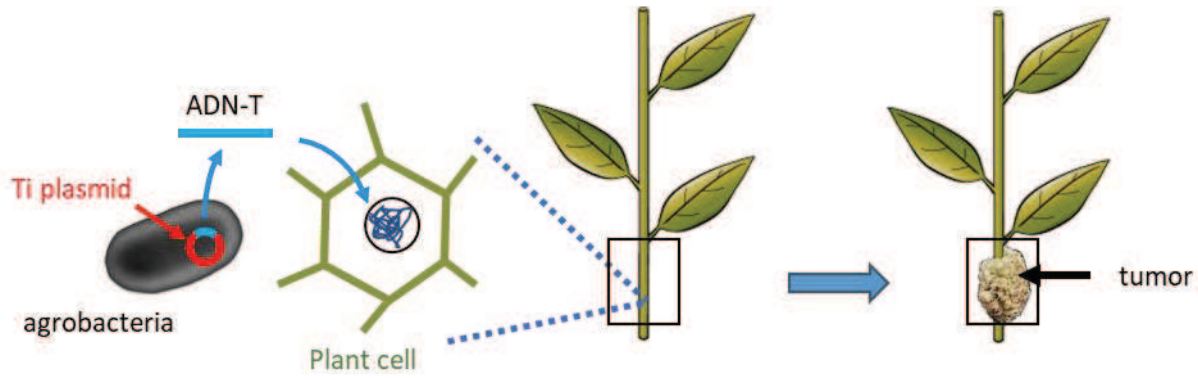


Fig. 1. Establishment of crown gall disease by agrobacteria. The Ti plasmid, inside the bacterial cell, contains the virulence genes. A portion of this plasmid, called T-DNA is transferred to the plant cell, where it will be integrated and expressed. This will cause the formation of the plant tumor.

The species complex "*Agrobacterium tumefaciens*" denotes agrobacteria which have both a great diversity of species and a great diversity intra-species. (Mougel et al. 2002; Portier et al. 2006). Field investigations showed that the same rhizosphere can host several species of agrobacteria living

in sympatry. So, in accordance with the principle of competitive exclusion, these species, although very related, must exploit different resources to escape competition and thus durably co-exist (ref). Therefore, according to the hypothesis formulated by Lassalle et al. in 2011, these species must have their own genes allowing them to exploit at least partially different ecological niches. The G8 species-specific genes, called Agrobacterium fabrum, have been identified by comparative genomics. These are 196 genes mainly clustered into 7 genomic regions called "G8 Specific Regions" or "SpG8" (named SpG8-1 to SpG8-7). One of these regions is located on the circular chromosome and the other six are located on the linear chromosome of *A. fabrum* (**Fig. 2**). Th annotation of these regions indicates that they encode hypothetical functional units related to metabolism or transport of plant compounds (sugar, phenolic compounds…). This allowed us to predict a specific ecology related to bacteria-plant interactions and the SpG8 functions seemed to collectively define an ecological specific niche of *A. fabrum*.

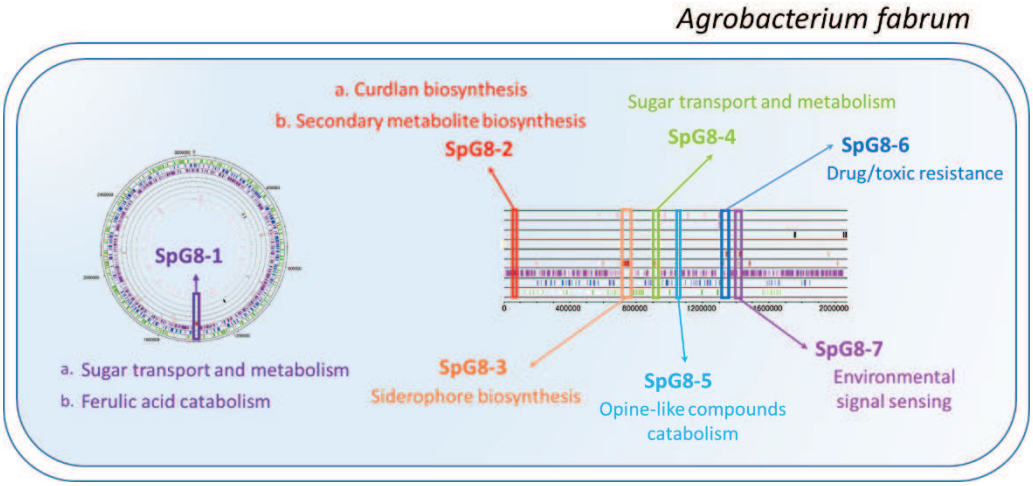


Fig. 2. Location of the *A. fabrum*-specific regions in chromosomes. Representation of the circular and linear chromosome of *A. fabrum* and the location of its specific regions with their hypothetical functions (after Lassalle et al. 2011).

The objective of this thesis is to assess the involvement *A. fabrum*-specific genes in the adaptation to the plant and how it manages to constructs its specific ecological niche. For this purpose, we have on the one hand measured the role played by specific regions in bacterial fitness in the plant, and on the other, the effect on plant secondary metabolites following the inoculation of *A. fabrum*.

Thereby, in a first part of bibliographic synthesis, we will discuss state-of-the-art knowledge on *Agrobacterium* taxonomy, the concept of bacterial species and two approaches methods to study a plant-bacteria interactions. Then, in a second experimental part, the results of this study will be presented in the form of two chapters. The first chapter will focus on the involvement of specific genes

on plant adaptation. We will present the methodological development that we used to measure bacterial fitness in planta that we applied to the two lifestyles of A. fabrum. The second chapter will concern the study of the effects of *A. fabrum* specific regions on the plant metabolome following the inoculation. We will present two metabolic studies carried out on the rhizosphere, to study root secondary metabolites, and in the tumor induced by *A. fabrum*, to study not only tumor secondary metabolites, but also opines produced by transformed cell plants. Furthermore, we present a new method of detection and quantification of these molecules. At the end of this experimental part, we will integrate the results obtained and discuss the importance of the *A. fabrum*-specific regions in the construction of the specific ecological niche of *A. fabrum* (**Fig. 3**).

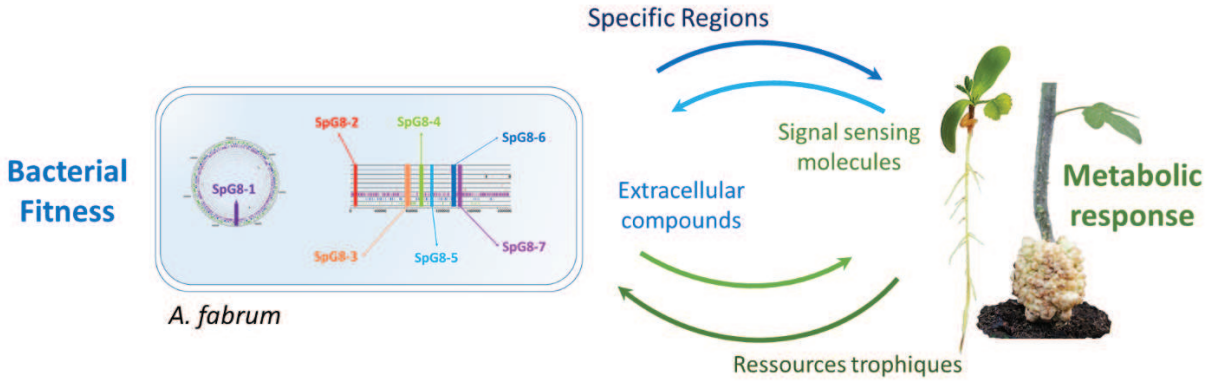


Fig. 3. Construction of the specific ecological niche of *A. fabrum*. Involvement of A. fabrum specific genes in the adaptation to the plant and the metabolic response of the latter.

Chapter

1

Bibliographic synthesis

I. *Agrobacterium* genus and taxonomy

The definition of *Agrobacterium* that we will use in this work does not take into account the presence of a plasmid, either as a pTi or a pRi. Indeed, it has already been demonstrated, it is an accessory element of this bacterium and a secondary ecological trait.

The family *Rhizobiaceae* belongs to the order *Rhizobiales* in the class α-proteobacteria. The *Rhizobiaceae* comprises six genera harboring plant-associated species, including among others the *Agrobacterium* genus that gather together different species of bacteria, mostly saprophytic and non-pathogenic. *Agrobacterium genus* harbors two megabase-sized chromosomes (Allardet-Servent et al. 1993) and more specifically, a chromid which is linear (Slater et al. 2009; Allardet-Servent et al. 1993; Goodner et al. 2001; Wood et al. 2001) Within the *Rhizobiaceae* family, the single acquisition event of a *telA* gene by an ancestor allowed the chromid linearization and maintenance of the linear geometry in the ancestor progeny **(Figure 1)** (Ramírez-Bahena et al. 2014). Thus, this constitutes a synapomorphy of this clade (Ramírez-Bahena et al. 2014).

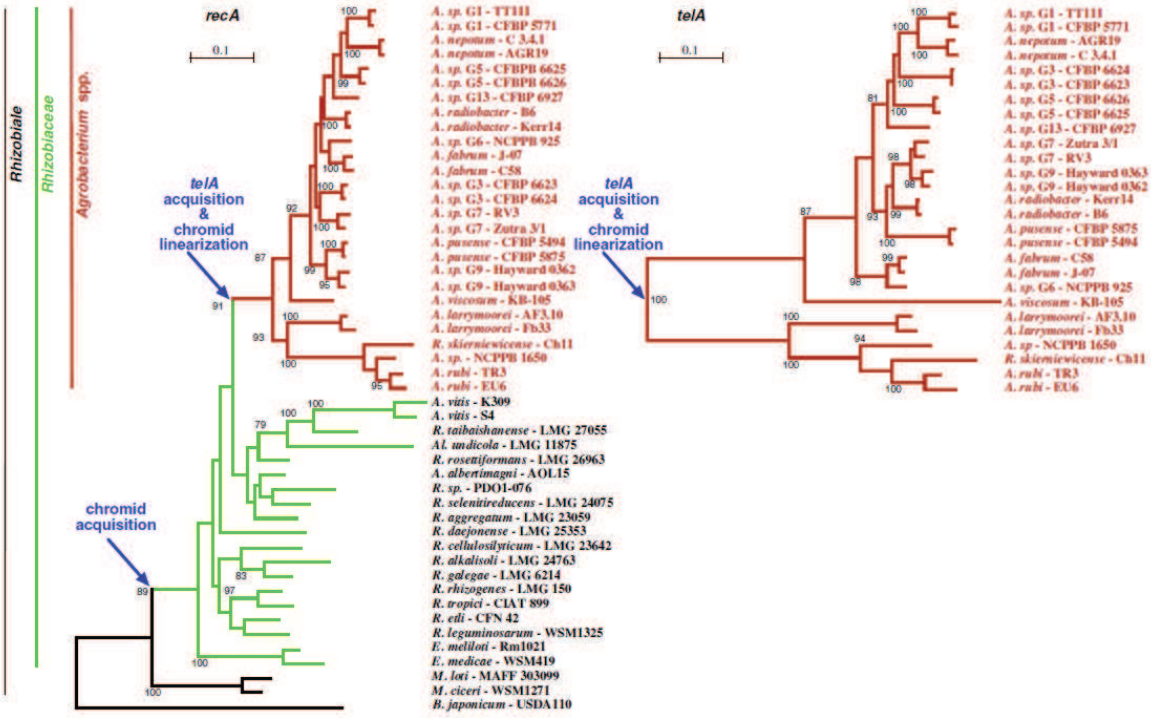


Figure 1. Phylogenetic tree of the Rhizobiaceae family based on the *recA* gene. The genus *Agrobacterium* is characterized by the presence of the *telA* gene and by the linearization of the chromide. All *Agrobacterium* therefore have a linear chromide and a circular chromosome while the other Rhizobiaceae have two circular replicons. (According to Ramírez-Bahena et al. 2014)

Gene density is very similar between the two chromosomes. However, essential genes are significantly overrepresented on the circular chromosome (Tatusov et al. 2001). This asymmetry is consistent with direct descent of the circular chromosome from the primordial α-proteobacterial genome, with a minority of essential genes moving to the linear chromosome (Goodner et al. 1999). The circular chromosome contains a putative origin of replication (Cori). The linear chromosome, on the other hand, has a plasmid-type replication system (Goodner et al. 2001).

The taxonomy of agrobacteria has long been confused and constantly changing over the past century. Historically, standard biochemical galleries and pathogenicity assays lead to the differentiation of agrobacteria into at least three biovars named biovar 1, 2 and 3 (Keane et al. 1970; Smith and Townsend 1907; Conn 1942) and mostly based on the symptomatology of host plants. But with the discovery of the plasmid Ti (Tumor-inducing), an accessory and transferable replicon carrying the virulence genes, this classification was inaccurate. Then, in 2001 Young et al. explored the history of agrobacteria taxonomic issue and proposed to transfer all members of the genera *Agrobacterium* and *Allorhizobium* to the genus *Rhizobium* (Young et al. 2001). But in 2003 Farrand et al. supported *Agrobacterium* as a proper name for this group of pathogenic rhizobial species (Mousavi et al. 2014; Farrand et al. 2003).

Hybridization studies have revealed, however, that biovar 1 is not a homogeneous species as it displays a too large genomic divergence (Popoff et al. 1984). Therefore thus, from a genomic point of view, biovar 1 was proposed to be called "*A. tumefaciens* complex" **(Figure 2)** (Costechareyre et al. 2010). Also, using multilocus sequence analysis (MLSA) it was shown that *A. vitis* and *A. rhizogenes* does not belong to the *Agrobacterium* genus since now they are reclassified as *Allorhizobium vitis*, and *Rhizobium rhizogenes* (Mousavi et al. 2014). Most of biovar 1 genomic species have not yet received accepted latin binomial names and are currently designated genomovar G1 to G9 (Mougel et al. 2002), G13 (Portier et al. 2006), G14 (Puławska and Kałużna 2012), G19 and G20 (Mafakheri et al. 2019) **(Figure 2)**. Five out of 11 genomospecies have valid species-level binominal names, i.e., *A. pusense* for G2 (Panday et al. 2011), *A. radiobacter* for G4 (Young et al. 2006), *A. nepotum* for G14 (Puławska and Kałużna 2012)), *A. fabrum* for G8 (Lassalle et al. 2011), *A. salinitolerans* for G9 (Yan et al. 2017b) and *A. deltaense* for G7 (Yan et al. 2017a).

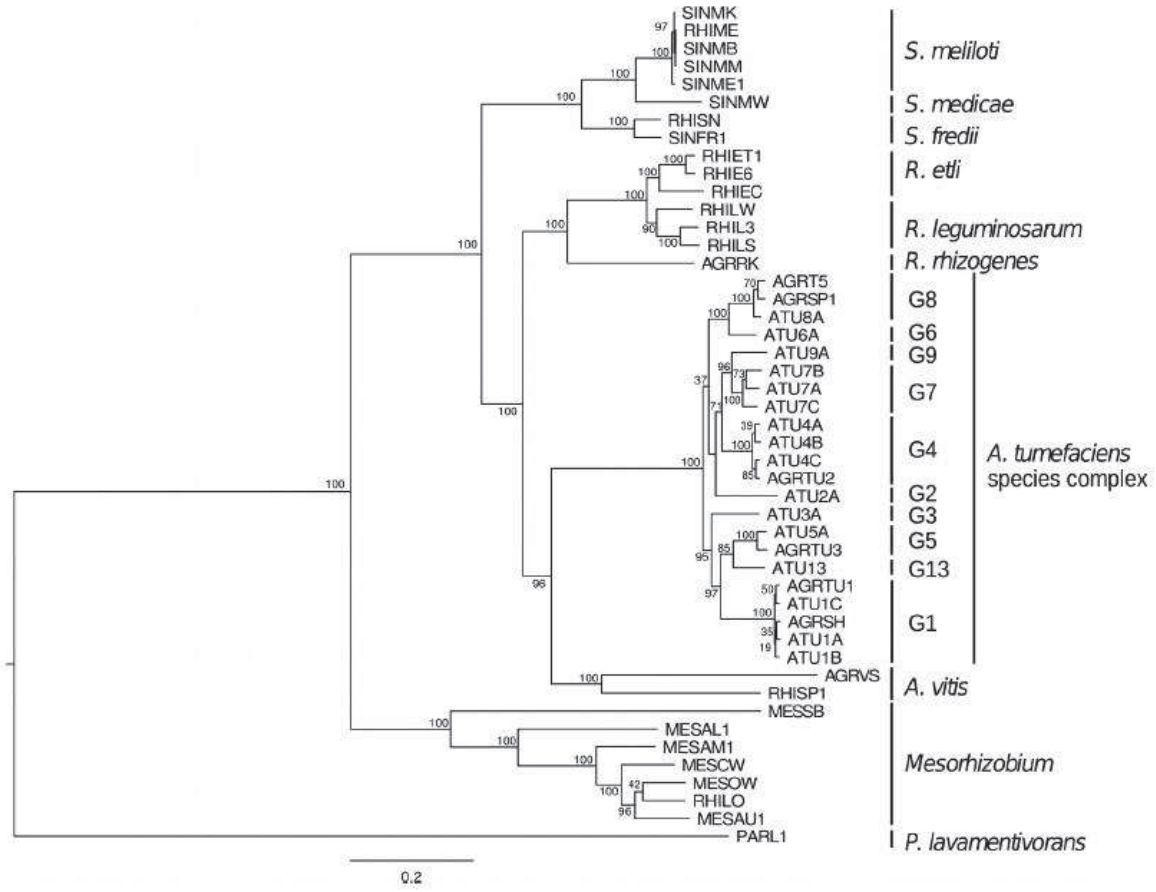


Figure 2. Reference phylogeny of Rhizobiales history (According to Lassalle et al. 2017)

Agrobacterium are native soil organisms. Although this bacterium is known to be phytopathogenic, it is capable of establishing commensal or even beneficial interactions of Plant Growth Promoting Bacteria (PGPB) type in the rhizosphere of plants (Hao et al. 2012; Walker et al. 2013; Naqqash et al. 2016; Bhattacharyya and Jha 2012; Chihaoui et al. 2012; Glaeser et al. 2016). The model bacteria *A. fabrum* C58 was originally isolated in 1958 by Robert Dickey from a cherry gall in upstate New York (Hamilton and Fall 1971). Its 5.67-megabase genome contains 5419 predicted protein-coding genes (Wood et al. 2001) and comprises four replicons: a primary circular chromosome of 3,000 kb, a chromid (Harrison et al. 2010)of 2,100 kb and two plasmids (the AtC58 of 450 kb and the TiC58 of 200 kb) (Wood et al. 2001; Allardet-Servent et al. 1993).

The overall GC content of the *A. fabrum* genome is 58%. The two largest gene families are composed of genes belonging to the adenosine triphosphatase (ATPase) and membrane-spanning components of the ATP binding cassette (ABC) transport family. The genome contains 53 transfer RNAs (tRNAs) that represent all 20 amino acids. These tRNAs are distributed unevenly between the circular and linear chromosomes. Transporters constitute 15% of the *A. fabrum* genome, with an abundance

of ABC transporters (60% of its total). This may reflect a need for high-affinity uptake systems for the acquisition of nutrients in the highly competitive soil and rhizosphere environments.

A. *Agrobacterium* primary ecology

Members of the genus *Agrobacterium* constitute a diverse group of organisms of high known diversity (Mougel et al. 2002; Portier et al. 2006; Costechareyre et al. 2010) that are ubiquitous inhabitants of soils and rhizospheres (Bouzar and Moore 1987; Farrand et al. 2003; Teixeira et al. 2010). They can be found in virtually every habitat, including water and sediments (D'Hondt et al. 2004; Süss et al. 2006). Agrobacteria are composed of several strains and genomic species commonly co-inhabiting in the same soil samples (Costechareyre et al. 2010), even at the very microscale (Vogel et al. 2003).

Since *Agrobacterium* populations live in different habitats (bare soil, rhizosphere, host plants), they face different environmental constraints. Agrobacteria have evolved the capacity to exploit diverse resources and to escape plant defense and competition from other microbiota (Dessaux and Faure 2018). Agrobacterial species are not geographically isolated and they have determinants for species-specific ecological niche. It is likely that speciations in the same habitat occurred as a consequence of local adaptations to host plants, as suggested by annotations of G8-specific functions (Lassalle et al. 2011). Lamovsek et al. found in Slovenia soil that the most frequently isolated genomic species was G1 (62%), followed by G4 (38%). Previously it had also been shown that the population of G1 was one of the most frequently isolated genomic species from non-contaminated soil from corn fields (Vogel et al. 2003). Also that both populations of both genomic species can be present at the same location at the same time (Lamovsek et al. 2014). The two species thus very likely inhabit different microhabitats (Lassalle et al. 2011).

The pathogenic forms of agrobacteria characteristically harbor a Ti plasmid, the key replicon that determines virulence (Van Larebeke et al. 1974). However, in soil or root ecosystems, most agrobacteria do not harbor such a plasmid and are avirulent (Bouzar and Moore 1987). Conversely, pathogenic agrobacteria have almost only been reported following isolation from diseased plants or from soils of contaminated areas with a history of crown gall disease (Krimi et al. 2002). Nonpathogenic agrobacteria are therefore generally saprophytic and inoffensive plant commensals (Costechareyre et al. 2010; Sanguin et al. 2006a).

In soil, plant roots are colonized by a wide range of bacteria (Sanguin et al. 2006b). Plant growth-promoting rhizobacteria (PGPR) play an important role, because they may enhance root system development, nutrient and water uptake, stress tolerance, or plant health (Raaijmakers et al. 2009). The first evidence of agrobacteria functioning as plant-beneficial bacteria in nonsterile soil was in the maize rhizosphere (Walker et al. 2013). *Agrobacterium* genus colonizes the maize rhizosphere extensively as a saprophyte because it does not induce disease in monocots (Sanguin et al. 2006b). Moreover, seed inoculation of maize with *A. fabrum* C58 enhance shoot and root biomass of certain type of maize. Maize phytostimulation by C58 was of a magnitude similar to that of the most effective *Azospirillum* PGPR strain, as well as the modifications induced in the secondary metabolite profile of maize. This suggests that these two bacteria were recognized by this maize cultivar as belonging to the same "type" of microorganism (Walker et al. 2013).

Another example of agrobacteria being a PGPR is with the formally called *Rhizobium radiobacter* F4 but belonging to G8 species that was originally characterized as an endofungal bacterium in the beneficial endophytic Sebacinalean fungus *Piriformospora indica* (Glaeser et al. 2016). RrF4 could be isolated from powdered fungal mycelia and propagated in axenic cultures, and the RrF4-colonized plants show increased biomass and enhanced resistance against bacterial leaf pathogens (Glaeser et al. 2016). So, it is possible that the beneficial biological activity previously assigned to P. indica may be at least partly allotted to the bacterium. So, RrF4's biological activity is in several aspects comparable with other efficient PGPR indicating that endofungal bacteria of the Sebacinalean symbiosis are a valuable source of beneficial bacteria with agronomical potential (Glaeser et al. 2016).

Plant-growth-promoting bacteria (PGPB) can also play a key role in host plant adaptation to metal-contaminated environments through triggering physiological changes in plant cell metabolism as in the case of *A. tumefaciens* CCNWGS0286, isolated from the nodules of *Robinia pseudoacacia* growing in zinc-lead mine tailings (Hao et al. 2012). This beneficial bacterium that assist rhizobia with legume nodulation (de Lajudie et al. 1999), displayed high metal resistance and the ability to enhance plant growth in a metal-contaminated environment, and thus accelerate phytoremediation (Hao et al. 2012)

Another nodule-endophytic PGPR agrobacteria is the non-pathogenic and non-symbiotic *Agrobacterium* sp. 10C2 that was previously isolated from root nodules of *Phaseolus vulgaris* and was shown to be able to colonize nodules and coexist with a symbiotic and infective rhizobial strain (Mhamdi et al. 2005). *Agrobacterium* sp. 10C2 promoted nodulation, and thereby, total soluble proteins, leghaemoglobin content, biomass production and nitrogen fixation are enhanced (Chihaoui

et al. 2012). Thus, Agrobacterium could retard the nodule senescence that is determinant for enhanced nitrogen fixation and increased yields (Van de Velde et al. 2006; Mhadhbi et al. 2011). Also, the inoculation with Agrobacterium sp. 10C2 affected the richness and structure of rhizosphere bacterial communities (Chihaoui et al. 2015). The fact that strain 10C2 is a nodule endophyte probably contributed to induced drastic changes in root exudation and so influence the composition of bacterial communities in the rhizosphere, stimulating those that are known for their plant growth-promoting abilities (Broeckling et al. 2008). This strain also affects soil bacterial populations in the uncultivated soil, as well as in the bulk soil of common bean. So, regulation of soil bacterial community structure is one of the plant growth-promoting mechanisms of Agrobacterium sp. 10C2 (Chihaoui et al. 2015). Another study showed that inoculation of *Phaseolus vulgaris*, *Medicago laciniate* and *Medicago polymorpha* with an endophytic *Agrobacterium* strain may enhance nodulation and shoot dry weight (Salem et al. 2012). Other studies showed that endophytic agrobacteria could strongly solubilize phosphates and produce growth hormones (Hameed et al. 2004).

As expected, agrobacteria have evolved a wide metabolic capability and exhibit several traits to exploit soil and rhizosphere resources. Pathways for use or degradation of plant metabolites typically found in the rhizosphere were also detected in Agrobacteria that can degrade a large range of oses, polyols, and sugar derivatives often from plant origin (Marasco et al. 1995; Ampomah et al. 2013). Bacteria that inhabit diverse environments also tend to have large complements of regulatory genes (Stover et al. 2000). Consistent with this, regulatory genes constitute a substantial proportion (9%) of the A. fabrum genome. This regulatory capacity likely facilitates survival of Agrobacterium within the dynamic soil and rhizosphere environments (Wood et al. 2001).

To survive and face microbial competitors, agrobacteria are armed for example, with a potent siderophores that allow an efficient recovery of iron in iron-deprived environments. Several types of siderophores have been identified as the agrobactin (Ong et al. 1979), a hydroxamate siderophore (Penyalver et al. 2001) and a C58 specific siderophore that remains unidentified (Rondon et al. 2004; Lassalle et al. 2011). Some agrobacteria also express a type VI secretion system (T6SS) (Ryu 2015) that drives the injection of at least three effectors with enzymatic activities (two DNase and one peptidoglycan amidase) into neighboring, competing bacteria (Ma et al. 2014). These findings suggest that agrobacteria are soil- and rhizosphere-adapted bacteria.

B. *Agrobacterium* secondary ecology

Biotrophic pathogens exploit the host as a living resource. To succeed, they have evolved traits to escape host defense, to exploit host resources and to compete with resident microbiota (Spanu and Panstruga 2017). Among them, the soil-borne Gram-negative Agrobacterium has the unique ability to construct an ecological niche by engineering the host genome during an interkingdom gene transfer (Gelvin 2017). This bacterium has a remarkably broad-host range especially considering other plant pathogens (Nester 2014), since it induces tumor formation (an uncontrolled plant cell division) on most dicotyledonous and some monocotyledonous species (De Cleene and De Ley 1976). Such tumors do not require the continuous presence of the bacteria for proliferation (White and Braun 1942), showing that the plant cells have been transformed genetically. These neoplastic diseases include crown gall (Agrobacterium tumefaciens complex and Agrobacterium vitis), hairy root (Agrobacterium rhizogenes), and cane gall (Agrobacterium rubi) (Gelvin 2009). The disease, which can affect nursery and mature plants, is regarded as one of the most economically important diseases of fruit trees and ornamental plants, including almond (Prunus dulcis), grapevine (Vitis vinifera), peach (Prunus persica), plum (Prunus domestica), rose (Rosa spp.), walnut (Juglans spp.) and weeping fig (Ficus benjamina) (Kennedy and Alcorn 1980). The economic losses caused by this disease are not only related to intrinsic damage in infected plants and crop yield reduction, but also to the prohibition of commercial use of plants with tumors (Epstein et al. 2008).

Agrobacteria are the etiological agent of the plant disease crown gall (Smith and Townsend 1907; Conn 1942) since in these bacteria an unusually large plasmid was discovered and its association with gall formation demonstrated (Zaenen et al. 1974). This was followed by the discovery that a piece of the plasmid was transferred and randomly integrated into the chromosome of the plant cell (Chilton et al. 1977; Lemmers et al. 1980; Thomashow et al. 1980). Moreover, most isolates do not contain such a plasmid and are capable of living independently of a plant host (Pitzschke and Hirt 2010). So, although this bacterium is known to be phytopathogenic by inducing crown gall, it is only a secondary ecological trait related to the presence of an accessory Ti plasmid (for tumor inducing) or Ri plasmid (for hairy root-inducing) that houses the virulence genes and the factors required for tumor formation (Suzuki et al. 2009). Ti plasmids are diverse replicons that show a mosaic structure composed of conserved and highly variable regions (Otten et al. 2008). They are about 200 kb and capable of vertical and horizontal transmission. They are maintained at one or a few copies per cell (Cevallos et al. 2008). This plasmid also serves as a source for the transfer DNA (T-DNA), the DNA region that is imported into plant cells and integrated into the genomes where it is expressed, resulting in the genetic plant

manipulation (Pitzschke and Hirt 2010). Ti plasmids are composed of the following elements: Transferred DNA (T-DNA), a virulence (vir) region, opine catabolism genes, a replication (rep) region, conjugative transfer genes (tra and trb loci), and uncharacterized and cryptic accessory regions **(Figure 3)**.

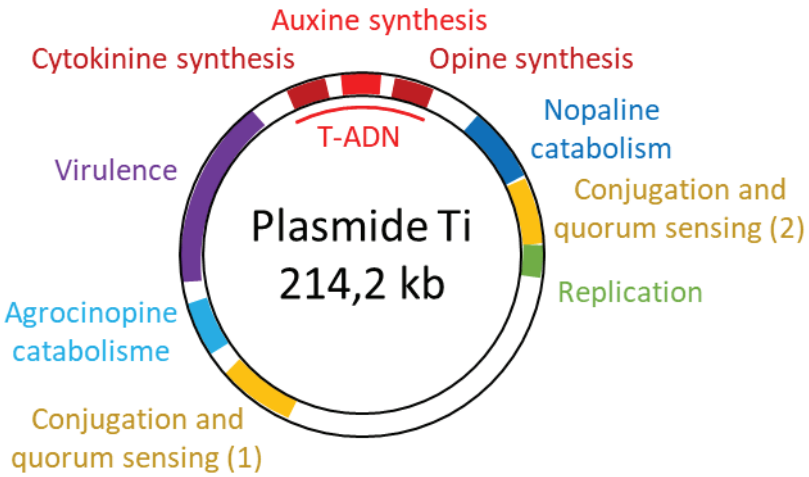


Figure 3. Agrobacterium Ti plasmid map. pTi regions involved in *A. fabrum* virulence and in plasmid dissemination (according to Quentin Duplay)

T-DNA transfer is mediated by agrobacterial virulence effectors proteins (Vir proteins), and involuntarily supported by proteins of the attacked host. The orchestration of both determines the success of transformation (Pitzschke 2013). Thus, agrobacteria abuse components of the host immunity system, mimics plant protein functions and manipulates hormone levels to bypass or override plant defenses (Pitzschke 2013). The interaction between Agrobacterium and plant cells can be divided into several steps: recognition of plant signals, virulence (Vir) genes activation, T-DNA synthesis and vir genes transcription, transport of T-DNA and vir factors by T4SS, nuclear import and T-DNA integration **(Figure 4)** (Pitzschke and Hirt 2010).

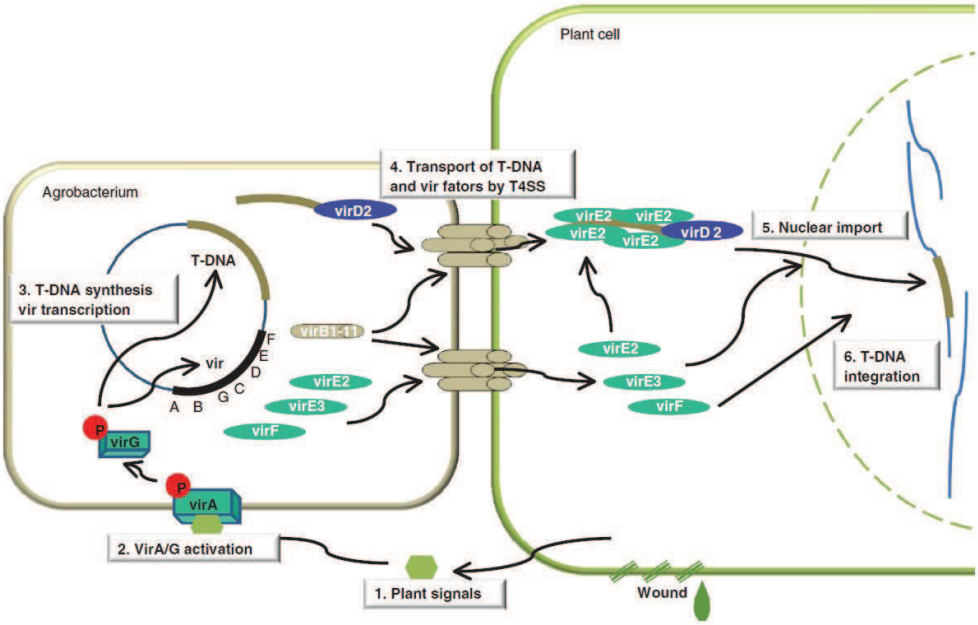


Figure 4. Overview of the Agrobacterium-plant interaction. The plant synthesizes compounds following wound **(1)**. The bacteria recognize plant compounds by the VirA/G system **(2)**. T-DNA synthesis and vir gene expression in *Agrobacterium* **(3)**. Transport of T-DNA to the plant cell **(4)** and then to the host cell nucleus **(5)** where it will be integrated into the chromosome **(6)**. (Pitzschke and Hirt 2010)

Agrobacterium attacks mainly wounded tissue (Braun 1952) that releases many chemical signals such as phenols, sugars and pH diminution, that induce vir gene expression and act also as chemotactic attractants of agrobacteria (Gelvin 2012). Subsequently, Vir proteins are produced, and single-stranded T-DNA molecules are synthesized (Alvarez-Martinez and Christie 2009). To export and inject both to the host cytoplams, a bacterial transfer machinery is subsequently produced and assembled, the type IV secretion system (T4SS) carried by the pTi and composed of 11 VirB proteins and VirD4 (Cascales and Christie 2003; Christie et al. 2005). An intimate association between pathogen and host cell is thus stablished. Quantitative-binding assays have revealed a non-specific interaction that is readily removed and a specific interaction (Neff and Binns 1985).

Once inside the plant cell, T-strands likely form complexes with other secreted vir proteins and supercomplexes with plant proteins as they traverse the cytoplasm and target its final destination, the host cell's nucleus **(Figure 5. step 1)** (Gelvin 2009). Once inside the nucleus, T-strands integrate randomly into the plant genome and express T-DNA-encoded transgenes **(Figure 5. step 2)**. Two classes of T-DNA genes mediate the pathology of Agrobacterium infection. The first group, the oncogenes, either effect phytohormone production such as auxin and cytokinin (Akiyoshi et al. 1984; Schröder et al. 1984) or sensitize the plant to endogenous hormone levels (Shen et al. 1988). This effect on

endogenous growth regulators will ultimately lead to an anarchic multiplication of plant cells, causing the agrobacterium-induced tumor colonized by the pathogen **(Figure 5. step 3)** (Gohlke and Deeken 2014).

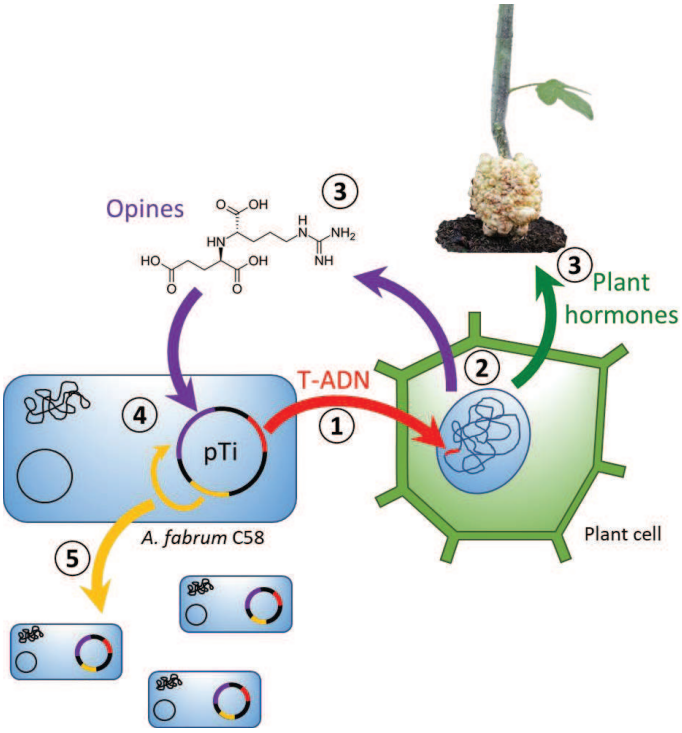


Figure 5. Virulence establishment during development of crown-gall induced by *A. fabrum*. 1. Transfer of T-DNA into the plant cell. **2.** Synthesis of hormones and opines by the plant cell. **3.** Uncontrolled proliferation of plant cells and use of opines by agrobacteria as a source of carbon and nitrogen. **4.** Activation of the quorum sensing system and of the replication and conjugation genes. **5.** Transfer of pTi to other agrobacteria which in turn become virulent. (inspired in Quentin Duplay)

A second set of genes directs the synthesis of peculiar metabolites, called opines **(Figure 5. step 3)**, that are compounds of low molecular weight resulting from the condensation of plant sugars, organic acids and amino acids (Dessaux et al. 1993). Opines fulfill two functions. Firstly, as opines can serve as specific carbon and sometimes nitrogen sources for the inciting agrobacteria **(Figure 5. step 4)** to the exclusion of most other microorganisms, they provide a selective advantage for this species (Tempé and Petit 1982; Veluthambi et al. 1989). This is call “the opine concept” which proposes that a parasitic agent may incite opine synthesis in its host to create a chemical environment favorable for growth and propagation of the pathogenic agent (Petit et al. 1983). The genes involved in opines catabolism are activated only in the presence of opines, via regulatory proteins (Kim et al 2008; Marines and White 1993). In *A. fabrum* C58 two types of opines are produced by transformed plant cells: nopaline and agrocinopine.

Secondly, opines are the triggering signal of the cellular communication or quorum sensing (QS), which controls the conjugative transfer of the Ti plasmid **(Figure 5. step 5)** (Lang and Faure 2014). The Ti plasmids are able to transfer by conjugation through a second T4SS, via the expression of tra-trb genes. They are transmissible to numerous *Rhizobiaceae* and more particularly to the genus

Agrobacterium (Hooykaas et al. 1977). The conjugative system of *A. fabrum* C58 pTi is regulated by TraI and TraR, homologous of the LuxI-LuxR system. The AHL synthase TraI allows the synthesis of N-3-oxo-octanoyl-L-homoserine lactone (OC_8HSL), the QS signal molecule. When the cell density is high enough, OC_8HSL reaches a critical concentration and can then bind to the transcriptional regulator TraR. As a homodimer, the complex TraR-OC_8HSL is active and will then regulate the different target genes by specifically binding to a region called "Tra boxes." These Tra boxes are consensus nucleotide regions present in the promoter region of the genes regulated by this QS system. This will increase the synthesis of TraI, resulting in a positive feedback loop (Zhang et al 1993). This conjugative system is locked by *accR*, the repressor of TraR. In the presence of conjugative opines such as agrocinopine, the repression exerted by *accR* is lifted. Other complementary mechanisms allow a strict control of the conjugation of pTi C58. So, in the presence of conjugative opine, the pTi is thus transferred to other agrobacteria which in turn become pathogenic and will subsequently contribute to the maintenance or spread of the disease (Kerr 1969).

II. The opine concept

Niche construction designates the process by which a living population modifies its environment and takes advantage of the induced environmental changes (Kylafis and Loreau 2011). Members of the genus *Agrobacterium* harboring Ti plasmids have evolved a unique interaction with certain plant species with their remarkable ability to generate specific nutrient niches in its host plants through the formation of a tumor (Chilton et al. 1977). Opines are compounds synthesized in crown-gall incited by phytopathogenic agrobacteria (Petit and Tempé 1978; Guyon et al. 1980). The causative bacteria can use opines as a source of carbon, nitrogen, and energy (Petit et al. 1983). Thus, *Agrobacterium* redirects plant cell metabolism to produce specific metabolites which the bacterium can use as growth substrates. Only agrobacteria containing a Ti-plasmid can catabolize opines, which are thus key players in *Agrobacterium* niche construction (Montoya et al. 1977). Opines are also important in the ecology of agrobacteria because some of these molecules induce the conjugative transfer of the pathogenic plasmid to saprophytic *Agrobacterium*. Thus, the crown gall tumor can be described as an ecological niche favoring multiplication of the pathogen and promoting dissemination of pathogenicity. This description of the biological role of opines is known as the opine concept (Dessaux et al. 1993).

Around 40 different types of opines have been structurally characterized (Dessaux et al. 1998) and each pTi allows the synthesis and catabolism of a few opines. So, the type of opines synthesized in the tumor depend on the particular Ti plasmid harbored by the inciting Agrobacterium strain (Ellis and Murphy 1981). Thus, only a subset of opines molecules can be detected at the same time in a given tumor. For the moment, the Ti plasmids have been classified into different types according to the main opine induced (Dessaux et al. 1988).

Chemically, opines are diverse molecules that fall into two structural classes. The first group is made up of secondary amine derivatives from the condensation of an amino acid, either with a keto acid or a sugar **(Figure 6)**. It contains the vast majority of opines such as octopine, nopaline, agropine and their derivatives. All of these are grouped in different families. The second group is made up of sugar phosphodiesters, called agrocinopines **(Figure 7)** (Moore et al. 1997; Brencic and Winans 2005).

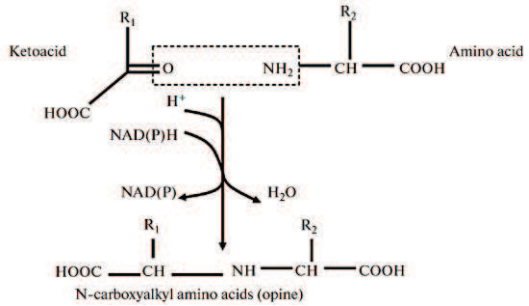


Figure 6. Opine biosynthesis. N-carboxyl acyl amino acids are formed by reductive condensation of amino acid and ketoacid; a reaction occurs between the amino group of an amino acid and a ketone group of the ketoacid. (Vladimirov 2014).

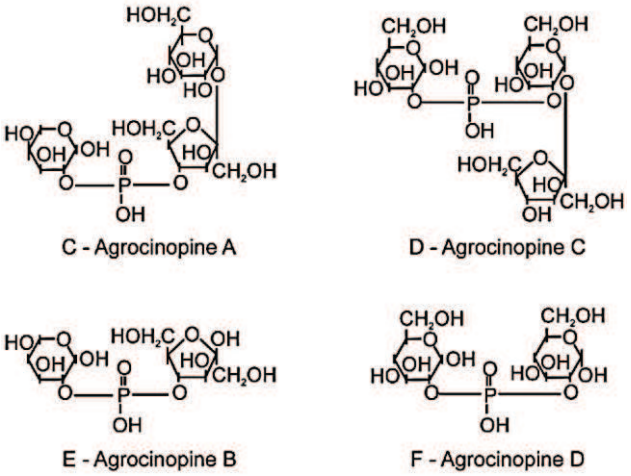


Figure 7. Structures of agrocinopines. (Anja Brencic 2005).

Most opine synthases are conjugases, allowing opine formation. Opine biosynthesis genes are well known for the high efficiency transcription promoters and terminators and some of them allow the formation of several opines thanks to low specificity for their substrates (Vladimirov et al. 2015). Thus, depending on the substrates available in transformed tumor cells, the genes encoding synthases of opines can produce different compounds (Flores-Mireles et al. 2012). Opines can be synthesized in a single stage so, one enzyme encoded by a single gene is required for their synthesis. Other opines however, require a multistage biosynthesis pathway. Some opines, depending on the physicochemical conditions of the plant cell, can be lactamized, such as nopaline, which spontaneously converts to pyronopaline at acid pH (Hall et al. 1983). Agrobacteria plasmids usually possess several opine genes, and they provide synthesis of several different opines (Vladimirov et al. 2015).

One of the families of opines in *Agrobacterium*, is the nopaline family **(Figure 8B)** composed by nopaline and nopalinic acid (Hall et al. 1983). This family is formed when alpha-ketoglutarate serves as the keto substrate in the condensation reaction. For example, nopaline and nopalinic acid come from the condensation of an α-ketoglutarate and respectively, an arginine or an ornithine. These two opines are synthesized in crown gall tumours by an enzyme encoded by the *nos* gene (nopaline synthase) of the T-DNA **(Table 2)** (Holsters et al. 1980).

Another family, the octopine family **(Figure 8A)**, comprises octopine (which is a conjugative opine), octopinic acid, lysopine and histopine (Biemann et al. 1960; Kemp 1977). They come from the condensation of pyruvate and, respectively, arginine, lysine, ornithine and histidine, catalyzed by the same enzyme encoded by the *ocs* gene (octopine synthase) **(Table 2)** (Dessaux et al. 1998; Flores-Mireles et al. 2012).

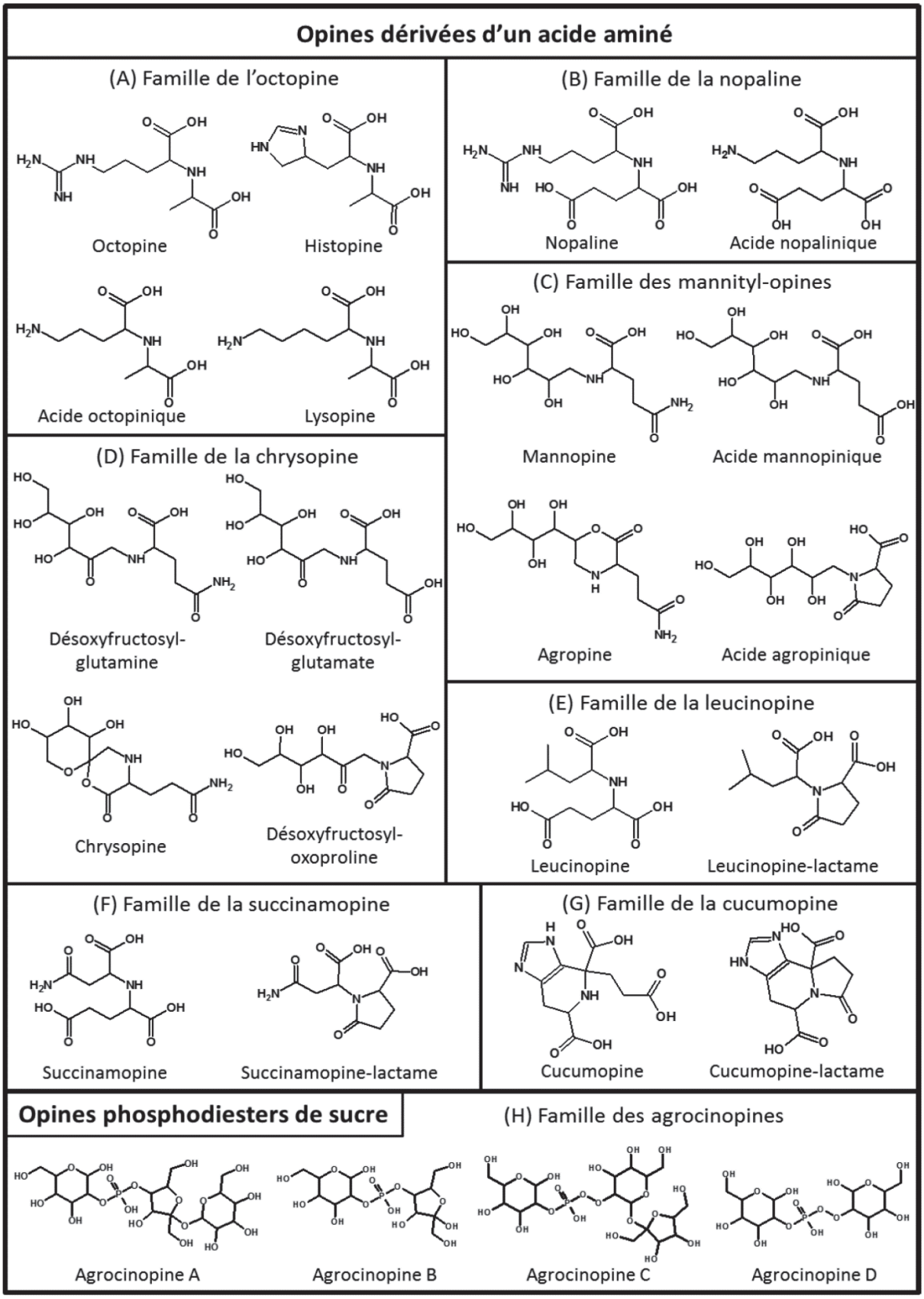


Figure 8. Chemical structure of opines induced by the genus *Agrobacterium.*
(According to Loïc Marty).

The chrysopine family **(Figure 8D)** and the mannityl family **(Figure 8C)** are formed by the condensation of an amino acid with mannose. The first one consists of five members: deoxy-fructosyl-glutamine (DFG) or santhopine, deoxy-fructosyl-glutamate (DFGA), chrysopine, chrysopinic acid and deoxy-fructosyl-oxoproline (DFOP) (Chilton et al. 1995). DFG is produced by the condensation of fructose and glutamine by a conjugase and the other opines in the chrysopin family are its derivatives. DFGA comes from an enzymatic or spontaneous condensation of glutamate with fructose, chrysopine is the product of the lactonization of DFG and DFOP is a lactamized form of DFG (Chilton et al. 1995). Chrysopine strains induce the synthesis of DFG and chrysopine in the tumor. They contain a common set of *mcts2-mas1-ags* genes **(Table 1)**. The chemical structures of these compounds are very close to the mannityl-opine family. The latter consists of four members: mannopine, agropine, mannopinic and agropinic acids **(Figure 8C)** (Dessaux et al. 1986). They are also known as the agropine family and were first detected in octopine tumours (Coxon et al. 1980). DPG is an intermediate product of mannopine and agropine synthesis. Agropine is a lactonized form of mannopine while agropinic acid is a lactamized form of mannopinic acid. Mannopine plasmids possess only two opine synthesis genes: *masl* and *mas2* (mannopine synthase) **(Table 1)** responsible for the production of mannopine and mannopinic acid. Agropine plasmids possess the *mas1* and *mas2* genes and the *ags* gene (agropine synthase) **(Table 1)**. The latter codes for a protein involved in the cyclization of mannopine to agropine. All three genes form a single pathway providing agropine synthesis from glutamine and glucose. Agropinic acid is spontaneously formed from agropine (Marty et al. 2019).

Table 1. Opine genes of pTi and pRi plasmids (Vladimirov 2014).

Plasmid type	Opine biosynthesis genes	Opines, synthesized by the tumor
Nopaline	*nos* (nopaline synthase), *acs* (agrocinopine synthase)	Nopaline, nopaline acid, agrocinopines A and B
Octopine	T_L: *ocs* (octopine synthase), T_R: *mas1*, *mas2* (mannopine synthases), *ags* (agropine sunthase)	Mannopine, mannopinic acid; agropine, agropinic acid; octopine, octopinic acid, lysopine, histopine, lysopinic acid
Succinamopine	T_L: *sus* (succinamopine synthase), T_R: *mas1*, *mas2* (mannopine synthases), *ags* (agropine sunthase)	Mannopine, mannopinic acid; agropine, agropinic acid; succinamopine, succinamopine- lactam
Chrysopine	T_L: *mas1*, *mas2* (mannopine synthases), *ags* (agropine sunthase)	DFG deoxy-fructosyl glutamate, chrysopine (DFG lacton)
Agropine	T_L: a*cs* (agrocinopine synthase), T_R: *mas1*, *mas2* (mannopine synthases), *ags* (agropine sunthase)	Mannopine, mannopinic acid; agropine, agropinic acid; agrocinopines C and D
Mannopine	*mas1*, *mas2* (mannopine synthases)	Mannopine, mannopinic acid
Mikimopine	*mis* (mikimopine synthase)	Mikimopine, mikimopine-lactam
Cucumopine	cus (cucumopine synthase)	Cucumopine, cucumopine-lactam

There are other families of opines **(Figure 8E, F, G)** and a vast variety of opines have been identified, including succinamopine (Chilton et al. 1984), leucinopine (Chang and Chen 1983), cucumopine (Davioud et al. 1988), mikimopine (Isogai et al. 1990) and the related lactams. Also, heliopine (Chilton et al. 2001), vitopine (whose inducing plasmids have exclusively been found in *Allobacterium vitis* (grapevine) isolates) (Faist et al. 2016), pseudo-nopaline (Chilton et al. 1995), and opine X (Chilton et al. 1984).

Agrocinopines **(Figure 8H)**, the conjugative opines, form a special family of opines because chemically they are not classic opines (Ellis and Murphy 1981). However, they have the same function and are therefore traditionally considered together with opines. These sugar phosphodiesters differ from each other by the nature of the phosphorylated monosaccharide and by the nature of the phosphate bond between the phosphate and fructose group (Kim et al. 2008). Two of these compounds, agrocinopines A and B, are phosphodiesters of arabinose and, respectively, sucrose or fructose (Scott et al. 1979). These two molecules are the only opines which contain a phosphorus atom being the only nitrogen-free opines, and their production is due to the product of the *acs* gene (agrocinopine synthase) **(Table 1)** (Dessaux et al. 1998). The pyranose-2-phosphate motif shared by the four agrocinopines is the motif recognized as a conjugative inducer (El Sahili et al. 2015).

Opine catabolism genes control the uptake of opines and their utilization. These genes are organized in operons located on the non-transferred part of the Ti plasmid (Kuzmanović and Puławska 2019) (Table 4). Transport of opines into the cytoplasm of agrobacteria is carried out by a tripartite system composed of a periplasmic binding protein (PBP), a transmembrane domain (TMD) and a nucleotide binding domain (NBD) (Lang et al. 2014). Opine catabolism is typically specific to the opine or opines produced in the induced tumor (Montoya et al. 1977). This nutritional specificity generally means that the inciting strain of Agrobacterium catabolizes only those opines produced by the incited tumor. For many opines, the enzyme in question is an opine oxidase, catalyzing the decondensation of the opine into amino acid and sugar or ketoacid (Dessaux et al. 1988).

The genes for the biosynthesis and catabolism of opines usually are not homologous, even if they catalyze the same reaction in the opposite directions. Sometimes, a chain of reactions is necessary for their degradation, as for their formation. For example, neither octopine nor nopaline oxidase is homologous for their biosynthetic genes (Zanker et al. 1994). The degradation of nopaline implicates several genes to uptake and cleave back nopaline to arginine and a-ketoglutarate (Montoya et al. 1977). Octopine degradation is determined by the octopine-inducible *occ* operon, which encodes genes involved in opine transport and breakdown (Montoya et al. 1978) and arc genes responsible for

arginine and ornithine degradation (Ellis et al. 1979). However, agropine and mannopine biosynthesis and catabolism are performed in opposite directions by the same pathway (Fig. 4), and genes of the biosynthesis and catabolism enzymes, which work at each stage, are pairwise homologous (Dessaux et al. 1987). The *ags* gene is homologous to the catabolic *agcA* gene (Hong et al. 1997), *mas1* is homologous to *mocC*, and *mas2* is homologous to *mocD* (Kim and Farrand 1996). In the case of agropinic acid, even if there are no genes involved in its synthesis, its degradation requires specific genes located within the mannityl opine catabolic region (Marty et al. 2019). Agrocinopines have a particular degradation pathway due to their nature. They involve a phosphodiesterase, which will cleave the phosphodiester bond between the two sugars, then a phosphatase, making it possible to dephosphorylate the sugar-phosphate (Kim et al. 2008).

Quantitative determination of opines is performed by several different methods, such as high-voltage paper electrophoresis (HVPE) (Savka and Farrand 1992), gas-liquid chromatography (Scott 1979), high-performance liquid chromatography (HPLC) (Firmin 1990), LC-QTOF analysis (Venter et al. 2017), ion exchange (Dessaux et al. 1993), enzymatic assay (Sato et al. 1988), isotachophoresis (Sandee et al. 1996) and reversed-phase high-performance liquid chromatography (RP-HPLC) (Zhang et al. 1998). Each assay method has unique features and although each of these methods has certain advantages, they also have disadvantages. Some methods are elaborate, others lack sensitivity or require a complicated derivatization (Sandee et al. 1996).

For example, HVPE is followed by chemical staining, using ninhydrin for opines with primary amines (lysopine, octopinic acid, nopalinic acid, rideopine) (Chilton et al. 2001), phenanthrenequinone for opines with guanidino groups (octopine, nopaline, agrocinopine) (Hernalsteens et al. 1984), alkaline silver nitrate for opines with a-diols (agropine, agropinic acid, mannopine, and mannopinic acid, vitopine) (Moore et al. 1997), or Pauly reagent (mikimopine, cucumopine) (Suzuki et al. 2001). The main disadvantages of these detection methods are the requirement of a partial purification of the extract, the inability to accurately quantify the compounds, and the limited sensitivity (Sato et al. 1988). The detection limit of mannityl opines with this method is approximately 1 mg/spot, the detection limit of vitopine is about 5µg (Szegedi 2003), while the detection limit of octopine and nopaline is approximately 2 to 5 mg/ spot (Zhang et al. 1998). The use of Sakaguchi's reagent enabled detection of octopine and nopaline down to levels as low as 1 µg (Dommisse et al. 1990). Later, the use of a fluorescent stain specific for guanidine groups allowed a sensitivity 1000 times more than that of Sakaguchi reagent and 700 times more than that of the pentacyanoaquo ferriate reagent, detecting as little as 0.085µg of octopine (Johnson et al. 1974).

The enzymatic assay has the advantage of monitoring the extract directly, but it requires a long assay period and does not distinguish between the acidic opines. The isotachophoresis method also has a low sensitivity (Sato et al. 1988). The gas-liquid chromatography method requires a lengthy pretreatment and analysis time. The HPLC method provides high sensitivity and prompt determination; however, identification of each peak in the chromatogram must be carefully considered because many compounds appear to elute in the area of the acidic opines (Sato et al. 1988). The detection limit with this method is about 0.05µg for cucumopine and mikimopine and at 0.02µg for agropine and mannopine (Sauerwein and Wink 1993). RP-HPLC with fluorescence detection were developed for the separation, detection and quantification of imine-linked opines, following 4-fluoro-7-nitrobenzoxadiazole (NBD-F) derivatization a sensitive reagent which reacts with both primary and secondary amines. All but agropine react with NBD-F and the sensitivity of detection range from 0.1 pmol for octopinic acid to 5 pmol for cucumopine (Zhang et al. 1998).

III. Concept of bacterial species and *A. fabrum*-specific genes

Everything that changes over time has, by definition, a history (Ernst Mayr 1982). But our understanding of bacterial history and ecology is fragmentary. Comparing genomes reveals historical signals that can be used to retrace genome evolution, by estimating their hypothetical ancestral state and the course of the evolutionary events that shaped them over time (Lassalle et al. 2017).

The species as basic taxonomic unit dates back to Carl Linnaeus and has since been universally used to describe all living organisms, including microbes. The classical Biological Species Concept (BSC) (Mayr 1999), which was originally defined for animals, places the sexual isolation of clades as the central condition for their divergence and therefore speciation. However, in asexually reproducing organisms, species are defined upon similarities of their members contrasted by interspecies genetic discontinuities (Lassalle et al. 2011).

In Bacteria, similarity discontinuities were first revealed through phenotypic traits and used to classify strains in different species by numerical taxonomy (Sneath and Sokal 1973). It was soon discovered that discontinuities also occur at the genomic level, leading to the current genomic species definition (Lassalle et al. 2011). So, a bacterial species is currently defined by their genomic homogeneity, leading to the conventional division of the bacterial world into genomically homogeneous units, genomic species, that attempt to reflect the natural occurrence of clusters of diversity (Moore et al. 1987; Lassalle et al. 2015). The current consensus for bacterial species identification (Wayne 1988) is based on DNA-DNA hybridization of whole genomes and indicates approximately 70% or greater DNA-DNA relatedness and with 5°C or less ΔTm" (Portier et al. 2006). Although this definition is operational, we still need to understand what mechanisms lead to differentiation of such genomic species (Fraser et al. 2009).

The term genomic species is applied to bacteria that are genetically so closely related that they can be classified as one species yet possess enough infra-species diversity to cluster into discrete groups of strains (Lamovsek et al. 2014). A genomic species descends from a single ancestor that speciated a long time ago as a result of adaptations to a new ecological niche. These adaptive mutations should have been conserved in the progeny as long as they continued to exploit the same primary niche (Lassalle et al. 2011). They have as a consequence clade-specific genes that likely provide a strong adaptive feature to a particular ecological niche (Lassalle et al. 2015). Niche-specifying traits are expected to provide higher differential fitness as they are less likely to be already present in

competing relatives and thus enable bacteria to escape competition. Consequently, it is crucial for these novel and complex traits to be gained at once for it to provide any kind of advantage, as it is in the case of the transfer of a complete operon, since clusters of genes participating to a same pathway are more likely to carry sufficient information to encode a new adaptive trait with a coherent biochemical function (Lassalle et al. 2017). The unique ecological properties of each bacterial clade should be revealed by identifying the genes specifically conserved inside a clade, and notably those grouped in clusters with related functions using comparative genomics (Lassalle et al. 2017).

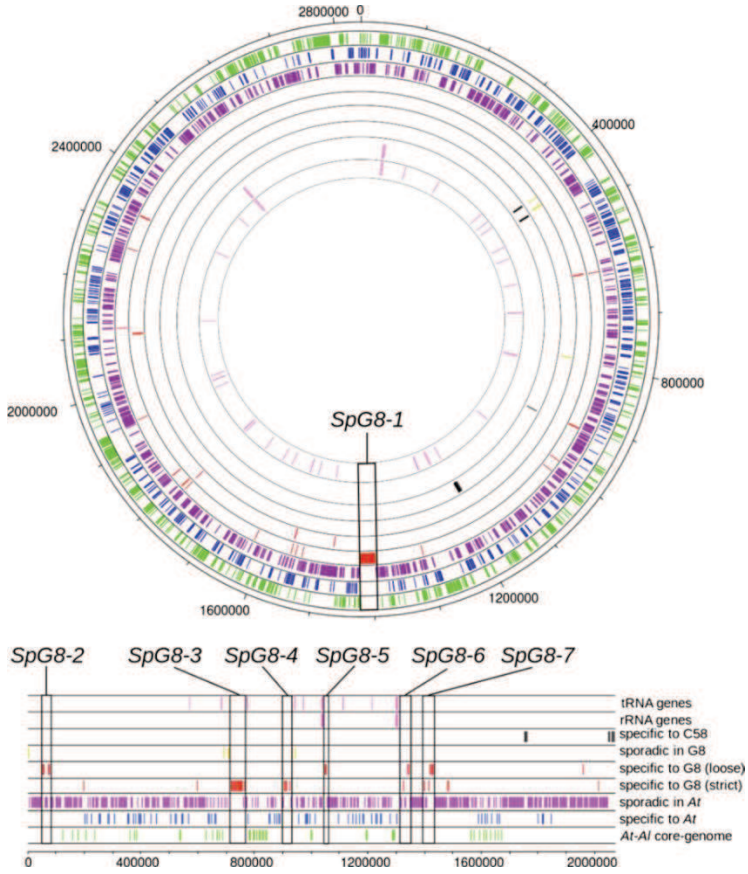
To prove this hypothesis researchers used the *Agrobacterium* genus or "*Agrobacterium tumefaciens* (At) complex" that displays high within-species diversity (Mougel et al. 2002; Portier et al. 2006), so as to capture the most common species characters, with the least possible divergence from their closest neighbors, thus maximizing the chance of detecting specific determinants (Lassalle et al. 2011). Agrobacteria are common inhabitants of soils and rhizospheres, with several strains and genomic species commonly found in the same soil samples (Vogel et al. 2003; Costechareyre et al. 2010). According to the competitive exclusion principle (Hardin 1960, Gause 1932), co-occurring species must be adapted to partly different ecological niches. Clade-specific blocks of cotransferred genes encoding coherent biochemical pathways were identified and recurrently presented the same classes of functions. These include transport and metabolism of phenolic compounds, aminoacids and complex sugars, and production of exopolysaccharides and siderophores, all of which can be related to bacterial life in the plant rhizosphere. This suggests that ecological diversification of Agrobacterium occurred through the partitioning of ecological resources available in plant rhizospheres (Lassalle et al. 2017). That is, specific genomes probably led each species to tap the same resources in different ways, avoiding any significant competition between them (Lassalle et al. 2015).

For example, G1 strains ant the [G5–G13] species group are able to specifically degrade aromatic compounds likely to be found in plant rhizospheres. Also, G8 species and the [G6–G8] clade present the ferulic acid degradation and siderophore biosynthesis operons that have been reported to provide a growth advantage with respect to cooccurring agrobacteria and to be expressed in a coordinated manner in a plant rhizosphere environment (Campillo et al. 2014; Baude et al. 2016). As well, the cluster encoding the nitrate respiration (denitrification) pathway acquired by the ancestor of the [G2–G4–G7–G9–G6–G8] clade, one of the two large clades that divide the At complex. Such an adaptation may have supported an early differentiation of At lineages towards the colonization of partitioned niches with anaerobic or micro-aerophilic conditions. Finally, the G1 and G8 species are thus likely to orchestrate the production of similar polysaccharides under different regulation schemes, as they both

have genes involved in the biosynthesis of curdlan—a cellulose-like polysaccharide—and the biosynthesis of O antigens of the lipopolysaccharide (LPS) (Lassalle et al. 2017).

In this work, we are focusing on the specific genes to the species G8, named *A. fabrum*, also gathered in clusters encoding coherent biological functions(Lassalle et al. 2011). *A. fabrum* has a total of 196 G8-specific genes (SpG8) that are unevenly dispersed on the two chromosomes. This SpG8 genes are mostly clustered into seven genomic islands on the C58 genome, numbered SpG8-1 to SpG8-7, one is located on the circular chromosome (SpG8-1) and the other six on the linear chromosome **(Figure 9)** (Lassalle et al. 2011). This suggest higher plasticity and a major adaptive role of the linear chromosome. Some SpG8 clusters were subsequently divided into sub-clusters encoding homogeneous functions. SpG8 functions seemed to collectively define an ecological niche of G8 agrobacteria related to commensal interactions with plants. Indeed, SpG8 clusters encoded functional units related to environ mental sensing (SpG8-7), secreted metabolite production (SpG8-2a, SpG8-3), detoxification (SpG8-6), and numerous catabolic pathways of carbohydrates, namely ferulic acid (SpG8-1b), diverse sugars (SpG8-1a, SpG8-4), amino acids (SpG8-1a, SpG8-5), and opine-like/Amadori compounds (SpG8-5). All these last ones are typical molecules that can be found in plant rhizospheres, exuded by plants or derived from plant degradation products. Also, phenolic compounds are of primary importance in the biology of G8 agrobacteria, being both metabolites and signals released by the host plant. As a result, the combination of G8-specific functions defines a hypothetical species primary niche for G8 related to commensal interaction with a host plant. This supports that the G8 ancestor was able to exploit a new ecological niche, maybe initiating ecological isolation and thus speciation (Lassalle et al. 2011).

.



G8-Specific Regions	C58 CDSs	Region Occurrence Outside G8	Main Predicted Functions
SpG8-1a	Atu1398–Atu1408	G6-NCCPB 925	Sugar and amino acid transport; sugar metabolism
SpG8-1b	Atu1409–Atu1423	G9-Hayward 0362	Ferulic acid uptake and catabolism
SpG8-2a	Atu3054–Atu3059	*r*	Curdlan EPS biosynthesis
SpG8-2b	Atu3069–Atu3073	*r*	Secondary metabolite biosynthesis
SpG8-3	Atu3663–Atu3691	G1-ICPPB TT111 G6-NCCPB 925	Siderophore biosynthesis; iron-siderophore uptake
SpG8-4	Atu3808–Atu3830	G6-NCCPB 925	Ribose transport; monosaccharide catabolism and carbohydrate metabolism
SpG8-5	Atu3947–Atu3952	*r*	Opine-like compounds catabolism
SpG8-6a	Atu4196–Atu4206	G1-CFBP 5771	Drug/toxic (tetracycline) resistance
SpG8-6b	Atu4213–Atu4221	*r*	Drug/toxic resistance
SpG8-7a	Atu4285–Atu4294	G6-NCCPB 925	Environmental signal sensing/transduction
SpG8-7b	Atu4295–Atu4307	Not present outside G8	Environmental signal sensing/transduction

NOTE.—*r*, rare occurrence of some CDSs outside G8.

Figure 9. *A. fabrum* C58-specific regions. A. Representation of the circular and linear chromosome of *A. fabrum* with the location of specific regions. **B.** Characteristics of specific regions (according to Lassalle *et al.* 2011).

The next section details quickly the most important aspects of the seven specific regions.

SpG-7 region: Environmental signal sensing / transduction

The annotation of this region **(Figure 10)** suggests its involvement in functions related to the detection of environmental signal sensing and transduction. Two genes (Atu4300 and Atu4305) encoding a two-component system and belonging to this region are homologous to genes in *Bradyrhizobium japonicum* whose proteins are involved in the recognition of the host plant during the nodulation process (Lang et al. 2008). These genes are also homologous to genes in *Pseudomonas putida* whose proteins recognize toluene and styrene, activating degradation pathways (Lau et al. 1997). Lassalle et al. 2011 hypothesized that this region is involved in the perception of signals from the environment that may be responsible for activation of other functions, including, perhaps, SpG8 functions such as phenolic metabolism

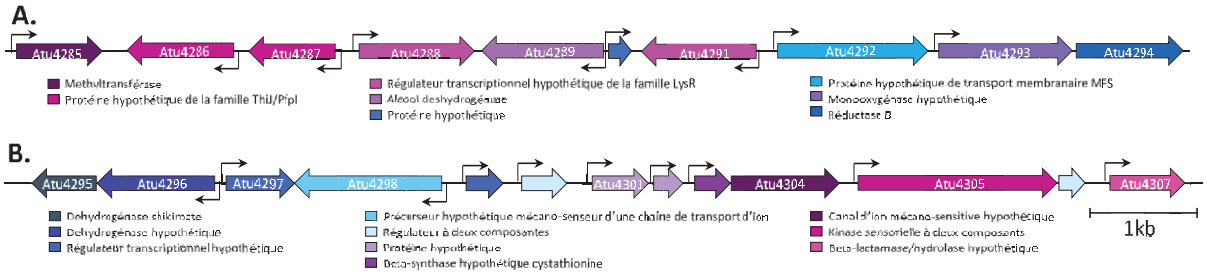


Figure 10. Organization of the SpG8-7 of *A. fabrum* C58 with their hypothetical functions. A. SpG8-7a. **B.** SpG8-7b. Promoters are shown with small arrows indicating the direction of gene transcription.

SpG8-6 region: Detoxification

This region **(Figure 11)** contains three hypothetical multi-drug transporter systems, including one that has been experimentally characterized for tetracycline resistance only in G8 species (Luo and Farrand 1999). It should be noted, however, that since tetracycline is not the natural inducer of these genes, the efflux system could allow the detoxification of other compounds that are still unknown (Luo and Farrand 1999). This region is still a putative defense mechanism provided by SpG8 loci by action of multidrug exporters.

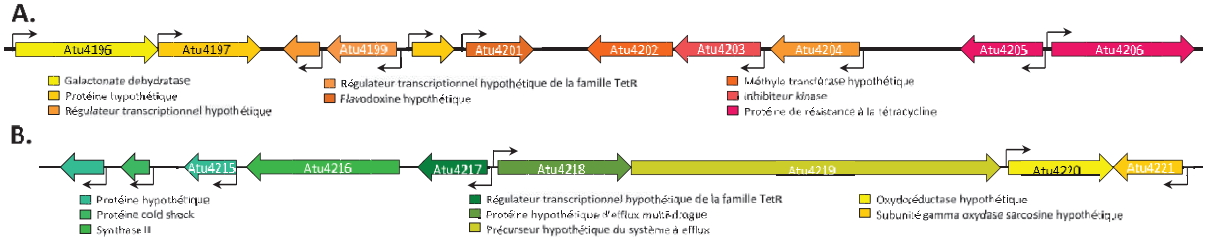


Figure 11. Organization of the SpG8-6 of *A. fabrum* C58 with their hypothetical functions. A. SpG8-6a. **B.** SpG8-6b. Promoters are shown with small arrows indicating the direction of gene transcription.

SpG8-5 region: Opine-like compounds catabolism

According to the annotations of Lassalle et al. 2011, this region **(Figure 12)** is thought to be involved in the catabolism of opine-like compounds, compounds specifically synthesized in tumors. In fact, this region encodes several enzymes, in particular an alanine racemase with a lectin-type binding domain which could be involved in the catabolism of opine-like compounds. Opines are phosphodiesters of sugars or condensates of an amino acid and a sugar or a ketonic acid. It is known that the presence of opines in the tumor provides an ecological niche favorable to the growth of pathogenic agrobacteria (Vaudequin-Dransart et al. 1995) thanks to the opine catabolism genes localized on the Ti plasmid. It is therefore remarkable that a chromosomal region such as SpG8-5 (and therefore present in agrobacteria lacking pTi) can be used for the exploitation of compounds related to those found almost exclusively in tumors.

The annotation of this region is only indicative and the substrate concerned could be a completely different group of amino acid condensates and / or sugars unrelated to opines, for example the Amadori compounds (Lassalle et al. 2011). Nevertheless, it could be a case of pre-adaptation, an exaptation, to the catabolism of the "real" opines synthesized in tumors.

Figure 12. Organization of the SpG8-5 of *A. fabrum* C58 with their hypothetical functions. Promoters are shown with small arrows indicating the direction of gene transcription.

SpG8-4 region: Transport and sugar metabolism

This region **(Figure 13)** is involved in the catabolism of carbohydrates. Indeed, this region seems to have a functional unit dedicated to monosaccharide uptake via a hypothetical ribose specific ABC transporter. Similarly, it has hypothetical enzymatic functions such as rhamnose mutarotase or D-galactarate dehydrogenase, involved in the metabolism of sugars. This region also has four LysR-like transcriptional regulators that may be involved in substrate-dependent regulation of certain metabolic pathways (Maddocks and Oyston 2008; Lassalle et al. 2011).

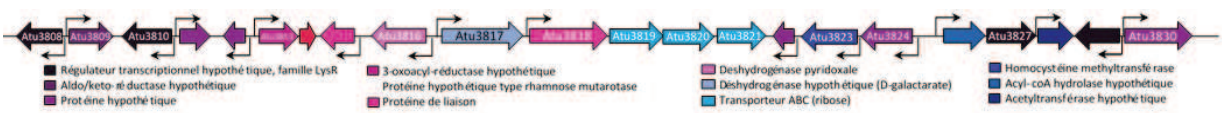


Figure 13. Organization of the SpG8-4 of *A. fabrum* C58 with their hypothetical functions. Promoters are shown with small arrows indicating the direction of gene transcription.

SpG8-3 region: Siderophore biosynthesis

This gene cluster **(Figure 14)**, the largest of the specific regions, has been shown to encode a wide range of enzymes involved in the biosynthesis of a siderophore required for *A. fabrum* C58 growth under iron limiting conditions, suggesting that this strain is highly efficient in iron scavenging (Rondon et al. 2004). This genomic region include genes for transporter proteins of the family MATE (Moriyama et al. 2008) perhaps involved in siderophore release. Also, genes involved in TonB-dependent reuptake of the siderophore when it is chelated to iron (Braun et al. 2006). And finally, genes that seem to form a cell surface signaling system proposed to regulate the whole system of biosynthesis, release, and reuptake of the siderophore (Braun et al. 2006; Lassalle et al. 2011). The siderophore produced by *A. fabrum* C58 has a unique chemical structure as the genes in this cluster do not show a strong similarity with known siderophore biosynthetic genes. This lack of similarity suggests that a novel siderophore is produced by *A. fabrum* C58 (Rondon et al. 2004). Thus, this siderophore biosynthesis locus might provide another general fitness gain in competition with other bacteria present in the biotope. Scavenging of limiting resources like iron is known to be a very potent means to outperform competitors, especially in habitats like rhizospheres with dense and diverse populations (Lassalle et al. 2011).

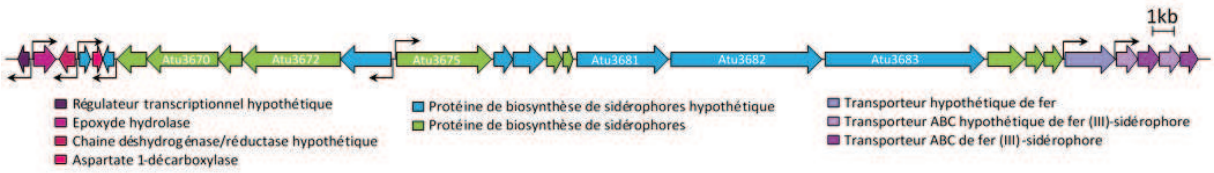


Figure 14. Organization of the SpG8-3 of *A. fabrum* C58 with their hypothetical functions. Promoters are shown with small arrows indicating the direction of gene transcription.

