UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
SCIENZE AGRARIE E AMBIENTALI

RESPONSE OF THE SYMBIONT *ENSIFER MELILOTI* TO THE PLANT FLAVONOID LUTEOLIN: NOT ONLY SYMBIOSIS

GIULIA SPINI

Anni 2012/2015

Table of Contents

Chapter 1

Introduction

Chapter 1

Introduction

1.1. The biological nitrogen fixation and the agronomic relevance of the rhizobium-legume symbiosis

Nitrogen is one of the most abundant elements on Earth and it is also one of the most limiting nutrient for plant growth because it is predominately found as atmospheric nitrogen (N_2), which is a chemically inert form. Nitrogen cycle is defined as a gaseous biogeochemical cycle (Figure 1.1) because the reserve pool of this chemical element is precisely the atmosphere, within which nitrogen occupies about 78% of the total volume [1]. The importance of the nitrogen cycle for living organisms is due to their need to assimilate nitrogen for the biosynthesis of essential organic compounds, such as proteins and nucleic acids. The atmospheric nitrogen does not constitute an available form and cannot be directly absorbed by organisms, except for nitrogen-fixing microorganisms thus representing a limiting factor for development of aquatic, as well as terrestrial ecosystems [2]. In agriculture the nitrogen is supplied by the use of various industrial fertilizers rich in nitrogen to achieve maximum productivity [3]. The production of nitrogenous fertilizers required a large amount of energy and fossil fuel, which is costly and consumes many natural resources. Furthermore, the carbon dioxide (CO_2) released during the process of combustion of fossil fuels and the nitric oxide released during the decomposition of fertilizers contribute to the increased greenhouse effect. The use of fertilizers has also resulted in increasing the risk of unacceptable levels of water pollution and the eutrophication of lakes and rivers [4].

International emphasis on environmentally sustainable development with the use of renewable resources is likely to focus attention on the potential role of biological nitrogen fixation (BNF) in supplying nitrogen for agriculture. Indeed, the largest input of nitrogen in the biosphere comes from the biological fixation of atmospheric nitrogen that is estimated to account for over half of the nitrogen fixed annually in terrestrial environments [5]. The BNF is a process by which chemically inert N_2 present in the atmosphere is enzymatically reduced to the metabolically usable form ammonia (NH_3) through the action of nitrogenase enzyme [5]. The ability to catalyze the conversion of N_2 to NH_3 has evolved only among prokaryotes. These prokaryotes include some strains of *Archea*, aquatic microorganisms, such as cyanobacteria, free-living soil bacteria, such as *Azotobacter, Bacillus, Clostridium and Klebsiella,* bacteria that form associative relationships with plants, such as *Azospirillum*, and most importantly, bacteria, such as *Rhizobia* and *Frankia*, that form symbioses with legumes and/or not legumes, respectively [6].

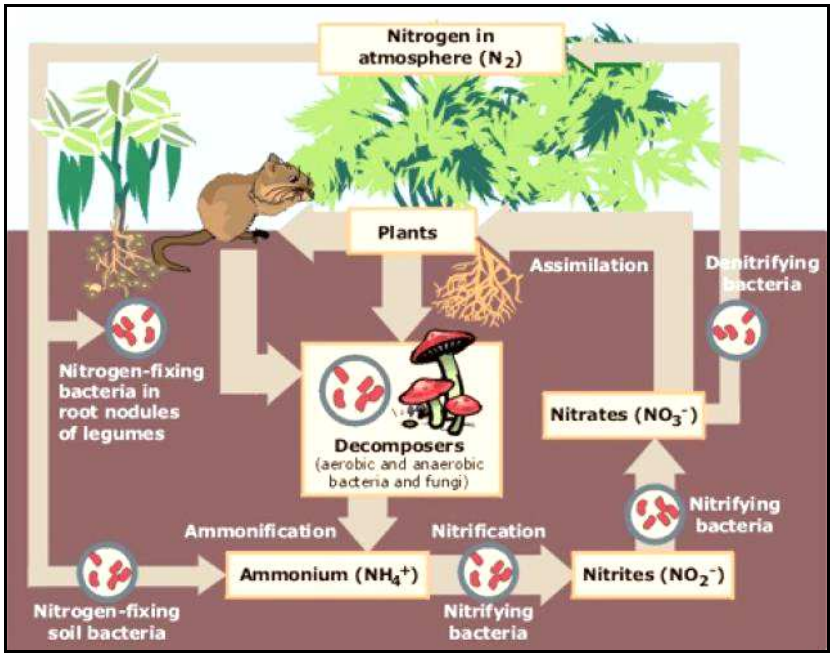


Figure 1.1. The nitrogen cycle
(modified by http://www.epa.gov/maia/html/nitrogen.html-Environmental Protection Agency)

Rhizobia, a group of Gram-negative soil bacteria, play a role of particular interest within the context of an intimate symbiosis established with legume plants. Symbiosis is defined as an intimate relationship between different biological species that are interdependent and gain reciprocal benefit [7]. Symbiotic interactions of microorganisms are widespread in nature, and support fundamentally important processes in several areas of biology that range from health and disease to agriculture and environment. Since plants and microorganisms coexisted for millions of years, they have evolved sophisticated strategies to perceive the presence of each other and respond appropriately. Plants release in the rhizosphere a series of molecules, which are recognized as signals by the microorganisms, inducing reciprocal responses [8]. Communications between plants and microorganisms are extremely complex and heterogeneous. Plants are able to recognize microbial-derived compounds and adjust their response according to the type of microorganism encountered. In some cases, the signal molecules exchanged can act as elicitors of defensive responses protecting plants against harmful organisms such as phytopathogenic fungi, bacteria, viruses, and nematodes. In other cases, plant-microbe interactions can be beneficial as in the context of the symbiosis between legumes and nitrogen-fixing bacteria (rhizobia) [9].

The rhizobium-legume symbiosis occurs in nitrogen limiting conditions through a multistep process in which, following an initial chemical signal released by the plant,

represented by flavonoids, rhizobia as response excreted the Nod Factor (NF). NF stimulates the entry of rhizobial cells into the roots, induces the formation of specialized structures, called root nodules (Figure 1.2) and the differentiation of rhizobia within the nodule into the bacteroid form. Bacteroids express genes encoding nitrogenase, which is the enzymatic complex responsible for atmospheric nitrogen fixation into ammonia. It is estimated that the rhizobium-legume symbiosis contributes to about half of the total biological fixation-nitrogen in the biosphere [10]. Amounts of N_2 fixed by the crop rhizobium-legume symbiosis were valued about 21 Tg annually and by the forage and fodder rhizobium-legume symbioses were valued 12-25 Tg annually. The symbiotic nitrogen-fixation promoted by rhizobia enhances the growth of legume plants, increases crop yield and reduces the dependence on chemical fertilizers [11]. The advantages granted by the symbiosis association endow the legume hosts with special significance among agricultural plants: their productivity is theoretically independent of soil nitrogen status and fertilizer application and they provide important grain and forage crops, both in temperate and in tropical zones.

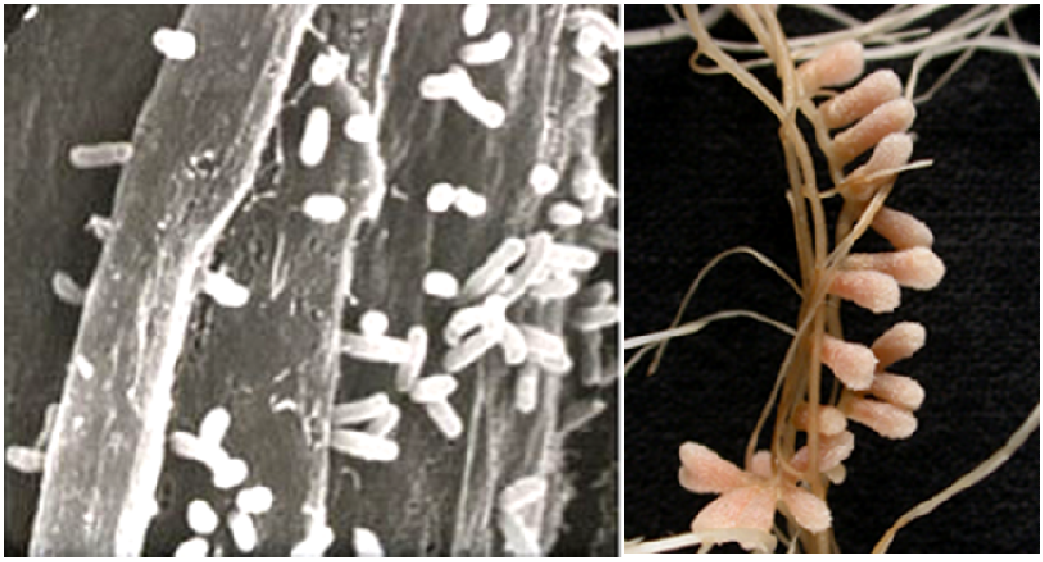


Figure 1.2. Symbiotic association between nitrogen-fixing rhizobia and legume plants. On the left, a magnification by scanning electron microscope of rhizobial cells colonizing legume root hairs is reported. On the right, root nodules containing rhizobia are shown [modified by bioinfo.bact.wisc.edu/themicrobialworld/Effects.html].

The host plants are not the only ones to benefit by the symbiotic interaction, because in exchange rhizobia receive nutrients from the host plant, such as sugars as well as protection within the nodule structure [12]. Nevertheless, rhizobia that not fixed nitrogen efficiently received the same nutrients from the host plant. In this case, rhizobia can be

considered as parasites rather than symbionts. Indeed, there are metabolic sanctions that plants can apply to non-efficient nodules, to limit the development of rhizobial strains which do not fix nitrogen efficiently [13].

The symbiosis of greatest importance both economically, agriculturally, basic and applied research involves the nitrogen-fixing bacterium *Ensifer meliloti* and compatible host legume [4] (Figure 1.3). Leguminous plant are the main crop commonly used as forage for cattle and in the crop rotation practice to supply soil with organic nitrogen. Moreover, the legume plants in phytoremediation could be used to capture and remove toxic compounds from contaminated soils and groundwater [11]. In the context of the rational use of resources, leguminous crops in symbiosis with *E. meliloti* provide considerable advantages. Nodulated plants, thanks to the nitrogen-fixing symbiosis with the bacterium *E. meliloti*, have higher yield than not nodulated and fertilized ones of the same species and have the ability to growth in nitrogen poor soil as well as in marginal areas where other more demanding plants, such as cereals, cannot be cultivated. The entire ecosystem can benefit of nitrogen fixed by rhizobia, which enter in the trophic network through the flow of elements between organisms [14]. The *E. meliloti* is significant also from an economic point of view. Indeed, the biological nitrogen fixation due to the *E. meliloti*-legume symbiosis is estimated to provide 90 million tons per year of assimilable-nitrogen worldwide, thus saving annually around $ 200 million in nitrogen chemical fertilizers. The economic value of leguminous crops in the U.S. is estimated about $ 8,1 billion per year [15]. For these reasons, the biological nitrogen fixation has been extensively exploited in agriculture for practical applications designed to improve the yield of leguminous crops used for human nutrition (beans, peas, peanuts, soybeans) and as forage plants as well as to accomplish other important functions such balancing the different components of the agro-system and maintaining the soil fertility.

The study of the nitrogen-fixing symbiosis between rhizobia and plants is one of the greatest contributions of the microbiology to agricultural applications aimed to improve growth of leguminous crops, their environmental sustainability and their cultivation as fodder plants, crops for bio-energy, to recover low fertility, degraded and contaminated areas [16]. An increasing interest of the research is addressed to the potential biotechnological exploitation of nitrogen-fixing rhizobia in symbiosis with legume plants, as urged by the Food and Agriculture Organization (FAO). The United Nations agency has expressed a position of interest in the innovative approaches and tools able to improve agricultural yields without impairing the input budget required for the production process (energy, water, soil). Trials sponsored by FAO are currently underway in several developing countries (www.fao.org/biotech/logs/) in the context of a "sustainable intensification and innovation" of agriculture. The employment of nitrogen-fixing rhizobia

targeted to optimize plant productivity and the process of plants growth, especially in arid and marginal areas, is a part of such trials. A concrete example of the increasing interest about the potential applications of nitrogen-fixing rhizobia is represented by foundation of biotechnology realities, such as Agradis in 2012 (http://www.agradis.com/, www.sribio.com and www.waterlooenvironmentalbiotechnology.com). Agradis is a biotechnology company to improve the sustainability and productivity of agriculture exploiting the interactions of plant species with beneficial microorganisms of the rhizophere.

1.2. General taxonomic features of rhizobia

The rhizobia are soil Gram-negative bacteria belonging to the rhizosphere microbial community, the region of soil characterized by the presence of living plant roots and closely associated soil. The word rhizobia comes from ancient Greek "rhiza" meaning "root" and "bios" meaning "life". Rhizobia are able to fix atmospheric nitrogen (N_2) through the formation of a symbiotic relationship with their host legume plants, belonging to the *Leguminosae* (*Fabaceae*) family. Although rhizobia require a plant host to fix nitrogen, they can survive in soil over periods of several years even in the absence of their legume hosts. The taxonomy and nomenclature of the rhizobia are the subjects of much debate and controversy. According to the latest version of taxonomy, rhizobia are divided into 13 genera, for a total of 76 species [17]. Rhizobia form a group that falls into two classes of the proteobacteria, α-proteobacteria and *β*-proteobacteria [18]. Most rhizobia belong to the *Bradyrhizobium, Mesorhizobium, Rhizobium* and *Ensifer* genera of the *α*-proteobacteria (order *Rhizobiales*) and are closely related to nonsymbiotic soil bacteria (Figure 1.3). Symbiotic nitrogen-fixing *β*-proteobacteria (*β*-rhizobia) also have been reported, and the evolution of other rhizobial species is attributed to the horizontal transfer of symbiotic genes into different types of bacteria [19].

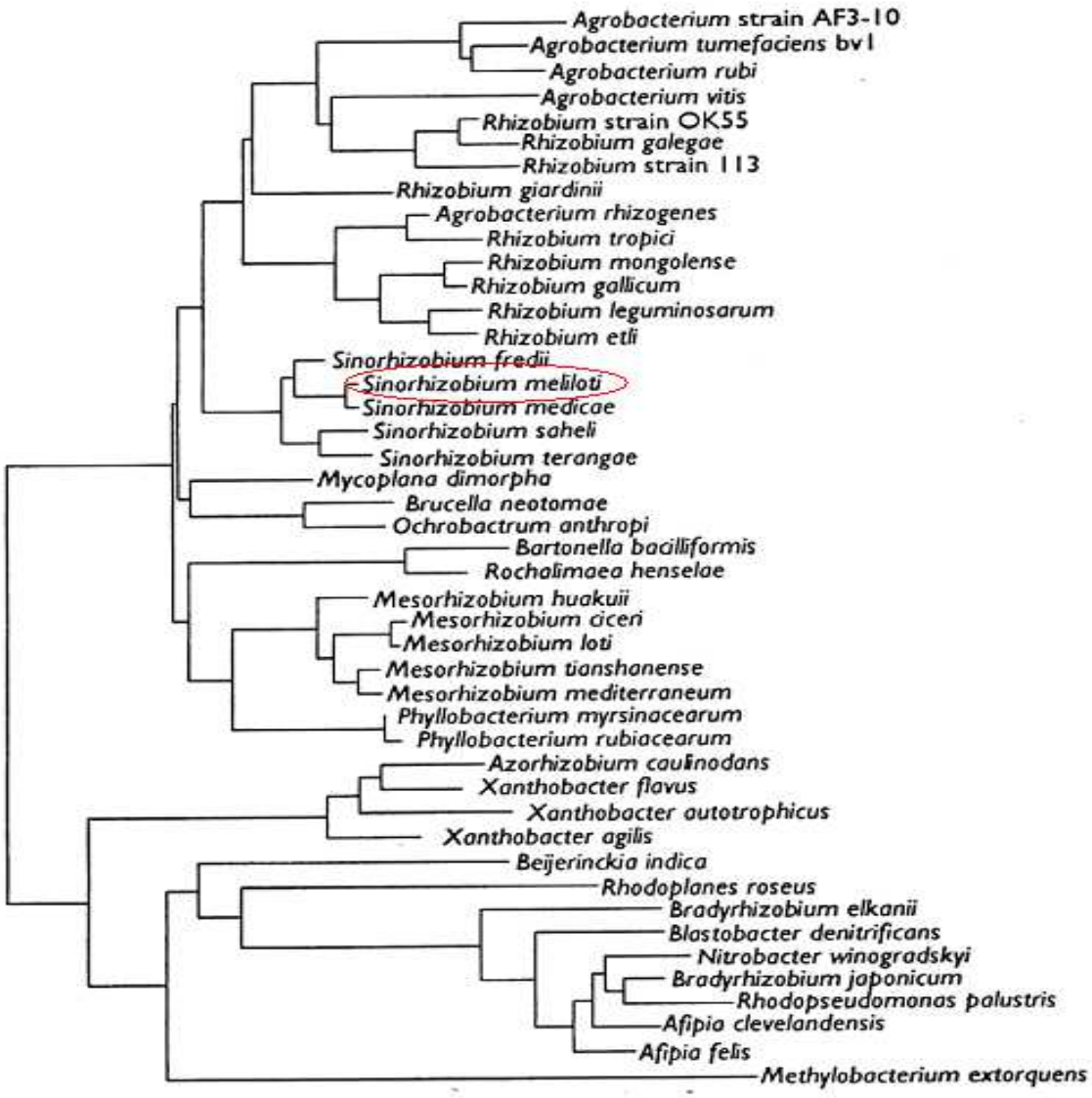


Figure 1.3. Phylogenetic tree that shows the main species belonging to the *α*-proteobacteria, based on homology of 16S ribosomal RNA sequence (modified by [20]). The *Ensifer meliloti* (*Sinorhizobium meliloti)* species, which includes the strains used in this thesis, is highlighted by a red circle.

1.2.1. Biology and genomics of the bacterial model organism *Ensifer meliloti*

Ensifer meliloti, a Gram-negative nitrogen-fixing proteobacterium that is distributed worldwide in temperate soils both in free-living and symbiotic form, is considered a model bacterium for the study of the rhizobium-legume symbiosis. *E. meliloti* specifically establishes symbiosis with species belonging to three genera of leguminous plants (*Melilotus*, *Medicago*, *Trigonella*) [21]. The infection of the host plant roots by *E. meliloti* induced the formation of specialized organs, the nodules within which bacteria carry out the process of nitrogen fixation. The genome of *E. meliloti* has been fully sequenced in

2001 [22] providing a solid basis for several genetic and molecular studies concerning the rhizobium-plant interaction and the bacterial response to environmental stimuli.
E. meliloti genome (6,69 Mb) contains 6204 genes predicted to encode proteins and distributed on three circular replication units: a chromosome (3,65 Mb) and two megaplasmids, pSymA (≈1,35 Mb) and pSymB (≈1,68 Mb) (Figure 1.4).