SpG8-2 region: Curdlan EPS biosynthesis and Secondary metabolite biosynthesis

This specific region is divided into 2 parts **(Figure 15)**. The first part called SpG8-2a **(Figure 15A)**is involved in the biosynthesis of an exopolysaccharide called curdlan, a water insoluble (1⟶3) β-glucan (Harada and Harada 1996). Curdlan is a linear glucose polymer with 400 to 500 residues with a molecular weight of 60kDa to 300kDa (McIntosh et al. 2005). This polysaccharide is obtained in good yields in an insoluble, microfibrillar form when *Agrobacterium* is grown on glucose-rich medium, and its synthesis is induced under stress conditions, such as depletion of the nitrogen source and a low pH (Phillips and Lawford 1983; Kim et al. 1999; Yu et al. 2015).

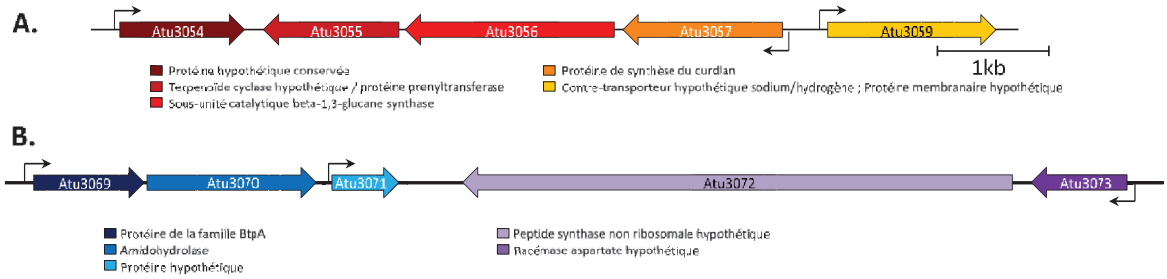


Figure 15. Organization of the SpG8-2 of *A. fabrum* C58 with their hypothetical functions. A. SpG8-2a. **B.** SpG8-2b. Promoters are shown with small arrows indicating the direction of gene transcription.

Curdlan is used in a wide-range of applications including the production of superworkable concrete (Sakamoto et al. 1991), as a gelling agent and a food texture modifier in the food industry (Harada et al. 1993), as a main component of potential anti-tumor and anti-HIV treatments (McIntosh et al. 2005) and in pharmaceutic products (Lehtovaara and Gu 2011). Despite the economic importance of curdlan, relatively little is known regarding the genetics and regulation of curdlan synthesis (Ruffing and Chen 2012).

The genes responsible for curdlan synthesis have been characterized in *A. fabrum* and especially in the reference strain C58, these are the *crdA*, *crdS* and *crdC* genes that seem to be arranged as an operon (Stasinopoulos et al. 1999). *crdS* codes for a putative beta-1,3-glucan synthase (curdlan synthase, CrdS) and *crdA* and *crdC* genes codes for CrdA and CrdC proteins that probably forms a multimeric complex with CrdS to assist the transit of curdlan across the two membranes of the cell envelope (Karnezis et al. 2003). Another gene, *crdR*, located outside the SpG8-2a region, regulates curdlan synthesis by activating expressions of its biosynthetic genes under oxidative stress, low pH and/or limited nitrogen with abundant sugar. This function of curdlan regulation indicates that curdlan biosynthesis under harsh conditions may have evolutionary origin for survival within a host plant (Yu et al. 2015). This exopolysaccharide seems to surround and protect bacteria. Thus, the production of curdlan is likely to increase the survival of bacteria in the soil (Ruffing and Chen 2012). Curdlan production may also play a role in attachment as attachment genes play a role in the ability of the bacteria to colonize roots, as well as in bacterial pathogenesis (Matthysse and McMahan 1998; Rodríguez-Navarro et al. 2007) and contact signaling.

Colonies of microorganisms capable of forming curdlan-type polysaccharide in glucose medium were found to stain blue with Aniline Blue, Brilliant Blue, or Trypan Blue, and red with Congo Red **(Figure 16A)** (Nakanishi et al. 1976). Aniline blue has thus been used both to detect curdlan production

by bacteria grown on agar medium and to identify mutants unable to produce curdlan (Nakanishi et al. 1976). Congo red was used to demonstrates that this function is specific to G8 species in agrobacteria **(Figure 16B)** (Lassalle et al. 2011)

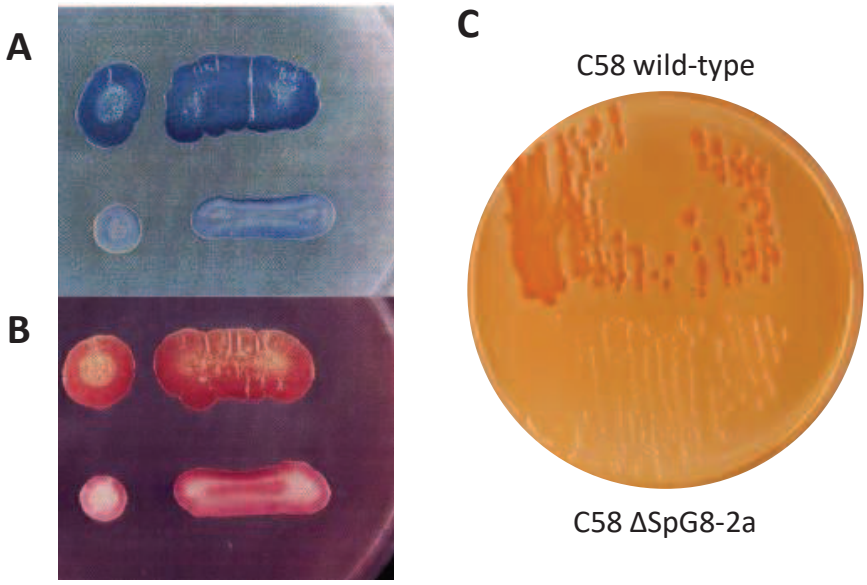


Figure 16. Experimental evidences of curdlan production. Colonies of organisms producing curdlan-type polysaccharide on agar plates containing (**A**) Aniline Blue (*Alcaligenes faecalis*) or (**B**) Congo Red (*Agrobacterium radiobacter*). (**C**) Curdlan production determined by SpG8-2a loci and revealed by red dye on Congo Red medium (*A. fabrum* wild-type C58, red colonies; ΔSpG8-2a, white colonies). (Nakanishi et al. 1976; Lassalle et al. 2011).

The second part of this region, called SpG8-2b **(Figure 15B)**, contains genes whose annotation suggests an involvement in the biosynthesis of secondary metabolites, the production of which would be NRPS (Non-Ribosomal Peptides Synthase).

SpG8-1 region: Sugar transport and metabolism / ferulic acid degradation pathway

This region has also been divided into two parts. The first, called SpG8-1a **(Figure 17A)** appears to be involved in the metabolism of sugars. This region contains several ABC transporters, including one specific for monosaccharides and two operons, whose homologs have been described as being specific to the import of amino acids (Hosie et al. 2002). Having key carriers of these nutrients, could give a competitive advantage to *A. fabrum*, allowing a better colonization to its main biotopes. This is to be taken with caution because the genome of *A. fabrum* includes a particularly large number of transporters (15% of the total genome of which 60% are ABC transporters). The SpG8-1a region contains another gene encoding a muconate lactonization enzyme, an intermediate of the highly conserved β-ketoadipate pathway in various bacteria, including several agrobacteria (Harwood and

Parales 1996). It is a degradation pathway for aromatic compounds frequently produced by plants and used as growth substrates by many members of the Rhizobiaceae family (Parke and Ornston 1986).

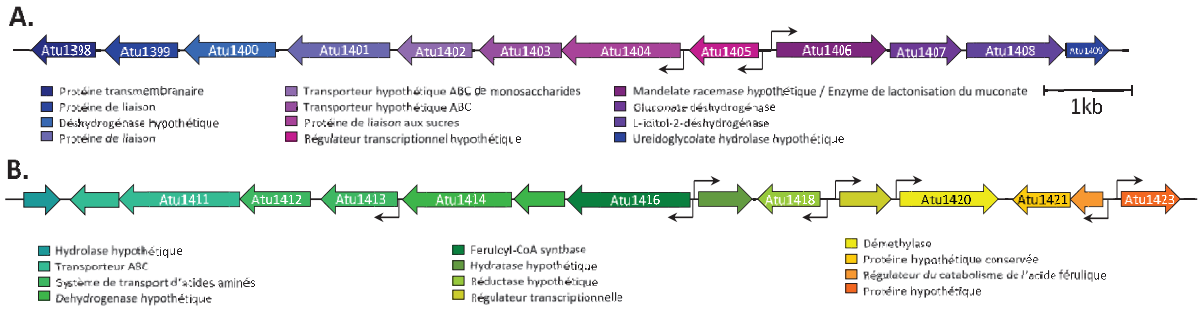


Figure 17. Organization of the SpG8-1 of *A. fabrum* C58 with their hypothetical functions. A. SpG8-1a. **B.** SpG8-1b. Promoters are shown with small arrows indicating the direction of gene transcription.

The second part of this region, SpG8-1b **(Figure 17B)** is involved in hydroxycinnamic acids (HCA) degradation (Campillo et al. 2014). HCA, such as ferulic acid and p-coumaric acid, are common plant secondary metabolites released in large amounts in soil during the decay of root cells (Whitehead et al. 1983). In rhizosphere, HCA serve a dual function of both repelling and attracting different organisms in the plant's surroundings (BHATTACHARYA et al. 2010). Indeed, while they have been demonstrated to generally inhibit the growth of bacteria (Sayadi et al. 2000; Ravn et al. 1989) or being toxic to most microorganisms (Perret et al. 2000), they can be chemotactic signals (Parke et al. 1987) and used as carbon sources by other soil bacteria (Deavours et al. 2006; Andreoni et al. 1995).

Campillo et al. shown that this region (SpG8-1b) encodes the complete degradation pathway of ferulic acid that will ultimately be used as a source of carbon and energy by *A. fabrum* via the incorporation of protocatechuic acid into the Krebs cycle. More specifically, ferulic acid is successively transformed into feruloyl-CoA, HMPHP-CoA, HMPKP-CoA, and vannillic acid by Atu1416, Atu1417, Atu1415, and Atu1421, respectively, and that Atu1420 associated with Atu1418 metabolized vannillic acid **(Figure 18A)** (Campillo et al. 2014). Similar pathway is observed with p-coumaric acid **(Figure 18B)** (Meyer et al. 2018). Thus, the *A. fabrum* pathway is more precisely an original coenzyme A-dependent -oxidative deacetylation (Campillo et al. 2014).

This pathway is finely regulated by atu1422 (hcaR) **(Figure 17B)** that encode HcaR, a transcriptional repressor of the MarR family (Deochand and Grove 2017) that regulates SpG8-1b gene expression and more precisely atu1416, atu1417, and its own transcription in the absence of HCA (Lassalle et al. 2011). Feruloyl-CoA and p-coumaroyl-CoA are the HcaR effectors that allow a positive feedback loop and induce HCA degradation gene expression (Meyer et al. 2018). As *A. fabrum* has two

contrasting lifestyles (the nonpathogen rhizosphere colonizer and the pathogen that causes crown-gall disease (Nester 2014)), the transition between lifestyles has to involve a coordinated modification of gene expression upon signal perception to express the appropriate gene at the right time (Duprey et al. 2014). So, HcaR regulation is decisive for the transition between the rhizospheric and pathogenic lifestyles. Indeed, in HCA-rich environments as rhizosphere (Mandal et al. 2009), it is important for *A. fabrum* to degrade and rapidly assimilate HCA to achieve competitive advantage over other agrobacteria. On the other hand, repression of HCA degradation by HcaR is important for providing *A. fabrum* a competitive advantage in colonizing tumors as HCA are not uniformly distributed in plant tissues, nor are they consistently available for bacteria (Mandal et al. 2009). So, the induction of SpG8-1b gene expression only after HCA detection is advantageous for bacteria to avoid the metabolic cost of constitutively expressing HCA degradation genes in plant environments with a low HCA content (Deochand and Grove 2017). Thus *A. fabrum* C58 can adapt its metabolism via HcaR regulation according to HCA availability (Meyer et al. 2018).

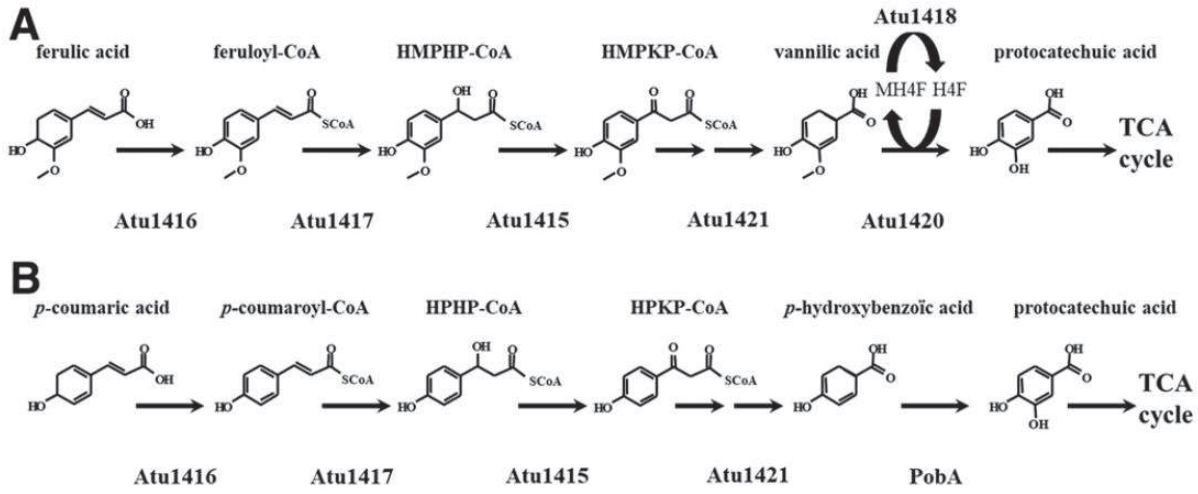


Figure 18. Ferulic acid and p-coumaric acid degradation pathway. A. Ferulic acid degradation pathway. **B.** p-coumaric acid degradation pathway (Meyer 2018).

Link between SpG8-1b and SpG8-3 genomic regions

A coordinated expression of several clade-specific genes resulting in conditional phenotypes has recently been observed in *A. fabrum* C58, strengthening the idea of the existence of an ecological niche to which species G8 is specifically adapted through the expression of a particular combination of clade-specific genes (Lassalle et al. 2017).

Transcriptional analyses highlighted the genes of the *A. fabrum* HCA degradation pathway (SpG8-1b) and the *A. fabrum*-specific iron acquisition pathway (SpG8-3) as being among the most

significantly upregulated genes in the presence of HCAs, either ferulic acid or p-coumaric acid **(Figure 19)** (Baude et al. 2016). Therefore, these two HCAs induce the expression of both *A. fabrum* specific gene clusters.

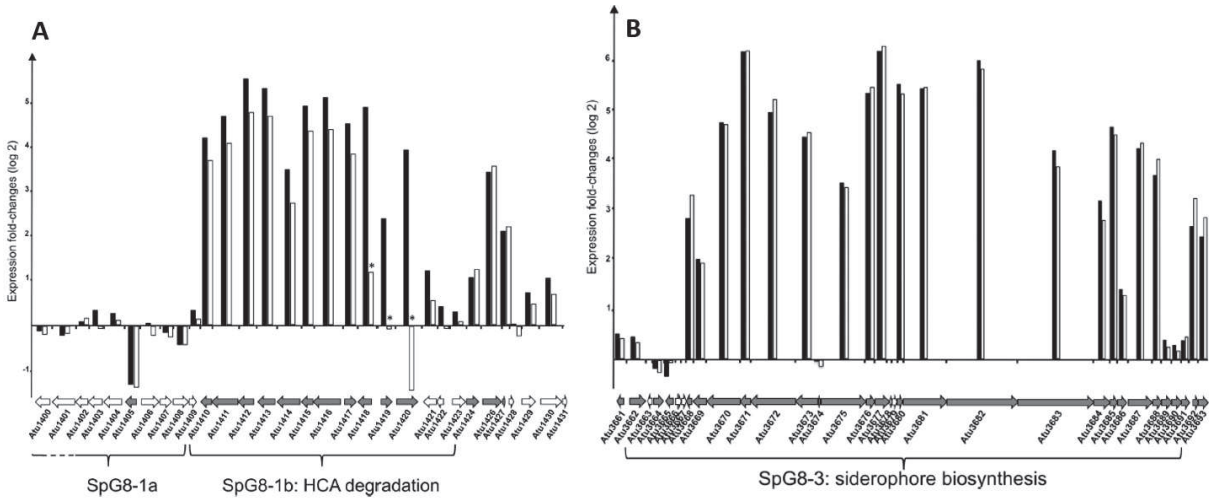


Figure 19. Upregulation of genes in the SpG8-1 and SpG8-3 genomic regions. Genomic regions regulated in the presence of ferulic acid and p-coumaric acid. For each gene, histograms represent logarithmic fold changes in expression levels measured from cells cultivated with versus without ferulic acid (black bars) or with versus without p-coumaric acid (white bars). **A.** SpG8-1 region. **B.** SpG8-3 region. Gray arrows, genes with significant expression differences **(Baude et al. 2016)**.

Baude et al. 2016 suspected the presence of a regulatory network coordinating the two functions since a consensus sequence has been identified in the promoter regions of almost all of the operons of both SpG8-1b and SpG8-3 clusters, suggesting the occurrence of an unknown DNA-binding protein that may bind to this consensus sequence. This putative transcriptional regulator would regulate both HCA degradation and iron uptake functions, in a coordinated manner (Baude et al. 2016). Furthermore, iron seems to be essential for ferulic acid degradation as a delay in its degradation under iron starvation conditions was observed **(Figure 20)**. There is also a possible control of iron homeostasis in the presence of HCAs since an excess in intracellular iron concentration was not found even though iron was not a limiting factor and that siderophore biosynthesis genes were induced (Baude et al. 2016). So, SpG8-1b and SpG8-3are interconnected and coregulated *A. fabrum* clusters, as the presence of one is require for the expression of the other and vice versa.

As we notice, *A. fabrum* has a very specific ability to capture iron in environments that are rich in HCAs but have extremely low iron levels, such as in decaying lignin-rich plant materials (Baude et al. 2016). From an ecological point of view, *A. fabrum* have an advantage over other agrobacteria in such environments as this two clusters cooperate together to determine a species-specific ecological niche.

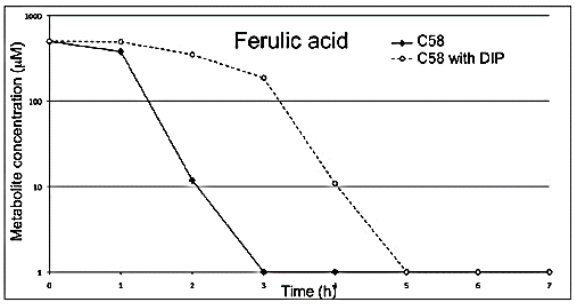


Figure 20. Ferulic acid degradation by A. fabrum C58 strain at a low iron concentration. Ferulic acid degradations were performed with 500 µM ferulic acid, with or without DIP (an iron chelator) (Baude 2016).

IV. Plant-bacteria interaction

A. Bacterial fitness and its measurement

Fitness has been used to indicate a measure of general adaptedness and to indicate a reproductive success (Partridge and Harvey 1988). The usage of fitness as a rate of increase was first introduced by Fisher who associated the term "fitness" with the "objective fact of representation in future generations" (Byerly and Michod 1991). Fitness involves the ability of organisms, populations or species to survive and reproduce in the environment in which they find themselves (Haldane 1932). Consequently, these organisms contribute genes to the next generation thus, having an evolutionary significance due to the subsequent genetic composition, but also an ecological importance through the population's growth (Williams 1992; Byerly and Michod 1991).

Fitness is a measure of natural selection, and plays a central role in evolutionary theory (Abrams 2012). Without fitness differences in genotypes, natural selection cannot operate and adaptation cannot occur (Gordo et al. 2011). Fitness is quantified by the number of offspring an organism produces throughout its life (Shaw et al. 2008). For organisms with binary fission such as bacteria, it becomes useful to define fitness for the genotype and estimate it from the genotype's growth rate in a particular environment (Gordo et al. 2011).

Fitness is often expressed relative to another genotype, such as the ancestor in evolution experiments. With microorganisms, relative fitness can be measured by allowing the ancestral and the evolutionarily derived line to compete with one another in a head-to-head competition (Elena and Lenski 2003). Thus, it is possible to reliably quantify evolutionary changes in fitness. Also, relative fitness depends not only on the genotypes but also on the environment in which it is measured (Orr 2009). Thus, it is possible to test the specificity of adaptation that occurred in an evolution experiment by measuring fitness in different environments at a given point in time (Mustonen and Lässig 2010).

Gene by gene approach (a targeted approach)

Two main types of assays to measure bacterial fitness have been generally used, both have different sensitivities and reflect different fitness components. The first one is growth curve assays **(Figure 21A)** where two or more strains (e.g. wild type and the mutant bacteria) grow in isolation over a period of time, and the dynamics of growth is determined. With this assay the ability of a genotype

to colonize an empty niche can be studied. The second one is competitive fitness **(Figure 21B)** where strains compete in a given environment and the comparison of the frequency of each genotype before and after competition leads to a competitive index (CI). Competitive fitness or mixed infection is more robust to small or subtle changes in the experimental conditions because they will affect both genotypes equally as both are subjected to the same conditions (Gordo et al. 2011; Segura et al. 2004; Macho et al. 2007).

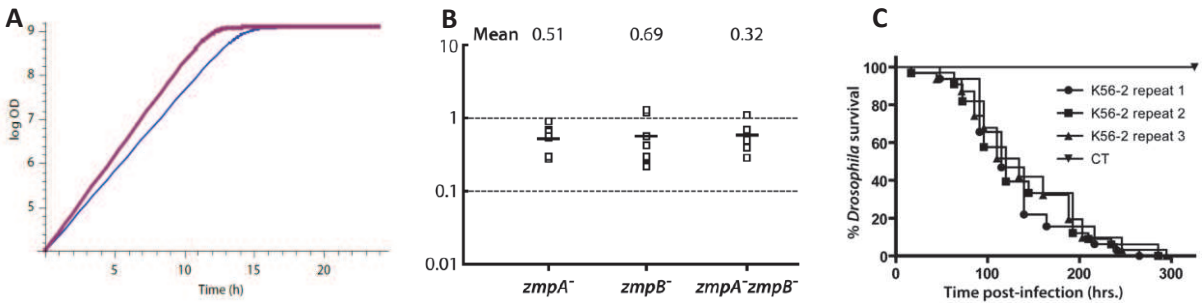


Figure 21. Types of assays to measure bacterial fitness (Gordo et al. 2011; Castonguay-Vanier et al. 2010). A. Growth curve assay. The wild-type (in pink) and mutant (in blue) genotypes are grown separately, and growth of each culture is estimated by taking periodic measurements of optical density (OD; e.g. every hour). OD is proportional to cell density. **B.** Competitive index (CI) analysis of three B. cenocepacia mutants in the D. melanogaster model. CI is defined as the ratio between the wild-type K56-2 and the mutant in the output (bacteria recovered from the fruit fly 96 h post infection) divided by their ratio in the input (inoculum). Each empty square represents the CI value obtained for one fly. A CI of less than 1 indicates a virulence defect. The mean of the CI is shown as a solid line. **C.** Survival curves for D. melanogaster flies challenged with B. cenocepacia K56-2. Pricking assays were performed in three independent replicates, each with a minimum of 30 flies.

Concerning bacterial virulence (the ability of a pathogen to cause disease), it has been classically evaluated in terms of lethality (lethal dose 50, LD50) which is defined as the number of bacteria required to kill 50% of the infected hosts **(Figure 21C)** (Reed and Muench 1938; Monk et al. 2008). This method yields data about the absolute virulence of the bacterium but if a gene deletion or mutation does not increase the LD50, it does not necessarily mean that the gene product does not play a role in the virulence of the bacterium (Castonguay-Vanier et al. 2010). On the other hand, with mixed infections a sensitive measure of virulence attenuation can be obtained with the identification of mutations or isolates with reduced competitive fitness (Falkow 2004). Besides, the variability inherent to the use of different hosts is avoided (Segura et al. 2004).

Mixed infections *in vivo* have thus proved a valuable tool in virulence studies (Lowe et al. 2007; Hunt et al. 2004) and have also been employed in screening methods and in the analysis of gene interactions (Beuzón and Holden 2001). So, bacteria fitness can be quantified using head-to-head

competitions by measuring the population growth rates that are achieved by each type as they compete for a pool of resources (Elena and Lenski 2003). Mixed infections provide thus information on the capacity of a mutant strain to compete with the wild-type strain.

Mixed infections have been broadly applied to the study of bacterial pathogens in animals or insects (Meynell and Stocker 1957) as in the case for example of virulence studies of *Burkholderia cepacia* in *Drosophila melanogaster* (Castonguay-Vanier et al. 2010). However, their application in plant pathogens has been very limited as bacterial growth does not seem to reflect growth in a single infection, because of growth interference between the co-inoculated strains and because some strains could be complemented by the wild-type. However, this is not intrinsic to growth within a plant host, but dependent on the dose of inoculation, the aggressiveness and the type of virulence factor of the co-inoculated strains as demonstrated with *Pseudomonas syringae* (Macho et al. 2007, 2010). Also, proportions of the two strains in the inoculum can be varied in order to counter complementation by the wild-type, this being another application of mixed infections i.e. the rescue of a mutant by the wild type (Segura et al. 2004).

Relative fitness in a mixed infection is measured by allowing two strains (e.g the ancestral and evolved types or the mutant strain and the wild type) to compete with one another in the same host **(Figure 22)** (Elena and Lenski 2003). The two competitors are mixed (usually at a 1:1 ratio) in the competition environment. To estimate the fitness cost of a mutation a competitive index (CI) must be calculated. The CI is defined as the ratio between the mutant strain and the wild type in the output (bacteria recovered from host post infection) divided by the ratio of the two strains in the input (inoculum) (Freter et al. 1981; Taylor et al. 1987). So, a CI = 1 would reflect that the mutant strain is able to grow as efficiently as the wild type, and a CI < 1 would indicate that growth of the mutant strain is attenuated (Macho et al. 2010). The CI is thus a relative measure of growth, and its value depends on when the measurement is taken (van Opijnen and Camilli 2013).

One crucial factor in mixed infections is to be able to discriminate the competitor strains, without adversely impacting on the natural fitness of the organism and with a lack of background activity in the environment to be studied (Wilson 1995). Different markers can be used to distinguish competitors, such as those that produce visible reactions with dyes, the use of antibiotic resistance markers, the use of plasmids encoding color markers (e.g., fluorescent proteins) or diagnostic PCR fragments (Segura et al. 2004; Elena and Lenski 2003).

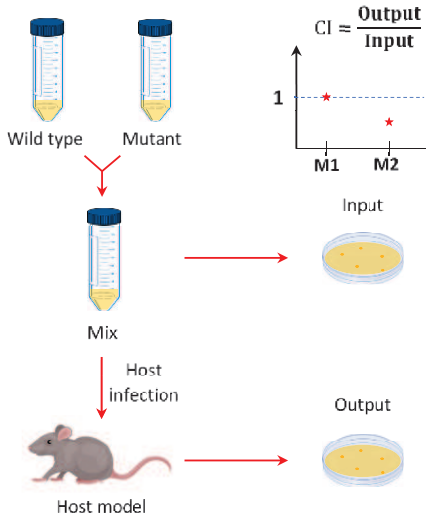


Figure 22. Determination of the competitive index (CI). The two competitors are grown separately and subsequently mixed. To calculate CI densities of both strains must be estimated. The CI is defined as the ratio between the mutant strain and the wild type in the output (bacteria recovered from host post infection) divided by the ratio of the two strains in the input (inoculum)(Freter et al. 1981; Taylor et al. 1987). A CI = 1 means the mutant strain has the same fitness as the wild type. A CI < 1 means the mutant strain has a lower fitness than the wild type (Macho et al. 2010).

Most reporter genes used in ecological studies allow detection of the marked organism by eye (Sessitsch et al. 1998). Identifying strains with color markers is easier and faster than other methods as no selective media is required and the strain proportions can be readily calculated (Figure). This reduced experimental variability, increase the accuracy and reproducibility of the essay and is potentially applicable to a variety of eukaryotic cell lines and intracellular pathogens. As in the case of the study of cell invasion and intracellular proliferation of *Salmonella enterica* in eukaryotic cell cultures where a green fluorescent protein (GFP) and a red fluorescent protein (RFP) were used to discriminate competitor strains **(Figure 23)** (Segura et al. 2004). Fluorescent proteins can also be measured by flow cytometry analysis **(Figure 24)** (Knöppel et al. 2014).

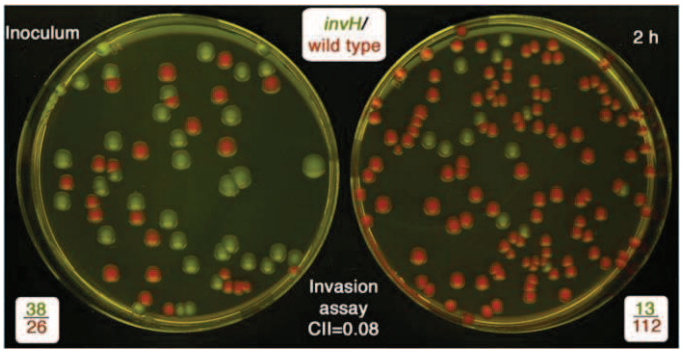


Figure 23. Competitive Index calculation with color markers (Segura et al. 2004). Different dilutions of a mix of the wild-type strain of *Salmonella enterica* (marked in red with a red fluorescent protein) and an invH mutant (in green with a green fluorescent protein) were plated on LB (on the left). Intracellular bacteria obtained 2h after infection of HeLa cells were also plated on LB (on the right).

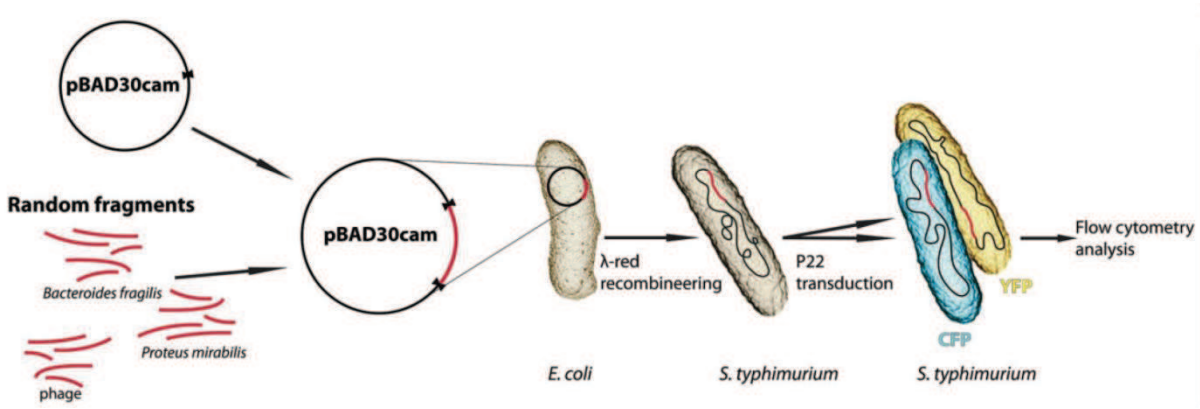


Figure 24. Fitness measure of HGT on the Salmonella chromosome by a flow cytometry analysis. Random inserts from Bacteroides fragilis, Proteus mirabilis, and human intestinal phage were cloned into a template plasmid, amplified together and then transformed into the Salmonella. The inserts were transduced into strains expressing cyan fluorescent protein (CFP) or yellow fluorescent protein (YFP), and the fitness of the inserts was measured through competition experiments followed by flow cytometry analysis.

The marker gene can encode an enzyme which gives rise to a colored product following incubation with a histochemical substrate (Sessitsch et al. 1998). One suitable marker gene for rhizobial competition studies is the *lacZ* gene, encoding ß-galactosidase, that has been used to assess rhizobial competition for the nodulation of soybean (Krishnan and Pueppke 1992) and to study root colonization by *Azospirillum* (Katupitiya et al. 1995). Another rhizobial compatible marker gene is the *E. coli gusA* gene, encoding ß-glucuronidase (GUS), a widely used reporter gene in plant molecular biology (Jefferson et al. 1987). It has also proved to be a highly suitable marker for studying plant-microbe interactions within nodules or on the root system (Streit et al. 1992, 1995) as GUS activity is not detected in many bacteria of agricultural importance such as *Rhizobium*, *Bradyrhizobium*, *Agrobacterium*, *Azospirillum* and *Pseudomonas* species or in their plant hosts (Sessitsch et al. 1997; Wilson et al. 1992). GUS cleaves glucuronide substrates such as X-glcA (5-bromo-4-chloro-3-indolyl-ß-D-glucuronide) or Magenta-glcA (5-bromo-6-chloro-3-indolyl-ß-D-glucuronide) releasing an indigo or magenta coloured precipitate by which the marked strain can be visualized. Quantitative assays for counts of soil bacteria based on the detection of gusA-marked cells on plates has been described (Wilson 1995). This marker gene was used to study rhizobial nodulation competitiveness, where strains were characterized for their competitive abilities in common bean (*Phuseolus vulgaris*) (Sessitsch et al. 1997).

Finally, the *celB* gene from the hyperthermophilic archaeon *Pyrococcus furiosus* encodes a thermostable ß-glucosidase with a high ß-galactosidase activity, which can be determined at temperatures up to 100°C (Voorhorst et al. 1995). Since the endogenous enzymes in both plant and

bacterium can be destroyed easily at high temperature, the thermostable ß-galactosidase has proved to be a suitable marker for rhizobial competition studies.

Several studies have been done in plants to study plant-microbe interactions. To better understand interactions between *Salmonella enterica* sv. Typhimurium and tomatoes competitive fitness was tested with fifty-one individual deletion mutants (Noel et al. 2010). Also, to explore the regulatory role of an orphan LuxR-type in *Sinorhizobium meliloti* mixed infections of a deletion mutant strain and a wild type where applied to the plants at various ratios and determined that the wild-type strain was more competitive for nodulation (Patankar and González 2009). Concerning competitiveness in rhizosphere colonization, a study with *Burkholderia ambifaria* found, using competition essays, a central difference in competitiveness between the wild-type and a phase variant in colonizing the rhizosphere suggesting the existence of two different phenotypic states of *B. ambifaria*, one of them more adapted to plant roots environment (Vial et al. 2010). To understand the role of the cell surface in nodulation competitiveness of *Rhizobium etli* mutants with altered colony morphology were studied in mixed infections and found a mutant which retained the ability to induce nitrogen-fixing nodules but substantially reduced its nodulation competitiveness and competitive growth on roots of common bean plants (Araujo et al. 1994).

Approach to the whole genome (an untargeted approach)

The availability of complete genome sequences provides the creation of relatively complete collections of strains and increased substantially the number of genes with unknown function (Saenz and Dehio 2005). Usually, to reveal the function of genes (e.g. in symbiosis and competitiveness in detail) gene disruption or deletion with subsequent phenotypic characterization is used. This method is efficient if one or a few genes have to be tested. However, it is very laborious and time consuming when used for screening of hundreds or thousands of mutants. A preselection of mutants often is performed to reduce the number of mutants to be tested but such a preselection reduces the chances to find unknown genes involved in the target phenotype (Milcamps et al. 1998; Trzebiatowski et al. 2001). An alternative strategy is a ‘‘negative selection’’ technology which generates mutant libraries of random whole-genome high-density transposon-insertion mutagenesis followed by massively parallel sequencing of transposon identification of insertion sites (Jacobs et al. 2003). This gene-disruption strategy produces a library of insertion mutants that can be tested in a host model to identify bacterial genes essential for growth or survival, thus mutations that reduce fitness under a particular condition (Saenz and Dehio 2005). The approach is cost-effective and applicable to a wide variety of microbes (Kumar et al. 2000).

A number of different genetic methods are currently available to unravel novel genes involved in bacterial interactions with their hosts, especially those that cannot be identified by bioinformatics genomic predictions. Signature tagged mutagenesis (STM) use hybridization to identify genes disrupted by transposon insertions combining with the ability to trace the fates of individual mutant strains within complex pools (Hensel et al. 1995; Matas et al. 2012). STM is a powerful negative selection method, predominantly used to identify ''niche-specific'' virulence genes in bacterial pathogens, thus required for a successful colonization (Saenz and Dehio 2005; Andrews-Polymenis et al. 2009). STM is based on transposon insertional mutagenesis that allows large numbers of mutants to be analyzed simultaneously. This is accomplished by using signature tags (i.e. unique short DNA sequences) inserted in the transposons to mark each mutant individually so that it can be distinguished from other mutants (Shimoda et al. 2008). A tag consisted on 40 bp variable central segment flanked by invariant 'arms' of 20 bp in length, which enable the co-amplification and labelling of the central portions by PCR **(Figure 25A)**. Therefore, the sequence tag acts as a molecular barcode to monitor the presence of each mutant in the mixed population before and after they have been subjected to selection (Mazurkiewicz et al. 2006). Mutants carrying distinct signature-tags are pooled and used as an inoculum into a host to test in parallel for their survival **(Figure 25B)**. This is advantageous as it minimizes both the work-load and the number of hots required (Shimoda et al. 2008). In the same way, tags also permit the identification of bacteria recovered from infected hosts as well as the selection of mutants with reduced or increased adaptation by analyzing the relative abundance of transposon insertions (Hensel et al. 1995). If a mutant does not survive in the host or compete with fully virulent mutants during pathogenesis, it will not be recovered following inoculation and incubation. Therefore, in theory, any mutant that does not grow as vigorously in the host as the wild-type strain will be identified as a candidate virulence mutant (Wang and Beer 2006). STM allows simultaneous testing of hundreds of mutants. Thus, large libraries of mutants can be screened quite easily in a single experiment using single selection screens capable of identifying fitness or competition mutants which do not cause a total phenotype knock-out and are usually missed by standard mutagenesis approaches (O'Sullivan et al. 2007; Pobigaylo et al. 2006, 2008).

STM has proved to be an invaluable tool to provide a better understanding of microbial behavior in vivo and also a robust and powerful high throughput screening technique for the analysis of genes that are not essential for life, but are needed for growth in specific environments (Saenz and Dehio 2005). The application of STM to a range of microbial pathogens has resulted in the identification of novel fitness-related genes and virulence determinants in each screen performed to date mainly in

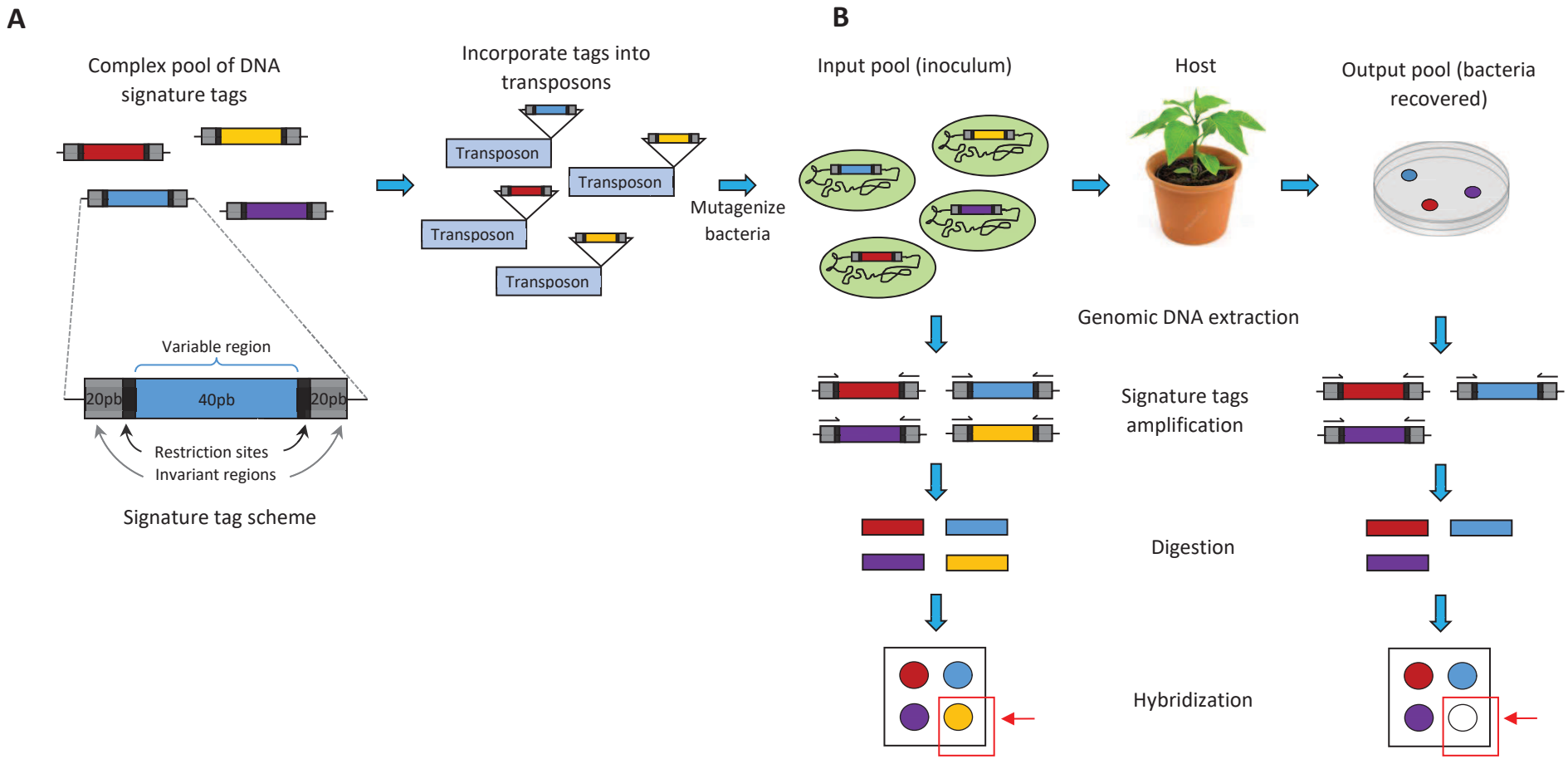


Figure 25. Signature-tagged mutagenesis (STM). (Mazurkiewicz et al. 2006). Each label contains a unique central sequence of 40 bp flanked by two invariant regions of 20 bp common to all the labels. The junctions of the variable and invariable regions are separated by restriction sites. A pool of labels (presented as colored rectangles) is ligated to the transposons which are then used for the mutagenesis of bacteria which, in turn, will serve as an inoculum on a host of interest. Genomic DNA is extracted from this input pool and bacteria recovered from the host (output pool). The labels thus obtained are amplified by means of primers specific to their invariant regions. These will then be removed by digestion with a restriction enzyme. Finally, a comparison of the hybridization between the input pool and the output pool is performed.

bacteria, but also in yeast and other fungi, viruses, parasites and, most recently, in mammalian cells (Autret and Charbit 2005; Mecsas 2002; Lawley et al. 2006; Bianconi et al. 2011).

STM can be successfully applied to ecologically important microbial interactions, defining the underlying genetic systems important for biotechnological fitness of environmental bacteria (O'Sullivan et al. 2007). In rhizobia, recent genome-wide analyses have provided various intriguing and useful starting points for more detailed functional studies (Bergès et al. 2003; Hoa et al. 2004). However, applications of this strategy to the study of plant–bacterium interactions have been scarce and have been mainly directed to identify bacterial genes relevant to symbiosis and competitiveness in the rhizosphere of plants (Pobigaylo et al. 2008; Shimoda et al. 2008).

This system was initially applied to a murine model of typhoid fever caused by *Salmonella typhimurium*, resulting in the identification of new virulence genes (Hensel et al. 1995). Examples of biological insights from STM screens are among others, the discovery of a specialized type 3 secretion (T3S) system on *Salmonella* typhimurium (Boucrot et al. 2005), the discovery of a complex cell wall lipid, the phthiocerol dimycocerosate (DIM), necessary for survival of *Mycobacterium tuberculosis* in the lung (Camacho et al. 1999) or the fact that the invasion of Shigella flexneri is dependent on the activity of a T3S system that delivers cell bacterial effectors into the host, eliciting dramatic rearrangements of the cytoskeleton (West et al. 2005).

When STM was used to identify genes required by *Salmonella enterica* serovar Typhimurium colonization of porcine intestines, up to 95 uniquely tagged transposon insertion mutants were analyzed and confirmed important roles for type III secretion systems (T3SS) and in intestinal colonization of pigs (Carnell et al. 2007). STM was also used in a more complex ecological interaction to identify genetic determinants of fitness associated with two key ecological processes mediated by *Burkholderia vietnamiensis*: phenol degradation as a model of bioremediation (Nelson et al. 1987) and rhizosphere interaction leading to growth promotion activity (Gillis et al. 1995). STM completely defined the components of the phenol degradation pathway, identified novel accessory genes not previously implicated in phenol utilization and confirmed genes compromised in their rhizosphere fitness (O'Sullivan et al. 2007). STM was also applied to a bacterial phytopathogen belonging to the *Pseudomonas syringae* complex, *P. savastanoi* pv. Savastanoi to identify metabolic pathways required for full fitness in olive (*Olea europaea*) knots and revealed novel mechanisms involved in the virulence of this pathogen, such as the type IV secretion system (Matas et al. 2012). Another phytopathogen where STM was used is *Erwinia amylovora* to identify genes that contribute to the virulence in plants finding a gene involved in the biosynthesis of an exopolysaccharide and a loci implicated in the

biosynthesis or transport of particular amino acids or nucleotides (Wang and Beer 2006). STM was also used in the identification of genes relevant to symbiosis and competitiveness in nitrogen-fixing symbiotic bacteria and even genes that influence symbiosis in a more subtle way (Pobigaylo et al. 2006, 2008; Shimoda et al. 2008).

Genes and their protein products are organized in complex networks. To increase the understanding of such networks, the development of high-throughput tools allowing the analysis of the genetic interaction at a genome-wide scale is necessary for the discovery of interactions in microorganisms (van Opijnen et al. 2009; Goodman et al. 2009). Recently, high-throughput Illumina sequencing has been combined with transposon mutagenesis bringing an unparalleled degree of resolution in robust, reproducible and sensitive techniques to quantitatively determine each gene's contribution to fitness simultaneously and for the discovery of quantitative genetic interactions in microorganisms on a genome-wide scale (Langridge et al. 2009; van Opijnen and Camilli 2013; van Opijnen et al. 2014). A re-analysis of an *Escherichia coli* transposon library in cattle revealed that transposon sequencing is superior to the original STM analysis in terms of sensitivity, allowing for the identification of hundreds of additional virulence genes (Eckert et al. 2011). These methods, based on the construction of a saturated transposon insertion library, are referred to as transposon sequencing (Tn-seq), high-throughput insertion tracking by deep sequencing (HITS), insertion sequencing (INSeq) and transposon-directed insertion site sequencing (TraDIS), making it possible to identify putative gene functions in a high-throughput manner.

These techniques allow the use of a single transposon, replacing the sets of individually barcoded variants needed for STM (Hensel et al. 1995). Thus, fitness can be rapidly determined in a specific environment without a pre-existing array of mutants (van Opijnen et al. 2009). Also, they provide an alternative to the species-specific DNA microarrays required for hybridization-based mutant profiling (Mazurkiewicz et al. 2006). And finally, with sequencing, the signal is of a ‘‘digital’’ nature; any sequence read is an indication of the exact position of a transposon insertion site and a count-based abundance readout of individual insertions (Langridge et al. 2009). In this way, independent insertions with shared behavior serve to validate gene-level fitness effects. The great number of insertions allow to distinguish between essential and nonessential genomic regions and the assay of nearly every gene in the genome for a particular growth condition (Langridge et al. 2009).

These techniques are mainly based on the properties of the Himar1 Mariner transposon for the construction of transposon insertion libraries, carrying out stable random insertions in a recipient genome without specific host factors (Gawronski et al. 2009). A particularity of this transposon is the

presence of a restriction site for the MmeI enzyme at the ends of the transposon. MmeI is an endonuclease which cuts 20 bp upstream of its recognition site, leaving a cohesive end of two base pairs. It was thus possible to generate banks with a theoretical insertion every 13 bp (van Opijnen and Camilli 2013). The ability of these transposons to serve as mutagenic agents is well established in members of all three domains of life (Lampe et al. 1996; Mazurkiewicz et al. 2006).

The approach works as follows. A gene disruption library is constructed after a first phase of in vitro transposition **(Figure 26A)** and a subsequently transformation of bacteria. The result gives a pool of strains each containing a single insertion of the transposon into its genome **(Figure 26B)**. DNA is isolated from a portion of the transformed bacterial pool (entry pool or t1). After passage of the library in the environment of interest, bacteria are recovered and DNA is isolated again (t2). The DNA thus obtained at both times is digested with the MmeI restriction enzyme **(Figure 26C)**. This cleavage generates fragments comprising the transposon plus 16 bp of flanking genomic DNA sufficient to precisely identify the site of insertion into the genome, leaving a two-base overhang which facilitate the ligation of an adapter (van Opijnen and Camilli 2013). By using a primer specific for this adapter and a primer specific for the transposon, it is possible to carry out a PCR whose fragments obtained are then sequenced in mass by Illumina-type approaches. The capture of adjacent chromosomal DNA performed by MmeI makes it possible to define the location of the transposons in the genome by mass sequencing (van Opijnen et al. 2009). Each 16 bp insertion is mapped on the genome, and the number of reads makes it possible to determine the relative abundance of the transposon at a given site **(Figure 26D)**. In other words, these reads are an estimate of the fitness conferred by each gene that could be used to calculate each mutant's fitness (Goodman et al. 2009). In this way, the contribution of each gene to the fitness of a microorganism in a given condition can be determined. Remarkably, what makes Tn-seq so robust is that the fitness of each gene is based on the average effect of multiple independent transposon insertions into that same gene **(Figure 26E)** (van Opijnen et al. 2014). To calculate fitness for a single gene, fitness must first be calculated for each transposon insertion, followed by averaging over all the insertions found within that gene to obtain a single fitness value.

So, upon growth of the library in the environment of interest, insertion mutants with a lower fitness decrease in frequency in the population, while other mutants, depending on the effect of the transposon insertion, remain the same or increase in frequency. Changes in frequency are determined by sequencing the transposon flanking regions en masse (van Opijnen et al. 2014). Thus, even mutants which have a fitness difference very little different from that of the wild (less than 5% difference compared to the wild strain) can be highlighted in this type of study (van Opijnen and Camilli 2013).

This method will open up the possibility to screen for genetic interactions across many different environments, strains and species (van Opijnen et al. 2009).

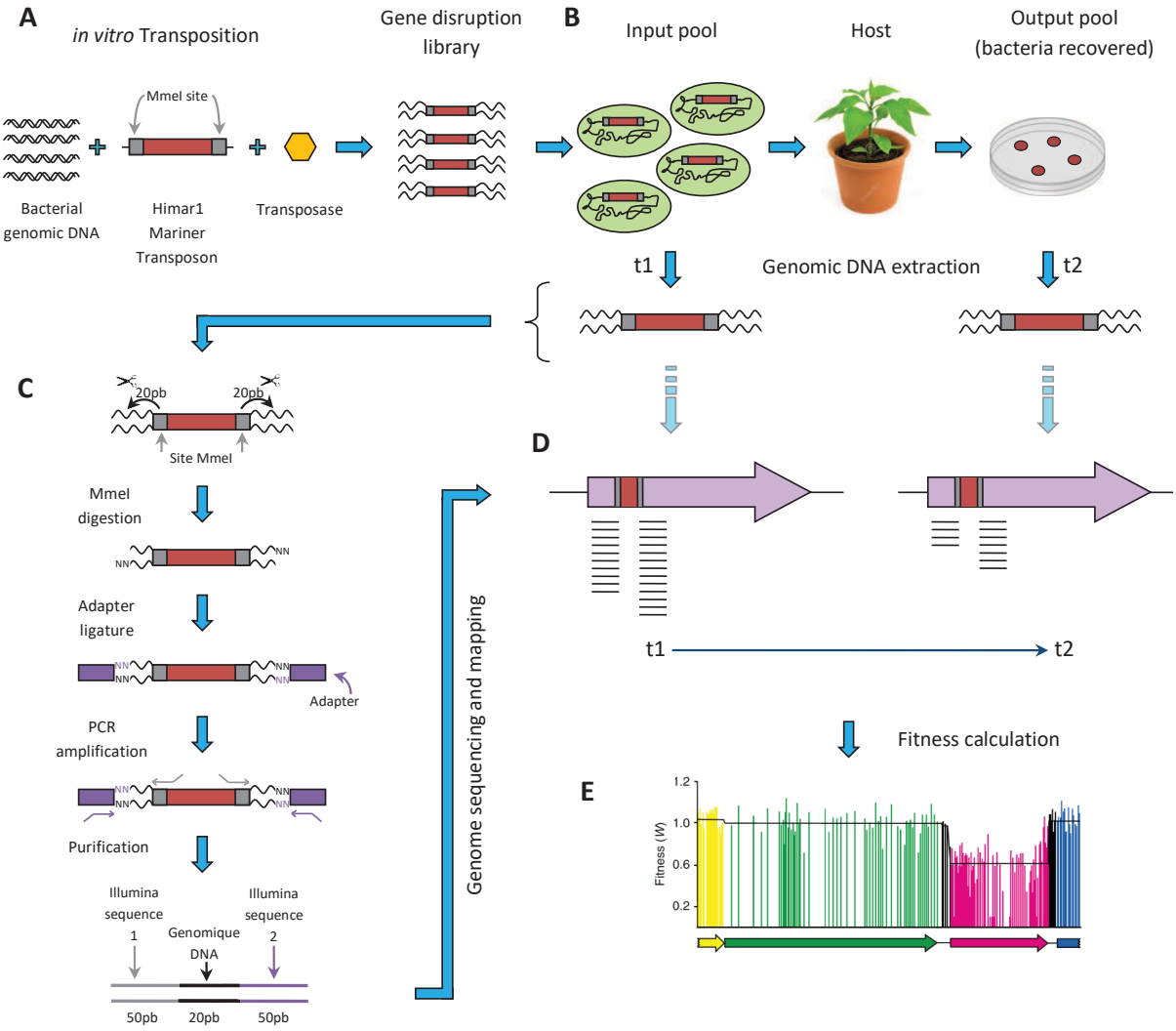


Figure 26. Transposon sequencing (Tn-seq) (van Opijnen et al. 2009). A. A gene disruption bank is produced by in vitro transposition on a pool of bacterial genomic DNA, using a transposase and the Himar1 Mariner transposon. This transposed DNA is then used to transform bacteria that will have only one insertion of the transposon into their genome. B. DNA is extracted from the pool of mutants (t1) and bacteria recovered from the host (t2). C. Digestion is carried out with the MmeI restriction enzyme, which cuts 20 bp downstream of its recognition site. The cohesive 2bp will be used for ligating an adapter. Amplification is carried out using a primer complementary to the adapter and another to the transposon. Purification is subsequently carried out to obtain fragments of 120 bp which contains 20 bp of bacterial genomic DNA flanked by 50 bp of Illumina sequences allowing sequencing. D. Fragments are then counted and mapped on the genome to allow fitness calculation. E. The different insertions of transposons are distributed over the entire length of the genes. By averaging all the insertions of a given gene, fitness value is obtained. A gene with a fitness equal to 1, is a neutral gene in the condition tested (does not contribute to fitness) (green gene). Genes with harmful insertions will decrease in the population and their fitness will be reduced (less than 1) (pink gene).

Using transposon sequencing in several in vivo environments has the potential to reveal the requirement for specific genes in specific niches. The method was originally developed for the Gram-positive bacterium *Streptococcus pneumoniae*, a causative agent of pneumonia and meningitis to determine gene's contribution to fitness and to identify genes essential for growth (van Opijnen et al. 2009) and genes important for in vivo colonization (van Opijnen and Camilli 2013; Mann et al. 2012) among others. Tn-seq has since been applied to several different microbial species, for example to Streptococcus pneumoniae to determine fitness for each gene and identifying those likely to be essential for basal growth (van Opijnen et al. 2009), to Pseudomonas aeruginosa to identify antibiotic resistance genes (Gallagher et al. 2011), to *Mycobacterium tuberculosis* to identify essential genes and pathways that are involved in the utilization of cholesterol (its crucial carbon source during infection) (Griffin et al. 2011) or to Porphyromonas gingivalis to identify essential genes (Klein et al. 2012).

HITS was applied to identify *Haemophilus influenzae* genes required to delay bacterial clearance in a murine pulmonary model. More than 96% of *H. influenzae* protein coding genes were analyzed. The profile of genes required in this environment provides a view of the host–pathogen interactions occurring during pulmonary pathogenesis and will provide insight into potential strategies for the design of vaccines or therapeutics to specifically target *H. influenzae* (Gawronski et al. 2009). INSeq was used to identify microbial genes critical for the fitness of a saccharolytic human gut bacterium *Bacteroides thetaiotaomicron*, in the presence or absence of other human gut commensals, revealing functions necessary for survival influenced by community composition and competition for nutrients (vitamin B12). The results provide evidence that human gut symbionts possess mechanisms critical for interaction with their host shaped by other members of the microbiota (Goodman et al. 2009). TraDIS was used to identify niche-specific essential gene set of the bacterium *Salmonella enterica* serovar Typhi (S. Typhi) under both standard laboratory and biologically relevant conditions. This was useful to understand how enteric bacteria resist the toxic effects of bile in the human gut and will also contribute to the future development of therapeutics targeted to the treatment of *S. Typhi* carriage (Langridge et al. 2009).

Deep sequencing of populations opens unprecedented opportunities to trace the genomic basis of adaptation by performing efficient fitness analysis within and across species, as well as in time series of evolution experiments in diverse contexts of host–microbe interactions. Understanding how a genotype leads to a specific phenotype is key to understanding how an organism functions and survives (Gawronski et al. 2009; Goodman et al. 2009).

B. Metabolomics contributions to the study of plant-bacteria interaction

Over the course of evolution, plants have developed many strategies to adapt to their environment (Popovici et al. 2010). Plants' growth and physiology are strongly influenced by the biotic interactions that plants establish with a highly diversified microbial community in soils (Berg and Smalla 2009). Concerning bacteria, plants are able to sense and respond to them by inducing defense pathways against pathogens or by stimulating the growth of root-associated beneficial or commensal bacteria through root exudation (Chamam et al. 2015). This response of plants to bacteria can impact resource allocation (Chiapusio et al. 2018). Secondary metabolism is thought to be one of the main tools for a plant to adapt to its environment. Plant secondary metabolites play a major role in mediating plant-soil interactions by affecting microbial community dynamics, acting as weapons or positive signals during biotic interactions and may ultimately impact ecosystem processes (Edreva et al. 2008; Iason et al. 2012).

Advancements in the field of metabolic profiling of plant secondary metabolites have resulted in recent discoveries related to the role of such metabolites in plant defense, chemical signaling processes, stress tolerance and symbiosis (Weston et al. 2015; Mandal et al. 2010). In particular, pathogens induce in host plants the biosynthesis of phenolic compounds with antimicrobial properties (phytoalexins), and of phenolic lignin polymers to strengthen cell wall (Vanholme et al. 2010). Phenolic compounds (also named polyphenols) are widely distributed in the plant kingdom and comprise a large chemical diversity, ranging from simple phenolic acids to complex polymerized tannins (Harborne and Williams 2000). Several works have outlined the indirect effects of phenolics on plant–plant interactions, particularly with respect to their influence on microbial community dynamics including that of plant symbionts (Cipollini et al. 2012; Putten et al. 2013).

Global approaches targeting the identification of partners' metabolites are needed to assess plant–bacteria interactions. Liquid or gas chromatography combined with tandem mass spectrometry (LC-MS or GC-MS) has become an indispensable tool for phytochemical profiling (Ferreres et al. 2008; Liu et al. 2009; Marczak et al. 2010), the latter being one of the most frequently used tools for profiling primary metabolites (Fiehn 2008). Until now, thousand kinds of plant secondary metabolites have been identified from plant kingdom. Thus, in recent years, metabolomics has become a powerful tool to understand the complexity of plant systems (Abdel-Farid et al. 2014; William Allwood et al. 2010; Pedras and Zheng 2010). Several metabolomic studies have been carried out to identify metabolites involved in bacteria-plant interactions and this, using several kind of plants and using different parts

of the plant, as the study of the responses of different *Frankia* strains to *Myrica gale* fruit exudates (Popovici et al. 2010) or analysis of xylem sap metabolites during plant-PGPR interactions using a PGPR strain *Azospirillum lipoferum* CRT1 and maize as a plant model system (Rozier et al. 2016). Even though, Much more information is available concerning plant secondary metabolic responses to bacterial pathogens (Chamam et al. 2015).

In roots, symbiotic associations resulting in nodules in which carbohydrates provided by the host plant are exchanged for nitrogen supplied by the bacteria have also been studied by metabolomic approaches. A recent study compares nodule metabolome from *Alnus glutinosa* to highlight modifications associated with in-planta sporulation using UV and mass spectrometric–based metabolite profiling (Hay et al. 2017). In the nodule, two types of *Frankia* strains have been described: strains able to sporulate inside plant cells (Sp+) and those which are not (Sp-) (Pozzi et al. 2015). UHPLC/UV/ESI-MS was used for secondary metabolites analyses and 50% of secondary metabolites were found significantly different between Sp+ and Sp− nodules. One secondary metabolite, identified as gentisic acid 5-O-β-d-xylopyranoside, was specific to Sp+ nodules and previously reported as involved in plant defenses against microbial pathogens (Fayos et al. 2006). These results suggested that the host plant could exert control over potentially less efficient in nitrogen fixing Sp+ symbionts (with a “cheater” behavior as they divert part of the plant energy resources for the production of many sporangia) through specific secondary metabolites (Hay et al. 2017).

Another metabolomic study was achieved this time to understand the diversity of the internal architectures on nitrogen-fixing root nodules infected with *Frankia*, comparing between the metabolomes of *Alnus glutinosa* and *Casuarina cunninghamiana* root nodules with roots from uninfected plants was performed (Brooks and Benson 2016). Using GC-MS on extracts of nodules and roots and it was found that between one-third to one-half of the metabolites significantly increased or decreased between roots and nodules. The authors concluded that *C. cunninghamiana* responds more robustly to the presence of *Frankia* than *A. glutinosa* with metabolite patterns consistent with different strategies used for compartmentalizing the symbiont from uninfected tissues (Brooks and Benson 2016). Another comparative metabolomic study was to explore the role for secondary metabolites comparing flavonoids from the metabolomes of roots and nodules of the actinorhizal host *Datisca glomerata* and of *Medicago sativa* (Gifford et al. 2018). The UHPLC/UV/ESI-MS analysis showed flavonoids of the same classes present in roots and nodules of both plants, including flavanones, flavonols, and isoflavonoids, suggesting similar roles for flavonoids during nodule development and symbiosis across lineages. An increase in flavonoids was observed in the nodules (indicating a likely major role) and while both hosts produced derivatives of naringenin, the metabolite profile in *D.*

glomerate indicated an emphasis on the pinocembrin biosynthetic pathway, and an abundance of flavonols with potential roles in symbiosis (Gifford et al. 2018). Another application of metabolomics in nodules was to identify the key metabolites that are present during biological nitrogen fixation. A combination of mass spectrometric techniques was performed using the *Medicago truncatula*–*Sinorhizobium meliloti* association as the model system (Gemperline et al. 2015). For this purpose, well-characterized plant and bacterial mutants capable or defective in nitrogen fixation were used to detect metabolic differences. Thereby, it was possible to have a better understanding of the underlying mechanisms of the biological nitrogen fixation process, providing an insight into the roles of metabolites in symbiotic nitrogen fixation and the responsible metabolic pathways (Gemperline et al. 2015).

Concerning denitrification process, metabolomic studies have also been performed to test if biological denitrification inhibition of *Fallopia* spp. may be explained by their secondary metabolites (Dassonville et al. 2011). In order to identify major phenolic compounds, *Fallopia* spp. extracts and their fractions were analyzed using an UHPLC/UV/ESI-MS approach and then tested on a complex soil microbial community to evaluate the biological denitrification inhibition effects on denitrifying bacteria. These results show that secondary metabolites from *Fallopia* spp. may reduce metabolism in denitrifying bacteria which could later result in a reduction in the abundance of denitrifying bacteria (Bardon et al. 2014). Even if the major molecules involved remain unidentified it has been suggested that polyphenolic compounds, such as catechin (the precursor of procyanidins reported to show antibacterial properties) could be involved (Lacombe et al. 2013). This provides new insight into plant–soil interactions and improves our understanding of a plant's ability to shape microbial soil functioning (Bardon et al. 2014).

Phenolic compounds are implied in plant-microorganisms interaction and can be also modified in response to plant growth-promoting rhizobacteria (PGPRs) (Cheynier et al. 2013; Chamam et al. 2013) as demonstrated with the beneficial bacterium *Paraburkholderia phytofirmans* PsJN known to modulate primary and secondary metabolism in different plants (Issa et al. 2018; Su et al. 2016; Fernandez et al. 2012). In this study, authors investigate the impact of PsJN root-inoculation on grapevine secondary metabolism by profiling root secondary metabolites (UHPLC-UV/DAD-MS QTOF) and comparing with control non-bacterized roots (Miotto-Vilanova et al. 2019). This metabolomic approach showed a decrease of glycosylated hydroxybenzoic acids derivatives as response of grapevine to PsJN inoculation. This is probably due to the secretion of aglycon forms (with an efficient antimicrobial activity) as plant defense mechanisms against microorganisms (Sánchez-Maldonado et al. 2011). Metabolomic analyses also showed that hydrolyzable tannins and flavonoids, well-known for

their antimicrobial effect, were accumulated in grapevine plantlets following PsJN inoculation. These compounds, tested on *Botrytis cinerea*, showed an inhibitory effect on fungal spore germination, demonstrating that PsJN is able to protect the plant against *B. cinerea* by a direct antifungal effect (Miotto-Vilanova et al. 2016). This finding suggests a supplementary biocontrol mechanism developed by this PGPR to fight against pathogens, giving to the plant a better resistance to future attacks by pathogenic microorganisms. From the ecological point of view, as exuded flavonoids are well-known signals, this may help the bacterium to efficiently establish in the grapevine rhizosphere through the use of plant metabolic abilities to manage its niche (Miotto-Vilanova et al. 2019).

Also in the roots, using proton nuclear magnetic resonance, mass spectrometry, and ultraviolet-visible absorbance analyses, another study aims at determine if the presence of *R. meliloti* around alfalfa roots alters the composition of root exudates either by increasing the amount of nod gene inducing activity or by promoting exudation of phytoalexins (Dakora et al. 1993). Results showed that inoculating alfalfa roots with symbiotic bacteria clearly increased total nod gene-inducing activity in root exudate and that three isoflavonoids appear (formononetin-7-0-(6”-0-malonylglycoside), a medicarpin-3-O-glycoside, and medicarpin) in these exudates. Both results could be related to a decline in 4’,7-dihydroxyflavone after *R. meliloti* inoculation. Indeed, this compound can inhibit the activity of a strong nod-gene inducer chalcone (Maxwell and Phillips 1990). Besides, this compound is an important storage product in young alfalfa roots and a major portion of this compound exuded by alfalfa roots is closely linked to concurrent synthesis (Maxwell and Phillips 1990). Presumably, redirecting carbon flow to isoflavonoids would decrease 4’,7-dihydroxyflavone in the exudate. Furthermore, formononetin-7-O-(6”-O-malonylglycoside) isolated in this study induced nod genes regulated by both NodD1 and NodD2 proteins in R. meliloti. Moreover, these results indicate that alfalfa responds to symbiotic *R. meliloti* by exuding a phytoalexin normally elicited by pathogens and that this symbiotic bacterium can use a precursor of the phytoalexin as a signal for inducing symbiotic nod genes (Dakora et al. 1993).

Another metabolomic application is a study documenting a non-symbiotic bacteria with PGPR effect on plant secondary metabolite profiles that shows a complex interaction between *Azospirillum* PGPR and maize (Walker et al. 2011). Indeed, using chromatographic profiling of root secondary metabolites, authors found that *Azospirillum* inoculation led to a significant impact on plant secondary metabolism. Results showed a major qualitative and quantitative modifications of the contents of secondary metabolites, especially benzoxazinoids, important for plant interactions that may serve as early markers of effective PGPR–maize interactions. Additionally, benzoxazinoids are defense molecules since they are repellents for aphids (Nicol et al. 1992), display antimicrobial properties (Sahi

et al. 1990) and inhibit vir genes of *Agrobacterium tumefaciens* (Zhang et al. 2000). Furthermore, secondary metabolites modifications founded depended on the PGPR strain and plant cultivar combination, suggesting the presence of fine-tuned interaction mechanisms (Walker et al. 2011). Another metabolomic work using RP-UHPLC/UV/ESI-MS also on maize, aimed at understanding the early impacts of seed inoculation on root secondary metabolite profiles by a singly inoculation of *Pseudomonas*, *Azospirillum* and *Glomus* with respect to the microbial consortium inoculation (Walker et al. 2012). Results showed that inoculation led to qualitative and quantitative modifications of root secondary metabolites, particularly benzoxazinoids and diethylphthalate. These modifications depended on fertilization level and microorganism(s) inoculated (Walker et al. 2012). Finally, with the same metabolomic approach, one more study on two maize lines sought to assess the plant-beneficial effects of the bacterial models *Agrobacterium fabrum* C58 and *Escherichia coli* K-12 compared to an *Azospirillum brasilense* positive PGPR control (Walker et al. 2013). Inoculation of C58 or K-12 resulted in a systemic impact on secondary metabolism depended on maize lines, with significant modifications of benzoxazinoids. Changes were similar between these two bacteria comparable to those of *Azospirillum* in one of the maize lines (a semi-late hybrid PR37Y15) having even a phytostimulatory effect. This suggests that all three strains were recognized by this maize cultivar as belonging to the same "type" of microorganism. In contrast, in the other maize line (a semi-early hybrid DK315) changes on secondary metabolites were different between the two bacteria plus, they had no effect on plant biomass (Walker et al. 2013).

Profiling of plant secondary metabolites can also be used to compared responses upon colonization of the same plant by cooperative, commensal and pathogenic bacteria in order to differentiate the different types of ecological interactions. This was possible in a metabolomic study using RP-UHPLC/UV/ESI-MS on two *Oryza sativa* L. cultivars, as rice is expected to be altered in response to bacterial root colonization (Chamam et al. 2015). The bacteria used in this study was *Azospirillum lipoferum* 4B (a PGPR), *Burkholderia glumae* AU6208 (a rice pathogen, causing agent of bacterial panicle blight) and *Escherichia coli* B6 (a commensal environmental bacteria). Comparison of metabolic profiles evidenced that each bacterial ecological interaction induced distinct qualitative and quantitative modifications of rice secondary metabolism that varied according to the cultivars and the interaction types (Chamam et al. 2015). Numerous flavonoid compounds and hydroxycinnamic acid derivatives were commonly affected in rice in response to PGPR and pathogen inoculation but they mostly varied in an opposite way between the two interactions. Plant response to the commensal bacterium is much closer to the PGPR-triggered response than to the pathogen ones in roots, where these two strains may colonize the root surface without penetrating the inner root and inducing cell damages. This study demonstrates the relevance of secondary metabolic profiling to differentiate

plant–bacteria interactions, allowing to decipher metabolic pathways induced in the host plant in response to ecologically distinct bacteria (Chamam et al. 2015).

A metabolomic study can also be carried out to detect fine differences on root secondary metabolome. Indeed, a work investigated, via a metabolic profiling approach (using UHPLC-DAD-QTOF), whether rice roots responded differently and with gradual intensity to the inoculation of a large panel of PGPR strains, isolated or not from different rice varieties (Valette et al. 2019). This, to see if there is a correlation plant response regarding the genetic distance of the inoculated plant and the original host plant used for bacterial isolation. Results revealed that instead of evidencing strain-specific metabolic changes, a common metabolomic signature of nine compounds was highlighted, with the reduced accumulation of three alkylresorcinols and increased accumulation of two hydroxycinnamic acid amides, identified as N-p-coumaroylputrescine and N-feruloylputrescine (Valette et al. 2019). Accumulation of hydroxycinnamic acid amides, that are potential antimicrobial compounds, might be considered as a primary reaction of plant to bacterial perception.

In conclusion, as we can see, metabolomics is a powerful tool to achieve an accurate study of plant secondary metabolism, in particular during bacteria-plant interactions. This approach can detect changes in secondary metabolites even when there are no visible effects on root architecture, that is, when there is no influence in the physiological functions linked to primary metabolism and development (Walker et al. 2011; Chamam et al. 2013). Also, when secondary metabolites modifications response is strain-dependent (Chamam et al. 2013) or even dependent to strain and plant cultivar combination, suggesting that this approach can detect fine-tuned mechanisms (Walker et al. 2011). With a metabolomic approach secondary metabolites modifications linked to fertilization level and to plant cultivar can also be detected (Walker et al. 2012, 2013).

Chapitre

2

Implication des gènes spécifiques d'*A. fabrum* sur l'adaptation à la plante

Préambule chapitre 2

La définition des espèces bactériennes est basée sur des similitudes génomiques, donnant naissance au concept d'espèce génomique, concept sur lequel est basée la définition opérationnelle et officielle des espèces bactériennes (ref. SVP). Ainsi, une espèce bactérienne est une population de bactéries génomiquement cohérentes dont tous les membres sont issus d'un même ancêtre commun. Pour comprendre la spéciation bactérienne et les adaptations particulières à des niches spécifiques de chaque espèce génomique, le complexe d'espèce *Agrobacterium tumefaciens* a été utilisé. Au sein de ce complexe, chez l'espèce génomique *Agrobacterium fabrum* (génomovar G8) ont été recherchés par génomique comparative de gènes adaptatifs potentiels, c'est-à-dire des gènes codant des adaptations écologiques spécifiques à cette espèce et éventuellement impliquées dans le processus de spéciation.

Suite à ces travaux, nous avons émis l'hypothèse que les gènes spécifiques d'*A. fabrum* groupés en régions dites « spécifiques » déterminent aux membres de cette espèce une adaptation particulière à la plante. Pour vérifier notre hypothèse, ce chapitre a pour but d'étudier la valeur adaptative conférée par chacune des régions spécifiques d'*A. fabrum*. Pour cela, nous avons utilisé des mutants de délétion de chacune des régions spécifiques d'*A. fabrum* C58 (souche-type de cette espèce) pour réaliser des compétitions *in planta* avec la souche sauvage.

Nous avons alors développé une approche méthodologique présentée dans la première partie de ce chapitre, pour mettre au point une méthode efficace, robuste et facile à mettre en œuvre pour repérer les plus fines différences de valeurs adaptatives susceptibles d'exister entre souches quasi isogéniques. Ainsi, dans la seconde et troisième partie de ce chapitre, nous avons estimé la valeur adaptative de chacun des mutants en calculant leurs indices de compétition (Macho et al. 2007) *in vitro* et *in planta* en conditions gnotoxéniques. Les tests *in planta* visaient à mesurer la valeur adaptative des mutants dans les deux styles de vie propres à *A. fabrum*. D'une part la vie commensale dans la rhizosphère, commune à la majorité des agrobactéries, et d'autre part la survie dans les tumeurs végétales que peuvent provoquer les agrobactéries pathogènes telles qu'A. fabrum C58. Afin de tenir compte du fait que la vie rhizosphérique s'exerce dans une communauté microbienne complexe, la valeur adaptative des mutants a été mesurée également en ajoutant d'autres espèces du complexe *A. tumefaciens* connues pour co-exister avec *A. fabrum* dans la rhizosphère des plantes (ref. SVP).

Pour des raisons de commodités méthodologiques, l'adaptation à la vie commensale a été testée avec *Medicago truncatula* (luzerne tronquée ou luzerne faux-tribule), une plante de petite taille facilement manipulable *in vitro.* Cela nous a permis de maitriser un maximum de paramètres, en s'affranchissant en particulier du recours à la culture dans un sol naturel pour lequel il est très difficile, voire impossible, d'avoir un microbiote identique d'un essai à l'autre. D'autre part, l'adaptation à la survie dans les tumeurs, requérant des temps de culture beaucoup plus longs (e.g. les tumeurs apparaissent 3 semaines après l'inoculation de plantes âgées d'un mois), a dû être testée en serre, dans des conditions gnotoxéniques. En effet, quand considère que les infections ont été réalisées sur la tige de la plante, l'inoculum apportant techniquement beaucoup plus d'agrobactéries qu'il ne pourrait y avoir de micro-organismes épiphylles aux points d'inoculation, mais toutefois pas en conditions totalement contrôlées. Ces essais ont été conduits sur des tomates, plantes usuellement utilisées pour tester le pouvoir pathogène des agrobactéries qui ont de plus l'avantage sur la luzerne, de produire des tumeurs de taille suffisante pour les besoins de l'étude.

I. Mesure de la valeur adaptative conférée par les gènes spécifiques d'*A. fabrum*

Introduction

Pour comprendre la grande diversité des bactéries, l'hypothèse courante est d'évoquer l'adaptation des bactéries à leur environnement. Pour cela, il faut considérer que la diversité existante est le résultat d'une évolution complexe. Chez une même espèce bactérienne, la trace de leur adaptation au cours de l'évolution est ancrée dans le génome de tous les membres de l'espèce. L'espèce bactérienne est actuellement définie comme un groupe d'isolats ayant une forte proximité génomique et auquel il est possible d'attribuer un phénotype particulier (Moore *et al.* 1987). Cette définition est parfaitement opérationnelle en taxonomie mais ne nous permet pas de comprendre comment et pourquoi les espèces se sont différenciées. Une hypothèse qui peut être émise pour expliquer la spéciation bactérienne est l'isolement écologique, où des espèces présentes interagissent peu les unes avec les autres, même si elles coexistent sur une même aire de répartition (Schluter et al. 2009). Des mutations adaptatives chez un individu peuvent donc être à l'origine de l'exploitation de nouvelles ressources et, par la suite, de l'adaptation à une nouvelle niche écologique par sa descendance (Cohan 2002).