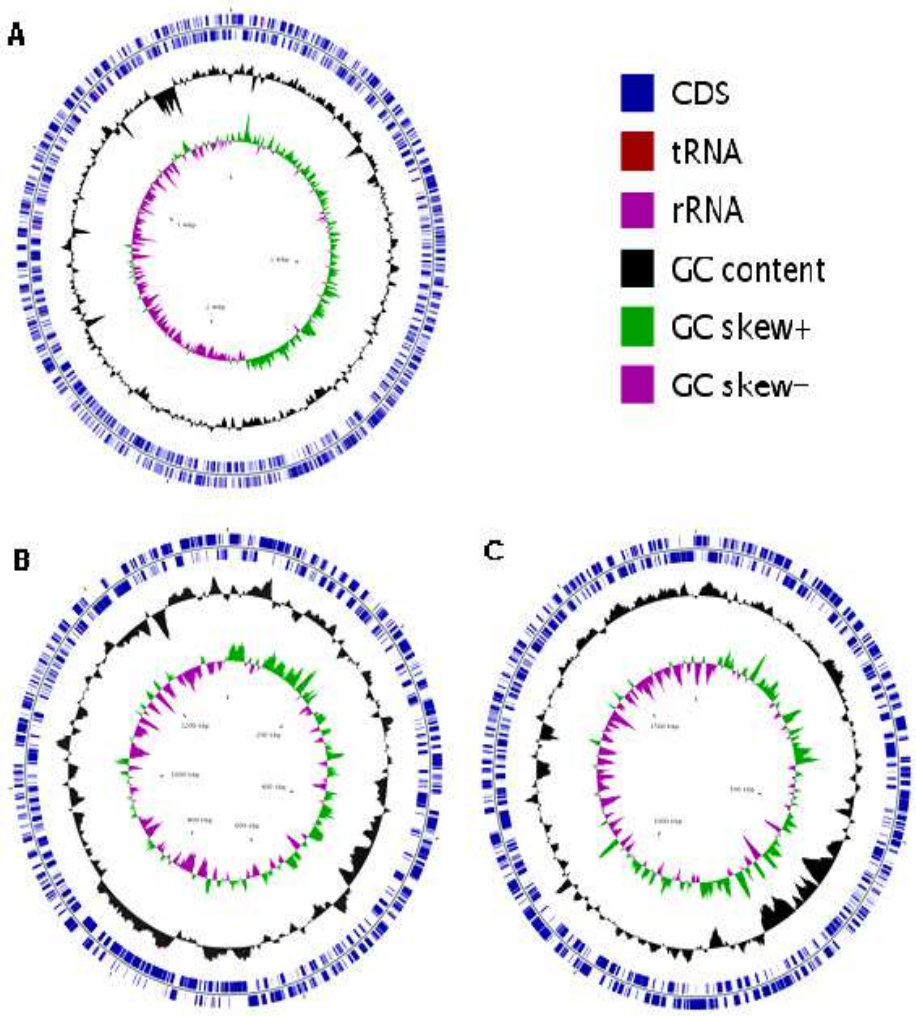


Figure 1.4. The genome of *Ensifer meliloti*. A) Chromosome (3,65 Mb); B) Megaplasmid pSymA (1,35 Mb); C) Megaplasmid pSymB (1,68 Mb) [modified by http://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi].

At the time of *E. meliloti* genome sequence determination, 40% of the genes on the chromosome could not be placed into a functional category. Moreover, 8% were orphan genes, defined as those not found in any other sequenced genome. Becker *et al*. [23] published a *E. meliloti* genome annotation update that incorporates information published from 2001 to 2008. The improved prediction tools allowed to identify 86 new

putative genes, to remove 66 previously predicted orphan genes and to adjust the start positions of 360 coding regions. As a result, more than 71% of genes have now a predicted function. The chromosomal genes include all the major housekeeping functions involved in essential processes of *E. meliloti* central metabolism. Accessory genes, usually encoding for proteins implicated in secondary metabolic pathways, are located on the megaplasmids pSymA and PsymB. The megaplasmid pSymA carries the most genes required for nodulation and nitrogen fixation (*nod*, *nif*, and *fix* genes), carbon metabolism, transport and stress responses, whereas pSymB reveals a high number of genes involved in exopolysaccharide biosynthesis. Both megaplasmids carry genetic loci for coniugation processes, such as *tra* and *tra2*, as well as *mob* and *oriT* loci.

TABLE 1

General genomic features of *Ensifer meliloti*

	Chromosome	**pSymA**	**pSymB**	**Genome**
Lenght (bp)	3,654,135	1,354,226	1,683,333	6,691,694
G+C Content (%)	62.7	60.4	62.4	62.1
Coding (%)	85.8	83.2	88.6	85.9
tRNAs	51	2	1	54
tmRNA*	1	0	0	1
Ribosomal RNA	3	0	0	3
Genes (ORFs)	3341	1293	1570	6204
Mean lenght of genes (bp)	938	871	950	927
Genes with annotated function	59%	56.5%	64.4%	59.7%
Genes orphans ** (% of total proteins-encoding genes)	5%	11.5%	12.3%	8.2%
Regulatory genes (% of total protein-encodng genes	7.2%	10.4%	10.5%	8.7%
Insertionalor phase sequences	2.2%	3.6%	0.9%	2.2%
Rhizobium-specific intergenic mosaic elements (RIME)	185	6	27	218
Palindromic sequences	253	0	5	258

* tmRNA are tRNA with two distinct domains: one functioning as tRNA and the other one functioning as mRNA.

** Genes with no homology with other sequenced genome

E. meliloti genome contains a high percentage of mobile genetic elements such as insertion sequences, mobile introns, trasposons, phage sequences, mosaic elements. The presence of ripetitive elements has been identified on the chromosome as palindromic sequences and RIME elements (*Rhizobium*-specific intergenic mosaic elements). Insertion and phage sequences has been mainly found on pSymA plasmid and on the chromosome (Table 1). Such genetic elements are associated to the genomic polimorphism of rhizobia and strongly contribute to the genetic diversity revealed within the *E. meliloti* populations. The most variable portion of the *E. melioti* genome is represented by the symbiotic megaplasmid pSymA containing genes required for host

nodulation and nitrogen-fixation [24]. Laboratory strains of *E. meliloti* as well as environmental strains that not contain the symbiotic plasmid pSymA or the nodulation genes are reported. The existence of such strains suggest that the genetic elements for nodulation and nitrogen-fixation could be acquired by processes of horizontal gene transfer. Most of the analysis of bacterial comparative genomic revealed large differences in genes content even between closely related strains leading to propose that non-essential genes are responsible for driving the evolutionary diversification between bacterial strains [25]. Based on such evidences the concept of "Pangenome" has been developed to describe the genome of a bacterial species. Pangenome is defined as composed by a "core genome" and by an "accessory genome". Core genome is the set of genes conserved in all strains, whereas genes variable among strains constituted the accessory genome [26]. It is extremely outstanding to outlined the core and the accessory genome because the corresponding gene sets include genes linked to the phenotypic similarities and to the phenotypic differences among strains, respectively. The comparative analysis performed on the available genomes of *E. meliloti* strains allowed to define the pangenome, which resulted in a core genome of 5196 genes and in an accesory genome of 3085 genes [24]. The accessory genome represent an about 38% of the total genome and therefore it constitutes a large portion of the total genetic repertoire. The symbiotic accessory genome was found to be highly variable. The most notable feature was a large variability in the so-called “microaerophilic” gene set, which includes the transcriptional regulator annotated as FixK-like, a third copy of electron transport chain (*fixNOQP*) and several genes related to nitrogen metabolism (*nos, nor, nir, nnr and nrt*). A comparable variability at regulatory level was also revealed. Presence and extent of polymorphism in *E. meliloti* regulons of transcription factors (NolR, NodD, FixJ, FixK, NifA, ChvL, Fur, NesR) involved in symbiotic interaction were also determined [27]. Regulatory interactions present in all the strains of *E. meliloti* constitute the core regulon and the regulatory interactions present in one or two strains constitute the accessory regulon. A large accessory symbiotic regulon of *E. meliloti* was found for most of the analyzed transcriptional regulators either because of the absence of the target gene or because of the absence of the predicted regulator binding site. About 31% of the putatively missing connections between regulator and regulated genes are due to the loss of DNA binding sites, the relative genes being still present in the genome. It can be conjectured that the presence of genes, which have lost (or not still acquired) the binding sites, may reflect a relatively recent evolutionary divergence, such as is expected among strains of the same species. The outlined data indicate that regulons are flexible, with a large number of accessory genes, suggesting that regulon polymorphism could also be a

key determinant in the variability of symbiotic performances among the analyzed strains of *E. meliloti*.

1.3. Evolution of symbiosis

The rhizobium-legume symbiosis, a relatively recent evolutionary adaptation, is thought to have evolved from the ancient arbuscular mycorrhizal symbiosis that is nearly ubiquitous throughout the plant kingdom [28]. This evolutionary relationship has been inferred based on findings that several host genes represent common requirements for the establishment of both rhizobial and mycorrhizal symbioses. Given that nearly all vascular plants interact with mycorrhizal symbionts, it remains unclear why the nitrogen-fixing symbiosis is strictly limited to legume species, with the exception of *Parasponia* (*Ulmaceae*). Current understandings of legume evolution and the appearance of nodulation indicate that the first symbiosis event involved bacterial invasion of roots via cracks in the host epidermis where lateral roots emerge. Subsequent to this, developmental mechanisms evolved, likely through the process of gene duplication, to craft the highly selective symbiosis described here. The emergence of a host-derived infection structure allows host control over the bacterial infection process. In this context, the symbiont is regarded as an "addomesticated" pathogen [29]. There is a strict specificity in the establishment of a symbiosis between host legume species and their nitrogen-fixing symbionts . Some rhizobia have very restricted host ranges, composed of only one or a few closely related legume species. *Rhizobium etli* is compatible only with species of *Phaseolus* (bean) genus and *Ensifer meliloti* is compatible with species of *Medicago* (alfalfa), *Melilotus* (sweetclover), and *Trigonella* (fenugreek) genera (Table 2). On the other hand, some rhizobia have a broader host range [30]. The *Rhizobium* strain NGR234 can nodulate legumes belonging to at least 112 genera of legumes (Table 2). Based on rhizobial phylogenetic relationships, it was suggested that the restricted host range symbiosis evolved from an ancestral broad host range symbiosis, and perhaps the specificity engendered by narrow host range interactions creates a finely tuned and more effective symbiosis [6].Several genomic evidences suggest that the symbiotic capacity of rhizobia have evolved in part through horizontal gene transfer events. Within the symbiotic rhizobial lineage, *Ensifer* is estimated to have diverged from *Bradyrhizobium* approximately 500 Mya, which is well before the initial appearance of legume species (60 Mya) [31]. The large and multipartite genomes of rhizobia consisting of a chromosome supplemented with one or more independent megaplasmids, contribute to an evolutionarily dynamic genome through the process of horizontal gene transfer. Moreover, rhizobial genes involved in symbiosis are often located within chromosomal

islands or on highly mobile plasmids. This is the case of the *Ensifer meliloti* genes involved in nodulation and nitrogen-fixation that are located on the megaplasmid pSymA. Horizontal transfer of these genomic elements has been observed among rhizobia within the rhizosphere and has the ability to convert a non symbiont into a symbiont through a single transfer event [32]. Other than symbiosis genes, there is no significant synteny shared between the plasmids of various rhizobial species or rhizobial chromosomes. Recently, has been discovered that two rhizobia belonging to the *Burkholderia* genus of *β*-proteobacteria are also able of establishing nitrogen-fixing symbioses with legumes [33], confirming the high mobility of the genetic elements necessary for the nitrogen-fixation symbiosis. Moreover, comparative phylogenetic analyses support the notion that the plasmid-borne symbiotic genes in rhizobia are derived from at least one horizontal gene transfer event [34].

TABLE 2

Some Species of Rhizobia and Their Legumes Hosts

Species[a,b]	**Hosts nodulated**
Rhizobium leguminosarum	
bv. phaseoli	*Phaseolus*
bv. trifolii	*Trifolium*
bv viciae	*Pisum, Lens, Vicia*
Rhizobium etli	*Phaseolus*
Rhizobium galegae	*Galega*
Rhizobium lupini	*Lupinus*
Rhizobium tropici	*Phaseolus, Leucaena*
Sinorhizobium fredii	*Glicine*
Ensifer meliloti	*Medicago, Melilotus, Trigonella*
Mesorhizobium meliloti	*Lotus, Astragalus*
Bradyrhizobium japonicum	*Glycine, Macroptilium, Vigna*
Azorhizobium caulinodans [c]	*Sesbania*
Rhizobium spp. [d]	*Vigna, Arachis, Desmodium, Lotus, etc.*
Bradyrhizobium spp.	*Sarothannus, Ulex, etc.*

[a,b] *Rhzobium, Sinorhizobium* and *Ensifer* species are fast growing in laboratory culture media. *Bradyrhizobium* species grow more slowly and *Mesorhizobium* species display an intermediate growth rate.

[c] Stem-nodulating and exceptional among rhizobia in nitrogen fixing in the free-living state.

[d] Includes strain NGR234, which can nodulate at least 112 legume genera.

1.4. The nodulation process of the host plant

1.4.1. Root infection and nodule development

Invasion of the host plant roots by rhizobia is a multistep process that begins with the signaling pre-infection events that take place in the rhizosphere [35].

Nitrogen-fixing rhizobia and leguminous plants have developed a complex molecular signal exchange that, in the early stages, involves the release from plant roots of flavonoid compounds (Figure 1.5), which are key signals for the organogenesis of nitrogen-fixing nodules. The secreted flavonoids are recognized by specific rhizobia through the NodD receptors. The flavonoid recognition leads to the transcriptional activation of rhizobial nodulation (*nod*) operon, resulting in the synthesis of the chito-oligosaccharide NF as response to the plant signal. The recognition of specific NFs by the plant root hairs elicits organized responses and differentiation programs in the plant roots leading to the rhizobium invasion and *de novo* formation of a specialized root organ, the nodule [36]. The whole process is tightly regulated at the genetic level and is developed in several stages,

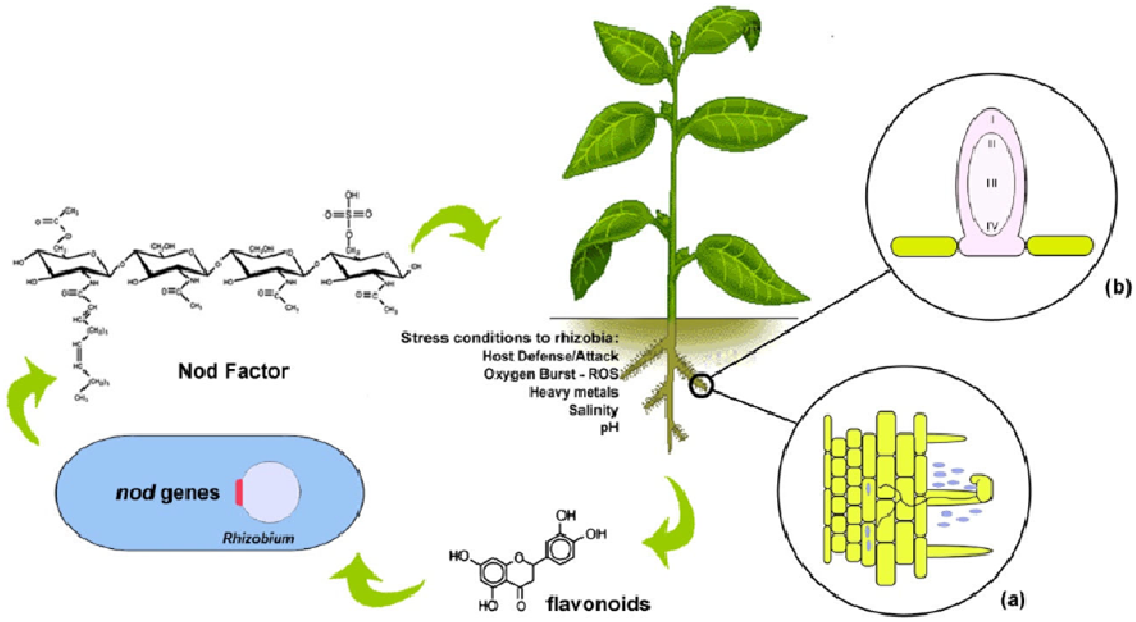


Figure 1.5. Signal exchange and root invasion in the rhizobium-legume symbiosis. Flavonoids released by the host plant induce rhizobial nodulation genes expression and lead to production of NFs. a-b) In turn, rhizobial NFs lead to the curling of hair root and formation of root nodules (modified by [6]).

The primary sites of infection for many rhizobia are young root hairs. Rhizobia respond to root exudates and move by chemotaxis toward specific sites localized on the legume roots. Apparently, chemotaxis is not a necessary requirement for nodulation although it has an influence on competition and organization in the rhizosphere [37]. The host lectines play an important role for the adhesion of rhizobia to the plant roots. Lectines are located in root hair apex and it is believed could help to maintain the host-symbiont specificity by binding simultaneously the plant cell wall and the carbohydrate portions of compatible bacteria outer surface. Recent studies suggest that cell-cell contact and specific binding of compatible bacteria to root hairs are important for early infection. Indeed, high localized concentration of NF is required to stimulate the curling of the root hair and root invasion [38]. When the bacteria adhering to the plant wall, the NF produced by rhizobia is absorbed by root cells an stimulates mitotic cell division both in root hairs and in root inner cortex of the host plant. The changes induce by NF lead to the root hair curling at the tip that entraps intimately associated rhizobial cells (Figure 1.6a). After the entrapment, a local lesion by rhizobium hydrolysis of the plant cell wall is formed (Figure 1.6b) [39]. The tubular intrusion structure formed by the ingrowth of the root hair cell walls from the point of penetration of rhizobia is called Infection Thread (IT). Inside IT, the rhizobial invasion proceed to the root cortical cells, by continued bacterial proliferation and new membrane synthesis at the tip of the developing IT (Figure 1.6c,d) [39]. Finally, bacteria within IT are deposited into the host cell cytoplasm in a process that resembles endocytosis [6]

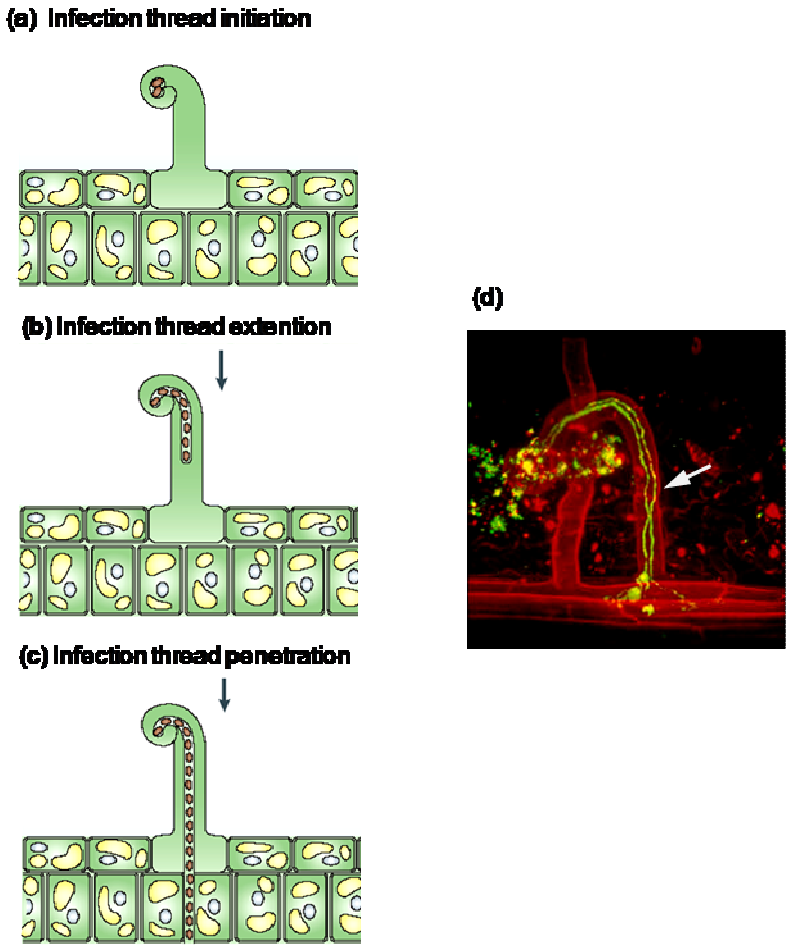


Figure 1.6. Root hair invasion by *Ensifer meliloti*. (a) Interaction between rhizobial cells and root hair; (b) Infection thread development; (c) Infection thread penetration into the underlying cell layers allowing rhizobial cells to reach the root cortex; (d) Invasion of the legume root (red) by cells of *E. meliloti* that over-express the green fluorescent protein (GFP). This root hair contains a rare double strand of infection (white arrow) (modified by [29].

The size and shape of the nodules are very different among the species of legume genera. The root nodules can be classified into two major categories from a morfological and hystological point of view: indeterminate nodules and determinate nodules. *Ensifer meliloti,* forms indeterminate nodules (Figure 1.7).

These indeterminate nodules are usually formed on temperate legumes (e.g. *Medicago sativa, Pisum sativum, Vicia hirsuta*) and are characterized by persistent meristematic activity, which causes an elongated shape of nodules. The central tissue of such nodules consists of a number of distinct zones containing both plant cells and invading

rhizobia at different stages of differentiation. Once inside nodule, the bacterial cells continue to differentiate and synthetize proteins required for nitrogen fixation.

Figure 1.7. Indeterminate nodules on legume roots [modified by http://biologia.campusnet.unito.it/cgi-bin/corsi.pl/Show?_id=2a87].

Structural studies of mature nodules led to distinguish the following spatially defined regions [40]:

Zone I meristematic, situated at the apex of the nodule, is a region of actively dividing plant cells devoid of bacteria.

Zone II is called the infection zone. Here the bacteria enter the root cells via infection threads and undergo differentiation into bacteroids.

Interzone II-III is a very restricted zone that contains only 3-4 layers of cells, separating the pre- nitrogen fixation zone II and nitrogen-fixing zone III.

Zone III contains fully differentiated nitrogen-fixing bacteroids and therefore it is characterized by an intense nitrogen-fixation activity. In zone III, the leghemoglobin is produced giving the typical pink color of the nitrogen-fixing nodules. Leghemoglobin is essential because of its binding of oxygen molecules, protecting oxygen-sensitive nitrogenase, the crucial bacterial enzyme catalyzing nitrogen-fixation.

The basal part of the indeterminate nodule is constituted by a senescence zone (Zone IV) containing rhizobia no more efficient in nitrogen-fixation processes.

Determinate nodules are formed on tropical and subtropical legumes (*Glycine max*, *Phaseolus vulgaris*, *Lotus japonicum*). The determinate nodules are characterized by disappearance of meristematic activity after nodulation and thus have a cylindrical shape. Inside determinate nodules, the three infection zones (recent, mature, senescent)

follow each other in time rather than space. This leads to the formation of a structure called nodule primordium [39].

1.4.2. Differentiation into bacteroids

Invading bacteria within the infection thread, once reached the target tissue that is the inner bark of the plant, are internalized in the cell cortex. Each bacterial cell undergoes endocytosis by a target cell in an individual vesicle in which the membrane is formed by the plasmalemm of plant cells. The entire unit, which consists of a single bacterium and the surrounding endocytic membrane is called symbiosome [39]. At this point rhizobial cells undergo into a series of changes and differentiate into a specialized symbiotic form referred as bacteroid (Figure 1.8). Bacteroids are surrounded by a modified plant membrane, greatly increased their size, assumed a club shape and lost the ability to replicate. Moreover the bacteroids membrane contains many invaginations to improve the metabolic exchanges between the two symbionts, cytoplasm is rich of nitrogenase and has more than one nucleoid. Bacteroids establish a chronic infection of the host cell cytoplasm and constitute the active form of rhizobia able to fix nitrogen. New lipidic and proteic material attached to the symbiosome membrane assigns a new chemical identity to this compartment [41].

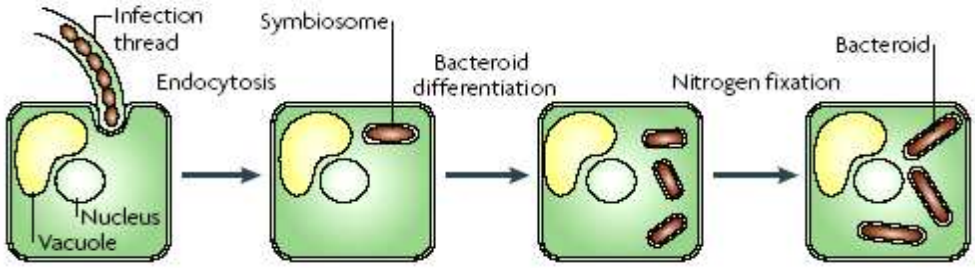


Figure 1.8. Endocytosis of rhizobia and bacteroids differentiation (modified by [42]).

1.4.3. Symbiotic relationship establishment and nitrogen-fixation

The *in planta* differentiation of rhizobia involves significant morphological and metabolic changes resulting from a fine-tuned modification of gene expression.
The transcriptional changes occuring into the bacteroids consist of a down-regulation of many metabolic processes in conjunction with the increased expression of gene products involved in nitrogen-fixation respect to the rhizobium cells. Bacteroids

undertake respiratory chain modification, which allows energy utilization under microaerobic conditions, repress glycolysis genes, activate C4-dicarboxylic acid utilization pathways for carbon metabolism and induce nitrogenase gene expression [43]. The ATP-dependent enzyme nitrogenase is responsible for catalyzing nitrogen-fixation by which rhizobial bacteroids provide nitrogen into an available form or the host plant [42]. This essential enzyme complex is constituted by an iron-protein -reductase and by a molibden-iron protein [44]. The metabolic product of the nitrogenase enzyme reaction is ammonia, assimilated by the host through its incorporation into the aminoacids glutamine and glutamate. In exchange, rhizobia are provided by the host plant with energy in the form of photosynthesis products (carbohydrates). The energy cost for rhizobia is about 16 molecules of ATP for reducing one molecule of atmospheric nitrogen into two molecules of ammonia [45]. The concentration levels of O_2 inside the nodule is critical for the nitrogenase enzymatic activity because oxygen strongly inhibits the nitrogenase. Thus, the O_2 levels must to be strictly controlled. Nevertheless, oxygen is required for the normal metabolic activities and for the cellular respiration of bacteria to provide ATP to nitrogenase. The control of the host microaerobic environment is dependent on structural aspects of the nodule that form an oxygen diffusion barrier in combination with high expression levels of plant leghemoglobin [46]. The leghemoglobin helps to limit the concentration of free oxygen and well simplifies the symbiotic relationship. Indeed leghemoglobin consists of a heme group synthesized by rhizobium and a globin part synthesized instead by plant cell [47]. Leghemoglobin provides sufficient oxygen for the metabolic functions of the bacteroids but prevents the accumulation of free oxygen that would destroy the activity of nitrogenase

1.4.4. Host defense response to symbiotic rhizobia

Plants in response to the microbial invasion can set up a complex defense responses mediated by signal molecules such as salicylic acid, reactive oxygen species (ROS: $O_2^{\cdot -}$, H_2O_2, and $HO^{\cdot}$), nitric oxide, jasmonic acid and ethylene (Figure 1.9) [48]. Therefore, rhizobia have in turn evolved strategies to avoid the accumulation of such signals once they have been recognized by the legume host. Several studies have revealed a striking similarity between the molecular mechanisms underlying the perception of nodulation factors of rhizobia and molecular structures that are associated with bacterial pathogens of plants (Figura 1.9) [49]. As with many host-microbe interactions, the rhizobium-legume symbiosis can be associated with a host-generated

release of ROS. The unsuccessful or aborted ITs display characteristics of the hypersensitive plant defense response that typically includes ROS production. Thus, the ROS efflux could play a role in limiting bacterial invasion [50]. Strategies to limit the synthesis of ethylene by the plant in response to microbial invasion are taken by some rhizobia and by plant-pathogenic bacteria. *Bradyrhizobium elkanii* and the plant pathogen *Burkholderia andropogonis* produce rizobitoxine [2-amino-4-(2-amino-3-hydroxypropoxy)-*trans*-but-3-enoic acid], an inhibitor of ethylene synthesis [51]. Several rhizobia produce the enzyme 1-aminocyclopropane-1-carboxylic acid (ACC) deaminase, which degrades the immediate precursor of ethylene [52]. Each strategy leads to an increase in efficiency of nodulation. Some pathogenic strains of *P. syringae* synthetize a phytotoxin (coronatine), which suppresses plant defenses based on salicylic acid inducing the jasmonic acid signaling pathways [53]. In addition to these strategies, rhizobia and plant-pathogenic bacteria use similar components, such as surface polysaccharides (EPS), antioxidant systems, ethylene inhibitors and specific virulence factors to control or actively suppress plant defenses [54]

Interestingly to note that hundreds of gene homologues to pathogen virulence factors are present in the available genomes of rhizobia. Moreover, the functional characterization of some of these genes, such as those that encode for type III and IV secretion systems, indicate a similar role in rhizobia-legume interaction. Thus, in plant-pathogenic bacteria and rhizobia are present factors such as surface polysaccharides, quorum sensing signals and secretion proteins, which play an important role modulating the plant defense response and in the outcome of the interaction [49].

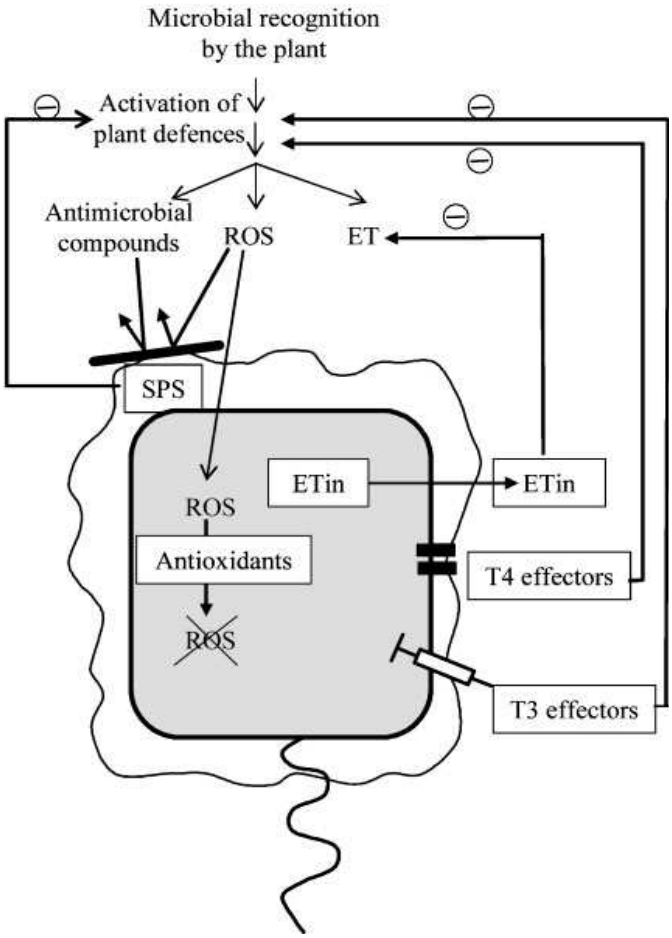


Figure 1.9. Bacterial components used to control plant defense responses. Surface polysaccharides (SPS) are able to suppress microbial-induced defence reactions and/or to act as shields protecting the bacterium against toxic compounds. Additionally, active suppression of defence reaction is achieved with ethylene inhibitors (ETin) and virulence factors such as type III and IV secretion systems. Antioxidant systems protect bacteria against reactive oxygen species (ROS) [49]).

The maintenance by rhizobia of a large number of genes required for symbiosis with their legume hosts is a question particularly relevant from an evolutionary point of view, especially in light of the recent observation that bacteroids within indeterminate nodules are terminally differentiated and unable to give rise to progeny [55;56]. A single symbiotic rhizobium is predicted to have a greater fitness if it successfully colonizes a nodule than its nonsymbiotic cousin living in the soil where growth can be severely limited by nutrient availability. Although there appears to be a fitness gain for rhizobia able to invade the nodule, it is also clear that the host has evolved mechanisms that prevent nitrogen-nonfixing rhizobia from parasitizing the legume nodule for energy. While the legume host controls the infection process and nodule morphology, the microsymbiont largely dictates the efficiency of nitrogen-fixation. Mathematical models suggest that if legumes deal with nitrogen-fixing and nonfixing rhizobial strains within the nodule without distinctions, then

nitrogen-nonfixing rhizobia would quickly outcompete the nitrogen-fixing ones [57]. For this reason, the host imposes effective sanctions on nitrogen-nonfixing rhizobial cheaters within the nodule [57;58]. So far, host sanctions have been found to take the form of severe O_2 limitation to nitrogen-nonfixing rhizobia within the nodule, which restricts bacterial viability and growth. Thus, the legume host imposes selective pressure on rhizobial cells that may affect the evolution of rhizobium populations in favor of efficient nitrogen-fixing rhizobia [59]. Despite the plant host sanctions, possible explanations for the persistance of less efficient rhizobial strains in nitrogen-fixation, could be the presence of mixed population inside nodules, systems of balanced selection, biochemical manipulations of the host by some rhizobial strains and differences in sanctions by different host genotypes [58;61]. The frequency of mixed nodules has rarely been measured in field. More than 32% of the nodules of soybean grown in field contains two strains [62], which allow to maintain the total nitrogen-fixation per nodule high enough to avoid sanctions if one strain less reduces nitrogen. The differentiation of the rhizobial cells into bacteroids avoids the redirection of resources from nitrogen-fixation to the bacterial reproduction and it avoids that they become pests and infect other tissues of the plant [56]. The bacteroids can be more easily lysed during senescence of the nodule, thereby facilitating the recovery of nutrients from the host. In addition, bacteroids have not direct benefit to accumulate reserve substances such as polyhydroxybutyrate, (PHB). The synthesis of PHB may reduce the total amount of carbon available for efficient rhizobia able to reproduce. Furthermore, diverting resources from nitrogen-fixation to the synthesis of PHB could trigger sanctions at the nodule that may damage undifferentiated rhizobia [60].

1.5. Genes and molecular signals in the rhizobium-legume symbiosis

1.5.1. The Host plant flavonoids

The high specificity of the rhizobium-legume symbiosis is mainly due to the molecular signals produced by the two symbionts. Different legumes secret different types of signals and different rhizobia have NodD proteins that recognize different root-exudated signals, allowing the establishment of a highly specific relationship. Compatible rhizobia are uniquely capable of gaining entry and invading the host roots based on a series of reciprocal signalling events. The early signals involved a diverse cocktail of flavonoids, which are actively exudated by the roots of leguminous plants into the soil

(Figure 1.10) [63]. Flavonoids released from plant roots are the key signals to trigger invasion and root nodules formation [64].

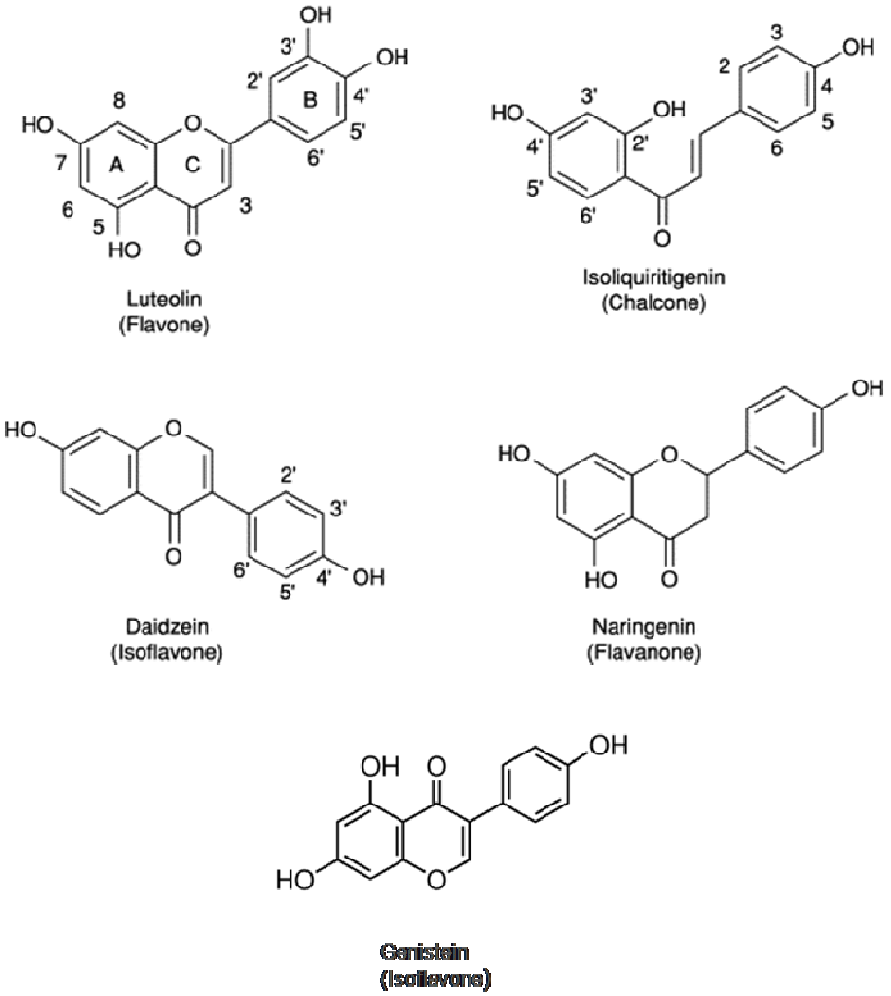


Figure 1.10. Major representative host flavonoids that are crucial signaling molecules for symbiosis: luteolin, isoliquiritigenin, daidzein, naringenin and genistein (modified by [64]).

Flavonoids are a class of plant secondary metabolites produced by the central phenylpropanoid pathway and the acetate-malonate pathway of plants. They consist of polyaromatic compounds with a skeleton of 15 carbon atoms, divided into subclasses according to their structure. The flavonoid skeleton can be modified by a diversity of substitutions that have important effects on flavonoid function, solubility, mobility and degradation in the soil. The main flavonoid subclasses (e.g., chalcones, flavones, flavanones, flavonols, proanthocyanidins, isoflavones, isoflavans, pterocarpans) contain numerous compounds involved in many plant functions including pigmentation,

protection against ultraviolet (UV) light, pollen fertility, regulation of auxin transport, and hydrogen peroxide scavenging, as well as interactions with symbiotic microorganisms or defense against microbial pathogens [65;66].

Flavonoids, acting as primary signals to rhizobia, have been found in legume seed coat and root exudates. When deposited on seed coat, flavonoids are simply released into the surrounding aqueous environment during imbibition without the involvement of any metabolic regulation. The storage of flavonoids in roots and their release from epidermal tissues are, however, subject to internal metabolic controls, and a strong evidence exists for a process of concurrent synthesis and release. Flavonoids may be released as aglycones or glycosidic conjugates. The latter are inherently more soluble and therefore may have a greater potential for diffusion from the root surface, before being hydrolyzed to the aglycone form by rhizobia, other soil microorganisms, or plant exoenzymes [67]. Rhizobia themselves may be able to alter the hydrophobicity of flavonoid aglycones; the complexation of the luteolin with cyclosophoraoses produced by *E. meliloti* markedly enhances the solubility of this nodulation inducer [68]. The presence of rhizobia in the legume rhizosphere also influences the quantity and the types of flavonoids released from roots. About that, root exudates of *Medicago sativa* inoculated with *E. meliloti* were found to be qualitatively different with respect to flavonoid content compared with exudates from sterile plants [69]. Moreover, flavonoids and isoflavonoids are not inert compounds, because rhizobia in legume rhizosphere are capable of metabolizing them to yield a plethora of polycyclic and monocyclic phenolic products [70]. These compounds and many other simple phenolics can be used by rhizobia as sole carbon and energy sources [71]. The concentration of flavonoids into rhizosphere varies widely and depends on plant growth conditions, nutrient supply and plant species. However, only few information on actual flavonoids concentrations in the soil and how these concentrations change in space and time are available [72].

Flavonoids generally have a fundamental role in protecting higher plants from biotic and abiotic stresses. In the rhizophere, plant-derived flavonoids play multiple roles (Figure 1.11), depending on their structure, such as to inhibit several phytopathogens, to stimulate mycorrhizal spore germination and hyphal branching, to mediate allelopathic interactions and to chelate soil nutrients [64]. Flavonoids can also alter the nutrient concentration and availability in the soil by acting as antioxidants and metal chelators. An isoflavonoid identified in *Medicago sativa* root exudates was able to dissolve ferric phosphate, thus making both phosphate and iron available to the plant [73]. In rhizobium-legume symbiosis, plant flavonoids have been shown to evoke a strong chemoattractant response of rhizobia toward plant roots, to act as microbial developmental regulators, as determinants of host specificity as well as regulators of

phytoalexin and rhizospheric compounds resistance. Moreover, flavonoids were demonstrated to inhibit the auxin transport thus causing accumulation of this phytohormone at the nodule initiation site to stimulate cell division and nodule organogenesis [74-76]. Several rhizobia genes have also been reported to be regulated in response to host flavonoids, including those for exopolysaccharide synthesis, which are important for modulate the defense responses in the host.