Comprendre la spéciation bactérienne, c'est-à-dire comprendre le phénomène qui a conduit à l'individualisation d'une telle population génomiquement homogène, revient donc à comprendre les facteurs qui ont favorisé la descendance de cet ancêtre commun. Pour des bactéries qui évoluent dans des écosystèmes riches et diversifiées comme le sol, l'hypothèse émise par Lassalle *et al.* 2011 est que l'ancêtre commun d'une espèce donnée avait un, ou des avantages adaptatifs favorisant sa descendance. Pour ce fait, il doit subsister dans le cœur génomique des espèces bactériennes, des traces de ces adaptations ancestrales (Lassalle et al. 2011). L'existence de telles traces ont été montrées dans le génome d'une espèce bactérienne modèle, *Agrobacterium fabrum*, à partir desquelles un modèle théorique de la niche spécifique de cette espèce a été proposé (Lassalle et al. 2011). L'hypothèse générale de ce travail est que la niche écologique initiée par ce modèle théorique confère encore aujourd'hui un avantage adaptatif spécifique aux membres de cette espèce en lui apportant une meilleure valeur adaptative.

La valeur adaptative est un concept central en théorie de l'évolution. Il est bien connu que sans différence de valeur adaptative, la sélection naturelle ne peut pas agir et donc, l'adaptation ne peut

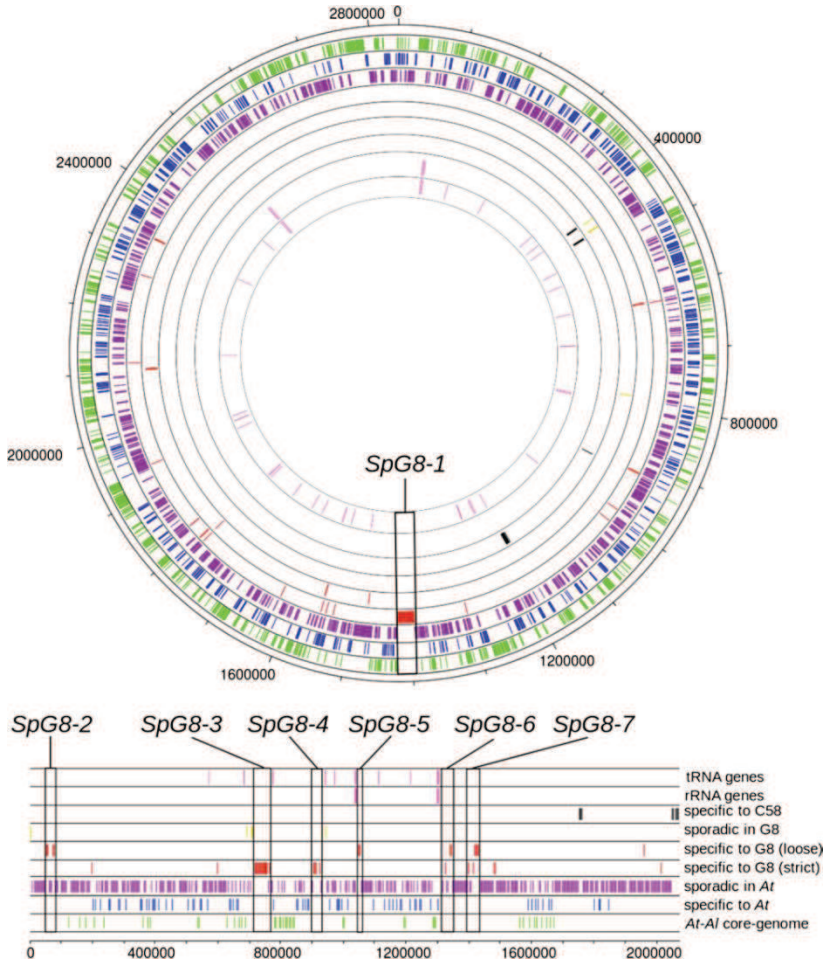
pas se produire. La valeur adaptative peut se référer à un gène, un génotype, un individu, une espèce ou une population (3, 4). Le terme valeur adaptative ou « fitness » a été utilisé pour indiquer une mesure d'adaptabilité et de succès reproducteur (1). Mais elle a aussi une signification évolutive, de par sa contribution génétique aux générations suivantes, et une importance écologique, par sa contribution à la croissance d'une population. La valeur adaptative est donc le succès de la reproduction d'un génotype dans un environnement particulier (6). Pour mesurer la valeur adaptative, la biologie évolutive s'est principalement appuyée sur des études comparatives qui sont de plus en plus puissantes avec l'émergence de nouvelles techniques. Par conséquent, c'est la différence de valeur adaptative, dite relative, dépendant non seulement du génotype, mais aussi de l'environnement et du moment de la mesure qui compte généralement. La plupart des expériences en génétique microbienne ont procédé par la perturbation des fonctions. Ceci a été productif pour identifier des gènes qui codent de nombreuses fonctions biochimiques et physiologiques (6).

Le complexe d'espèces « *Agrobacterium tumefaciens »* désigne les agrobactéries qui possèdent à la fois une grande diversité d'espèce et une grande diversité intraespèce (Mougel et al. 2002; Portier et al. 2006). Les agrobactéries sont avant tout des organismes indigènes du sol capables d'établir des interactions commensales voire bénéfiques dans la rhizosphère des plantes. Bien que les agrobactéries soient connues pour être phytopathogènes en induisant la galle du collet, la pathogénie n'est qu'un trait écologique secondaire lié à la présence d'un plasmide accessoire, le plasmide Ti *(Tumor inducing)* qui porte les gènes de virulence (Watson *et al.* 1975). Des études ont montré qu'une même rhizosphère pouvait héberger plusieurs espèces d'agrobactéries vivant en sympatrie (ref). En accord avec le principe d'exclusion compétitive, ces espèces, pourtant très apparentées, doivent exploiter des ressources différentes pour échapper à la compétition et ainsi durablement co-exister (Garrett Hardin 1960). Ainsi, selon l'hypothèse émise par Lassalle *et al.* en 2011, ces espèces doivent posséder des gènes propres leur permettant d'exploiter des niches écologiques au-moins en partie différentes. Les gènes spécifiques de l'espèce G8, appelée *Agrobacterium fabrum* ont été mis en évidence par génomique comparative. Il s'agit de 196 gènes groupés en 7 régions génomiques appelées « régions Spécifiques de G8 » ou « SpG8 » (nommées SpG8-1 à SpG8-7). Une de ces régions est située sur le chromosome circulaire et les six autres sont situées sur le chromosome linéaire d'*A. fabrum* **(Figure 1A)**. L'annotation de ces régions **(Figure 1B)** indique qu'elles codent des unités fonctionnelles hypothétiques dont certaines (SpG8-1b, SpG8-2a et SpG8-3) ont été confirmées expérimentalement. Outre qu'elles définissent à priori la niche spécifique d'*A. fabrum* certaines de ces régions déterminent des traits phénotypiques propres à cette espèce et utilisables pour sa caractérisation morpho-biochimique (Campillo et al. 2012). De plus, ces régions spécifiques suggèrent une implication étroite

avec la plante, ce pourquoi nous faisons l'hypothèse que les gènes spécifiques d'*A. fabrum* lui confèrent une meilleure valeur adaptative dans ses interactions avec cette dernière.

L'objectif de cette étude est d'évaluer l'implication des régions spécifiques d'*A. fabrum* dans ses deux styles de vie dans la plante : l'adaptation à la vie commensale dans la rhizosphère (avec pour modèle la rhizosphère de la luzerne, majoritairement colonisée par cette bactérie) (Ponsonnet 1994)) et l'adaptation dans la tumeur produite par cette bactérie (symptôme de la galle du collet, avec pour modèle la tomate). Pour cela nous avons étudié la valeur adaptative conférée par chacune des régions spécifiques dans ces deux biotopes lors de compétitions avec la souche sauvage C58. En première instance nous avons dû mettre en place une méthodologie capable de calculer la valeur adaptative toute en distinguant les différentes souches mises en compétition. Nos résultats permettent de mieux comprendre la niche écologique spécifique d'*A. fabrum* et donc comment son écologie spécifique lui permet d'échapper à la compétition avec les espèces sympatriques les plus apparentées.

A



B

Région spécifique	Fonction prédite
SpG8-1a	Transport et métabolisme de sucres
SpG8-1b	Dégradation d'acides hydroxycinnamiques
SpG8-2a	Biosynthèse d'exopolysaccharides (curdlan)
SpG8-2b	Biosynthèse de métabolites secondaires
SpG8-3	Biosynthèse de sidérophores
SpG8-4	Transport et métabolisme de sucres
SpG8-5	Catabolisme de composés de type opines
SpG8-6	Détoxification (tétracycline)
SpG8-7	Systèmes mécano-senseurs

Figure 1. Localisation des régions spécifiques dans les chromosomes d'*A. fabrum* C58 avec leurs fonctions hypothétiques (d'après Lassalle *et al.* 2011). A. Représentation du chromosome circulaire et linéaire d'*A. fabrum* avec la localisation des régions spécifiques. B. Annotation des fonctions des régions spécifiques

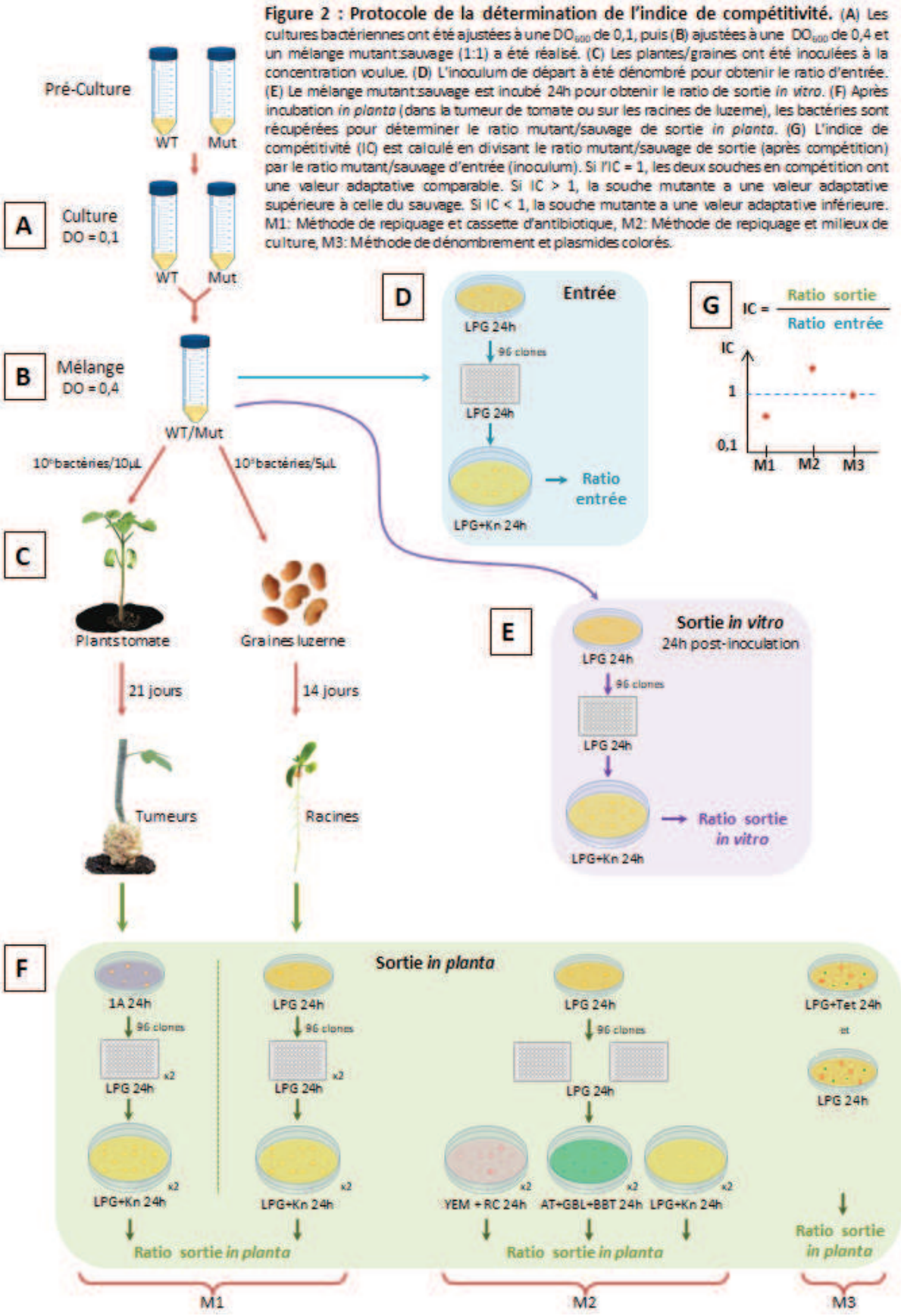


Figure 2 : Protocole de la détermination de l'indice de compétitivité. (A) Les cultures bactériennes ont été ajustées à une DO_{600} de 0,1, puis (B) ajustées à une DO_{600} de 0,4 et un mélange mutant:sauvage (1:1) a été réalisé. (C) Les plantes/graines ont été inoculées à la concentration voulue. (D) L'inoculum de départ à été dénombré pour obtenir le ratio d'entrée. (E) Le mélange mutant:sauvage est incubé 24h pour obtenir le ratio de sortie *in vitro*. (F) Après incubation *in planta* (dans la tumeur de tomate ou sur les racines de luzerne), les bactéries sont récupérées pour déterminer le ratio mutant/sauvage de sortie *in planta*. (G) L'indice de compétitivité (IC) est calculé en divisant le ratio mutant/sauvage de sortie (après compétition) par le ratio mutant/sauvage d'entrée (inoculum). Si l'IC = 1, les deux souches en compétition ont une valeur adaptative comparable. Si IC > 1, la souche mutante a une valeur adaptative supérieure à celle du sauvage. Si IC < 1, la souche mutante a une valeur adaptative inférieure. M1: Méthode de repiquage et cassette d'antibiotique, M2: Méthode de repiquage et milieux de culture, M3: Méthode de dénombrement et plasmides colorés.

Matériels et Méthodes

Souches bactériennes, plasmides, conditions de culture et préparation des inocula

Les souches bactériennes et plasmides utilisés lors de cette étude sont présentés dans le **Tableau 1**. Les agrobactéries ont été cultivées à 28°C en trois milieux différents : LPG (5 g/L d'extrait de levure, 10 g/L de glucose et 5 g/L de peptone, à pH 7,2) liquide sous agitation, ou LPGA (LPG amendé de 15 g/L d'agar) supplémentés le cas échéant avec de la kanamycine à 25 µg/mL et néomycine à 25 µg/mL. Le milieu minimum AT **(Annexe 1)** supplémenté en azote (40 mM) et additionné d'une source de carbone, la γ-butyrolactone (43 mg/L), et d'un colorant, le bleu de bromothymol (50mg/L). Et le milieu YEM-Rouge Congo pH 6,8 (10 g/L mannitol, 0,5 g/L K2HPO4, 0,1 g/L NaCl, 0,2 g/L MgSO4 7H2O, 0,4 g/L extraits de levure, 3 g/L CaCO3, agar 15 g/L, 0,025g/L Rouge Congo).

Les inoculum ont été préparés à partir des pré-cultures incubés pendant la nuit à 28°C dans 5 mL de milieu LPG liquide, puis ajustées à une DO_{600} de 0,1 et réincubées à 28°C pour atteindre la phase exponentielle de croissance **(Figure 2A)**. Les cultures ont été ensuite ajustées à une DO_{600} de 0,4 avant de mélanger chaque mutant individuellement avec la souche sauvage avec un rapport 1:1 **(Figure 2B)**.

Cultures de *Medicago truncatula*

Les graines de luzerne (*Medicago truncatula* cv. Jemalong, lignée J5), aimablement fournies par Ch. Mougel [INRAe], ont été stérilisées par un bain de 25 min dans une solution d'hypochlorite de sodium à 3,5% avant d'être abondamment rincées avec de l'eau UP stérile et vernalisées par un passage à 4°C pendant 7 jours. Deux jours avant l'inoculation, les graines ont été scarifiées à l'aide d'un scalpel stérile, puis déposées sur milieu gélosé à 0,8% (Agar-plante, Sigma Aldrich, St Louis, Etats-Unis) stérile supplémenté avec une solution nutritive commerciale diluée au 1/1000 (Plant-Prod 15-15-30, Fertil s.a.s, Le Syndicat, France) dans des boîtes de Pétri stériles. La partie inférieure des boites a ensuite été recouverte par du papier aluminium pour mettre les racines à l'abri de la lumière. La germination des graines s'est déroulée dans les boîtes disposées verticalement pendant 2 nuits à l'obscurité et à température ambiante. Les plantes au stade « racine émergente » ont ensuite été inoculées individuellement par 10^3 cellules apportées par 5 µL de l'inoculum mutant:sauvage en proportion (1:1) **(Figure 2C)** puis mises à incuber verticalement dans un phytotron pendant 14 jours à 20°C et une photopériode de 16 h de jour et 8 h de nuit. Le phytotron a été ajusté à 75% d' humidité afin de limiter l'évaporation des boites où l'humidité était à saturation. Après 14 jours d'incubation,

les racines ont été récupérées, rincées avec de l'eau stérile, broyées et mises en suspension dans 1mL de NaCl 0,8% avant de procéder au comptage des bactéries présentes.

Cultures de tomate et vérification du pouvoir pathogène des souches

Des plants de tomates (*Solanum lycopersicum* variété Marmande) âgés d'un mois ont été inoculés avec 10^6 cellules apportées par 10 µL de l'inoculum mutant:sauvage (1:1) au point d'inoculation effectué par une incision avec un scalpel stérile au niveau du collet **(Figure 2C)**. Les plants ont été ensuite incubés en serre pendant 3 semaines avant d'être sacrifiés pour collecter les tumeurs qui ont été ensuite broyées individuellement dans 1 mL de NaCl 0,8% pour récupérer les bactéries. Chaque essai comportait au moins 4 réplicats. La pathogénie de chaque souche mutante a été vérifiée individuellement en procédant de la même façon. Cette expérience a été réalisée en quadruplicatas pour chacune des souches utilisées.

Calcul de l'indice de compétitivité (IC)

Les tests de compétition ont été réalisés en utilisant des mélanges mutant:sauvage (1:1). Pour chaque essai, la proportion des souches dans l'inoculum a été déterminée afin de vérifier l'exactitude du ratio d'entrée **(Figure 2D)**. Pour cela, les mélanges ont été dilués en série, successivement au 1/10, dans un volume final de 1 mL de NaCl 0,8% (10^{-1} à 10^{-5}). Afin d'obtenir des colonies isolées, 100 µL de la dilution 10^{-5} ont été étalés sur LPGA avec l'ensemenceur automatique EasySpiral (Interscience, France) et incubés 48 h à 28°C. Pour chaque essai, 192 colonies ainsi obtenues ont été transférées à l'aide de cure-dents stériles dans des microplaques 96 puits contenant 200 µL de milieu LPG liquide par puits et ensuite mises à incuber 24 h à 28°C. Pour l'identification des colonies de souches mutantes (qui possèdent le gène nptII), l'ensemble des colonies de chaque microplaque a ensuite été transféré à l'aide d'un réplicateur constitué d'un peigne métallique à 96 dents sur un milieu LPGA amendée en kanamycine 25 µg.mL^{-1} et néomycine 25 µg.mL^{-1}. Une incubation de 24 h à 28°C a été réalisée pour déterminer finalement, la proportion relative des souches mises en compétition et ainsi obtenir le ratio mutant/sauvage d'entrée. Le mélange mutant:sauvage (1:1) a ensuite été réincubé à 28°C pendant 24 h pour ainsi obtenir de la même façon le ratio de sortie *in vitro* **(Figure 2E)**.

De la même manière, près la phase d'incubation *in planta* (dans la tumeur de tomate ou sur les racines de luzerne), les bactéries ont été récupérées pour déterminer le ratio mutant/sauvage de sortie *in planta* **(Figure 2F)**. Les milieux 1A **(Annexe 2)** ou LPGA ont été utilisés pour étaler la dilution 10^{-3} du broyat de tumeurs de tomates ou de racines de luzerne, respectivement. Ceci, afin d'éviter les éventuels biais de comptage qui pourraient résulter de l'adjonction d'un antibiotique dans le milieu,

l'identification des souches sur la base de leur capacité à croître sur un milieu contenant l'antibiotique sélectif a toujours été effectuée seulement après une première phase de culture non sélective. Après 24h d'incubation à 28°C, 192 colonies ont été récupérées avec des cure-dents stériles et chacune transférée dans un puits différent d'une plaque de 96 puits contenant du LPG liquide. Après 24h d'incubation à 28°C sous agitation, l'ensemble de colonies de chaque microplaque a ensuite été transféré à l'aide d'un réplicateur constitué d'un peigne métallique à 96 dents sur différents milieux gélosés afin de poursuivre avec l'identification des souches. Pour les compétitions avec uniquement des dérivés d'*A. fabrum* C58, les colonies provenant des plaques de 96 puits ont été transférées dans du LPG additionnée de kanamycine (25µg/mL) et de néomycine (25µg/mL) et incubées 24h à 28°C pour être ensuite dénombrées (**Figure 2F**). Pour les compétitions en présence d'autres souches appartenant à d'autres espèces (G1 et G4), trois milieux différents ont été utilisés pour l'identification des souches : une gélose de YEM supplémentée de Rouge Congo (permettant de différentier G8 et G1 de G4), une gélose d'AT supplémentée avec du GBL et du BBT (permettant de différentier G8 des G1 et G4) et une gélose de LPG supplémenté avec de la néomycine et de la kanamycine (permettant de différentier *A. fabrum* WT des souches mutantes). Les boites de Pétri ont ensuite été incubées 24h à 28°C puis dénombrées (**Figure 2F**). Pour les compétitions utilisant des souches marquées par des plasmides codant la production d'un fluorophore, le milieu LPGA supplémentée de 2µg/mL de tétracycline a été utilisée pour étaler directement la dilution 10^{-3} du broyat de racine de luzerne et ainsi récupérer que les souches contenant le plasmide construit (*A. fabrum* C58 et les mutants de délétion). Le milieu LPGA sans antibiotique a été utilisé pour étaler la dilution 10^{-5} du broyat de racine de luzerne et ainsi récupérer l'ensemble des souches mises en compétition (G1, G4 et les G8). De cette manière, le dénombrement de toutes les souches a été possible. Toutes les boites ainsi étalées ont été incubées 24h à 28°C pour être ensuite dénombrées (**Figure 2F**).

L'IC **(Figure 2G)** est défini comme le changement dans le ratio de deux souches après leur mise en compétition. Il est calculé grâce au ratio de sortie (bactéries récupérées en fin d'essais après compétition) entre les deux souches testées, divisé par leur ratio d'entrée (mélange mutant:sauvage (1:1)). Si l'IC est égale à 1, les deux souches en compétition ont une valeur adaptative comparable. Si IC > 1, la souche mutante est plus compétitive, indiquant que sa valeur adaptative est supérieure à celle du sauvage. Si IC < 1, la souche mutante est moins compétitive que la sauvage, et donc, sa valeur adaptative est inférieure.

Construction de marqueurs plasmidiques (*uidA*, pME 6010 *gfp* /*red*)

Le gène *uidA* a été amplifié et introduit au plasmide pBBR entre les sites HindIII et XbaI, puis introduit chez *A. fabrum* par électroporation. Les plasmides pME6010-gfp et pME6010-red ont été introduits chez *A. fabrum* par électroporation.

Analyse statistique

Les indices de compétitivité ont été analysés par un test de Wilcoxon en prenant comme hypothèse nulle : la moyenne réelle n'est pas significativement différente de 1 ($P<0,05$). L'ensemble de ces tests a été réalisé à l'aide du logiciel R-studio et du package Agricolae.

Mise au point Méthodologique

Pour pouvoir étudier la valeur adaptative conférée par les différents régions spécifiques d'*A. fabrum*, nous avons du tout d'abord mettre en place une méthode fiable et efficace pour la calculer.

a. Repiquage de colonies

La première méthode utilisée à été le repiquage de colonies à l'aide des cure-dents stériles. En effet, après l'étape d'étalement, soit de l'inoculum d'entrée (**Figure 2D**) ou du broyat végétal après compétition (**Figure 2F**), 192 colonies ont été prélevées à l'aide des cure-dents stériles pour être placées individuellement dans un puits d'une plaque de 96 puits.

Il a été possible d'utiliser cette méthode grâce à la présence de *nptII*, un gène de résistance à la kanamycine portée dans le génome chez tous les mutants de délétion des régions spécifiques d'*A. fabrum*. Ainsi, après l'étape de repiquage et d'incubation en plaques de 96 puits, les colonies sont transférées sur le milieu LPGA contenant de la kanamycine où seulement les souches mutantes d'*A. fabrum* vont pousser. Ceci nous permet alors de distinguer les deux types de souches mises en compétition, étape cruciale lors des analyses de valeur adaptative.

i) Coût de la cassette *nptII*

Pour pouvoir utiliser cette méthode, nous avons dû nous assurer que la cassette codant la résistance à la kanamycine ne conférait aucun coût supplémentaire à la bactérie, et de ce fait, lui apporter un poids à sa valeur adaptative. Pour tester cela, nous avons mis en compétition une souche sauvage C58 avec une autre souche C58 contenant le gène *nptII* dans son génome, d'abord *in vitro*

dans un milieu riche (LPG), puis *in planta* (dans la tumeur de tomate induite par *A. fabrum*). Les résultats de ces tests nous montrent que l'indice de compétitivité (IC) *in vitro* et in planta ne sont pas significativement différent de 1 (**Figure 3**). Nous en concluons que la présence du gène *nptII* sur le génome bactérien n'a pas d'effet significatif sur la valeur adaptative d'*A. fabrum*.

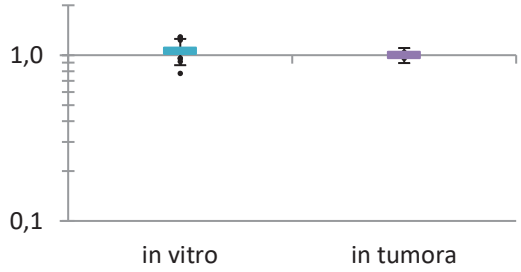


Figure 3. Indices de compétition *in vitro* et *in planta.* Les données présentées correspondent aux indices de compétition *in vitro* et *in planta* (dans la tumeur de tomate induite par *A. fabrum*) des souches sauvages C58 ,dont une qui porte le gène nptII, avec un ratio 1:1. Les barres d'erreur correspondent aux intervalles de confiance.

Cette méthode a été très utile et a démontré être très efficace pour faire nos premiers dénombrements et calculs de valeur adaptative, mais elle s'est révélée très couteuse en termes de temps. En effet, nous avons réalisé 2 plaques de 96 puits par replicat et au moins 4 replicats par condition, ce qui fait une quantité considérable de colonies à répliquer. Pour nous épargner du repiquage et ainsi calculer des indices de compétitions de façon plus rapide mais toute aussi efficace, nous avons cherché d'autres méthodes. En plus, nous voulions réaliser également des compétitions en présence d'autres souches appartenant à d'autres espèces génomiques du genre *Agrobacterium* et ainsi étudier la réponse des gènes spécifiques dans une communauté plus diverse. Pour cela, une méthode distinguant l'ensemble de souches mises en compétition est nécessaire.

b. Utilisation de milieux de culture

Pour développer une méthode rapide et efficace pour le calcul de la valeur adaptative, il est crucial de pouvoir distinguer facilement l'ensemble de souches mises en compétition. En effet, nous souhaitons utiliser un mélange complexe contenant plusieurs espèces d'agrobactéries (l'espèce génomique G8 à laquelle appartient *A. fabrum*, mais aussi les espèces G1 et G4). Pour cela nous avons tout d'abord cherché à utiliser leurs propriétés biologiques intrinsèques en utilisant plusieurs milieux de culture contenant différentes sources de carbone. En effet, nous allons profiter des différentes propriétés des souches utilisées pour les distinguer lors de la mise en compétition. Les trois milieux utilisés sont les suivants :

Milieu YEM-Rouge Congo. Ce milieu a été utilisé car *A. fabrum* produit un exo-polysaccharide particulier, le curdlan (grâce à sa région spécifique SpG8-2a), qui peut être mis en évidence par l'ajout de Rouge Congo au milieu de culture (Lassalle et al. 2011), même s'il s'est avéré plus tard que l'espèce G1 possède également cette propriété. Les bactéries produisant ce glucane peuvent donc fixer le colorant, ce qui se traduit par la coloration de la colonie en rouge. Nous avons obtenu donc sur ce milieu les espèces G1 et G8 (WT et mutants) colorées en rouge, tandis que l'espèce G4 est incolore **(Figure 4A)**.

Milieu AT-GBL/BBT. Nous avons utilisé un milieu AT (milieu minimum), additionné d'une source de carbone, la γ-butyrolactone, et d'un colorant, le bleu de bromothymol. L'utilisation de γ-butyrolactone comme source de carbone est codée par le plasmide At chez C58. Les essais ont montré que l'ensemble des agrobactéries utilisées lors de cette étude poussent sur ce milieu. Cependant, il a été possible de les différentier grâce à la couleur des colonies. En effet, *A. fabrum* (G8) reste incolore sur ce milieu, tandis que les espèces G1 et G4 l'acidifient et le colorent en jaune **(Figure 4B)**.

Milieu LPG-Neo/Kn. Les mutants de délétion des régions spécifiques portent le gène nptII codant la résistance à la kanamycine. De ce fait, c'est seulement ceux-ci qui poussent sur ce milieu **(Figure 4C)**.

L'ensemble de ces résultats rend possible donc d'identifier l'identité du mélange des espèces utilisées lors de l'étalement sur boîte à la fin de chaque compétition. Cependant cette approche ne nous permet pas de nous épargner de l'étape de repiquage de colonies.

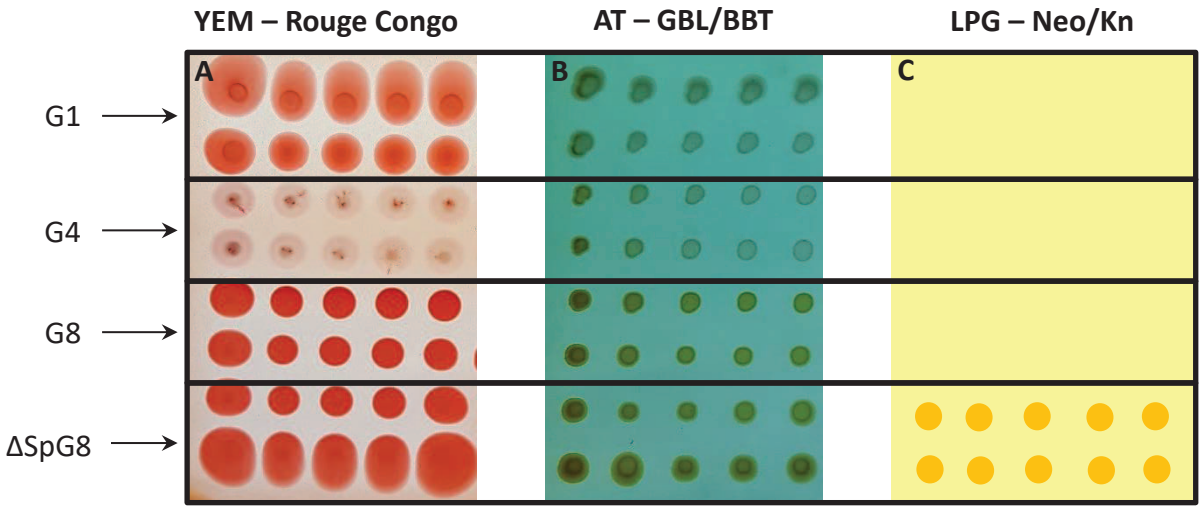


Figure 4. Milieux différentiels utilisés. Trois milieux différents ont été utilisés pour repérer l'identité de chacune des espèces utilisées. Les colonies de l'espèce G1 deviennent rouges sur milieu YEM-RC, restent incolores sur milieu AT-BGL/BBT et ne peuvent pas pousser sur milieu LPG-Neo/Kn. Les colonies de l'espèce G4 restent incolores à la fois sur le milieu YEM-RC et sur milieu AT-BGL/BBT et ne peuvent pas pousser non plus sur milieu LPG-Neo/Kn. Les colonies de l'espèce G8 (*A. fabrum*) deviennent rouges sur milieu YEM-RC et jaune sur milieu AT-BGL/BBT. La souche sauvage de cette espèce ne peut pas pousser sur milieu LPG-Neo/Kn, mais les mutants de délétion des régions spécifiques utilisés le peuvent grâce à une cassette de résistance à la kanamycine introduit dans leur génome.

Le ratio d'entrée des souches que nous avons utilisé a été le suivant :

Espèce	Souche	Ratio
G1	CFBP 5771	1/6
G4	CFBP 5621	1/6
G8	C58 WT	1/3
(*A. fabrum*)	Mutants de délétion des régions spécifiques	1/3

Cette approche n'a pas pu être appliquée pour nos tests de valeur adaptative. Nous avons trouvé que l'espèce G4 (souche CFBP 5621) exerce une très grande pression de sélection sur le reste des souches utilisées en affectant sévèrement leur compétitivité. En effet, lors de la sortie *in vitro* et *in planta*, la souche G4 était beaucoup plus abondante que les autres espèces utilisées. Ainsi, près de la moitié des colonies récupérées appartenaient à l'espèce G4 pour la sortie *in vitro* **(Figure X)** et sur 96 colonies, 85 à 90 étaient des G4 dans les sorties *in planta*. Ces résultats montrent qu'il est difficile voire impossible de récupérer une quantité suffisante des autres espèces, et en particulier de G8, pour calculer convenablement leur IC.

Figure X. Nombre de colonies obtenues après compétitions in vitro. Pour les compétitions, trois espèces différentes d'agrobactéries ont été utilisés. La souche CFBP 5771 appartenant à l'espèce G1, la souche CFBP 5621 appartenant à l'espèce G4 et la souche sauvage C58 et les mutants de délétion des régions spécifiques appartenant tous à l'espèce G8.

c. Utilisation de marqueurs plasmidiques

Pour palier les difficultés rencontrées avec l'utilisation de milieux de culture, nous proposons l'utilisation de marqueurs plasmidiques dont les gènes ou protéines exprimées sont visibles à l'œil nu. Ainsi, nous nous épargnons de l'étape de repiquage de colonies, mais aussi cela permet la distinction de souches, étape cruciale lors des essais de compétitions surtout au sein d'un mélange complexe d'espèces. De plus, la cassette d'antibiotique présente dans le plasmide, permet de récolter seulement les souches le contenant choisies préalablement.

Utilisation du pBBR

Pour développer une méthode rapide, efficace et capable de distinguer facilement un mélange de souches mises en compétition, nous avons envisagé l'utilisation de marqueurs plasmidiques dont les protéines exprimées sont visibles à l'œil nu. En effet, un élément essentiel est de pouvoir repérer chacune des souches mises en compétition dans un temps de manipulation réduit.

Pour cet effet, nous avons utilisé le gène de la β-glucuronidase (GUS) codé par le locus *uidA* chez *E. coli* (Novel and Novel, 1973) que nous avons introduit dans le plasmide pBBR déjà disponible dans l'équipe et utilisé pour la construction de souches de complémentation. Ce gène est connu pour être utilisé comme marqueur d'expression génique, cependant, nous nous intéressons à son activité enzymatique. En effet, il s'agit d'une hydrolase qui catalyse le clivage d'une grande variété de β-glucuronides et produit un précipité bleu insoluble, conférant une coloration bleue à la bactérie. C'est grâce à cette propriété que le temps de manipulation est largement réduit, car les colonies peuvent être comptées à l'œil nu après leur étalement sur boîte. Nous avons utilisé une deuxième construction (construite au sein de l'équipe) en utilisant le même vecteur d'expression, le plasmide pBBR, auquel le gène *M-Cherry* a été introduit. Ce gène est également utilisé comme gène rapporteur pour l'étude de l'expression des gènes. Néanmoins, nous nous intéressons à la propriété du fluorophore produit. En effet, cette protéine fluorescente est également visible à l'œil nu, ce qui nous permettra de faire la distinction des souches mises en compétition par le même principe qu'avec le gène *uidA*.

ii) Compétitions *in vitro*

Les tests de compétition *in vitro* nous montrent une nette distinction entre les souches contenant ou pas le vecteur construit avec le gène de la β-glucuronidase **(Figure 5A)**. En effet, nous pouvons observer une coloration bleue des colonies bactériennes portant le gène *uidA*. Ensuite, l'indice de compétitivité (IC) *in vitro* n'étant pas significativement différent de 1 **(Figure 6)**, nous en concluons que la construction réalisée n'a pas d'effet significatif sur la valeur adaptative de la bactérie. De même, les tests de compétition *in vitro* nous montrent une coloration rouge des colonies bactériennes porteuses du plasmide avec le gène *M-Cherry*, en contraste de celles qui ne le possèdent pas **(Figure 5B)**. L'IC *in vitro* mesuré n'étant pas significativement différent de 1 **(Figure 6)**, nous en concluons que la construction réalisée n'a pas d'effet significatif sur la valeur adaptative de la bactérie.

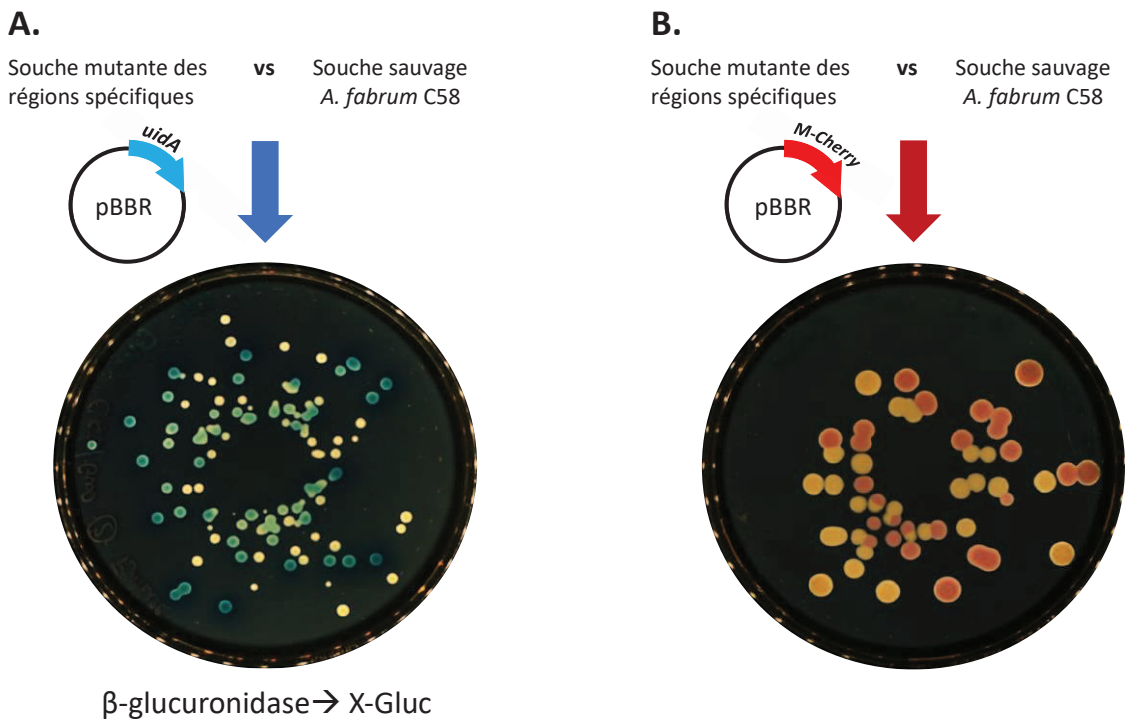


Figure 5. Construction de marqueurs plasmidiques sur le pBBR pour la distinction des souches. A. Le gène *uidA* a été introduit au plasmide pBBR et utilisé chez les souches mutantes de délétion des régions spécifiques d'*A. fabrum* (coloration bleue liée à la formation d'un précipité bleu insoluble grâce à l'ajout de X-Gluc dans le milieu de culture) lors de la mise en compétition avec la souche sauvage. **B.** Le gène *M-Cherry* a été introduit au plasmide pBBR et utilisé chez les souches mutantes de délétion des régions spécifiques d'*A. fabrum* (coloration rouge liée à l'expression du fluorophore produit) lors de la mise en compétition avec la souche sauvage. La souche C58 WT utilisée avec chacune des compétitions ne possédant pas de plasmide et reste donc incolore.

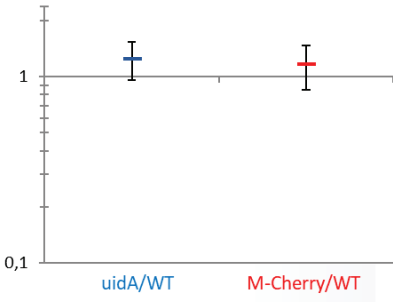


Figure 6. Indices de compétition *in vitro*. Les données présentées correspondent aux indices de compétition *in vitro* des constructions réalisées (pBBR-*uidA* et pBBR-*M-Cherry*) co-inoculées individuellement avec la souche sauvage C58 avec un ratio 1:1. Les barres d'erreur correspondent aux intervalles de confiance.

i) Stabilité du plasmide *in planta*

Nous avons poursuivi avec l'analyse *in planta*. Pour ce faire, comme avec toute utilisation de plasmides, nous devions nous assurer avant tout de la stabilité ou conservation des plasmides *in planta*. Pour cela, nous avons inoculé des plants de tomate avec les souches construites et nous avons obtenu un pourcentage de conservation des plasmides très faible. En effet, le plasmide pBBR vide possède un pourcentage de conservation de 5%, le plasmide pBBR-uidA 9% et le plasmide pBBR-M-Cherry uniquement 3%. Cela nous indique que l'utilisation de ces constructions *in planta* n'est pas envisageable, car la perte du plasmide est très importante. D'après d'autres tests effectués au sein de

l'équipe, cette instabilité est propre au plasmide pBBR. Nous devons donc éviter d'utiliser *in planta* ce plasmide ou ses dérivés pour *A. fabrum*. Nous devons nous orienter vers l'utilisation d'une autre technique pour la différentiation des souches mises en compétition ou l'utilisation d'un autre plasmide très stable *in planta*.

Utilisation du pME6010

Nous avons utilisé le plasmide pME6010 qui possède une cassette de résistance à la tétracycline auquel deux marqueurs plasmidiques différents, soit la *gfp*, soit *red*, ont été introduits. Ainsi, grâce à l'exposition sous la lumière bleue de la boite de Pétri sur laquelle ont été étalées ces constructions, nous pouvons voir à l'œil nu les deux couleurs : vert fluo pour la GFP et rouge pour la red, et de ce fait, faire la distinction des deux souches mise en compétition. Après avoir introduit les plasmides chez *A. fabrum* C58, soit le pME6010-*gfp* soit le pME6010-*red*, nous avons réalisé des tests préliminaires.

i) Compétition *in vitro*

Nous avons réalisé des compétitions *in vitro* indépendantes entre une bactérie *A. fabrum* C58 WT possédant le plasmide pME6010 *red* et une autre bactérie *A. fabrum* C58 WT possédant le pME6010 *gfp* (**Figure 7**). Nous avons trouvé un IC moyen du pME6010 *red* de 1,0058 (**Figure 8**). Nous avons alors constaté que ni la casette de *gfp* ni celle de *red*, apportent un poids supplémentaire à la bactérie.

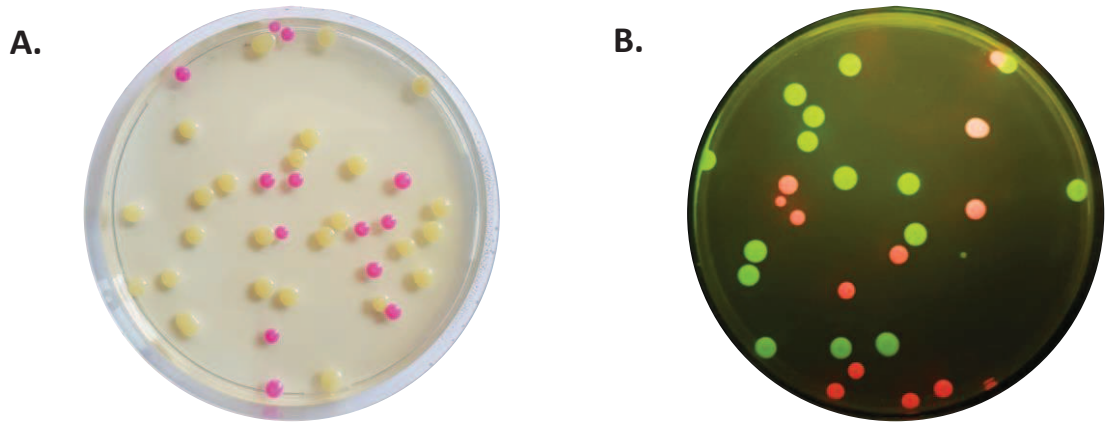


Photo à l'œil nu | Photo sous la lumière bleue

Figure 7. Construction de marqueurs plasmidiques sur le pME6010 pour la distinction des souches. Les gènes *gfp* et *red* ont été introduits séparément au plasmide pME6010 et ensuite introduits chez *A. fabrum C58*. Des compétitions *in vitro* ont ensuite été réalisées entre ces deux souches avec un ratio d'entrée de 1:1. **A.** Photo de l'étalement de la compétition à l'œil nu. **B.** Photo de l'étalement de la compétition sous la lumière bleue.

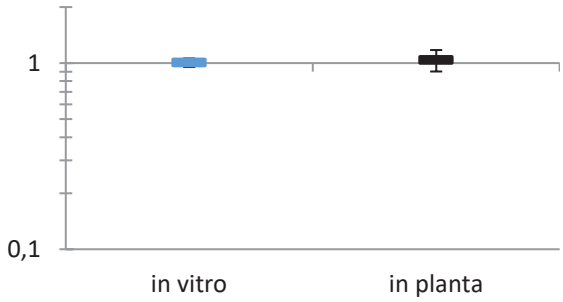


Figure 8. Indices de compétition *in vitro* et *in planta*. Les données présentées correspondent aux indices de compétition *in vitro* et *in planta* des souches sauvages A. fabrum C58 portant les constructions réalisées (pME6010 *gfp* et pME6010 *red*) co-inoculées avec un ratio 1:1. Les barres d'erreur correspondent aux intervalles de confiance.

iii) Stabilité du plasmide *in planta*

Nous avons poursuivi avec l'analyse *in planta* pour nous assurer de la stabilité du plasmide. Pour cela, nous avons inoculé des plantules de *Medicago* avec les souches WT construites et nous avons obtenu 100% de conservation des plasmides. Cela nous indique que l'utilisation de ces constructions est possible car le plasmide est très stable *in planta*.

iv) Compétition *in planta*

Nous avons également réalisé des compétitions indépendantes *in planta* (dans la rhizosphère de *M. truncatula*) entre une bactérie *A. fabrum* C58 WT possédant le plasmide pME6010 *red* et une autre bactérie *A. fabrum* C58 WT possédant le pME6010 *gfp* (**Figure 9A**). Nous avons trouvé un IC moyen du pME6010 *red* de 1,0386 (**Figure 8**). Nous avons alors constaté que ni la casette de *gfp* ni celle de *red*, apportent un poids supplémentaire à la bactérie *in planta*. Nous avons également réalisé de tests d'étalement avec le mélange complexe d'espèces testées (G1, G4 et les G8) pour établir la dilution à étaler lors des entrées *in vitro* et lors des sorties *in planta* (**Figure 9B**).

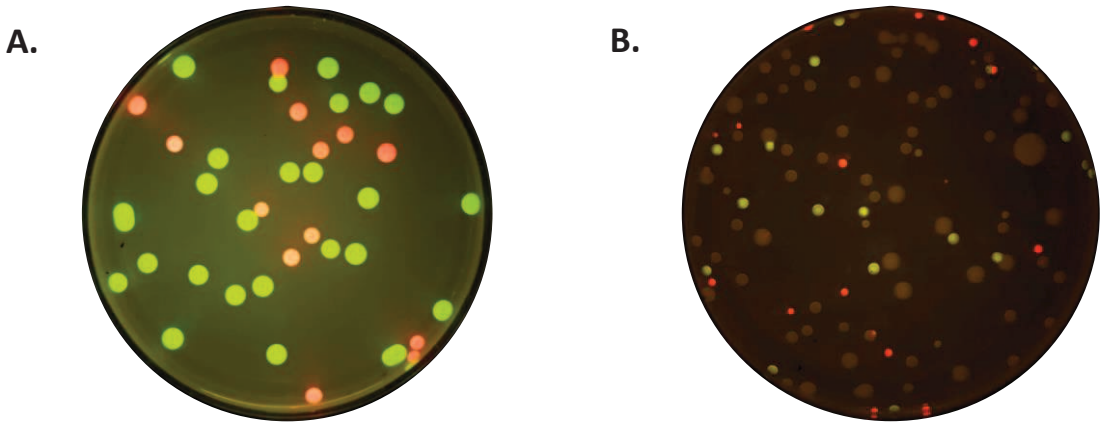


Figure 9. Construction de marqueurs plasmidiques sur le pME6010 pour la distinction des souches. Les gènes *gfp* et *red* ont été introduits séparément au plasmide pME6010 et ensuite introduits chez *A.*

fabrum C58. Des compétitions *in planta* (dans la rhizosphère de *M. truncatula*) ont ensuite été réalisées entre ces deux souches avec un ratio d'entrée de 1:1. **A.** Photo de l'étalement de la compétition sous la lumière bleue. **B.** Photo de l'étalement de la compétition avec le mélange complexe d'espèces (G1, G4 et les G8) sous la lumière bleue pour déterminer la dilution à étaler lors des entrées *in vitro* et sorties *in planta*.

v) Test concentration de tétracycline *in planta*

Etant donné que le plasmide pME6010 possède une cassette de résistance à la tétracycline, nous avons réalisé des tests pour établir la concentration de tétracycline idéale à utiliser lors de nos compétitions. Après avoir inoculé séparément chaque souche sur des plantules de *M. truncatula*, nous avons étalé nos bactéries sur des différentes boites de LPG contenant des concentrations croissantes en tétracycline.

	Nombre de colonies moyennes trouvées		
Milieu	Red	GFP	C58
LPG	38,3	108	31
LPG tet2	28	86,5	0
LPG tet4	15,3	57,8	0
LPG tet8	0,25	3,83	0
LPG tet10	0,25	1,5	0

Nous avons conclu qu'une concentration qui était pertinente pour réaliser nos compétitions était de 2µg/mL.

Finalement, pour la mesure de la valeur adaptative d'*A. fabrum* conférée par ses régions spécifiques, nous proposons donc l'utilisation d'un plasmide, le pME6010, qui possède une résistance à la tétracycline et dont la conservation *in planta* a été testée et est de 100%. Nous pourrons donc l'introduire seulement sur l'espèce *A. fabrum*, ce qui permettra de ne récupérer que cette espèce à la fin de la mise en compétition.

II. Effet des gènes spécifiques d'*A. fabrum* sur l'adaptation à la plante

Dans cette étude, la valeur adaptative d'*A. fabrum* conférée par ses régions spécifiques a été mesurée lors du style de vie commensale de la bactérie, mais aussi lors de son style de vie pathogène. L'adaptation à la vie commensale a été testée avec *Medicago truncatula*, une plante de petite taille facilement manipulable *in vitro*. L'adaptation à la survie dans les tumeurs a été testée sur des tomates, plantes usuellement utilisées pour tester le pouvoir pathogène des agrobactéries. Pour mesurer la valeur adaptative nous avons utilisé des mutants de délétion des régions spécifiques que nous avons mis individuellement en compétition avec la souche sauvage *A. fabrum* C58. Pour cela, des indices de compétitivité (IC) *in vitro* et *in planta* ont été calculés en divisant le ratio mutant/sauvage de sortie (après compétition) par le ratio mutant/sauvage d'entrée (inoculum).

Valeur adaptative d'*A. fabrum* dans son style de vie commensale dans la rhizosphère

La valeur adaptative d'*A. fabrum* conférée par ses régions spécifiques a été mesurée dans la rhizosphère de *Medicago truncatula* en utilisant les différents mutants de délétion de chacune des régions spécifiques qui ont été mis en compétition avec la souche sauvage C58. Pour cela nous avons calculé d'abord la valeur adaptative *in vitro* dans un milieu riche puis *in planta*. Nous avons utilisé la méthode de repiquage de colonies en profitant de la présence d'une cassette de résistance à la kanamycine chez toutes les souches mutantes. Ainsi, nous avons pu étaler le mélange bactérien d'entrée (inoculum) ou sortie (après compétition) sur un milieu contenant de la kanamycine où seulement les souches mutantes pourront pousser.

Ensuite, pour prendre en compte l'effet de la pression de sélection exercée par des espèces compétitrices comme c'est le cas dans la rhizosphère, nous avons ajouté à nos essais de compétiteurs appartenant à d'autres espèces d'agrobactéries, couramment retrouvées dans les mêmes biotopes qu'*A. fabrum*. Pour cela, nous avons utilisé la méthode utilisant le plasmide pME6010 qui s'est avéré être la meilleure option pour nos tests de calcul de valeur adaptative au sein d'un mélange complexe de souches.

a. Compétition bactérienne *in vitro*

La valeur adaptative des souches testées a été étudiée *in vitro* dans un milieu de croissance riche (LPG) (**Figure 2B**). Une co-inoculation de chaque souche mutante avec la souche sauvage a été réalisée en utilisant un rapport 1:1 (**Figure 2A**). Après 24h d'incubation à 28°C sous agitation, le rapport mutant/sauvage a été établi et un IC dit *in vitro* a ainsi pu être calculé (**Figure 2E**). Les indices de compétitivité *in vitro* **(Figure 10)** chez les souches testées, ne sont pas significativement différents de 1. Il n'y a donc pas de différences de valeur adaptative entre la souche sauvage et les différents mutants de délétion des régions spécifiques d'*A. fabrum*.

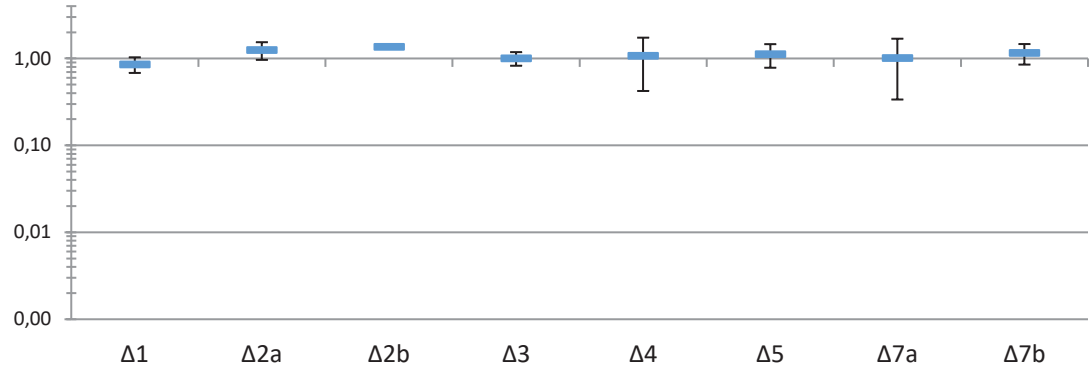


Figure 10. Indices de compétition *in vitro.* Les données présentées correspondent aux indices de compétition *in vitro* des mutants de délétion des régions spécifiques mis en compétition individuellement avec la souche sauvage C58 avec un ratio 1:1. Les barres d'erreur correspondent aux intervalles de confiance. Méthode de dénombrement utilisée : repiquage de souches.

b. Compétition bactérienne *in planta*

Après avoir validé par des tests de compétition que les différentes régions spécifiques *d'A. fabrum* n'ont pas d'effet significatif sur la valeur adaptative de la bactérie *in vitro* **(Figure 10)**, nous avons réalisé des mesures de valeur adaptative *in planta* avec un ratio 1:1 au niveau de la rhizosphère de *M. truncatula*.

Après 14 jours post inoculation chez *M. truncatula*, uniquement les indices de compétitivité de la souche ΔSpG8-2a sont significativement différentes de 1 (P=0,01563) **(Figure 11)**. Il s'agit de la seule souche présentant un IC significativement inférieur lors de la colonisation racinaire et donc à avoir une valeur adaptative plus faible par rapport à celle de la souche sauvage. Ce résultat montre l'avantage adaptatif conféré par le curdlan à *A. fabrum* dans son style de vie commensale dans la plante. Cet avantage peut-être en relation avec un effet protecteur de ce polysaccharide vis-à-vis des défenses de la plante. Le test statistique ne permet pas de mettre en évidence une différence significative pour le reste des souches, indiquant que leur valeur adaptative est semblable à celle du sauvage.

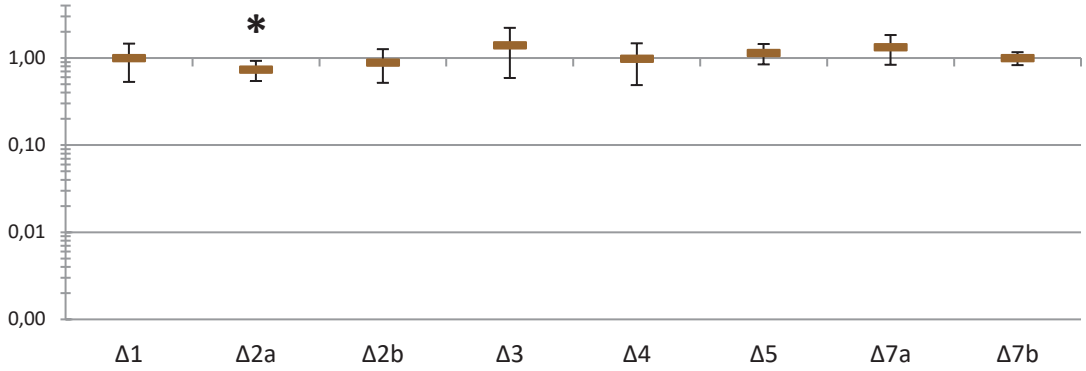


Figure 11. Indices de compétition *in planta*. Les données présentées correspondent aux indices de compétition sur les racines de *M. truncatula* des mutants de délétion des régions spécifiques mis en compétition individuellement avec la souche sauvage C58 avec un ratio 1:1. Les barres d'erreur correspondent aux intervalles de confiance. Le symbole * représente les valeurs significativement différentes de 1 établies par un test t de Wilcoxon (P<0,05). Méthode de dénombrement utilisée : repiquage de souches.

Nous avons appliqué la méthode utilisant le plasmide pME6010 avec l'ensemble des souches finalement à disposition. En effet, nous avons fait également des compétitions avec les souches mutantes ΔSpG8-1a et ΔSpG8-1b (pas disponibles au moment des premiers essais de compétition par la méthode de repiquage) chacune avec la souche sauvage. L'IC calculé à partir des bactéries récupérées après 14 jours post inoculation chez *M. truncatula* montre des valeurs significativement inférieures à celle de la souche sauvage pour les deux mutants testés : ΔSpG8-1a (P=0,01415) et ΔSpG8-1b (P=0,00781) **(Figure 12)**. Finalement, nous avons également testé trois autres souches mutantes : ΔSpG8-1, ΔSpG8-5 et ΔSpG8-7b. L'IC calculé après 14 jours chez *M. truncatula* nous montre une valeur adaptative significativement inférieure par rapport à la souche sauvage pour les souches : ΔSpG8-5 (P=0,00781) et ΔSpG8-7b (P=0,1403) mais pas pour la souche ΔSpG8-1 (P=0,25) **(Figure 12)**. Ces quatre régions codant le transport et métabolisme de sucres (SpG8-1a), la dégradation d'acides hydroxycinnamiques (SpG8-1b), le catabolisme de composés de type opines (SpG8-5) et les systèmes mécano-senseurs (SpG8-7b), en plus de la région SpG8-2a codant la biosynthèse de curdlan (repérée avec la méthode de repiquage de colonies), semblent donc favorables à la survie de la bactérie dans la racine.

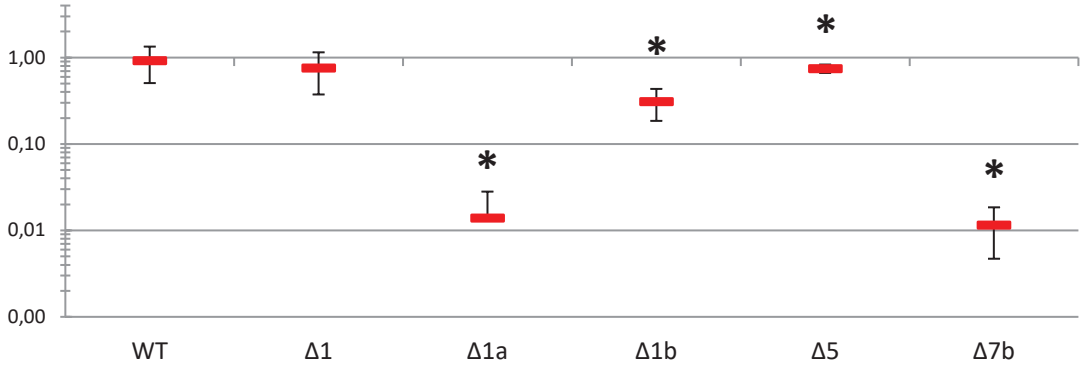


Figure 12. Indices de compétition *in planta.* Les données présentées correspondent aux indices de compétition sur les racines de *M. truncatula* des mutants de délétion des régions spécifiques mis en compétition individuellement avec la souche sauvage C58 avec un ratio 1:1. Les barres d'erreur correspondent aux intervalles de confiance. Le symbole * représente les valeurs significativement différentes de 1 établies par un test t de Wilcoxon (P<0,05). Méthode de dénombrement utilisée : utilisation du plasmide pME6010.

Grâce aux compétitions avec ces dernières souches qui avaient été testées également avec la méthode de repiquage de colonies, nous pouvons mettre en évidence que la méthode utilisant le plasmide pME6010 est beaucoup plus fine. En effet, avec la méthode de repiquage de colonies, nous avons mis en évidence que la souche mutante ΔSpG8-2a était la seule qui avait une valeur adaptative inférieure à celle de la souche sauvage. En revanche, avec la méthode utilisant le plasmide pME6010 nous avons pu montrer que 5 mutants de délétion de plus (les régions SpG8-1a, SpG8-1b, SpG8-5 et SpG8-7b) avaient également une valeur adaptative inférieure à celle de la souche sauvage. Cela nous montre que nous avons fait le bon choix de la méthode à effectuer car, en plus d'être plus rapide, elle est plus précise et capable de mieux détecter les fines différences de valeur adaptative entre les souches mises en compétition. Les régions SpG8-2b, SpG8-3, SpG8-4 et SpG8-7a n'ont pas été testées avec la méthode utilisant le plasmide pME6010 avec ce type de compétitions.

c. Compétition bactérienne *in planta* au sein d'un mélange complexe d'espèces

Pour se rapprocher le plus possible et d'une manière simple aux conditions réelles de vie bactérienne, nous avons ajouté une microflore compétitive lors des compétitions. Nous avons utilisé les espèces génomiques G1 (souche CFBP 5771) et G4 (souche CFBP 5621), espèces les plus couramment retrouvées avec *A. fabrum* (G8) dans la rhizosphère des plantes. L'apport d'autres espèces était destiné à prendre en compte l'effet de la pression de sélection exercée par des espèces compétitrices comme c'est le cas dans la rhizosphère.

Nous avons appliqué la méthode utilisant le plasmide pME6010 qui s'est avéré être la meilleure option pour nos tests. Ce plasmide possède une cassette de résistance à la tétracycline ce qui nous permet de différencier, après étalement sur un milieu LPGA amendé de 2 µg/mL de tétracycline, l'espèce G8 des espèces G1 et G4 qui ne possèdent pas cette résistance et ne sont donc pas capables de pousser sur ce milieu. De plus, nous avons introduit séparément au plasmide deux marqueurs plasmidiques différents, soit la *gfp*, soit la *red*. Ceci nous a permis de différencier les souches d'*A. fabrum*. En effet, nous avons introduit le plasmide possédant la cassette *gfp* à la souche sauvage, tandis qu'aux souches mutantes, nous avons introduit le plasmide possédant la cassette *red*. Ainsi, grâce à l'exposition sous la lumière bleue de la boite de Pétri sur laquelle a été étalé le mélange bactérien d'entrée (inoculum) ou le broyat racinaire de sortie (après compétition), nous pouvons voir à l'œil nu les deux couleurs : vert fluo pour la souche sauvage et rouge fluo pour l'ensemble de souches mutantes. De ce fait, la distinction et donc le dénombrement des deux souches d'*A. fabrum* mises en compétition a été possible de manière simple et rapide.

L'IC calculé à partir des bactéries récupérées après 14 jours post inoculation chez *M. truncatula* montre des valeurs significativement différentes de 1 pour les souches ΔSpG8-1a (P=0,0131), ΔSpG8-1b (P=0,03906), ΔSpG8-2a (P=0,01038), ΔSpG8-4 (P=0,0425), ΔSpG8-5 (P=0,04688) et ΔSpG8-7b (P=0,01415) **(Figure 13)**. En effet, la valeur adaptative de ces souches est inférieure à celle de la souche sauvage lors de la colonisation dans la racine de *M. truncatula*, notamment pour la souche ΔSpG8-2a qui possède une valeur nettement plus faible.

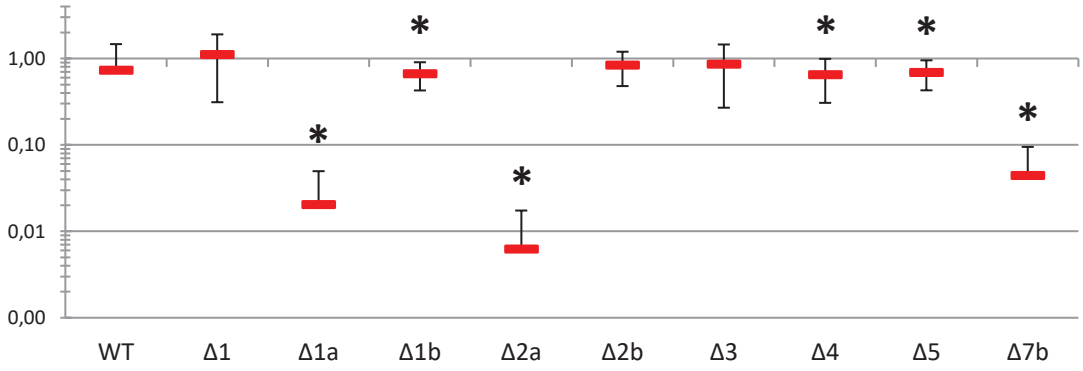


Figure 13. Indices de compétition *in planta.* Les données présentées correspondent aux indices de compétition sur les racines de *M. truncatula* des mutants de délétion des régions spécifiques mis en compétition individuellement avec la souche sauvage C58 avec un ratio 1:1, en plus des compétiteurs des espèces G1 et G4. Les barres d'erreur correspondent aux intervalles de confiance. Le symbole * représente les valeurs significativement différentes de 1 établies par un test t de Wilcoxon (P<0,05). Méthode de dénombrement utilisée : utilisation du plasmide pME6010.

Nous observons donc qu'en plus des 5 régions spécifiques déjà trouvées comme étant favorables à la survie de la bactérie dans la racine lors des compétitions uniquement entre la souche sauvage et chacun des mutants (SpG8-1a, SpG8-1b, SpG8-2a, SpG8-5 et SpG8-7b), nous trouvons en plus une autre région (SpG8-4, codant le transport et métabolisme de sucres) qui confère une meilleure valeur adaptative, notamment dans le cas où d'autres espèces sont mises en compétition également, comme c'est le cas dans une rhizosphère.

Valeur adaptative d'*A. fabrum* dans son style de vie pathogène dans la tumeur

La valeur adaptative d'*A. fabrum* conférée par ses régions spécifiques a été mesurée ensuite dans la tumeur de tomate induite par *A. fabrum* afin d'étudier leur implication dans le style de vie pathogène de la bactérie. Pour ce but, nous avons utilisé également les différents mutants de délétion de chacune des régions spécifiques. Des inoculations ont été effectuées sur des plants de tomate âgées d'un mois. Le mélange bactérien d'entrée (inoculum) était composé également de la souche sauvage et de chacun des mutants de délétion des régions spécifiques individuellement avec un ratio 1:1. La méthode de repiquage de colonies a était utilisé pour ces essais.

L'IC calculé à partir des bactéries récupérées à l'intérieur de la tumeur après 21 jours post inoculation montre des valeurs significativement différentes de 1 pour les souches ΔSpG8-2a (P=0,0078), ΔSpG8-3 (P=0,002516) et ΔSpG8-5 (P=0,00488) **(Figure 14)**. En effet, la valeur adaptative de ces souches est inférieure à celle de la souche sauvage lors de la colonisation dans la tumeur de tomate, notamment pour la souche ΔSpG8-2a qui possède une valeur nettement plus faible. Ces trois régions codant la biosynthèse de curdlan (SpG8-2a), de sidérophores (SpG8-3) et le catabolisme de composés de type opines (SpG8-5) semblent donc favorables à la survie de la bactérie dans la tumeur. En revanche, les autres souches mutantes testées semblent avoir une valeur adaptative identique à celle de la souche sauvage chez la tumeur de cette plante.

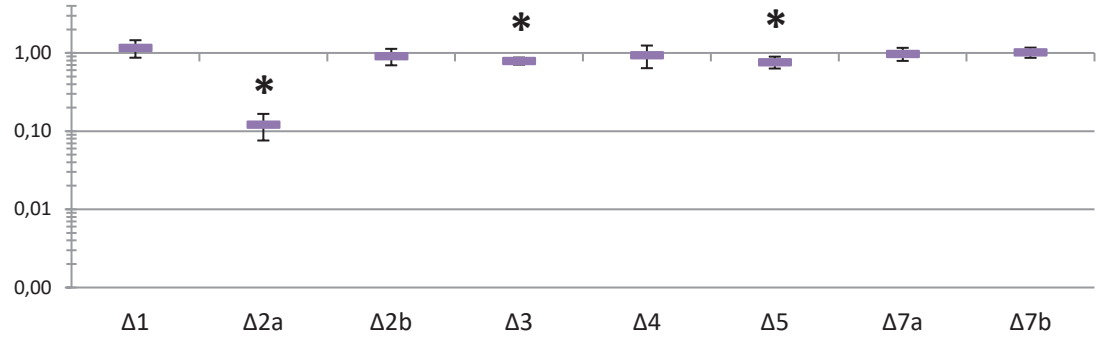


Figure 14. Indices de compétition *in planta.* Les données présentées correspondent aux indices de compétition des mutants de délétion des régions spécifiques mis en compétition individuellement avec la souche sauvage C58 avec un ratio 1:1 dans la tumeur de tomate induite par *A. fabrum*. Les barres d'erreur correspondent aux intervalles de confiance. Le symbole * représente les valeurs significativement différentes de 1 établies par un test t de Wilcoxon (P<0,05). Méthode de dénombrement utilisée : repiquage de souches.

Discussion

Cette étude s'est intéressée à déterminer si les différentes régions spécifiques d'*A. fabrum* lui conféraient une meilleure valeur adaptative dans ses biotopes les plus couramment colonisés, la rhizosphère de *M. truncatula* et la tumeur (chez une plante modèle, ici la tomate). Une démarche méthodologique a été réalisée pour pouvoir mesurer finement des indices de compétition pour les différents mutants de délétion des régions spécifiques. Ainsi, nous avons conçu une méthode qui nous permet de trouver les fines différences de valeur adaptative entre les souches lors de leur mise en compétition dans la même plante.

Habituellement, la valeur adaptative des souches est mesurée par des comparaisons indépendantes où la souche sauvage et la souche mutante poussent de manière isolée dans des milieux différents. Cette approche apporte des données absolues de survie dans un milieu donné, mais il peut refléter une situation dans laquelle l'effet recherché n'est pas suffisant pour être détecté (2). Nous avons voulu utiliser une méthode alternative, où la souche sauvage et la souche mutante sont étudiées simultanément dans un même environnement (la tumeur ou la rhizosphère). Ce type d'approche permet de mesurer un indice de compétitivité (IC), qui reflète la capacité concurrentielle des deux souches testées (10). Cette méthode possède un pouvoir discriminant supplémentaire en améliorant la sensibilité de détection entre les deux souches testées. De plus, c'est une approche plus robuste aux petits changements des conditions expérimentales (16).

Lors d'une mise en compétition entre deux ou plusieurs bactéries, un facteur crucial est de veiller à pouvoir distinguer les bactéries mise en compétition (par exemple, une souche mutante d'une souche sauvage), sans que cela ait un impact (soit négatif ou positif) sur leur valeur adaptative (3). L'approche la plus classique consiste à distinguer les deux souches par une résistance à un antibiotique, approche que nous avons utilisée lors des premières compétitions réalisées. Il s'agit d'une approche très longue, car elle nécessite une étape minutieuse, de repiquage de colonies **(Figure 2D, 2E, 2F)**. D'autres approches peuvent être utilisées, comme le marquage des souches avec des gènes codant des protéines fluorescentes, comme la GFP (*green fluorescent protein*). C'est cette deuxième approche qui a été envisagée pour la suite des compétitions, car elle nous permet de nous épargner des étapes très longues de repiquage de colonies. En effet, certaines protéines fluorescentes sont visibles à l'œil nu et donc le comptage de colonies devient accessible très rapidement.

Méthode de repiquage de colonies avec comme différentiation une cassette d'antibiotique

Pour mesurer la valeur adaptative bactérienne liée aux régions spécifiques d'*A. fabrum*, nous avons choisi tout d'abord d'utiliser une méthode de dénombrement basée sur le repiquage de colonies. Pour différencier les deux souches mises en compétition, une cassette de résistance à la kanamycine a été utilisée. En effet, les différentes souches mutantes de délétions des régions spécifiques possèdent une cassette de résistance à la kanamycine dans leur génome. De ce fait, lors des compétitions entre la souche sauvage et les différentes souches mutantes de délétions des régions spécifiques, ce sont ces dernières qui peuvent être différenciées à la fin de la compétition en utilisant un milieu contenant de la kanamycine. Après s'avoir assuré que cette cassette de kanamycine ne conférait aucun coût supplémentaire ni un avantage compétitif à la bactérie (**Figure 3**), nous avons montré que toutes les souches mutantes induisaient bien la formation des tumeurs chez la tomate (données non montrées). Ce résultat pouvait être attendu puisque nous ne travaillons que sur des souches mutées sur des gènes chromosomiques, mais devait être vérifié afin de pouvoir interpréter par la suite d'éventuelles observations d'absence de formation de tumeur.

Pour pouvoir étudier une valeur adaptative *in planta*, il est nécessaire de vérifier tout d'abord qu'il n'y a pas un effet *in vitro* sur la valeur adaptative de nos mutants, même si ce type de vérification n'est pas souvent réalisé. Nos résultats *in vitro* montrent qu'il n'existe pas d'effet sur la valeur adaptative *in vitro* lié aux régions spécifiques (**Figure 8**).

Les résultats obtenus des indices de compétition *in planta* (**Figure 12**) ont permis de montrer que 3 des régions spécifiques étudiées (SpG8-2a, SpG8-3 et SpG8-5) sont impliquées dans la valeur adaptative d'*A. fabrum* dans la tumeur de tomate, en plus, une de ces régions (SpG8-2a) est également impliquée au niveau la racine de *M. truncatula* (**Figure 9**). L'absence d'effet observé pour les autres régions spécifiques au niveau de la rhizosphère est à considérer avec précaution et à mettre en relation avec les conditions expérimentales, en effet, les tests ont été réalisés en conditions gnotoxéniques, c'est-à-dire en absence de microflore tellurique. En effet, c'est probable que la microflore du sol fasse compétition aux agrobactéries pour l'accès aux nutriments sécrétés par la plante. L'absence de compétiteurs dans nos tests n'a probablement pas permis de mettre en évidence l'importance de la capacité à utiliser un substrat carboné particulier codée par exemple, par la région SpG8-4 (transport et métabolisme de sucres) ou SpG8-5 (catabolisme de composés de type opines). C'est pourquoi nous avons cherché à réaliser des compétitions en incluant cette fois-ci, deux autres espèces d'agrobactéries. Ainsi, nous visons à nous rapprocher, d'une manière simple, d'une microflore compétitive. Il est possible que ce test plus proche des conditions réelles de vie de la bactérie permette

la mise en évidence de l'avantage adaptatif de régions spécifiques qui n'avaient pas été révélées dans les tests précédents, ni dans la racine, ni dans la tumeur.

Méthode de dénombrement de colonies avec marqueurs plasmidiques colorés

Nous avons remarqué, lors de nos premiers tests de compétitions, que pour pouvoir mettre en évidence la valeur adaptative des régions spécifiques et donc leur implication lors de l'adaptation d'*A. fabrum* à la plante, nous devions utiliser des compétiteurs qui accentuent l'utilité de posséder ces régions spécifiques. Nous avons choisi d'utiliser les espèces génomiques G1 et G4, espèces les plus couramment retrouvées avec *A. fabrum* (G8) dans la rhizosphère des plantes.