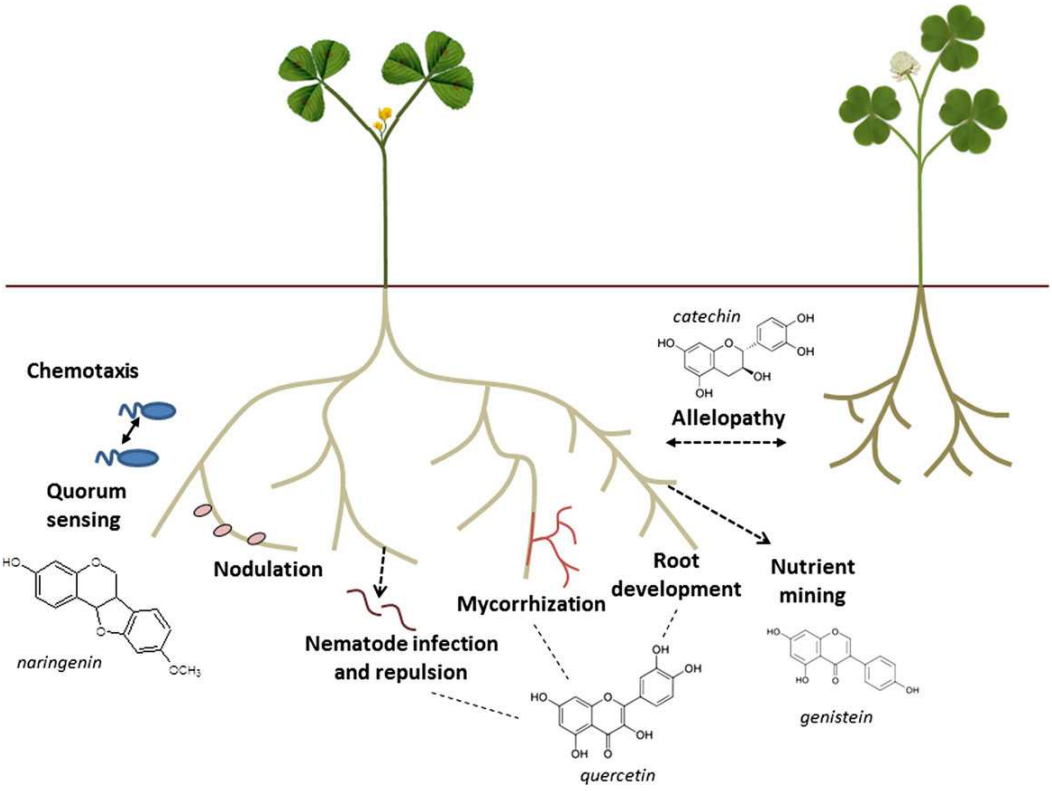


Figure 1.11. Schematic overview of flavonoid functions in the rhizosphere. Flavonoid functions in the rhizosphere range from *nod* gene inducers and chemoattractants in rhizobia, stimulators of mycorrhizal spore germination and hyphal branching, possible quorum-sensing regulators in bacteria, repellents for parasitic nematodes, nutrient mining, and as allelochemicals in plant–plant interactions. They can also affect root development. Examples of biologically active flavonoids mediating the different interactions are shown (modified by [72]).

Rhizobia exhibit positive chemotaxis both toward legume exudates [77] and to individual compounds that are present in exudates, including a number of flavonoids. Luteolin, 4,40-dihydroxy-20-methoxychalcone, 7,40-dihydroxyflavone, and 7,40-dihydroxyflavanone from alfalfa induce positive chemotaxis in *E. meliloti*. Apigenin, luteolin, umbelliferone, and acetosyringone act as chemoattractans for *Rhizobium leguminosarum* bv. *phaseoli* [78]; naringenin, kaempferol (3,4,5,7-tetrahydroxyflavonol) and apigenin are chemoattractants for *R. leguminosarum bv. viciae* [79]. Conversely, *Bradyrhizobium japonicum* showed no chemotaxis to isoflavonoids from its soybean

host; however, hydroxycinnamic acids were strong chemoattractants [80]. In addition to plant flavonoids, rhizobia display to be positively chemoattracted to many other compounds such as sugars, common aminoacids [81] as well as aromatic acids, hydroxyaromatic acids and simple phenolic compounds [78]. Flavonoids depending on their concentration are potentially toxic to bacteria and inhibitory effects on rhizobial growth have been reported. Soybean rhizobia are sensitive to the phytoalexin glyceollin [82], whereas medicarpin and kievitone from soybean roots were strong inhibitors of growth for *B. japonicum* and *R. lupini*. Flavonoids also act as growth stimulators. Lameta and coworkers [83] reported a stimulatory effect on the growth of *B. japonicum* by daidzein at low concentrations. Genistein, naringenin, chrysin, and apigenin promoted the growth of *Sinorhizobium fredii* in late *log* phase [84]. Quercitin was found to exert a positive effect on the growth of *E. meliloti* [85].

1.5.2. Flavonoids as inducers of nodulation (*nod*) genes

The most relevant and crucial role of root-exudated flavonoids is as inducers of host nodulation in rhizobium-legume symbiosis.The first inducing flavonoids to be discovered were the luteolin [86] and 7,4 -dihydroxyflavone [87]. The former has been isolated from the seed coat of *Medicago sativa* and the latter from roots of *Trifolium repens*. They are nodulation inducers for *Ensifer meliloti* and *Rhizobium leguminosarum* bv. *trifolii*, respectively (Table 3).

TABLE 3

Rhizobial nod Genes Inducers Isolated from Legumes

Host legume	Inducer
Medicago sativa	Luteolin (5,7,3'4'-tetrahydroxyflavone)
	Chrysoeriol (3'-methoxy-5,7,4'-trihydroxyflavone)
	Liquiritigenin (7,4'-dihydroxyflavanone)
	7,4'-dihydroxyflavone
	Methoxychalcone
	Stachydrine (betaine)
	Trigonelline (betaine)
Pisum sativum	Apigenin-7-O-glucoside
	Eriodictyol (5,7,3',4'-trihydroxyflavanone)
Vicia sativa	3,5,7,3'-tetrahydroxyflavone-4'-methoxyflavanone
	7,3'-dihydroxyflavone-4'-methoxyflavanone
	Four more partially characterized flavanones

Trifolium repens	7,4'-dihydroxyflavone
	Geraldone (7,4'-dihydroxy-3'-methoxyflavone)
	4'-hydroxyflavone-7-methoxyflavone
Glycine max	Daidzenin (7,4'-dihydroxyisoflavone)
	Genistein (5,7,4'-trihydroxyisoflavone)
	Coumestrol (3,9-dihydroxycoumestan)
	Isoliquiritigenin (4,2',4'-trihydroxychalcone)
	Genistein-7-O-glucoside
	Genistein-7-O-(6-O-malonylglucoside)
Phaseolus vulgaris	Delphinidin (3,5,7,3',4',5'-hexahydroxyflavylium)
	Kaempferol (3,5,7,4'-tetrahydroxyflavonol)
	Malvidin (3,5,7,4'-pentahydroxy-3',5'-dimethoxyflavylium)
	Myricetin (3,5,7,3',4',5'-hexahydroxyflavone)
	Petunidin (3,5,7,4',5'-pentahydroxy-3'-methoxyflavylium)
	Quercitin (3,5,7,3',4'-pentahydroxyflavonol)
	Eriodictyol (5,7,3',4'-trihydroxyflavanone)
	Genistein (5,7,4'-trihydroxyisoflavone)
	Naringenin (5,7,4'-trihydroxyflavanone)
	Daidzein (7,4'-dihydroxyisoflavone)
	Liquiritigenin (7,4'-dihydroxyflavanone)
	Isoliquiritigenin (4,2',4'-trihydroxychalcone)
	Coumestrol (3,9-dihydroxycoumestan)
Vigna	Daidzein (7,4'-dihydroxyisoflavone)
	Genistein (5,7,4'-trihydroxyisoflavone)
	Coumestrol (3,9-dihydroxycoumestan)
Sesbania rostrata	Liquiritigenin (7,4'-dihydroxyflavanone)
Lupinus albus	Erythronic acid (aldonic acid)
	Tetronic acid (aldonic acid)
Galega orientalis	uncharacterized chalcone

Plant-derived flavonoids, which passively diffuse across the bacterial membrane, are perceived by the rhizobial regulator NodD, thereby eliciting a significant transcriptional response from compatible bacteria within the rhizosphere, which results in host nodulation.

NodD proteins belong to the LysR family of transcriptional regulators [88]. They are constituted of an N-terminal-ligand-binding domain thought to function as flavonoid receptor that regulates the activity of the associated C-terminal DNA-binding domain. In the presence of suitable plant inducer, NodD regulators bind a conserved regulatory DNA sequence, called *nod* boxes, thereby inducing the expression of nodulation genes (*nod, nol, noe*). The protein products of nodulation genes (Table 4) are collectively involved in biosynthesis of lipo-oligosaccharide signal, known as NF [89]. However,

interesting nuances to NodD-dependent regulation are beginning to emerge, including the identification of genes unrelated to the NF biosynthesis within the NodD regulon and a temporal progression to flavonoid-induced gene expression that implies NodD coordinates a complex regulatory hierarchy [90]. The molecular basis for the NodD activation are not yet completely understood. The binding of an appropriate flavonoid presumably induces a conformational change in NodD that enhances the RNA polymerase access at the promoter of target genes [91]. NodD regulatory system is present in all *Rhizobium*, *Bradyrhizobium* and *Azorhizobium* strains studied so far. However, there are variations between species in the number of *nodD* copies present and one to five copies have been observed in the sequenced rhizobia genomes. The species that have only one copy of this gene, such as *Rhizobium leguminosarum* bv. *trifolii*, a mutation usually results in the abolition of nodulation [92]. In *E. meliloti*, *R. leguminosarum* bv. *phaseoli* and *B. japonicum*, which have multiple *nodD* copies, the nodulation is not completely suppressed by a mutation in a single *nodD* gene [93]. In some rhizobia, such as *E. meliloti* and *B. japonicum*, *nod* gene induction appears to involve more complicated regulatory mechanisms. In *E. meliloti* another symbiotic regulatory gene, *syrM* which is flavonoid independent, acts in conjunction with *nodD3* to provide self-amplifying positive regulation of nodulation genes in developing root nodules [94]. *B. japonicum* possesses two supplementary genes, *nodV* and *nodW*, which are distinct from *nodD* and involved in the regulation of the NFsynthesis via isoflavonoid inducers. This two-component system relies on NodV, a sensory kinase, to recognize flavonoids that do not normally interact with NodD, whereas NodW activates gene transcription [95]. Another regulatory system is present in *S. fredii* and involves the *nolJ*, *nolBTUV*, and *nolX* transcriptional units [96]. Regulation of nodulation gene expression is also subjected to negative control by repressor proteins such as NolR and NolA whose production is flavonoid independent [97]. An excess of NFs in the rhizosphere is apparently harmful to efficient nodulation and can affect the spectrum of hosts that are nodulated. It may also trigger unwanted host defense reactions [98]. In *R. leguminosarum* bv. *viciae* the single *nodD* gene is negatively autoregulated by its own product, NodD [99]. The specific perception of a certain mixture of exudated flavonoids by NodD proteins of different rhizobia is responsible for the host range determination. Moreover, the spectrum of flavonoid specificity of the endogenous NodD protein appeared to correlate with the broadness of the host range [100]. The NodD regulators of broad host range rhizobia respond to a wider range of flavonoid species than those present in restricted host range bacteria [63]. NodD1 from the broad host range symbiont *Rhizobium sp*. NGR234 responds positively to a structurally diverse arrayof compounds, including phenolics (vanillin and isovanillin) that are inhibitors for other

rhizobia [101]. Rhizobia with narrower host range appear to require a more specific pattern of substitutions in the basic flavonoid structure to ensure interaction with NodD.
The activation of the transcriptional regulator NodD by suitable host flavonoids results in the expression of the *Rhizobium* nodulation genes that are essential for host infection and nodulation. The nodulation genes can be divided in four major classes: *nod, nif, fix* and *enf* genes (Table 4). The structural *nod* genes are in turn classified into two groups, the common and host-specific *nod* genes. The common genes, as *nodABC* and *nodIJ,* have been found in *Azorhizobium*, *Rhizobium* and *Bradyrhizobium* [102]. These genes are designated as common *nod* genes because are structurally conserved and functionally interchangeable between the species outlined above without altering the host range [103]. The host-specific *nod* genes are not conserved among rhizobia and are necessary for the nodulation of a particular host plant [104]. In most cases, mutations of host-specific genes cannot be fully complemented by the introduction of ortholog genes from other rhizobia and such mutations often result in alteration or enlargement of the host range [105]. Common *nod* genes encode for the enzymes responsible for synthesis of the NF chitin backbone. In contrast, host-specific *nod* genes introduce various modifications of the NF basal structure in order to confer specificity for nodulation of a particular host [63]. The symbiotic nitrogen-fixation requires the coordinate interaction of the *nif* and *fix* classes of genes present in rhizobia. The *nif* genes are involved in the synthesis, functioning and regulation of the nitrogenase enzymatic complex [106], which catalyzes nitrogen-fixation in symbionts and free-living microorganisms. Environmentally, *nif* genes expression can be repressed in the presence of high soil ammonia and high oxygen concentrations [107]. The f*ix* genes coordinate and regulate the nitrogen-fixation process inside the nodule and therefore are essential for its proper functioning. The *enf* genes influence the kinetic and efficiently of the host nodulation.
In most *Rhizobium* species studied to date, the *nod* genes reside on large symbiotic plasmids (pSym) that also carry the *nif* and *fix* genes [25]. The symbiosis-related genes are localized on the chromosome, in *Rhizobium loti*, *Bradyrhizobium spp.* and *Azorhizobium s*pp..

TABLE 4

Nodulation Genes Products Required For Synthesis and Release of Nod Factor (NF)

Protein	Function
Biosynthesis of glucosamine (chitin) oligosaccharide backbone	
NodM	Glucosamine synthase
NodC	*N*-acetyl-glucosamine transferase
NodB	Deacetylase, acting at the non reducing end of glucosamine oligosaccharide
Biosynthesis and transfer of fatty acid moiety at nonreducing end	
NodF	Acyl carrier protein
NodE	*β*-Ketoacyl synthase
NodA[a]	Acyl transferase involved in *N*-acylation od deacetylated nonreducing end of glucosamine oligosaccharide
Modification of nonreducing end	
NodS	Methyl transferase
NodU	Carbamoyl transferase
NolO	Carbamoyl transferase
NodL	*O*-acetyl transferase, *O*-acetylates at 6-C position
Modification of reducing end	
NodP,Q	ATP suphurylase and APS kinase, provide activated sulphur for sulphated Nod Factors
NodH	Sulphotransferase
NoeE	Sulphotransferase involved in sulphation of fucose
NolK	GDP fucose synthesis
NodZ	Fucosyl transferase
NolL	*O*-acetyltransferase; involved in acetyl-fucose formation
NodX	*O*-acetyltransferase, specifically *O*-acetylglucosamine of of *R. leguminosarum*
NoeI	2-*O*-methyltransferase involved in 2-*O*-methylation of fucose
Secretion of Nod Factors	
NodI [a]	ABC transporter component carrying an ATPase domain
NodJ [a]	ABC transporter sub-unit

[a] Present in all rhizobia

Some flavonoids that are inducers of nodulation genes for some rhizobia can act as anti-inducers (antagonists) for others [108]. The isoflavones genistein and daidzein are inducers of *nod* genes expression in *B. japonicum* and *Rhizobium* sp. NGR234, but they are anti-inducers for *R. leguminosarum* bv. *trifolii* and *viciae*. This antagonist effect may be based on competitive inhibition because it can be overcome by increasing inducer concentration [109]. The fact that inducers and anti-inducers often are present in the exudates of a single legume species has prompted the suggestion that *in vivo* levels of *nod* gene induction are the net outcome of positive and negative flavonoid effects on the process [108].

1.5.3. Rhizobial response to host flavonoids: the Nod Factors (NFs)

The rhizobial response to the inducing flavonoids from the plant is represented by production and secretion of the NFs. The specific mechanisms, that characterize the NF discharge are still unknown. However, recent studies indicated that *nodI* and *nodJ* genes are implicated in the production of proteins that affect the lipo-oligosaccharides excretion from rhizobial cells [110]. Bacterial NFs are considered to function as a key for rhizobial entry into legume roots, and the success or otherwise of the infection process is in large part determined by their structural features [111]. Indeed, there is a high degree of stringency for NF chemical structure that determines whether the host allows bacterial invasion to proceed. NF is a complex signaling molecule secreted from the cell as a lipo-oligosaccharide with a backbone of β-(1,4)-linked N-acetyl-D-glucosamine (GlcNAc) residues varying from three to five units [112]. The chitin backbone is modified on the nonreducing terminal residue at the C2 position by a fatty acid. However, the size and saturation-state of this lipid chain varied in a species-specific manner. NF can be further modified with a variety of chemical substituents, including acetyl, arabinosyl, carbamoyl, fucosyl, methyl, and sulfuryl additions. In fact, a given rhizobial species produces a mixture of NF compounds, anywhere from 2 to 60 distinct molecules, and this is true of broad host range bacteria [63] (Figure 1.12).

The NF also plays a role in *E. meliloti* biofilm development and in a host-independent manner [113]; thus, the NF appears to perform a significant role in both free-living and symbiotic lifestyles.

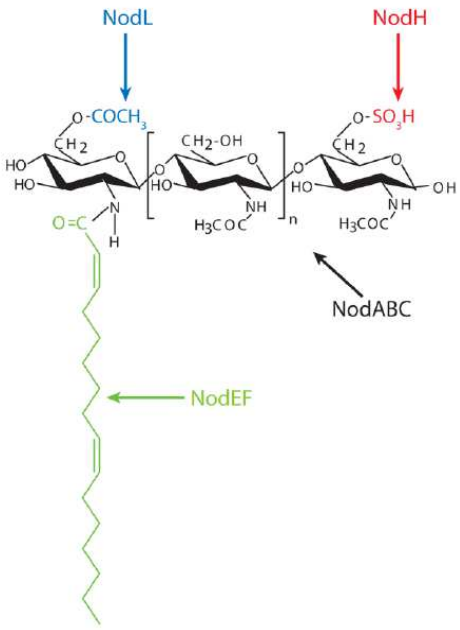


Figure 1.12. Composite structure of Nod Factor and biosynthetic enzymes (Nod proteins) responsible for synthesis and structural modifications on the oligochitin backbone (modified by [6]).

The host signaling-transduction pathway that leads from the perception of NFs to symbiosis-related gene activation is yet the subject of intensive research [114]. The plant receptors *Medicago truncatula* Nod factor perception (MtNFP) could be directly involved in the perception and transduction of the rhizobial NFs signals [114]. The MtNFP receptors belong to the lysin motif (LysM) receptors-like kinase family that contains multiple extracellular domains (Figure 1.13). Host plant responses to specific NF structures depend on the LysM domain specifically and one amino acid difference within this motif can alter the range of rhizobia recognized for symbiosis [115].

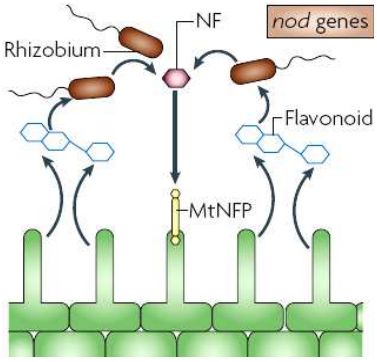


Figure 1.13. Initial dialogue between rhizobium and *Medicago truncatula*. The induction of *nod* genes of rhizobium demands plant flavonoids; *nod* genes then lead to the production of the NF, which initially is perceived by the receptor *Medicago truncatula* Nod factor perception (MtNFP) (modified by [42]).