Pour pouvoir distinguer chacune des souches mises en compétition, nous avons essayé plusieurs approches. Celui qui a été concluant c'est l'utilisation d'un plasmide, le pME6010 avec deux marqueurs plasmidiques colorés à la fois, soit la *gfp*, soit la *red* (**Figure 7**). En effet, ce plasmide possède une cassette de résistance à la tétracycline, et étant donné qu'*Agrobacterium* ne possède pas cette résistance, uniquement les souches possédant le plasmide pourront pousser dans un milieu contenant de la tétracycline. Nous avons ensuite introduit le plasmide pME6010 *gfp* chez *A. fabrum* C58 et le plasmide pME6010 *red* à chacun des mutants des régions spécifiques. Ainsi, à la fin de la compétition, l'ensemble des bactéries sont récupérées de la plante et mises sur le milieu LPGA contenant de la tétracycline. Uniquement les souches possédant le plasmide vont pousser (*A. fabrum* C58 et les mutants de délétion des régions spécifiques) et pas les autres (les compétiteurs : G1 et G4) (**Figure 9**). En plus, en mettant le milieu gélosé sous la lumière bleu, nous pourrons facilement compter la quantité de colonies appartenant à la souche sauvage (en vert) et celles appartenant à la souche mutante (en rouge).

Les résultats obtenus des indices de compétition dans les racines de *M. truncatula* ont permis de montrer que 5 des régions spécifiques étudiées (en plus de la région SpG8-2a, biosynthèse de curdlan, déjà trouvée comme étant impliquée avec la méthode de repiquage) sont impliquées dans la valeur adaptative d'*A. fabrum* au niveau la racine de *M. truncatula* (**Figure 13**). Il s'agit des régions codant le transport et métabolisme de sucres (SpG8-1a), la dégradation d'acides hydroxycinnamiques (SpG8-1b), le transport et métabolisme de sucres (SpG8-4), le catabolisme de composés de type opines (SpG8-5) et les systèmes mécano-senseurs (SpG8-7b). Nous avons donc réussi à mettre en évidence l'importance de posséder ces régions spécifiques en ajoutant à la compétition d'autres espèces d'agrobactéries qui exercent une pression de sélection sur *A. fabrum*. Finalement nous pouvons dire que toutes ces régions spécifiques (SpG8-1a, SpG8-1b, SpG8-2a, SpG8-4, SpG8-5 et SpG8-7b) d'*A.*

fabrum semblent donc favorables à la survie de la bactérie dans la racine, puisqu'elles semblent lui apporter une meilleure valeur adaptative dans la rhizosphère de *M. truncatula*.

Finalement, nous avons appliqué cette méthode utilisant le plasmide pME6010 avec l'ensemble des souches à disposition. En effet, avec la méthode de repiquage, nous n'avons pas réalisé les compétitions avec les souches mutantes ΔSpG8-1a et ΔSpG8-1b, chacune uniquement avec la souche sauvage sans compétiteurs. L'IC calculé à partir des bactéries récupérées après 14 jours post inoculation chez *M. truncatula* montre des valeurs significativement inférieures à celle de la souche sauvage pour les deux mutants testés. Nous avons également testé trois autres souches mutantes (ΔSpG8-1, ΔSpG8-5 et ΔSpG8-7b) sans compétiteurs avec la nouvelle méthode (**Figure 10**). L'IC calculé après 14 jours chez *M. truncatula* nous montre une valeur adaptative significativement inférieure pour les souches ΔSpG8-5 et ΔSpG8-7b mais pas pour la souche ΔSpG8-1 comparées à celle de la souche sauvage. Cela nous montre encore une fois que cette nouvelle méthode avec le plasmide pME6010 est une méthode puissante pour repérer les fines différences de valeur adaptative entre les bactéries mises en compétition.

Région SpG8-2a : biosynthèse du curdlan

Cette région est la seule qui serait aussi impliquée dans la colonisation des tumeurs que dans la colonisation racinaire de *M. truncatula* lors des compétitions **(Figure 11, 13 et 14)**. Lassalle *et al* 2011 (15) ont montré que cette région est impliquée dans la biosynthèse du curdlan, un exopolysaccharide linéaire entièrement composé de résidus de D-glucose enchaînés les uns aux autres par des liaisons β (1-3), et qui forme une matrice recouvrant les bactéries sur les racines de plantes (17). En effet, l'annotation de cette région comporte les gènes de *atu3054* à *atu3059* **(Figure 15C)** dont les gènes *atu3057* et *atu3056* sont impliqués dans la biosynthèse du curdlan (35). La production de divers exopolysaccharides (EPS) est connue chez les bactéries pour assurer leur survie dans des conditions difficiles. Ainsi, le curdlan, lui, est connu pour être produit de manière efficace par *A. fabrum* dans des milieux à faible pH et en carence d'azote (35), et aussi pour être susceptible d'augmenter la survie des bactéries dans le sol (17). Par contre, il n'y a pas d'information concernant son expression et son rôle à l'intérieur de la tumeur ou à la surface des plantes (26). Pourtant, nos résultats par une approche d'IC, montrent que la délétion de cette région confère un désavantage à la bactérie, aussi bien dans son style de vie commensale que dans son style de vie pathogène, ce qui suggère que cette région a un rôle majeur dans la colonisation et la survie dans la plante. Nous avons montré également, que l'implication de cette région dans la valeur adaptative d'*A. fabrum* dans la plante est beaucoup plus forte lorsqu'on ajoute des compétiteurs lors de la mise en compétition entre la souche sauvage et la souche mutante. Ce deuxième cas se rapproche beaucoup plus des conditions de vie réelles de la

bactérie dans le sol ou dans la rhizosphère. Ce qui nous amène à penser que cette région est bien utile pour *A. fabrum* lors de son style de vie commensale, où elle se trouve couramment en contact avec d'autres espèces d'*Agrobacterium*.

Région SpG8-1 (SpG8-1a et SpG8-1b) : Transport et métabolisme de sucres et dégradation d'acides hydroxycinnamiques

La région SpG8-1 a été divisée en deux sous-régions, la sous-région SpG8-1a qui va du gène *atu1398* au gène *atu1409* et la sous-région SpG8-1b qui va du gène *atu1410* au gène *atu1423* **(Figure 15A et 15B)**. Lors des tests de compétition, nous avons utilisé trois mutants différents pour cette région, le mutant ΔSpG8-1, le mutant ΔSpG8-1a et le mutant ΔSpG8-1b. Le mutant ΔSpG8-1 a été testé en compétition avec la souche sauvage au niveau des tumeurs de tomate. Les valeurs d'IC nous montrent que ce mutant possède une valeur adaptative semblable à celle de la souche sauvage **(Figure 14)**. Ce qui nous amène à penser que cette région toute entière ne semble pas être impliquée dans la colonisation ou la survie dans la tumeur. Les mutants des sous-régions ΔSpG8-1a et ΔSpG8-1b n'ont pas été testé dans la tumeur de tomate. Ensuite, les résultats des différentes compétitions dans la racine de *M. truncatula* avec ce même mutant (ΔSpG8-1), nous montrent également que cette région toute entière n'est pas impliquée dans la colonisation rhizosphérique de cette plante **(Figure 11-13)**. Nous pouvons donc observer que la région SpG8-1 entière ne semble pas conférer un avantage adaptatif à *A. fabrum* lors de son interaction avec la plante dans les conditions testées.

En revanche, lors des compétitions avec les sous-régions ΔSpG8-1a et ΔSpG8-1b, nous avons pu mettre en évidence qu'elles peuvent conférer un avantage adaptatif à la bactérie lors de son style de vie commensale. En effet, les deux sous-régions possèdent des valeurs d'IC significativement inférieurs à celles de la souche sauvage lors des différentes compétitions dans la rhizosphère de *M. truncatula* **(Figure 12 et 13)**. La région ou « sous-région » SpG8-1a semble être impliquée dans le transport et métabolisme des sucres. Cette région contient plusieurs transporteurs ABC, dont un spécifique aux monosaccharides et deux opérons, dont les homologues ont été décrits comme étant spécifiques à l'importation d'acides aminés (Hosie et al. 2002). La région SpG8-1a contient également un gène codant pour une enzyme de lactonisation du muconate, intermédiaire de la voie hautement conservée du β-cétoadipate dans diverses bactéries, dont plusieurs agrobactéries (Harwood et Parales 1996). Il s'agit d'une voie de dégradation des composés aromatiques fréquemment produits par les plantes et utilisés comme substrats de croissance par de nombreux membres de la famille des *Rhizobiaceae* (Parke et Ornston 1986). Avoir cette région semble conférer un avantage compétitif à *A. fabrum* permettant une meilleure colonisation de ses principaux biotopes. En effet, lors de la délétion de cette région, la bactérie perd la capacité d'importer et par la suite utiliser un substrat utile.

La région ou « sous-région » SpG8-1b est impliquée dans la dégradation des acides hydroxycinnamiques (HCA) (Campillo et al. 2014) comme par exemple l'acide férulique et l'acide p-coumarique, des métabolites secondaires communs chez les plantes et libérés en grandes quantités dans le sol lors de la décomposition cellulaire (Whitehead et al., 1983). Dans la rhizosphère, les HCA sont des signaux chimiotactiques (Parke et al. 1987) et utilisés comme source de carbone par les agrobactéries (Deavours et al. 2006; Andreoni et al. 1995). La région SpG8-1b code plus spécifiquement pour la voie de dégradation de l'acide férulique qui sera finalement utilisé comme source de carbone et d'énergie par *A. fabrum* (Meyer et al.2018) (Campillo et al.2014). Cette voie est finement régulée par HcaR, un répresseur transcriptionnel codé par atu1422 (*hcaR*) qui réprime également sa propre transcription en l'absence de HCA (Lassalle et al. 2011). Les HCA sont les effecteurs de HcaR et permettent donc l'induction de l'expression des gènes de dégradation des HCA (Meyer 2018). Ainsi, l'induction de l'expression de cette voie uniquement après la détection du HCA est avantageuse pour les bactéries afin d'éviter le coût métabolique de l'expression constitutive de ces gènes (Deochand et Grove 2017). En effet, dans les environnements riches en HCA comme la rhizosphère (Mandal et al. 2009), il est important pour *A. fabrum* de dégrader et d'assimiler rapidement les HCA pour obtenir un avantage concurrentiel sur les autres agrobactéries. D'un autre côté, la répression de cette voie par HcaR est importante pour fournir à *A. fabrum* un avantage compétitif dans un environnement à faible teneur en HCA (Mandal et al.2009). En l'absence donc de cette région dans la rhizosphère de *M. truncatula*, nous observons une diminution de la valeur adaptative d'*A. fabrum* **(Figure 12 et 13)**, qui n'est donc plus capable de dégrader l'acide férulique qui devient à son tour toxique.

Nous pouvons également remarquer qu'il existe un lien entre ces deux sous-régions. En effet, leurs valeurs d'IC mettent en évidence qu'elles confèrent un avantage adaptatif à la bactérie séparément, contrairement à la délétion de la région en entière.

Région SpG8-7b : systèmes mécano-senseurs

Cette région, qui comporte les gènes de *atu4295* à *atu4307* **(Figure 15K)** semble impliquée dans la colonisation racinaire de *M. truncatula* **(Figure 12 et 13)** mais pas lors de la colonisation ou de la survie dans la tumeur **(Figure 14)**. Grâce à la dernière méthode proposée pour mesurer la valeur adaptative des souches avec l'utilisation du pME6010, nous pouvons constater que nos résultats d'IC montrent une valeur significativement inférieure à celle de la souche sauvage pour les différents type de compétitions réalisées dans la rhizosphère (avec ou sans compétiteurs d'autres espèces d'agrobactéries), nous montrant la finesse de la méthode. En effet, avec la méthode de repiquage,

nous n'avons pas pu montrer une différence de valeur adaptative entre cette souche mutante et la souche sauvage. Néanmoins, l'implication de cette région au contact avec d'autres espèces d'*Agrobacterium* serait plus forte.

L'annotation de cette région suggère son implication dans les fonctions liées à la détection et transduction des signaux environnementaux. En effet, Lassalle *et al.* 2011 ont émis l'hypothèse que cette région est impliquée dans la perception des signaux de l'environnement qui peuvent être responsables de l'activation d'autres fonctions, y compris, peut-être, les fonctions SpG8 telles que le métabolisme phénolique. Deux gènes (*atu4300* et *atu4305*) codant pour un système à deux composants et appartenant à cette région sont homologues aux gènes de *Bradyrhizobium japonicum* dont les protéines sont impliquées dans la reconnaissance de la plante hôte pendant le processus de nodulation (Lang et al.2008). Ces gènes sont également homologues aux gènes de *Pseudomonas putida* dont les protéines reconnaissent le toluène et le styrène, activant les voies de dégradation (Lau et al. 1997).

Région SpG8-5 : catabolisme de composé type opines

Cette région, qui comprend les gènes *atu3947* à *atu3952* **(Figure 15G)** est impliquée dans le catabolisme de composés type opines, composés spécifiques aux tumeurs (31). Il s'agit de phosphodiesters de sucres ou de condensats d'un acide aminé et d'un sucre ou d'un acide cétonique (32), et leur présence dans la tumeur fournit une niche écologique favorable à la croissance d'*A. fabrum* (7). Bien qu'il existe des régions de gènes localisées dans le plasmide Ti bien identifiées pour cataboliser certains types d'opines (14), la région SpG8-5 pourrait servir à l'utilisation de certains types d'opines, pas forcément produites par le pTi, mais retrouvées dans la tumeur. Cependant, l'annotation de cette région n'est pas suffisamment précise. Il est donc possible que le substrat concerné appartienne à une autre classe de condensats d'acides aminés et de sucres, comme par exemple les composés Amadori, une classe de molécules produites dans la matière végétale en décomposition et donc commune dans le sol. Toutefois, les résultats de cette étude montrent lors de nos compétitions dans la tumeur que le mutant dépourvu de cette région présente une valeur adaptative inférieure à celle de la souche sauvage **(Figure 14)**. Cette région spécifique semble donc conférer un certain avantage adaptatif à cette bactérie lors de la colonisation ou de la survie dans la tumeur.

Les résultats des compétitions au sein de la rhizosphère de *M. truncatula*, nous montrent que cette région est aussi importante pour l'adaptation dans la rhizosphère. En effet, lors des différentes compétitions dans la racine (avec ou sans compétiteurs d'autres espèces), cette région confère un avantage adaptatif à la bactérie (**Figure 12 et 13**). Ceci a été repéré seulement avec la méthode utilisant

le plasmide pME6010 (**Figure 12**), qui s'est avéré une méthode plus fine et pas avec la méthode de repiquage (**Figure 11**).

Région SpG8-4 : transport et métabolisme de sucres

Cette région semble impliquée dans la colonisation de la rhizosphère de *M. truncatula* uniquement au sein des compétitions avec d'autres agrobactéries **(Figure 13)**. De plus, cette région est la seule à conférer une valeur adaptative lors de la présence de compétiteurs d'autres espèces d'agrobactéries, mais pas lors de compétitions au sein d'uniquement l'espèce G8 (**Figure 11**). En effet, nous pouvons penser que lorsqu'il y a la présence d'autres agrobactéries, l'avantage compétitif qui apporte cette région est bien mise à l'avant. Cette région comporte les gènes *atu3808* à *atu3830* **(Figure 15F)**. Cette région est impliquée dans le catabolisme des glucides. En effet, cette région semble avoir une unité fonctionnelle dédiée à l'absorption des monosaccharides via un transporteur ABC hypothétique spécifique au ribose. De même, il a des fonctions enzymatiques hypothétiques telles que la rhamnose mutarotase ou la D-galactarate déshydrogénase, impliquées dans le métabolisme des sucres. Cette région possède également quatre régulateurs transcriptionnels de type LysR qui peuvent être impliqués dans la régulation de certaines voies métaboliques (Maddocks et Oyston 2008; Lassalle et al.2011). Le mutant de cette région, le ΔSpG8-4 ne possède plus la capacité de transport et métabolisme de ces glucides hypothétiques présents dans la racine.

Région SpG8-3 : biosynthèse de sidérophores

Les résultats nous permettent de démontrer l'implication d'une autre région dans la survie et/ou la colonisation d'*A. fabrum* dans la plante, notamment dans la tumeur induite par cette bactérie **(Figure 14)**. Cette région spécifique SpG8-3 comporte les gènes *atu3663* à *atu3691* **(Figure 15E)** (24). Il a été démontré que cette région possède le potentiel génétique pour produire des sidérophores (24), petites molécules qui lient le fer avec une grande affinité (5). Ils sont produits par certaines bactéries lorsqu'elles sont soumises à des faibles concentrations de fer comme c'est le cas chez la plante (8, 19). Une croissance plus lente, par rapport à une souche sauvage en présence d'un chélateur de fer, a déjà été montrée chez des mutants de délétion des gènes appartenant à cette région (24). Nos résultats sont en accord avec cette étude, car ils montrent que le mutant de délétion de cette région possède un indice de compétition plus faible dans la tumeur. La région spécifique SpG8-3 peut donc fournir une augmentation de la valeur adaptative lors des compétitions dans un biotope donné, comme suggéré par Lassalle *et al*. 2011 (15).

En revanche, cette région ne semble pas conférer un avantage sur la valeur adaptative dans la rhizosphère de *M. truncatula* **(Figure 11 et 13)**. En effet, on n'observe pas une différence de valeur

adaptative entre la souche sauvage et la souche mutante de cette région lors de leur mise en compétition avec aucune des méthodes utilisées au cours de cette étude. Avec ces résultats, nous supposons que le fer n'est pas un facteur limitant dans la rhizosphère de *M. truncatula* comme il pourrait être le cas dans la tumeur.

Régions avec peu ou pas d'effet sur la valeur adaptative d'*A. fabrum in planta*

Les résultats des IC obtenus à partir des compétitions réalisées, nous montrent qu'il existe 3 régions ou sous-régions (SpG8-1, SpG8-2b et SpG8-7a) qui ne semblent pas impliquées dans l'interaction d'*A. fabrum* avec la plante ni dans son style de vie commensale ni dans son style de vie pathogène (**Figure 11-14**). En effet, nous n'avons pas réussi à montrer un quelconque effet sur la valeur adaptative de la bactérie apporté avec l'une de ces régions spécifiques. Néanmoins, comme déjà montré ci-dessus, quand nous regardons les sous-régions séparément appartenant à la région SpG8-1, nous sommes capables de montrer un effet sur la valeur adaptative de la bactérie avec les méthodes proposées (**Figure 12 et 13**). De même, nous n'avons pas testé en compétition les souches mutantes ΔSpG8-2b et ΔSpG8-7a avec la souche sauvage sur la racine de *M. truncatula* en utilisant la nouvelle méthode des plasmides colorés (**Figure 12**). Cela nous laisse supposer que potentiellement peuvent finalement être impliquées dans la valeur adaptative d'A. fabrum dans la rhizosphère. Le cas échéant, ces régions pourrait être importantes dans un autre biotope colonisé par la bactérie comme par exemple, le sol nu.

Optimisation des approches proposées

Récemment, des techniques dites *transposon sequencing* ont été développées en combinant les nouvelles technologies de séquençage massif avec la mutagenèse traditionnelle de transposition pour la construction de banques (28). Ce sont des méthodes robustes et sensibles à haut débit, avec un grand potentiel pour déterminer la contribution de chaque gène d'un organisme sur sa valeur adaptative dans une condition donnée (30). Ainsi, même des mutants avec une valeur adaptative très peu différente de celle du sauvage (moins de 5% de différence) peuvent être mis en évidence dans ce type d'approche (29). La construction et la validation d'un transposon pour l'utilisation chez les *Rhizobiaceae* ont été récemment développées (21). Ce nouvel outil peut être appliqué chez *A. fabrum* pour étudier le rôle de l'ensemble des gènes spécifiques d'espèce dans la colonisation de la rhizosphère ou de la tumeur. Il s'agirait donc d'une approche moins lourde que celle utilisée, nécessitant la comparaison des souches une à une.

Conclusion

Ce travail nous a permis de mettre en évidence les régions spécifiques d'*A. fabrum* impliquées dans la valeur adaptative de la bactérie lors de son interaction avec la plante dans ses deux styles de vie, commensale et pathogène. Le style de vie commensale d'*A. fabrum* a été étudié dans la rhizosphère de *M. truncatula* avec ou sans compétiteurs d'autres espèces d'agrobactéries. Le style de vie pathogène d'*A. fabrum* a été étudié dans la tumeur de plantes de tomates induite par cette bactérie. Pour mettre en évidence l'avantage compétitive de ces régions, nous avons réussi à développer une méthode qui repère de façon rapide et précise les différences de valeur adaptative entre les souches mises en compétition.

Nous avons pu constater que la quasi-totalité des régions spécifiques d'*A. fabrum* confèrent un avantage adaptatif à la bactérie. Certaines régions sont plus impliquées dans le style de vie commensale d'*A. fabrum* (SpG8-1a, SpG8-1b, SpG8-4 et SpG8-7b), une région est plus impliquée dans son style de vie pathogène (SpG8-3) et deux régions sont impliquées dans ces deux styles de vie (SpG8-2a et SpG8-5). Il y a aussi certaines régions dont nous n'avons pas réussi à mettre en évidence une quelconque implication dans la valeur adaptative d'*A. fabrum* comme c'est le cas des régions SpG8-1, SpG8-2b et SpG8-7a. Cela est peut-être dû aux conditions expérimentales utilisées, aux souches des compétiteurs interspécifiques utilisées ou aux biotopes étudiés. Ces travaux nous permettent de mieux comprendre l'écologie d'*A. fabrum* et la construction de sa niche écologique spécifique.

Chapter 3

Effects of *A. fabrum*-specific regions on plant metabolomic profile

Preamble chapter 3

In soil, plant roots are colonized by a wide range of bacteria and many of them can play an important role influencing secondary plant metabolism. As well as pathogens, it has been shown that PGPR (plant growth-promoting rhizobacteria) can lead to qualitative and quantitative modifications of the contents of secondary metabolites, notably phenolic compounds, resulting in changes in their biosynthetic pathways. This also proves that plant secondary metabolism is of great importance when it comes to plant-bacteria interactions and plant ecological adaptation.

Agrobacterium fabrum colonize rhizospheres extensively as a saprophyte but it is also a model strain for pathogenicity studies as it causes crown-gall disease. The hypothesis of this chapter is that *A. fabrum* modulates, as well as other soil-borne bacteria, plant secondary metabolites and more specifically phenolic compounds in order to construct its specific ecological niche. Indeed, we demonstrated in Chapter 2 that *A. fabrum*-specific genes gives a competitive advantage to the bacteria in the plant. So, we aimed at understand the implication of these genes in the establishment of this complex bacteria-plant interaction. Thereby, the objective of this work was to assess the *A. fabrum* effect on secondary plant metabolism during the two lifestyles of this bacterium: the commensal in the root plants and the pathogen in the plant tumor that it itself induces. Furthermore, we also studied the involvement of the *A. fabrum*-specific genes using mutant deletion strains, making our work original. Indeed, several studies raise the significant changes in secondary metabolic plant profiles following bacteria inoculation but, to the best of our knowledge, there are no investigations dealing with mutant strains.

In the first part of this chapter, we studied the impact of *A. fabrum* inoculation on root plant secondary metabolites and, more particularly, the involvement of its species-specific regions, during the commensal lifestyle in the rhizosphere. Metabolomic analyses showed that *A. fabrum* inoculation modulates phenolic compounds content in roots, particularly flavonoids. These molecules play a key role when it comes to plant-bacteria interactions as they can be signaling molecules for bacteria or serve as markers in the adaptation of *A. fabrum* to the plant. Furthermore, we showed that this modulation was essentially link to *A. fabrum*-specific genes. This work is presented in the form of a publication which will be submitted to the journal Molecular Plant-Microbe Interactions.

The second part of this chapter concerns the pathogenic lifestyle of *A. fabrum.* We studied not only the plant response to bacteria inoculation, but also the involvement of *A. fabrum*-specific regions

on secondary metabolites of tumors (symptom of crown-gall disease) induced by these bacteria and on opines, low molecular weight metabolites produced by the transformed plant cells. Although the complete analysis of discriminating compounds is not finished, metabolic analyses highlight the importance of secondary metabolites, especially HCAAs, in the bacteria-plant interaction. Furthermore, we showed the metabolic plant response exclusively related to the *A. fabrum*-specific regions. Then, to study the influence of specific genes on opine production, a method allowing fast and precise assays was required. However, the existing methods are not sensitive enough, are limited to only a few types of opines and require long protocols for sample preparation. Therefore, a general, rapid, specific and sensitive analytical method was developed for overall opine detection based on ultra-high-performance liquid chromatography–electrospray ionization quadrupole time-of-flight–mass spectrometry (UHPLC-ESI-MS-QTOF). As this method is useful not only to detect but also to quantify a wide range of opines in different plant extracts, it represents a powerful tool to perform analyses of plant galls to diagnostic crown gall and hairy roots. The development and validation of this method is presented below in the form of an article that has been submitted to the journal Analytical and Bioanalytical Chemistry. It is followed by the application of the method to study the influence of *A. fabrum*-specific genes on the production of opines in the tumor.

All these results will contribute to a better understanding of the ecological niche construction of *A. fabrum* not only by the formation of plant tumors, but also on its rhizospheric lifestyle, highlighting the importance of its specific genes in the establishment of this fine-tuned interaction.

Part I

Metabolomic study of *A. fabrum* commensal lifestyle on root secondary metabolites

INTRODUCTION

The rhizosphere of plants is highly colonized by a wide range of soil microorganisms, establishing different types of interactions with plant roots. The effect of these interactions can be beneficial, neutral or detrimental to the plant (Walker et al. 2013; Sanguin et al. 2006). Approaches targeting partner metabolic changes are needed to evaluate plant–bacteria interactions. Indeed, plants are able to synthesize diversity of secondary metabolites, which are key components to interact with their environment and for the adaptation to both biotic and abiotic stress conditions (Bennett and Wallsgrove 1994). Some metabolomic studies have been carried out to identify metabolites involved in bacteria-plant interactions as it is a powerful tool to achieve an accurate study of plant secondary metabolism. Recent studies have assessed bacteria effects on plants (e.g. maize and rice), especially PGPR bacteria as *Azospirillum*, *Pseudomonas*, *Agrobacterium* and *Paraburkholderia* (Walker et al. 2011, 2012, 2013; Rozier et al. 2016; Valette et al. 2019; Miotto-Vilanova et al. 2019), as well as pathogenic bacteria as *Burkholderia* (Chamam et al. 2015). Researchers focused on plant secondary metabolism and found significant changes on root profiles in particular phenolic compounds. These molecules, known to be involved in plant-bacteria interactions, comprise a large chemical diversity, ranging from simple phenolic acids to complex polymerized tannins. In plants, they are described among others, as signaling molecules for bacteria (Mandal et al. 2010).

Agrobacteria are ubiquitous soil borne and rhizospheric bacteria of the Alphaproteobacteria family, capable of establishing commensal or even beneficial interactions with plant roots and whose ecology and distribution are well known (Costechareyre et al. 2010; Shams et al. 2012; Bhattacharyya and Jha 2012; Naqqash et al. 2016). Field investigations consistently showed that several species of agrobacteria generally co-exist in the same biotopes (bare soil, rhizosphere, host plants) (Nesme et al. 1987; Vogel et al. 2003; Shams et al. 2012; Bouri et al. 2016). This means that co-existing species must exploit different resources to avoid competitions with their closest relatives, and thus have at least partly different ecological niches. The ecological functions required for adaptation to such environment could be encoded by species-specific genes (Lassalle et al. 2011).

Using comparative genomics, it was possible to reveal the species-specific genes of the genomic species *Agrobacterium fabrum* (*i.e. Agrobacterium* species genomovar G8) a species often found together with other members of the *A. tumefaciens* species complex in several biotopes (Lassalle et al. 2011). In C58, the reference strain of *A. fabrum*, these authors found 196 species-specific genes that were mostly clustered into seven genomic islands on its genome, called "specific-regions" or SpG8 (for

G8-specific coding DNA sequences), going from SpG8-1 to SpG8-7. Some specific-regions were subsequently divided into subclusters encoding homogeneous functions. The annotation of these specific regions leans towards a close connection with the plant (Lassalle et al. 2011), since they encode coherent putative pathways related, for example, to the metabolism of trophic resources derived from plants (SpG8-1a, SpG8-1b, SpG8-4, SpG8-5), to the expression of environmental sensing systems (SpG8-7), or to the synthesis of extracellular compounds, such as siderophores (SpG8-3) or curdlan (SpG8-2a). The latter, a glucose polymer induced under stress conditions, surround and protect bacteria, increasing its survival in soil (Harada and Harada 1996; Kim et al. 1999; McIntosh et al. 2005). Curdlan production may also play a role in attachment to colonize roots, as well as in bacterial pathogenesis and contact signaling (Matthysse and McMahan 1998; Rodríguez-Navarro et al. 2007).

One well-studied specific-region, the SpG8-1b, encodes the complete degradation pathway of ferulic acid (an hydroxycinnamic acid, HCA) that will ultimately be used as a source of carbon and energy by *A. fabrum* (Campillo et al. 2014; Meyer et al. 2018). HCA are common plant secondary metabolites released in large amounts in soil during the decay of root cells and described as chemotactic signals for agrobacteria (Whitehead et al. 1983; Parke et al. 1987). This pathway is finely regulated by a transcriptional repressor (HcaR) that allow a positive feedback loop in the presence of HCA, inducing their degradation and giving to *A. fabrum* a competitive advantage over other agrobacteria in HCA-rich environments such as rhizosphere (Mandal et al. 2009; Lassalle et al. 2011; Meyer et al. 2018). Another well-known specific-region is the SpG8-3, that has been shown to encode the biosynthesis of a siderophore required for *A. fabrum* C58 growth under iron limiting conditions, suggesting that this strain is highly efficient in iron scavenging (Rondon et al. 2004). This genomic region includes genes for siderophore release and reuptake as well as the regulation system (Braun et al. 2006; Lassalle et al. 2011). This siderophore has a unique chemical structure (Rondon et al. 2004), thus providing another fitness advantage to outperform competitors, especially in habitats like rhizospheres with dense and diverse populations (Lassalle et al. 2011). There is even a link between these last two specific-regions (SpG8-1b and SpG8-3) since a coordinated expression has been observed, both being induced in the presence of HCAs thanks to a transcriptional regulator (Baude et al. 2016). Furthermore, the presence of one of these specific-regions is required for the expression of the other and vice versa (Baude et al. 2016). This strengthen the idea of the existence of an ecological niche to which species G8 is specifically adapted through the expression of a particular combination of species-specific genes (Lassalle et al. 2017).

Given the strong interconnection between *A. fabrum*-specific regions and the plant, they are thus suspected to have an impact on root secondary metabolites during bacteria-plant interactions. To this end, a metabolomic approach is here performed using deletion mutant strains of each *A.*

fabrum-specific regions inoculated singly on *M. truncatula*, a model plant for studying rhizospheric interactions with bacteria (Cook 1999; Farag et al. 2008). Analyzes were performed by ultra–high pressure liquid chromatography with ultraviolet and electrospray ionization-mass spectrometry (UHPLC-UV/DAD-ESI MS QTOF), a powerful tool for the global study of constituents in plant extracts such as phenolics (Qi et al. 2008; Niessen and Tinke 1995). A comparison of root metabolic profiles was then performed between plants inoculated with the mutant strains with those inoculated with the Wild-type strain (WT). The latter was also compared to a Non-inoculated condition (NI). Finally, discriminating metabolites were characterized and some identified performing tandem mass spectrometry (MS/MS), some of which are reported here for the first time. This will bring new elements to evaluate the bacterial influence on plants through its specific-genes and thus achieve the construction of its own and specific niche during bacterial-plant interactions.

RESULTS

All *A. fabrum* specific genes are expressed on *M. truncatula* roots

In order to verify the expression *in planta* of the *A. fabrum*-specific regions, *M. truncatula* seedlings were inoculated with the WT bacterial strain C58 harboring singly transcriptional fusions of one gene per region selected on the basis of their likely involvement in region regulations. As shown by the green fluorescent of reporter bacteria (**Fig. S1**), the microscopic observations of five plants per condition 14 days post inoculation revealed that all the constructions were expressed in the wild-type C58 in contact with the root system of *M. truncatula*. This indicated that the native genes targeted by fusions were also expressed in the WT strain during the colonization of *M. truncatula* roots. As it is very likely that all specific genes behave accordingly, we assume that changes in the metabolite profiles between plants inoculated with strains deleted of specific regions and plants inoculated the wild-type strain are related to the presence of these genes in the wild-type.

Global comparison of the root phenolic compounds profiles

Root plant extracts obtained 14 dpi were analyzed by UHPLC-DAD-ESI-MS QToF. The quality control (QC) led to the detection of a total of 92 peaks **(Fig. 1)**. Only compounds with mean relative abundances higher than 1% in the total ion current (TIC) chromatogram were considered thereafter in this study (*i.e.* 28 peaks). Changes detected in root secondary metabolites in all conditions were in terms of relative intensity rather than in appearance or disappearance of peaks. Chromatographic

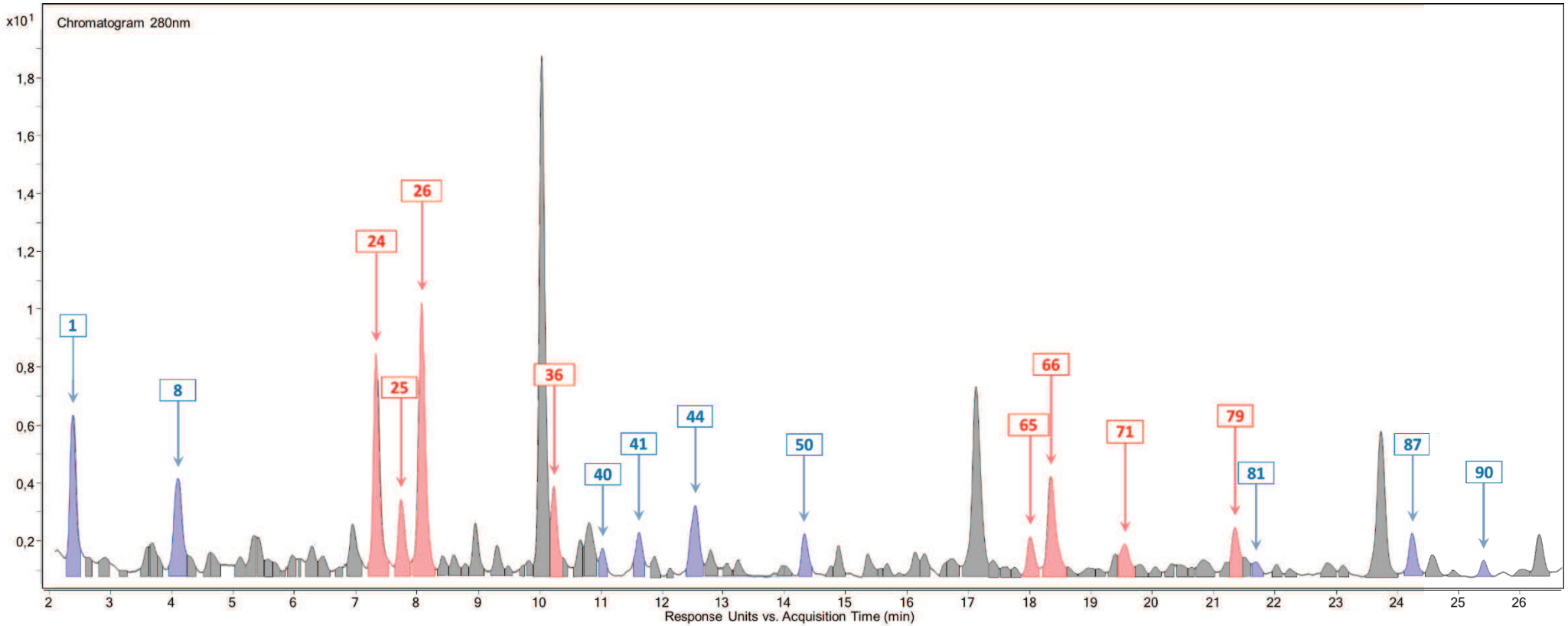


Fig. 1. Example of chromatogram (280nm) obtained by UHPLC-UV/DAD analysis of phenolic compounds extracted from barrel medic roots inoculated or not with the wild-type strain or the deletion mutant strains of A. fabrum C58. The integrated chromatographic peaks were taken into account in the metabolite profiling data study. Peak numbering corresponds to metabolite numbering in **Table 1** and in the heatmap in **Fig. 4**. Peaks with an arrow are discriminating with a significant difference (P <0.05, Student t-test) between plants inoculated with the wild-type strain compared to each of the other conditions (the Non-Inoculated condition and the deletion mutants of the *A. fabrum*-specific regions conditions). Peaks in blue correspond to the down-accumulated compounds, peaks in red correspond to the over-accumulated compounds compared to the WT condition.

profiles were further compared using Partial Least Squares Discriminating Analysis (PLS-DA) (**Fig. 2**) or Principal Component Analysis (PCA) (**Fig. 3**). Two *A. fabrum* WT strains were used in this study (WT1 and WT2). The first one is a native *A. fabrum* WT strain and the second one has the *ntpII* kanamycin resistance gene amplified from plasmid pKD4 (Datsenko and Wanner 2000). A first metabolic profile comparison and a statistical analysis was made with these two strains inoculated in *Medicago truncatula* in order to measure the effect of the *ntpII* gene. Indeed, this gene was used to perform the construction of each deletion mutant of *A. fabrum*-specific regions used in this study. The PCA of metabolic profiles of the two WT conditions show no discrimination between the two profiles with PC1 and PC2 scores of 10.97 and 3.78, respectively **(Fig. 3A)** and the statistical analysis highlighted only one metabolite significantly different in its relative abundance (**39**). So, we decided to exclude this metabolite from the analysis and to group the two wild-type conditions in one, that will be called from now as the WT condition.

The PLS-DA performed on metabolic profiles data showed a clear discrimination between all the deletion mutants of *A. fabrum*-specific regions taken altogether compared to the WT condition and the NI condition (**Fig. 2**). While the latter was found to be clearly separated from all the other profiles of the conditions tested in this study, profiles of the *A. fabrum* deletion mutant strains conditions show a gradient when compared to the WT condition (**Fig. 2**). The SpG8-Δ2a mutant strain condition has the closest profile to that of the WT condition. On the contrary, SpG8-Δ3 mutant strain condition has the further profile to that of the WT condition. The other *A. fabrum* deletion mutant strains profiles are in the middle of these last two.

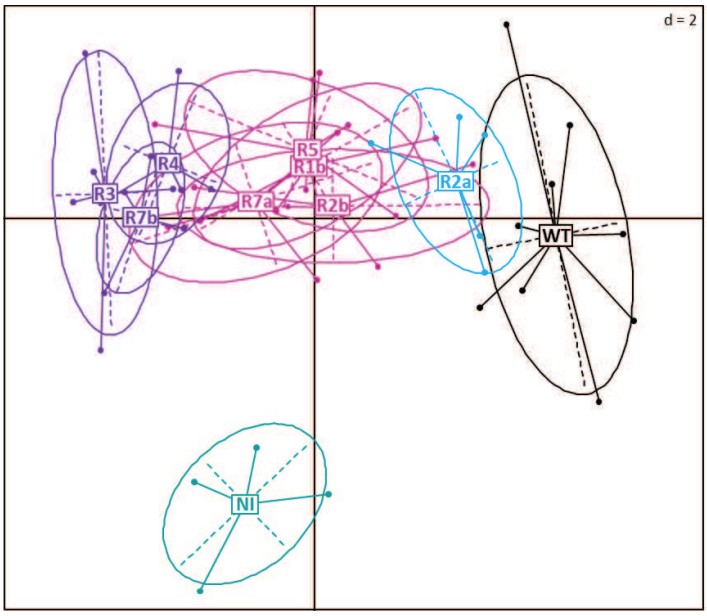


Fig. 2. Comparison of roots secondary metabolites profiles between all our tested conditions (WT, mutants and NI). PLS-DA were performed on chromatographic data at 280 nm obtained for each methanolic extract of barrel medic roots based on peak areas and retention time (data matrix of 92 peaks). Plants were inoculated or not with *A. fabrum* C58 wild-type or deletion mutant strains of *A. fabrum*-specific regions. WT: plants inoculated with the wild-type strain, NI: non-inoculated condition, RX: plants inoculated singly with each of the deletion mutant of *A. fabrum*-specific regions.

Evidence of bacterial effects on root plant secondary metabolites content

The PCA of metabolic profiles of the WT condition and the non-inoculated condition (NI) showed a discrimination between the two profiles with PC1 and PC2 scores of 9.49 and 5.78, respectively **(Fig. 3F)**. A total of 11 metabolites significantly varied in their relative abundances between both conditions. Five of which are underabundant (**8**, **40**, **44**, **87** and **90**), the remaining six are overabundant (**24**, **25**, **26**, **36**, **65** and **79**) in the NI condition, compared to the WT condition (**Fig. 4**). Among them, there are two compounds (**8** and **25**, underabundant and overabundant, respectively) that are exclusively discriminating between the NI and the WT condition. Indeed, these two compounds are not discriminating when comparing the conditions of *A. fabrum*-specific regions with the WT condition.

Effect of specific genes on root plant secondary metabolites content

Metabolic profiles of each deletion mutant condition were compared with the WT condition (**Fig. 3B**-**E**, **G**-**J**). Only the SpG8-Δ2a mutant strain condition was found to have a metabolic profile indistinguishable from that of the WT condition (**Fig. 3C**) and as it could be expected, no metabolites were found to be significantly different when compared to the WT condition (**Fig. 4**). To the contrary, all other mutant strains metabolic profiles clearly separated from that of the WT condition (**Fig. 3B**, **D**, **E**, **G**-**J**). Taken all comparisons together (NI and mutant conditions), a total of 16 metabolites were found to display a significantly different relative abundance compared to the WT condition (**Fig. 4**). Remarkably, metabolites fell only into two groups, one containing exclusively overabundant compounds in the mutant conditions when compared to the WT condition (**24**, **26**, **36**, **65**, **66**, **71** and **79**), the other containing exclusively underabundant compounds (**1**, **40**, **41**, **44**, **50**, **81**, **87** and **90**) (**Fig. 4**).

In the same vein concerning deletion mutant strains, three groups stand out. The first one includes one mutant condition corresponding to the ΔSpG8-2a deletion mutant strain, since is the only one that have no discriminating compounds when compared to the WT condition (**Fig. 4**). The second group includes four conditions which correspond to the ΔSpG8-1b, ΔSpG8-2b, ΔSpG8-5 and ΔSpG8-7a deletion mutant strains, as they have between two and seven discriminating metabolites. The third group includes three conditions which are the ΔSpG8-3, ΔSpG8-4 and ΔSpG8-7b deletion mutant strains, as they have the highest number of discriminating metabolites (between eight and thirteen) (**Fig. 4**).

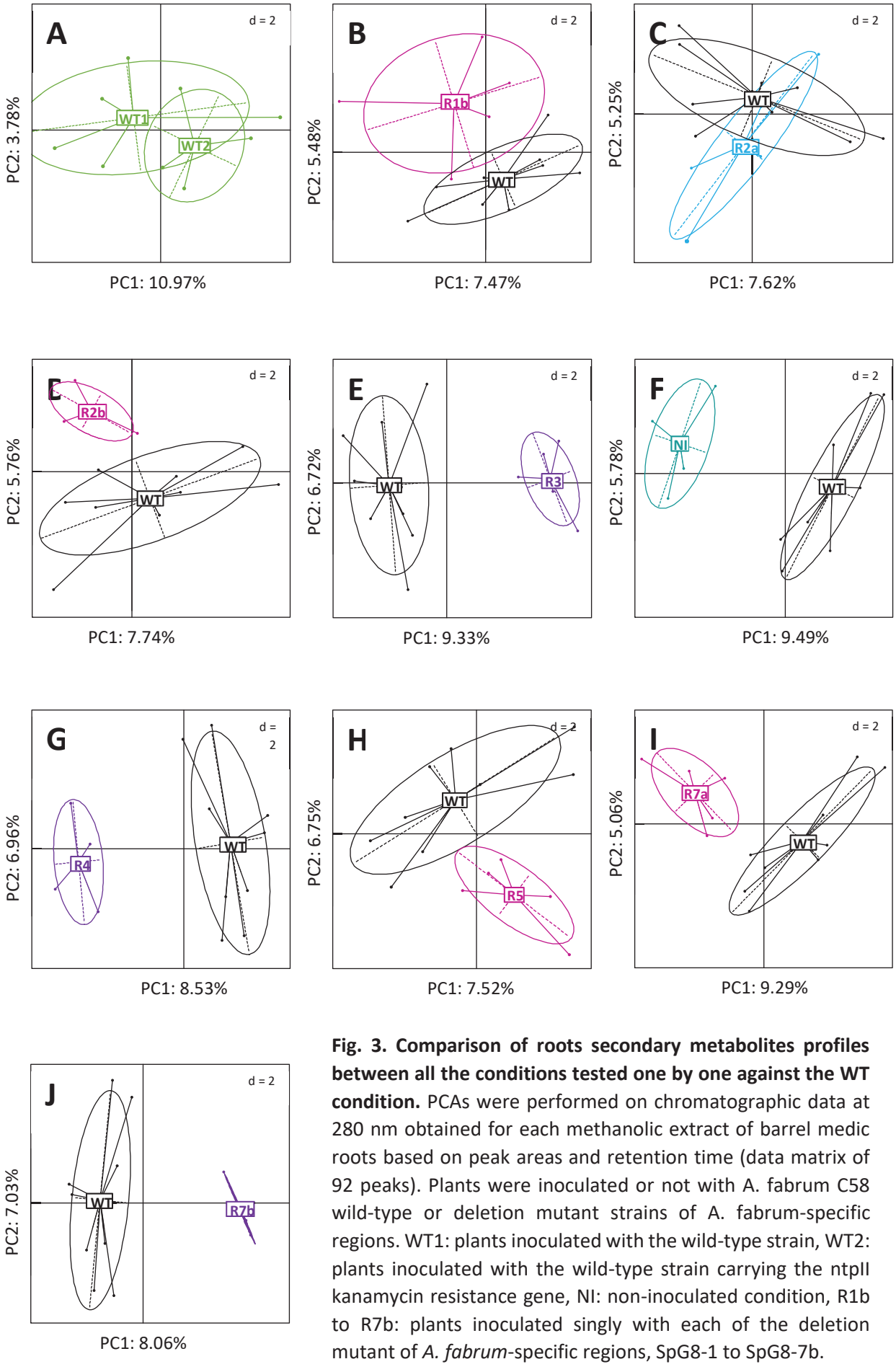


Fig. 3. Comparison of roots secondary metabolites profiles between all the conditions tested one by one against the WT condition. PCAs were performed on chromatographic data at 280 nm obtained for each methanolic extract of barrel medic roots based on peak areas and retention time (data matrix of 92 peaks). Plants were inoculated or not with A. fabrum C58 wild-type or deletion mutant strains of A. fabrum-specific regions. WT1: plants inoculated with the wild-type strain, WT2: plants inoculated with the wild-type strain carrying the ntpII kanamycin resistance gene, NI: non-inoculated condition, R1b to R7b: plants inoculated singly with each of the deletion mutant of *A. fabrum*-specific regions, SpG8-1 to SpG8-7b.

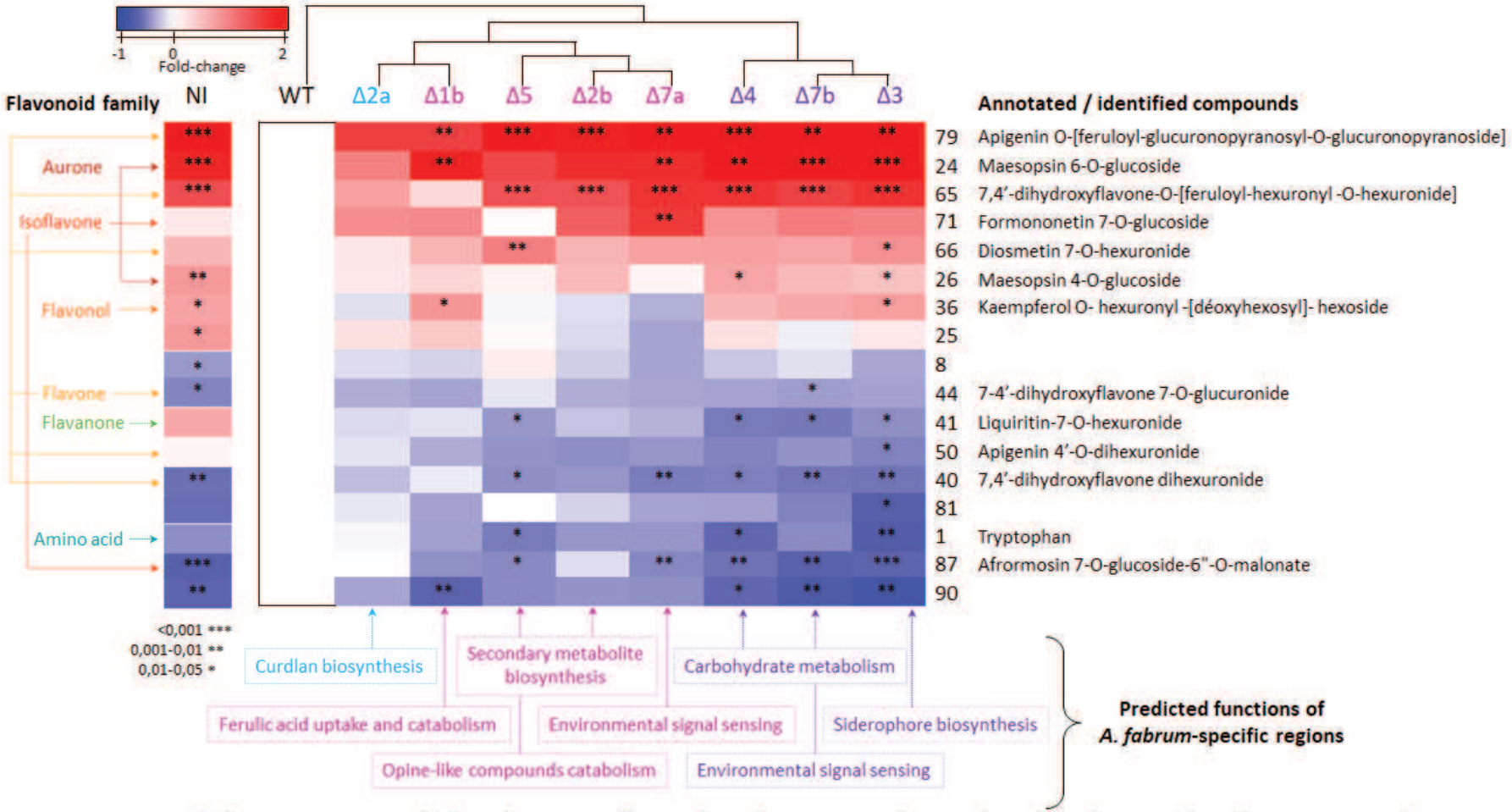


Fig 4. Heat-map of discriminating metabolites between all tested conditions according to their abundance in the plant. Compounds are over-abundant (red) or underabundant (blue) in non-inoculated (NI) or inoculated plants with each of the deletion mutant strains of *A. fabrum*-specific regions compared to the wild-type condition. Colors of mutant conditions are the same as in **Fig. 2** and **3** in order to respect the color code of each group. WT: plants inoculated with the wild-type strain, NI: non-inoculated condition, Δ2a to Δ3: plants inoculated singly with each of the deletion mutants of the *A. fabrum*-specific regions SpG8-2a to SpG8-3. Boxes indicate putative functions of *A. fabrum*-specific regions as proposed by Lassalle et al. (2011) based on gene annotation. Data were analyzed using analysis of variance. Statistical differences are indicated with the symbol * ($P < 0.05$).

Identification or annotation of discriminating metabolites

The UHPLC-UV/DAD-MS/MS QTOF data were explored in order to identify the discriminating compounds highlighted by statistical analyses. Study of the spectral data (UV-vis maxima; accurate mass; MS and MS/MS in positive and negative ionization mode) allowed the annotation and/or the identification of 13 compounds in different classes of metabolites by comparison to bibliographical data and analyses of standard compounds when available. Chemical data of the annotated discriminating compounds are shown in Table 1.

All but one discriminating metabolites are phenolic compounds belonging to the flavonoid family. We found two aurones (**24** annotated as maesopsin 6-O-glucoside and **26** maesopsine 4-O-glucoside), one flavanone (**41** annotated as liquiritin-7-O-hexuronide), two isoflavones (**71** formononetin-7-O-glucoside and **87** annotated as afrormosin-7-O-glucoside-6'-O-malonate), six flavones (**40** annotated as 7,4'-dihydroxyflavone dihexuronide, **44** annotated as 7-4'-dihydroxyflavone 7-O-glucuronide**, 50** annotated as apigenin 4'-O-dihexuronide, **65** annotated as 7,4'-dihydroxyflavone-O-[feruloyl-hexuronyl -O-hexuronide], **66** annotated as diosmetin 7-O-hexuronide and **79** annotated as apigenin O-[feruloyl-glucuronopyranosyl-O-glucuronopyranoside]) and one flavonol (**36** annotated as kaempferol O-hexuronyl-[déoxyhexosyl]-hexoside). There is also one amino acid (**1** tryptophan), whose accumulation was significantly different in some conditions.

Amino acid. **Compound 1** (Table 1, Fig. 1) had an *m/z* 205 $[M+H]^+$ ion that dissociated to give *m/z* 188 $[M+H-H_2O]^+$ ion in MSMS indicating the loss of a water molecule and a *m/z* 118 $[M+H-87]^+$ showing the presence of the indole moiety of the molecule. On the basis of retention time, UV-vis spectrum, fragmentation pattern, literature data (Fletcher et al. 2013; Jiang et al. 2011) and a standard comparison, this compound was identified as tryptophan.

Flavonoids. *Aurones*. Despite of the fact that **compound 24** and **26** (Table 1, Fig. 1) eluted at different times they had identical *m/z* 449 $[M-H]^-$ ions that fragmented to give product ions at *m/z* 287 $[M-H-162]^-$, 269 $[M-H-162-18]^-$, 259 $[M-H-162-18-10]^-$, indicating consecutive losses, first of an hexose, then a loss of carbonyl and finally a loss of $-H_2O$ respectively. According to the UV-vis spectrum, MS and MSMS spectral data, literature data (Stochmal et al. 2009; Yoshikawa et al. 1998; Thuy et al. 2004; Ye et al. 2009) and a standard comparison, X26 was identified as maesopsin 4-O-glucoside. Compound X24 was tentatively identify as isomer of compound X26 as maesopsin 6-O-glucoside (Li et al. 1997; Thuy et al. 2004).

Flavonoids. *Flavanone*. **Compound 41** (Table 1, Fig. 1) had an *m/z* 593 $[M-H]^-$ ion that dissociated to give a minor ion at *m/z* 417 $[M-H-176]^-$ indicating the loss of a glucuronic acid and a major ion at *m/z* 255 $[M-H-176\ -162]^-$ indicating the loss of a hexose and corresponding to aglycone of liquiritigenin; therefore, combined to literature data (Qiao et al. 2012; Zhou et al. 2017; Ma et al. 2016) and the aglycone standard comparison, this compound was tentatively identified as liquiritin-7-O-hexuronide.

Flavonoids. *Flavonol*. **Compound 36** (Table 1, Fig. 1) had an *m/z* 771,1950 $[M+H]^+$ ion that yielded fragments at *m/z* 625,1323 $[M+H-146]^+$, 463,0841 $[M+H-162]^+$ and 287,0542 $[M+H-176]^+$ indicating respectively the loss of a rhamnose, a hexose and a glucuronic acid. The MSMS negative mode showed an *m/z* 769,1841 $[M-H]^-$ that dissociated to give an ion at *m/z* 593,1523 $[M-H]^-$ indicating the loss of a glucuronic acid and another ion at *m/z* 285,0404 $[M-H]^-$ indicating a loss of a hexose and a rhamnose. On the basis of UV-vis spectrum, the fragmentation pattern and literature data, this compound was tentatively annotated as kaempferol O-hexuronyl-[déoxyhexosyl]-hexoside (Budzanowski 1991).

Flavonoids. *Flavone*. **Compound 40** (Table 1, Fig. 1) was tentatively identified as 7,4'-dihydroxyflavone dihexuronide according to the UV-vis spectrum, fragmentation pattern, literature data (Park et al. 2003; Wong et al. 2009; Ibrahim and Abul-Hajj 1990; Marczak et al. 2016; Singh et al. 2010) and aglycone standard comparison. Indeed, analysis of the fragmentation of the *m/z* 607 $[M+H]^+$ ion yielded fragments at *m/z* 431 $[M+H-176]^+$, 255 $[M+H-176-176]^+$ indicating successive losses of two molecules of glucuronic acid and showing the fragment aglycone 7,4'-dihydroxyflavone. The MSMS negative mode led to the conclusion that the two molecules of glucuronic acid are attached by the loss of 352amu (Marczak et al. 2016). Finally, the glucuronidation must be on the 4'-hydroxyl group since a hypsochromic shift is observed in band I (Singh et al. 2010). To the best of our knowledge, this flavone conjugate is reported for the first time in the plant kingdom.

Compound 44 (Table 1, Fig. 1) had an *m/z* 429 $[M+H]^+$ ion that fragmentated to give an ion at *m/z* 253 $[M+H]^+$ by losing one molecule of glucuronic acid. On the basis of the fragmentation pattern, literature data and the aglycone standard comparison, this compound was tentatively identified as 7-4'-dihydroxyflavone 7-O-glucuronide (Saleh et al. 1982; Singh et al. 2010; Staszków et al. 2011).

Compound 65 (Table 1, Fig. 1) was tentatively identified as 7,4'-dihydroxyflavone-O-[feruloyl-hexuronyl -O-hexuronide] according to the UV-vis spectrum, fragmentation pattern and aglycone standard comparison. Indeed, the fragmentation of the *m/z* 783 $[M+H]^+$ ion yielded two fragments at *m/z* 431 $[M+H-352]^+$ and 353 $[M+H-430]^+$ showing each the presence of half of the molecule, that is the aglycone 7,4'-dihydroxyflavone and a glucuronic acid for the first, and a glucuronic acid and a

ferulate for the second. A loss of one molecule of glucuronic acid was then produced to each of these ions, giving ions at *m/z* 255 $[M+H-352-176]^+$ and 177 $[M+H-430-176]^+$. The structure of the molecule is confirmed by the MSMS negative mode that showed ions at *m/z* 527 $[M-H-254]^-$ and 333 $[M-H-254-194]^-$ showing the loss of the aglycone followed by the loss of a ferulate (Guy et al. 2009; Arni et al. 2010).This compound has the exactly fragmentation pathway as compound **79** (Stochmal et al. 2001b). To the best of our knowledge, this flavone conjugate is reported for the first time in the plant kingdom.

The elution order of these phenolics followed indeed, the sequence of decreasing polarity (Farag et al. 2007), therefore **compound 40** eluted first, followed by compound **44** and **65**.

Compound 50 (Table 1, Fig. 1) had an *m/z* 623 $[M+H]^+$ ion that yielded fragments at *m/z* 447 $[M-H-176]^+$ and 271 $[M-H-176-176]^+$ indicating successive losses of molecules of glucuronic acid. This is confirmed in the MSMS negative mode by the loss of 270amu (the aglycone apigenin) and the formation of an *m/z* 351 ion indicating that the two molecules of glucuronic acid are attached. A hypsochromic shift is observed in band I compared to that of the apigenin, which seems to show that the substituent is on the nucleus B, that means that the glucuronidation must be on the 4'-hydroxyl group (Mabry et al. 1970). Indeed, the apigenin 7-O-diglucuronide, already found in *Medicago truncatula* (Jasiński et al. 2009), is also found in our extracts at RT 13.242 min, corresponding to compound **47**. Its UV spectrum, however, show an absorption spectra from of 266/340, indicating substitution occurring in ring A of the aglycone, unlike compound 50 (Stochmal et al. 2001a). Therefore, this data combined to literature data and the aglycone standard comparison, this compound was tentatively identified as apigenin 4'-O-dihexuronide. This compound is described for the first time in the plant kingdom.

Compound 66 (Table 1, Fig. 1) had an *m/z* 475 $[M-H]^-$ ion that dissociated to give a major ion at *m/z* 299 $[M-H-176]^-$ indicating the loss of a glucuronic acid and a minor ion at *m/z* 284 $[M-H-176-15]^-$ indicating the loss of the $-CH_3$ from the aglycone diosmetin. The analysis data, literature data (Beninger and Hall 2005; Li et al. 2016b; Ferreres et al. 2014; Greenham et al. 2003; Yu et al. 2015a) and the aglycone standard comparison, are consistent with this compound being tentatively identified as diosmetin 7-O-hexuronide.

Compound 79 (Table 1, Fig. 1) gave a protonated ion at *m/z* 799 $[M+H]^+$, which further fragmented to produce ions at *m/z* 353 $[M+H-446]^+$ and 177 $[M+H-446-176]^+$ suggesting the loss of a glucuronic acid and the aglycone apigenin followed by the loss of a second glucuronic acid. This compound is confirmed by a parallel fragmentation of the same *m/z* 799 $[M+H]^+$ ion giving product ions at *m/z* 447

$[M+H-352]^+$ and 271 $[M+H-352-176]^+$ indicating a loss of a glucuronic acid and a ferulate followed by the loss of a second glucuronic acid. This compound couple to literature data (Stochmal et al. 2001b; Jasiński et al. 2009; Fu and Wang 2015; Marczak et al. 2016; Kera et al. 2018; Marczak et al. 2010) and the aglycone standard comparison could therefore be apigenin O-[feruloyl-glucuronopyranosyl-O-glucuronopyranoside].

Flavonoids. Isof*lavone*. **Compound 71** (Table 1, Fig. 1) gave a deprotonated ion at *m/z* 475 $[M-H+HCOO]^-$ that fragmented to give ions at *m/z* 267 $[M-H-162]^-$ corresponding to a loss of a hexose and the formic acid and at *m/z* 252 $[M-H-162-15]^-$ indicating the loss of the $-CH_3$ from the aglycone formononetin. On the basis of retention time, UV-vis spectrum, fragmentation pattern, literature data (Park et al. 2003; Wen et al. 2007; Jung et al. 2013; Wu et al. 2003; Farag et al. 2007) and a standard comparison, this compound was identified as formononetin-7-O-glucoside (Ononin).

Compound 87 (Table 1, Fig. 1) gave a deprotonated ion at *m/z* 547 $[M+H]^+$ that fragmented to give ions at *m/z* 299 $[M-H-248]^+$ and 284 $[M-H-248-15]^+$ corresponding to a loss of a glucoside malonate and the loss of the $-CH_3$ from the aglycone afrormosin. On the basis of UV-vis spectrum, fragmentation pattern and literature data (Tibe et al. 2011; Zhang et al. 2014; Farag et al. 2007, 2008; Zhang et al. 2007b) this compound was tentatively identified as afrormosin-7-O-glucoside-6'-O-malonate.

The mass of all the discriminating plant metabolites was research in a bacterial extract. None of these metabolites was detected (Data not shown). Likewise, the bacterial metabolites found in the bacterial extract are not those identified in the plant.

Table 1. Metabolites identified by UPHLC-DAD-ESI-MS Q-ToF in *Medicago truncatula* in interaction with *A. fabrum* strains

No.	RT (min)	UV λmax (nm)	UHPLC-MS Q-ToF Analysis					UHPLC-MSMS Q-ToF Analysis		Proposed annotation	Identification procedure	References
			Ionization mode	Theoretical m/z	Observed ions m/z	Molecular formula	Δppm	Collision energy (V)	Fragments MS/MS (% base peak)			
1	2.388	218, 278, 288sh	+	205.097154	205.0969 $[M+H]^+$; 227.0789 $[M+Na]^+$	$C_{11}H_{12}N_2O_2$	-1.24	10	146.0593 (100); 188.0697 (74) indole moiety; 118.0644 (41) $[M-H_2O]^+$; 205.0851 (5);	Tryptophan	a, c	Fletcher et al. 2013 Jiang et al. 2011
			-	203.082601	203.0824 $[M-H]^-$;		-0.99	10	116.0505 (100) indole moiety; 203.0826 (68) 142.0669 (27); 159.0918 (17); 186.0553 (4) $[M-H_2O]^-$			
8	4.093	260, 296sh	+		449,1369 $[M+H]^+$					Unknown		
24	7.332	296, 346sh	+	473.105433	473.1110 $[M+Na]^+$	$C_{21}H_{22}O_{11}$	11.77	20	473.1020 (100) $[M+Na]^+$; 311.0519 (28) $[M-hex]^+$;	Maesopsin 6-O-glucoside	a	Li et al. 1997 Thuy et al. 2004
			-	449.108935	449.1094 [M-H]-		1.03	30	269.0461 (100) $[M-hex-H_2O]^-$; 259.0605 (20) $[M-hex-CO]^-$; 449.1073 (14) $[M-H]^-$ 287.0579 (5) $[M-hex]^-$;			
25	7.735	232, 278, 290sh								Unknown		
26	8.062	292; 343sh	+	473.105433	473.1035 $[M+Na]^+$	$C_{21}H_{22}O_{11}$	-4.08	30	311.0514 (100) $[M-hex]^+$; 473.1038 (21) $[M+Na]^+$; 185 (19);	Maesopsin 4-O-glucoside (Hovetrichoside C)	a, c	Li et al. 1997 Thuy et al. 2004
			-	449.108935	449.1094 $[M-H]^-$		1.03	20	259.0635 (100) $[M-hex-CO]^-$; 287.0588 (82) $[M-hex]^-$; 269.0483 (78) $[M-hex-H_2O]^-$; 449.1132 (65) $[M-H]^-$			
36	10.208	248sh, 262, 321sh, 354	+	771.197835	771.1935 $[M+H]^+$	$C_{33}H_{38}O_{21}$	-5.62	20	463.0841 (100) $[M-rhamnose-hex]^+$; 287.0542 (22) $[M-rhamnose-hex-hexu]^+$; 625.1323 (6) $[M-rhamnose]^+$; 771.1950 (1) $[M-H]^+$;	Kaempferol O-hexuronyl-[desoxyhexose]-hexoside	a	
			-	769.183282	769.1844 $[M-H]^-$		1.45	30	593.1523 (100) $[M-hexu]^-$; 258.0404 (20) $[M-hexu-hex-rhamnose]^-$; 769.1841 (17) $[M-H]^-$			
39	10.792	232, 273	+		515,0582 $[M+H]^+$					Unknown		
40	11.003	228, 252, 324	+	607.129361	607.1271 $[M+H]^+$	$C_{27}H_{26}O_{16}$	-3.72	20	255.0639 (100) $[M-hexu-hexu]^+$; 607.1260 (31) $[M+H]^+$; 431.0944 (3) $[M-hexu]^+$	7,4'-dihydroxyflavone dihexuronide	a, b	Park et al. 2003 Wong et al. 2009 Ibrahim and Abul-Hajj 1990 Marczak et al. 2016 Singh et al. 2010
			-	605.114808	605.1149 $[M-H]^-$		0.15	30	351.0571 (100) $[M-Aglycon]^-$; 175.0246 (30) $[M-Aglycon-hexu]^-$; 253.0514 (15) $[M-2hexu]^-$			
41	11.607	216, 228, 276, 310	+	617.147691	617.1441 $[M+Na]^+$	$C_{27}H_{30}O_{15}$	-5.82	40	441.1136 (100) $[M-hexu]^+$; 617.3226 (2) $[M-Na]^+$;	Liquiritin-7-O-hexuronide	a, b	Qiao et al. 2012 Zhou et al. 2017 Ma et al. 2016
			-	593.151194	593.1514 $[M-H]^-$		0.35	30	255.0669 (100) $[M-hexu-hex]^-$; 593.1507 (11) $[M-H]^-$;			

									417.1188 (2) [M-hexu]$^-$			
44	12.523	228, 252,	+	431.097273	431.0960 [M+H]$^+$	$C_{21}H_{18}O_{10}$	-2.95	20	255.0636 (100) [M-hexu]$^+$; 431.0973 (2) [M+H]$^+$;	4′,7-dihydroxyflavone-7-O-glucuronide	a, b	Saleh et al. 1982 Singh et al. 2010
		312sh, 330	-	429.082720	429.0830 [M-H]$^-$		0.65	20	253.0506 (100) [M-hexu]$^-$; 429.0818 (1) [M-H]$^-$;			
50	14.313	268, 286, 326	+	623.124276	623.1209 [M+H]$^+$	$C_{27}H_{26}O_{17}$	-5.42	20	271.0592 (100) [M-2hexu]$^+$; 623.1211 (44) [M+H]$^+$; 447.0920 (10) [M-hexu]$^+$	Apigenin 4′-O-dihexuronide	a, b	Mabry et al. 1970 Jasiński et al. 2009 Stochmal et al. 2001a
			-	621.109723	621.1103 [M-H]$^-$		0.93	20	351.0574 (100) [M-Aglucone]$^-$; 621.1100 (3) [M-H]$^-$			
65	17.992	254, 282sh, 322	+		783.1660 [M+H]$^+$	$C_{37}H_{33}O_{19}$		20	255.0610 (100) [M-hexu-ferulate-hexu]$^+$; 783.1640 (62) [M+H]$^+$; 353.0832 (18) [M-hexu-Aglycone]$^+$; 431.0900 (12) [M-hexu-ferulate]$^+$ 177.0518 (11) [M-2hexu-Aglycone]$^+$	7,4′-dihydroxyflavone-O-[feruloyl-hexuronyl -O-hexuronide]	a, b	Guy et al. 2009 Arni et al. 2010
			-		781.1711 [M-H]$^-$			20	527.1112 (100) [M-Aglycone]$^-$; 333.0519 (21) [M-Aglycone-ferulate]$^-$; 781.1765 (8) [M-H]$^-$			
66	18.328	251, 268, 346	+	477.102753	477.1009 [M+H]$^+$	$C_{22}H_{20}O_{12}$	-3.88	50	286.0456 (100) [M-hexu-CH_3]$^+$; 301.0707 (57) [M-hexu]$^+$; 258.0500 (51)	Diosmetin 7-O-hexuronide	a, b	Beninger and Hall 2005 Li et al. 2016b; Ferreres et al. 2014; Greenham et al. 2003 Yu et al. 2015 Park et al. 2003 Wen et al. 2007 Jung et al. 2013 Wu et al. 2003 Farag et al. 2007
			-	475.088200	475.0886 [M-H]$^-$		0.84	30	299.0570 (100) [M-hexu]$^-$; 284.0334 (52) [M-hexu-CH_3]$^-$;			
71	19.517	230, 252, 296	+	431.133659	431.1348 [M+H]$^+$	$C_{22}H_{22}O_9$	2.65	50	269.0794 (100) [M-hex]$^+$; 431.1348 (2) [M-H]$^+$	Formononetin-7-O-glucoside (Ononin)	a, c	
			-	475.124585	475.1284 [M-H+HCOOH]$^-$		8.03	50	252.0449 (100) [M-hex-HCOO-CH_3]$^-$; 267.0694 (16) [M-hex-HCOO]$^-$; 475.1284 (2) [M-H+HCOOH]$^-$			
79	21.305	270, 296sh, 324	+	799.171620	799.1615 [M+H]$^+$	$C_{37}H_{34}O_{20}$	-12.6	20	271.0570 (100) [M-2hexu-ferulate]$^+$; 353.0836 (62) [M-hexu-Aglycone]$^+$; 177.0525 (31) [M-2hexu-Aglycone]$^+$; 447.0876 (25) [M-hexu-ferulate]$^+$; 799.1616 (18) [M+H]$^+$	Apigenin 4′-O-[feruloyl-glucuronopyranosyl-O-glucuronopyranoside]	a, b	Stochmal et al. 2001b Jasiński et al. 2009 Fu and Wang 2015 Marczak et al. 2016
			-	797.157067	797.1673 [M-H]$^-$		12.84	30	527.1112 (54) [M-Aglycone]$^-$; 269.0490 (42) [Aglycone-H]$^-$; 797.1673 (1) [M-H]$^-$			
81	21.642	282, 324	+		563,1482 [M+H]$^+$					Unknown		
87	24.233	212sh, 224sh, 258, 318	+		547.1517 [M+H]$^+$	$C_{26}H_{26}O_{13}$		50	299.0928 (100) [M-glucoside malonate]$^+$; 284.0707 (47) [M-glucoside malonate-CH_3]$^+$;	Afrormosin-7-O-glucoside-6′-O-malonate	a	Tibe et al. 2011 Zhang et al. 2014 ; Farag et al. 2007, 2008 Zhang et al. 2007b
90	25.422	220, 274	+		515,6927 [M+H]$^+$ 992,4288					Unknown		
			-		990,4068							

hexu: hexuronic acid or glucuronic acid; hex: hexose or glucose.

a, comparison with literature MS, MS/MS and UV data. b, comparison with the aglycone moiety showing identical retention, UV and mass data. c, comparison with the pure compound showing identical retention, UV and mass data.

DISCUSSION

Phenolic compounds are often involved in plant-bacteria interactions (Miotto-Vilanova et al. 2019) and their content has been already shown to be affected in plant response to PGPR (Walker et al. 2011; Chamam et al. 2013). In this context, this study was focused on phenolic compounds using UHPLC-UV/DAD-ESI MS QTOF analyses to identify plant compounds whose relative abundances were significantly modified in plants inoculated with bacterial mutants. Therefore, the present study aimed at understanding the effect of *A. fabrum* and particularly that of its species-specific genes through its specific regions on root secondary metabolic profiles of *Medicago truncatula*. Our ultimate goal is to better understand *A. fabrum* interaction with plants during its rhizospheric lifestyle.

In the present study, the 92 secondary metabolites detected are present in all the conditions studied **(Fig. 2)**, suggesting that *A. fabrum*-specific regions would not induce novel metabolite biosynthesis, nor suppress their production by the plant. Considering all conditions combined, 17 secondary metabolites were found significantly different, of which 8 are overabundant and 9 are underabundant in root plants compared to the WT condition **(Fig. 4)**. So, as observed in other bacteria-plant interaction models (Valette et al. 2019), there is indeed a modulation of compounds in the plant incited by the bacteria. Even if it is hard to elucidate, one possible explanation of such modulations could be an induction or repression of certain biosynthetic pathways for these compounds. Another explanation could be the transformation by the plant of some compounds into others under the influence of bacteria or it is even possible that the bacteria degrades some of these compounds. Among phenolic compounds, almost all discriminating metabolites are flavonoids belonging to the subclasses of aurones, flavanones, isoflavones, flavones and flavonols.

Flavonoids, one of the major constituents in legume crops, are mostly found in plants as glycoconjugates (Golawska et al. 2010; Fu and Wang 2015; Marczak et al. 2010; Harborne and Williams 2000). The sugar moiety can be mono or oligosaccharides, the most frequently glucose, rhamnose or galactose. In some plant species such as *Medicago*, a species rich in flavonoids (Saleh et al. 1982), the glycosidic moiety contains also glucuronic acid and could be acylated by phenylpropenoic acids derivates such as p-coumaric, caffeic, ferulic or sinapic acids (Stochmal et al. 2001b; Marczak et al. 2010). Accordingly, in our study all flavonoids found were flavonoid O-glycosides containing a hexose or a glucuronic acid and two of our flavones were acylated with ferulic acid. Flavonoid glycosides plays an important role in plant physiology and biochemistry (Taylor and Grotewold 2005). These compounds can be involved in pathogenic and symbiotic interactions with microorganisms (Dixon et

al. 1994; Marczak et al. 2010), as for example, in the formation of the nitrogen-fixing nodules (Kobayashi et al. 2004). There is even a strong evidence that recognition of specific flavonoids by rhizobia is an important basis of host specificity (Spaink et al. 1987; Spaink 1994). Likewise, some of these compounds are modulated by the presence of the C58 strain which could mean that they are also important during the interaction with *Agrobacterium* (the *Rhizobiaceae* family).

The metabolic plant response upon *A. fabrum* inoculation

Among the 11 discriminating compounds highlighted on the metabolic profile comparison between the NI and the WT condition, two overabundant compounds are aurones and probably isomers, one annotated as maesopsin 6-O-glucoside (**24**), the other identified as maesopsin 4-O-glucoside (**26**) (**Fig. 4**). These two isomers are auronols, a small set of aurone derivatives. Maesopsin 4-O-glucoside (**26**) was first isolated from *Hovenia trichocarea* (Yoshikawa et al. 1998), then also isolated from *Artocarpus tonkinensis* leaves (Thuy et al. 2004), *Chaenomeles sinensis* fruits (Lee et al. 2002), *Medicago truncatula* roots (Stochmal et al. 2009), *Pseudolarix kaempferi* roots (Ye et al. 2009) and from stems and bark of *Jaffrea xerocarpa* (Muhammad et al. 2017). The biological activity of this molecule is quite diverse since a low cytotoxic activity against human KB cell lines was found (Muhammad et al. 2017), also a strong immunosuppressive activity in human cells (Thuy et al. 2004) and a strong TF inhibitory activity in rats, making it an attractive anti-coagulant drug (Lee et al. 2002). Instead, maesopsin 6-O-glucoside (**24**) has not been much studied and has just been reported as isolated from root bark of *Ceanothus americanus* (Li et al. 1997). Thus, even if its activity has not been tested on micro-organisms, this molecule is biologically active and it could be possible that this compound also has a toxic activity against bacteria. In such a case, it may be that the large amount of this compound in NI plants can be reduced in the presence of *A. fabrum* thanks to a particular capacity of this bacterium to reduce the production by plants of this toxic molecule.