A number of physiological responses to NFs perception were observed in the host plant by biochemical, molecular and microscopical analysis. Initial root epidermal responses include an alkalinization of the cytosol and a depolarization of the plasma membrane. These two responses appear to depend on a brief NF-induced Ca^{2+}-influx that precedes them, and they are closely followed by a prolonged Ca^{2+}-spiking response. NFs also elicit root hair deformation and root hair curling, leading the rhizobia invasion into the root during the infection thread formation [116]. Root hair deformation likely relies on Ca^{2+}-induced changes to the organization of the actin cytoskeleton, which produces a reorientation of cell growth. In fact, NFs can accumulate within the host plasma membrane [117], and appear to provide a direct positional cue to the host such that the tip of the root hair grows toward the site of the greatest NF concentration [118]. Inhibition of the reactive oxygen-generating system in *Medicago truncatula* roots by NFs, indicating a plant defense suppression function, has recently been reported. NF promotes early morphological responses in root hairs that initiate the process of rhizobial invasion. NFs also induce significant changes in the expression of host genes, including those referred

to as nodulin or *ENOD* genes [119;120] that are induced early in nodule development. More globally, transcriptome profiles reveal that plant genes predicted to be involved in responses to abiotic and biotic stresses, as well as cell reorganization and proliferation, are rapidly induced by rhizobia and largely in a NF-dependent manner [121]. NFs can function as a mitogen modifying the plant hormone balance to elicit the primordium formation that ultimately gives rise to a mature nodule tissue [119;120;122].

1.5.4. Flavonoids as inducers of rhizobial genes other than *nod* genes

In addition to the nodulation genes, transcriptional and proteomic analyses have identified other rhizobial genes and proteins, whose expression is flavonoid-dependent but whose functions have not yet be defined [89;123-127]. They are located both on chromosome and on the symbiotic plasmid pSym. Several flavonoid-inducible transcripts have been found on the symbiotic plasmid of *Rhizobium* sp. NGR234 that shared no homologies with known nodulation genes but strong homologies to a number of other prokaryotic genes and proteins [128]. Further, detailed studies showed that the flavonoid daidzein enhanced the transcription of 147 previously silent ORFs in *Rhizobium* NGR234, and that genes involved in NF biosynthesis were more rapidly induced than some others whose products are required at a later stage of interaction with a host plant [90]. In *B. japonicum*, soybean flavonoid genistein was found to induce flagellar cluster and several genes that are involved in transport processes, in addition to nodulation-associated genes [125]. Similarly, other genes are induced by the perception of the plant flavonoid luteolin in *E. meliloti* besides those related to nodulation [89]. The repertoire of luteolin-regulated genes includes genes encoding the EmrAB efflux system, the conjugal transfer protein TraA, a NTPase essential for the *E. meliloti* infective phenotype, three genes involved in iron metabolism as well as the SyrM transcriptional regulator [89]. Moreover, studies aiming to assess the expression of small non-coding RNA (sRNA) in *E. meliloti* have pointed out that at least three RNA transcripts involved in modulating gene expression are controlled by the flavonoid luteolin [129].

Proteomic analysis has identified new proteins whose expression levels are influenced by the presence of *nod* gene-inducing flavonoids in the bacterial growth medium. Two proteins that did not show sequence matches with any known *nod* gene products were induced in *R. leguminosarum* bv. *trifolii* by dihydroxyflavone [130]. The expression of several proteins that appeared to be encoded by pSymA of *E. meliloti* was positively regulated by luteolin and none of them matched the products of any previously identified luteolin-regulated gene [127]. Other proteins were down-regulated in the presence of

luteolin, or expressed only in the absence of pSymA, or accumulated in maximum amounts when pSymA was either present or absent. At the level of protein expression, it is clear therefore that luteolin exerts both positive and negative regulatory effects on plasmid and chromosomal genes in *E. meliloti*. Two proteins with homologies to a molecular chaperone, GroEL, which is thought to assist partially folded proteins to acquire a correct configuration [131], were up-regulated by luteolin. It was suggested that chaperone up-regulation fulfilled the need for specific folding requirements of other luteolin induced proteins and that another up-regulated, 30S ribosomal protein, was indicative of a luteolin influence on the cell translational machinery [127]. It has been established that flavonoid inducers are also required for the transcription of type III secretion system (TTSS) genes. TTSS genes occur in several plant and animal Gram-negative bacterial pathogens as well as in rhizobia such as *Rhizobium sp.* NGR234 [132], *Mesorhizobium loti* [133]*, Bradyrhizobium japonicum* USDA110 [134], *Rhizobium etli* CFN42 and *Rhizobium fredii* USDA257 [135]. Exceptions are *R. leguminosarum* and *E. meliloti* that lack homologues of genes encoding a TTSS [22]. The TTSS systems are characterized by secretion of proteins into the extracellular environment or directly into eukaryotic cytoplasm when in contact with the host cells, thereby eliciting a defense response [136]. In rhizobia, protein production and secretion by TTSS were dependent on the presence of flavonoid inducers (e.g, genistein, daidzein, luteolin, naringenin) and export occurred without N-terminal processing. Proteins secreted by rhizobial TTSSs were referred as nodulation outer proteins (Nops). TTSS systems, governing the delivery and reception of Nops, appear to make an important contribution to the formation of successful symbiosis by influencing nodulation [137]. TTSS mutants exhibit inconsistent symbiotic phenotypes compared to the wild-type strains, ranging from changes in nitrogen-fixation capacity of nodule to alteration in nodule number and host specificity. To cite a few examples: the wild-type *Rhizobium sp*. NGR234 and wild-type *S. fredii* HH103 form (nonfixing) nodule-like structures on the roots of *Crotalaria juncea* and *Erythrinia variegata*, respectively, but TTSS defective mutants form effective nodules. *B. japonicum* USDA257 displayed an altered phenotype after TTSS disruption, consisting in the ability to efficiency nodulate a soybean cultivar which cannot be nodulated at all by the wild-type strain [135]. Another flavonoid-inducible rhizobial gene, encoding a secreted protein NodO, has been found, but only in *R. leguminosarum* bv. *viciae* [138] and a broad host range strain, *Rhizobium* sp. BR816, isolated from *Leucaena leucocephala* [100]. NodO is a Ca^{2+}-binding protein with partial homology to *Escherichia coli* hemolysin [139] and is released by a different, type I, rhizobial secretion system. NodO can suppress nodulation defects brought about by the absence of fatty acid and carbamoyl NFsubstituens, in several rhizobial species [140]. Original proposals

for the mode of action of NodO invoked a capacity to form ion channels that permit cation movement across and concomitant depolarization of the plasma membrane of plant cells [141]. Such changes are among the first to be observed when roots are challenged with NFs. More recently *nodO* has been identified as a gene that promotes infection thread development in root hairs [142].

1.5.5. Additional bacterial components required in rhizobium-legume symbiosis

Plant flavonoids, in addition to the processes outlined in the above paragraphs, influence the biosynthesis and the structural features of other rhizobial components that contribute to the symbiotic development. Included in this category are the various secreted and surface polysaccharides of rhizobia that fulfill defense functions and influence rhizobial invasion: extracellular polysaccharides (EPS), lipopolysacchardides (LPS), K-antigen polysaccharides (KPS) and cyclic glucans [143;144]. Similar to NF, several EPS, LPS, KPS and cyclic glucans exert their effects on symbiosis in a structurally dependent manner, arguing that they may function as signals between invading bacteria and their hosts [145]. Evidences suggest that bacterial EPSs (Figure 1.14a) suppress a potentially lethal host defense response, and in the absence of EPSs, the unproductive response may cause a block in infection thread formation [146]. The cyclic *β*-glucans (unbranched polymers of glucose) may play a role in modulating a host defense response to bacterial invasion. Specifically, *M. loti* cyclic *β*-glycans are required to suppress high-level production of antimicrobial phytoalexins during symbiotic development with *L. japonicus* [147]. The LPS, as a component of the Gram-negative outer membrane, plays an important role in promoting rhizobial adaptation and persistance within the particular environment of the host cell cytoplasm (Figure 1.14b) [143;148]. Indeed, the rhizobial cell surface is in intimate association with its host throughout symbiotic development but this is particularly true for the microsymbiont within the root nodule (symbiosome).

Flavonoids in some rhizobia appear to brought directly structural changes on LPS [149;150]. In turn, the alteration of the LPS carbohydrate core and content has an aberrant effect in a variety of symbioses. For example, an *E. meliloti lpsB* mutant has a dramatically altered LPS core and is unable to establish a chronic host infection [151]. Moreover, defects in LPS can sensitize bacteria to membrane-disrupting agents and antimicrobial peptides. Thus, the LPS layer may provide a protective barrier against environmental stress and host defense responses. There are indications that the rhizobial LPS may also play an active role by suppressing the release of Reactive Oxygen Species (ROS) [152]. The LPS component of *E. meliloti* is able to suppress the

oxidative burst and to damp the plant transcriptional response [152], indicating that an interaction between the rhizobial LPS and its host plant could suppress any potential immune response to intracellular bacteria. This could be particularly important for bacteroids within the root nodule (symbiosome) as they no longer express genes for the biosynthesis of succinoglycan that appears to dampen a potential plant defense response to bacteria within the infection thread [42].

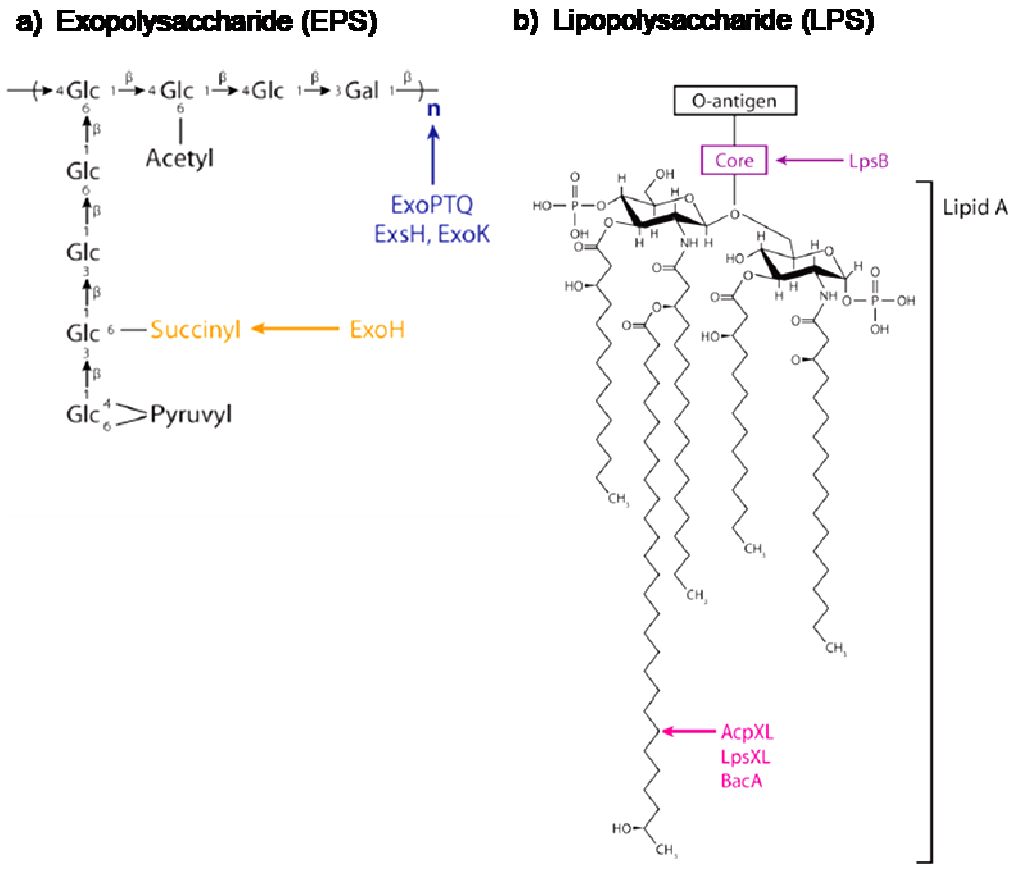


Figure 1.14. Schematic representation of additional rhizobial molecules involved in symbiosis with legume plants and enzymes responsible for the substitutions on the core structure. a) The exopolysaccharide (EPS) and b) lipopolysaccharide (LPS) of *Ensifer meliloti.* ExoH is responsible for succinyl modification; succinoglycan molecular weight is controlled by ExoPTQ and two extracellular glycosylases, ExsH and ExoK. (*d*) Schematic representation of *E. meliloti* lipopolysaccharide (LPS). LpsB is a glycosyltransferase with broad substrate specificity involved in synthesis of the LPS core. AcpXL, LpsXL, and BacA are required for the Very-Long-Chain Fatty Acid (C28) modification of lipid A (modified by [6]).

1.5.6. Bacterial quorum sensing as a strategy to modulate symbiotic interaction

Several studies revealed that many plant associated bacteria (PAB) such as rhizobia coordinate the gene expression in response to changes in bacterial cell density, a process known as *Quorum Sensing* (QS) [153]. QS is employed as a strategy by phytopathogens for the regulation of virulence associated functions and by plant-growth-promoting bacteria for beneficial traits. In the case of rhizobia, QS signaling allows to regulate the expression of important genes for host colonization and invasion to establish compatible association with their hosts [154;155] (Figure 1.15).

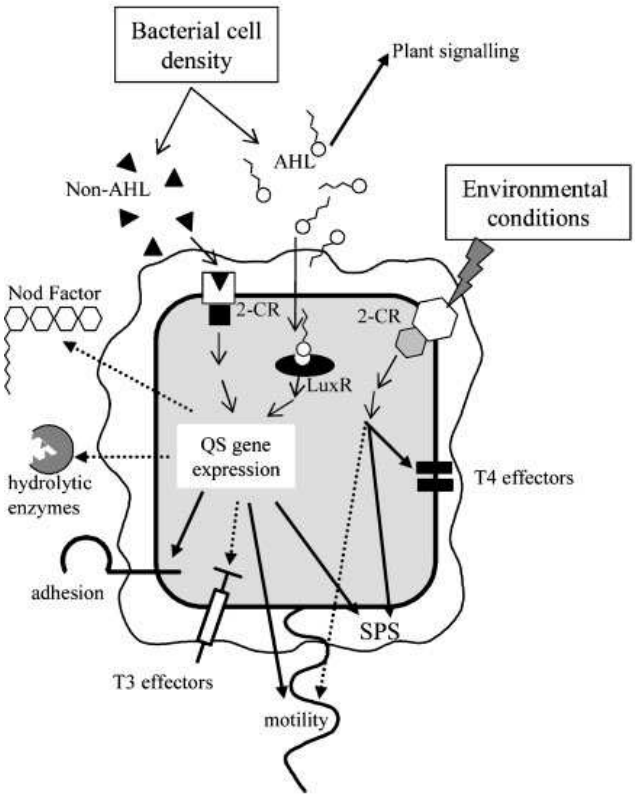


Figure 1.15. Coordination of genes expression for host colonization and invasion mediated by quorum sensing (QS) signals and two component regulatory systems (2-CR). Detection of N-acyl homoserine lactones (AHL, loop and tail) by cytosolic LuxR-type transcriptional activators (black oval) and non-AHL (black triangles) by 2-CR systems (white and black squares), allow plant-interacting bacteria to coordinate the expression of important genes for host colonization and invasion in response to cell density. AHLs play an additional role in plant signaling. Regulation of bacterial factors required during the infection process is also accomplished in plant-interacting bacteria by 2-CR systems (white and grey hexagons), which are activated by environmental conditions usually encountered during the invasion process. Common rhizobia and pathogenic bacteria responses are shown by bold arrows, whereas responses observed only in one or other are represented by dotted arrows (modified by [49]).

The QS is mediated by small diffusible signal molecules referred as autoinducers, which can differ in their chemical and structural properties but they share a common role [156]. Once the threshold concentration of autoinducers is reached, the bacterial population detects the signaling molecules and responds with a population-wide alteration in gene expression. The most common signals of QS belong to the class of acyl-homoserin lactones (AHLs), which contain a conserved homoserin ring tied to a variable acyl chain. The QS signals are detected through two-component regulatory systems. Several AHLs as well as two-component regulatory systems (2-CR) were identified both in bacterial plant pathogens and in rhizobia, and they are essential for a successful interaction with the host plants [157;158]. A canonical QS regulatory system consists of a LuxI-family synthase responsible for synthesizing the AHL signal (autoinducer), which then interacts with the cognate LuxR-family transcription factors (response regulator). The response elicited by AHL signals consists in the expression of hundreds of bacterial genes, including genes responsible for biofilm formation, nodulation, nitrogen-fixation, synthesis of toxins, as well as motility and conjugation [157;158]. Shortcomings in QS lead to a reduction or a loss of virulence in plant pathogenic bacteria and to an alteration in the nodulation efficiency and nitrogen-fixation in rhizobia [154;158;159]. In addition to the common response regulators of QS, other LuxR-type proteins have been found. They were defined as orphan LuxR regulators because they have the same modular structure of QS LuxRs but are devoid of a cognate LuxI synthase associated with them in the genome [160-162]. LuxR orphans have been shown to be responsive to exogenous AHLs produced by neighboring cells as well as to endogenously produced AHLs [163]. It is now also evident that some LuxR orphans proteins have evolved the ability to respond to other molecular signals different from AHLs. Recently a group of LuxR orphans that do not bind AHLs and instead respond to low-molecular weight plant compounds have been discovered in plant associated bacteria (PAB) [164;165]. Then, the QS signaling is not restricted to bacterial cell-to-cell communication, but also allows an interkingdom signaling between microorganisms and their host. On the other hand, plants have been shown to synthesize QS mimics compounds that can both inhibit and stimulate AHL-dependent genes, although most of these remain unidentified [166;167]. The so-far identified mimic signals from plants include flavonoid catechin that are present in the rhizosphere [168]. Another flavonoid with inhibitory effects on QS-regulated genes is naringenin, which is exudated by some legume roots [169]. Naringenin was shown to inhibit QS in *Escherichia coli* and *Vibrio fisheri* as well as *Pseudomonas aeruginosa* [170]. Interestingly, the flavonoid pathway is activated in legume plants by exposure to QS signals from rhizobia. It has been also show that bacterial AHLs can stimulate the production of AHL mimics by *M. truncatula* [171]. Inducing flavonoids were reported to

increase AHL synthesis in the three species of rhizobia, *Sinorhizobium fredii, Rhizobium etli* and *Rhizobium sullae*, concomitant with enhanced expression of AHL synthesis genes [172]. These evidences suggest a link between nodulation genes induction by plant flavonoid, QS activation of the flavonoid pathway and possible feedback on bacteria by production of possible AHL mimics.

References

1. J.R. Postgate, Biological nitrogen fixation, Nature 226 (1970) 25-27.

2. J.I. Sprent and S.M. Defaria (1988), Mechanisms of infection of plants by nitrogen-fixing organisms, Plant and Soil 110:157-165.

3. J.W. Erisman, A. Bleeker, J. Galloway, and M.S. Sutton (2007), Reduced nitrogen in ecology and the environment, Environmental Pollution 150 140-149.

4. H.H. Zahran, Rhizobia from wild legumes: diversity, taxonomy, ecology, nitrogen fixation and biotechnology (2001), Journal of Biotechnology 91 143-153.

5. B. M. Hoffman, D. Lukoyanov, Z. Yang, D.R. Dean and L.C. Seefeldt, Mechanism of nitrogen fixation by nitrogenase: the next stage (2014), Chemical Reviews 114, 4041−4062.

6. K.E. Gibson, H. Kobayashi, and G.C. Walker, Molecular determinants of a symbiotic chronic infection (2008), Annual Review of Genetic 42 413-441.

7. D.A. Relman, 'Til death do us part': coming to terms with symbiotic relationships-Foreword (2008), Nature Reviews Microbiology 6 721-724.

8. S.R. Long, Genes and signals in the rhizobium-legume symbiosis (2001), Plant Physiololy 125 69-72.

9. D.V. Badri, T.L. Weir, D. van der Lelie, and J.M. Vivanco, Rhizosphere chemical dialogues: plant-microbe interactions (2009), Current Opinion in Biotechnology 20 642-650.

10. N. Gruber and J.N. Galloway, An Earth-system perspective of the global nitrogen cycle (2008), Nature 451 293-296.

11. L.A. Newman and C.M. Reynolds, Bacteria and phytoremediation: new uses for endophytic bacteria in plants (2005), Trends in Biotechnology 23 6-8.

12. E. Lodwig and P. Poole, Metabolism of Rhizobium bacteroids, (2003) Critical Reviews in Plant Sciences 22 37-78.

13. R. Oono, R.F. Denison, and E.T. Kiers, Controlling the reproductive fate of rhizobia: how universal are legume sanctions? (2009), New Phytologist 183 967-979.

14. L.R. Boring, W.T. Swank, J.B. Waide, and G.S. Henderson, Sources, fates, and impacts of nitrogen inputs to terrestrial ecosystems-review and synthesis (1988), Biogeochemistry 6 119-159.

15. J. Bouton, The economic benefits of forage improvement in the United States (2007), Euphytica 154 263-270.

16. G. Carlsson and K. Huss-Danell, Nitrogen fixation in perennial forage legumes in the field (2003), Plant and Soil 253 353-372.

17. F. Zakhia and P. de Lajudie, Taxonomy of rhizobia (2001), Agronomie 21 569-576.

18. M. Martens, M. Delaere, R. Coopman, P. De Vos, M. Gillis, and A. Willems, Multilocus sequence analysis of *Ensifer* and related taxa (2007), International Journal of Systematic and Evolutionary Microbiology 57 489-503.

19. P. de Lajudie, E. Laurent-Fulele, A. Willems, U. Torck, R. Coopman, M.D. Collins, K. Kersters, B. Dreyfus, and M. Gillis, *Allorhizobium undicola* gen. nov., sp. nov., nitrogen-fixing bacteria that efficiently nodulate *Neptunia natans* in Senegal (1998), International Journal of Systematic Bacteriology 48 1277-1290.

20. P. Bueno, M.J. Soto, M.P. Rodriguez-Rosales, J. Sanjuan, J. Olivares, and J.P. Donaire, Time-course of lipoxygenase, antioxidant enzyme activities and H_2O_2 accumulation during the early stages of rhizobium-legume symbiosis (2001), New Phytologist 152 91-96.

21. M. McIntosh, S. Meyer, and A. Becker, Novel *Sinorhizobium meliloti* quorum sensing positive and negative regulatory feedback mechanisms respond to phosphate availability (2009), Molecular Microbiology 74 1238-1256.

22. F. Galibert, T.M. Finan, S.R. Long, A. Puhler, P. Abola, F. Ampe et al. The composite genome of the legume symbiont *Sinorhizobium meliloti* (2001), Science 293 668-672.

23. A. Becker, M.J. Barnett, D. Capela, M. Dondrup, P.B. Kamp, E. Krol, B. Linke, S. Ruberg, K. Runte, B.K. Schroeder, S. Weidner, S.N.Yurgel, J .Batut, S.R. Long, A. Puhler, and A. Goesmann, A portal for rhizobial genomes: RhizoGATE integrates a *Sinorhizobium meliloti* genome annotation update with postgenome data (2009), Journal of Biotechnology 140 45-50.

24. M. Galardini, A. Mengoni, M. Brilli, F. Pini, A. Fioravanti, S. Lucas, A. Lapidus, J.F. Cheng et al., Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti* (2011), BMC Genomics 12 235.

25. X. Bailly, I. Olivieri, B. Brunel, J.C. Cleyet-Marel, and G. Bena, Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of *Medicago* species (2007), Journal of Bacteriology 189 5223-5236.

26. D. Medini, C. Donati, H. Tettelin, V. Masignani, and R. Rappuoli, The microbial pangenome (2005), Current Opinion in Genetics & Development 15 589-594.

27. M. Galardini, M. Brilli, G. Spini, M. Rossi, B.Roncaglia, A. Bani, M.C hiancianesi, M. Moretto, K. Engelen, G. Bacci, F. Pini, E.G.Biondi, M. Bazzicalupo, and A.

Mengoni, Evolution of Intra-specific regulatory networks in a multipartite bacterial genome (2015), Plos Computational Biology 11.

28. A.M. Schmitz and M.J. Harrison, Signaling events during initiation of arbuscular mycorrhizal symbiosis (2014), Journal of Integrative Plant Biology 56 250-261.

29. D.J. Gage, Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes (2004), Microbiology and Molecular Biology Reviews 68 280.

30. S.G. Pueppke and W.J. Broughton, *Rhizobium* sp. strain NGR234 and *Rhizobium fredii* USDA257 share exceptionally broad, nested host ranges (1999), Molecular Plant-Microbe Interactions 12 293-318.

31. J. Sprent, 60 Ma of legume nodulation? What's new? What's changing?, Comparative Biochemistry and Physiology A-Molecular & Integrative Physiology 146 (2007) S215-S216.

32. F.G. Barcellos, P. Menna, J.S.D. Batista, and M.Hungria, Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium (Ensifer) fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil (2007), Applied and Environmental Microbiology 73 2635-2643.

33. L. Moulin, A. Munive, B. Dreyfus, and C. Boivin-Masson, Nodulation of legumes by members of the beta-subclass of *Proteobacteria* (2001), Nature 411 948-950.

34. W.M. Chen, L. Moulin, C. Bontemps, P. Vandamme, G.B ena, and C. Boivin-Masson, Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature (2003), Journal of Bacteriology 185 7266-7272.

35. S.R. Long, Rhizobium-legume nodulation: life together in the underground (1989), Cell 56 203-214.

36. J.P. Nap, Developmental biology of a plant-prokaryote symbiosis - the legume root nodule." (1990), Science 948-954.

37. W. Malek, The role of motility in the efficiency of nodulation by *Rhizobium meliloti* (1992), Archives of Microbiology 158 26-28.

38. C.L. Diaz, T.J.J. Logman, H.C. Stam, and J.W. Kijne, Sugar-binding activity of *Pea* lectin expressed in white clover hairy roots (1995), Plant Physiology 109 1167-1177.

39. N.J. Brewin, Plant cell wall remodelling in the rhizobium-legume symbiosis (2004), Critical Reviews in Plant Sciences 23 293-316.

40. B. Lotocka, J. Kopcinska, and M. Skalniak, Review article: The meristem in indeterminate root nodules of *Faboideae* (2012), Symbiosis 58 63-72.

41. V. Oke and S.R. Long, Bacteroid formation in the Rhizobium-legume symbiosis (1999), Current Opinion in Microbiology 2 641-646.

42. K.M. Jones, H. Kobayashi, B.W. Davies, M.E.Taga, and G.C.Walker, How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model (2007), Nature Reviews Microbiology 5 619-633.

43. J. Prell and P. Poole, Metabolic changes of rhizobia in legume nodules (2006), Trends in Microbiology 14 161-168.

44. R.J. Norby, Nodulation and nitrogenase activity in nitrogen-fixing woody-plants stimulated by CO_2 enrichment of the atmosphere (1987), Physiologia Plantarum 71 77-82.

45. J. White, J. Prell, E.K. James, and P. Poole, Nutrient sharing between symbionts (2007), Plant Physiology 144 604-614.

46. J.A. Downie, C.D. Knight, A.W.B. Johnston, and L. Rossen, Identification of genes and gene products involved in the nodulation of *Peas* by *Rhizobium leguminosarum* (1985), Molecular & General Genetics 198 255-262.

47. J.D. Tjepkema and C.S. Yocum, Measurement of oxygen partial pressure within soybean nodules by oxygen microelectrodes (1974), Planta 119 351-360.

48. X.N. Dong, SA, JA, ethylene, and disease resistance in plants (1998), Current Opinion in Plant Biology 1 316-323.

49. M.J. Soto, J. Sanjuan, and J. Olivares, Rhizobia and plant-pathogenic bacteria: common infection weapons (2006), Microbiology 152 3167-3174.

50. N. Pauly, C. Pucciariello, K. Mandon, G. Innocenti, A. Jamet, E. Baudouin, D. Herouart, P. Frendo, and A. Puppo, Reactive oxygen and nitrogen species and glutathione: key players in the legume-Rhizobium symbiosis (2006), Journal of Experimental Botany 57 1769-1776.

51. R.E. Mitchell, E.J. Frey, and M.H. Benn, Rhizobitoxine and L-threo-hydroxythreonine production by the plant pathogen *Pseudomonas andropogonis* (1986), Phytochemistry 25 2711-2715.

52. W.B. Ma, D.M. Penrose, and B.R. Glick, Strategies used by rhizobia to lower plant ethylene levels and increase nodulation (2002), Canadian Journal of Microbiology 48 947-954.

53. R.B.Abramovitch and G.B.Martin, Strategies used by bacterial pathogens to suppress plant defenses (2004), Current Opinion in Plant Biology 7 356-364.

54. M.A. Van Sluys, C.B. Monteiro-Vitorello, L.E.A. Camargo, C.F.M. Menck et al., Comparative genomic analysis of plant-associated bacteria (2002), Annual Review of Phytopathology 40 169-189.

55. R.F. Denison and E.T.K iers, Why are most rhizobia beneficial to their plant hosts, rather than parasitic? (2004), Microbes and Infection 6 1235-1239.

56. P. Mergaert, T. Uchiumi, B. Alunni, G.E vanno, A. Cheron, O. Catrice, A.E. Mausset, F.Barloy-Hubler, F. Galibert, A. Kondorosi, and E. Kondorosi, Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis (2006), Proceedings of the National Academy of Sciences of the United States of America 103 5230-5235.

57. S.A. West, E.T. Kiers, E.L. Simms, and R.F. Denison, Sanctions and mutualism stability: why do rhizobia fix nitrogen? (2002), Proceedings of the Royal Society B-Biological Sciences 269 685-694.

58. E.T. Kiers, R.A. Rousseau, S.A. West, and R.F. Denison, Host sanctions and the legume-rhizobium mutualism (2003), Nature 425 78-81.

59. R.F. Denison and E.T. Kiers, Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis (2004), Fems Microbiology Letters 237 187-193.

60. E.L. Simms and J.D. Bever, Evolutionary dynamics of rhizopine within spatially structured rhizobium populations (1998), Proceedings of the Royal Society B-Biological Sciences 265 1713-1719.

61. E.T. Kiers and R.F. Denison, Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms (2008), Annual Review of Ecology Evolution and Systematics 39 215-236.

62. M. Moawad and E.L. Schmidt, Occurrence and nature of mixed infections in nodules of field-grown soybeans (*Glycine max*) (1987), Biology and Fertility of Soils 5 112-114.

63. X. Perret, C.S taehelin, and W.J. Broughton, Molecular basis of symbiotic promiscuity (2000), Microbiology and Molecular Biology Reviews 64 (2000) 180.

64. J.E. Cooper, Multiple responses of rhizobia to flavonoids during legume root infection (2004), Advances in Botanical Research Incorporating Advances in Plant Pathology, 41 1-62.

65. H.H. Woo, G. Kuleck, A.M. Hirsch, and M.C. Hawes, Flavonoids: Signal molecules in plant development (2002), Flavonoids in Cell Function 505 51-60.

66. B.W. Shirley, Flavonoid biosynthesis: 'New' functions for an 'old' pathway (1996), Trends in Plant Science 1 377-382.

67. U.A. Hartwig and D.A. Phillips, Release and modification of *nod gene* Inducing flavonoids from alfalfa seeds (1991), Plant Physiology 95 804-807.

68. S.H. Lee, D.H. Seo, H.L. Park, Y.J. Choi, and S.H. Jung, Solubility enhancement of a hydrophobic flavonoid, luteolin by the complexation with cyclosophoraoses

isolated from *Rhizobium meliloti* (2003), Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 84 201-207.

69. F.D. Dakora, C.M. Joseph, and D.A. Phillips, Alfalfa (*Medicago Sativa)* root exudates contain isoflavonoids in the presence of *Rhizobium meliloti* (1993), Plant Physiology 101 819-824.

70. S. Latha and A. Mahadevan, Role of rhizobia in the degradation of aromatic substances (1997), World Journal of Microbiology & Biotechnology 13 601-607.

71. S. Vela, M.M. Haggblom, and L.Y. Young, Biodegradation of aromatic and aliphatic compounds by rhizobial species (2002), Soil Science 167 802-810.

72. S. Hassan and U. Mathesius, The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions (2012), Journal of Experimental Botany 63 3429-3444.

73. Y. Masaoka, M. Kojima, S. Sugihara, T. Yoshihara, M. Koshino, and A. Ichihara, Dissolution of ferric phosphate by Alfalfa (*Medicago sativa*) root exudates (1993), Plant and Soil 155 75-78.

74. U. Mathesius, H.R.M. Schlaman, H.P. Spaink, C. Sautter, B.G. Rolfe, and M.A. Djordjevic, Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides (1998), Plant Journal 14 23-34.

75. K.J.M. Boot, A.A.N. van Brussel, T. Tak, H.P. Spaink, and J.W. Kijne, Lipochitin oligosaccharides from *Rhizobium leguminosarum* bv. *viciae* reduce auxin transport capacity in *Vicia sativa* subsp *nigra* roots (1999), Molecular Plant-Microbe Interactions 12 839-844.

76. A.P. Wasson, F.I. Pellerone, and U. Mathesius, Silencing the flavonoid pathway in *Medicago truncatula* inhibits root nodule formation and prevents auxin transport regulation by rhizobia (2006), Plant Cell 18 1617-1629.

77. M. Bacilio-Jimenez, S. guilar-Flores, E. Ventura-Zapata, E. Perez-Campos, S. Bouquelet, and E. Zenteno, Chemical characterization of root exudates from rice (*Oryza sativa*) and their effects on the chemotactic response of endophytic bacteria (2003), Plant and Soil 249 271-277.

78. J.M.M. Aguilar, A.M .Ashby, A.J.M. Richards, G.J. Loake, M.D. Watson, and C.H. Shaw, Chemotaxis of *Rhizobium leguminosarum* biovar *phaseoli* towards flavonoid inducers of the symbiotic nodulation genes(1988), Journal of General Microbiology 134 2741-2746.

79. J.P. Armitage, A. Gallagher, and A.W.B. Johnston, Comparison of the chemotactic behavior of *Rhizobium Leguminosarum* with and without the nodulation plasmid (1988), Molecular Microbiology 2 743-748.

80. R. Kape, M. Parniske, and D. Werner, Chemotaxis and *nod* gene activity of *Bradyrhizobium japonicum* in response to hydroxycinnamic acids and isoflavonoids (1991), Applied Environmental Microbiology 57 316-319.

81. W.W. Currier and G.A. Strobel, Chemotaxis of *Rhizobium spp.* to a glycoprotein produced by birdsfoot trefoil roots (1977), Science 196 434-436.

82. M. Parniske, B. Ahlborn, and D. Werner, Isoflavonoid-inducible resistance to the phytoalexin glyceollin in soybean rhizobia (1991), Journal of Bacteriology 173 3432-3439.

83. A.D. Lameta and M. Jay, Study of Soybean and Lentil Root Exudates .3. Influence of soybean isoflavonoids on the growth of rhizobia and some rhizospheric microorganisms (1987), Plant and Soil 101 267-272.

84. C.C. Lin, Y.C. Chen, S.C. Song, and L.P.Lin, Flavonoids as inducers of extracellular proteins and exopolysaccharides of *Sinorhizobium fredii* (1999), Biology and Fertility of Soils 30 83-89.

85. V. Jain and H.S. Nainawatee, Flavonoids influence growth and saccharide metabolism of *Rhizobium meliloti* (1999), Folia Microbiologica 44 311-316.

86. N.K. Peters, J.W. Frost, and S.R. Long, A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes, Science 233 (1986) 977-980.

87. J.W. Redmond, M. Batley, M.A. Djordjevic, R.W. Innes, P.L. Kuempel, and B.G. Rolfe, Flavones induce expression of nodulation genes in rhizobium (1986), Nature 323 632-635.

88. M.A. Schell, Molecular biology of the LysR family of transcriptional regulators (1993), Annual Review of Microbiology 47 597-626.

89. D. Capela, S. Carrere, and J. Batut, Transcriptome-based identification of the *Sinorhizobium meliloti* NodD1 regulon (2005), Appl. Environ. Microbiol. 71 4910-4913.

90. H. Kobayashi, Y. Naciri-Graven, W.J. Broughton, and X. Perret, Flavonoids induce temporal shifts in gene expression of *nod* box controlled loci in *Rhizobium* sp NGR234 (2004), Molecular Microbiology 51 335-347.

91. F.Q. Li, B.H. Hou, L. Chen, Z.J. Yao, and G.F. Hong, *In vitro* observation of the molecular interaction between NodD and its inducer naringenin as monitored by fluorescence resonance energy transfer (2008), Acta Biochimica et Biophysica Sinica 40 783-789.

92. M.A. Djordjevic, P.R. Schofield, and B.G. Rolfe, Tn5 mutagenesis of *Rhizobium trifolii* host-specific nodulation genes result in mutants with altered host-range ability (1985), Molecular & General Genetics 200 463-471.

93. M. Hungria, A.W.B.J ohnston, and D.A. Phillips, Effects of flavonoids released naturally from bean (*Phaseolus vulgaris*) on NodD regulated gene transcription in *Rhizobium leguminosarum* bv *Phaseoli* (1992), Molecular Plant-Microbe Interactions 5 199-203.

94. J.A. Swanson, J.T. Mulligan, and S.R.Long, Regulation of SyrM and NodD3 in *Rhizobium meliloti* (1993), Genetics 134 435-444.

95. J. Loh, M. Garcia, and G. Stacey, NodV and NodW, a second flavonoid recognition system regulating *nod* gene expression in *Bradyrhizobium japonicum* (1997), Journal of Bacteriology 179 3013-3020.

96. C.M. Bellato, P.A. Balatti, S.G. Pueppke, and H.B.Krishnan, Proteins from cells of *Rhizobium fredii* bind to DNA sequences preceding *noIX*, a flavonoid-inducible *nod* gene that is not associated with a *nod* box (1996), Molecular Plant-Microbe Interactions 9 457-463.

97. H.R. Schlaman, B.J. Lugtenberg, and R.J. Okker, The NodD protein does not bind to the promoters of inducible nodulation genes in extracts of bacteroids of *Rhizobium leguminosarum* biov*ar viciae* (1992), Journal of Bacteriology 174 6109-6116.

98. A. Savoure, C. Sallaud, J. ElTurk, J. Zuanazzi, P. Ratet, M. Schultze, A. Kondorosi, R. Esnault, and E. Kondorosi, Distinct response of *Medicago* suspension cultures and roots to Nod factors and chitin oligomers in the elicitation of defense-related responses (1997), Plant Journal 11 277-287.

99. L. Rossen, C.A. Shearman, A.W. Johnston, and J.A. Downie, The *nodD* gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the *nodABC* genes (1985), EMBO Journal 4 3369-3373.

100. P.Vanrhijn, E.Luyten, K.Vlassak, and J.Vanderleyden, Isolation and characterization of a pSym locus of *Rhizobium sp* BR816 that extends nodulation ability of narrow host range *Phaseolus vulgaris* symbionts to L*eucaena leucocephala* (1996), Molecular Plant-Microbe Interactions 9 74-77.

101. K.K. Lestrange, G.L. Bender, M.A. Djordjevic, B.G. Rolfe, and J.W. Redmond, The Rhizobium strain NGR234 *nodD1* gene product responds to activation by the simple phenolic compounds vanillin and isovanillin present in wheat seedling extracts (1990), Molecular Plant-Microbe Interactions 3 214-220.