The only discriminating flavonol, which is also overabundant, found in this study was annotated as kaempferol O- hexuronyl -[déoxyhexosyl]- hexoside (compound **36**). Kaempferol is a flavonoid widely distributed in plants, found in many edible plants or botanical products commonly used in traditional medicine. Sugars such as glucose, rhamnose, galactose and rutinose are usually bound to kaempferol to form glycosides. Kaempferol has been identified in many botanical families and has been found in *Pteridophyta*, *Pinophyta* and *Magnoliophyta* (Calderón-Montaño et al. 2011). Kaempferol and some glycosides of kaempferol have a wide range of pharmacological properties, among them anti-inflammatory (Park et al. 2009; Kim et al. 2010; De Melo et al. 2009), antioxidant (Kampkötter et al. 2007; Bonina et al. 2002; Sanz et al. 1994), and cardioprotective activities (Calderón-Montaño et al. 2011; Imran et al. 2019). Also, the scrutiny of kaempferol extraordinary list of cancer-

fighting properties highlights its full potential (Cui et al. 2008; Garcia-Closas et al. 1999; Nöthlings et al. 2007; Gates et al. 2007; Imran et al. 2019). Another property of kaempferol is its antimicrobial activity. Numerous papers have reported that kaempferol, its glycosides, or plants containing kaempferol have antibacterial (Kataoka et al. 2001; Habbu et al. 2009), antiviral (Lyu et al. 2005; Jeong et al. 2009; Min et al. 2001), antifungal and antiprotozoal activities (Barbosa et al. 2007; Calzada 2005). The only case of kaempferol O- hexuronyl -[déoxyhexosyl]- hexoside was reported in 1990 isolated from perianths of *Tulipu yesneriana* L. cv 'Paradae' and identified as kaempferol O-glucuronopyranosyl-[rhamnopyranosyl]-glucopyranoside (Budzanowski 1991).

The metabolic profile comparison between the NI and the WT condition also revealed four different flavones glycosides to be discriminating between them. Half of them are underabundant (**40** and **44**), half of them are overabundant (**65** and **79**) in the NI condition. Flavones, present in many higher plants and distributed in almost all plant tissues, have diverse functions in plants such as co-pigments, as antioxidants, and as antimicrobials (Zhang et al. 2007a). They even have various roles in the interaction with other organisms, for example, as signal molecules during the establishment of symbiosis between legumes and nitrogen fixing rhizobia (Dixon 1986; Schmelzer et al. 1988; Zhang et al. 2007a; Martens and Mithöfer 2005). Some glucuronidic derivatives of flavonoids have been shown to exert antioxidant and anti-inflammatory activities (Pawlak et al. 2010; Granica et al. 2013). Over 20 flavones are present in *M. truncatula* aerial tissues (Kowalska et al. 2007), being apigenin one of the aglycones of the majority of these compounds. Furthermore, many of these flavones are only glycosylated by glucuronic acid (Bruijn 2019). Diglucuronoflavones are particularly rare, they have been identified in a few species from Fabaceae (*Medicago truncatula*, *M. radiata*, *M. sativa*) (Kowalska et al. 2007; Stochmal et al. 2001a; Marczak et al. 2010, 2016).

Diglucuronoflavones seem indeed to be involved in the interaction with bacteria as three of the discriminating flavones contain diglucuronic acids in their glycosidic moiety (**40**, **65** and **79**), among them, two are acylated by ferulic acid (**65** and **79**). It is noteworthy that the four discriminating flavones are only glycosylated by glucuronic acid (**44**). Among these flavones glycosides, three discriminating 7,4'-dihydroxyflavones have been identified, one monoglucuronide (7-4'-dihydroxyflavone 7-O-glucuronide, compound **44**), one dihexuronide (7-4'-dihydroxyflavone 4'-dihexuronide, compound **40**) and one acylated dihexuronide (7-4'-dihydroxyflavone-O-[feruloyl-hexuronyl-O-hexuronide], compound **65**). The first two (compounds **40** and **44**) are underabundant in NI condition and the third one (compound **65**) is overabundant, when compared to the WT condition. Thus, it could be possible that compound **40** and **44** are intermediates leading to the biosynthesis of compound **65** (**Fig. 5**), as the glycoside and acylated moieties may attach one by one to the aglycone as it has already been

observed by (Farag et al. 2007). Compound **44** (7-4'-dihydroxyflavone 7-O-glucuronide) was already been found in *M. truncatula* roots but its biological activity was not described (Staszków et al. 2011). The other two compounds (**40** and **65**) found as such in our study have never been described before in the plant kingdom. The 7,4'-dihydroxyflavone has been demonstrated to enhance spore germination of two *Glomus* species (Tsai and Phillips 1991) and to accumulate in vesicular arbuscular mycorrhizal roots of *M. truncatula* after *Glomus versiforme* colonization. Thus, this compound is suspected to be required for signaling internal growth of the fungus (Harrison and Dixon 1993). Also, the 7,4'-dihydroxyflavone is one of the most potent *nod*-gene inducer of rhizobia in *Medicago truncatula* (Maxwell et al. 1992; Wasson et al. 2006; Zhang et al. 2007a, 2009) since it accumulates in its root exudates under nitrogen starved conditions (Zuanazzi et al. 1998). It thus disserves to be noted that this molecule is involved in a tight interaction established between *M. truncatula* and a bacterium which is a close relative of *A. fabrum*.

Compound **79** is overabundant in NI condition and was annotated as apigenin O-[feruloyl-glucuronopyranosyl-O-glucuronopyranoside], either acylated on sugar at C-4' or at C-7. Either of these compounds were previously reported in *Medicago truncatula* flower, leaf, stem and roots (Jasiński et al. 2009; Fu and Wang 2015; Kera et al. 2018; Marczak et al. 2010), in *M. sativa L* aerial parts (Shirley 1996; Stochmal et al. 2001b; Golawska et al. 2010) and in *Axyris amaranthoides* herb (Marczak et al. 2016) among others. To date, many types of apigenin glucuronides have been reported from *M. truncatula* (Kowalska et al. 2007; Marczak et al. 2010). Apigenin glycosides are one of the major flavones found in alfalfa aerial parts (Goławska et al. 2010). They have an important role in plant development and physiology, especially during their biotic interactions as with the pea aphid feeding behavior on alfalfa. Indeed, a negative correlation was found between pea aphid phloem sap ingestion and the concentration of apigenin glycosides (Golawska et al. 2010). Thus, this compound has been shown to be a feeding deterrent to herbivores and to have antifeedant and growth inhibitory effects on insects, especially on the pea aphid (Honda 1986; Feeny et al. 1988; Golawska et al. 2010).

While flavonoids are found throughout the plant kingdom, isoflavonoids have a more limited distribution and are particularly prevalent in the *Papilionoideae* subfamily of the *Leguminosae* (Farag et al. 2007; Gou et al. 2016; Guadalupe Soto-Zarazúa et al. 2016). The only isoflavone glycoside found as discriminating between the NI and the WT condition is compound **87**, annotated as afrormosin 7-O-glucoside-6"-O-malonate. Isoflavones are phenolic compounds with important implications for plant, animal and human health (Dixon and Steele 1999). They are involved in plant defense and UV protection (Dixon and Pasinetti 2010; Dixon and Steele 1999) and have estrogenic activity mainly

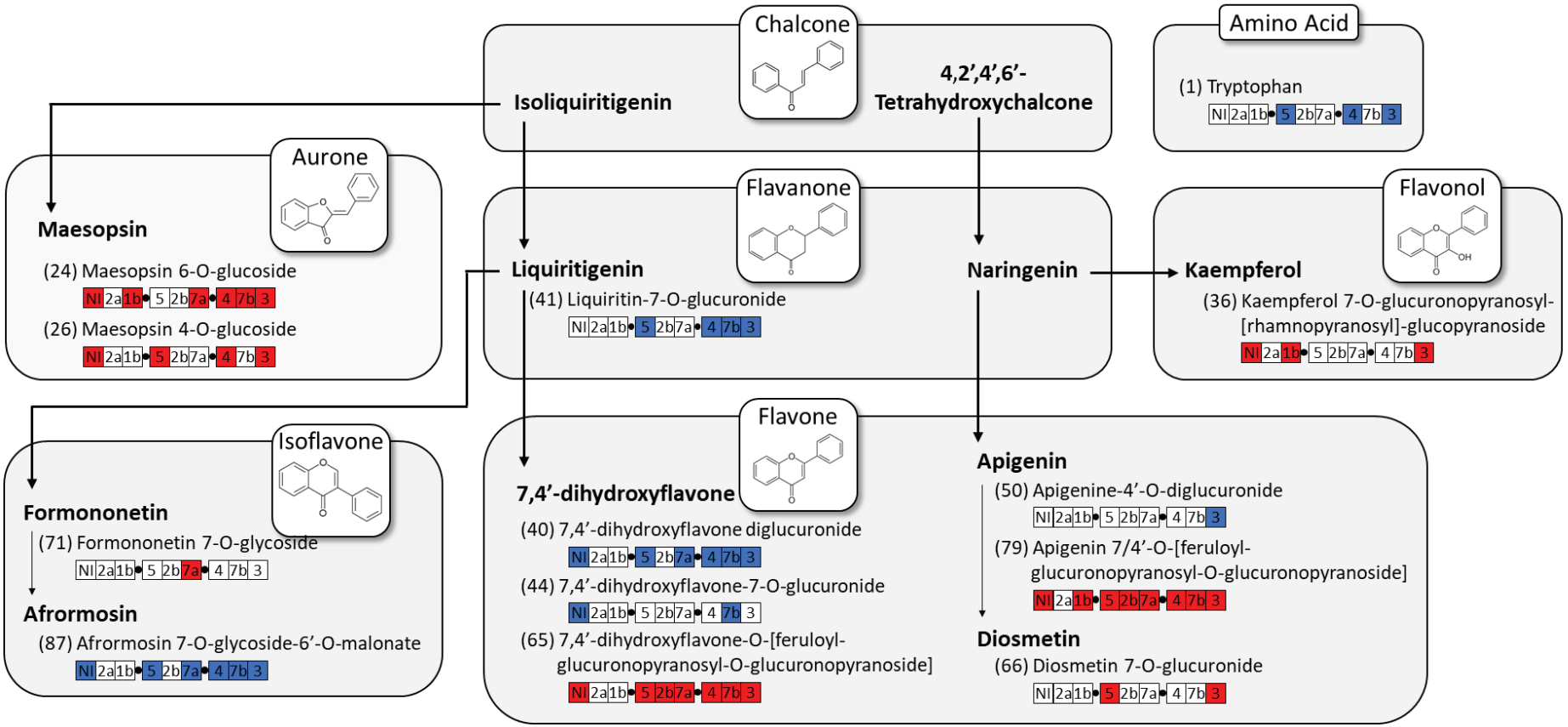


Fig 5. Flavonoid biosynthesis pathways in plants. Peak numbering corresponds to metabolite numbering in **Table 1** and in heatmap in **Fig. 4**. Discriminating compounds are arranged by flavonoid family according to the flavonoid biosynthesis pathway in plants. Boxes indicate plant conditions inoculated or not with each of the *A. fabrum* mutant strains. Colors in boxes correspond to the over-abundance (red) or under abundance (bleu) of each metabolite when compared to the abundance of the wild-type condition. NI: non-inoculated condition, Δ2a to Δ3: plants inoculated singly with each of the deletion mutants of the *A. fabrum*-specific regions SpG8-2a to SpG8-3.

attributed to their chemical structure similar to estradiol (Bedell et al. 2014; Guadalupe Soto-Zarazúa et al. 2016; Polasek et al. 2007; Wu et al. 2003). They possess antifungal, antimicrobial and antioxidant activities (Dakora and Phillips 1996; Ryan-Borchers et al. 2006) and can function as allelopathic agents or as signaling molecules mediating interactions with symbiotic bacteria like *Rhizobium*, and pathogenic microorganisms (Farag et al. 2007; Phillips and Kapulnik 1995).

Afrormosin (the aglycone in compound **87**) is found in a wide range of legume species (Dewick 1978; Al-Ani and Dewick 1980). This molecule might have a defensive role against both insects and fungi, but not an antimicrobial activity (Farag et al. 2008; Arnoldi et al. 1986). Afrormosin has been shown to be toxic to the moth *Pseudoplusia includens* or soybean loopers (Caballero et al. 1986) and several afrormosin glucosides possibly acts as a preformed insect deterrent (Wittstock and Gershenzon 2002). Furthermore, malonates are of biological interest in plants because this conjugated form can be utilized to store the less soluble isoflavone aglycones, and upon microbial infection, the aglycones are generated from the malonate conjugates (Sumner et al. 1996; Edwards et al. 1997). Afrormosin, afrormosin glucoside and afrormosin glucoside malonate are the major isoflavonoids in *M. truncatula* cell cultures (Farag et al. 2008, 2007) and have also been isolated from roots and cell suspension cultures of *M. sativa* L. (Kessmann et al. 1990). Afrormosin-7-O-glucoside-6''-O-malonate has also been isolated from *Trifolium pratense* L. (red clover) (Klejdus et al. 2001; Tava et al. 2016) and from roots of *Astragalus membranaceus* (Zhang et al. 2007b, 2014).

All discriminating compounds found have a biological activity and even if they have not been tested on *A. fabrum*, we can suppose, as they are modulated in *M. truncatula* roots, that these compounds also have an activity over bacteria.

Going further: root metabolome modulation by *A. fabrum*-specific regions

Our approach allows to go further and highlight the modulation of discriminating compounds according to *A. fabrum*-specific genes. Remarkably, all the metabolites modulated during inoculation are also modulated by *A. fabrum*-specific genes (except **8** and **25**). Besides, there are six more discriminating compounds modulated by *A. fabrum*-specific genes exclusively (**1**, **41**, **50**, **66**, **71** and **81**) when compared to the WT condition. Four of them are underabundant and two are overabundant **(Fig. 4)**. This could be compounds closely involved in the interaction between *A. fabrum* and the plant. At the same time, nine discriminating compounds (**24**, **26**, **36**, **40**, **44**, **65**, **79**, **87** and **90**) are shared between the NI condition and the plant conditions inoculated with the *A. fabrum* deletion mutant strains when they are all singly compared to the WT condition. This means that, regarding these specific metabolites, when the *A. fabrum*-specific regions are not present, there is no reaction from

the plant and its metabolism is not modified. *A. fabrum*-specific regions would therefore be involved or important in the recognition of the bacteria by the plant or could be directly linked to the plant-agrobacteria interaction, especially regarding these compounds.

While all *A. fabrum*-specific regions are expressed *in planta* in contact with *M. truncatula* rhizosphere (**Fig. S1**), which suggest that changes in plant metabolites are related to genes disrupted in bacterial strains, not all of them modify root metabolism. Indeed, even if the ΔSpG8-2a condition (i.e. plants inoculated by the ΔSpG8-2a strain) is expressed on roots, it highlights no discriminating compounds when compared to the WT condition (**Fig. 4**). SpG8-2a region main predicted function is curdlan biosynthesis, an exopolysaccharide induced under stress conditions, such as depletion of the nitrogen source and a low pH (Phillips and Lawford 1983; Kim et al. 1999; Yu et al. 2015b; Harada and Harada 1996). This exopolysaccharide surround and protect bacteria and its production is likely to increase the survival of bacteria in the soil (Ruffing and Chen 2012). Our first hypothesis to explain such a result is that the absence of curdlan has no direct role in the perception of the bacteria by the plant. Thus, the latter does not change its metabolism nor its phenolic content in the presence of the ΔSpG8-2a strain or the wild-type. Another hypothesis is that the lack of curdlan could be compensated by another exopolysaccharide produced by *A. fabrum*, the succinoglycan. Indeed, it is known that succinoglycan protects *A. fabrum* from environmental stress and that is produced in large amounts in this bacteria (Matthysse 2018).

A potential involvement in the flavonoid biosynthesis pathway (Metabolites the most modulated by *A. fabrum*-specific regions)

Compounds **24**, **65** and **79** are the most overabundant in *A. fabrum* mutant conditions when compared to the WT condition **(Fig. 4)**. Concerning compound **24** and its isomer (compound **26**), the first one seems to be more involved in plant-bacteria interactions because it discriminates for most of the *A. fabrum* deletion mutant conditions (ΔSpG8-1b, ΔSpG8-3, ΔSpG8-4, ΔSpG8-7a and ΔSpG8-7b) (**Fig. 4**). It could be likely that this set of *A. fabrum*-specific regions are involved in the pathway that leads to the reduction of compound **24** by plants. Either way, the plant produces less of these three compounds in contact with *A. fabrum* wild-type strain, or maybe the bacteria degrades it. Another hypothesis is that the plant transforms these compounds into other kind of molecules, following the pathway of flavonoid biosynthesis (**Fig. 5**). In contrast, compounds **40**, **87** and **90** are underabundant in most of the *A. fabrum* deletion mutant strains conditions when compared to the WT condition **(Fig. 4)**. This means that the relative abundance in the WT condition is higher for these compounds and that there is a plant response following bacteria inoculation.

A potential involvement in bacteria perception by the plant (*A. fabrum*-specific regions modifying the highest number of root secondary metabolites)

A. fabrum mutant conditions were grouped in three groups. One containing only a mutant condition (ΔSpG8-2a) with no discriminating compounds, a second group containing conditions with between two and seven discriminating metabolites (ΔSpG8-1b, ΔSpG8-2b, ΔSpG8-5 and ΔSpG8-7a) and a third containing conditions with the highest number of discriminating metabolites (ΔSpG8-3, ΔSpG8-4 and ΔSpG8-7b) when compared to the WT condition(**Fig. 4**). This means that the latter is probably composed of the specific regions involved in the bacterial recognition by the plant, since it has the largest number of metabolites.

Among this last group there is the ΔSpG8-3 mutant strain condition that induces in the plant an under abundance of seven compounds (**1**, **40**, **41**, **50**, **81**, **87** and **90**) and an overabundance of six compounds (**24**, **26**, **36**, **65**, **66**, **79**) (**Fig. 4**). This region (SpG8-3) encode the biosynthesis of a siderophore of a unique chemical structure helping *A. fabrum* to grow under iron limiting conditions (Rondon et al. 2004). In the rhizosphere, plant and bacteria compete for iron. But if the bacteria are no longer able to use it (as in the case of the ΔSpG8-3 mutant strain), then the plant will not lack iron and therefore will react differently compared to its reaction to the wild-type. Indeed, the strain profile lacking of siderophore production is the furthest from the WT condition profile and approximates to the NI condition profile (**Fig. 2**). Moreover, this specific region has two underabundant compounds that are exclusively discriminating for this region. They are compound **50** annotated as apigenin 4'-O-dihexuronide and compound **81** non identified (**Fig. 4**). This could mean that the plant reacts in an own specific way when it perceives this *A. fabrum*-specific region.

Furthermore, it is known that the SpG8-3 and the SpG8-1b have a confirmed link between each other. Indeed, a coordinated expression has been observed following the induction with HCAs and the fact that the presence of one of them is required for the expression of the other and vice versa (Baude et al. 2016). The close link between these two specific regions strengthen some more with our findings. Indeed, even if the mutation of these two specific regions do not cause the same metabolic changes in root, there is one compound (**36**, annotated as kaempferol O-hexuronyl-[déoxyhexosyl]-hexoside) overabundant exclusively in these two conditions (**Fig. 4**). This means that there is no plant reaction concerning this compound when there is no presence of either SpG8-1b or SpG8-3. One of our hypotheses is that these two *A. fabrum*-specific regions could be involved in the bacteria recognition. Another hypothesis is that *A. fabrum* could modulate this plant metabolite to construct its own niche through the expression of a particular combination of species-specific genes as compound **36** is known to have an antimicrobial activity (Kataoka et al. 2001; Habbu et al. 2009; Lassalle et al. 2017).

On the other hand, the ΔSpG8-4 mutant strain condition has a similar profile pattern to that of the ΔSpG8-3 mutant strain condition. Indeed, all nine discriminating compounds belonging to the ΔSpG8-4 mutant strain condition (**1**, **24**, **26**, **40**, **41**, **65**, **79**, **87** and **90**) are also discriminating for the ΔSpG8-3 mutant strain condition, even if the latter has four more discriminating compounds. This could mean that carbohydrate metabolism (the predicted function of the SpG8-4 region) can also be involved in the perception of bacteria by the plant.

The ΔSpG8-7b mutant strain condition modifies the relative abundance of eight compounds. Indeed, there is three overabundant (**24**, **65** and **79**) and five underabundant compounds (**40**, **41**, **44**, **87** and **90**) when plants are inoculated with this mutant strain (**Fig. 4**). SpG8-7b region main predicted function is environmental signal sensing and transduction. This region modifies root secondary metabolites that much probably because signal detection is essential to modulate bacterial metabolism which will therefore modify the perception of the bacteria by the plant. Indeed, compounds **40**, **44** and **65** have 7,4'-dihydroxyflavones as aglycones, already been suggested to be required for signaling internal growth of *Glomus versiforme* during *M. truncatula* colonization (Harrison and Dixon 1993). Moreover, compound **44** (annotated as 7,4'-dihydroxyflavone-7-O-hexuronide) is underabundant exclusively in this mutant strain (ΔSpG8-7b) reinforcing the idea of an involvement in the signal exchange between the bacteria and the plant.

Glucuronic acids, an important brick in the niche construction

The second group of *A. fabrum*-specific regions is composed of four regions that moderately modifies root secondary metabolites. These specific regions are SpG8-1b, SpG8-2b, SpG8-5 and SpG8-7a. One region of this group, the SpG8-5, modifies the relative abundance of five out of seven discriminating metabolites containing glucuronic acids (or hexuronic acids). Three of these five metabolites are overabundant and two are under abundant. The growth of this *A. fabrum*-specific region deletion mutant has been tested in presence of glucuronic acid and it has been found that it has a decreased growth compared to the wild-type (data not shown). This means this mutant is less prominent to use and degrade glucuronic acid. So, *A. fabrum* WT strain can uses this uronic acid as carbon source and its inoculation to the plant can diminish the relative abundance of these compounds as in the case of compounds **65**, **66** and **79**. We hypothesize that this is an important feature on the *A. fabrum* adaptation to the plant and to its niche construction in *M. truncatula*. Indeed, the sugar chains of numerous flavones in *M. truncatula* are composed of glucuronic acid (Kowalska et al. 2007; Bruijn 2019; Staszków et al. 2011; Stochmal et al. 2001a). Moreover, 8 metabolites out of the 13 identified or annotated in this study contain glucuronic acids (or hexuronic acids). Flavonoid glucuronides,

produced from flavonoids by UDP-glucuronosyltransferases (UGTs) in tissues of plants belonging to various taxa (Manach et al. 2004; Huang et al. 1999; Kowalska et al. 2007; Ringl et al. 2007), form a particular group among the flavonoid glycosides as they contain glucuronic acid moiety (or moieties) either instead or in addition to sugar units commonly present in plants (Marczak et al. 2010). Concerning the aerial parts of *M. truncatula*, one study found that among 26 flavonoid glycosides identified, there were 22 flavones having glucuronic acid moiety (or moieties) (Marczak et al. 2010) and another study found that, the moieties of flavones identified were composed exclusively of glucuronic acid (Kowalska et al. 2007). Finally, in another study, the most abundant compounds found in seedlings roots of *M. truncatula* were mono- and diglucuronides of isoflavones and/or flavones (Staszków et al. 2011).

Another sous-section

A compound in common with these three *A. fabrum*-specific regions (ΔSpG8-3, ΔSpG8-4 and ΔSpG8-7b), in addition to ΔSpG8-5, is compound **41** which is underabundant in these conditions when compared to the WT condition. Moreover, this compound is not modified in the NI condition. This metabolite is the only flavanone found as discriminating compound in this study and is annotated as liquiritin 7-O-hexuronide (or liquiritigenin 4'-hexosyl-7-O-hexuronide) **(Fig. 4)**. Flavanones are ubiquitous intermediates leading to the biosynthesis of all other flavonoid subclasses and, in particular, liquiritigenin serve as entry point into the flavone and isoflavone biosynthetic pathways (Farag et al. 2008) **(Fig. 5)**. Liquiritigenin has been shown to be a *nod* inducer gene, being then beneficial for the symbiosis legume-*Rhizobium* (Recourt et al. 1991). Liquiritigenin has been isolated from *M. truncatula* leaves, roots and seeds (Banasiak et al. 2013; Farag et al. 2007; Ishiga et al. 2015; Li et al. 2016a), from *Vitextrifolia L.*, *Glycyrrhizae* (Ma et al. 2016), alfalfa seeds and root exudates of *Vicia sativa* subsp. *nigra* inoculated with infective *Rhizobium leguminosarum* biovar *viciae* (Recourt et al. 1991). This flavanone glycoside has already been described (Ma et al. 2016; Qiao et al. 2012; Zhou et al. 2017) but not in the context of interaction between bacteria and plants.

The third group of *A. fabrum*-specific regions is composed of four regions that moderately modifies root secondary metabolites. These specific regions are SpG8-1b, SpG8-2b, SpG8-5 and SpG8-7a. Two metabolites share by this two group of regions are compound **1** and **66**. The first one is underabundant in ΔSpG8-3, ΔSpG8-4 and ΔSpG8-5 mutant strain conditions, while the second is overabundant in ΔSpG8-3 (siderophore biosynthesis) and ΔSpG8-5 (opine-like compounds catabolism). Compound **1** is the only one that is not a flavonoid but an amino acid and identified as tryptophan. Inoculation of the *A. fabrum* WT strain induced an over-abundance of tryptophan in *M. truncatula* roots (**Fig. 4**). Tryptophan is a central precursor of secondary metabolites and auxins in plants (Ishihara

et al. 2011; Bennett and Wallsgrove 1994). It has already been observed that PGPR can increase tryptophan accumulation, consequently causing a better PGPR effect (Chamam et al. 2013). This is consistent with our results as *A. fabrum* can be considered as a PGPR (Walker et al. 2013). Compound **66**, a flavone glycoside, was annotated as diosmetin O-hexuronide. This compound has already been isolated from the fruits of *Luffa cylindrica* ((Du and Cui 2007), from *Chrysanthemum morifolium* L. v. Ramat leaves (Beninger and Hall 2005), from aerial parts of *Grindelia robusta* (Ferreres et al. 2014) and as one of the major components of *Dracocephalum moldavica* L (Li et al. 2016b; Yu et al. 2015a).

Almost all metabolites modulates by this *A. fabrum*-specific regions group are part of the metabolites which are modified by the second group of mutants already described, with the exception of compound **71**, which is overabundant exclusively when the plants are inoculated with the mutant of the SpG8-7a region (environmental signal sensing and transduction). Compound **71**, identified as formononetin-7-O-glucoside (ononin). Ononin was isolated for the first time from flowers of *Ononis speciosa* (Barrero et al. 1989). Formononetin is a phytoestrogen, which had many well-known pharmacological activities (Xu et al. 2006), such as immunological enhancement actions (Zhang and Han 1994) and an inhibitory effect on mouse brain monoamine oxidase (Hwang et al. 2005). Formononetin has not been shown to have antimicrobial activities but is a precursor of the isoflavonoid phytoalexins produced in alfalfa in response to stresses and microbial infections (Dalkin et al. 1990; Deavours et al. 2006). There is evidence that formononetin and ononin have neuroprotective and antioxidant effects (Yu et al. 2005, 12). Concerning mycorrhization on *M. truncatula* roots with *Glomus intraradices*, it was shown that in the late stages, there is enhanced levels of isoflavonoids, and in particular the contents of ononin were significantly higher (Schliemann et al. 2008). That is also the case on alfalfa roots with *Glomus intraradix*, indeed ononin levels are significantly greater in inoculated roots (Volpin et al. 1994). In contrast, ononin was highly accumulated at the beginning of *M. sativa* root colonization by *Glomus mosseae* and *Glomus intraradices*, but not with *Gigaspora rosea*. Ononin seem to play a role at the beginning of root colonization and to have possibly, different genus-specific signaling requirements of arbuscular mycorrhizal fungus (Larose et al. 2002). This isoflavone glycoside has already been isolated from the leaves of *M. truncatula* (Marczak et al. 2010; Schliemann et al. 2008) and is one of the most predominant flavonoids in the roots of this plant (Farag et al. 2007). This metabolite has also been isolated from the roots of several plants, such as *Pueraria lobata* Ohwi (Jun et al. 2003; Zhang et al. 2005), *Astragalus membranaceus* (Jung et al. 2013; Zhang et al. 2007b), *Astragalus mongholicus* (Zhang et al. 2007b), *Sophora subprostrata* (Park et al. 2003) and *Glycyrrhiza uralensis* Fisch. (Qiao et al. 2012). Also, from the extracts of several species of *Trifolium* (Polasek et al. 2007; Wu et al. 2003; Tava et al. 2016), and described in a plant decoction (Xiao et al. 2005; Xu et al. 2006; Qi et al. 2006; Wen et al. 2007).

CONCLUSION

Metabolomic analyses showed that *A. fabrum* inoculation modulates phenolic compounds content in roots, particularly flavonoids. These molecules play a key role when it comes to plant-bacteria interactions as they can be signaling molecules for bacteria or serve as markers in the adaptation of *A. fabrum* to the plant. Furthermore, we showed that this modulation was essentially link to *A. fabrum*-specific genes. These results contribute to a better understanding of the ecological niche construction of *A. fabrum* highlighting the importance of its specific genes in the establishment of this fine-tuned interaction.

MATERIALS AND METHODS

Bacterial strains, plasmids and culture conditions

The *Agrobacterium fabrum* strains used in this study are listed in Table 2. The strains were grown over-night with shaking (160rpm) at 28°C in YPG-rich medium (yeast extract, 5g.L^{-1}; peptone, 10g.L^{-1}; glucose, 10g.L^{-1}; pH 7.2). Tryptophane was obtained from Agilent Technologies (Waidbronn, Germany). Flavonoids (i.e. 7-4'-dihydroxyflavone, liquiritigenin, diosmetin, ononin, apigenin, kaempferol and kaempferol 3-O-glucoside) were obtained from Extrasynthese (Genay, France). Maesopsin 4-O-glycoside was kindly provided by Laurence Voutquenne-Nazabadioko from the Institut de Chimie moléculaire de Reims (ICMR) UMR CNRS 7312.

Construction of the deletion mutants and transcriptional fusions

A. fabrum deletion mutants were constructed by mutagenic polymerase chain reaction (PCR) according to a strategy previously described by Lassalle et al. (2011). Briefly, the recombinant region containing the flanking fragments of the sequence to be deleted of C58 (primers listed in Table X) and a fragment encoding the nptII kanamycin resistance gene amplified from plasmid pKD4 (Datsenko and Wanner 2000) were cloned into pLQ200SK (Quandt and Hynes 1993) and then inserted in C58 by electroporation. Single-crossover integration was selected by kanamycin and neomycin resistance on YPG medium plates. Double crossover events were identified by sucrose resistance on YPG media. Deletion mutants were verified by diagnostic PCR with appropriate primers.

eGFP transcriptional fusions were obtained as follows. The promoter regions of the targeted genes in each *A. fabrum* specific region were PCR amplified with specific primers listed in Table S1. Then, PCR fragments obtained were cloned into the pGEM-T vector (Promega, Madison, WI) according to

Table 2. Strains and plasmids used in this study

Strains	Relevant genotype and features	Reference
Escherichia coli		
JM109	*endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB+ Δ(lac-proAB) e14- [F'traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)*	NEB catalog
Agrobacterium fabrum		
C58	Wild-type	CFBP 1903
C58 Kan	Wild-type, Kan^R	
ΔSpG8-1b	C58 deleted of *atu1409* - *atu1423* cluster, Kan^R	Lassalle *et al.* 2011
ΔSpG8-2a	C58 deleted of *atu3054* - *atu3059* cluster, Kan^R	Lassalle *et al.* 2011
ΔSpG8-2b	C58 deleted of *atu3069* – *atu3073* cluster, Kan^R	Present work
ΔSpG8-3	C58 deleted of *atu3663* – *atu3693* cluster, Kan^R	Baude *et al.* 2016
ΔSpG8-4	C58 deleted of *atu3808* – *atu3830* cluster, Kan^R	Present work
ΔSpG8-5	C58 deleted of *atu3947* – *atu3952* cluster, Kan^R	Present work
ΔSpG8-7a	C58 deleted of *atu4285* – *atu4294* cluster, Kan^R	Present work
ΔSpG8-7b	C58 deleted of *atu4295* – *atu4307* cluster, Kan^R	Present work
Plasmids		
pGEM-T	Cloning vector for PCR fragments; Amp^R; lacZ	Promega
pGEM-Teasy	Cloning vector for PCR fragments; Amp^R; lacZ	Promega
pOT1e	Promoter-probe vector based on pBBR1MCS-5 replicon; contains promoterless eGFP and MCS between two transcriptional terminators; Gm^R	Allaway *et al.* 2001
pJQ200SK	Suicide vector; P15A sacB; Gm^R	Quandt *et al.* 1993
pGEM-T P*atu1416*	Upstream region of *atu1416* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3057*	Upstream region of *atu3057* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3069*	Upstream region of *atu3069* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3675*	Upstream region of *atu3675* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3817*	Upstream region of *atu3817* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3948*	Upstream region of *atu3948* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu4292*	Upstream region of *atu4292* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu4299*	Upstream region of *atu4299* gene cloned into pGEM-T; Amp^R	Present work
pOT1E P*atu1416*	Upstream region of *atu1416* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3057*	Upstream region of *atu3057* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3069*	Upstream region of *atu3069* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3675*	Upstream region of *atu3675* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3817*	Upstream region of *atu3817* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3948*	Upstream region of *atu3948* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu4292*	Upstream region of *atu4292* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu4298*	Upstream region of *atu4298* gene cloned into pOT1E; Gm^R	Present work

manufacturer's instructions. After digestion of the resulting plasmids with HindIII and SalI, fragments were subcloned into pOT1e (Allaway et al. 2001) digested with the same enzymes. Finally, transcriptional reporter constructions were inserted into C58 by electroporation and verified by diagnostic PCR with appropriate primers.

Plant material and plant bacteria co-culture procedures

All plant experiments (confocal observation and metabolomic study) were conducted on *Medicago truncatula* Gaertn (cultivar Jemalong line A17). Barrel medic seeds were surface sterilized by soaking in a 3.5% sodium hypochlorite solution before being thoroughly rinsed 5 times with sterile distilled water. In order to vernalize the seeds, they were put in the dark at 4°C overnight before use. The seeds were then scarified and placed on 0.8% agar plant medium (A7921, Sigma Aldrich, St Louis, MO, USA) supplemented with nutrient solution (1.5 g/L, Plant-Prod 15-15-30, Fertil s.a.s., Le Syndicat, France). For confocal observation, bacterial cells harboring the transcriptional fusions constructed were inoculated with 10µL of overnight culture ($2x10^5$ bacteria.mL^{-1}) on the *M. truncatula* seeds placed on plates containing 0.8% agar plant cell culture supplemented with the nutrient solution at 1.5 g/liter. To perform the metabolomic study, *Agrobacterium* strains were inoculated in the supercooled agar at a concentration of $1x10^7$ bacteria.mL^{-1} of agar plant medium. One group of barrel medic seeds was cultivated without bacterial inoculation to carry out the non-inoculated (NI) condition. All petri dishes were placed vertically and the lower part was covered to protect the root development from light. The plants were incubated for 14 days in a phytotron with a 16 hours light daily and a temperature between 24°C and 28°C.

Confocal microscopy observation

Microscope observation of the constructed reporter bacteria harboring the transcriptional fusions on *M. truncatula* roots was performed using a confocal laser scanning microscope (LSM 800 Meta Confocal Microscope, Zeiss, Oberkochen, Germany). eGFP is expressed under the control of the promoter of each targeted gene of each *A. fabrum* specific region and the green color indicates when the induction takes place. At 14 dpi, *M. truncatula* roots were mounted between a slide and a coverslip in a commercial mounting fluid (Aqua Poly/Mount, Polysciences, Inc., Warrington, PA). eGFP was excited with argon laser at 488 nm, and fluorescence was captured at 528nm. Analyses of images (five plants per condition) were performed with the LSM 800 software (Zeiss, Oberkochen, Germany).

Roots harvesting and extraction of phenolic compounds

After 15 days of incubation, the roots were harvested, pooled, frozen in liquid nitrogen and freeze-dried 24 hours. They were next grinded using a bead mill TissueLyser II system (Qiagen, Hilden,

Germany). Four or five replicates were performed per condition, each replicate consisting of root pools of several seedlings. For the extraction targeting phenolic compounds, samples of 100 mg of dried powder were first extracted twice with MeOH 80% and then extracted twice with MeOH 100% with homogenization and sonication at room temperature. The combined extracts were then centrifuged (14 500 rpm, 10 min) and the supernatants, particle free, were evaporated to dryness at 30°C in a Centrivap concentrator (Labconco, USA) to constitute the dried crude extracts. These extracts were weighed, dissolved in MeOH 60% at 10 mg/ml and conserved at -80°C prior to analysis. A quality control (QC) sample was prepared containing 20µL of each of the extracts obtained.

Metabolites analysis by UHPLC-UV/DAD-ESI-MS QToF

Metabolites analysis was performed on an UHPLC Agilent 1290 coupled to a UV-vis Diode Array Detector (Agilent 1290 Infinity series) and an Accurate-Mass Q-ToF 6530 spectrometer (Agilent Technologies). Liquid chromatography was carried out using a Nucleodur Sphinx RP C_{18} column (2mm × 100mm, 1.8µm, Macherey-Nagel) maintained at 50°C with an injection volume of 3 µL of sample.
The mobile phase was a mixture of acetonitrile and acidified ultrapure water (0.4% acetic acid) applied at a flow-rate of 0.5mL/min, and with a gradient (acidified H_2O:CH_3CN, [v/v]) starting at 98:2 for 1.5min, increasing to 76:24 at 26min, going to 0:100 in 6min and maintained for 3min, before returning to the starting conditions in 1min and equilibrating for 2min. The QToF-MS instrument was operated for MS analyses under the following conditions : the ion source ESI (Agilent Jet Stream thermal gradient focusing Technology) in positive ionization mode, with a 320°C gas temperature, a 11-liter/min gas flow, a nebulizer pressure at 40 psi, a 360°C sheath gas temperature with 12-liter/min flow rate, and with the capillary, nozzle, and fragmentor voltages at 3,000 V, 500 V, and 150 V, respectively. The acquisition mass range was of m/z 100 to 2000. The QC sample was initially analyzed and then inhected after every 8-9 samples in the run sequence to monitor the repeatability of the analysis. Chromatograms were explored at 280 nm with Mass Hunter qualitative analysis B.07.00 software. For characterization of discriminating compounds, complementary MS and tandem MS experiment were performed in positive and negative ionization modes and with different collision energies (10, 20, 30 or 40V) in AutoMS/MS mode. The UHPLC-UV/DAD-MS QToF device was managed by the Mass Hunter Workstation Acquisition B.07.00 software and the resulting data was reprocessed with the Mass Hunter Qualitative Analysis B.07.00 software (Agilent Technologies).

Metabolite profiling data treatment and statistical analyses

First, integration of each peak of the UV-chromatograms at 280nm was performed and a pick alignment was made between all samples based on peak retention time, resulting in a metabolites matrix with samples in rows and absolute areas (A_{abs}) of detected metabolites in columns. Then for

each metabolite within a given sample, its relative peak area (i.e. relative abundance in the sample= A_{rel}) was calculated over the total area of the sample chromatogram (ΣA_{abs}) as follows: $\%A_{rel} = A_{abs} / \Sigma A_{abs} \times 100$. Only the most abundant metabolites (relative abundance >1%) were selected to continue the study. Partial least squares discriminating analysis (PLS-DA) was made from this matrix to visualize the global effect of bacterial inoculation (WT strain, mutant strains and the NI condition) on root phenolic profiles of *M. truncatula*. Principal Component Analyzes (PCAs) were made between the WT strain and each mutant strain to stand out the effect of each *A. fabrum* specific regions on root secondary metabolites profiles. The statistical study was oriented to a t-test to compare each condition with the WT strain condition in order to highlight significantly different metabolites. Statistical analyses were performed using R software (R Studio version 0.99.491).

Identification of discriminating root plant secondary metabolites

Identification or annotations of discriminating metabolites were performed by the analysis of the UV-vis spectrum and the positive and negative MS and MSMS spectra. These data were compared to literature data, first on compounds already described in *M. truncatula* but also on chemical data of metabolites from other plants. When possible, comparisons with standards (the entire molecule or the aglycone moiety) allowed confirming metabolite identity or at least the aglycone nature.

SUPPLEMENTARY DATA

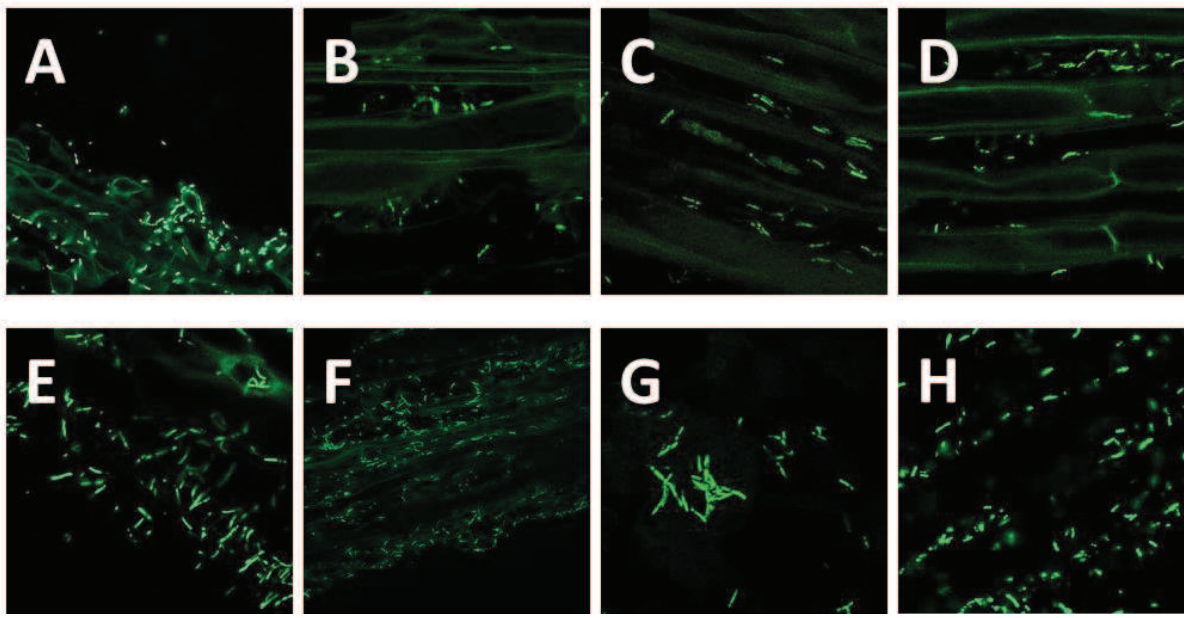


Fig S1. *In planta* expression of the *A. fabrum* specific regions through the expression of their targeted gene on *Medicago truncatula* roots. A. pOT1e *Patu1416* belonging to SpG8-1b. B. pOT1e *Patu3057* belonging to SpG8-2a. C. pOT1e *Patu3073* belonging to SpG8-2b. D. pOT1e *Patu3675* belonging to SpG8-3. E. pOT1e *Patu3817* belonging to SpG8-4. F. pOT1e *Patu3948* belonging to SpG8-5. G. pOT1e *Patu4292* belonging to SpG8-7a. H. pOT1e *Patu4299* belonging to SpG8-7b. Gene expression was monitored using transcriptional fusions by confocal microscopy at 14 dpi. Representative pictures from five plants per specific region are shown. Green fluorescence showed bacteria able to express the corresponding transcriptional fusion. Plant auto fluorescence represented in green allowed to distinguish plant cells. All transcriptional fusions are induced in contact with the entire *M. truncatula* root system.

Table S1. Primers used in this study

A. fabrum-specific region	Analysis and region	Forward primer (5' to 3')	Reverse primer (5' to 3')	Reference
Transcriptional fusions				
SpG8-1b	*Atu1416* promoter			Meyer *et al.* 2018
SpG8-2a	*atu3057* promoter	AAGCTTTGACGGTTGTGAACAGCACT	GAGGTGAATACAGGCGGAAA	This study
SpG8-2b	*atu3073* promoter	AAGCTTACTGAAACCGACATGAACGC	CCGCCGAGAATTCCGATAGT	This study
SpG8-3	*atu3675* promoter	AAGCTTAAATCGCGTTCTCCAGATGG	GGGTCTCTACTGGCTCGATG	This study
SpG8-4	*atu3817* promoter	TGGTTCCCAGACGTCGTTTA	CAGTTTGATGTAGCGAGCCA	This study
SpG8-5	*Atu3948* promoter	AAGCTTGAAGGTCGGTGGCATTGT	ACGGCTCTTGCTTGTTCG	This study
SpG8-7a	*atu4292* promoter	ATCGATGTGCAGAGCTTGCTGACG	CAGGGTGATGTGGAGATCGT	This study
SpG8-7b	*atu4299* promoter	AAGCTTCGAGCCATTTCATGAGTGCT	AGGAGACAATGCAACCCGTA	This study
	Insertion verification in pGEMT	GTTTTCCCAGTCACGAC	CAGGAAACAGCTATGAC	Promega
	Insertion verification in pOT1e	CGGTTTACAAGCATAAAGC	CATTTTTTCTTCCTCCACTAG	Pothier *et al*. 2007
Construction of the *A. fabrum*-specific regions deletion mutants (inactivation gene clusters)				
ΔSpG8-1b	Upstream region of *atu1409*			Lassalle *et al.* 2011
	Downstream region of *atu1423*			
ΔSpG8-2a	Upstream region of *atu3054*			Lassalle *et al.* 2011
	Downstream region of *atu3059*			
ΔSpG8-2b	Upstream region of *atu3069*			This study
	Downstream region of *atu3073*			
ΔSpG8-3	Upstream region of *atu3663*	CCGGTTCTACATCCTGGAAA	CCTGCTCAACAGGCTACTCC	Baude *et al.* 2016
	Downstream region of *atu3693*	GACAACATGCCCCTCTCCTA	TCTGGAACGTCACCGACATA	
ΔSpG8-4	Upstream region of *atu3808*			This study
	Downstream region of *atu3830*			
ΔSpG8-5	Upstream region of *atu3947*			This study
	Downstream region of *atu3952*			
ΔSpG8-7a	Upstream region of *atu4285*			This study
	Downstream region of *atu4294*			
ΔSpG8-7b	Upstream region of *atu4295*			This study
	Downstream region of *atu4307*			
	nptII gene	TTGCTGCGCGCGACATCAAGGTTTCGACCGAGGAGT AGCCTGTTGAGCAGGTGTGTAGGCTGGAGCTGCTTC	TGAAAATGCCGCTGTATTTCCTCGATCACGTAGGA GAGGGGCATGTTGTCCATATGAATATCCTCCTTA	Baude *et al.* 2016
	Region for inactivation diagnosis	GAGAGTGACGCTTTGGCTCT	GGTTGATCTGGTGCAGCTTT	Baude *et al.* 2016

Part II

Metabolomic study of *A. fabrum* pathogenic lifestyle in plant tumors

A. Secondary metabolites

INTRODUCTION

The objective of this part is to determine whether the *A. fabrum*-specific regions are likely to modify secondary metabolite profiles produced in the tumor of plants inoculated with *A. fabrum*. Indeed, given the strong interconnection between *A. fabrum*-specific regions and the plant, they are thus suspected to have an impact on its pathogenic lifestyle specially via the modulation of secondary metabolites. To this end, a metabolomic approach is here performed using deletion mutant strains of each *A. fabrum*-specific regions inoculated singly on the plant. We used tomato plants (*Solanum lycopersicum*, Marmande variety) cultivated in a greenhouse, inoculated or not with the wild-type strain or each of the deletion mutants of the *A. fabrum*-specific regions.

Analyzes were performed by ultra–high pressure liquid chromatography with ultraviolet and electrospray ionization-mass spectrometry (UHPLC-UV/DAD-ESI MS QTOF). A comparison of tumor profiles was then performed between plants inoculated with the mutant strains with those inoculated with the Wild-type strain. The latter was also compared to a Non-inoculated condition (NI). Finally, discriminating metabolites were characterized performing tandem mass spectrometry (MS/MS). An analysis of the identified metabolites will be performed afterwards. This will bring new elements to evaluate the bacterial influence on plants through its specific-genes and thus achieve the construction of its own and specific niche during bacterial-plant interactions.

MATERIALS AND METHODS

Bacterial strains, plasmids and culture conditions

The *Agrobacterium fabrum* strains used in this study are listed in **Table 1**. *A. fabrum* strains were grown over-night with shaking (160rpm) at 28°C in YPG-rich medium (yeast extract, 5g.L^{-1}; peptone, 10g.L^{-1}; glucose, 10g.L^{-1}; pH 7.2).

Table 1. Strains and plasmids used in this study

Strains	Relevant genotype and features	Reference
Escherichia coli		
JM109	*endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB+ Δ(lac-proAB) e14- [F'traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)*	NEB catalog
Agrobacterium fabrum		
C58	Wild-type	CFBP 1903
C58 Kan	Wild-type, Kan^R	
ΔSpG8-1b	C58 deleted of *atu1409* - *atu1423* cluster, Kan^R	Lassalle *et al.* 2011
ΔSpG8-2a	C58 deleted of *atu3054* - *atu3059* cluster, Kan^R	Lassalle *et al.* 2011
ΔSpG8-2b	C58 deleted of *atu3069* – *atu3073* cluster, Kan^R	Present work
ΔSpG8-3	C58 deleted of *atu3663* – *atu3693* cluster, Kan^R	Baude *et al.* 2016
ΔSpG8-4	C58 deleted of *atu3808* – *atu3830* cluster, Kan^R	Present work
ΔSpG8-5	C58 deleted of *atu3947* – *atu3952* cluster, Kan^R	Present work
ΔSpG8-7a	C58 deleted of *atu4285* – *atu4294* cluster, Kan^R	Present work
ΔSpG8-7b	C58 deleted of *atu4295* – *atu4307* cluster, Kan^R	Present work
Plasmids		
pGEM-T	Cloning vector for PCR fragments; Amp^R; lacZ	Promega
pGEM-Teasy	Cloning vector for PCR fragments; Amp^R; lacZ	Promega
pOT1e	Promoter-probe vector based on pBBR1MCS-5 replicon; contains promoterless eGFP and MCS between two transcriptional terminators; Gm^R	Allaway *et al.* 2001
pJQ200SK	Suicide vector; P15A sacB; Gm^R	Quandt *et al.* 1993
pGEM-T P*atu1416*	Upstream region of *atu1416* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3057*	Upstream region of *atu3057* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3069*	Upstream region of *atu3069* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3675*	Upstream region of *atu3675* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3817*	Upstream region of *atu3817* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu3948*	Upstream region of *atu3948* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu4292*	Upstream region of *atu4292* gene cloned into pGEM-T; Amp^R	Present work
pGEM-T P*atu4299*	Upstream region of *atu4299* gene cloned into pGEM-T; Amp^R	Present work
pOT1E P*atu1416*	Upstream region of *atu1416* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3057*	Upstream region of *atu3057* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3069*	Upstream region of *atu3069* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3675*	Upstream region of *atu3675* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3817*	Upstream region of *atu3817* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu3948*	Upstream region of *atu3948* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu4292*	Upstream region of *atu4292* gene cloned into pOT1E; Gm^R	Present work
pOT1E P*atu4298*	Upstream region of *atu4298* gene cloned into pOT1E; Gm^R	Present work

Deletion mutant constructions

Five deletion mutants (corresponding to the SpG8-2b, SpG8-4, SpG8-5, SpG8-7a and SpG8-7b regions) were constructed for this study as described previously (Lassalle et al. 2011) using the primers listed in **Table S1**.

Plant material and plant bacteria co-culture procedures

All plant experiments (confocal observation and metabolomic study) were conducted on *Solanum lycopersicum*, Marmande variety. *A. fabrum* strains were inoculated at a concentration of $1x10^6$ bacteria.mL^{-1} on one-month-old tomato plants after an incision with a sterile scalpel blade in the lower part of the stem (the collar zone). For confocal observation, bacterial cells harboring transcriptional fusions of the *A. fabrum*-specific genes were used. To perform the metabolomic study, deletion mutants of the *A. fabrum*-specific regions were used. Two so-called "non-inoculated" conditions were used, the NI condition (neither injured nor inoculated plants) and the NaCl condition (plants inoculated with saline water). Tomato plants were incubated for 21 days in a greenhouse with a temperature between 24°C and 28°C (**Fig. 1**).

Confocal microscopy observation

Microscope observation on tomato tumors of *A. fabrum* harboring the transcriptional fusions of the *A. fabrum*-specific genes was performed using a confocal laser-scanning microscope (LSM 800 Meta Confocal Microscope, Zeiss, Oberkochen, Germany). eGFP is expressed under the control of the promoter of each targeted gene of each *A. fabrum*-specific region and the green color indicates when the induction takes place. At 21 dpi, tumor tissue was mounted between a slide and a coverslip in a commercial mounting fluid (Aqua Poly/Mount, Polysciences, Inc., Warrington, PA). eGFP was excited with argon laser at 488 nm, and fluorescence was captured at 528nm. Analyses of images (five plants per condition) were performed with the LSM 800 software (Zeiss, Oberkochen, Germany).

Tumors harvesting and extraction of phenolic compounds

After 21 days of incubation, tumors were harvested, pooled, frozen in liquid nitrogen and freeze-dried 24 hours (**Fig. 1**). A small portion of each tumor was taken to carry out to calculate bacterial concentration. Tumors were next grinded using a bead mill TissueLyser II system (Qiagen, Hilden, Germany). Five replicates were performed per condition, each replicate consisting of pools of 4 tumors. Extractions targeting phenolic compounds were performed with samples of 100 mg of dried powder with MeOH 80% twice and then MeOH 100% twice with homogenization and sonication at room temperature (**Fig. 1**). The combined extracts were then centrifuged (14 500 rpm, 10 min) and the supernatants, particle free, were evaporated to dryness at 30°C in a Centrivap concentrator (Labconco, USA) to constitute the dried crude extracts. These extracts were weighed, dissolved in MeOH 60% at 10 mg/ml and conserved at -80°C prior to analysis. A quality control (QC) sample was prepared containing 20µL of each of the extracts obtained.

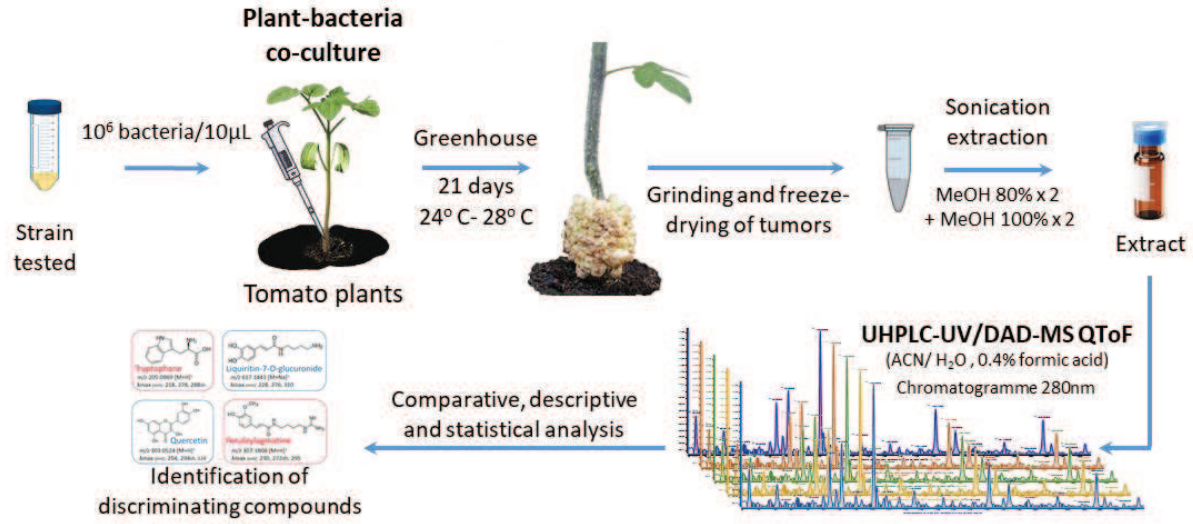


Fig. 1. Extraction and analysis of phenolic compounds from tomato tumors. Plants were inoculated singly with each of the deletion mutants of the *A. fabrum*-specific regions on one-month-old tomato plants after an incision in the collar zone. After 21 days of incubation, tumors were harvested and extracted targeting phenolic compounds. A metabolomic analysis was then performed using UHPLC-UV/DAD-ESI MS QTOF before the data analysis.

Metabolites analysis by UHPLC-UV/DAD-ESI-MS QToF

Metabolites analysis was performed on an UHPLC Agilent 1290 coupled to a UV-vis Diode Array Detector (Agilent 1290 Infinity series) and an Accurate-Mass Q-ToF 6530 spectrometer (Agilent Technologies). Liquid chromatography was carried out using a Nucleodur Sphinx RP C_{18} column (2mm × 100mm, 1.8µm, Macherey-Nagel) maintained at 50°C with an injection volume of 3 µL of sample.

The mobile phase was a mixture of acetonitrile and acidified ultrapure water (0.4% acetic acid) applied at a flow-rate of 0.5mL/min, and with a gradient (acidified H_2O:CH_3CN, [v/v]) starting at 98:2 for 1.5min, increasing to 76:24 at 26min, going to 0:100 in 6min and maintained for 3min, before returning to the starting conditions in 1min and equilibrating for 2min. The QToF-MS instrument was operated for MS analyses under the following conditions : the ion source ESI (Agilent Jet Stream thermal gradient focusing Technology) in positive ionization mode, with a 320°C gas temperature, a 11-liter/min gas flow, a nebulizer pressure at 40 psi, a 360°C sheath gas temperature with 12-liter/min flow rate, and with the capillary, nozzle, and fragmentor voltages at 3,000 V, 500 V, and 150 V, respectively. The acquisition mass range was of m/z 100 to 2000. The QC sample was initially analyzed and then inhected after every 8-9 samples in the run sequence to monitor the repeatability of the analysis. Chromatograms were explored at 280 nm with Mass Hunter qualitative analysis B.07.00 software. For characterization of discriminating compounds, complementary MS and tandem MS experiment were performed in positive and negative ionization modes and with different collision energies (10, 20, 30 or 40V) in AutoMS/MS mode. The UHPLC-UV/DAD-MS QToF device was managed by the Mass Hunter

Workstation Acquisition B.07.00 software and the resulting data was reprocessed with the Mass Hunter Qualitative Analysis B.07.00 software (Agilent Technologies).

Metabolite profiling data treatment and statistical analyses

First, integration of each peak of the UV-chromatograms at 280nm was performed and a pick alignment was made between all samples based on peak retention time, resulting in a metabolites matrix with samples in rows and absolute areas (A_{abs}) of detected metabolites in columns. Then for each metabolite within a given sample, its relative peak area (i.e. relative abundance in the sample= A_{rel}) was calculated over the total area of the sample chromatogram (ΣA_{abs}) as follows: $\%A_{rel} = A_{abs} / \Sigma A_{abs} \times 100$. Only the most abundant metabolites (relative abundance >0.5%) were selected to continue the study. Partial least squares discriminating analysis (PLS-DA) was made from this matrix to visualize the global effect of all our conditions tested (WT strain, mutant strains and the "non-inoculated" conditions) on tumor phenolic profiles of tomato. Two predominant saturated chromatographic peaks were removed from the analysis (peak **17** and **62**) and analyzed separately. Principal Component Analyzes (PCAs) were made between the WT strain and each mutant strain to stand out the effect of each *A. fabrum*-specific regions on tumor secondary metabolites profiles. The statistical study was oriented to a t-test to compare each condition with the WT strain condition in order to highlight discriminating metabolites. Statistical analyses were performed using R software (R Studio version 0.99.491).

Identification of discriminating root plant secondary metabolites

Annotations of discriminating metabolites were performed by the analysis of the UV-vis spectrum and the positive and negative MS and MSMS spectra. These data were compared to literature data, first on compounds already described in tomato but also on chemical data of metabolites from other plants.

RESULTS

Pathogenicity verification of *A. fabrum* mutant strains

All *A. fabrum* mutant strains were tested for their pathogenicity, i.e. their ability to form tumors on tomato plants. For this purpose, the dry weigh of tumors was compared between the WT condition and each mutant condition singly. A decrease on tumor weight was observed in the ΔSpG8-1a, ΔSpG8-1b, ΔSpG8-2a, ΔSpG8-2b, ΔSpG8-3 and ΔSpG8-4 mutant strains (**Fig. 2**).

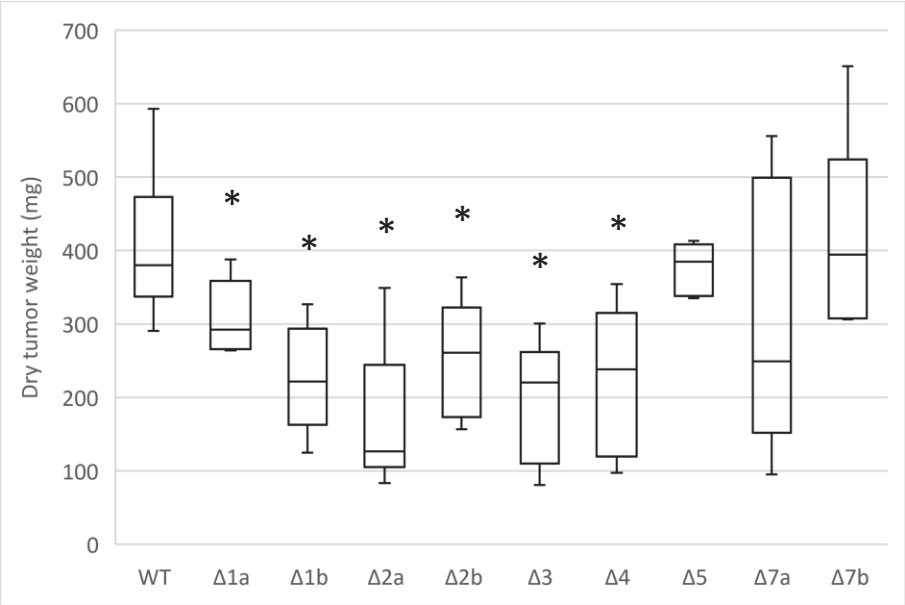


Fig. 2. Pathogenicity of deletion mutants of A. fabrum-specific regions. Dry weight of tomato tumors induced with *A. fabrum* wild-type C58 and mutant strains. Data were analyzed using a t test. Statistical differences are indicated with the symbol * (P<0.05).

All *A. fabrum* specific genes are expressed on tomato tumors

In order to verify whether *A. fabrum*-specific regions are expressed on tomato tumors, their expression was studied using one-month-old tomato plants inoculated with the WT bacterial strain C58 harboring singly transcriptional fusions of one gene per region selected on the basis of their likely involvement in region regulations. Microscopic observations of five plants of each reporter bacteria at 21 dpi revealed that all the constructions were expressed in contact with the tumor tissue, as shown by the green fluorescent cells (**Fig. 3**). Thus, *A. fabrum* is able to express the specific genes targeted by fusions during tumor colonization and it is very likely that all specific genes behave accordingly. Based on these results we are able to assume that changes in plant metabolite profiles between our different conditions and plants inoculated with the wild-type strains are related to the presence of the specific genes in the wild-type strain.

Bacterial survival in tumors is similar in all the conditions tested

Bacterial concentration per milligram of dry tumor was calculated in each condition. No difference in bacterial concentrations was found between the WT condition and each of the inoculated conditions with the deletion mutants of the *A. fabrum*-specific regions (**Fig. 4**).

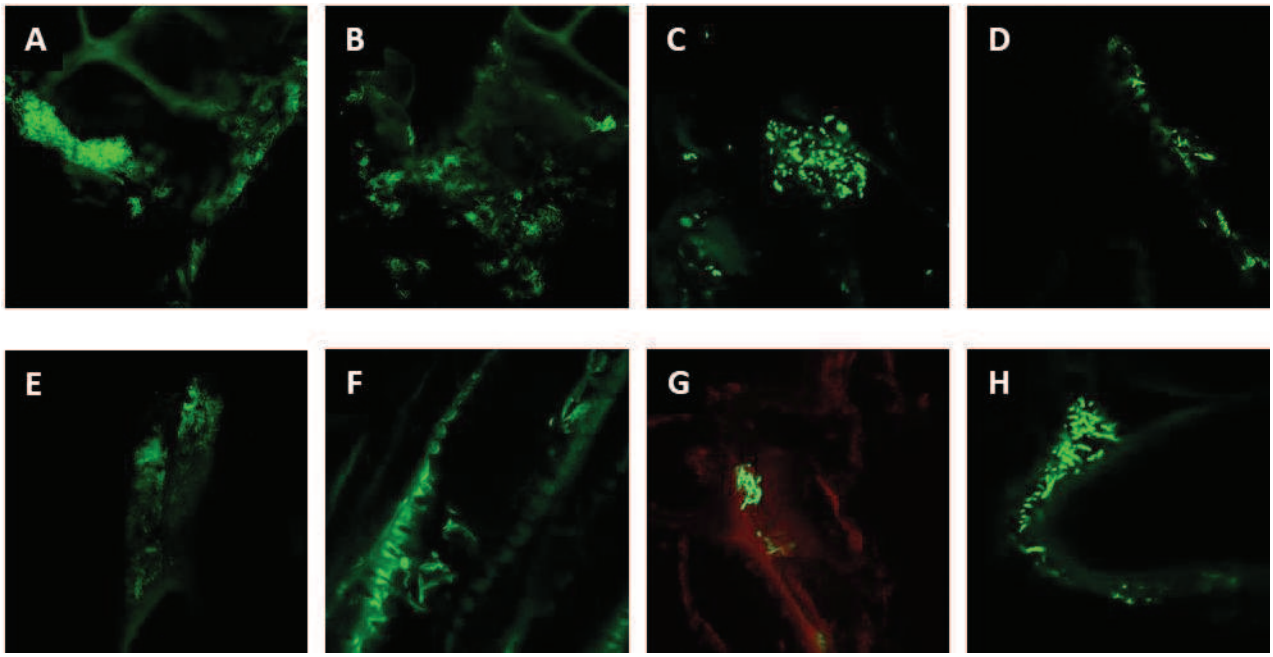


Fig. 3. *In planta* expression of the *A. fabrum*-specific regions through the expression of their targeted gene on tumors. A. pOT1e *Patu1406* belonging to SpG8-1a. B. pOT1e *Patu3057* belonging to SpG8-2a. C. pOT1e *Patu3073* belonging to SpG8-2b. D. pOT1e *Patu3675* belonging to SpG8-3. E. pOT1e *Patu3817* belonging to SpG8-4. F. pOT1e *Patu3948* belonging to SpG8-5. G. pOT1e *Patu4292* belonging to SpG8-7a. H. pOT1e *Patu4299* belonging to SpG8-7b. Gene expression was monitored using transcriptional fusions by confocal microscopy at 21 dpi. Representative pictures from five plants per specific region are shown. Green fluorescence shows bacteria able to express the corresponding transcriptional fusion. Plant auto fluorescence represented in green or red allows distinguishing plant cells. All transcriptional fusions are induced in contact with the tumor tissue.

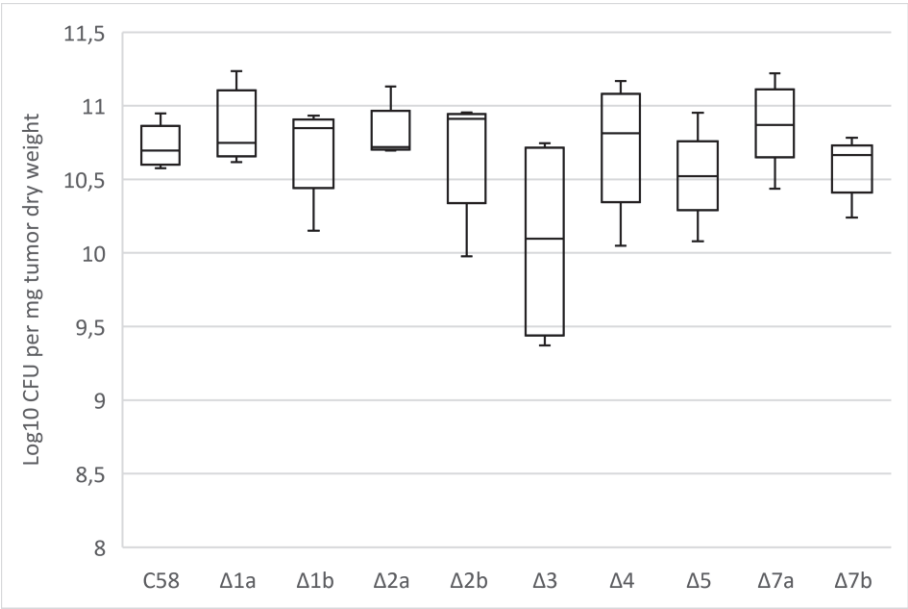


Fig. 4. Bacterial survival in tumor. Log10 UFC of *A. fabrum* wild-type and mutant deletion strains per tumor dry weight. Data were analyzed using a t test ($p>0.05$).

Global comparison of the tumor phenolic compounds profiles

The quality control (QC) was analyzed by UHPLC-DAD-ESI-MS QToF and led to the detection of a total of 90 peaks. Only secondary metabolites with mean relative abundances higher than 0.5% in the total ion current (TIC) chromatogram were considered thereafter in this study (*i.e.* 73 peaks). Changes detected in tumor metabolites between the wild-type condition and all mutant conditions singly were in terms of relative intensity rather than in appearance or disappearance of peaks. Chromatographic profiles were further compared using Partial Least Squares Discriminating Analysis (PLS-DA) (**Fig. 5**) and Principal Component Analysis (PCA) (**Fig. 5**). Two predominant saturated chromatographic peaks were removed from the analysis (peak **17** caffeoyl putrescine and **62** tris-(dihydrocaffeoyl)spermine) and analyzed separately.

The wild-type condition (plants inoculated with the *A. fabrum* C58 wild-type strain) was performed twice to ensure repeatability of our metabolomic study. A first metabolic profile comparison and a statistical analysis was made with both repeats inoculated in tomato plants. The PCA of metabolic profiles of the two wild-type conditions show no discrimination between the two profiles with PC1 and PC3 scores of 20.416 and 10.228, respectively **(Fig. 7A)**. However, the statistical analysis highlighted 10 metabolites significantly different in its relative abundance. So, we decided to exclude these metabolites from the analysis and to group the two wild-type conditions in one, that will be called from now as the “WT condition”.

The PLS-DA performed on metabolic profiles data showed a clear discrimination on one side all the inoculated plants (deletion mutants of *A. fabrum*-specific regions conditions and the WT condition) and on the other side the two so-called “non-inoculated” conditions (NI and NaCl conditions) separated from each other (**Fig. 5A**). In order to better visualize all the inoculated plant profiles (mutants and WT), we removed the “non-inoculated” conditions (NI and NaCl conditions) (**Fig. 5B**). Thereby, plant profiles can be clearly distinguishable by forming three distinct groups. The first one is composed of SpG8-2a and SpG8-2b conditions, the second group is composed of SpG8-7a and SpG8-7b conditions. Profiles of these two groups are closer to the WT profile than the profiles of the conditions that compose the third condition which are SpG8-1a, SpG8-1b, SpG8-3, SpG8-4 and SpG8-5 conditions (**Fig. 5B**).

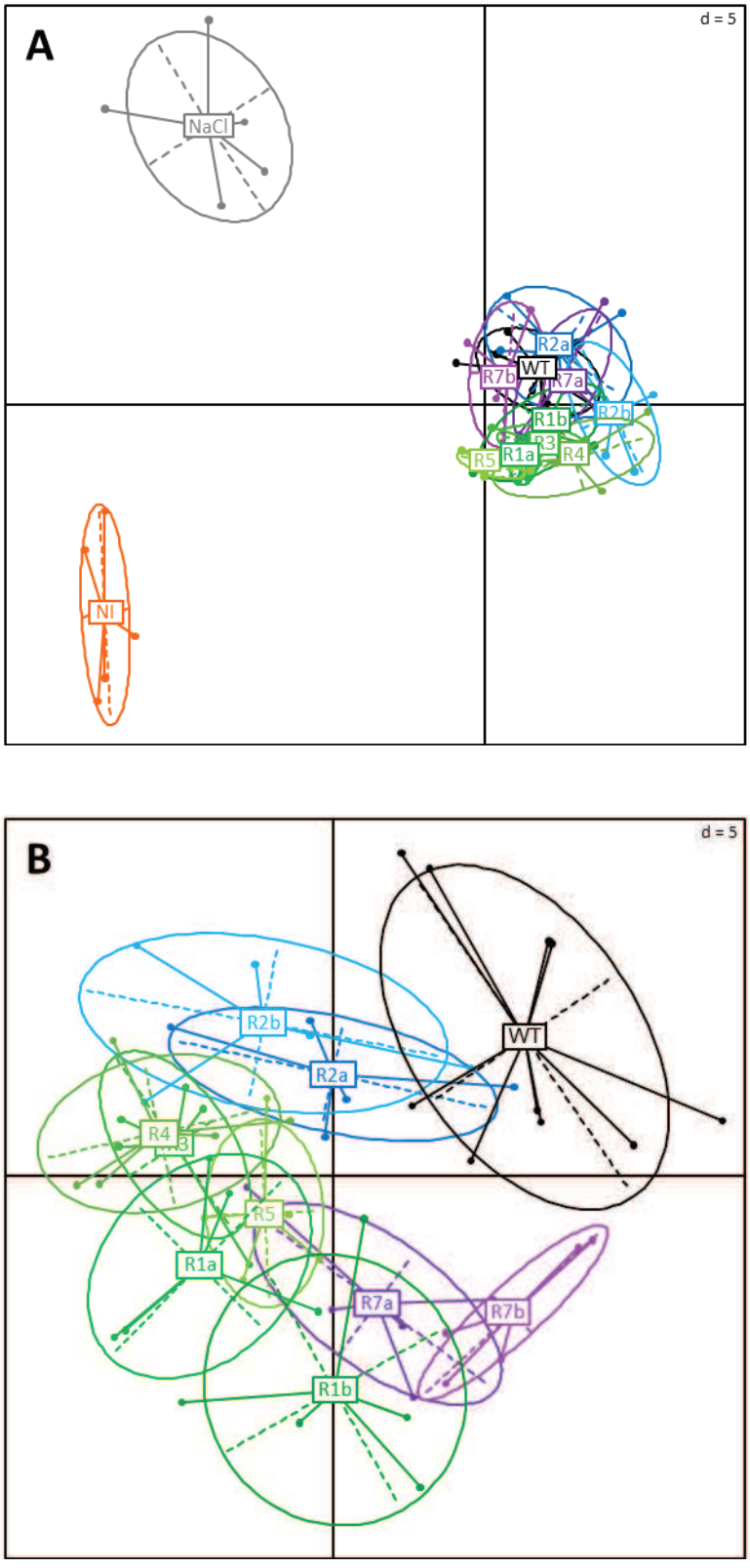


Fig. 5. Comparison of roots secondary metabolites profiles between all our tested conditions (WT, mutants and NI). PLS-DA were performed on chromatographic data at 280 nm obtained for each methanolic extract of barrel medic roots based on peak areas and retention time (data matrix of 92 peaks). Plants were inoculated or not with *A. fabrum* C58 wild-type or deletion mutant strains of *A. fabrum*-specific regions. WT: plants inoculated with the wild-type strain, NI: non-inoculated condition, RX: plants inoculated singly with each of the deletion mutant of *A. fabrum*-specific regions.

Evidence of bacterial effects on tumor secondary metabolites content

A total of 53 and 56 metabolites significantly varied in their relative abundances in the NI and the NaCl condition respectively when compared to the WT condition. 16 metabolites are over-abundant and 37 are underabundant in the NI condition, while 23 metabolites are over-abundant and 33 are underabundant in the NaCl condition (**Fig. 6**). Among them, there are 7 compounds that, while they are underabundant for one of these conditions, they are over-abundant for the other condition (**Fig. 6**).

Effect of specific genes on root plant secondary metabolites content

A metabolic profile comparison was done between the WT condition and each of the deletion mutants of *A. fabrum*-specific regions condition. Each of the PCA of metabolic profiles are shown in **Fig. 7B**-**J**. All mutant conditions have their metabolic profile separated from that of the WT condition. Taken all mutant conditions together, a total of 60 metabolites were found to display a significantly different relative abundance compared to the WT condition (**Fig. 8**). These discriminating metabolites formed three groups, one containing exclusively over-abundant compounds (23 compounds) in the mutant conditions when compared to the WT condition, a second group containing exclusively underabundant compounds (29 compounds) and a third group containing both over- and under abundant compounds (8 compounds) (**Fig. 8**). Concerning deletion mutant strain groups, the first one (composed of SpG8-2a and SpG8-2b conditions) has the least number of discriminating metabolites among the three groups of mutants with 16 and 21 respectively. Then, the second group (comprising SpG8-7a and SpG8-7b conditions) has 19 and 22 discriminating metabolites respectively (**Fig. 8**). Finally, the third group of deletion mutant strains composed of SpG8-1a, SpG8-1b, SpG8-3, SpG8-4 and SpG8-5 conditions, has the greatest number of discriminating metabolites with 28 discriminating metabolites each except SpG8-5 which has 29 (**Fig. 8**).

Identification or annotation of discriminating metabolites

The UHPLC-UV/DAD-MS/MS QTOF data were explored in order to characterize the discriminating compounds highlighted by statistical analyses. Study of the spectral data (UV-vis maxima; accurate mass; MS and MS/MS in positive and negative ionization mode) allowed the annotation of 23 compounds of different classes of metabolites by comparison to bibliographical data. Chemical data of the annotated discriminating compounds are shown in **Table 2**. Three out of the 23 compounds annotated are amino acids (**4** glutathione disulfide, **10** phenylalanine and **19** tryptophan), one is a nucleoside (**7**), one a flavonoid (**81** quercetin), one an hydroxycinnamic acid (**15** chlorogenic acid) and 16 are hydroxycinnamic acid amides (**12** caffeoylputrescine, **14** and **18** paucine 3'-

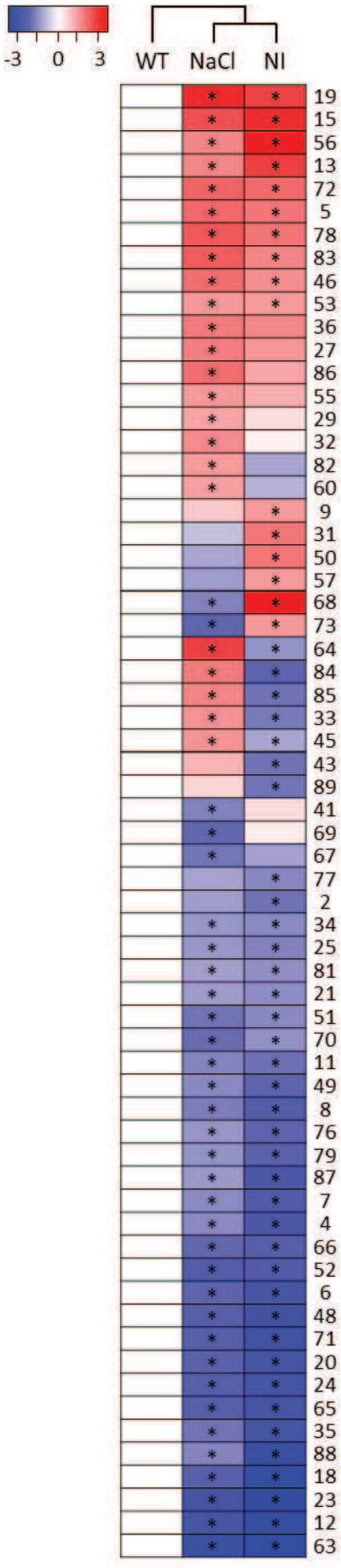


Fig. 6. Heat-map of discriminating metabolites between WT and the "non-inoculated" conditions (NaCl and NI) according to their abundance in the plant. Compounds are over-abundant (red) or underabundant (blue) in non-injured (NI) or inoculated with saline water (NaCl) plants compared to the wild-type condition (WT). WT: plants inoculated with the wild-type strain, NI: non-injured condition, NaCl: plants inoculated with saline water. Data were analyzed using analysis of variance. Statistical differences are indicated with the symbol * ($P < 0.05$).

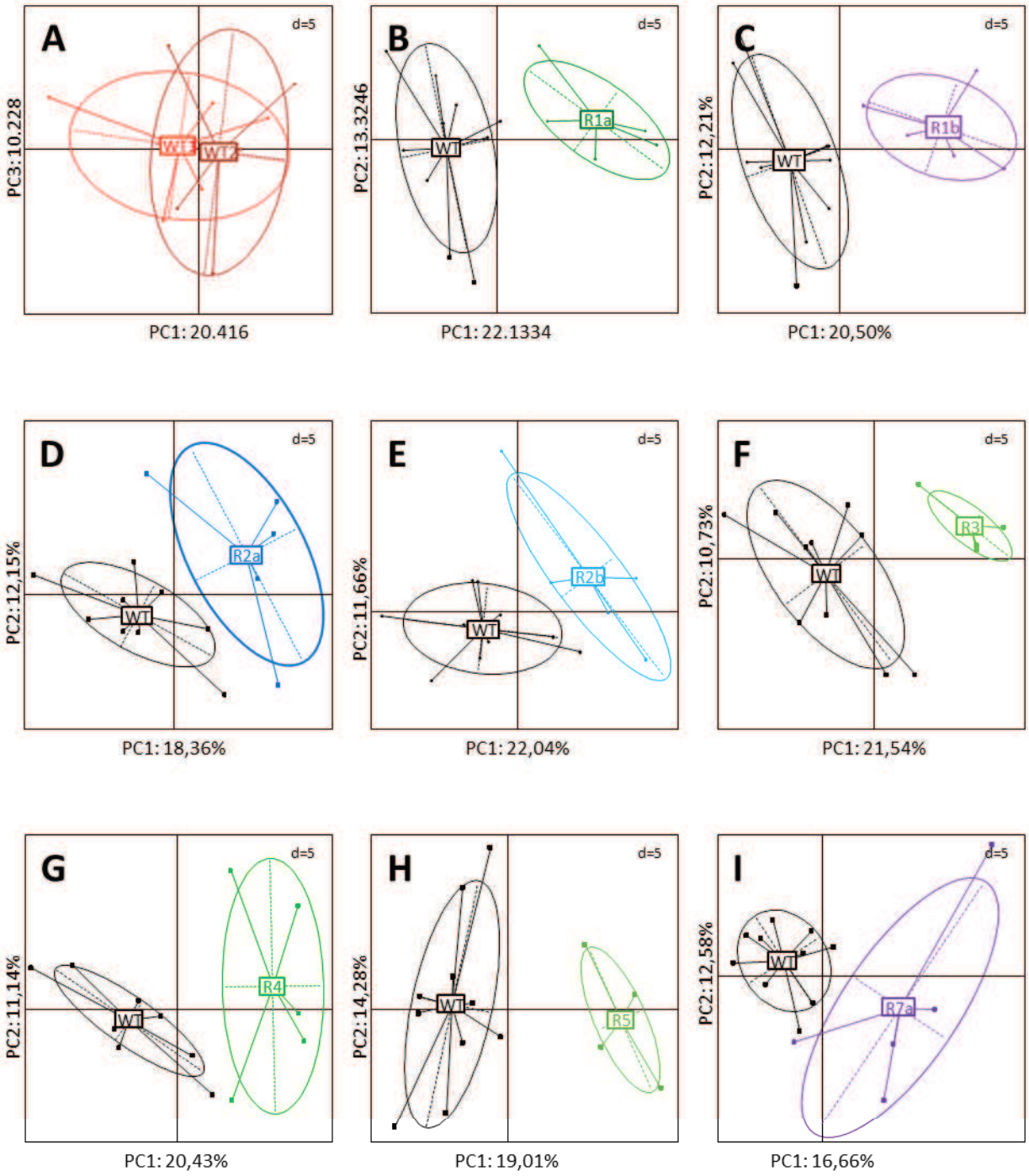


Fig. 7. Comparison of roots secondary metabolites profiles between all the conditions tested singly against the WT condition. ACPs were performed on chromatographic data at 280 nm obtained for each methanolic extract of barrel medic roots based on peak areas and retention time (data matrix of 92 peaks). Plants were inoculated or not with *A. fabrum* C58 wild-type or deletion mutant strains of *A. fabrum*-specific regions. WT1: plants inoculated with the wild-type strain, WT2: plants inoculated with the wild-type strain carrying the *ntpII* kanamycin resistance gene, NI: non-inoculated condition, RX: plants inoculated singly with each of the deletion mutant of *A. fabrum*-specific regions.

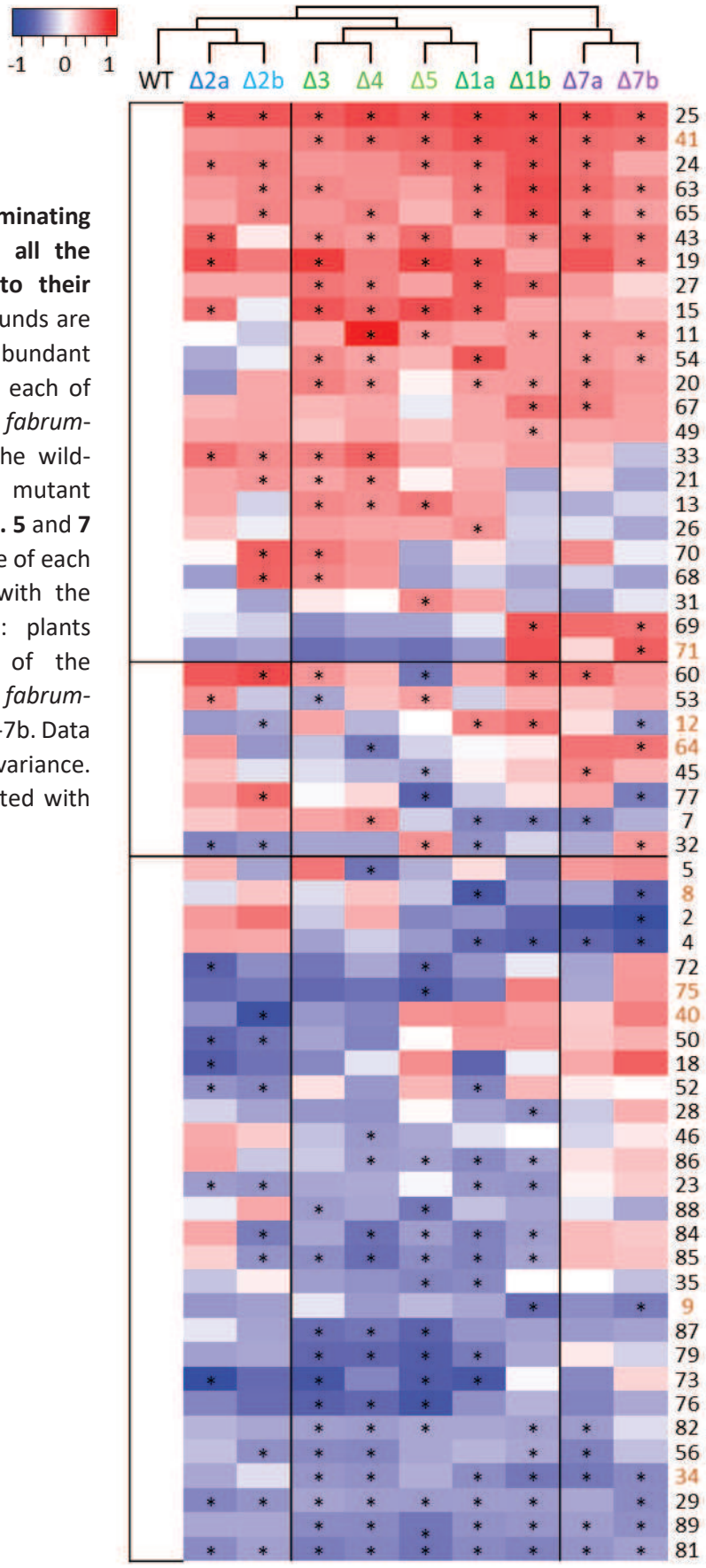


Fig. 8. Heat-map of discriminating metabolites between WT and all the mutant conditions according to their abundance in the plant. Compounds are over-abundant (red) or underabundant (blue) in inoculated plants with each of the deletion mutant strains of *A. fabrum*-specific regions compared to the wild-type condition. Colors of mutant conditions are the same as in **Fig. 5** and **7** in order to respect the color code of each group. WT: plants inoculated with the wild-type strain, Δ2a to Δ7b: plants inoculated singly with each of the deletion mutants of the *A. fabrum*-specific regions SpG8-2a to SpG8-7b. Data were analyzed using analysis of variance. Statistical differences are indicated with the symbol * ($P < 0.05$).

glucopyranoside, **23** coumaroylputrescine, **24** and **41** bis-(dihydrocaffeoyl)spermine, **31** feruloylputrescine, **32** and **43** feruloylagmatine, **44** and **48** hydrocaffeoyl-caffeoylspermidine, **51** feruloylquinic acid, **52** dimethyl-5-methoxytryptamine, **63** and **67** methyl 5-(6-caffeoyl-glucopyranosyl)-caffeoylquinic acid and **64**).

Two predominant peaks

Two predominant saturated chromatographic peaks (**17** and **62**) were removed from the initial analysis. The first one (peak **17**) annotated as caffeoyl putrescine was revealed by the statistical analysis to be over-abundant in all the mutant conditions tested and underabundant in the two "non-inoculated" conditions (NI and NaCl conditions). The second peak (**62**) annotated as tris-(dihydrocaffeoyl)spermine was also revealed by the statistical analysis to be underabundant in the two "non-inoculated" conditions (**Fig. 9**).

A total of 53 and 56 metabolites significantly varied in their relative abundances in the NI and the NaCl condition respectively when compared to the WT condition. 16 metabolites are over-abundant and 37 are underabundant in the NI condition, while 23 metabolites are over-abundant and 33 are underabundant in the NaCl condition (**Fig. 6**). Among them, there are 7 compounds that, while they are underabundant for one of these conditions, they are over-abundant for the other condition (**Fig. 6**).

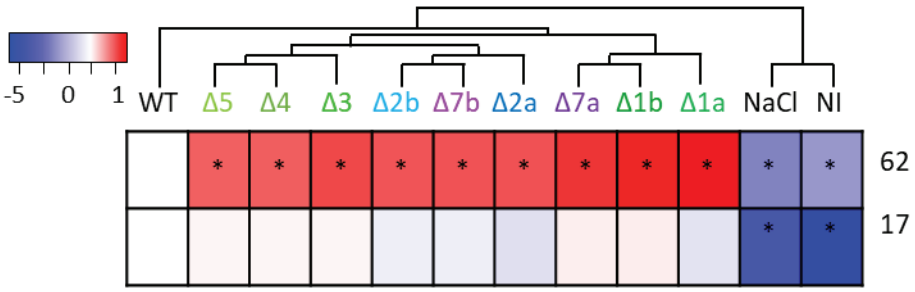


Fig. 9. Heatmap of peak 17 and 62 between WT and all the conditions tested. Compounds are over-abundant (red) or underabundant (blue) in non-inoculated (NaCl and NI) or inoculated plants with each of the deletion mutant strains of *A. fabrum*-specific regions compared to the wild-type condition. Colors of mutant conditions are the same as in **Fig. 5**, **7** and **8** in order to respect the color code of each group. WT: plants inoculated with the wild-type strain; Δ5 to Δ1a: plants inoculated singly with each of the deletion mutants of the *A. fabrum*-specific regions SpG8-5 to SpG8-1a; NaCl: plants inoculated with saline water; NI: non-injured condition. Data were analyzed using analysis of variance. Statistical differences are indicated with the symbol * ($P < 0.05$).

Table 2. Metabolites identified by UPHLC-DAD-ESI-MS Q-ToF in tomato tumors in interaction with *A. fabrum* strains

No.	RT (min)	UV λmax (nm)	UHPLC-MS Q-ToF Analysis					UHPLC-MSMS Q-ToF Analysis		Conditions										Proposed annotation	Chemical family	Reference
			Theoretical m/z	Ionization mode	Observed ions m/z	Molecular formula	Δppm	Collision energy (V)	Fragments MS/MS (% base peak)	1a	1b	2a	2b	3	4	5	7a	7b	NaCl	NB			
2	1.460	nd	nd		nd	nd			nd									*		*	nd	nd	
4	1.688	262	612	+	613.1635 [M+H]$^+$; 307.0867	$C_{20}H_{32}N_6O_{12}S_2$		10	613.1586; 484.1169; 355.0740; 231.0475	*	*						*	*	*	*	Glutathione disulfide	Amino acid	
				-	611.1427 [M-H]$^-$; 306.0572			20	611.1492; 338.0457; 306.0757; 272.0901													
5	1.753	220, 274sh		+	613.1687; 365.1095; 307.0887; 229.1578	nd			nd						*				*	*	nd	nd	
				-	633.1237; 611.1424; 306.0571; 227.0273																		
6	1.98	275, 283sh		+	nd	nd			nd										*	*	nd	nd	
				-	795.1732; 773.1955; 404.1009; 317.0742; 260.0442; 232.0066																		
7	2.468	258	267	+	290.0883 [M+Na]$^+$; 268.1064 [M+H]$^+$	$C_{10}H_{13}N_5O_4$		20	268.1012; 136.0619; 119.0351	*	*				*		*		*	*	Adenosine / Deoxyguanosine	Nucleoside	Lennon *et al*. 1998
				-	312.0936 [M+HCOO]$^-$; 266.0883 [M-H]$^-$			30	266.0879; 135.0384; 134.0464; 107.0355													
8	2.695	280		+	284.1002; 194.1184	nd			nd	*								*	*	*	nd	nd	
				-	282.0832; 191.9456																		
9	3.228	nd	nd		nd	nd			nd		*							*		*	nd	nd	
10	3.555	248	165	+	166.0884 [M+H]$^+$; 120.0824	$C_9H_{11}NO_2$		30	120.0810; 103.0546; 91.0551; 77.0394	-	-	-	-	-	-	-	-	-	-	Phenylalanine	Amino acid	Nozaki 1990 ; Shakya and Navarre 2006,
				-	164.0710 [M-H]$^-$			30	147.0444; 121.0248; 103.0518; 78.9591													
11	3.665	280	252	+	253.1582 [M+H]$^+$; 236.1315	nd		20	253.1543; 236.1290; 165.0547; 123.0443		*				*	*	*	*	*	*	nd	nd	
				-	251.1396 [M-H]$^-$			30	251.1382; 129.1024; 121.0294; 112.0757													
12	3.76	270, 292	250	+	251.1425 [M+H]$^+$; 273.1254 [M+Na]$^+$; 163.0411	$C_{13}H_{18}N_2O_3$		30	163.0392; 145.0281, 135.0442; 123.0452; 117.0338	*	*		*					*	*	*	Caffeoylputrescine	Hydroxycinnamic acid amide	Gális et al. 2004 ; Shakya and Navarre 2006; Singh et al. 2009
				-	249.1241 [M-H]$^-$			20	249.1241; 207.1126; 148.0510; 135.0450													
13	3.933	300sh, 322	384	+	407.1597 [M+Na]$^+$; 385.2012 [M+H]$^+$	nd		30	407.1520; 317.1510					*	*	*			*	*	nd	nd	
				-	429.1594 [M+HCOO]$^-$; 383.1544 [M-H]$^-$			30	383.1555; 301.0542; 251.1357; 168.0034													
14	4.213	296, 322	412	+	413.1985 [M+H]$^+$	$C_{19}H_{28}N_2O_8$		20	413.1907; 329.1433; 234.1143; 163.0395	-	-	-	-	-	-	-	-	-	-	-	Paucine hexoside	Hydroxycinnamic acid amide	Wang *et al*. 2010
				-	411.1746 [M-H]$^-$			40	303.2376; 208.1149; 157.0299; 135.0449													
15	4.242	300, 326	354	+	355.1058 [M+H]$^+$	$C_{16}H_{18}O_9$		20	355.0988; 324.0790; 195.0363; 185.0029; 163.0388	*		*		*	*	*			*	*	Chlorogenic acid (3-Caffeoylquinic acid)	Hydroxycinnamic acid	Cornard et al. 2008; Gális et al. 2004; Guo et al. 2008; Petrus 2004; Singh et al. 2009
				-	nd				nd														

17	3.76	270, 292	250	+ -	251.1425 [M+H]+; 273.1254 [M+Na]+; 163.0411 249.1241 [M-H]-	$C_{13}H_{18}N_2O_3$	30 20	163.0392; 145.0281; 135.0442; 123.0452; 117.0338 249.1241; 207.1126; 148.0510; 135.0450										*	*	Caffeoylputrescine	Hydroxycinnamic acid amide	Gális et al. 2004 ; Shakya and Navarre 2006; Singh et al. 2009
18	4.625	250, 294, 314	412	+ -	413.1953 [M+H]+; 435.1770 [M+Na]+; 251.1413 411.1761 [M-H]-; 249.1241	$C_{19}H_{28}N_2O_8$	50 20	180.1016; 163.0391; 150.0914; 145.0279; 135.0442; 117.0335 321.1467; 291.1319; 249.1229; 135.0447			*							*	*	Paucine hexoside	Hydroxycinnamic acid amide	Wang *et al.* 2010
19	4.773	276, 288sh	204	+ -	205.1003 [M+H]+; 188.0734 203.0820 [M-H]-;	$C_{11}H_{12}N_2O_2$	10 20	188.0678; 146.0596; 188.0681; 91.0571 184.8576; 142.0651; 116.0505; 74.0243	*		*		*		*		*	*	*	Tryptophan	Amino acid	Fletcher et al. 2013; Jiang et al. 2011
20	4.855	314	286	+ -	309.1971 [M+Na]+; 267.1740; 251.1430 285.0608 [M-H]-	nd	30 20	309.1862; 236.1276; 165.0544; 123.0446; 72.0814 285.0613; 152.0112; 108.0216	*	*			*	*		*		*	*	nd	nd	
21	4.967	255			nd	nd		nd				*	*	*				*	*	nd	nd	
23	5.108	228, 291, 308	234	+ -	235.1459 [M+H]+ 233.1284 [M-H]-	$C_{13}H_{18}N_2O_2$	10 30	235.1445; 218.1184; 147.0432; 119.0490 233.0670; 224.3603; 132.0575; 119.0482	*	*	*	*						*	*	Coumaroylputrescine	Hydroxycinnamic acid amide	
24	5.128	280, 305	530	+ -	531.3267 [M+H]+ 529.3017 [M-H]-	$C_{28}H_{42}N_4O_6$	40 40	531.3161; 293.1871; 222.1131; 165.0553; 123.0447 529.3020; 407.2653; 365.2551; 121.0294	*	*	*	*			*	*		*	*	Bis-(dihydrocaffeoyl) spermine (Kukoamine A)	Hydroxycinnamic acid amide	Yingyongnarongkul et al. 2008; Nzeuwa et al. 2017
25	5.217	nd	nd		nd	nd		nd	*	*	*	*	*	*	*	*	*	*	*	nd	nd	
26	5.303	nd	nd		nd	nd		nd	*											nd	nd	
27	5.39	288, 315	328	+ -	329.3295 [M+H]+; 351.3130 [M+Na]+ nd	$C_{18}H_{40}N_4O$	20	329.3292; 255.2439; 184.1699; 129.1388; 112.1124	*	*			*	*				*		nd	nd	
28	5.448	nd	nd		nd	nd		nd		*										nd	nd	
29	5.532	222, 268	265		nd	nd		nd	*	*	*	*	*	*	*		*	*		nd	nd	
31	5.687	236, 296, 324	264	+ -	287.1416; 265.1590 [M+H]+ 263.1397 [M-H]-	$C_{14}H_{20}N_2O_3$	30 10	177.0548; 149.0599, 145.0290; 134.0364, 117.0341; 106.0414 263.1432; 248.1161; 175.0735							*				*	Feruloylputrescine	Hydroxycinnamic acid amide	Li et al. 2018; Torras-Claveria et al. 2018
32	5.832	230, 272sh, 295	306	+ -	307.1808 [M+H]+ 305.1606 [M-H]-	$C_{15}H_{22}N_4O_3$	20 30	307.1763; 290.1505; 248.1307; 177.0549; 163.0372; 145.0285; 131.1290; 114.1028 289.1296; 272.1045; 248.1121; 204.0641; 148.0537; 134.0386	*		*	*			*		*	*		Feruloylagmatine	Hydroxycinnamic acid amide	Li et al. 2018; Piasecka et al. 2015 ; Nzeuwa et al. 2017
33	5.983	264 ; 290	nd		nd	nd		nd			*	*	*	*				*	*	nd	nd	

34	6.035	nd	nd		nd	nd		nd	*	*			*	*		*	*	*	*	nd	nd	
35	6.088	nd	nd		nd	nd		nd	*						*			*	*	nd	nd	
36	6.172	nd	nd		nd	nd		nd										*		nd	nd	
40	6.550	nd	nd		nd	nd		nd				*								nd	nd	
41	6.72	280, 324sh	473	+ -	474.2661 $[M+H]^+$ 472.2437 $[M-H]^-$	$C_{25}H_{35}N_3O_6$	40 40	236.1280; 222.1130; 165.0550; 136.0757; 123.0443; 100.0762 472.2439; 350.2078; 308.1976; 228.1713; 186.1609; 121.0292	*	*			*	*	*	*	*	*		Bis-(dihydrocaffeoyl) spermidine	Hydroxycinnamic acid amide	Yingyongnarongkul et al. 2008; Nzeuwa et al. 2017
43	6.997	220, 234, 293, 316	306	+ -	307.1786 $[M+H]^+$ 305.1613 $[M-H]^-$	$C_{15}H_{22}N_4O_3$	40 30	177.0548; 149.0597; 145.0287; 134.0363; 117.0337 289.1280; 272.1020; 247.1091; 204.0650; 148.0525; 134.0375		*	*		*	*	*	*	*		*	Feruloylagmatine	Hydroxycinnamic acid amide	Li et al. 2018; Piasecka et al. 2015 ; Nzeuwa et al. 2017
44	7.267	288, 320	471	+ -	472.2528 $[M+H]^+$ 470.2280 $[M-H]^-$	$C_{25}H_{33}N_3O_6$	30 30	472.2452; 310.2124; 293.1856; 236.1275; 220.0973; 177.0543; 163.0386; 145.0300; 123.0438 472.2423; 334.1747; 308.1951; 291.1007; 161.0239; 135.0444	-	-	-	-	-	-	-	-	-	-	-	Hydrocaffeoyl-caffeoylspermidine	Hydroxycinnamic acid amide	Cho et al. 2013; Nzeuwa et al. 2017; Sun et al. 2015; Zhang et al. 2015
45	7.303	282, 310	nd		nd	nd		nd							*	*		*	*	nd	nd	
46	7.382	228sh, 290, 328sh	nd		nd	nd		nd						*				*	*	nd	nd	
48	7.553	292, 324	471	+ -	472.2515 $[M+H]^+$ 470.2277 $[M-H]^-$	$C_{25}H_{33}N_3O_6$	30 30	472.2456; 455.2217; 310.2127; 293.1859; 222.1130; 163.0383; 145.0269; 123.0438 455.8883; 334.1756; 291.1635; 135.0458										*	*	Hydrocaffeoyl-caffeoylspermidine	Hydroxycinnamic acid amide	Cho et al. 2013; Nzeuwa et al. 2017; Sun et al. 2015; Zhang et al. 2015
49	7.663	292, 318	320	+ -	321.1963 $[M+H]^+$ 319.1765 $[M-H]^-$	$C_{16}H_{24}N_4O_3$	20 10	321.1929; 290.1495; 248.1318; 177.0551; 145.0281; 128.1181 319.1786; 304.1541; 263.1354; 149.0604		*								*	*	nd	nd	
50	7.725	286, 322sh	nd		nd	nd		nd			*	*							*	nd	nd	
51	7.907	234, 298sh, 323	368	+ -	369.1210 $[M+H]^+$ 367.1021 $[M-H]^-$	$C_{17}H_{20}O_9$	40 40	309.0917; 264.1187; 194.1108; 145.0271 206.9785; 191.0558; 173.0400; 134.0365										*	*	Feruloylquinic acid	Hydroxycinnamic acid amide	Caderby et al. 2013; Ferreres et al. 2009; Jaiswal and Kuhnert 2010; Kammerer et al. 2004; Lin and Harnly 2010; Matsui et al. 2007; Parejo et al. 2004; Piasecka et al. 2015; Shakya and Navarre 2006; Sun et al. 2015

52	8.092	276	218	+ -	219.1503 [M+H]$^+$ nd	$C_{13}H_{18}N_2O$	10	219.1488; 202.1223; 131.0492; 103.0542; 77.0380; 72.0817 nd	*		*	*						*	*	Dimethyl-methoxytryptamine	Hydroxycinnamic acid amide	Huhn et al. 2005; Sharma et al. 2007; Vermeulen et al. 2004
53	8.385	221, 274	nd		nd	nd		nd			*		*		*			*	*	nd	nd	
54	8.685	224, 284	nd		nd	nd		nd	*				*	*		*	*			nd	nd	
55	8.772	nd	nd		nd	nd		nd										*		nd	nd	
56	8.91	279, 328	501	+ -	502.2305 [M+H]$^+$ 500.2108 [M-H]$^-$	$C_{24}H_{31}N_5O_7$	20 20	502.1179; 325.1235; 181.0888; 163.0758 500.2112; 359.1299; 299.1137; 197.0812		*		*	*	*		*		*	*	nd	nd	
57	9.020	nd	nd		nd	nd		nd											*	nd	nd	
60	9.67	224, 296, 306	299	+ -	322.1069 [M+Na]$^+$ 298.1077 [M-H]$^-$	$C_{17}H_{17}NO_4$	20 20	322.1033; 281.0229; 257.9177; 129.1392 280.0990, 273.0362; 160.0418; 145.0292; 134.0607; 119.0498		*		*	*		*	*		*		Tetrahydroxyberbine		Grobe et al. 2010; Liscombe and Facchini 2007
62				+ -																Tris-(dihydrocaffeoyl) spermine		
63	10.645	288, 320	692	+ -	693.3595 [M+H]$^+$ 691.3337 [M-H]$^-$	$C_{32}H_{37}O_{17}$	40 50	693.3511; 531.3188; 457.2342; 293.1864; 220.0975; 163.0392 555.2822; 529.3018; 433.2436; 391.2357; 135.0451	*	*		*	*			*	*	*	*	Methyl-(caffeoyl-glucopyranosyl)caffeoylquinic acid	Hydroxycinnamic acid amide	Wu et al. 2013
64	10.808	222, 242sh, 294, 317	329	+ -	352.1213 [M+Na]$^+$; 681.2523 [2M+Na]$^+$ 328.1180 [M-H]$^-$	$C_{18}H_{19}NO_5$	30 20	201.0058; 123.0415; 112.1104; 84.0817 310.1082; 295.0834; 252.0759; 175.0372; 161.0231; 149.0601; 133.0521						*			*	*	*	Feruloyloctopamine / Feruloyldopamine	Hydroxycinnamic acid amide	
65	10.87	284, 322	692	+ -	693.3602 [M+H]$^+$ 691.3335 [M-H]$^-$	nd	50	293.1859; 222.1130; 163.0387 nd	*	*		*		*		*	*	*	*	nd	nd	
66	11.162	nd	nd		nd	nd		nd										*	*	nd	nd	
67	11.487	286, 318	692	+ -	693.3603 [M+H]$^+$ 691.3333 [M-H]$^-$	$C_{32}H_{37}O_{17}$	40 50	531.3153; 457.2313; 293.1854; 222.1132 555.2804; 529.3024; 512.1773; 391.2373; 348.2287, 135.0452		*						*		*		Methyl-(caffeoyl-glucopyranosyl)caffeoylquinic acid	Hydroxycinnamic acid amide	Wu et al. 2013
68	11.572	nd	nd		nd	nd		nd				*	*					*	*	nd	nd	
69	11.700	nd	nd		nd	nd		nd		*							*	*		nd	nd	
70	11.933	296, 318	505	+ -	528.1872 [M+Na]$^+$; 506.2063 [M+H]$^+$ 504.1860 [M-H]$^-$	nd		nd				*	*					*	*	nd	nd	
71	11.938	284, 336	nd		nd	nd		nd									*	*	*	nd	nd	

72	12.067	292, 330sh	nd		nd	nd		nd			*				*			*	*	nd	nd	
73	12.340	280, 320	nd		nd	nd		nd	*		*		*		*			*	*	nd	nd	
75	12.800	nd	nd		nd	nd		nd							*					nd	nd	
76	13.288	284	nd		nd	nd		nd					*	*	*			*	*	nd	nd	
77	15.100	nd	nd		nd	nd		nd				*			*		*		*	nd	nd	
78	15.968	286, 332	nd		nd	nd		nd										*	*	nd	nd	
79	16.388	219, 240sh, 292, 316	313	+ -	314.1437 [M+H]+; 336.1260 [M+Na]+ 312.1236 [M-H]-	$C_{18}H_{19}NO_4$	20 20	177.0547; 145.0285; 121.0649; 117.0337 312.1240; 297.0992; 190.0502; 178.0503; 148.0522; 135.0449	*				*	*	*			*	*	nd	nd	
81	17.362	222sh, 254, 294sh, 324, 370	302	+ -	303.0524 [M+H]+; 325.1211 [M+Na]+ 301.0351 [M-H]-	$C_{15}H_{10}O_7$	30 30	303.0496; 285.0418; 257.0454; 229.0486; 201.0527; 183.0451; 165.0213; 153.0175; 137.0215; 111.0087 178.9965; 151.0033; 121.0290; 107.0147	*	*	*	*	*	*	*	*	*	*	*	Quercetin	Flavonoid	Chae et al. 2002; Cornard et al. 2005; Du et al. 2006; Momić et al. 2007; Scigelova et al. 2011; Shakya and Navarre 2006
82	17.597	252, 266, 288, 346	286	+ -	287.0582 [M+H]+ 285.0398 [M-H]-	$C_{15}H_{10}O_6$	nd 30	nd 285.0404; 243.0298; 199.0419; 175.0419, 151.0048, 133.0291		*			*	*	*	*		*		nd	nd	
83	17.777	290, 334sh	nd		nd	nd		nd										*	*	nd	nd	
84	18.255	288sh, 300sh, 320	nd		nd	nd		nd	*	*		*		*	*			*	*	nd	nd	
85	18.387	284sh, 304sh, 320	nd		nd	nd		nd	*	*		*	*	*	*			*	*	nd	nd	
86	18.842	280, 320	nd		nd	nd		nd	*	*				*	*			*		nd	nd	
87	20.670	nd	nd		nd	nd		nd					*	*	*			*	*	nd	nd	
88	21.312	224, 242sh, 288sh, 328	nd		nd	nd		nd					*		*			*	*	nd	nd	
89	21.815	285, 318	nd		nd	nd		nd	*	*			*	*	*	*	*		*	nd	nd	

nd: non determined; sh: spectral shoulder

DISCUSSION

Tomato (*Solanum lycopersicum*) is one of the most important and widespread vegetables worldwide and has served also as a model vegetal organism (Lin et al., 2014). These crops, however, suffer attacks of various pathogens such as viruses, bacteria, fungi and nematodes, causing substantial production losses (Galeano). Evidence has emerged during the past decades demonstrating the importance of compounds from the phenylpropanoid pathway in plant defense response against pathogens (Dixon et al., 2002; Jahangir et al., 2009). Extensive research of secondary metabolites has been carried out in some plant-pathogen interactions such as *Arabidopsis/Pseudomonas syringae* (Hagemeier et al., 2001; Tan et al., 2004), tobacco/tobacco mosaic virus (Choi et al., 2006) or saskatoons/*Entomosporium mespili* (Wolski et al., 2010). However, less is known regarding changes of the metabolic profile associated with the response of tomato plants to pathogens (Bednarek et al., 2004; Zacarés et al., 2007).

In the present study, we were able to assessed the presence and relative abundance of metabolites within the infected plant tumor compared to noninfected stems. Several metabolites were highlighted to be accumulated under infection of *A. fabrum*, some others to be under-accumulated. The 90 secondary metabolites detected are present in all the mutant and in the wild-type conditions but not in the "non-inoculated" conditions. So, as expected, we can see that *A. fabrum* inoculation changes plant metabolome. Therefore, as well as in our root metabolome analysis with *A. fabrum*, we observed a modulation of compounds in the plant incited by this bacteria. Unlike that first analysis, the progress of this study is in the metabolite identification stage.

Based on the spectral data, we managed to annotate 23 compounds belonging to a different classes of metabolites **Table 2**. Most metabolites belong to the class of hydroxycinnamic acid amides, annotated as caffeoylputrescine (**12**), paucine 3'-glucopyranoside (**14** and **18**), coumaroylputrescine (**23**), bis-(dihydrocaffeoyl)spermine (**24** and **41**), feruloylputrescine (**31**), feruloylagmatine (**32** and **43**), hydrocaffeoyl-caffeoylspermidine (**44** and **48**), feruloylquinic acid (**51**), dimethyl-5-methoxytryptamine (**52**), methyl 5-(6-caffeoyl-glucopyranosyl)-caffeoylquinic acid (**63** and **67)** and feruloyloctopamine or feruloyldopamine (**64**).

Hydroxycinnamic acid amides (HCAAs) are a diverse and widely distributed group of plant secondary metabolites synthesized via the phenylpropanoid pathway (Bassard et al. 2010). Hydroxycinnamic acids (namely cinnamic acid, coumaric acid, caffeic acid, ferulic acid, and sinapic acid)

forming HCAAs bind to polyamines (El- Seedi et al. 2012). The latter, can be distinguished in two types based on their chemical and physical properties: basic and neutral. Basic amides contain primary amine function and are water-soluble aliphatic polyamines such as cadaverine, spermidine, putrescine, and spermine. On the other hand, neutral amides possess weak ionizable functions, are water-insoluble, and mostly contain aromatic amines such as tyramine, octopamine, and tryptamine (Facchini et al. 2002). Amino groups of aliphatic polyamines could be N-acylated with mono-, bis-, and tri- substituted HCAs (same or different HCA moieties) (Li 8). In addition, modifications of HCA, such as O-glycosylation, widely occur in the plant (Li 9). All of these lead to structural diversity of HCAAs. So far in our annotation analysis we found both types of amides present in tomato tumors as we detected putrescine, spermidine, spermine octopamine and tryptamine.

In contrast to the extensive literature on the chemistry and distribution of HCAAs in plants, relatively little is known about the biological function of these compounds. It is know that HCAAs play essential roles in plant growth and developmental processes throughout the plant kingdom including root growth, cell division, flowering, sexual differentiation, tuberization, leaf senescence and adaptation to stress due to antioxidant properties (Ten Chen and Huei Kao 1991; Hurng and Kao 1993; Martin-Tanguy 1997; Bouchereau et al. 1999; Alca´zar et al. 2006, 2010; Kuznetsov et al. 2006). But their main role seems to be plant defense responses to pathogen challenge and wounding and, in addition to lignin, HCAAs are recognized as important constituents of plant cell walls (Fry 1986; Iiyama et al. 1994, Facchini).

Accumulation of HCAA has been observed upon several abiotic stresses, such as mineral deficiencies, water excess, and heat shock (Edreva et al. 2007). The accumulation of HCAAs in response to pathogen challenge was first reported in potato tubers infected with *Phytophthora infestans* where amides of tyramine and octopamine were shown by microscopic examination to bind rapidly to cell walls following exposure of potato tubers to the oomycete (Clark 1982). Later, feruloyltyramine and feruloyloctapamine have also been shown to be covalently linked components of the cell wall in both natural and wound periderms of potato (Negrel et al. 1996). These compounds seemingly contribute to the formation of a poly-phenolic barrier which makes the cell walls more resistant to enzyme attack (King). The pathogen or stress-induced biosynthesis of HCAAs has since been reported in many plants, especially in members of the Solanaceae (Negrel et al. 1996; Miyagawa et al. 1998; Newman et al. 2001; Pearce et al. 1998). HCAA content of rice leaves revealed to increase upon infection by pathogenic microorganisms as several tyramine conjugates, serotonin conjugates, as well as N-feruloylagmatine and N-feruloylputrescine displayed induced accumulation in leaves infected with the fungus *Cochliobolus miyabeanus* and with the bacterium *Xanthomonas oryzae* (Morimoto et al. 2018).

Accumulation of HCAA, seems thus to comfort their role as chemical defense barriers and might be considered as a primary reaction of plant to bacterial perception. The accumulation of high abundances of HCAAs and flavonoid glycosides also reinforce cell walls to contain the pathogen to initial infection area (Karre).

HCAAs are also induced during the interaction of the tomato plant with pathogens. Upon infection by *Pseudomonas syringae*, it has been shown that four HCAAs (N-p-coumaroyltyramine, N-feruloyltyramine, N-pcoumaroyldopamine and N-feruloyldopamine) strongly accumulated in leaves 3 days post-inoculation (Zacarés et al. 2007). Also, in response to exposure to the pathogen *Pseudomonas syringae* pv. tomato, a rapid accumulation of HCAAs of noradrenaline (cis/trans N-p-coumaroylnoradrenaline and cis/trans N-feruloylnoradrenline) and octopamine (cis/trans N-p-coumaroyloctopamine and cis/trans N-feruloyloctopamine) was detected (López-Gresa). Wounding was also shown to increase the accumulation of feruloyltyramine and 4-coumaroyltyramine in tomato (Pearce et al. 1998).

The metabolic plant response upon *A. fabrum* inoculation

As expected, there is a strong modulation of secondary metabolites between the wild-type condition and the "non inoculated" conditions (NI and NaCl conditions) as the formation of the tumor induced by *A. fabrum* change plant metabolism (Mashiguchi 2019, Deeken et al., 2006; Lang et al., 2017, Gonzalez-Mula 2018). Indeed, statistical analysis showed that there are 64 discriminating compounds between the WT condition and the "non-inoculated" conditions (**Fig. 6**). Among them, 37 discriminating compounds are accumulated in tomato following the inoculation of *A. fabrum*, of these, only 28 are share by the two "non-inoculated" conditions. The annotation of 15 out of the 37 discriminating compounds was possible, which are glutathione disulfide (**4**), adenosine or deoxyguanosine (**7**), caffeoylputrescine (**12**), paucine 3'-glucopyranoside (**18**), coumaroylputrescine (**23**), di-(hydrocaffeoyl)spermine (**24**), bis(dihydrocaffeoyl) spermidine (**41**), feruloylagmatine (**43**), hydrocaffeoyl-caffeoyl spermidine (**48**), feruloylquinic acid (**51**), dimethyl-methoxy tryptamine (**52**), methyl-(caffeoyl-glucopyranosyl)-caffeoylquinic acid (**63**), feruloyloctopamine or feruloyldopamine (**64**), methyl-(caffeoyl-glucopyranosyl)-caffeoylquinic acid (**67**) and quercitin (**81**).

Interesting, our study showed that *A. fabrum* let to a reducer accumulation of some compounds. Indeed, 16 discriminating metabolites are under-accumulated in response to *A. fabrum* inoculation of which five could be annotated and are chlorogenic acid (**15**), tryptophan (**19**), feruloylputrescine (**31**), feruloylagmatine (**32**) and tetrahydroxyberbine (**60**) (**Fig. 6** and **Table 2**). An under-accumulation of secondary metabolites and particularly HCAAs has already been reported in rice in response to

pathogens (Chamam 2015, Morimoto 2018, Valette 2019). Indeed, following application of hormones involved in plant stress response (ethylene and 6-benzylaminopurine) decreased the content of feruloylagmatine and feruloylputrescine in rice roots (Morimoto 2018). Also, rice roots interaction with the pathogen *Burkholderia glumae* decreased the relative intensities of HCA derivatives, especially feruloylhexose, and feruloylquinic acid, p-coumaroylputrescine, feruloylputrescine, feruloylcadaverine and feruloylglycine (Chamam 2015 and Valette 2019). It seems thus that ferulic acid plays an important role as it confers rigidity to cell walls by cross-linking cell wall polysaccharides and lignin (Bennett and Wallsgrove 1994). So, variations in the content of ferulic acids derivates can suggest their use in the biosynthesis of lignin to strengthen cell wall and protect the plant against pathogen invasion as pathogenic bacteria increased the thickness of roots (Chamam 2015). These important structural modifications of roots correspond to the first line of defense against pathogens (Freeman and Beattie 2008).

The progress of our analysis allows us to see so far that wounding also increase the accumulation of seven compounds (**33**, **45**, **64**, **68**, **73**, **84** and **85**) (**Fig. 6**). One of them (**64**) was annotated as feruloyloctopamine or feruloyldopamine (**Table 2**). Besides, profiles comparisons (**Fig. 5**) showed that all the inoculated plants are grouped together away from the "non-inoculated" conditions but the fact that the NaCl condition and the NI condition are not overlapped indicates that wounding also provoked a reaction of the plant via the modification of its metabolism.

These results could mean that all these compounds may be chemical signatures of the crown-gall disease induced by *Agrobacterium fabrum*. To the best of our knowledge, this is the first study of secondary metabolites in tumor induced by *A. fabrum*. Some compounds reported in literature are detected more than once in our chromatograms, e.g. paucine hexoside, caffeoylputrescine, feruloylagmatine, hydrocaffeoyl-caffeoylspermidine, methyl-(caffeoyl-glucopyranosyl)caffeoylquinic acid (**Table 2**). Apparently, these metabolites can exist as different constitutional isomers in tomato fruit. The position and/or nature of the sugar substitution can influence the polarity and therefore the retention time of the compound (Moco). From the literature it is often unclear which particular isomer is mentioned.

Tumor metabolic modulation by A. fabrum-specific regions

Our approach allows to observe not only the effect of *A. fabrum* inoculation, but also the involvement of its specific regions in tumor metabolome. Results suggest that there is an impact of specific regions on tumor secondary metabolites of tomato plants. Indeed, the statistical analysis, taking into account all the mutant conditions, resulted in a total of 60 discriminating metabolites when

compared to the WT condition (**Fig. 8**). Therefore, we observed a modulation of compounds in the plant incited by these regions. The progress of this study is in the metabolite identification stage. The analysis of the compounds identified will be done afterwards. Still, our results showed a stronger modulation for some regions than others (**Fig. 8**).

Remarkably, there are 13 discriminating compounds modulated by *A. fabrum*-specific genes exclusively (**27**, **32**, **36**, **40**, **41**, **54**, **55**, **60**, **67**, **69**, **75**, **82** and **86**) when compared to the WT condition that were not discriminating in the bacterial inoculation comparison (**Fig. 8**). Four of these compounds could be annotated which are feruloylagmatine (**32**), bis-(dihydrocaffeoyl) spermidine (**41**), tetrahydroxyberbine (**60**), methyl-(caffeoyl-glucopyranosyl)caffeoylquinic acid (**67**).

CONCLUSION

Although the complete analysis of discriminating compounds is not finished, we managed to highlight the importance of secondary metabolites, especially HCAAs, in the bacteria-plant interaction, and more precisely, between *A. fabrum* and tomato. Furthermore, we still manage to show that there are even some discriminating compounds exclusively related to the *A. fabrum*-specific regions. These results will contribute to a better understanding of the ecological niche construction of A. fabrum by the formation of plant tumors, highlighting the importance of its specific genes in the establishment of this fine-tuned interaction.

ANNEXES

Table S1. Primers used in this study

A. fabrum-specific region	Analysis and region	Forward primer (5' to 3')	Reverse primer (5' to 3')	Reference
Transcriptional fusions				
SpG8-1b	*Atu1416* promoter			Meyer *et al.* 2018
SpG8-2a	*atu3057* promoter	AAGCTTTGACGGTTGTGAACAGCACT	GAGGTGAATACAGGCGGAAA	This study
SpG8-2b	*atu3073* promoter	AAGCTTACTGAAACCGACATGAACGC	CCGCCGAGAATTCCGATAGT	This study
SpG8-3	*atu3675* promoter	AAGCTTAAATCGCGTTCTCCAGATGG	GGGTCTCTACTGGCTCGATG	This study
SpG8-4	*atu3817* promoter	TGGTTCCCAGACGTCGTTTA	CAGTTTGATGTAGCGAGCCA	This study
SpG8-5	*Atu3948* promoter	AAGCTTGAAGGTCGGTGGCATTGT	ACGGCTCTTGCTTGTTCG	This study
SpG8-7a	*atu4292* promoter	ATCGATGTGCAGAGCTTGCTGACG	CAGGGTGATGTGGAGATCGT	This study
SpG8-7b	*atu4299* promoter	AAGCTTCGAGCCATTTCATGAGTGCT	AGGAGACAATGCAACCCGTA	This study
	Insertion verification in pGEMT	GTTTTCCCAGTCACGAC	CAGGAAACAGCTATGAC	Promega
	Insertion verification in pOT1e	CGGTTTACAAGCATAAAGC	CATTTTTTCTTCCTCCACTAG	Pothier *et al*. 2007
Construction of the *A. fabrum*-specific regions deletion mutants (inactivation gene clusters)				
ΔSpG8-1b	Upstream region of *atu1409*			Lassalle *et al.* 2011
	Downstream region of *atu1423*			
ΔSpG8-2a	Upstream region of *atu3054*			Lassalle *et al.* 2011
	Downstream region of *atu3059*			
ΔSpG8-2b	Upstream region of *atu3069*			This study
	Downstream region of *atu3073*			
ΔSpG8-3	Upstream region of *atu3663*	CCGGTTCTACATCCTGGAAA	CCTGCTCAACAGGCTACTCC	Baude *et al.* 2016
	Downstream region of *atu3693*	GACAACATGCCCCTCTCCTA	TCTGGAACGTCACCGACATA	
ΔSpG8-4	Upstream region of *atu3808*			This study
	Downstream region of *atu3830*			
ΔSpG8-5	Upstream region of *atu3947*			This study
	Downstream region of *atu3952*			
ΔSpG8-7a	Upstream region of *atu4285*			This study
	Downstream region of *atu4294*			
ΔSpG8-7b	Upstream region of *atu4295*			This study
	Downstream region of *atu4307*			
	nptII gene	TTGCTGCGCGCGACATCAAGGTTTCGACCGAGGAGT AGCCTGTTGAGCAGGTGTGTAGGCTGGAGCTGCTTC	TGAAAATGCCGCTGTATTTCCTCGATCACGTAGGA GAGGGGCATGTTGTCCATATGAATATCCTCCTTA	Baude *et al.* 2016
	Region for inactivation diagnosis	GAGAGTGACGCTTTGGCTCT	GGTTGATCTGGTGCAGCTTT	Baude *et al.* 2016

B. Opines: a new method of detection and quantification

Crown gall and hairy roots: development of a new diagnostic tool based on the UHPLC-ESI-MS-QTOF analysis of opines, natural substances produced by infected plants

ABSTRACT

Opines are low molecular weight metabolites specifically biosynthesized by plant cells transformed by agrobacteria in crown gall and hairy roots, plant diseases provoking tissue uncontrolled overgrowth. As transferred DNA is sustainably incorporated into the genomes of transformed plant cells, opines in plant tissue constitute a persistent biomarker of plant infection by pathogenic agrobacteria targetable for crown gall/hairy roots diagnostic. In the present study, a general, rapid, specific and sensitive analytical method was developed for overall opine detection based on ultra-high-performance liquid chromatography–electrospray ionization quadrupole time-of-flight–mass spectrometry (UHPLC-ESI-MS-QTOF), requiring no derivatization and with easy sample preparation. Based on MS, MS/MS and chromatographic data, detection selectivity of a wide range of standard opines was validated in pure solution and in different plant extracts. It was successfully used for the detection of different structural type of opines, including opines without standard compounds available, in real tumors or hairy roots induced by pathogenic strains. As this method can detect a wide range of opines in a single run, it represents a powerful tool to perform analyses of plant tumors for the diagnosis of crown gall and hairy roots. Using an appropriate dilution of plant extract and a matrix-based calibration curve, the quantification ability of the method was validated for three opines belonging to different families (nopaline, octopine, mannopine) allowing an accurate quantification in plant tissue extracts.

KEYWORDS

Opines, detection, quantification, UHPLC-ESI-MS-QTOF, diagnostic, agrobacterium

INTRODUCTION

Crown gall and hairy roots are plant neoplastic diseases inducing uncontrolled overgrowth of plant tissue (tumors) or anarchically proliferating roots, respectively. These diseases can affect nursery and mature plants, and are considered to be among the most important ones in economic terms for fruit trees and ornamental plants, including almond (*Prunus dulcis*), grapevine (*Vitis vinifera*), peach (*Prunus persica*), plum (*Prunus domestica*), rose (*Rosa* spp.), walnut (*Juglans* spp.) and weeping fig (*Ficus benjamina*) (Kennedy 1980; Puławska 2010; Tarkowski and Vereecke 2014). The economic losses are not only related to intrinsic damage to infected plants, but also to their prohibited commercial use (Tarkowski and Vereecke 2014; Epstein et al. 2008).

Both pathologies are reported to be caused by agrobacteria, common soil Alphaproteobacteria belonging to the *Rhizobiaceae* family. They usually live saprophytically in many plant rhizospheres, but occasionally adopt a phytopathogenic lifestyle and cause infections in wounded plants. This occasional virulence is determined by the presence of a large (at least 200 kb) DNA plasmid: the tumor-inducing (Ti) or the root-inducing (Ri) plasmids provoking crown gall and hairy root diseases, respectively (Gelvin 1990). These plasmids often occur in bacteria of the genus *Agrobacterium*, but they are also harbored by other genera of *Rhizobiaceae*, *e.g*. *Allorhizobium, Rhizobium, Pararhizobium* or *Neorhizobium* (Mousavi et al. 2015). These agrobacteria generally exhibit a broad host-plant range (Nester 2014), as they can induce tumor or hairy root formation on most dicotyledonous species, and on a few monocots and gymnosperms (De Cleene and De Ley 1976). Pathogenicity is due to a small portion (around 20 kb) present on both plasmids. It is a DNA section transferred (T-DNA) from bacteria to the nuclear DNA of the host plant cells where it is incorporated into the plant genome. T-DNA genes are thus expressed in these transformed plant cells (Chilton et al. 1980, 1982; Gordon and Christie 2014). Besides causing the disease, the T-DNA encodes information for enzymes that catalyze the biosynthesis of unusual low-molecular-weight metabolites by the transformed plant cells, referred to as opines (Dessaux et al. 1993).

More than 30 crown gall opines have been characterized structurally (Dessaux et al. 1998; Flores-Mireles et al. 2012; Dessaux and Faure 2018). Opines are a heterogeneous group of molecules divided into two structure classes: (i) sugar phosphodiesters, called agrocinopines, and (ii) the majority of opines, made of primary and secondary amine derivatives from the condensation of one amino acid with an alpha-ketoacid or a sugar (Dessaux et al. 1993; Brencic and Winans 2005). This latter chemical class is subdivided into 8 different opine families, depending on their composition and their biosynthesis pathways; the most frequently encountered are the nopaline, octopine, and mannityl-

opine families (Moore et al. 1997). For a review of opine families and structures see Dessaux *et al.* (Dessaux et al. 1998; Dessaux and Faure 2018). There are several types of Ti and Ri plasmids, and each plasmid allows for the biosynthesis and catabolism of a few of these compounds. Opines serve as specific carbon, nitrogen or sulfur sources for the Ti or Ri plasmid-harboring agrobacteria, whereas most other microorganisms are unable to use these compounds (Tempé and Petit 1982; Veluthambi et al. 1989). Consequently, the inciting bacterium diverts the plant cell metabolism to provide itself with an ecological niche with abundant nutrient supply (Dessaux and Faure 2018; Lang et al. 2014). Furthermore, some opines are inducers of conjugative transfer of these plasmids to neighboring non-pathogenic agrobacteria and thus promote the dissemination of pathogenicity. They play thereby a key role in the ecology of the plant-agrobacteria interaction (Klapwijk et al. 1978; Petit et al. 1978; Ellis et al. 1982; Teyssier-Cuvelle et al. 1999, 2004). As no curative methods are available for these plant diseases, an early diagnosis followed by sanitation procedures (*e.g*. uprooting) can prevent pathogenic agrobacteria from disseminating and settling in crops (Tarkowski and Vereecke 2014). Consequently, a rapid, sensitive and specific method to diagnose this bacterial infection in plant material is required.

Opines are not all biosynthesized concomitantly in abnormal tissue outgrowths (tumor or hairy roots) because their production depends on the type of plasmid harbored by the inciting agrobacterium. However, as T-DNAs are sustainably incorporated into the genomes of the transformed plant cells, the presence of opines in plant tissue constitutes a persistent biomarker of plant infection by a pathogenic agrobacterium targetable for crown gall/hairy roots diagnostic. For example, an opine-based biosensor has recently been proposed to specifically detect nopaline and octopine, and identify if tumors are induced by *Agrobacterium* or not (Choi et al. 2019). Nevertheless, according to the type of plasmid, each inciting agrobacterium transfers diverse types of opine biosynthesis genes, and generally more than one (Ellis and Murphy 1981). Opines produced in transformed plants also seem to be dependent on the precursors available in the plant (Flores-Mireles et al. 2012). Therefore, the development of a general method to detect a wide range of opines in plant extracts is needed to use opine-based diagnostic techniques for these plant diseases.

Different analytical techniques and methods have been developed and used to detect and analyze opines, such as enzymatic assay (Grieshaber 1976; Storey and Dando 1982), paper/thin layer chromatography (TLC) (Firmin and Fenwick 1977) or high-voltage paper electrophoresis (HVPE) followed by chemical staining (Ellis and Murphy 1981; Chilton et al. 2001; Zhang et al. 1998; Savka et al. 1990; Swain et al. 2010), high-performance liquid chromatography (HPLC) with diverse detectors (fluorescence, UV absorbance at 254 nm, conductivity detection, refractometry) associated or not to a previous step of derivatization specific to some types of opines (Sato et al. 1988; Zhang et al. 1998;

Sandee et al. 1996; Sauerwein and Wink 1993; Firmin 1990), gas chromatography coupled to mass spectrometry (GC-MS) with previous derivatization specific to non-guanidino opines (Scott et al. 1979), or liquid chromatography coupled to mass spectrometry (LC-MS) with previous derivatization by butylation (Venter et al. 2017). Each of these methods has unique features that make it possible to study a few types of opines only. However, some can present drawbacks: TLC and HPVE have limited sensitivity (high detection limits) and do not provide accurate quantification (Zhang et al. 1998), whereas GC-MS and HPLC coupled to various detectors are more sensitive and allow for a quantification procedure. Furthermore, they all present main disadvantages such as requiring a partial purification pretreatment of the extract before analysis, or the need to use reagents to reveal the compounds or to carry out additional steps of derivatization to detect opines. All these previous steps not only lead to a time-consuming process before analysis, but also fail to detect all opine types in a one-go analysis, as they require different chemical staining (cited in (Zhang et al. 1998)) or various derivatizations according to the targeted opine structures (Sato et al. 1988; Zhang et al. 1998; Sandee et al. 1996; Sauerwein and Wink 1993; Firmin 1990).

In the present study, an analytical method was developed for overall opine detection based on ultra-high-performance liquid chromatography–electrospray ionization quadrupole time-of-flight–mass spectrometry (UHPLC-ESI-MS-QTOF), requiring no derivatization and with easy sample preparation. We showed that this method is useful not only to detect but also to quantify different types of opines in plant extracts. The quantification method was validated for opines belonging to three different families, *i.e.* the nopaline-, octopine-, and mannityl-opine families.

MATERIALS AND METHODS

1. Chemicals and reagents

Nopaline and octopinic acid were purchased from Sigma-Aldrich Laboratory (Saint Louis, Missouri, USA), and octopine from ICN Biomedicals Inc. (Aurora, Ohio, USA). Mannopine, mannopinic acid, agropine, agropinic acid, cucumopine, histopine, alanopine, pyronopaline, agrocinopine A, and some opine analogs like allo-octopine, nor-mannopine, galactopine, galactopinic acid, glucopine and glucopinic acid, were kindly supplied by Y. Dessaux, D. Faure and S. Morera (Institute for Integrative Biology of the Cell, Gif-sur-Yvette, France). The chemical reagents and solvents such as acetonitrile, water and formic acid (UHPLC-MS grade) were obtained from Fisher Scientific (Fair Lawn, New Jersey, USA).

2. Bacterial strains and plant materials

The bacteria used in this study are listed in **Table S1**. Plant experiments were conducted on the following plant species grown in a greenhouse: *Solanum lycopersicum* (tomato), *Rosa* sp., *Kalanchoe daigremontiana,* and *in vitro* cultures of wild-type and transgenic *Lotus corniculatus* (bird's-foot trefoil) modified to produce nopaline and mannopine (Oger et al. 2004, 1997).

Table S1. Bacterial strains used in this study

Bacterial strains[a]		Relevant features[b]	Reference
Agrobacterium sp.	CFBP 2407	Wild-type strain harboring an octopine/cucumopine-type pTi.	(Ridé et al. 2000)
	CFBP 2788	Wild-type strain harboring a chrysopine/nopaline-type pTi.	(Chilton et al. 1995)
A. fabrum	C58	Wild-type strain harboring the pTiC58 (nopaline-type pTi) and the pAtC58.	CFBP 1903
	CFBP 1898	C58 derivative strain cured of pTiC58, and harboring an octopine-type pTi (pTiA66) and the pAtC58.	CFBP 1898
	UIA5	C58 derivative strain lacking both the pTiC58 and the pAtC58.	(Kim et al. 1996; Nair et al. 2003)
	AB150	UIA5 isogenic strain harboring the pAtC58.	(Nair et al. 2003)
	AB152	UIA5 isogenic strain harboring thepTiC58.	(Nair et al. 2003)
	AB153	AB150 isogenic strain harboring the pTiC58.	(Nair et al. 2003)
A. radiobacter	CFBP 296	Wild-type strain harboring a nopaline-type pTi.	(Pionnat et al. 1999)
A. sp. G1	TT111	Wild-type strain harboring an octopine-type pTi.	(Vigouroux et al. 2017)
	CFBP 2712	Wild-type strain harboring a mannityl-opine-type pTi.	(Ridé et al. 2000)
A. sp. G3	CFBP 4424	Wild-type strain harboring a succinamopine-type pTi.	(Pionnat et al. 1999)
Allorhizobium vitis	S4	Wild-type strain harboring a vitopine/heliopine-type pTi.	(Szegedi et al. 1988)
	CFBP 2736	Wild-type strain harboring an octopine/cucumopine-type pTi.	(Ridé et al. 2000)
Rhizobium rhizogenes	NIAES 1724	Wild-type strain harboring a mikimopine-type pRi.	(Isogai et al. 1990)
	CFBP 2692	Wild-type strain harboring a mannityl-opine-type pRi.	(Ridé et al. 2000)
	K599 (NCPPB 2659)	Wild-type harboring a cucumopine-type pRi.	(Failla et al. 1990)

[a] Taxonomy according to Mousavi *et al.* (Mousavi et al. 2015). *A. fabrum* and *A. radiobacter* correspond to *A.* sp. G8 and *A.* sp. G4, respectively, using the provisional nomenclature of Costechareyre *et al.* (Costechareyre et al. 2010), while *A.* sp. G1 and *A.* sp. G3 are two other *bona fide* species of the *Agrobacterium tumefaciens* species complex that have not yet received accepted Latin binominal names.

[b] "pTi" for tumor-inducing plasmid and "pRi" for root-inducing plasmid.

3. Inoculation, culture, harvesting and extraction

Prior to plant infection, bacterial strains were grown overnight under shaking (160 rpm) at 28°C in YPG-rich medium (yeast extract 5 g/L; peptone 10 g/L; glucose 10 g/L; pH 7.2). The bacteria were then washed and suspended in sterile water at a concentration of 10^8 *per* mL. Ten microliters of the suspensions (i.e. 10^6 bacteria) were inoculated on the stem collar zone of the plants after incising with a scalpel blade. The plants were then incubated 21 days in a greenhouse with 16 hours illumination, at 22°C (day) and 19°C (night). Each plant tumor was harvested and immediately frozen in liquid nitrogen (metabolism quenching) and stored at -80°C. Freeze-dried tumors were ground with a benchtop homogenizer (FastPrep-24™, MP Biomedicals™, Fisher Scientific, UK), and 30 mg of powdered samples were subjected to two successive extractions by adding 1 mL of a methanol/H_2O 20:80 (v/v) solvent mix and 15 min sonication at each extraction step. After centrifugation (10 min, 19 745 × g), the supernatant was evaporated to obtain the crude extract. For the *in vitro*-cultivated transgenic *Lotus corniculatus,* the extraction process was applied to entire plants. All these extracts were solubilized at 10 mg/mL in UHPLC-MS grade water and stored at -20°C until analyses. Plant blank matrices consisted of non-transformed, opine-less plant tissue extracts. They were obtained following the same extraction procedure, but using stem samples of non-inoculated plants (for tomato and kalanchoe) or entire plants for *in vitro*-cultivated wild-type *L. corniculatus.*

4. Preparation of standard solutions, calibration standards and quality control samples

Each opine standard stock solution was prepared at 1 mg/mL in UHPLC-MS grade water and stored at -20°C. Two mixed standard solutions were also prepared. They contained equal quantities of each compound, that were either all 16 tested opines (*Mix16*) or just three of them: nopaline, octopine and mannopine (*Mix3*). These mixed standard solutions were prepared twice, in UHPLC-MS grade water or in plant blank matrix solutions at extract concentrations of 10 mg/mL, 0.5 mg/mL and 0.1 mg/mL. For the linearity study, the *Mix3* standard solution was also diluted to get concentrations of each opine ranging from 10 to 15 000 ng/mL using either UHPLC-MS grade water or the appropriate concentration of plant blank matrix solution. All opines were also separately spiked (1 µg/mL) in the same plant blank matrix samples for quality control (QC).

5. UHPLC-ESI-MS-QTOF analysis: instruments and analytical conditions

An Agilent 1290 UHPLC system (Agilent Technologies, Santa Clara, CA, USA) coupled with an Accurate-Mass Q-TOF 6530 spectrometer (Agilent Technologies, Santa Clara, CA, USA) was used. Liquid chromatography was carried out using a Poroshell 120 EC-C_8 column (2.7 µm particle size, 100 mm x 3 mm, Agilent Technologies, Newport, CA, USA) at 30°C. The mobile phase was a mixture of acetonitrile and acidified water (0.4% formic acid) with a flow rate of 0.6 mL/min. The gradient started with 100%

acidified water for 3 min, and then linearly increased to 100% acetonitrile in 4 min. This condition was maintained for 1 min before cycling back to the initial conditions, and then kept for post-acquisition column equilibration (2 min). A volume of 10 µL of extracts or standard solutions was injected. The Quadrupole Time-of-Flight Mass Spectrometer (QTOF-MS) equipped with an electrospray ionization source (ESI Agilent Jet Stream thermal gradient focusing technology, Santa Clara, CA, USA) was used in positive ionization mode for MS and tandem MS analyses under the following conditions: a drying gas (N_2) flow of 11 L/min at 310°C, a nebulizer pressure of 40 psi, a sheath gas flow rate of 11 L/min at 350°C, with the capillary, nozzle, and fragmentor voltages set at 3000 V, 500 V, and 100 V, respectively. The acquisition mass range was *m/z* 80 to 2000 with a scan rate of 2 spectra/s, and the MS^2 experiment was done with the collision energy set at 10, 20, 30 or 40 V. A reference solution containing a standard compound (HP-0921) at *m/z* 922.00979 $[M+H]^+$ was constantly infused as an accurate mass reference. The UHPLC-ESI-MS-QTOF device was managed by Mass Hunter Workstation Acquisition B.07.00 software, and the data was reworked with Mass Hunter Qualitative Analysis B.07.00 software (Agilent Technologies).

6. Quantification method validation

The validation process of the quantification method was carried out as described in the literature (ICH. Harmonized Tripartite Guideline. Validation of Analytical Procedures: Text and Methodology Q2(R1). n.d.; Bioanalytical Method Validation. Guidance for Industry. US FDA, 2018. n.d.; Guideline on Bioanalytical Method Validation, European Medicines Agency, 2011. n.d.; Garofolo 2004; Tiwari and Tiwari 2010; Moein et al. 2017). Validation was performed on *Mix3,* containing nopaline, octopine and mannopine. The validated parameters were selectivity, linearity, sensitivity, intra- and inter-assay precision and accuracy, extraction efficiency (recovery), stability, carry-over, and matrix effect. Selectivity was also verified for the detection of all the opines studied in this work.

6.1. Selectivity (specificity)

Selectivity refers to the ability of the method to precisely detect and unequivocally distinguish a particular compound from all the other endogenous components present in a complex mixture or in the solvent. For each opine, selectivity was assessed by analyzing UHPLC-grade water samples (the solvent used for sample preparation, n=5) and plant blank matrices (n=5 for tomato and kalanchoe; n=3 for bird's-foot trefoil) compared to standard solutions and spiked plant blank matrices (QC). The presence of potential interfering components co-eluting with each opine was investigated based on retention time and accurate mass. MS/MS fragmentations were studied when required. For each opine, the specificity of the method was validated if no interfering peak was observed for its characteristic ions at the expected retention times.

6.2. Linearity (Calibration curve)

Linearity was evaluated by analyzing three or four different calibration curves for each opine. Twelve concentration levels were tested: 10, 50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10 000 and 15 000 ng/mL (prepared as described in 2.4). For each opine standard, the slope and the y-intercept of the calibration curve were calculated, as well as the coefficient of determination (R^2). The linearity range was established as the interval between the limit of quantification (LOQ) and the highest concentration of the compound which still maintained a good linearity. It was considered satisfactory when the R^2 of the calibration curves was higher than 0.990 for all the curves.

6.3. Sensitivity (LOD and LOQ)

The limit of detection (LOD) for each opine was defined as the lowest detectable concentration and was calculated as $3\sigma/S$ (σ is the standard deviation of the y-intercepts of the regression lines and S the slope of the calibration curve) (ICH. Harmonized Tripartite Guideline. Validation of Analytical Procedures: Text and Methodology Q2(R1). n.d.). The LOQ was also defined for each opine as the lowest concentration that could be quantified and calculated as $10\ \sigma/S$ (ICH. Harmonized Tripartite Guideline. Validation of Analytical Procedures: Text and Methodology Q2(R1). n.d.).

6.4. Intra- and inter-assay precision and accuracy

Accuracy and precision were assessed by analyzing *Mix3* at four concentrations (250, 500, 750 and 1000 ng/mL). Intra-assay precision and accuracy were estimated by repetitive measurements (n=4) in a single run. Inter-assay precision and accuracy were estimated, by injecting the same standard solutions on three different days (n=3). The mean, the standard deviation, the relative standard deviation (% RSD) and the percent deviation of the calculated concentrations (% bias) were obtained for each standard level. Precision (% RSD) and accuracy (% bias) were considered as acceptable when below 15% (<20% at the LOQ level) (Bioanalytical Method Validation. Guidance for Industry. US FDA, 2018. n.d.; Garofolo 2004).

6.5. Extraction efficiency (Recovery)

Extraction efficiency was determined by comparing the peak areas obtained from a known amount of each opine (1 µg/mL) spiked in a plant blank matrix before or after extraction (n=4). Results were expressed in percentages of recovery between the two signals (Bioanalytical Method Validation. Guidance for Industry. US FDA, 2018. n.d.; Garofolo 2004).

6.6. Matrix effect

In quantitative LC-MS bioanalysis, matrix effects can cause ion suppression or enhancement because the complex matrix composition can influence the ionization of the targeted compound, providing incorrect data that negatively affect the performance of the study. The matrix effects were assessed by performing the following two experiments: (1) the matrix effect of plant blank matrix (crude extract at 10 mg/mL) on opine ionization was evaluated by comparing the peak areas of the targeted opine spiked at 500 ng/mL in the plant blank matrix to the peak areas obtained by injecting the neat standard solution at the same concentration in water; (2) quantifications were performed for a set of four opine concentrations (250, 500, 1000 and 2500 ng/mL) spiked in a 100-fold diluted plant blank matrix (corresponding to 0.1 mg/mL of crude extract). The quantified values were calculated from either a solvent-based calibration curve or a matrix-based calibration curve (with a 100-fold diluted plant blank matrix). The matrix effects on the analytical recovery rates of each opine were evaluated by comparing these quantification results (n=3). Recovery rates were expressed as percentages of the expected value and considered as acceptable when above 85%. Relative standard deviations (% RSD) were also calculated.

6.7. Stability

Three opines at low, medium and high concentrations (250, 750 and 1000 ng/mL) were analyzed at three different dates within a 12-day period to evaluate their stability, with storage at -20°C and freeze-thaw cycles between each analysis. The analytical signals obtained for each opine concentration were compared between days, and the relative standard deviations (% RSD) were calculated and considered as acceptable when below 15% (<20% at the LOQ level) (Tiwari and Tiwari 2010).

6.8. Carry-over

Carry-over was evaluated by analyzing a solvent blank sample right after analyzing the highest concentration level of the calibration curve or the sample containing the greatest concentration of the targeted opine. The extracted ion chromatograms (EIC) were then studied to detect if any trace residues from the first injection of the targeted opine were found in the second injection. If a carry-over phenomenon was detected, it was considered as acceptable for quantification when below 20% of LOQ (Garofolo 2004; Moein et al. 2017).

7. Statistical analysis

A normality test was performed with a Shapiro-Wilk test (p>0.05) to analyze quantification data. Data were also analyzed with a Bartlett test to verify the homoscedasticity of the values (p>0.05). A parametric test (ANOVA) was performed to compare the data from all conditions (statistical

significance set at $p<0.05$). All statistical analyses and plots were performed with the open-source software program RStudio (Version 1.1.383 – © 2009-2017 RStudio, Inc.) using packages “ade4”, “agricolae”, “car”, “carData”, “ggplot2”, “GraphR” and “RVAideMemoire”.

RESULTS AND DISCUSSION

1. Opine detection by UHPLC-ESI-MS-QTOF: towards a new diagnostic tool

1.1. A wide range of opines detected by UHPLC-ESI-MS-QTOF

Solutions of commercial or purified opines belonging to five different opine families were prepared in UHPLC-grade water and analyzed to test their ability to ionize and be detected by UHPLC-ESI-MS-QTOF. All eighteen molecules tested were detected (**Table 1**), and all but agrocinopine A were observed with a major ion corresponding to the protonated molecule $[M+H]^+$, with sometimes the presence of ions corresponding to the adduct ions $[M+Na]^+$, $[M+K]^+$ and/or the cluster ion $[2M+Na]^+$. Agrocinopine A was the only one observed in the form of ionic species $[M+Na]^+$ and $[M+K]^+$. All these polar compounds were eluted between 0.845 and 1.440 min within our 10 min chromatography gradient. The accurate ion mass obtained for each opine was consistent with the theoretical mass calculated from the molecular formula of the corresponding ionic species. All the compounds were also submitted to MS/MS analyses to determine their characteristic MS^2 fragmentation data (**Table 1** and **Fig. S1**). Glucopinic acid was directly fragmented in the ESI source with a daughter ion at *m/z* 148.0627 observed together with the $[M+H]^+$ ion at *m/z* 312.1321 in the MS spectrum; it thus appeared as a possible unstable opine. This fragment ion was suspected to correspond to the glutamic acid moiety of glucopinic acid (theoretical mass of the $[M+H]^+$ ion: 148.06043, calculated for $C_5H_{10}NO_4$; mass error, 15.3 ppm). To our knowledge, few opine fragmentation data are available in the literature for LC-MS/MS analysis, but such data are useful for the identification of these compounds. For example, comparisons with MS/MS data of synthesized standard compounds were used to identify butylated opines (alanopine, lysopine, strombine, tauropine) in untargeted analysis of derivatized abalone extracts (Venter et al. 2017). With similar product ions, our results are in agreement with their octopine standard fragmentation data (Venter et al. 2017). The MS/MS fragmentation data presented in **Table 1** will thus be helpful for further opines studies.

Each opine was then independently spiked in tomato blank matrix extract to evaluate whether they could also be detected within a complex mix of phytochemicals (*i.e*. a plant extract). All opines tested were detected with retention times, accurate masses and MS/MS fragmentations similar to

those obtained in standard solutions. Thus, all natural opines tested were detected and identified by UHPLC-ESI-MS/MS-QTOF analyses in pure solution and in plant extract, without requiring prior derivatization.

A previous work used synthetic non-natural opine analogs to analyze their bacterial degradation in culture media (Nautiyal et al. 1991). Such opine analogs were also included in the present work. They were galactopine and glucopine (isomers of mannopine), galactopinic and glucopinic acids (isomers of mannopinic acid), allo-octopine (a diastereoisomer of octopine) and nor-mannopine (a synthetic opine). All of them were detected by UHPLC-ESI-MS/MS-QTOF analyses in pure solution and when spiked in tomato blank matrix, but the isomers exhibited retention times, accurate masses and MS/MS fragmentations similar to their natural opine analogs. Hence our UHPLC-ESI-MS-QTOF method could also be used to monitor the fate of each analogous compound independently spiked in culture broth.

Table 1. Results of UHPLC-ESI-MS/MS-QTOF analysis of standard compounds belonging to different opine families.

Opines		UHPLC-ESI-MS-QTOF analysis					UHPLC-ESI-MS/MS-QTOF analysis
			[M+H]$^+$ or [M+Na]$^+$				
Opine family	Compound	RT (min)	Measured *m/z*	Calculated *m/z*	Δppm	Molecular formula	**Main product ions observed** *m/z* (relative abundance)
Nopaline	Nopaline	0.970	305.1481	305.1456	8.3	$C_{11}H_{21}N_4O_6$	305(32) 287(28) 228(20) **200(100)** 175(41) 158(24) **130(86)** 70(66) 60(52)
	Pyronopaline	1.440	287.1371	287.1350	8.4	$C_{11}H_{19}N_4O_5$	287(49) 270(6) **228(90)** 227(27) 200(9) 184(10) **182(47)** 181(41) 156(9) 142(7) **130(100)** 114(9) 97(8) 95(8) 84(17) 71(9) 70(17)
Cucumopine	Cucumopine	0.860	284.0905	284.0877	9.8	$C_{11}H_{14}N_3O_6$	238(8) 220(14) **205(100)** 204(10) 194(16) 192(8) **177(57)** **176(57)** 163(7) 149(21) 148(39) 135(12)
Octopine	Octopine	0.916	247.1419	247.1401	7.4	$C_9H_{19}N_4O_4$	230(3) 188(7) 175(12) 158(7) **142(100)** **130(41)** 114(13) 98(27) 70(39) 60(38)
	Allo-octopine	0.917	247.1422	247.1401	8.6	$C_9H_{19}N_4O_4$	247(5) 188(7) 175(9) 158(5) **142(100)** **130(36)** 116(5) 114(11) 98(28) 87(10) 72(9) 70(38) 60(38)
	Octopinic acid	0.845	205.1203	205.1183	9.8	$C_8H_{17}N_2O_4$	205(6) **188(30)** 187(8) **142(100)** **141(40)** 116(11) 70(23)
	Histopine	0.869	228.0999	228.0979	8.8	$C_9H_{14}N_3O_4$	228(44) **182(85)** **138(100)** 111(1) 95(33) 83(3) 68(1)
	Alanopine	0.984	162.0775	162.0761	8.7	$C_6H_{12}NO_4$	162(4) **116(100)** 88(10) 84(5) **70(87)**
Mannityl-opine	Nor-mannopine	0.852	297.1319	297.1292	8.9	$C_{10}H_{21}N_2O_8$	297(23) 280(2) 238(27) 234(12) 220(12) 216(12) 192(11) 174(8) **164(25)** 133(6) 111(6) **88(100)** 69(17) 60(12)
	Mannopine	0.872	311.1479	311.1449	9.4	$C_{11}H_{23}N_2O_8$	311(17) 294(5) 276(12) 222(13) 194(11) **182(41)** 142(27) **130(100)** 129(16) 114(17) **111(47)** 98(26) 84(31) 81(10) 69(7)
	Mannopinic acid	0.898	312.1325	312.1289	11.5	$C_{11}H_{22}NO_9$	312(65) 294(9) 276(9) **266(100)** **248(40)** 230(16) 194(15) 164(16) 160(19) 156(12) 148(12) **130(48)** 129(13) 128(14) 111(22) 102(20) 98(29) 84(74) 69(37) 61(13)
	Glucopine	0.873	311.1480	311.1449	10.0	$C_{11}H_{23}N_2O_8$	311(18) 294(4) 276(25) 258(13) 230(18) 222(19) 194(12) **182(48)** 176(16) 164(8) 142(46) **130(100)** 129(20) 114(27) **111(58)** 98(20) 84(36) 81(15) 69(9)
	Glucopinic acid	0.886	312.1321	312.1289	10.2	$C_{11}H_{22}NO_9$	nd
	Galactopine	0.872	311.1480	311.1449	10.0	$C_{11}H_{23}N_2O_8$	311(17) 294(4) 276(11) 258(7) 230(15) 222(22) 194(14) **182(44)** 176(15) 142(37) **130(100)** 129(19) 114(27) **111(52)** 98(26) 84(33)
	Galactopinic acid	0.883	312.1319	312.1289	9.6	$C_{11}H_{22}NO_9$	312(62) 294(12) 276(14) **266(100)** **248(44)** 230(17) 160(19) 148(18) 142(34) **130(36)** 116(23) 114(61) 111(27) 98(36) 85(21) 84(63) 81(16) 69(26) 68(18)
	Agropine	0.907	293.1370	293.1343	9.1	$C_{11}H_{21}N_2O_7$	293(38) 275(12) 257(6) 171(16) 153(20) 141(7) **129(39)** 109(14) 95(37) 85(8) **83(100)** 69(10) 61(7)
	Agropinic acid	1.283	294.1208	294.1183	8.4	$C_{11}H_{20}NO_8$	276(7) 258(9) 230(21) 222(30) 194(13) 184(13) 176(38) 172(19) 148(15) 142(61) **130(68)** 129(32) 126(23) **114(75)** **111(100)** 98(6) 84(59) 83(36) 81(17) 69(20) 57(11)
Agrocinopine	Agrocinopine A	0.951	577.1194	577.1140	9.3	$C_{17}H_{31}O_{18}PNa$	**577(100)** **559(58)** 541(16) 492(6) 445(16) 397(14) 379(33) 320(6) **253(73)** 235(22) 216(6)

1.2 Selectivity (specificity) of opine detection

To detect putative interfering compounds co-eluting with each opine, plant blank matrix extracts were analyzed. Standard compound solutions and spiked plant blank matrix extracts were also assessed as controls. Based on the retention time and accurate mass expected for each opine, no potential interfering substance was found in the solvent blank or in the three plant blank matrices (tomato, kalanchoe, bird's-foot trefoil). Examples of typical chromatograms obtained from these investigations are shown in **Fig.1** and **Fig.S2**, in which the targeted opines were nopaline and mannopine, respectively. To verify the absence of putative interfering substances in tumors (which are more complex than non-infected stem), the accurate mass of a given opine was searched in an extract of a tumor that did not biosynthesize this opine (*e.g.* in **Fig.1-G** nopaline was sought in the extract of a tumor induced by an octopine-type strain). No significant interference at the same expected retention time and with the same expected accurate mass was detected in these tumor matrices. Thus, the method specificity was validated for all opines.