102. K. Goethals, M. Gao, K. Tomekpe, M. Vanmontagu, and M. Holsters, Common *nodABC* genes in *nod* locus of *Azorhizobium caulinodans* - Nucleotide sequence and plant-Inducible expression (1989), Molecular & General Genetics 289-298.

103. B. Horvath, E. Kondorosi, M.J ohn, J. Schmidt, I. Torok, Z. Gyorgypal, I. Barabas, U. Wieneke, J. Schell, and A. Kondorosi, Organization, structure and symbiotic function of *Rhizobium Meliloti* nodulation genes determining host specificity for alfalfa (1986), Cell 46 335-343.

104. F. Debelle, C. Rosenberg, J. Vasse, F. Maillet, E. Martinez, J. Denarie, and G. Truchet, Assignment of symbiotic developmental phenotypes to common and specific nodulation (*nod*) genetic loci of *Rhizobium meliloti* (1986), Journal of Bacteriology 168 1075-1086.

105. C. Faucher, S. Camut, J. Denarie, and G. Truchet, The *nodH* and *nodQ* host range genes of *Rhizobium meliloti* behave as virulence genes in *R. leguminosarum* bv *viciae* and determine changes in the production of plant-specific extracellular signals (1989), Molecular Plant-Microbe Interactions 2 291-300.

106. P. Mylona, K. Pawlowski, and T. Bisseling, Symbiotic nitrogen fixation (1995), Plant Cell 7 869-885.

107. M.J. Merrick and R.A. Edwards, Nitrogen control in bacteria (1995), Microbiology Review 59 604-622.

108. J.A.S. Zuanazzi, P.H. Clergeot, J.C. Quirion, H.P. Husson, A. Kondorosi, and P. Ratet, Production of *Sinorhizobium meliloti nod* gene activator and repressor flavonoids from *Medicago sativa* roots (1998), Molecular Plant-Microbe Interactions 11 784-794.

109. N.K. Peters and S.R. Long, Alfalfa root exudates and compounds which promote or Inhibit induction of *Rhizobium meliloti* nodulation genes (1988), Plant Physiology 88 396-400.

110. H.P. Spaink and B.J.J. Lugtenberg, Role of rhizobial lipo-chitin oligosaccharide signal molecules in root-nodule organogenesis (1994), Plant Molecular Biology 26 1413-1422.

111. B. Relic, X. Perret, M.T. Estradagarcia, J. Kopcinska, W. Golinowski, H.B. Krishnan, S.G. Pueppke, and W.J. Broughton, Nod factors of rhizobium are a key to the legume door (1994), Molecular Microbiology 13 171-178.

112. J.E. Cooper, Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue (2007), Journal of Applied Microbiology 103 1355-1365.

113. N.A. Fujishige, M.R.Lum, P.L.De Hoff, J.P.Whitelegge, K.F.Faull, and A.M.Hirsch, Rhizobium common *nod* genes are required for biofilm formation (2008), Molecular Microbiology 67 504-515.

114. E. Limpens and T. Bisseling, Signaling in symbiosis (2003), Current Opinion in Plant Biology 6 343-350.

115. S. Radutoiu, L.H. Madsen, E.B. Madsen, A. Jurkiewicz, E. Fukai, E.M.H. Quistgaard, A.S. Albrektsen, E.K. James, S. Thirup, and J. Stougaard, LysM domains mediate lipochitin-oligosaccharide recognition and *Nfr* genes extend the symbiotic host range (2007), Embo Journal 26 3923-3935.

116. L.J. Shaw, P. Morris, and J.E. Hooker, Perception and modification of plant flavonoid signals by rhizosphere microorganisms (2006), Environmental Microbiology 8 1867-1880.

117. E. Jumas-Bilak, S. Michaux-Charachon, G. Bourg, M. Ramuz, and A.lardet-Servent, Unconventional genomic organization in the alpha subgroup of the *Proteobacteria* (1998), Journal of Bacteriology 180 2749-2755.

118. J. Goedhart, J.J. Bono, T. Bisseling, and T.W.J. Gadella, Identical accumulation and immobilization of sulfated and nonsulfated Nod factors in host and nonhost root hair cell walls (2003), Molecular Plant-Microbe Interactions 16 884-892.

119. W. D'Haeze and M. Holsters, Nod factor structures, responses, and perception during initiation of nodule development (2002), Glycobiology 12 79R-105R.

120. G.E.D. Oldroyd and J.M. Downie, Coordinating nodule morphogenesis with rhizobial infection in legumes (2008), Annual Review of Plant Biology 59 519-546.

121. D.P. Lohar, N. Sharopova, G. Endre, S. Penuela, D. Samac, C. Town, K.A.T. Silverstein, and K.A. VandenBosch, Transcript analysis of early nodulation events in *Medicago truncatula* (2006), Plant Physiology 140 221-234.

122. F.Foucher and E.Kondorosi, Cell cycle regulation in the course of nodule organogenesis in *Medicago* (2000), Plant Molecular Biology 43 773-786.

123. F. Ampe, E. Kiss, F. Sabourdy, and J. Batut, Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis (2003), Genome Biology 4 (2003) R15.

124. M.J. Barnett, C.J. Tolman, R.F. Fisher, and S.R. Long, A dual-genome symbiosis chip for coordinate study of signal exchange and development in a prokaryote-host interaction (2004), Proceedings of the National Academy of Sciences of the United States of America 101 16636-16641.

125. K. Lang, A. Lindemann, F. Hauser, and M.G ottfert, The genistein stimulon of *Bradyrhizobium japonicum* (2008), Molecular Genetics and Genomics 279 203-211.

126. J.S.D. Batista and M. Hungria, Proteomics reveals differential expression of proteins related to a variety of metabolic pathways by genistein-induced *Bradyrhizobium japonicum* strains (2012), Journal of Proteomics 75 1211-1219.

127. H.C. Chen, J. Higgins, I.J. Oresnik, M.F. Hynes, S. Natera, M.A. Djordjevic, J.J. Weinman, and B.G. Rolfe, Proteome analysis demonstrates complex replicon and luteolin interactions in pSymA-cured derivatives of *Sinorhizobium meliloti* strain 2011 (2000), Electrophoresis 21 3833-3842.

128. X. Perret, C. Freiberg, A. Rosenthal, W.J. Broughton, and R. Fellay, High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium sp.* NGR234 (1999), Molecular Microbiology 32 415-425.

129. C. del Val, E. Rivas, O. Torres-Quesada, N. Toro, and J.I. Jimenez-Zurdo, Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics (2007), Molecular Microbiology 66 1080-1091.

130. N. Guerreiro, J.W. Redmond, B.G. Rolfe, and M.A.D jordjevic, New Rhizobium leguminosarum flavonoid-induced proteins revealed by proteome analysis of differentially displayed proteins (1997), Molecular Plant-Microbe Interactions 10 506-516.

131. J. Ogawa and S.R. Long, The Rhizobium-Meliloti Groelc Locus Is Required for Regulation of early *nod* genes by the transcription activator NodD (1995), Genes & Development 9 714-729.

132. V. Viprey, A. Del Greco, W. Golinowski, W.J. Broughton, and X. Perret, Symbiotic implications of type III protein secretion machinery in rhizobium (1998), Molecular Microbiology 28 1381-1389.

133. T.Kaneko, Y.Nakamura, S.Sato, E.Asamizu, T.Kato, S.Sasamoto et al., Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti* (2000), Dna Research 7 (2000) 331-338.

134. A. Krause, A. Doerfel, and M. Gottfert, Mutational and transcriptional analysis of the type III secretion system of *Bradyrhizobium japonicum* (2002), Molecular Plant-Microbe Interactions 15 1228-1235.

135. H.B. Krishnan, J. Lorio, W.S. Kim, G.Q. Jiang, K.Y. Kim, M. DeBoer, and S.G. Pueppke, Extracellular proteins involved in soybean cultivar-specific nodulation are associated with pilus-like surface appendages and exported by a type III protein secretion system in *Sinorhizobium fredii* USDA257 (2003), Molecular Plant-Microbe Interactions 16 617-625.

136. G.R. Cornelis, Type III secretion: a bacterial device for close combat with cells of their eukaryotic host (2000), Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 355 681-693.

137. C. Marie, W.J. Deakin, V. Viprey, J. Kopcinska, W. Golinowski, H.B. Krishnan, X. Perret, and W.J. Broughton, Characterization of Nops, nodulation outer proteins, secreted via the type III secretion system of NGR234 (2003), Molecular Plant-Microbe Interactions 16 743-751.

138. R.A. Demaagd, A.H.M. Wijfjes, H.P.S paink, J.E. Ruizsainz, C.A. Wijffelman, R.J.H. Okker, and B.J.J. Lugtenberg, NodO, a new *nod* gene of the *Rhizobium leguminosarum* biovar *viciae* Sym Plasmid-Prl1Ji, encodes a secreted protein (1989), Journal of Bacteriology 171 6764-6770.

139. J.M. Sutton, E.J.A. Lea, and J.A. Downie, The Nodulation-signaling protein NodO from *Rhizobium leguminosarum* biovar *viciae* forms Ion channels in membranes

(1994), Proceedings of the National Academy of Sciences of the United States of America 91 9990-9994.

140. K.M. Vlassak, E. Luyten, C. Verreth, P. van Rhijn, T.Bisseling, and J.Vanderleyden, The *Rhizobium* sp. BR816 *nodO* gene can function as a determinant for nodulation of *Leucaena leucocephala, Phaseolus vulgaris*, and *Trifolium repens* by a diversity of *Rhizobium* spp (1998), Molecular Plant-Microbe Interactions 11 383-392.

141. A. Economou, A.E. Davies, A.W.B. Johnston, and J.A. Downie, The *Rhizobium leguminosarum* biovar *viciae nodO* gene can enable a *nodE* mutant of *Rhizobium leguminosarum* biovar *trifolii* to nodulate vetch (1994), Microbiology-Uk 140 2341-2347.

142. S.A. Walker and J.A. Downie, Entry of *Rhizobium leguminosarum* bv. *viciae* into root hairs requires minimal Nod factor specificity, but subsequent infection thread growth requires *nodO* or *nodE* (2000), Molecular Plant-Microbe Interactions 13 754-762.

143. A. Becker, N. Fraysse, and L. Sharypova, Recent advances in studies on structure and symbiosis-related function of rhizobial K-antigens and lipopolysaccharides (2005), Molecular Plant-Microbe Interactions 18 899-905.

144. A. Skorupska, M. Janczarek, M. Marczak, A. Mazur, and J.Krol, Rhizobial exopolysaccharides: genetic control and symbiotic functions (2006), Microbial Cell Factories 5.

145. S. Simsek, T. Ojanen-Reuhs, S.B. Stephens, and B.L. Reuhs, Strain-ecotype specificity in *Sinorhizobium meliloti-Medicago truncatula* symbiosis is correlated to succinoglycan oligosaccharide structure (2007), Journal of Bacteriology 189 7733-7740.

146. H.P. Cheng and G.C. Walker, Succinoglycan is required for initiation and elongation of infection threads during nodulation of *alfalfa* by *Rhizobium meliloti* (1998), Journal of Bacteriology 180 5183-5191.

147. A.L. D'Antuono, T. Ott, L. Krusell, V. Voroshilova, R.A. Ugalde, M .Udvardi, and V.C. Lepek, Defects in rhizobial cyclic glucan and lipopolysaccharide synthesis alter legume gene expression during nodule development (2008), Molecular Plant-Microbe Interactions 21 50-60.

148. C. Erridge, E. nett-Guerrero, and I.R. Poxton, Structure and function of lipopolysaccharides (2002), Microbes and Infection 4 837-851.

149. K.D. Noel, D.M. Duelli, H. Tao, and N.J. Brewin, Antigenic change in the lipopolysaccharide of *Rhizobium etli* CFN42 induced by exudates of *Phaseolus vulgaris* (1996), Molecular Plant-Microbe Interactions 9 180-186.

150. N. Fraysse, F. Couderc, and V. Poinsot, Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis (2003), European Journal of Biochemistry 270 1365-1380.

151. G.R.O. Campbell, B.L. Reuhs, and G.C. Walker, Chronic intracellular infection of alfalfa nodules by *Sinorhizobium meliloti* requires correct lipopolysaccharide core (2002), Proceedings of the National Academy of Sciences of the United States of America 99 3938-3943.

152. H. Scheidle, A. Gross, and K. Niehaus, The Lipid A substructure of the *Sinorhizobium meliloti* lipopolysaccharides is sufficient to suppress the oxidative burst in host plants (2005), New Phytologist 165 559-565.

153. C. Fuqua, S.C. Winans, and E.P. Greenberg, Census and consensus in bacterial ecosystems: The LuxR-LuxI family of quorum sensing transcriptional regulators (1996), Annual Review of Microbiology 50 727-751.

154. J.E. Gonzalez and N.D. Keshavan, Messing with bacterial quorum sensing (2006), Microbiology and Molecular Biology Reviews 70 859-+.

155. H.H. Hoang, A. Becker, and J.E. Gonzalez, The LuxR homolog ExpR, in combination with the *Sin* quorum sensing system, plays a central role in *Sinorhizobium meliloti* gene expression (2004), Journal of Bacteriology 186 5460-5472.

156. M. Sanchez-Contreras, W.D. Bauer, M.S. Gao, J.B. Robinson, and J.A. Downie, Quorum sensing regulation in rhizobia and its role in symbiotic interactions with legumes (2007), Philosophical Transactions of the Royal Society B-Biological Sciences 362 1149-1163.

157. J.E. Gonzalez and M.M. Marketon, Quorum sensing in nitrogen-fixing rhizobia (2003), Microbiology and Molecular Biology Reviews 67 574-+.

158. S.B. von Bodman, W.D. Bauer, and D.L. Coplin, Quorum sensing in plant-pathogenic bacteria (2003), Annual Review of Phytopathology 41 455-482.

159. J. Loh, E.A. Pierson, L.S. Pierson, G. Stacey, and A. Chatterjee, Quorum sensing in plant-associated bacteria (2002), Current Opinion in Plant Biology 5 285-290.

160. A.V.Patankar and J.E.Gonzalez, An Orphan LuxR Homolog of *Sinorhizobium meliloti* affects stress adaptation and competition for nodulation (2009), Applied and Environmental Microbiology 75 946-955.

161. A.V. Patankar and J.E. Gonzalez, Orphan LuxR regulators of quorum sensing (2009), Fems Microbiology Reviews 33 739-756.

162. S. Subramoni and V. Venturi, LuxR-family 'solos': bachelor sensors/regulators of signaling molecules (2009), Microbiology-Sgm 155 1377-1385.

163. B.M.M. Ahmer, Cell-to-cell signalling in *Escherichia coli* and *Salmonella enterica* (2004), Molecular Microbiology 52 933-945.

164. J.F. Gonzalez and V.Venturi, A novel widespread interkingdom signaling circuit (2013), Trends in Plant Science 18 167-174.

165. V. Venturi and C. Fuqua, Chemical signaling between plants and plant-pathogenic bacteria (2013), Annual Review of Phytopathology, 51 (2013) 17-37.

166. M. Teplitski, J.B. Robinson, and W.D. Bauer, Plants secrete substances that mimic bacterial N-acyl homoserine lactone signal activities and affect population density-dependent behaviors in associated bacteria (2000), Molecular Plant-Microbe Interactions 13 637-648.

167. M.S. Gao, M. Teplitski, J.B. Robinson, and W.D. Bauer, Production of substances by *Medicago truncatula* that affect bacterial quorum sensing (2003), Molecular Plant-Microbe Interactions 16 827-834.

168. O.M. Vandeputte, M. Kiendrebeogo, S. Rajaonson, B. Diallo, A. Mol, M. El Jaziri, and M. Baucher, Identification of catechin as one of the flavonoids from *Combretum albiflorum* bark extract that reduces the production of quorum sensing controlled virulence factors in *Pseudomonas aeruginosa* PAO1 (2010), Applied and Environmental Microbiology 76 243-253.

169. K. Novak, P.C hovanec, V. Skrdleta, M. Kropacova, L.Lisa, and M.Nemcova, Effect of exogenous flavonoids on nodulation of pea (*Pisum sativum L*.) (2002), Journal of Experimental Botany 53 1735-1745.

170. A.Vikram, G.K. Jayaprakasha, P.R. Jesudhasan, S.D.Pillai, and B.S.Patil, Suppression of bacterial cell-cell signalling, biofilm formation and type III secretion system by citrus flavonoids (2010), Journal of Applied Microbiology 109 515-527.

171. U. Mathesius, S. Mulders, M.S. Gao, M. Teplitski, G. Caetano-Anolles, B.G. Rolfe, and W.D. Bauer, Extensive and specific responses of a eukaryote to bacterial quorum-sensing signals (2003), Proceedings of the National Academy of Sciences of the United States of America 100 1444-1449.

172. F .Perez-Montano, B. Guasch-Vidal, S. Gonzalez-Barroso et al., Nodulation-gene-inducing flavonoids increase overall production of autoinducers and expression of N-acyl homoserine lactone synthesis genes in rhizobia (2011), Research in Microbiology 162 715-723.

Chapter 2

Aim

The symbiosis between soil rhizobia and host legumes is of great importance at environmental, agricultural and ecological level. The potential exploitation and improvement of the nitrogen-fixing rhizobia in symbiosis with legume plants to increase agriculture productivity is one of the main focus of the worldwide scientific community. The nitrogen-fixing rhizobia are capable to access atmospheric N that represents a renewable and environmentally sustainable source of nitrogen. Moreover, the N_2 fixing is a cheaper and very effective natural effective agronomic practice to ensure an adequate supply of N in agroecosystems than the application of nitrogen fertilizers.

The establishment of a successful rhizobium-legume symbiosis depends on a complex molecular signal exchange between the two partners. In the early stage, it involves the release of flavonoids from plant roots, which in turn induce the expression of rhizobium nodulation (*nod)* genes required for both root infection and nodule development. In particular, the flavonoid luteolin is the key plant inducer for *Ensifer meliloti,* which represents a model bacterium for rhizobium-legume symbiosis. A more comprehensive understanding of the cascade triggered by the plant signal luteolin should be exploited to improve the symbiosis performance of the bacterial partner. This could potentially lead to an increase of the efficiency of the nitrogen-fixing process itself for agricultural applications. To date, the plant flavonoid luteolin has been the subject of a number of transcriptional and molecular studies, which led to the characterization of luteolin-triggered bacterial response mainly at the transcriptional level. However, the effective role of luteolin on bacterial physiology cannot be fully described using only molecular approach, because a number of the genes modulated by luteolin still have unknown function. In addition, the plant signal luteolin could also exert regulatory activity at post-transcriptional level. Therefore, a global analysis of phenotypic responses induced by luteolin in *Ensifer meliloti* is still lacking. The present work was mainly addressed to provide an extensive phenotypic investigation of the luteolin effects on the phenotypes of *E. meliloti* to get an interpretative framework in modeling luteolin-induced metabolic switches. In the context of the luteolin-responsive phenotypes, the changes dependent or independent from the NodD regulation (i.e the major luteolin sensor) were elucidated using a deletion *nodD* mutant of *E. meliloti* and the possible contribution of other luteolin mediators, beyond NodD, was then investigated.

Chapter 3

Effect of the plant flavonoid luteolin on *Ensifer meliloti* 3001 phenotypic responses

Published on *Plant and Soil*, 2015, DOI 10.1007/s11104-015-2659-2.