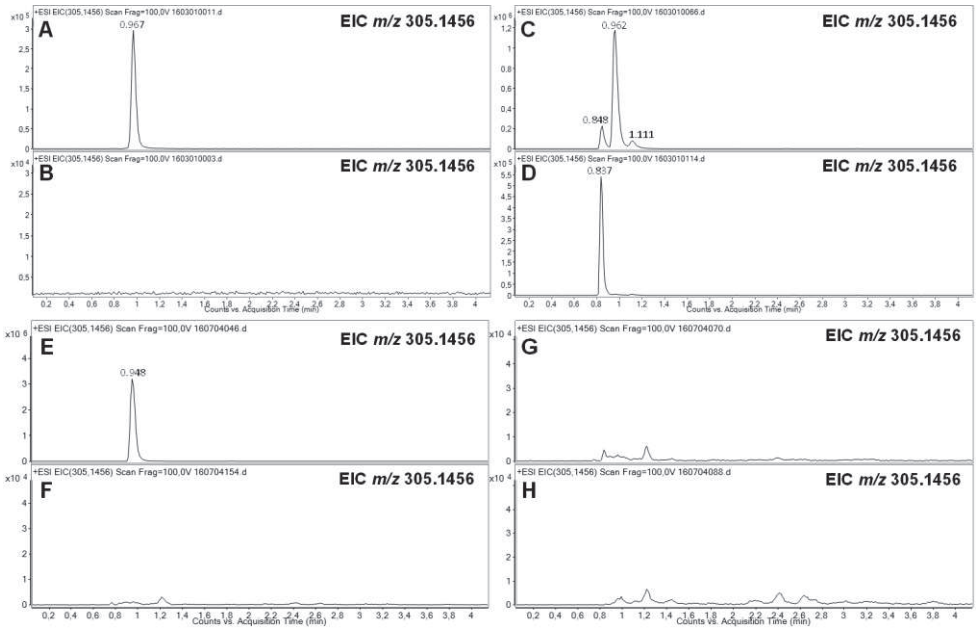


Fig. 1. Typical chromatograms of the search for nopaline in different types of samples. Each chromatogram is an extracted ion chromatogram (EIC) at *m/z* 305.1456 (expected value for the ionic species $[M+H]^+$ for nopaline). A- injection of a standard solution of nopaline at 500 ng/mL in water; B- injection of water used for preparing the samples and standard solutions (blank solvent injection); C and D- injection of crude extracts (10 mg/mL) of *K. daigremontiana* plant tissues: tumor induced by the C58 strain harboring a nopaline-type Ti plasmid (C), and non-inoculated plant stem tissue (kalanchoe blank matrix) (D); E to H- injection of crude extracts (10 mg/mL) of *Solanum lycopersicum* plant tissues: tumor induced by the C58 strain harboring a nopaline-type Ti plasmid (E), non-inoculated plant stem tissue (tomato blank matrix) (F), tumor induced by the C58 derivative strain (CFBP 1898) harboring an octopine-type Ti plasmid (G), and stem sample above the latter tumor (H). In C and D, the chromatographic peak with a retention time around 0.84 min corresponds to an ion at *m/z* 305.1375 which is not nopaline, whereas in A and C the chromatographic peak around 0.96 min is an ion at *m/z* 305.1490 consistent with nopaline. At the retention time of nopaline (observed in the standard

solution analyses in the same sequence), no potential interfering substance was detected in the analyses of blank injection (B) or of blank matrix extracts of kalanchoe (D) and tomato (F and H) plants or in tumors induced by a strain harboring an octopine-type Ti plasmid not expected to induce nopaline biosynthesis in the tumor (G).

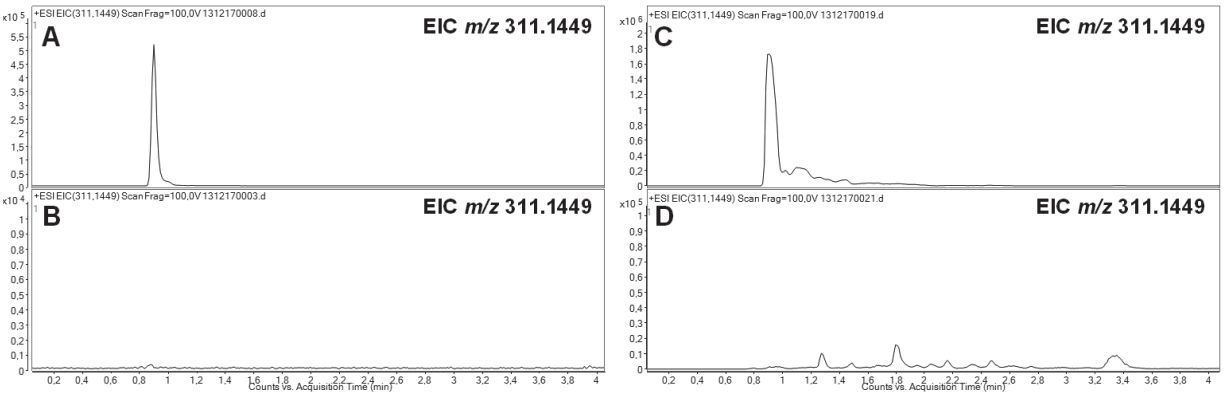


Fig. S2. Typical chromatograms of the search for mannopine in different types of samples. Each chromatogram is an extracted ion chromatogram (EIC) at *m/z* 311.1449 (expected value for the ionic species $[M+H]^+$ for mannopine). A- injection of a standard solution of mannopine at 500 ng/mL in water; B- injection of water used for preparing the sample and standard solutions (blank solvent injection); C- injection of a crude plant extract (10 mg/mL) of transgenic *Lotus corniculatus* expected to produce mannopine; D- injection of a crude plant extract (10 mg/mL) of wild-type *Lotus corniculatus* (bird's-foot trefoil blank matrix). At the retention time of mannopine (observed in the analyses of the standard solution in the same sequence), no potential interfering substance was detected in the analyses of solvent blank injection (B) or of blank matrix extracts (D).

1.3. Applications of the detection method

1.3.1 Search for opines in tumors induced by strains with different types of Ti/Ri plasmids

Strains harboring different types of Ti/Ri plasmids were inoculated in tomato plants. Tumors were extracted 21 days post inoculation and analyzed to assess whether our UHPLC-ESI-MS-QTOF method indeed detected the expected opines according to literature data on opine synthase genes or previous reported opines (**Table 2**). Extracted ion chromatograms (EIC) were performed based on the accurate masses of the putative ionic species of the expected opines. For each analysis, a solvent blank, tomato blank matrix extract and extracts of tumors not expected to produce the targeted opines were used as controls to make sure that the detected signals were not caused by interfering substances. When available, standard solutions of the expected opines were also analyzed to confirm identification with a better confidence level (Schymanski et al. 2014). Whatever the inoculated strain, at least one of the expected opines (if not all) was detected in the tumor extracts (**Table 2**).

Table 2. Detection of opines in extracts of tomato tumors induced by strains harboring diverse Ti/Ri plasmids

Bacterial species	Strain	Literature data		UHPLC-ESI-MS-QTOF analysis of tomato tumor extracts				
		Expected opines in the tumor	Reference	RT (min)	$[M+H]^+$ Measured *m/z*	$[M+H]^+$ Ion molecular formula	$[M+H]^+$ Δppm	Opine detected
Agrobacterium sp.	CFBP 2407	Cucumopine, octopine	(Ridé et al. 2000)	0.934	247.1422	$C_9H_{19}N_4O_4$	8.6	Octopine [b]
Agrobacterium fabrum	CFBP 2788	Chrysopine, nopaline, santhopine (DFG[a]), DFOP[a]	(Chilton et al. 1995)	1.178	291.1242	$C_{11}H_{19}N_2O_7$	18.9	Chrysopine
				1.020	305.1497	$C_{11}H_{21}N_4O_6$	13.6	Nopaline [b]
				0.940	309.1324	$C_{11}H_{21}N_2O_8$	10.2	Santhopine (DFG)
Agrobacterium fabrum	C58	Nopaline, agrocinopines A & B	(Montoya et al. 1977; Ellis and Murphy 1981)	0.939	305.1472	$C_{11}H_{21}N_4O_6$	5.4	Nopaline [b]
Agrobacterium radiobacter	CFBP 296	Nopaline	(Pionnat et al. 1999)	0.940	305.1518	$C_{11}H_{21}N_4O_6$	20.4	Nopaline [b]
Agrobacterium sp. G1	TT111	Octopine	(Vigouroux et al. 2017)	0.872	247.1450	$C_9H_{19}N_4O_4$	19.9	Octopine [b]
Agrobacterium sp. G1	CFBP 2712	Agropine, mannopine	(Ridé et al. 2000)	0.928	293.1400	$C_{11}H_{21}N_2O_7$	19.3	Agropine [b]
				0.894	311.1502	$C_{11}H_{23}N_2O_8$	17.1	Mannopine [b]
Agrobacterium sp. G3	CFBP 4424	Succinamopine	(Pionnat et al. 1999)	0.983	263.0900	$C_9H_{15}N_2O_7$	10.0	Succinamopine
Allorhizobium vitis	S4	Vitopine, rideopine	(Szegedi et al. 1988; Chilton et al. 2001)	0.996	219.0987	$C_8H_{15}N_2O_5$	5.3	Vitopine
				0.837	219.1350	$C_9H_{19}N_2O_4$	4.9	Rideopine
Allorhizobium vitis	CFBP 2736	Cucumopine, octopine	(Ridé et al. 2000)	0.962	247.1429	$C_9H_{19}N_4O_4$	11.4	Octopine [b]
Rhizobium rhizogenes	NIAES 1724	Mikimopine	(Isogai et al. 1990)	0.935	284.0955	$C_{11}H_{14}N_3O_6$	27.4	Mikimopine / cucumopine [b]
Rhizobium rhizogenes	CFBP 2692	Agropine, mannopine	(Ridé et al. 2000)	0.918	293.1419	$C_{11}H_{21}N_2O_7$	25.8	Agropine [b]
				0.895	311.1497	$C_{11}H_{23}N_2O_8$	15.5	Mannopine [b]
Rhizobium rhizogenes	K599 (NCPPB 2659)	Cucumopine	(Failla et al. 1990)	0.935	284.0935	$C_{11}H_{14}N_3O_6$	20.4	Cucumopine / mikimopine [b]

[a] DFG = deoxy-fructosyl-glutamine = santhopine; DFOP = deoxy-fructosyl-oxo-proline

[b] These compounds were identified according to retention time and spectral data, and confirmed by comparison to the data of authentic standard obtained in the same sequence of analyses.

Most of opines with available standard compounds were detected in tumors when their presence was expected, except in three cases: cucumopine was not detected in extracts from tumors induced by strains CFBP 2407 and CFBP 2736, and agrocinopine A was not detected in the tumor caused by strain C58. As all opines were detected when they were spiked in a plant extract (see above 3.1.1), several hypotheses can be proposed. For instance, they may have been preferentially consumed by the bacteria inside the tumor or they may not have been produced (or in too small quantities) under our experimental conditions (incubation time, plant species ...). However, cucumopine was detected in the K599-induced tumor, which confirms that cucumopine can indeed be detected in an induced plant tumor. The result foreseen for strain NIAES 1724, expected to induce mikimopine biosynthesis in the hairy roots, was also confirmed. As cucumopine and mikimopine are diastereoisomers (Isogai et al. 1990), they were not distinguished by MS analyses and showed similar retention times in our separation conditions. The use of a chiral column or another specific stationary phase (Goda and Sakamoto 1993; Suzuki et al. 2001) might separate them on the chromatogram. In any case, they are not expected to be produced together, as their biosynthesis is driven by different genes currently described as being harbored by different Ri plasmids (Suzuki et al. 2001).

Even when standard opines were not available, we detected several expected opines based on the consistency of accurate mass *vs* theoretical mass and on the absence of similar signals detected in any of the control analyses performed in the same run sequence. Chrysopine, deoxy-fructosyl-glutamine (santhopine), succinamopine, vitopine and rideopine were thus unequivocally detected in the tumor extracts (**Table 2**). Two kinds of succinamopines (D,L- and L,L-succinamopines) encoded by different synthases have been described in tumors (Chilton et al. 1985; Blundy et al. 1986; Shao et al. 2019). As they have the same molecular formula, they are undistinguishable by MS, and the one we detected was annotated with the generic term succinamopine. However, it was most likely to be D,L-succinamopine because the opine synthase gene sequence of the Ti plasmid of strain CFBP4424 (*susD* gene) is identical to the *susD* sequence of the Ti plasmid of the reference strain EU6 (Shao et al. 2019). *Allorhizobium vitis* S4 strain induces vitopine (=heliopine) and rideopine biosynthesis by transformed plant cells (Chilton et al. 2001). These two compounds have the same molecular weight (218 Da), but owing to their different molecular formulae we differentiated them by UHPLC-ESI-MS/MS-QTOF based on the accurate masses of their adduct ions $[M+H]^+$. Furthermore, vitopine MS/MS data confirmed this identification with a higher confidence level, consistent with its structure (**Fig. S3**).

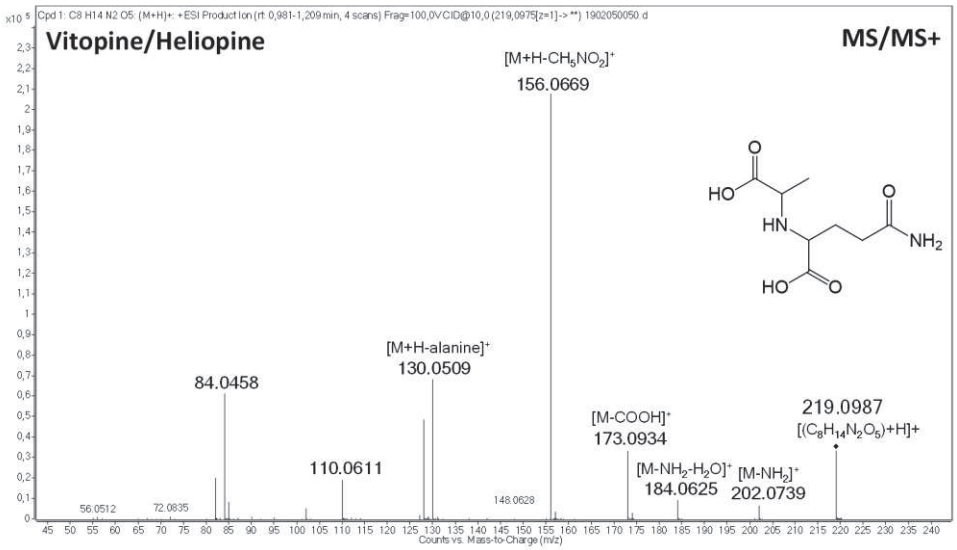


Fig. S3: MS² spectrum of vitopine (=heliopine) detected by UHPLC-ESI-MS/MS-QTOF analysis (positive ionization mode) in tomato tumor induced by *Allorhizobium vitis* S4 strain.

All these results demonstrate that our analytical method allowed for the detection of opines of different structural types in real tumors or hairy roots induced by pathogenic strains. Whatever the inoculated strain, at least one expected opine was detected in the tumor extracts. Therefore, this procedure appears to be a useful diagnostic tool.

1.3.2 Diagnostic tool of tumors: the case of *Rosa* sp. plants suspected to have crown-gall disease

Two rose plants from a horticultural nursery with abnormal stem tissue outgrowths similar to crown-gall disease symptoms were subjected to diagnostic analyses to determine if these outgrowths were caused by pathogenic agrobacteria. For each rose plant, gall tissues were harvested, extracted and analyzed by UHPLC-ESI-MS-QTOF, and the accurate masses of the putative ionic species of known opines were searched for in the MS data from each tumor extract.

Nopaline was detected in one of the two tumors (**Fig. 2**), and its identification was confirmed by standard compound analysis. Therefore, only the tumor containing nopaline was caused by a pathogenic agrobacterium. The results of the conventional diagnostic process (with strain isolation on agrobacteria-selective media followed by pathogenicity tests on model plants (Campillo et al. 2012)) were consistent with the opine detection results: a pathogenic agrobacterium was only isolated in the nopaline-positive tumor (data not shown). These results demonstrate that (i) opines can be detected in rose tissue (one more plant matrix), and (ii) our UHPLC-ESI-MS-QTOF method is efficient as a diagnostic tool to identify abnormal tissue outgrowths caused by crown-gall disease.

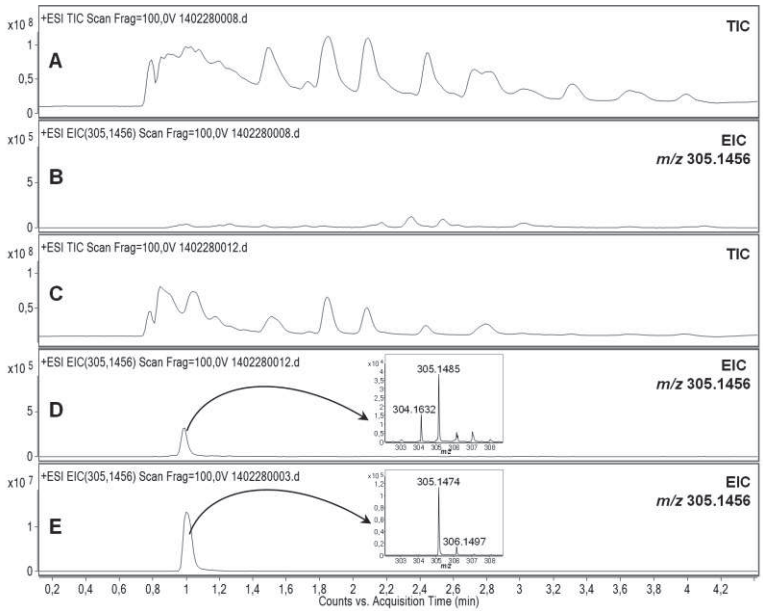


Fig. 2. Diagnostic analysis of tumors harvested on *Rosa* sp. plants suspected to have crown-gall disease. A and C- Total ion chromatograms (TICs) of crude extracts (10 mg/mL) of tumors harvested on two different *Rosa* sp. plants; B, D and E- Extracted ion chromatograms (EICs) at *m/z* 305.1456 (expected value for the ionic species $[M+H]^+$ of nopaline) of the crude tumor extracts from *Rosa* sp. plants (B and D) and a standard solution of nopaline at 1 µg/mL in water (E). The spectra show the accurate mass of the ions detected in D and E.

Detecting opines in plant extracts by UHPLC-ESI-MS-QTOF would be helpful for researchers and plant growers to reveal crown-gall disease. The necessary sanitary measures could then be taken to avoid dissemination and perennial contamination, and thus prevent pathogenic agrobacteria establishing in crops. Quantifying opines produced in gall tissues and hairy roots, or translocated in other plant parts (Savka et al. 1996), would also be a useful tool in more academic research to better understand this pathogenic plant-bacterium interaction. This is the reason why the ability and the sensitivity of our method to quantify opines are evaluated and validated below.

2. Validation of the quantification method

The quantification method was validated on nopaline, octopine and mannopine. These three opines represent different types of opine families commonly described in crown gall (Moore et al. 1997) and with standard compounds available in the laboratory.

2.1. Linearity (Calibration curve)

Calibration curves were constructed by plotting the peak areas as a function of standard compound concentrations (ng/mL) and by using a linear regression model. The calibration curves, correlation coefficients (R^2) and linear ranges of each standard opine are shown in **Table 3.** Whatever

the matrix, the calculated R^2 values were higher than 0.995, so the linearity was considered as acceptable. The linear range was opine-dependent, and slightly restrained for nopaline. The slope of these calibration curves revealed that relative "area/concentration" responses were opine-dependent too, highlighting different ionizations of the compounds. For example, at a similar concentration, octopine had an ionization intensity higher than nopaline and mannopine. Thus, to quantify a given opine precisely, quantification should not be expressed relatively to the calibration curve of another opine (unless it has been previously proven that they ionize similarly in the same conditions of analysis).

Table 3. Calibration curves, correlation coefficients, LOD, LOQ, and linear ranges of nopaline, octopine and mannopine in two matrices. For each calibration curve, each point corresponds to the mean ± SD (n=4 for the solvent blank; n=3 for tomato blank matrix). LOD, LOQ and linear range data are expressed in ng/mL and in nmol/mL so as to be easily compared with literature data.

Matrix	Opine	Calibration	R^2	LOD		LOQ		Linear range	
				ng/mL	nmol/mL	ng/mL	nmol/mL	ng/mL	nmol/mL
Solvent blank (water)	Nopaline	y = 1883x – 50975	0.9984	74.3	0.24	247.8	0.81	248 - 2500	0.81 - 8.22
	Octopine	y = 5541x – 225067	0.9994	68.7	0.28	229.1	0.93	229 - 5000	0.93 - 20.30
	Mannopine	y = 2514x – 52160	0.9997	81.9	0.26	273.1	0.88	273 - 5000	0.88 - 16.11
Tomato blank matrix 100-fold diluted	Nopaline	y = 1775x + 80364	0.9982	134.7	0.44	449.0	1.48	449 - 2500	1.48 - 8.22
	Octopine	y = 2418x + 212965	0.9956	119.7	0.49	399.0	1.62	399 - 5000	1.62 - 20.30
	Mannopine	y = 692x + 87112	0.9966	78.5	0.25	261.7	0.84	262 - 5000	0.84 - 16.11

2.2. Sensitivity (LOD and LOQ)

The LOD and LOQ of our method were expressed as the lowest concentrations that could be detected or quantified, respectively. The results in water solution and in 100-fold diluted tomato blank matrix are presented in **Table 3**.

Previous works on opine quantification methods usually report detection limits based on analyses of standard solutions in solvent and expressed the results as the lowest quantity of opine detected in the test. To our knowledge, LOQs have not been reported. Using an injection volume of 10 µL for the standard solution in water, we detected 0.7 to 0.8 ng of opines (2.4 to 2.8 pmol, depending on the opine) and quantified at least 2.3 to 2.7 ng (8.1 to 9.3 pmol). Thus, our method was sensitive enough to quantify a wide range of opines. As a comparison, a LOD of 1µg/spot (3 nmol) was reported using high-voltage paper electrophoresis (HPVE) with silver staining for mannityl-opines, whereas HPVE with phenanthrenequinone staining yielded LODs of 2 and 5 µg/spot (8-10 nmol) for octopine and nopaline, respectively (cited in (Zhang et al. 1998)). Several HPLC methods for opine detection have been described with various detection techniques associated or not to a derivatization step,

specific to some opine types (Sato et al. 1988; Zhang et al. 1998; Sandee et al. 1996; Sauerwein and Wink 1993; Firmin 1990). For example, LODs of 0.05 µg for cucumopine/mikimopine and 0.02 µg for agropine and mannopine have been reported for HPLC coupled to a refractometer and without derivatization (Sauerwein and Wink 1993). Thus, our method is more sensitive. Zhang *et al.* reported an HPLC method with fluorescence detection and prior derivatization with detection limits between 0.1 to 5 pmol for opines, among which nopaline, octopine and mannopine [33]. Even though the authors did not attribute a value to each individual opine, our method displayed similar sensitivity with a LOD of 2.4-2.8 pmol with standard solutions, but without requiring the time-consuming derivatization step. With the biosensor approach for nopaline and octopine detection (Choi et al. 2019), researchers reported linearity ranges of 0-100 nM and 0-50 nM for nopaline and octopine, respectively. However, they did not precisely determine the LOD and LOQ of their method. As the first concentration they tested was of 1 nM in a volume of 2 mL, their biosensor seemed to be able to detect at least 2 pmol, which is in a range of sensitivity similar to ours and the one of Zhang *et al.* [33]. However, their method was restricted to octopine and nopaline.

Other methods have been applied to quantify certain opines in crown gall or transformed plant tissues. Mannopine has been quantified between 0.5 to 2.3 µg/cm² of tobacco leaf (Zhang et al. 1998), whereas the amount of nopaline and octopine have been suggested to be around 10-20 and 70-74 mg/g dry weight of potato and tomato tumor tissue, respectively (Choi et al. 2019). Johnson *et al.* showed that the amounts of octopine in tobacco tumors induced by various *A. tumefaciens* strains ranged from 0.75 to 6.25 mg/g of dry tissue (Johnson et al. 1974), whereas another work reported between 24 and 411 µg/g dry weight of crown gall tissue in various plant species for octopine, and between 5.5 and 19.9 mg/g for nopaline (Scott et al. 1979). Given these reported amounts of opine in infected plant tissue, the sensitivity of our method towards nopaline, octopine and mannopine seems to be adequate for routine detection and quantification requirements.

2.3. Intra- and inter-assay precision and accuracy

Precision (%RSD) and accuracy (%bias) results are shown in **Table 4**. Intra-assay precision and accuracy were 0.5-1.3% and 0.01-8.0%, respectively, whereas inter-assay precision and accuracy were 1.3-11.9% and 0.5-10.7%, respectively. For each opine and concentration level, precision and accuracy were below 15%. Therefore, the method was considered robust and reproducible (Bioanalytical Method Validation. Guidance for Industry. US FDA, 2018. n.d.; Garofolo 2004).

Table 4. Results of intra- and inter-assay precision and accuracy of the UHPLC-ESI-MS-QTOF method developed to quantify opines.

Compound	Expected concentration (ng/mL)	Intra-assay (n = 4)			Inter-assay (n = 3)		
		Concentration (ng/mL)	Precision	Accuracy	Concentration (ng/mL)	Precision	Accuracy
		Mean ± SD	%RSD	%bias	Mean ± SD	%RSD	%bias
Nopaline	1000	939.9 ± 4.8	0.5	6.1	966.7 ± 91.3	9.4	3.3
	750	727.6 ± 8.6	1.2	3.0	728.6 ± 36.8	5.1	2.9
	500	473.2 ± 5.9	1.3	5.4	510.2 ± 19.5	3.8	2.0
	250	239.4 ± 1.2	0.5	4.2	261.0 ± 26.7	10.2	4.4
Octopine	1000	948.0 ± 5.4	0.6	5.2	936.2 ± 74.7	8.0	6.4
	750	772.4 ± 9.8	1.3	3.0	746.1 ± 38.7	5.2	0.5
	500	487.4 ± 3.1	0.6	2.5	486.9 ± 58.1	11.9	2.6
	250	250.0 ± 2.4	1.0	0.01	225.2 ± 6.7	3.0	9.9
Mannopine	1000	1,001.1 ± 7.5	0.8	0.1	941.9 ± 13.6	1.4	5.8
	750	809.9 ± 5.1	0.6	8.0	712.9 ± 9.0	1.3	5.0
	500	518.1 ± 5.1	1.0	3.6	446.3 ± 15.4	3.4	10.7
	250	244.7 ± 1.0	0.4	2.1	274.0 ± 6.7	2.4	9.6

2.4. Extraction efficiency (recovery)

The efficiency of opine extraction was determined by comparing the analytical signal obtained for each opine in a blank tomato matrix spiked with known amounts of opines before and after extraction. The results ranged between 103 and 108% (**Table S2**).

Table S2. Opine extraction efficiency in tomato blank matrix. Opine peak areas were obtained from tomato blank matrix with known amounts of opines spiked before (Condition A) or after (Condition B) extraction.

Opines	Conditions	Area under curve (AUC)						Recovery %	Precision %RSD
		1	2	3	4	Mean	SD		
Nopaline	**A**	77992056.8	58330739.4	68998063.5	68272390.0	68398312.4	8036821.8	**107.95**	11.8
	B	69306330.3	59909286.5	58017145.7	66211952.7	63361178.8	5289801.1		8.3
Octopine	**A**	157180.2	143318.9	149395.4	153201.2	150773.9	5899.4	**103.44**	3.9
	B	157293.4	135915.6	135009.7	154841.6	145765.1	11944.0		8.2
Mannopine	**A**	1784715.9	1398753.8	1606992.6	1713956.1	1626104.6	168254.9	**105.76**	10.3
	B	1674511.7	1478305.4	1467405.1	1529893.0	1537528.8	95301.9		6.2

2.5. Carry-over

The EIC of the targeted opines obtained from the solvent blank analyzed right after a sample with a high opine concentration (the highest concentration level of the calibration curve or a plant sample with the greatest concentration) did not reveal any peak indicating the presence of trace residues of these opines. Consequently, the method was considered as reliable for routine analysis.

2.6. Stability

Opine stability during storage at -20°C for 12 days with freeze-thaw cycles between each analysis was assayed at low, medium and high concentrations (250, 750 and 1000 ng/mL). The %RSD values of the analytical signals were less than 15%, and were considered acceptable (**Table S3**). At a concentration close to the LOQ, octopine proved to be the least stable opine with a %RSD of 14.96%. Investigations beyond 12 days at -20°C were not performed.

Table S3. Opine stability at three different concentrations. Solutions of opines at three different concentrations corresponding to low, medium and high concentration levels were analyzed on three different days, with storage at -20°C and freeze-thaw cycles between each analysis. The resulting areas under curves were compared.

Opines	Concentration (ng/mL)	Area under curve (AUC)					%RSD
		11-October	22-October	23-October	Mean	SD	
Nopaline	1000	6808957	6482722	7000605	6764095	261840	**3.87**
	750	4631758	4361450	4819080	4604096	230066	**5.00**
	250	1200786	1262937	1176878	1213534	44423	**3.66**
Octopine	1000	28151832	29532382	26710945	28131720	1410826	**5.02**
	750	20726905	22989135	20084658	21266899	1525678	**7.17**
	250	6696381	6760428	5124262	6193690	926706	**14.96**
Mannopine	1000	18675537	22165368	17927793	19589566	2261824	**11.55**
	750	13937275	17189990	13811329	14979531	1915349	**12.79**
	250	3856973	4830844	3974463	4220760	531604	**12.59**

2.7. Matrix effect

Matrix effect on opine ionization was evaluated directly in the crude extract (tomato blank matrix extract at 10 mg/mL) by comparing the peak area of the targeted opine in the spiked tomato blank matrix with the one obtained by injecting the neat standard solution at the same opine concentration. This preliminary test showed that all opines were detected, but with analytical recoveries ranging between 5 and 20% according to the opine (data not shown). This crude extract concentration revealed an influence of the complex composition of the matrix on opine ionization (decreased signal intensity) and established the need to dilute the crude extract to reach precise quantification.

Then, the matrix effect was assessed by quantification tests performed with a set of four concentration levels of opines spiked in diluted extract of tomato blank matrix. The quantified values were calculated from either a solvent-based or a matrix-based calibration curve, and the recovery rates were expressed as percentages of the expected values. The results for the three opines in 100-fold diluted tomato blank matrix are presented in **Table 5**.

Table 5. Results of matrix effect assessment: recovery (%) of opine standards spiked into a diluted tomato blank matrix, calculated from solvent-based or matrix-based standard curves. Mean recovery

rates were expressed as percentages of the expected values, and relative standard deviations (% RSD) were calculated (n=3).

Opine	Plant blank matrix	Opine concentration (ng/mL)	Quantification			
			Solvent-based standard curve (water)		Tomato matrix-based standard curve (100-fold diluted)	
			Recovery (%)	%RSD	Recovery (%)	%RSD
Nopaline	Tomato blank matrix, 100-fold diluted	2500	99.0	6.2	96.2	6.6
		1000	109.3	7.3	105.4	7.4
		500	108.8	4.3	103.6	5.1
		250	105.0	2.8	97.1	0.8
Octopine	Tomato blank matrix, 100-fold diluted	2500	47.4	6.2	95.7	6.4
		1000	56.1	4.9	107.0	5.5
		500	62.9	2.1	109.7	2.7
		250	66.2	2.2	93.2	3.4
Mannopine	Tomato blank matrix, 100-fold diluted	2500	29.1	3.4	97.7	3.6
		1000	34.5	5.4	107.8	6.9
		500	39.0	2.7	108.3	4.3
		250	44.4	4.5	96.6	7.3

These results showed good recovery rates for nopaline with both calibration curves (solvent-based and tomato matrix-based), showing that nopaline quantification could be done precisely with any of them. On the other hand,the octopine and mannopine quantification results obtained from the solvent-based calibration curve showed a matrix effect. However, their quantifications were accurate when performed with the tomato matrix-based calibration curve which allowed taking this matrix effect into account.

Therefore, lower ionization resulting from a matrix effect of the extract can be solved or decreased by using an appropriate dilution in the sample preparation procedure before quantitative analysis. Furthermore, matrix-based calibration curves have to be used if these three opines (nopaline, octopine and mannopine) have to be quantified in the same run. However, if nopaline is the only opine targeted in a tomato sample, it could be quantified precisely with a simple solvent-based calibration curve.

These tests were also performed with kalanchoe blank matrix, and similar results were obtained. They also showed the need to use a 100-fold dilution of the crude extract before quantitative analysis and to use a matrix-based calibration curve to obtain accurate results. In this case, the use of an appropriate dilution made it possible to avoid an artificial enhancement of the signal. Linearity and matrix effect results for kalanchoe blank matrix are presented in **Table S4** and **S5**.

Using an appropriate dilution of the plant extract and a matrix-based calibration curve, UHPLC-ESI-MS-QTOF analysis permits an accurate quantification of opines present in plant tissues, which could in turn be useful for researchers to quantify opines produced in gall tissues or translocated to other plant parts.

Table S4. Calibration curves, correlation coefficients, LOD, LOQ, and linear ranges of three opines in kalanchoe matrix. For each calibration curve, each datum corresponds to the mean ± SD (n=3). LOD, LOQ and linear range data were expressed in ng/mL and nmol/mL so as to be easily compared with literature data.

Matrix	Opine	Calibration	R^2	LOD		LOQ		Linear range	
				ng/mL	nmol/mL	ng/mL	nmol/mL	ng/mL	nmol/mL
Kalanchoe blank matrix, 100-fold diluted	Nopaline	y = 2070x + 256635	0.9958	97.9	0.32	326.2	1.07	326 - 2500	1.07 - 8.22
	Octopine	y = 6267x + 925406	0.9966	44.3	0.18	147.5	0.60	148 - 1000	0.60 - 4.06
	Mannopine	y = 2930x + 463641	0.9922	45.7	0.15	152.2	0.49	152 - 1000	0.49 - 3.22

Table S5. Matrix effect assessment: recovery (%) of opine standards spiked into diluted kalanchoe blank matrix, calculated from solvent-based or matrix-based standard curves. Mean recovery rates were expressed as percentages of the expected values, and relative standard deviations (% RSD) were calculated (n=3).

Opine	Plant blank matrix	Opine concentration (ng/mL)	Quantification			
			Solvent-based standard curve (water)		Kalanchoe matrix-based standard curve (100-fold diluted)	
			Recovery (%)	%RSD	Recovery (%)	%RSD
Nopaline	Kalanchoe blank matrix, 100-fold diluted	2500	118.5	0.1	99.5	0.2
		1000	141.4	12.5	113.4	13.9
		500	138.3	0.2	100.4	1.1
		250	154.1	5.2	94.0	8.5
Octopine	Kalanchoe blank matrix, 100-fold diluted	1000	136.3	7.7	103.1	8.5
		500	163.7	2.0	109.1	2.4
		250	204.0	3.5	108.1	5.9
Mannopine	Kalanchoe blank matrix, 100-fold diluted	1000	134.4	7.4	97.8	12.7
		500	165.1	0.9	107.1	6.4
		250	205.8	4.4	108.3	8.5

2.8. Application of the quantification method

2.8.1 Influence of the At plasmid on the opine content of tumors induced by isogenic strains

The *A. fabrum* C58 strain contains a circular chromosome, a linear chromosome, and two large plasmids – the Ti and At plasmids (Wood et al. 2001; Goodner et al. 2001). The ecological role of the At plasmid (pAt) is not defined as well as that of the Ti plasmid (pTi) (Rosenberg and Huguet 1984). However, it may have a positive effect on virulence gene induction, and pAt-free strains have been

reported to be less virulent (inducing the formation of smaller tumors) (Nair et al. 2003). These observations led us to investigate if the presence of the pAt had an impact on the opine content of tumors of infected plants. Opine quantifications were carried out on non-inoculated plants and on plants inoculated with the *A. fabrum* wild-type strain (C58), as well as with isogenic strains harboring or not the pTiC58 and/or the pAtC58 (strains UIA5, AB150, AB152 and AB153 described in **Table S1**) (Nair et al. 2003). Tumors or stem tissues were extracted 21 days post inoculation, and the crude extracts were diluted 100-fold and analyzed by UHPLC-ESI-MS-QTOF to detect and quantify the expected nopaline. Quantifications were performed from a 100-fold diluted tomato matrix-based calibration curve as described above. First, as expected, nopaline was not detected in the control samples (non-inoculated plants, and plants inoculated with the UIA5 plasmid-less strain and the AB150 strain harboring only pAtC58), confirming additionally the absence of an interfering substance. Secondly, nopaline quantification in the tumors induced by the strains harboring both, the pTiC58 and the pAtC58 (strains C58 and AB153) or only the pTiC58 (strain AB152) showed nopaline amounts ranging from 1.64 to 2.30 µg *per* mg of crown gall tissue dry weight (**Fig.3**). These results were consistent with literature data, *e.g.* a previous work reports nopaline amounts of 10-20 µg *per* mg of tumor dry weight in tomato tumors obtained 5-6 weeks post inoculation (Choi et al. 2019). Finally, our results revealed significantly higher nopaline contents in the tumors of plants inoculated with the strains harboring both plasmids (AB153, C58) than in the tumors induced by strain AB152 that only harbored the pTiC58 (**Fig.3**). In addition, the quantities of bacteria still persisting in the three types of tumors were similar (data not shown), indicating that the variation of the opine content did not result from a difference in *A. fabrum* concentration in the tumor tissue. The lower nopaline abundance in the tumors induced by At plasmid-free agrobacteria is consistent with previous works suggesting that the At plasmid increases *A. fabrum* C58 virulence (Nair et al. 2003).

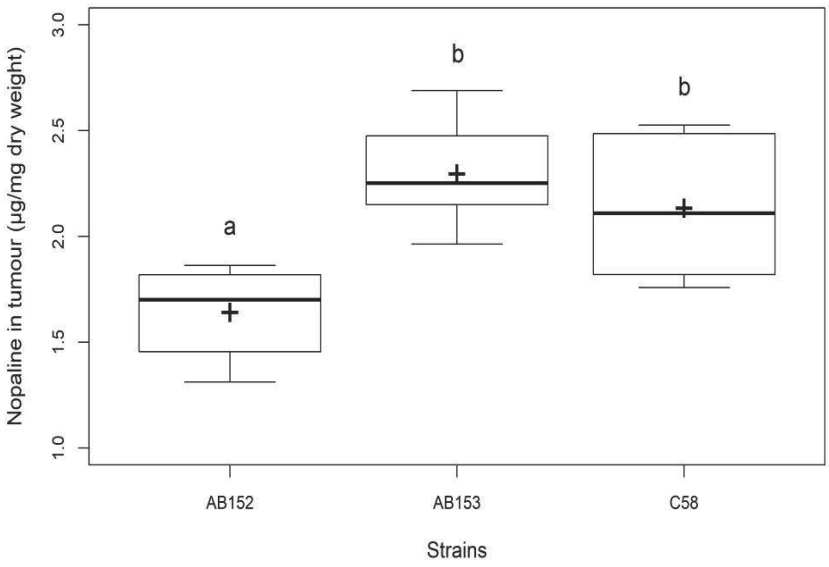


Fig. 3. Box-plots illustrating the nopaline contents of tumors induced by isogenic strains harboring pAt or not. Strains C58 (wild-type) and AB153 (isogenic to C58) harbored pTiC58 and pAtC58, while strain AB152 (isogenic to C58 without pAtC58) contained only the pTiC58. Nopaline quantification was expressed in µg/mg of tumor dry weight. Boxes cover 50% of the data. Central lines represent the medians and whiskers represent the minimum and maximum values among non-atypical data. The cross (+) denotes the mean value of the data (n=6). Different letters indicate significant differences between strains (ANOVA, $p<0.05$).

CONCLUSION

In the present work, a general, rapid, specific and sensitive method was developed and validated for opine analysis. Using a simple process of sample preparation (solid-liquid tumor extraction, without time-consuming steps of purification and/or derivatization) followed by an UHPLC-ESI-MS/MS-QTOF analysis (run-time of 10 min), opines of different structural types were detected unequivocally in pure or mixed solutions, and in extracts of tumors/hairy roots induced by strains harboring various Ti/Ri plasmids. Detection and identification were based on the accurate mass of the precursor ion, unequivocal monoisotopic mass, accurate mass of MS/MS fragment ions, and the matching of MS spectra and retention times with data obtained from standard compounds when available. The absence of similar signals detected in all the control analyses (blank solvent, plant blank matrix, extracts of tumors not expected to produce the targeted opine) further validated opine detection and identification in plant extracts. Moreover, our method was successfully used with several plant species (tomato, rose, kalanchoe, and bird's-foot trefoil). Taking all these results into account, at least one molecule of each opine family commonly described in infected plants (Dessaux and Faure 2018) was shown to be detectable. Indeed, opines belonging to the nopaline-, octopine-, mannityl-opine-,

agrocinopine-, cucumopine-, succinamopine-, chrysopine- and heliopine/vitopine families were detected using this method.

The nature of the opines produced by the abnormal tissue outgrowths depends on the plasmid harbored by the pathogenic infecting strain. These opines constitute persistent biomarkers of plant infection by pathogenic agrobacteria. The UHPLC-ESI-MS/MS-QTOF method presented in this study represents a powerful tool to perform diagnostic analyses of plant tumor tissue because it can detect a wide range of opines in one run. It will be useful for researchers and plant growers to determine if a gall observed on a plant is caused by infection by pathogenic agrobacteria or not. As no curative methods are available for this plant disease, early diagnosis of crown-gall is essential for appropriate sanitation precautions to be taken to avoid dissemination and perennial contamination, and thus prevent pathogenic agrobacteria establishing in crops.

Based on this UHPLC-ESI-MS-QTOF method, a quantitative assay was also developed and validated for three selected opines representing different types of opine families commonly described in crown gall (nopaline, octopine and mannopine). With an appropriate dilution of the plant extract during sample preparation and the use of a matrix-based calibration curve, this method proved to be sensitive enough to allow for accurate and reproducible quantification of these opines in plant extracts (tomato and kalanchoe). As the ionization was found to be opine-dependent, opine quantification based on only one selected standard should be avoided because it could lead to quantification errors. The method was successfully applied to study the nopaline content of tumors induced by isogenic strains harboring the At plasmid or not, confirming the positive effect of this plasmid on *A. fabrum* C58 virulence highlighted in previous works.

Quantitative monitoring of opines in plant samples offers the opportunity to accurately quantify opine production in crown-gall tissues or hairy roots, or to evidence the spatiotemporal distribution of one opine in an infected plant. Thus, from a more academic perspective, this method can also constitute a useful tool for researchers to better understand these pathogenic plant-bacterium interactions.

C. Influence of *A. fabrum*-specific regions on opine production

INTRODUCTION

In this part we present another application of our new quantification method to assess whether *A. fabrum*-specific regions influence opine production in tomato tumors. As *A. fabrum* C58 harbors a nopaline pTi, it was thus the targeted opine in this study. We showed once again that this method is useful not only to detect but also to quantify nopaline in plant extracts. As the method details, we used ultra-high-performance liquid chromatography–electrospray ionization quadrupole time-of-flight–mass spectrometry (UHPLC-ESI-MS-QTOF) for nopaline quantification with no derivatization required, resulting in a rapid, specific and sensitive essay.

The ecological role of *A. fabrum*-specific regions is not yet very well defined. We already showed that these regions modulate tumor secondary metabolites, but opines are particular molecules specifically biosynthesized by transformed plant cells and their production are not determined by their presence. However, there may also be an effect on nopaline production. Indeed, we already showed that *A. fabrum*-specific regions can modulate the content of amino acids in tumors, being these the precursors of opines. As opine production depends not only on the type of plasmid harbored by the inciting bacteria, but are also dependent on the precursors available in the plant (Flores-Mireles et al. 2012), our goal was to investigate whether the presence of the *A. fabrum*-specific regions had an impact on the opine content on tumors of infected plants.

MATERIALS AND METHODS

Opine quantifications were carried out on one-month-old tomato plants (*Solanum lycopersicum*, Marmande variety) inoculated with the deletion mutants of *A. fabrum*-specific regions and with the wild-type strain listed on **Table 1** (**Chapter 3**, **Part II.A**). Inoculation procedures, tumors harvesting and

extractions are detailed on **Chapter 3**, **Part II.A** and **II.B**). Tumors were extracted 21 days post inoculation, and the crude extracts were diluted 100-fold and analyzed by UHPLC-ESI-MS-QTOF to detect and quantify nopaline. Quantifications were performed from a 100-fold diluted tomato matrix-based calibration curve.

RESULTS AND DISCUSSION

Nopaline quantification in the tumors induced by *A. fabrum* strains revealed significantly higher nopaline contents in the tumors of plants inoculated with the ΔSpG8-7b (environmental signal sensing) mutant strain and significantly lower nopaline contents with the ΔSpG8-2a (curdlan biosynthesis) mutant strain when compared with the wild-type (**Fig. 1**). Furthermore, as we already showed (**Chapter 3**, **Part II.A**), bacterial survival in tumor was similar for all the *A. fabrum* mutant strains and for the wild-type, thus, the changes in nopaline concentration are not linked to *A. fabrum* concentration in the plant tumor.

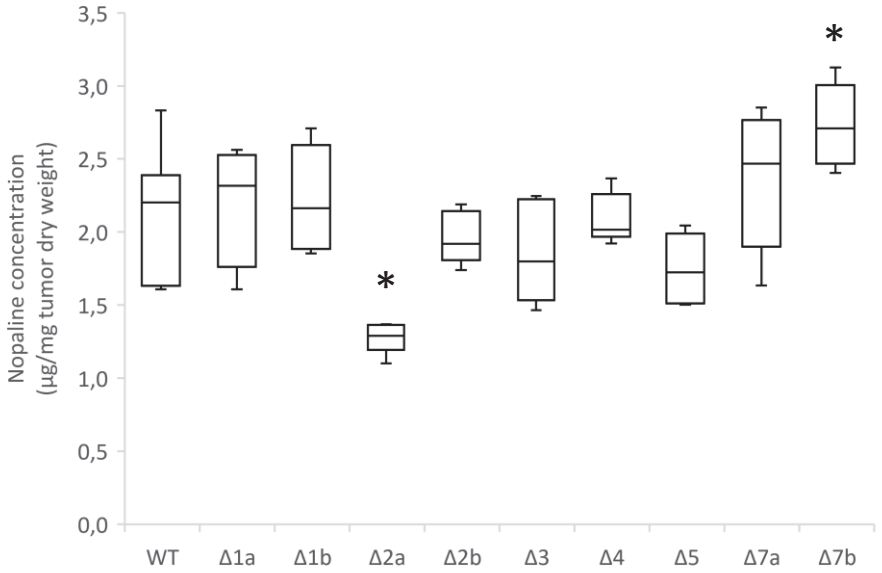


Fig. 1. Nopaline concentration on tumors induced by *A. fabrum* strains. Nopaline quantification on dry weight tomato tumors induced by deletion mutants of A. fabrum-specific regions and the wild-type strain. Data were analyzed using a t test. Statistical differences are indicated with the symbol * ($P<0.05$).

Our results showed that using an appropriate dilution of plant extract (100-fold diluted) and also a 100-fold diluted tomato matrix-based calibration curve, an accurate nopaline quantification from plant extracts is possible.

We also show that the SpG8-2a and SpG8-7b regions influence opine production in tomato tumors. These two regions seem to modulate plant compounds in opposite ways. One possible explanation is that, while one seems to modulate the opine precursors in a positive way, and thus the plant can produce a greater quantity of opine, the other instead seems to do the contrary so the opine precursors are present in a less quantity in the plant and the opine production is lower.

CONCLUSION

In this last part of this study, we present the application of our new method to the *A. fabrum* niche construction. Indeed, as our opine quantification method is sensitive enough to accurately quantify opine production in plant tumors, we were able to highlight the influence in opine production by *A. fabrum*-specific regions. Thereby, these results can lead us to a better understanding of the construction of the ecological specific niche of *A. fabrum*.

Discussion générale et Perspectives

Ces travaux de thèse se sont intéressés à l'adaptation d'*A. fabrum* à la plante et plus particulièrement à l'implication de ses gènes spécifiques sur l'interaction entre la bactérie et la plante. Deux approches différentes et complémentaires ont été réalisées dans le but d'étudier l'importance de ces gènes bactériens lors des différents styles de vie d'*A. fabrum* et surtout, leur contribution dans la construction de sa niche écologique spécifique. La première approche se base sur des mesures de valeur adaptative conférée par chacune des régions spécifiques afin de comprendre leur rôle adaptatif dans la plante. La deuxième approche a consisté en une analyse des métabolites secondaires végétaux dans le but d'étudier non seulement la réponse métabolique de la plante au contact d'*A. fabrum*, mais aussi l'impact sur celle-ci des régions spécifiques.

Style de vie commensale d'*A. fabrum*

A travers les différents chapitres de ce manuscrit, nous avons abordé les deux styles de vie *d'A. fabrum*. Le style de vie un peu moins connu de cette bactérie, le commensalisme, est tout aussi riche et complexe que son style de vie pathogène. Grâce à l'étude des gènes spécifiques d'*A. fabrum*, nous avons pu montrer leur implication dans la capacité de cette bactérie à se multiplier et se maintenir au sein de la rhizosphère en compétition non seulement avec la souche sauvage, mais aussi avec des compétiteurs d'autres espèces d'agrobactéries. Pour arriver à cela, nous avons mis en place une méthode fine et rapide pour réaliser des tests de compétition grâce à l'utilisation de marqueurs plasmidiques dont les protéines exprimées sont visibles à l'œil nu. Nous avons pu démontrer que les régions spécifiques d'*A. fabrum* confèrent, pour la plupart, une valeur adaptative à la bactérie au sein de la rhizosphère. Ces régions sont les SpG8-1a (transport et métabolisme de sucres), SpG8-1b (dégradation d'acides hydroxycinnamiques), SpG8-2a (biosynthèse de curdlan), SpG8-5 (catabolisme de composés de type opines) et SpG8-7b (systèmes mécano-senseurs). La région SpG8-4 (transport et métabolisme de sucres) confère également une valeur adaptative quand des compétiteurs d'autres espèces d'agrobactéries sont également mis en compétition avec l'espèce *A. fabrum*. Ces résultats nous montrent donc que les régions spécifiques d'*A. fabrum* lui confèrent effectivement un avantage adaptatif en lien avec la plante, comme il a été proposé par Lasalle *et al.* en 2011.

D'autre part, l'étude métabolomique au sein de la rhizosphère nous a permis de repérer les métabolites secondaires impliqués dans l'interaction avec *A. fabrum* via ses gènes spécifiques. Ces métabolites appartiennent essentiellement aux composés phénoliques et plus particulièrement à la famille des flavonoïdes appartenant aux sous-classes d'aurones, flavanones, isoflavones, flavones et flavonols. Nous nous sommes intéressé aux composés phénoliques car ils sont souvent impliqués dans

les interactions plantes-bactéries et leur modulation en réponse à la présence de bactéries a déjà été observé (Chamam et al. 2013, 2015; Walker et al. 2011, 2012; Valette et al. 2019; Miotto-Vilanova et al. 2019), et ce, même en présence d'*Agrobacterium* (Walker et al. 2013). La reconnaissance de flavonoïdes par les rhizobiums est même une base importante de la spécificité de l'hôte (Spaink et al. 1987; Spaink 1994).

Mais ce qui est novateur dans notre étude, c'est l'utilisation de souches mutantes pour étudier la réponse métabolique de la plante et la comparer à sa réponse face à la souche sauvage. Ainsi, nous nous sommes focalisés dans l'effet des régions spécifiques d'*A. fabrum* sur le métabolome racinaire de *Medicago truncatula*. Notre objectif final était de mieux comprendre l'interaction d'*A. fabrum* avec la plante au cours de son style de vie rhizosphérique ou commensale.

Les flavonoïdes glycosidés jouent un rôle important dans la physiologie et la biochimie des plantes (Taylor and Grotewold 2005) et peuvent être impliqués dans les interactions pathogènes et symbiotiques avec les micro-organismes (Dixon et al. 1994; Marczak et al. 2010). Dans certaines espèces végétales telles que *Medicago*, une espèce riche en flavonoïdes (Saleh et al. 1982), leur fraction glycosidique peut contenir de l'acide glucuronique et en plus, être acylée par des dérivés d'acides phénylpropanoïques tels que les acides p-coumariques ou féruliques (Stochmal et al. 2001b; Marczak et al. 2010). Plus de 20 flavones sont présentes dans *M. truncatula* (Kowalska et al. 2007) et beaucoup d'entre elles ne sont glycosylées que par de l'acide glucuronique (Bruijn 2019; Staszków et al. 2011; Stochmal et al. 2001a). De plus, malgré le fait que les diglucuronoflavones sont particulièrement rares, elles ont été identifiées chez quelques espèces de *Fabaceae* dont *M. truncatula* (Kowalska et al. 2007; Stochmal et al. 2001a; Marczak et al. 2010, 2016). L'acide glucuronique semble en effet impliqué dans l'interaction d'*A. fabrum* avec *M. truncatula*, car huit métabolites discriminants sur les 17 identifiés dans cette étude contiennent ce composé dans sa fraction glycosidique (apigenin O-[feruloyl-glucuronopyranosyl-O-glucuronopyranoside], 7,4'-dihydroxyflavone-O-[feruloyl-hexuronyl -O-hexuronide], diosmetin 7-O-hexuronide, kaempferol O- hexuronyl -[déoxyhexosyl]-hexoside, 7-4'-dihydroxyflavone 7-O-glucuronide, liquiritin-7-O-hexuronide, apigenin 4'-O-dihexuronide, 7,4'-dihydroxyflavone dihexuronide) dont quatre sont des diglucuronoflavones et deux sont acylés avec de l'acide férulique. Trois de ces composés (7-4'-dihydroxyflavone 7-O-glucuronide, 7,4'-dihydroxyflavone dihexuronide et 7,4'-dihydroxyflavone-O-[feruloyl-hexuronyl -O-hexuronide]) semblent appartenir à la même voie de biosynthèse des flavones et ainsi, les deux premiers pourraient être des intermédiaires conduisant à la biosynthèse du 7,4'-dihydroxyflavone-O-[feruloyl-hexuronyl -O-hexuronide] car les fragments glycoside et acylé peuvent se fixer un à un à l'aglycone comme cela a déjà été observé (Farag et al. 2007). De plus, il a été constaté que le mutant ΔSpG8-5 (catabolisme de

composés de type opines) est moins capable d'utiliser l'acide glucuronique. Cela suggère que la souche sauvage peut utiliser cet acide uronique comme source de carbone et son inoculation à la plante peut diminuer l'abondance relative de ces composés. Nous émettons ainsi l'hypothèse qu'il s'agit d'une caractéristique importante sur l'adaptation d'*A. fabrum* à la plante et à sa construction de niche écologique spécifique chez M. truncatula. Parmi ces métabolites glycosylés par de l'acide glucuronique, cette étude nous a permis de trouver trois composés jamais décrits dans la littérature qui sont : 7,4'-dihydroxyflavone dihexuronide, apigenin 4'-O-dihexuronide et 7,4'-dihydroxyflavone-O-[feruloyl-hexuronyl -O-hexuronide].

Même s'il est difficile à élucider, une explication possible aux modulations des métabolites observés dans notre étude pourrait être l'effet de l'induction ou de la répression de certaines voies de biosynthèse pour ces composés par la présence de la bactérie et plus particulièrement de ses gènes spécifiques. Une autre explication pourrait être la transformation par la plante de certains composés dans d'autres molécules en suivant la voie de biosynthèse des flavonoïdes sous l'influence des bactéries suite à la perception de cette dernière, à des mécanismes de défense ou à l'utilisation de ces composés par la plante pour interagir avec les bactéries. Il est aussi tout à fait possible que les bactéries dégradent certains de ces composés. Tous les composés discriminants détectés ont une activité biologique et même s'ils n'ont pas été testés sur *A. fabrum*, nous pouvons supposer, comme ils sont modulés dans les racines de *M. truncatula*, que ces composés ont également une activité sur les bactéries.

Notre approche permet d'aller plus loin et de mettre en évidence la modulation des composés discriminants liée aux gènes spécifiques d'*A. fabrum*. Nous avons donc trouvé des métabolites importants dans la reconnaissance spécifique de cette bactérie par la plante et des métabolites étroitement impliqués dans l'interaction particulière entre *A. fabrum* et la plante. Nous avons également mis en évidence des régions plus impliquées que d'autres dans cette interaction commensale avec la plante. Tandis qu'il y a des régions qui modulent plus les métabolites secondaires racinaires (SpG8-3, SpG8-4 et SgP8-7b), il y en a d'autres qui les modulent moins (SpG8-1b, SpG8-2b, SpG8-5 et SpG8-7b) voire pas du tout (SpG8-2a).

L'intégration de l'ensemble de ces résultats et sachant que toutes les régions spécifiques sont exprimées dans la rhizosphère, nous pouvons émettre un classement du rôle des gènes spécifiques. Tout d'abord nous pouvons évoquer la région SpG8-2a (biosynthèse de curdlan), impliquée dans la valeur adaptative d'*A. fabrum* dans la rhizosphère mais pas dans la modification du métabolisme racinaire. Le rôle de cette région peut être donc uniquement de protection de la bactérie dans le

contexte du commensalisme. En effet, la principale fonction prévue de cette région est la biosynthèse d'un exopolysaccharide (curdlan) qui entoure et protège les bactéries, et sa production est susceptible d'augmenter la survie des bactéries dans le sol (Ruffing et Chen 2012). C'est donc un rôle essentiel à la survie et persistance de la bactérie dans la rhizosphère puisque c'est cette région qui confère une valeur adaptative plus importante. Notre hypothèse est aussi que l'absence de curdlan n'a pas de rôle direct dans la perception des bactéries par la plante. Ainsi, cette dernière ne modifie ni son métabolisme, ni sa teneur en composés phénoliques, que ce soit en présence de la souche sauvage ou du mutant.

De plus, nous avons toutes les autres régions spécifiques étudiées qui interviendraient à tour de rôle à la construction de la niche écologique spécifique d'*A. fabrum*. Tout d'abord, nous avons les régions SpG8-2b (biosynthèse de métabolites secondaires), SpG8-3 (biosynthèse de sidérophores) et SpG8-7a (Systèmes mécano-senseurs), qui ne confèrent pas de valeur adaptative à la bactérie dans la rhizosphère (ou du moins, avec les méthodes proposées, nous n'avons pas repéré une quelconque implication), mais en revanche, elles sont capables de moduler les métabolites secondaires racinaires. Le rôle de ces régions peut être donc la préparation à la mise en place de la construction de la niche écologique spécifique d'*A. fabrum*. En effet, *A. fabrum* pourrait, dans un premier temps, utiliser certaines de ses fonctions écologiques, codées par ces gènes spécifiques en particulier, pour moduler son environnement à son avantage, notamment en synthétisant des sidérophores (SpG8-3) pour s'assurer l'approvisionnement en fer tout au long de l'interaction avec la plante, mais aussi en synthétisant des métabolites secondaires (SpG8-2b) pour assurer la mise en place de la communication et l'interaction avec la plante. En effet, la région SpG8-3 est celle qui module le plus les métabolites racinaires, notamment deux dérivés d'apigénine. En effet, cette molécule est connue comme étant une molécule signal lors des interactions entre les *Rhizobiaceae* et les légumineuses (Falcone Ferreyra et al., 2012; Hassan and Mathesius, 2012). Cela suggère que cette première étape de construction de niche est importante et essentielle car elle modifie de manière significative le métabolome racinaire. Ces changements peuvent également suggérer que la plante réagit quand elle aperçoit la bactérie et se prépare à son tour à la mise en place de l'interaction avec la bactérie. La région SpG8-7a pourrait fournir des signaux à la bactérie concernant son environnement pour déclencher l'interaction, comme l'état ou développement de la plante, mais aussi la présence qualitative et quantitative des microorganismes présents dans la rhizosphère.

Ensuite, avec son environnement préétabli, *A. fabrum* pourrait par la suite installer une interaction plus étroite avec la plante en employant les régions spécifiques qui sont impliquées à la fois dans sa valeur adaptative et dans la modulation de la réponse métabolique de la plante. Ces

régions seraient SpG8-1b (dégradation d'acides hydroxycinnamiques), SpG8-4 (transport et métabolisme de sucres), SpG8-5 (catabolisme de composés de type opines) et SpG8-7b (systèmes mécano-senseurs), et ainsi, finaliser la construction de sa niche écologique spécifique. En effet, la région SpG8-1b est impliquée dans la dégradation des acides hydroxycinnamiques (HCA) (Campillo et al. 2014), des métabolites secondaires communs des plantes, exsudés dans la rhizosphère (Mandal et al. 2010; Zhalnina et al. 2018; Badri et al. 2013), mais aussi libérés en grandes quantités dans le sol lors de la décomposition des cellules racinaires (Whitehead et al., 1983). Même s'il a été démontré que les HCA inhibent généralement la croissance des bactéries (Sayadi et al. 2000; Ravn et al. 1989) ou qu'ils sont toxiques pour la plupart des micro-organismes (Perret et al. 2000), ils peuvent être utilisés comme source de carbone par d'autres bactéries du sol (Deavours et al.2006; Andreoni et al.1995). Il a été montré que cette région (SpG8-1b) code pour la voie de dégradation complète de l'acide férulique (Campillo et al.) qui sera finalement utilisée comme source de carbone et d'énergie par *A. fabrum*. Une voie similaire est observée avec l'acide p-coumarique (Meyer et al.2018). La dégradation des HCAs peut donc permettre non seulement leur transformation en une molécule moins toxique, mais aussi leur utilisation comme ressource trophique. Le feruloyl-CoA et le p-coumaroyl-CoA sont les effecteurs de HcaR (un répresseur transcriptionnel) et permettent ainsi une boucle de rétroaction positive en induisant la dégradation des HCA (Meyer et al.2018). Dans les environnements riches en HCA comme la rhizosphère (Mandal et al. 2009), il est donc avantageux pour *A. fabrum* d'avoir une voie finement régulée et ainsi pouvoir dégrader et assimiler rapidement les HCA et, par ce moyen, obtenir un avantage concurrentiel sur les autres agrobactéries. De plus, une induction des gènes de dégradation de HCA uniquement après la détection de ceux-ci, est avantageuse pour les bactéries afin d'éviter le coût métabolique de leur expression constitutive dans des environnements végétaux à faible teneur en HCA (Deochand et Grove 2017). La région SpG8-1b est donc utile pour la colonisation d'environnements riches en HCAs, en les détoxifiants, en les utilisant comme ressource trophique, ou comme signal de perception de la plante.

Par ailleurs, une expression coordonnée des régions SpG8-1b et SpG8-3 a récemment été observée chez *A. fabrum* C58 en présence des HCAs (Baude et al. 2016), renforçant l'idée de l'existence d'une niche écologique à laquelle l'espèce G8 est spécifiquement adaptée (Lassalle et al. 2017). De plus, le fer semble être essentiel à la dégradation de l'acide férulique car un retard dans sa dégradation dans un milieu pauvre en fer a été observé (Baude et al. 2016). Il semble donc que la première étape de construction de niche où la bactérie produit des sidérophores pour capturer le fer soit essentielle pour la mise en place, par la suite, des autres fonctions écologiques. Postérieurement, la présence des HCA assurent à leur tour une homéostasie du fer intracellulaire (Baude et al. 2016). Nous pouvons ainsi observer que ces deux régions spécifiques sont interconnectées et corégulés, car la présence de l'une

est nécessaire pour l'expression de l'autre et vice versa. D'un point de vue écologique, *A. fabrum* a un avantage sur d'autres agrobactéries en ayant des régions qui s'expriment de manière coordonnée pour déterminer une niche écologique spécifique à l'espèce.

Sur les trois régions restantes, les régions SpG8-4 et SpG8-5 sont impliquées dans le catabolisme de carbohydrates. En effet, la région SpG8-4 est particulièrement impliquée dans le transport de monosaccharides et dans le métabolisme de sucres, et la région SpG8-5 code des enzymes probablement impliquées dans le catabolisme de condensats d'acides aminés et de sucres. Ces condensats sont possiblement des composés de type opine ou des composés Amadori (une classe de molécules produites dans la matière végétale en décomposition et donc communes dans le sol humique). Cela suggère que ces régions vont probablement intervenir une fois que la réponse métabolique de la plante face à la bactérie commence à se mettre en place et qu'elle commence à exsuder des composés utiles à *A. fabrum*. Une autre région qui pourrait être groupée ici, est la région SpG8-1a. En effet, cette région semble également impliquée dans le métabolisme de sucres car elle possède des transporteurs ABC hypothétiques spécifiques aux monosaccharides et à des acides aminés. Cependant, la réponse métabolique de la plante face à cette région spécifique reste à vérifier.

La dernière région qui semble établir une interaction étroite avec la plante est la région SpG8-7b, qui, de la même manière que la première moitié de la région, code des systèmes mécano-senseurs. Il est probable que les signaux fournis cette fois-ci à la bactérie soient plus spécifiques à l'interaction. En effet, cette région modifie une grande quantité de métabolites secondaires des racines, probablement parce que la détection du signal est essentielle pour moduler le métabolisme bactérien lié à l'interaction, qui modifiera à son tour la perception des bactéries par la plante. Cette région est donc impliquée dans l'échange de signaux entre la bactérie et la plante lors de l'interaction. Il serait alors possible de parler d'un dialogue moléculaire entre la plante et la bactérie.

Les caractères des agrobactéries qui favorisent l'adaptation à la rhizosphère des plantes restent encore à identifier plus précisément. Cependant, notre étude montre que des approches combinées peuvent nous donner des informations précieuses concernant les caractéristiques qui pourraient permettre à *A. fabrum* d'exploiter les ressources du sol et de s'adapter à la rhizosphère des plantes, un environnement très compétitif, mais plus encore, de moduler son hôte non seulement lors de son style de vie pathogène, mais aussi lors de son style de vie rhizosphérique. Cette modulation est possible grâce à l'expression combinée et coordonnée de ses gènes spécifiques dans l'interaction affinée avec la plante et ainsi, construire sa propre niche écologique spécifique.

Style de vie pathogène d'*A. fabrum*

Tout au long de ce manuscrit, nous avons traité également le style de vie pathogène *d'A. fabrum*. Ce style de vie se caractérise par la formation d'une tumeur induite par cette bactérie chez les plantes infectées en modifiant leur génome. Cette tumeur se caractérise par l'accumulation d'opines qui jouent un rôle essentiel dans ce mode de vie d'*A. fabrum*. Ainsi, la bactérie présente la remarquable capacité de construire et d'exploiter la niche écologique des tumeurs végétales qu'elle incite. Nous avons montré que ses régions spécifiques sont aussi impliquées à la fois dans sa valeur adaptative et à la modulation de la réponse métabolique de la tumeur induite par cette bactérie. En effet, nous avons montré que les régions spécifiques d'*A. fabrum* sont impliquées dans sa capacité à se multiplier et à se maintenir au sein de la tumeur.

Références

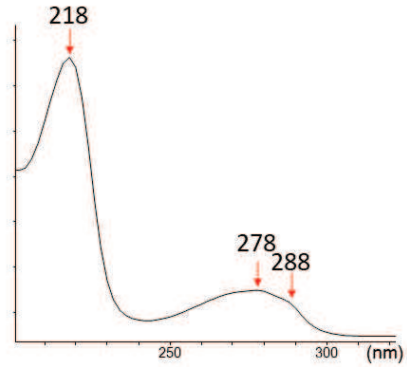
Annexes

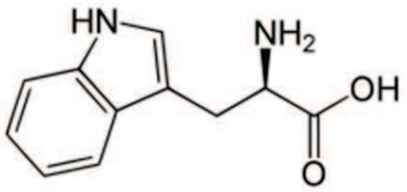
Compound X1

MW = 204Da
Formula: $C_{11}H_{12}N_2O_2$ score: 99,5

Comparison to literature: OK
Standard Analysis: OK

UV Spectrum

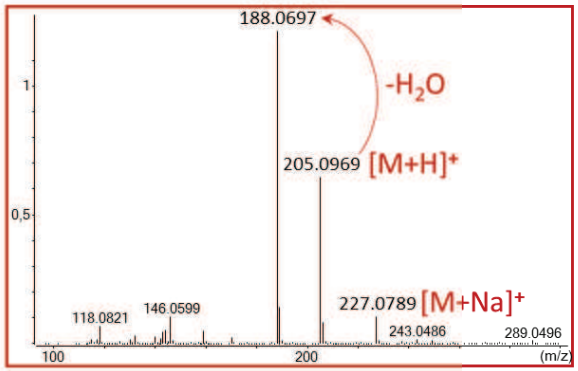




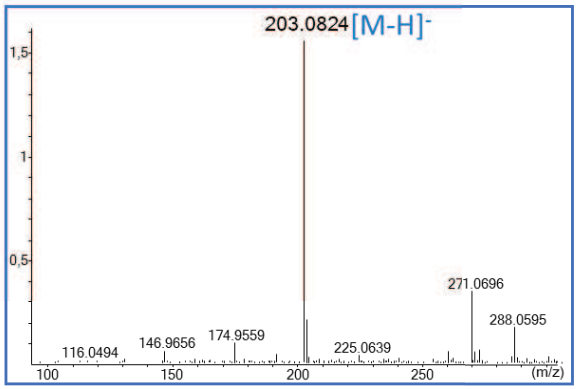
Tryptophan

MS Spectrum

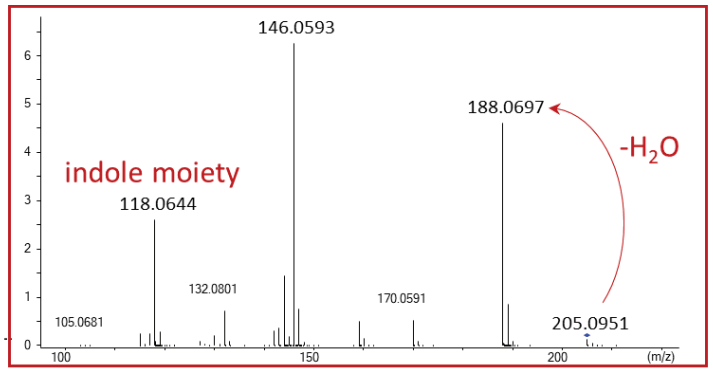
Positive

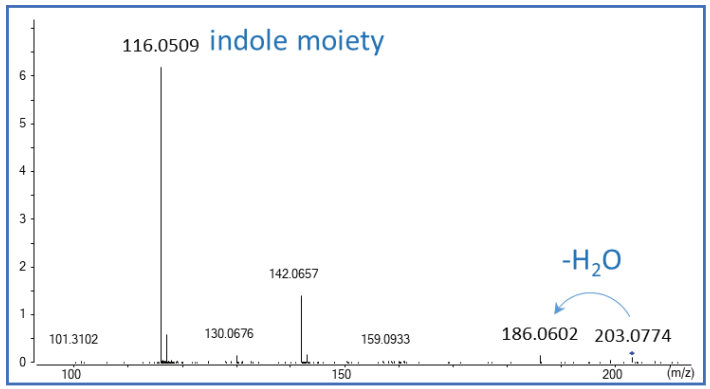


Negative



MS/MS Spectrum

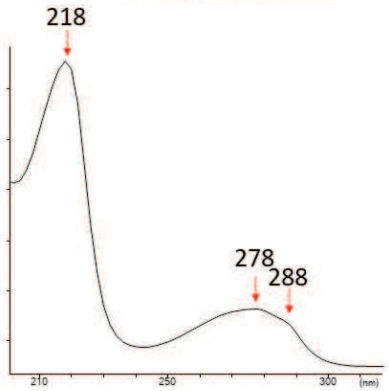


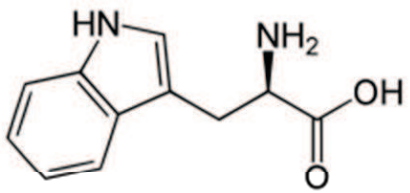


Tryptophan

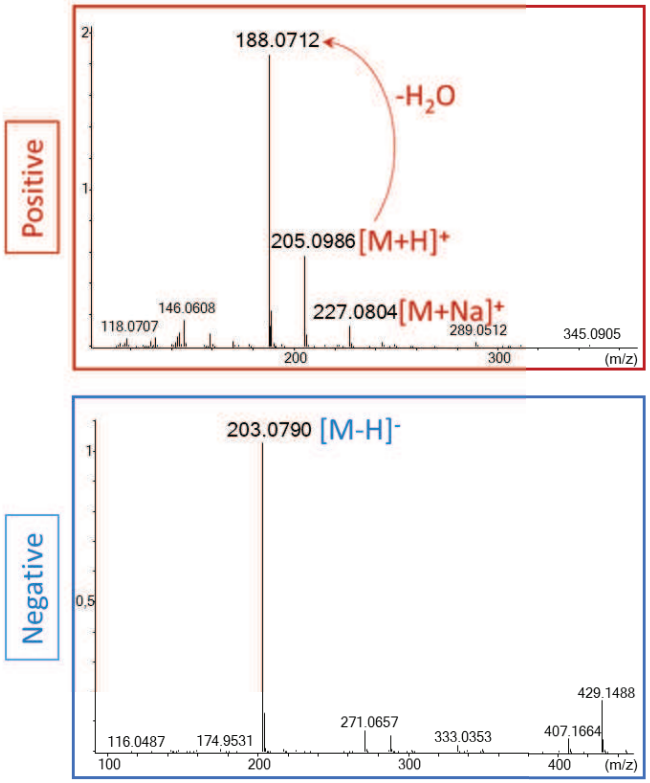
MW = 204Da
Formula: $C_{11}H_{12}N_2O_2$

UV Spectrum

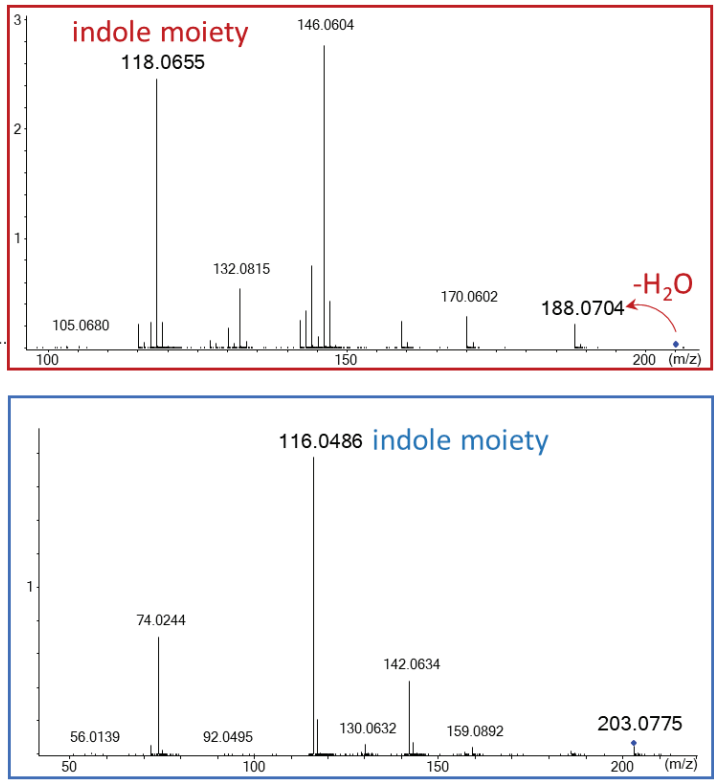




MS Spectrum



MS/MS Spectrum

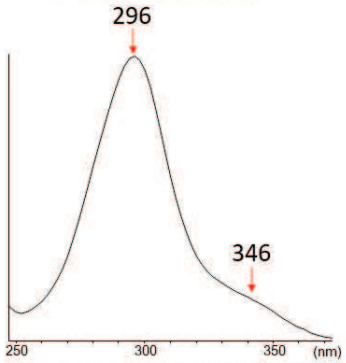


Compound X24

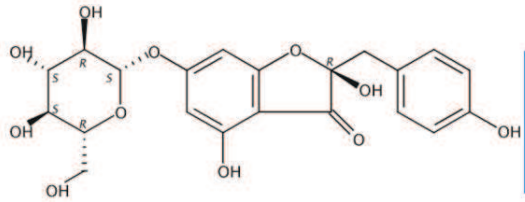
MW = 450Da
Formula: $C_{21}H_{22}O_{11}$ score: 99,79

Comparison to literature: OK
Aglycon Analysis: OK

UV Spectrum



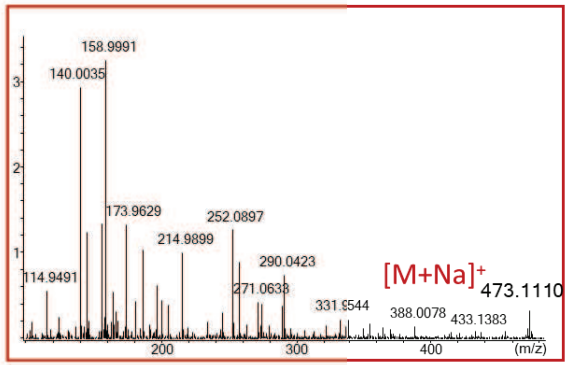
Aglycon: aurone, maesopsin type



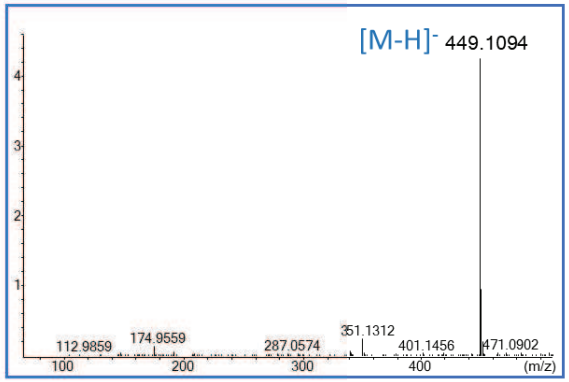
Maesopsin 6-O-glucoside

MS Spectrum

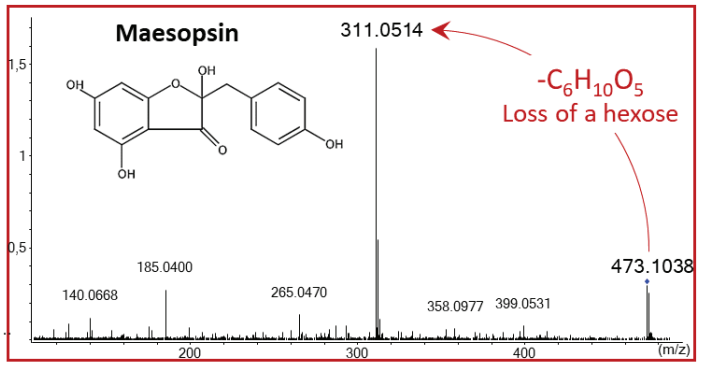
Positive

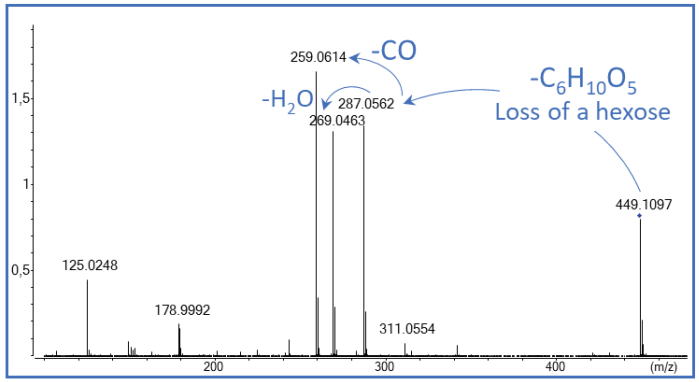


Negative



MS/MS Spectrum



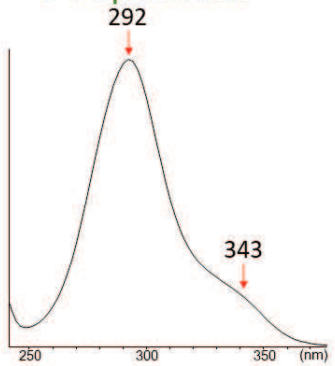


Compound X26

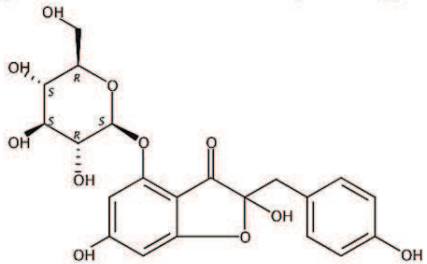
MW = 450Da
Formula: $C_{21}H_{22}O_{11}$ score:

Comparison to literature: OK
Standard Analysis: OK

UV Spectrum



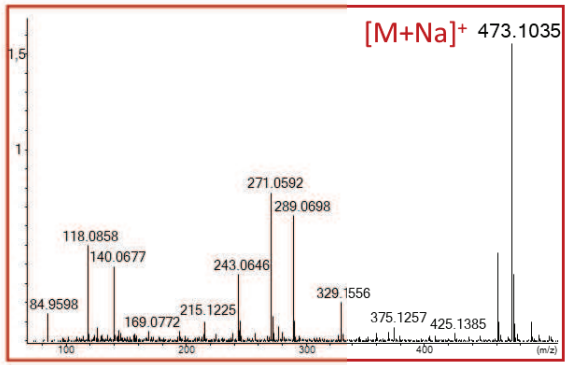
Aglycon: aurone, maesopsin type



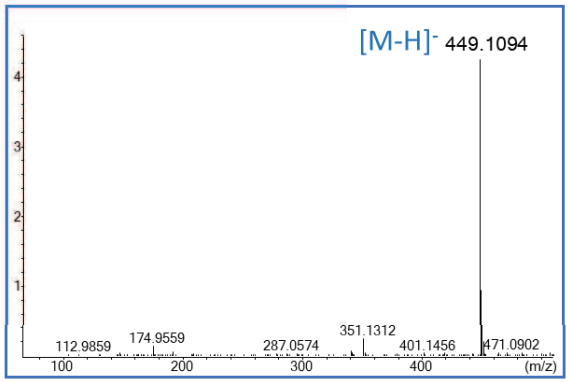
Maesopsin 4-O-glucoside
Hovetrichoside C

MS Spectrum

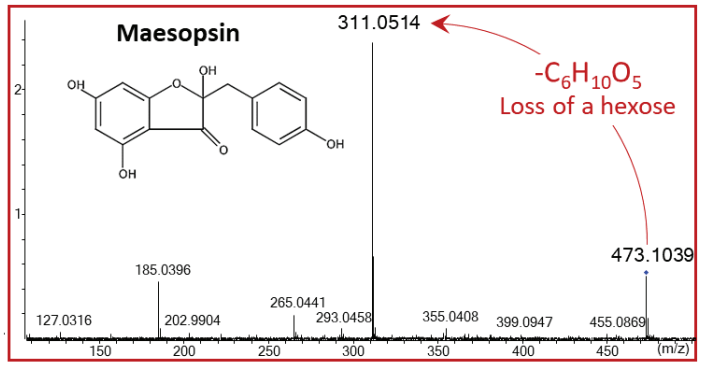
Positive

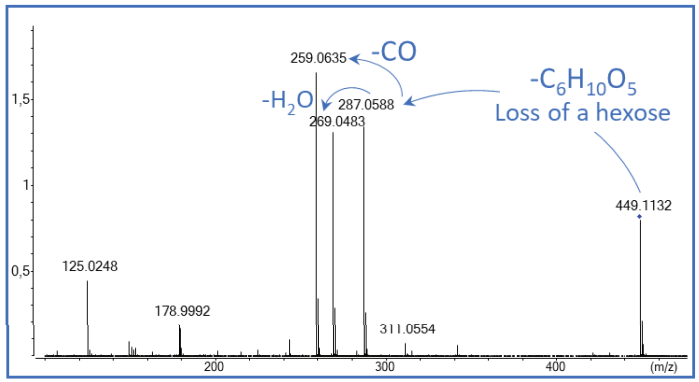


Negative



MS/MS Spectrum

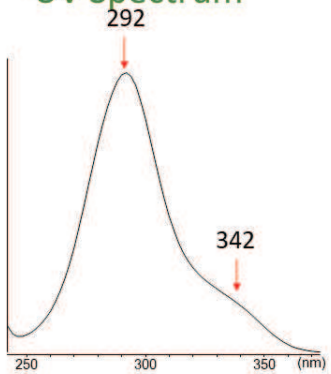




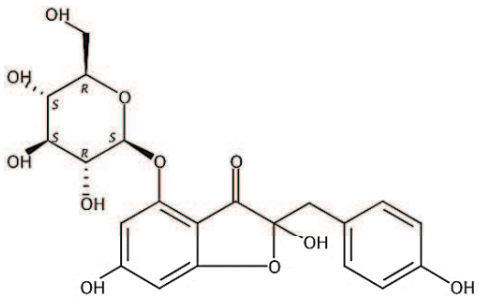
Maesopsin-4-O-glycoside

MW = 450Da
Formula: $C_{21}H_{22}O_{11}$ score:

UV Spectrum

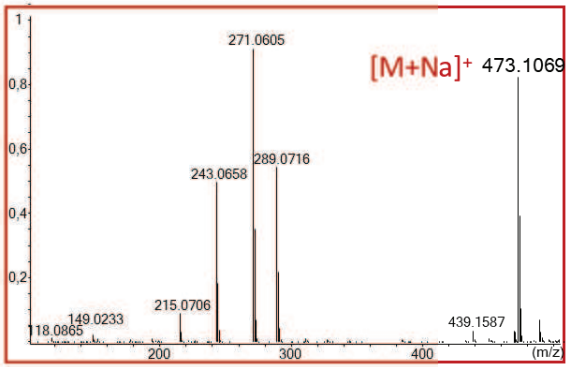


Aglycon: aurone, maesopsin type

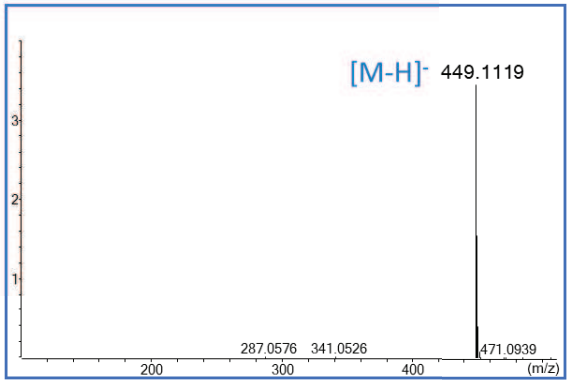


MS Spectrum

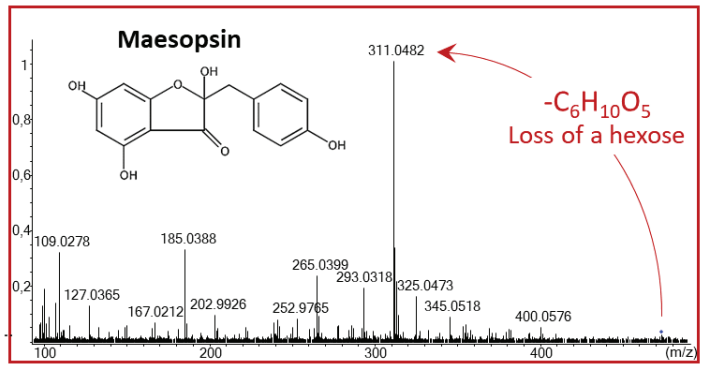
Positive

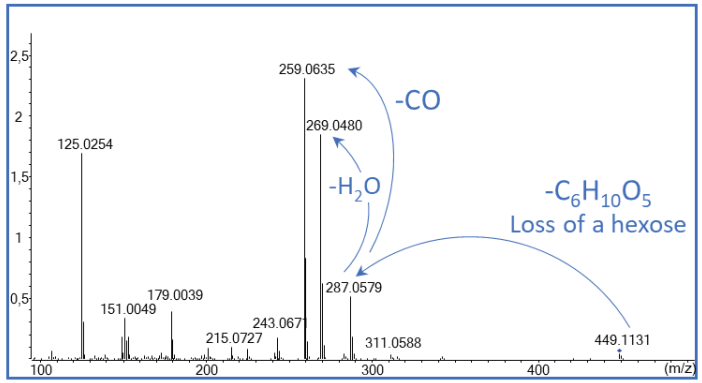


Negative



MS/MS Spectrum



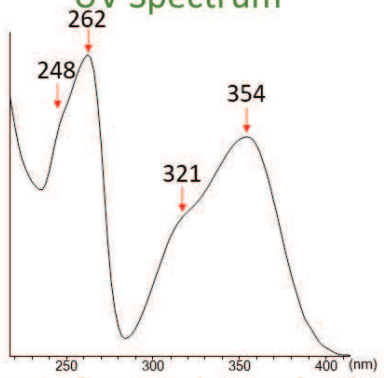


Compound X36

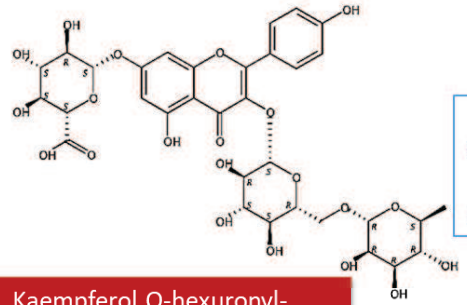
MW = 770Da
Formula: $C_{33}H_{38}O_{21}$ score:

Comparison to literature: OK
Aglycon Analysis: OK

UV Spectrum



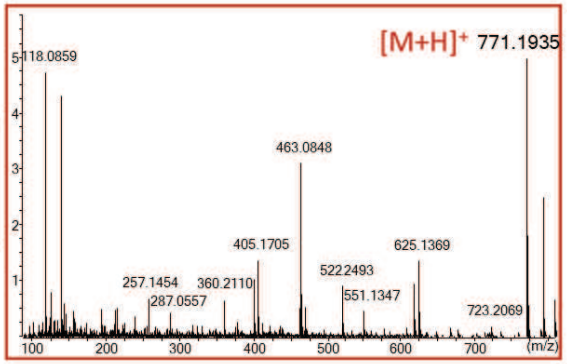
Aglycon: flavone, kaempferol type



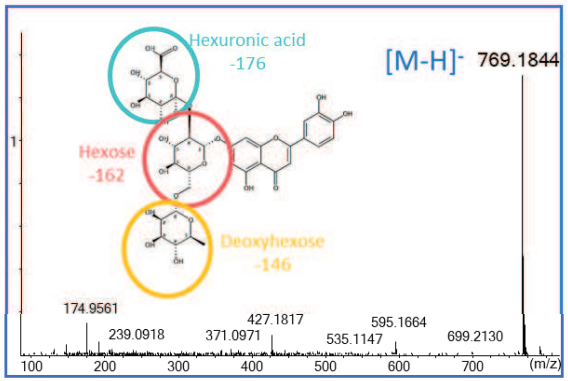
Kaempferol O-hexuronyl-[desoxyhexose]-hexoside

MS Spectrum

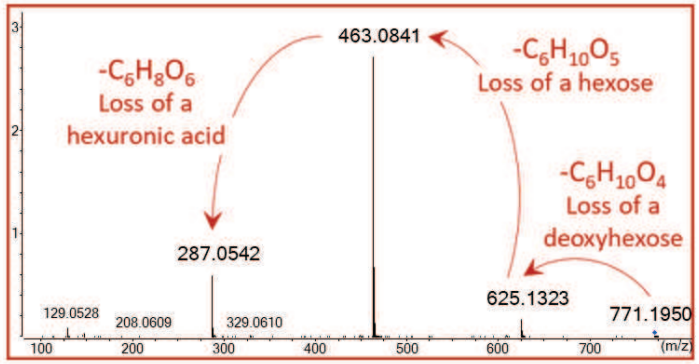
Positive

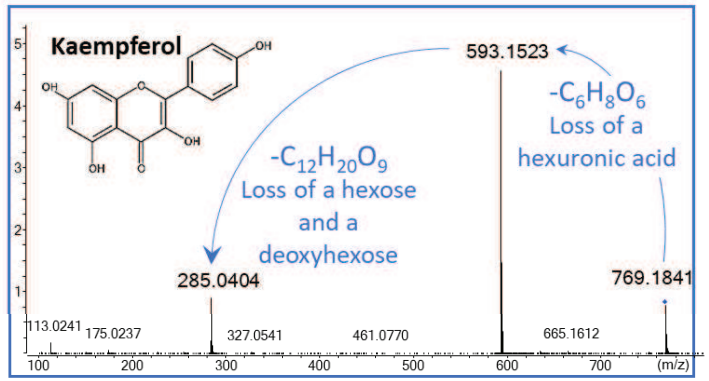


Negative



MS/MS Spectrum



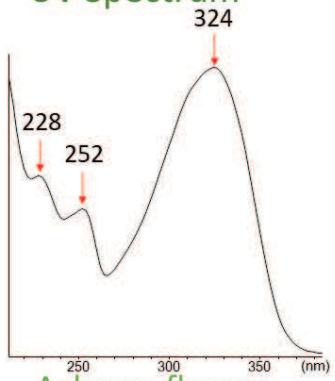


Composé X40

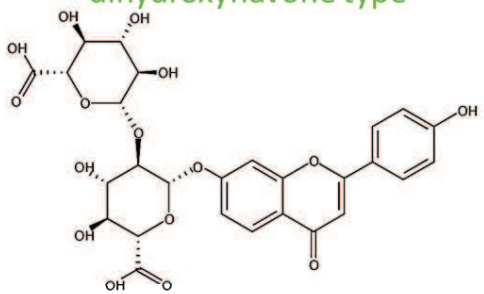
MW = 606Da
Formula: $C_{27}H_{26}O_{16}$ score: 99,79

Comparison to literature: OK
Aglycon Analysis: OK

UV Spectrum



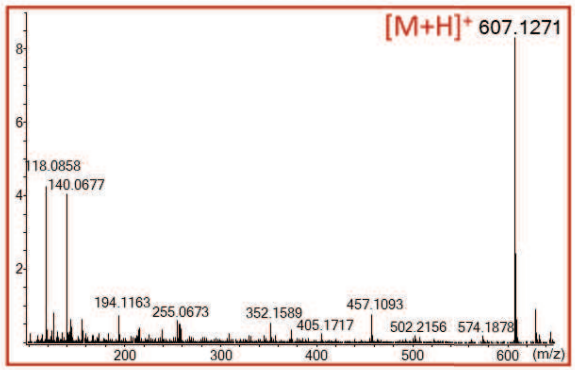
Aglycon: flavone, dihydroxyflavone type



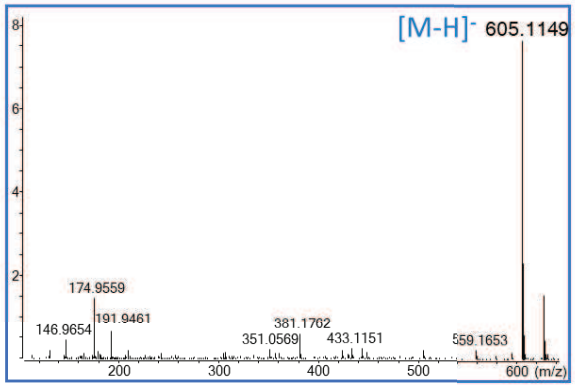
7,4'-dihydroxyflavone dihexuronide

MS Spectrum

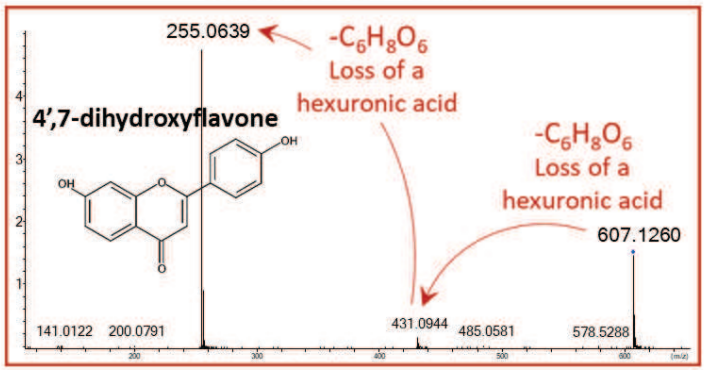
Positive

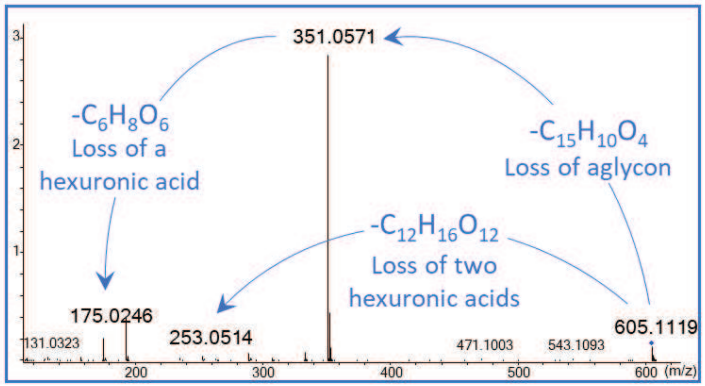


Negative



MS/MS Spectrum

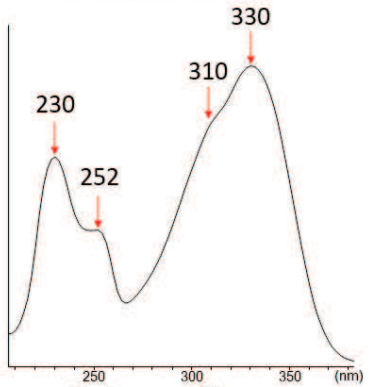




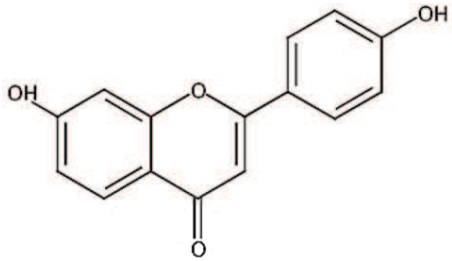
4’,7-dihydroxyflavone

MW = 254Da
Formula: $C_{15}H_{10}O_4$

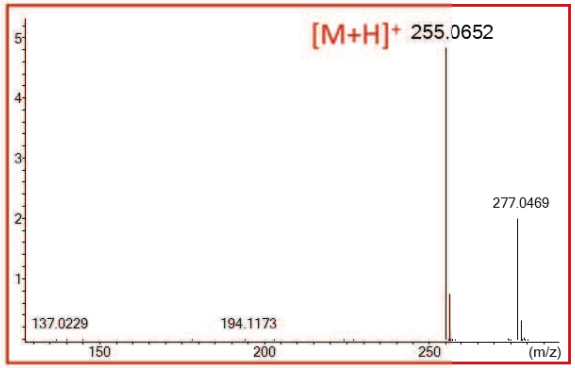
UV Spectrum

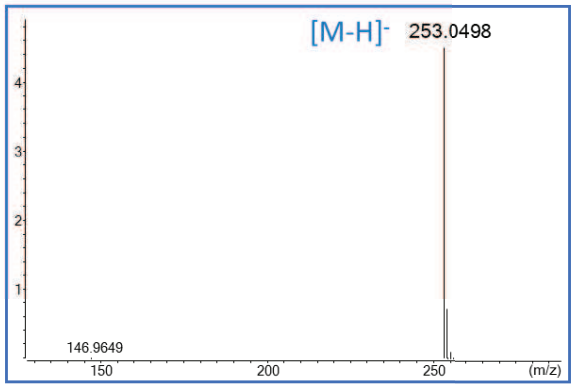


Aglycon: flavone

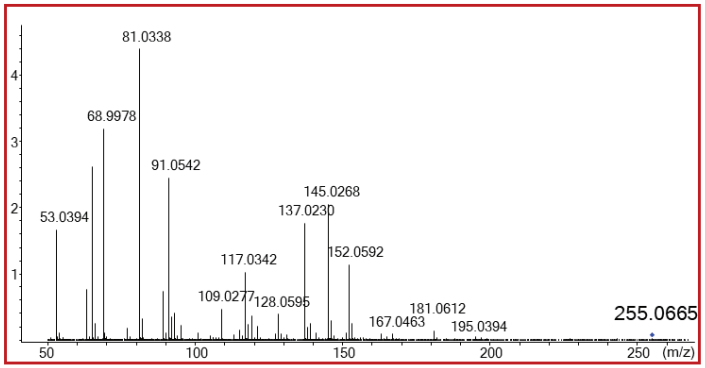


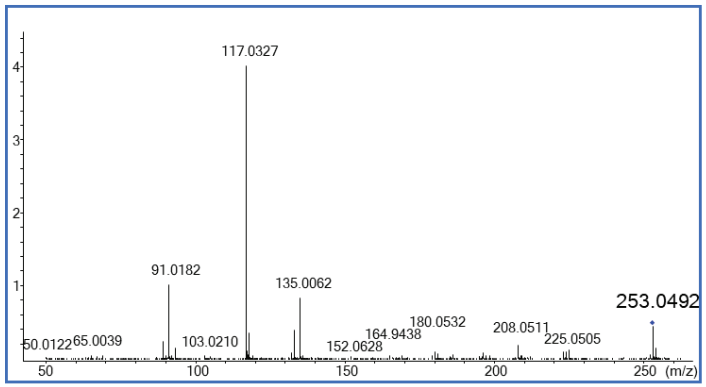
MS Spectrum





MS/MS Spectrum



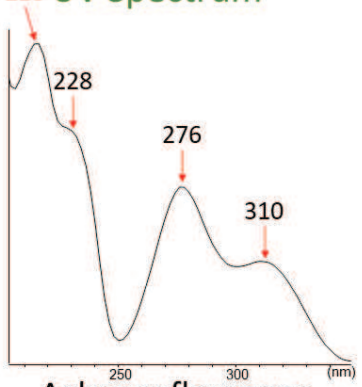


Composé X41

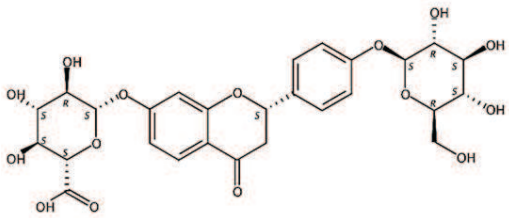
MW = 594Da
Formula: $C_{27}H_{30}O_{15}$ score: 99,3

Comparison to literature: OK
Aglycon Analysis: OK

216 UV Spectrum



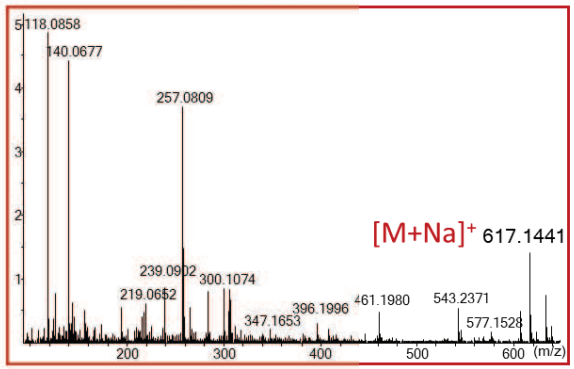
Aglycon: flavanone, Liquiritigenin type



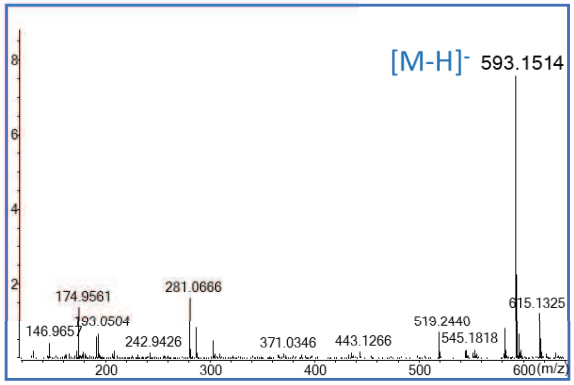
Liquiritin-7-O-hexuronide

MS Spectrum

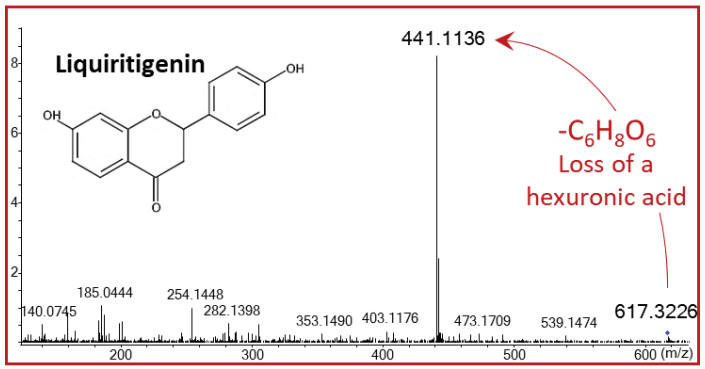
Positive

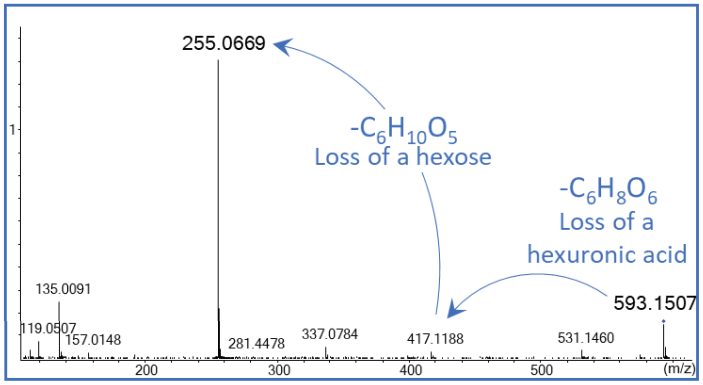


Negative



MS/MS Spectrum

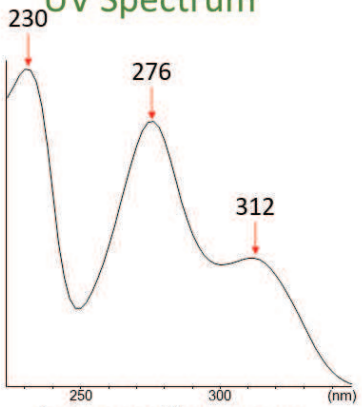




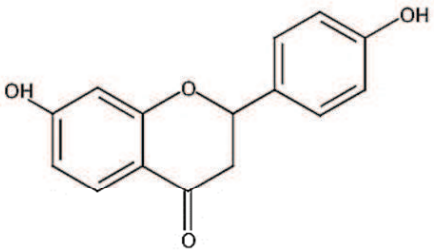
Liquiritigenin

MW = 256Da
Formula: $C_{15}H_{12}O_4$

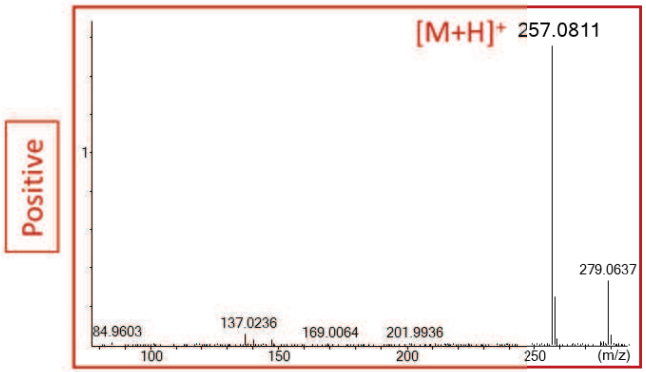
UV Spectrum

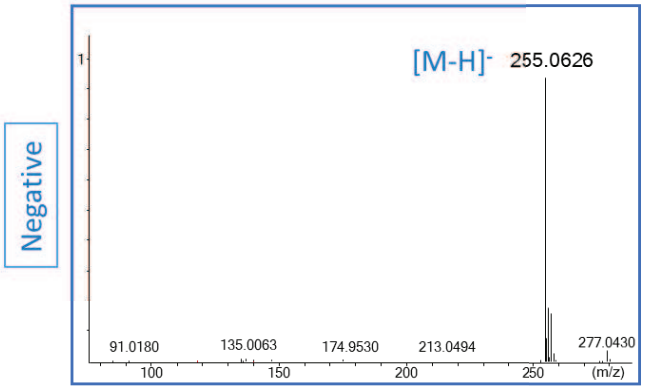


Aglycon: flavanone, Liquiritigenin type

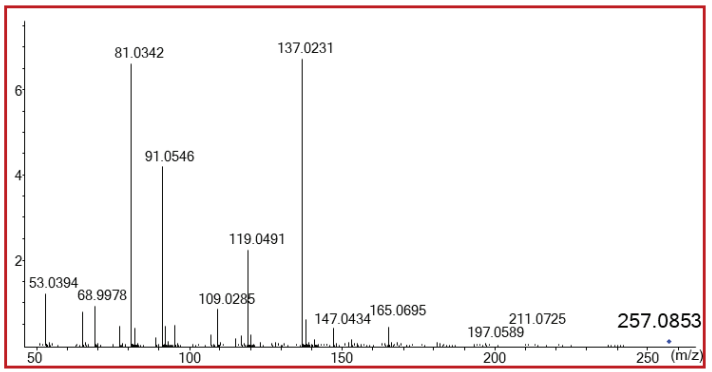


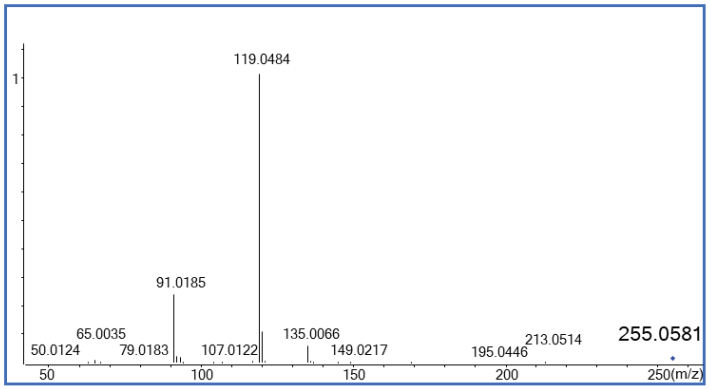
MS Spectrum





MS/MS Spectrum



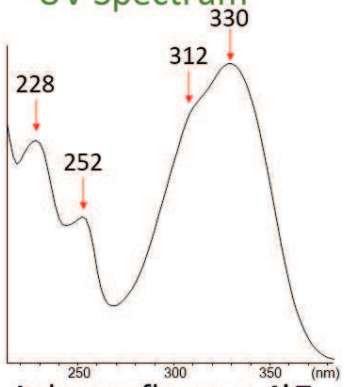


Composé X44

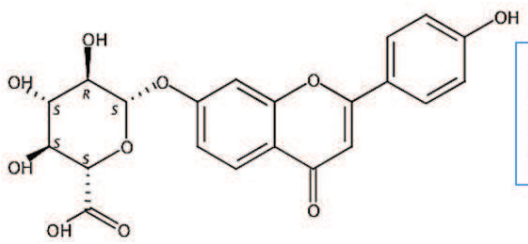
MW = 430Da
Formula: $C_{21}H_{18}O_{10}$ score: 95,27

Comparison to literature: OK
Aglycon Analysis: OK

UV Spectrum



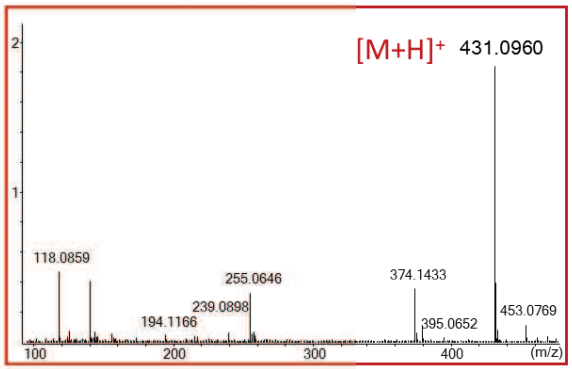
Aglycon: flavone, 4',7-dihydroxyflavone type



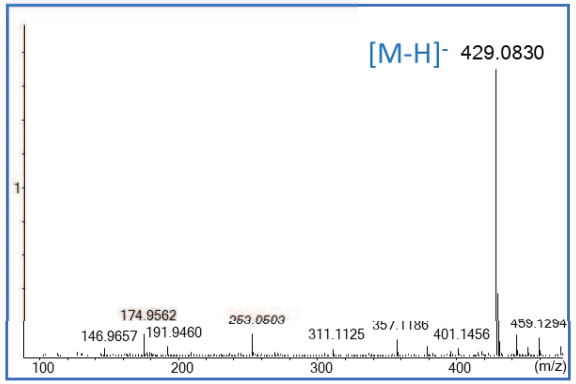
4',7-dihydroxyflavone-7-O-glucuronide

MS Spectrum

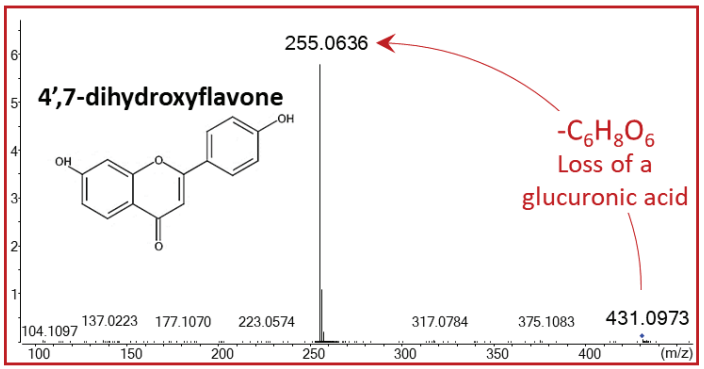
Positive

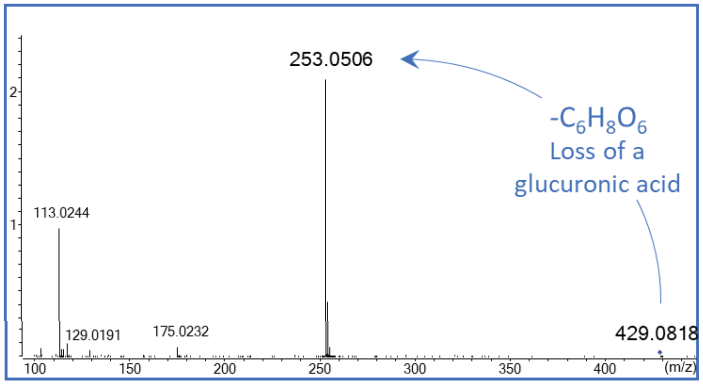


Negative



MS/MS Spectrum



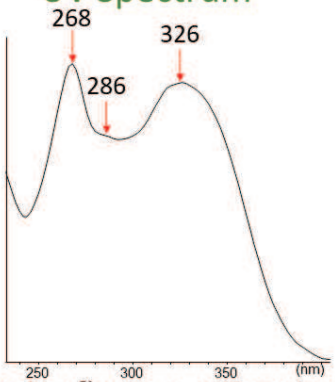


Composé X50

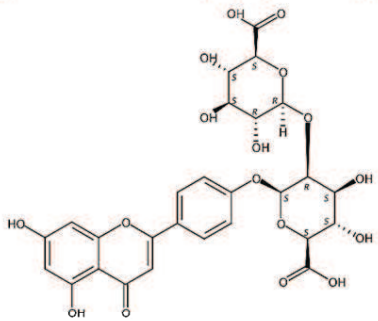
MW = 622Da
Formula: $C_{27}H_{26}O_{17}$ score: 94,14

Comparison to literature: OK
Aglycon Analysis: OK

UV Spectrum



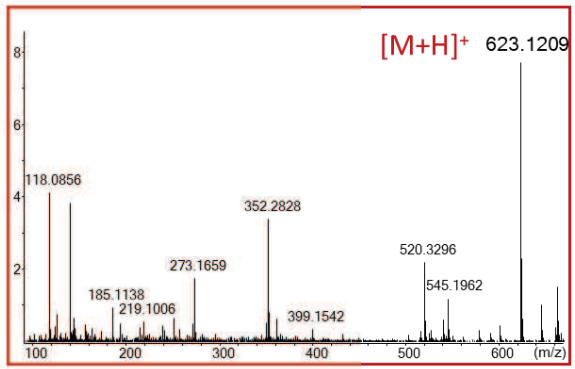
Aglycon: flavone, apigenin type



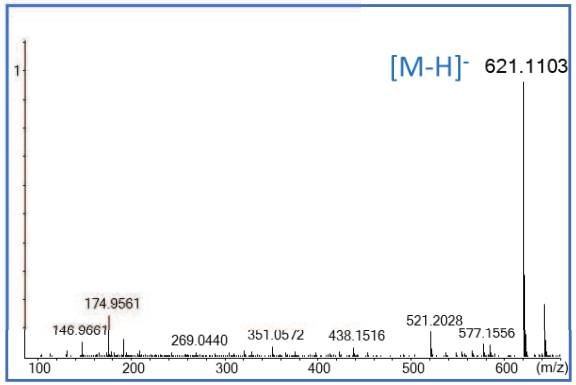
Apigenin 4'-dihexuronide

MS Spectrum

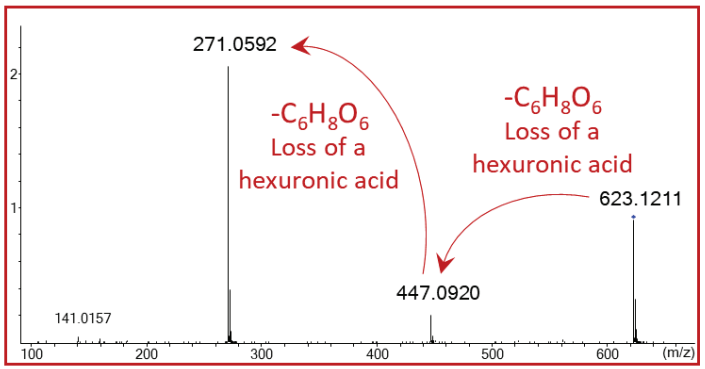
Positive

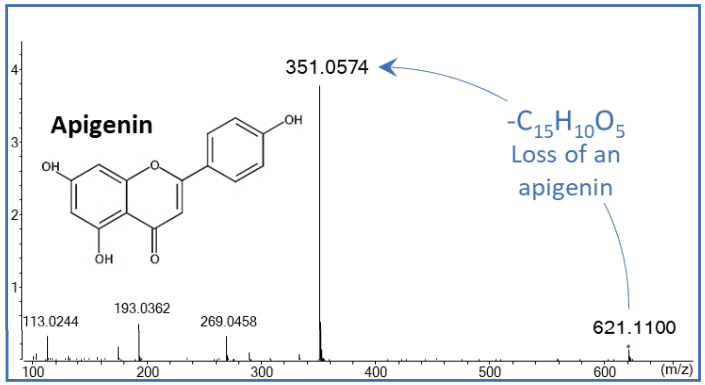


Negative



MS/MS Spectrum



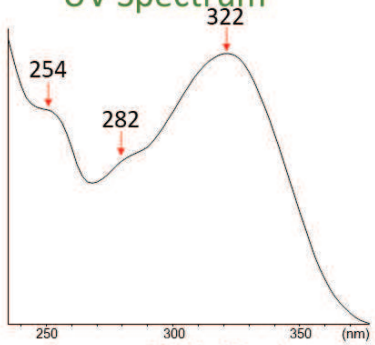


Composé X65

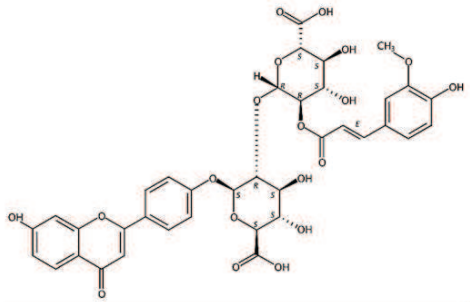
MW = 782Da
Formula: $C_{37}H_{33}O_{19}$ score:

Comparison to literature: OK
Genin Analysis : OK

UV Spectrum



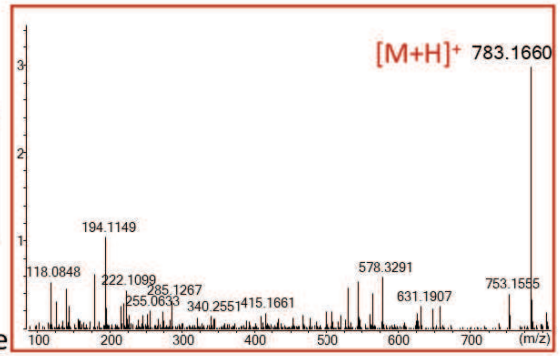
Genine: flavone, 7,4'-dihydroxyflavone type



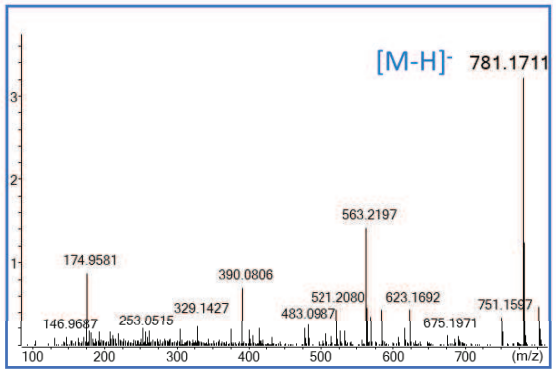
7,4'-dihydroxyflavone-O-[feruloyl-hexronyl-O-hexuronide]

MS Spectrum

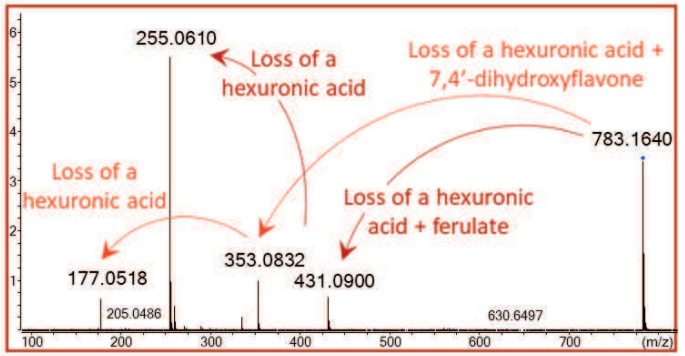
Positive

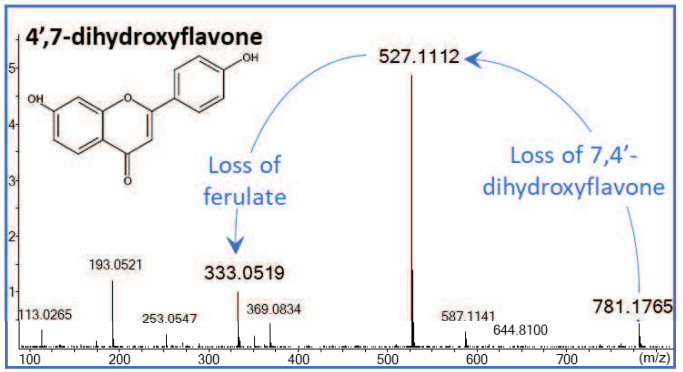


Negative



MS/MS Spectrum



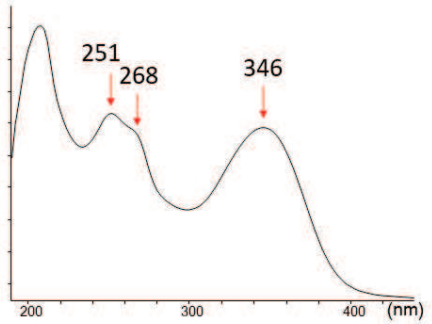


Compound X66

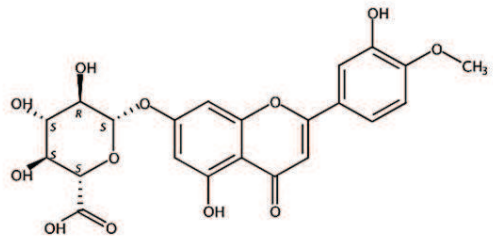
MW = 476Da
Formula: $C_{22}H_{20}O_{12}$ score: 98,7

Comparison to literature: OK
Aglycon Analysis: OK

UV Spectrum



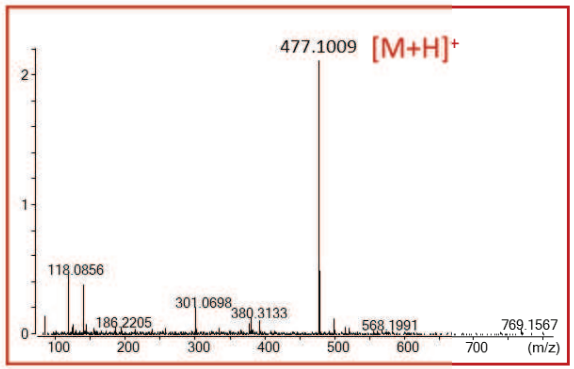
Aglycon: flavone, luteolin type



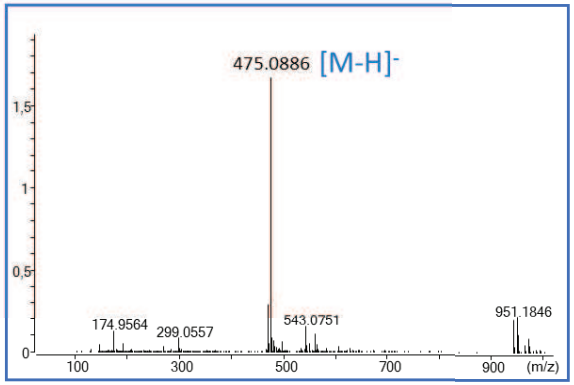
Diosmetin 7-O-hexuronide

MS Spectrum

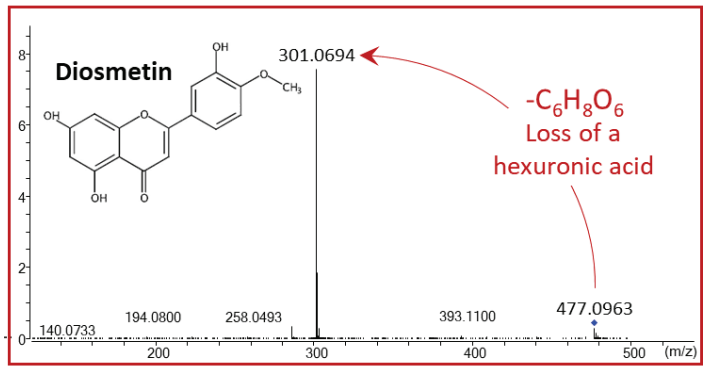
Positive

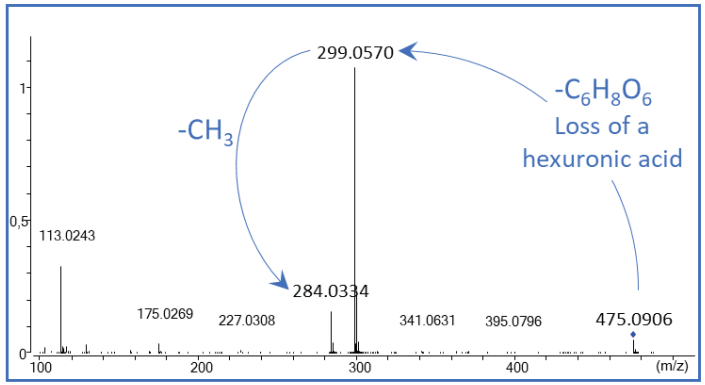


Negative



MS/MS Spectrum

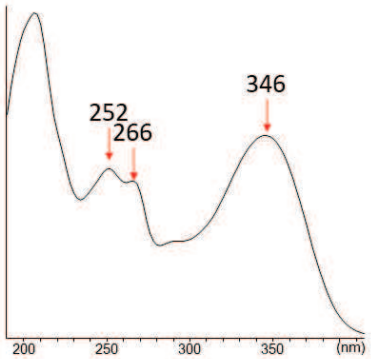




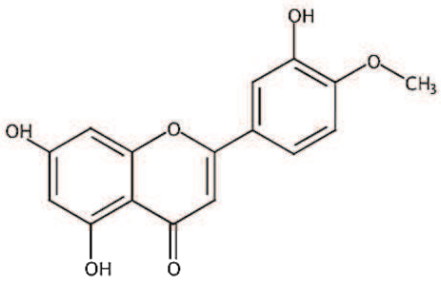
Diosmetin

MW = 300Da
Formula: $C_{16}H_{12}O_6$

UV Spectrum

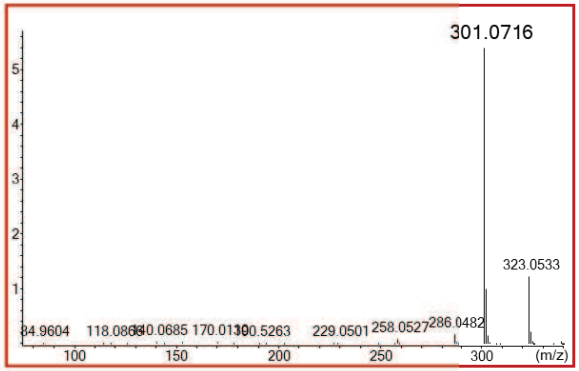


Aglycon: flavone, luteolin type

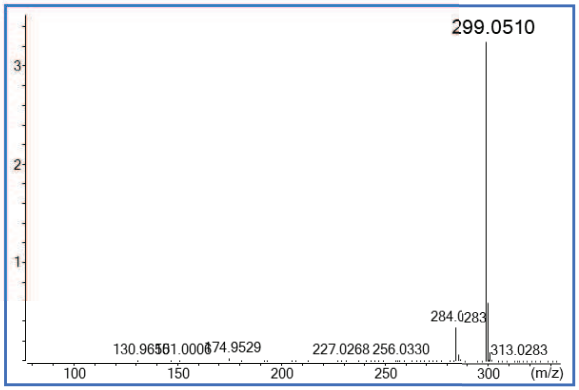


MS Spectrum

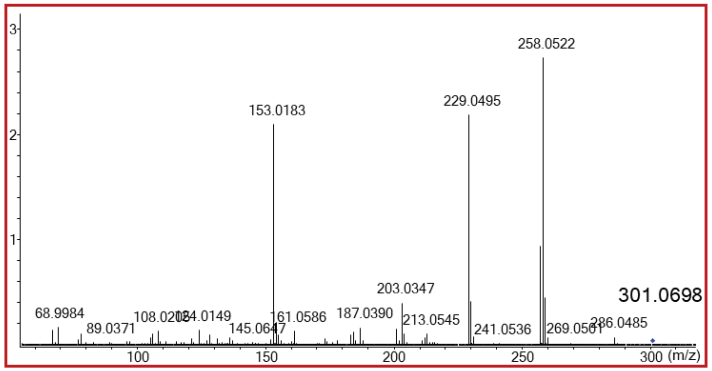
Positive

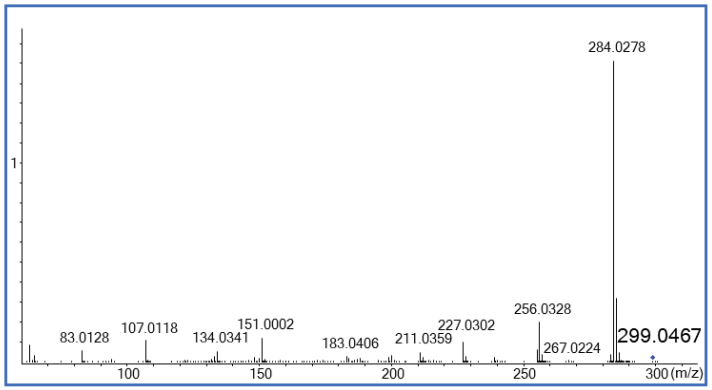


Negative



MS/MS Spectrum



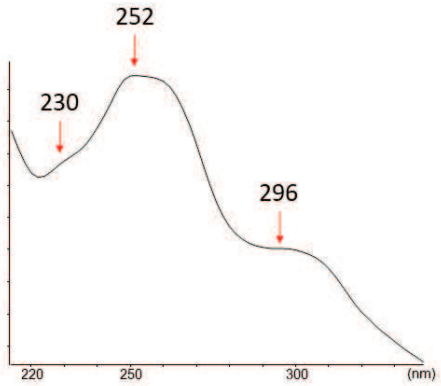


Composé X71

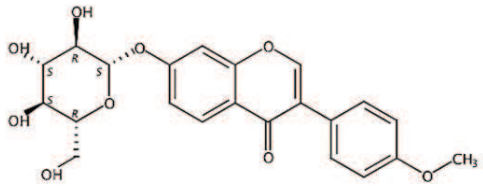
MW = 430Da
Formula: $C_{22}H_{22}O_9$ score:

Comparison to literature: OK
Standard Analysis: OK

UV Spectrum



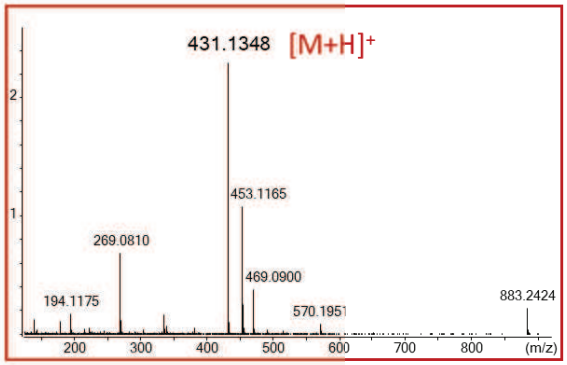
Aglycon: flavone, formononetin type



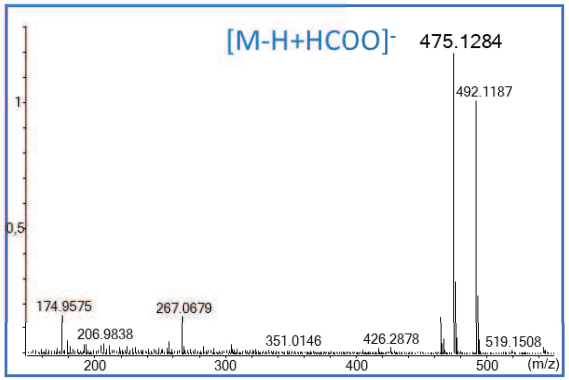
Formononetin-7-O-glucoside = Ononin

MS Spectrum

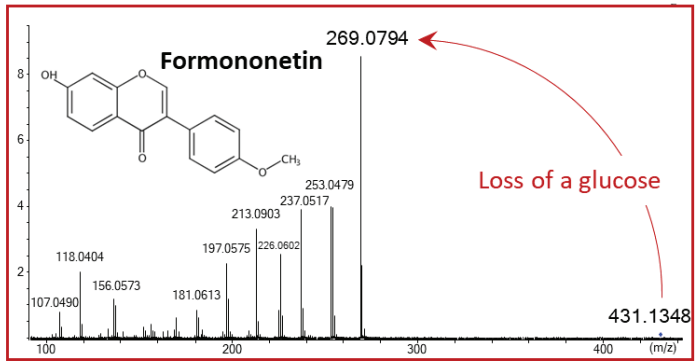
Positive

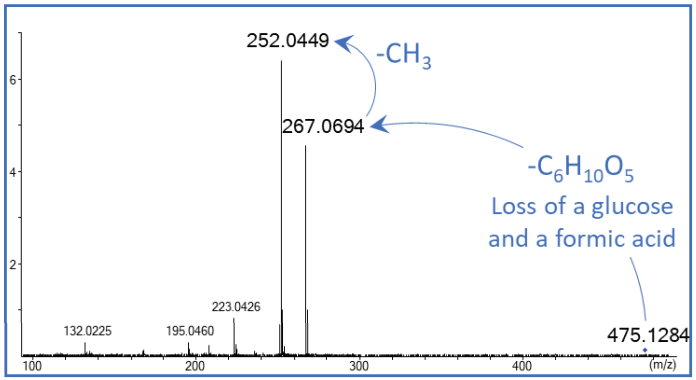


Negative



MS/MS Spectrum

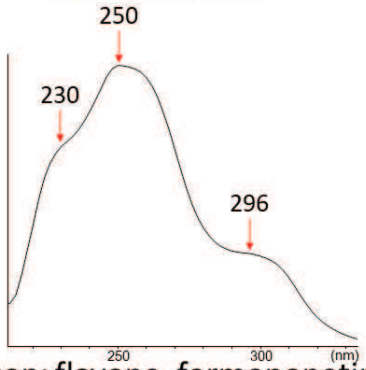




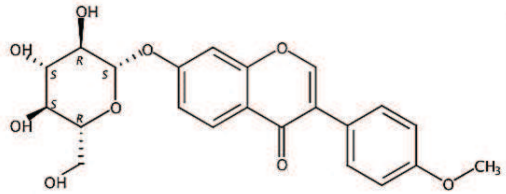
Ononin

MW = 430Da
Formula: $C_{22}H_{22}O_9$

UV Spectrum

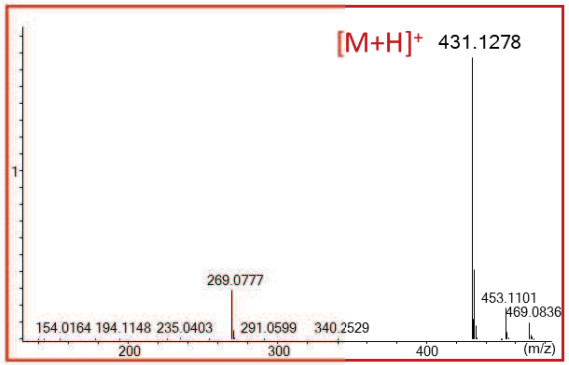


Aglycon: flavone, formononetin type

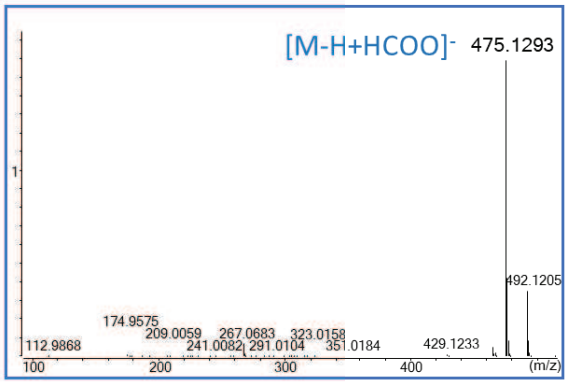


MS Spectrum

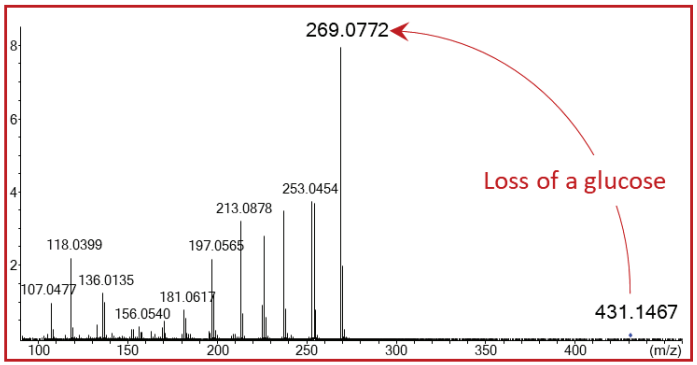
Positive

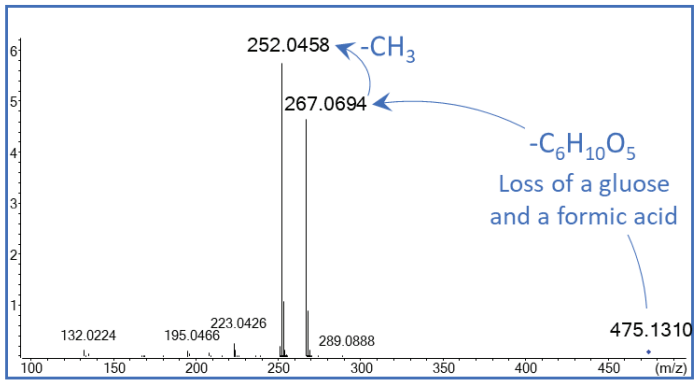


Negative



MS/MS Spectrum



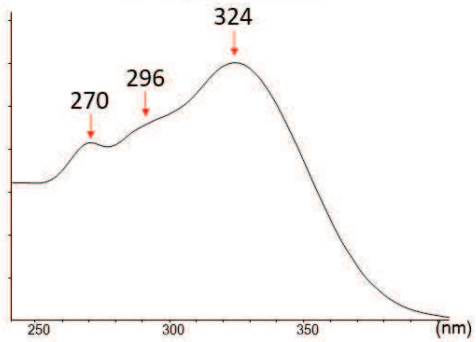


Composé X79 → 6 Regions

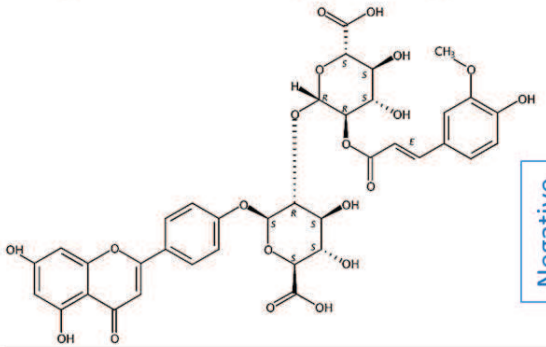
MW = 798Da
Formula: $C_{37}H_{34}O_{20}$ score:

Comparison to literature: OK
Aglycon Analysis: OK

UV Spectrum



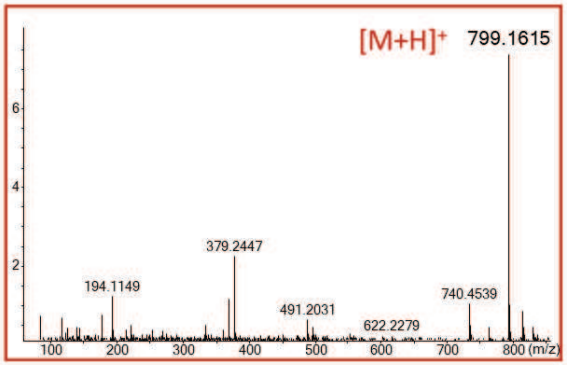
Aglycon: flavone, apigenin type



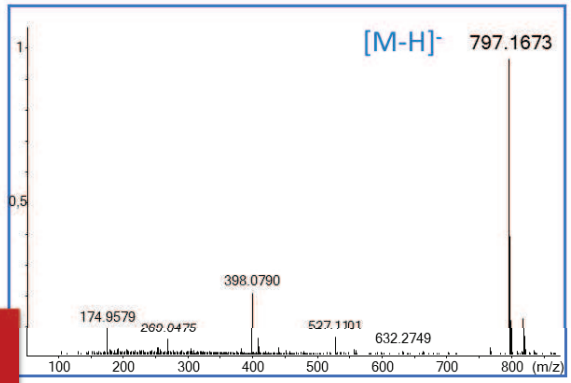
Apigenin 7/4'-O-[feruloyl-glucuronopyranosyl-O-glucuronopyranoside]

MS Spectrum

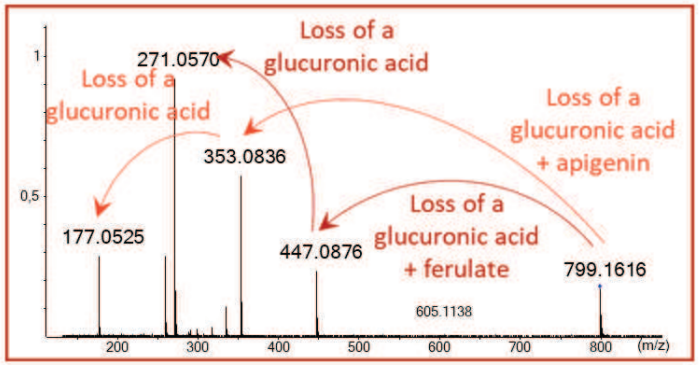
Positive

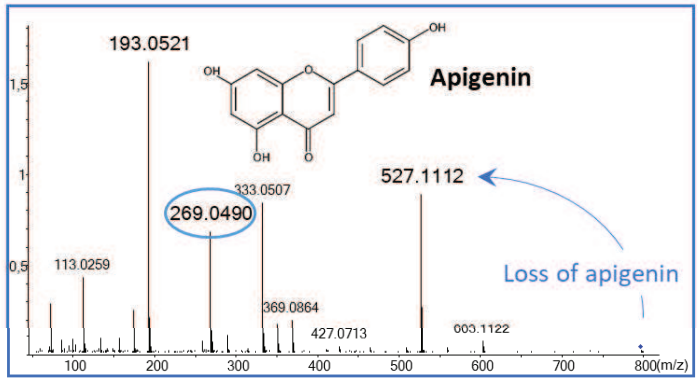


Negative



MS/MS Spectrum

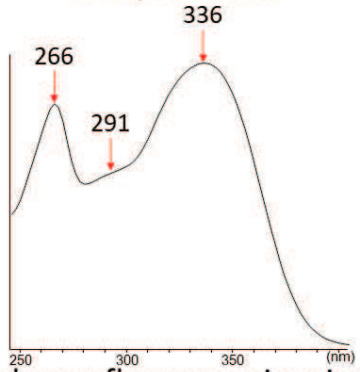




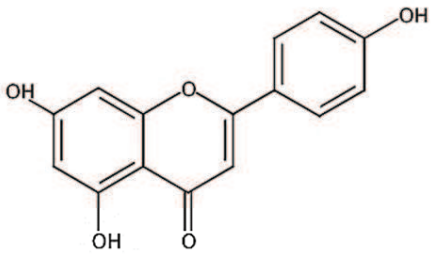
Apigenin

MW = 270Da
Formula: $C_{15}H_{10}O_5$

UV Spectrum

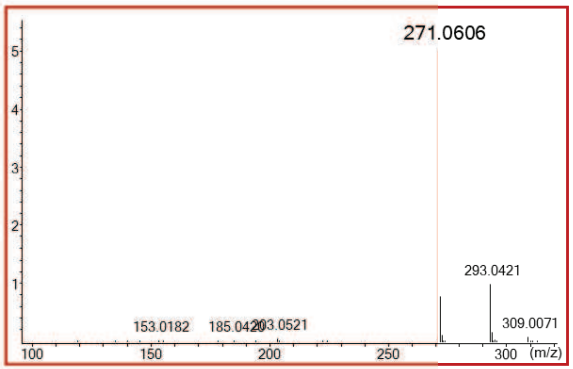


Aglycon: flavone, apigenin type

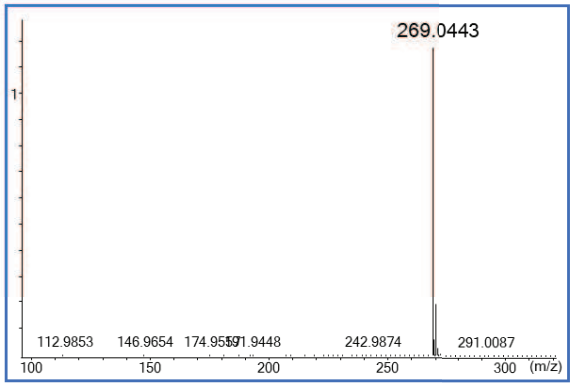


MS Spectrum

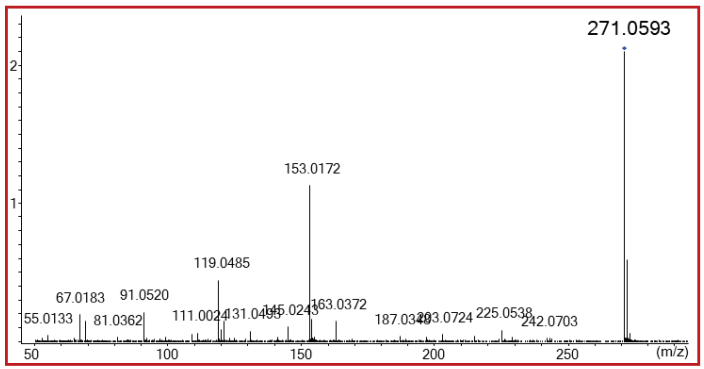
Positive

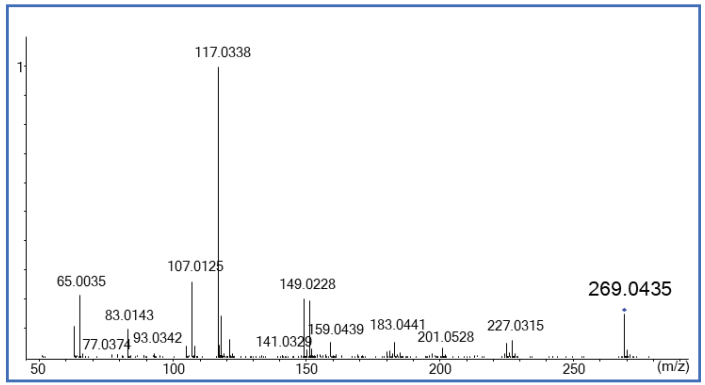


Negative



MS/MS Spectrum



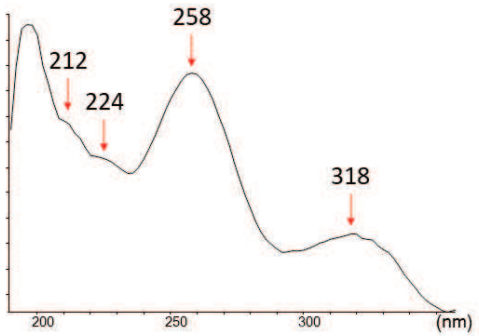


Composé X87

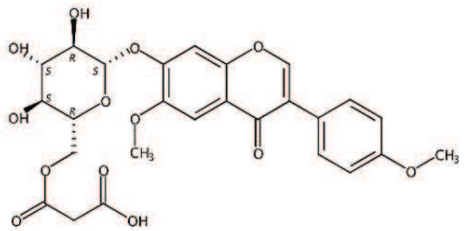
MW = 546Da
Formula: $C_{26}H_{26}O_{13}$ score:

Comparison to literature: OK
Aglycon Analysis :

UV Spectrum



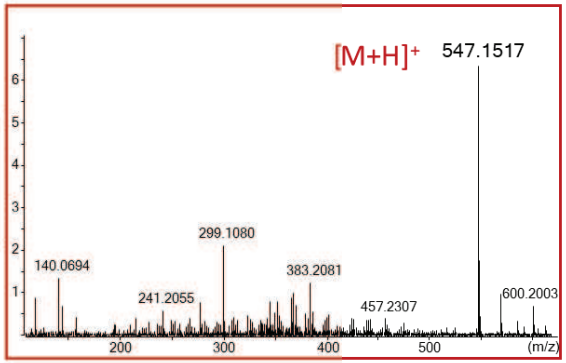
Aglycon: isoflavone, afrormosin type



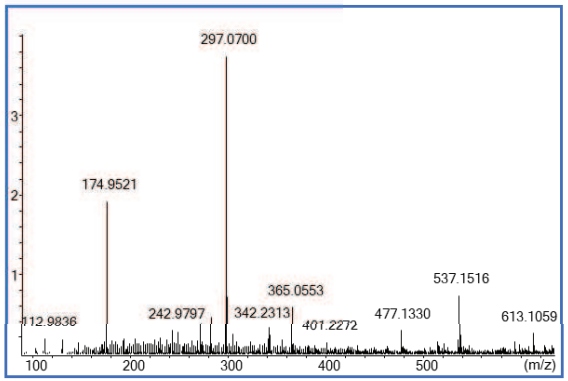
Afrormosin-7-O-glucoside-6'-O-malonate

MS Spectrum

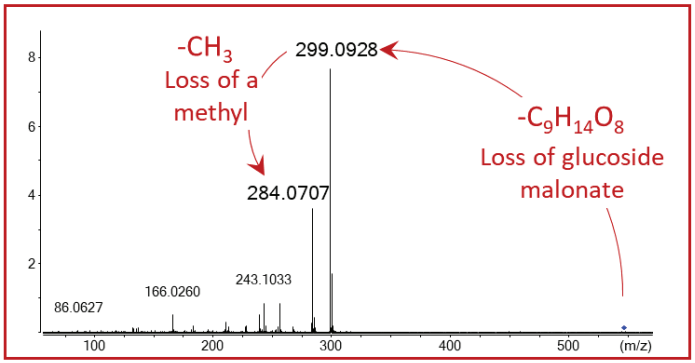
Positive



Negative



MS/MS Spectrum



Afrormosin