Plant Soil
DOI 10.1007/s11104-015-2659-2

REGULAR ARTICLE

Effect of the plant flavonoid luteolin on *Ensifer meliloti* 3001 phenotypic responses

G. Spini · F. Decorosi · M. Cerboneschi · S. Tegli · A. Mengoni · C. Viti · L. Giovannetti

Received: 2 May 2015 / Accepted: 31 August 2015

Abstract
Background and aims The establishment of a successful symbiosis between the nitrogen-fixing bacterium *Ensifer meliloti* and compatible host legumes (*Medicago spp.*) depends on a complex molecular signal exchange. The early stage of signaling involves the release from plant roots of the flavonoid luteolin, which in turn induces the expression of rhizobia nodulation (*nod*) genes required for root infection and nodule development. To date, the bacterial response to the luteolin perception has been characterized in detail as far as gene expression is concerned. Nevertheless, despite this molecular information, a global view on *E. meliloti* phenotypes affected by the plant signal luteolin is still lacking. Therefore, an extensive phenotypic investigation of luteolin effect on the nitrogen-fixing *E. meliloti* 3001 has been performed.
Methods A thousand different growth conditions, including sensitivity to osmolites, pH stresses, antibiotics and toxic compounds, were tested by the application of the high-throughput Phenotype MicroArray (PM) technology, as well as by several dedicated assays to evaluate growth stimulation, motility, biofilm formation, N-acyl homoserine lactones and Indole-3- acetic acid (IAA) production.
Results Results revealed that the plant signal luteolin affected a wide spectrum of *E. meliloti* 3001 phenotypes. *E. meliloti* 3001 displayed an enhanced resistance phenotype in the presence of luteolin toward a broad set of chemicals including several antibiotics, toxic ions, respiration inhibitors, membrane damagers, DNA intercalants and other potential antimicrobial agents. Moreover, the presence of luteolin significantly reduced overall AHLs production, as well as the lag phase in relation to the starting cellular density, the motility and biofilm formation under nutrient-limited growth conditions. An effect on *E. meliloti* indole-3-acetic acid (IAA) production was also detected in vitro as a response to luteolin.
Conclusions Overall, these findings suggest that the plant signal luteolin triggers a broad response in *E. meliloti* 3001, which was shown to be dependent on nutritional conditions sensed by the bacterium, pointing out a wide role in modifying rhizobial phenotypes, possibly in relation to plant root association and then symbiotic interaction.

Responsible Editor: Katharina Pawlowski.

Electronic supplementary material The online version of this article (doi:10.1007/s11104-015-2659-2) contains supplementary material, which is available to authorized users.

G. Spini · F. Decorosi · C. Viti (✉) · L. Giovannetti
Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente (DISPAA) – Laboratorio Genexpress, University of Florence, Florence, Italy
e-mail: carlo.viti@unifi.it

M. Cerboneschi · S. Tegli
Dipartimento di Scienze delle Produzioni Agroalimentari e dell'Ambiente (DISPAA), Sezione Patologia Vegetale, University of Florence, Florence, Italy

A. Mengoni
Dipartimento di Biologia, University of Florence, Florence, Italy

Published online: 30 September 2015

Keywords Flavonoid luteolin · *Ensifer meliloti* · Rhizobium-legume symbiosis · Long-chain N-acyl homoserine lactones (AHLs) · Phenotype MicroArray (PM) · Motility

Introduction

Rhizobia are well known for their ability to establish a successful nitrogen-fixing symbiosis with legume plants by root infection and formation of specialized structures known as root nodules (Young and Johnston 1989). Within these nodules, rhizobia differentiate into a bacteroid form that is able to convert the atmospheric nitrogen (N_2) into ammonia (NH_3), thus metabolically available for the plant. In exchange, the host plant supplies rhizobia with carbon compounds and other nutrients. As a result of this interaction, rhizobia can play a significant role in promoting plant growth and in improving the fertility of low-N soils (Gibson et al. 2008; Long 2001).

The rhizobia-legumes symbiosis development depends on a complex molecular signal exchange between the two partners that, in the early stages, involves the release from plant roots of flavonoids, which are essential for a successful infection and serve as key signals for the organogenesis of nitrogen-fixing nodules (Cooper 2004; Cooper 2007; Perret et al. 2000; Shaw et al. 2006). This finding is definitely intriguing when considering that the flavonoids generally have a fundamental role in protecting higher plants from biotic and abiotic stresses (Agati and Tattini 2010; Winkel-Shirley 2002). As demonstrated also in other rhizospheric microorganisms, plant-derived flavonoids play multiple roles, depending on their structure, such as to inhibit several phytopathogens, to stimulate mycorrhizal spore germination and hyphal branching, to mediate allelopathic interactions, and to chelate soil nutrients (Cooper 2004). Plant flavonoids have been shown to evoke a strong chemoattractant response toward plant roots in rhizobia (Caetano-Anolles et al. 1988). Subsequently, flavonoids are perceived by the rhizobial regulator NodD, a transcription factor belonging to the LysR family (Peck et al. 2006). In turn, NodD induces the expression of the nodulation genes (*nod* genes) implicated in biosynthesis of lipo-chito-oligosaccharidic signaling compounds, which are known as Nod factors (NFs) (Denarie and Cullimore 1993; Spaink 1996). These bacterial lipo-chito-oligosaccharidic signaling compounds redirect root hair growth to support rhizobial cell entry via infection threads, leading to nodule organogenesis. Flavonoids were also demonstrated to be involved in the production of symbiotically active exopolysaccharides as well as in changes occurring on surface polysaccharides (Ardissone et al. 2011; Broughton et al. 2006; Le Quere et al. 2006; Lopez-Baena et al. 2008; Schmeisser et al. 2009). Moreover, the biosynthesis of rhizobial proteins secreted via the Type III Secretion System (Galan et al. 2014), which are called Nops (nodulation outer proteins), is also induced by plant flavonoids.

This wide repertoire of adaptive changes concerning rhizobial gene expression and physiology towards the establishment of a successful symbiosis must be tightly concerted and temporally coordinated, accordingly to the density of bacterial population and to the presence of several host molecular signals. A major portion of these processes are regulated and modulated through the Quorum Sensing (QS) system, which in *Ensifer meliloti* (formerly *Sinorhizobium meliloti*) is composed by the response regulators SinR and ExpR, and the autoinducer synthase SinI. SinI is the enzyme responsible for catalyzing the synthesis of the diffusible autoinducer signals N-acyl homoserine lactones (AHLs) (Llamas et al. 2004; Sanchez-Contreras et al. 2007). The SinR and AHL-activated ExpR mediate positive and negative regulatory feedback mechanisms that modulate the *sinI* expression and consequently the AHLs levels (McIntosh et al. 2009). *SinI* expression requires the transcription activator SinR and is strongly enhanced by the LuxR-type regulator ExpR in the presence of AHLs, resulting in a positive feedback. ExpR also represses transcription of *sinR* at high AHL levels, providing a negative feedback regulation of *sinI* (Krol and Becker 2014).

Plant flavonoids in the rhizosphere have been demonstrated to increase the production of AHLs by rhizobia (Perez-Montano et al. 2011). AHLs in turn trigger specific responses in the host to positively influence nodule numbers (Veliz-Vallejos et al. 2014). *E. meliloti*, a Gram-negative nitrogen-fixing proteobacterium that is distributed worldwide in temperate soils both in free-living and symbiotic form, is considered a model bacterium for legume-rhizobium symbiosis (McIntosh et al. 2009). *E. meliloti* specifically establishes symbiosis with species belonging to three genera of leguminous plants (*Melilotus*, *Medicago*, *Trigonella*), but can be also found on other host species (Mnasri et al. 2009). In the early stages of symbiosis plant roots release a cocktail of nodulation-inducing molecules composed predominantly by flavonoids

(Peck et al. 2006). In particular, the flavone luteolin was found to be the most active plant inducer of the *E. meliloti nod* genes (Honma et al. 1990; Peters et al. 1986). The repertoire of luteolin-regulated genes in *E. meliloti* was identified by a combination of computational predictions (e.g. sequence homologies together with structural conservation analyses) and experimental approaches (Fisher and Long 1993; Peck et al. 2006; Peck et al. 2013; Roux et al. 2014; Schluter et al. 2013; Tolin et al. 2013). Genes differentially expressed in response to the plant luteolin, as well as those related to the various stages of the symbiotic interaction, were identified (Ampe et al. 2003; Barnett et al. 2004; Becker et al. 2014; Roux et al. 2014). Notably, the NodD regulon, comprising luteolin-inducible and NodD-dependent genes (Barnett et al. 2004; Capela et al. 2005), showed that the biosynthesis of NFs is one of the major pathways induced by the luteolin, but not the only one. Indeed, other additional genes are induced by the perception of plant flavonoid signals or show in their promoter region putative binding sites for NodD besides those related to NF biosynthesis (Batista and Hungria 2012; Galardini et al. 2011; Lang et al. 2008; Guerreiro et al. 1997; Tolin et al. 2013). Similarly to other plant flavonoids, luteolin was found to induce in *E. meliloti*, in addition to *nod* genes, the expression of genes encoding the EmrAB efflux pump, the GroES and GroEL chaperonins, as well as the SyrM transcriptional regulator and three hypothetical proteins (Capela et al. 2005). Five other genes were also significantly overexpressed following luteolin treatment, coding for the conjugal transfer protein TraA, for a NTPase essential for the *E. meliloti* infective phenotype, and three genes involved in iron metabolism (Ampe et al. 2003). Furthermore, luteolin induced the accumulation of three small non-coding RNAs in *E. meliloti*, probably having a regulatory function, but whose biological role remains to be explored (del Val et al. 2007).

Overall these data suggest that the plant signal luteolin induces the expression of a number of rhizobial genes and sRNA whose functions are unknown or still to be examined in detailed (Jimenez-Zurdo et al. 2013; Perez-Montano et al. 2011; Perret et al. 2000). Despite the valuable molecular information outlined above, a global analysis of the impact of luteolin on *E. meliloti* metabolic phenotypes is still lacking. Therefore, the purpose of the present study was to provide an extensive investigation on the effects of luteolin on the phenotypic changes occurring in *E. meliloti*, to get at an interpretative framework for modeling luteolin-induced metabolic switches. The model of our investigation was the *E. meliloti* 3001 (*expR*$^{+}$) strain (Nogales et al. 2012), which is derived from *E. meliloti* 2011 (*expR*$^{-}$) and harbors a complete ExpR/Sin QS system, enabling us to investigate the effect of luteolin on QS related phenotypes.

Material and methods

Bacterial strains and growth conditions

E. meliloti 3001 and *E. meliloti* 3001 *sinI*/*sinR*$^{-}$ were grown at 30 °C either in Luria-Bertani medium (LB) or Vincent minimal medium (VMM) (Vincent 1970), supplemented with 0.2 % of several carbon sources (D-glucose, D-fructose, D-ribose, Na-succinate, Na-pyruvate, L-glutamine, L-histidine). The *E. coli* biosensor strain JM109 pSB1142 was grown overnight at 37 °C in LB medium with the addition of 125 μg/ml of tetracycline.

The stock solution of luteolin was prepared at a final concentration of 3.18 mM in a 9 mM NaOH dissolving solution. The working concentration of luteolin was 10 μM, known to be the required concentration in vitro for NodD induction (Kapulnik et al. 1987; Peters et al. 1988). Untreated cultures were supplemented with an equal volume of the 9 mM NaOH dissolving solution rather than luteolin.

The long chain N-acy-homoserine lactones (AHLs) were detected by the bioreporter *E. coli* strain JM109 pSB1142, carrying the *P. aeruginosa lasR* and *lasI* genes fused to *luxCDABE* (Winson et al. 1998). *E. meliloti* 3001 *sinI*/*sinR*$^{-}$ mutant strain, which is impaired in the synthesis and perception of AHLs, was used as a negative control.

Phenotype MicroArray (PM)

The growth of *E. meliloti* 3001 was tested in 1437 different culture conditions using PM metabolic (PM 01-03) and chemical sensitivity panels (PM09-PM20). The tested conditions included carbon and nitrogen sources, several concentrations of ions and osmolytes, pH stresses, and a wide variety of antibiotics, antimetabolites, heavy metals and other inhibitors. PM11-PM20 allowed assaying for the sensitivity to 240 chemical agents at four concentrations. The complete list of the compounds assayed can be obtained at http://www.biolog.com/pdf/PM1-PM10.pdf.

PM uses tetrazolium violet reduction as a reporter of active metabolism. The reduction of the dye causes the formation of a purple color that, recorded every 15 min, provides quantitative and kinetic information about the response of the cells in the PM plates (Bochner et al. 2001; Viti et al. 2015). *E. meliloti* 3001 was grown at 30 °C on Biolog Universal Growth agar (BUG) (Biolog Inc., Hayward CA, US) for two days. Then colonies were picked up with a sterile cotton swab and suspended in 1× IF-0 (Biolog) until OD_{600} = 0.1. Inoculation fluid for PM1 and PM2 inoculation was obtained diluting the cellular suspension (OD_{600} = 0.1) 10 times in an appropriate volume of VMM supplemented with 1× Dye Mix A as final concentration (Biolog). The inoculation fluid for PM3 was prepared diluting the cellular suspension (OD_{600} = 0.1) 10 times in VMM without ammonium chloride, and supplemented with 0.2 % Na-succinate as a carbon source and with 1× Dye Mix A (Biolog). The inoculation fluid for PM9-10 was prepared diluting the cellular suspension (OD_{600} = 0.1) 10 times in VMM supplemented with 0.2 % Na-succinate as carbon source and with 1× Dye Mix A (Biolog). The inoculation fluid for PM11-PM20 was prepared diluting the cellular suspension (OD_{600} = 0.1) 13.64 times in VMM supplemented with 0.2 % Na-succinate as carbon source. PM plates were inoculated with 100 μl per well. To test the influence of luteolin on the phenotype of the strain, luteolin was added to the inoculation fluids, at a final concentration of 10 μM, according to other *in vitro* induction assays (Barnett et al. 2004; Capela et al. 2005; Peters et al. 1986; Schluter et al. 2013). All the PM experiments were performed in duplicate, as two independent experiments. PM panels were incubated statically at 30 °C in an Omnilog Reader (Biolog) for 96 h. The kinetic profiles for *E. meliloti* 3001 strain in presence and in absence of luteolin were analyzed by inspecting kinetic curves and compared using the Omnilog-PM software (release OM_PM_109M). In order to discard possible false errors, the set of criteria reported by Khatri et al. (Khatri et al. 2013) were applied to the PM data analysis. The difference of area under the curve (Δarea) was used as discriminating parameter for comparing the kinetic curves obtained in order to identify different responses on metabolic panels (PM1-2–3–9–10) in presence of luteolin. The selection filter of the Omnilog-PM software allows to highlight all the wells, and thus the related compounds, in which the chosen parameter (Δarea) exceeds the standard threshold set up at 10,000 ($|\Delta area| \geq 10{,}000$) (Arioli et al. 2014). All the compounds showing a difference of area (Δarea) above the threshold value were considered as significant. The parameter IC50 was chosen to compare kinetic curves obtained on PM11-20. IC50, which was determined on the response on the four concentrations of each tested chemical, is defined as the well or fraction of well at which a particular per-well parameter is at half of its maximal value over the concentration series (Biolog, personal communication), and is expressed by PM software in "well units". IC50 values range between a minimum of 0.60 (no metabolic activity in any of the wells) and a maximum of 4.40 (optimal growth in all the wells). When a strain shows high metabolic activity in all the wells containing the four concentrations of a compound without showing any decrease of activity, the IC50 value is established as >4.40, whereas, when metabolic activity is absent in the four wells, IC50 is evaluated as <0.60. In order to compare the chemical resistance/sensitivity profiles and to highlight differences, a $|\Delta IC50|$ parameter threshold equal to 0.3 was used.

Carbon source utilization in relation to the inoculum cellular density (ICD)

E. meliloti 3001 was grown at 30 °C on LB agar plates for 48 h. Then colonies were picked up with a sterile cotton swab and suspended in VMM without any carbon source until OD_{600} = 1.0 was reached. Then the cellular suspension was serially diluted with a dilution ratio 1:4 in VMM without carbon sources for 10 times (cellular concentrations of the suspensions were valued by the plate count method on LB). The bacterial suspensions were used to inoculate 1:10 the following media supplemented with 1× Dye Mix A (Biolog): VMM supplemented with 0.2 % D- glucose, VMM supplemented with 0.2 % D-fructose, VMM supplemented with 0.2 % D-ribose, VMM supplemented with 0.2 % Na-succinate, VMM supplemented with 0.2 % Na-pyruvate, VMM supplemented with 0.2 % L-glutamine, VMM supplemented with 0.2 % L-histidine, and LB. The inoculated media were dispensed in microplates (250 μl each well). The experiment was done in absence (control) and in presence of 10 μM luteolin. Three replicates for each treatment were performed.

Microplates were statically incubated at 30 °C in the Omnilog Reader (Biolog) for 96 h. and monitored automatically every 15 min for color changes in the wells. Readings were recorded for 48 h and data were analyzed with Omnilog-PM software (release OM_PM_109M)

(Biolog). lag phase time was defined as the time of incubation (expressed in hours) at which the color intensity of 50 Arbitrary Omnilog Units (AOU) in VMM medium and 150 AOU for LB medium were reached, respectively. These threshold values were three times higher than that observed in the not inoculated wells (negative controls). To analyze the effect of luteolin on the lag phase of the cultures in relation to the inoculum cellular density (ICD) a two-way ANOVA was performed, considering the lag phase as the dependent variable and as explanatory variables the luteolin presence/absence (two levels) and the different inoculum cellular density (eleven levels). When the ANOVA analysis was significant, the post hoc *t* test was performed to locate the significant differences ($P < 0.01$). The lag phase time was expressed as mean of three independent replicates ± standard deviation.

Long-chain N-acyl homoserine lactones (AHLs) quantification

The biosensor *E. coli* JM109 pSB1142 was used to detect the production of long chain N-acyl homoserine lactones (AHLs) by *E. meliloti* 3001 and its *sinIR* mutant, which was used as a negative control. These strains were grown at 30 °C on LB agar for two days. Then colonies were picked up with a sterile cotton swab and suspended in VMM without carbon source until OD_{600} = 1.0 was reached. An aliquot of 0.2 ml of cellular suspension was used to inoculate 1.8 ml of the following media: VMM supplemented with 0.2 % D-glucose, VMM supplemented with 0.2 % D-fructose, VMM supplemented with 0.2 % D-ribose, VMM supplemented with 0.2 % Na-succinate, VMM supplemented with 0.2 % Na-pyruvate, VMM supplemented with 0.2 % L-glutamine, VMM supplemented with 0.2 % L-histidine, and LB. The experiments were done in the absence (negative control) and in the presence of 10 μM luteolin. Four independent experiments were performed for each growth condition, and each one of the four replicates was tested in quadruplicate. Cultures were incubated at 30 °C with shaking (100 rpm) until exponential phase (OD_{600} = 0.4) or stationary phase (OD_{600} = 1.0) were reached. Then the cultures were centrifuged at 5000x*g* for 15 min at 4 °C. One hundred μl of the supernatant of each culture was dispensed, in black microplate with clear bottom (Greiner bio-one International GmbH, Austria). An overnight culture of *E. coli* JM109 biosensor, resuspended in fresh medium at OD_{600} = 0.3, was added for each well to an equal volume of the *E. meliloti* 3001 supernatant. Microplates were then incubated at 37 °C for 24 h in the spectrophoto/fluoro/luminometer Infinite M200 Pro (Tecan Group Ltd., Switzerland), which measured luminescence and turbidity values every 30 min in each well. In order to avoid the possibility that quantitative changes in AHL production observed could depend by differences in bacterial growth, and not to any effect on QS depending from the presence of luteolin, the measured bioluminescence values were normalized for the biosensor optical density and for the tested strain optical density and expressed as r_{AHLs}. The sum of r_{AHLs} along the 24 h was calculated (R_{AHLs}). R_{AHLs} was expressed as mean ± standard deviation. Significant differences between treated and untreated cultures were determined by *t* test ($P < 0.05$).

Siderophores production assay

Siderophores were detected in cultures of *E. meliloti* of 3001 using the chrome azurol S (CAS) solution prepared according to Alexander and Zuberer (1991). CAS solution has a blue color due to the CAS-iron complexes, but when a strong iron chelator such as a siderophore removes iron from the dye complex, the color changes from blue to orange inducing a decrease in absorbance at 670 nm.

E. meliloti 3001 was inoculated (OD_{600} = 0.05) in VMM medium supplemented with 0.2 % Na-succinate prepared without iron chloride in order to stimulate siderophore production. Cultures were supplemented or not with 10 μM luteolin, performed in three independent experiments, and incubated at 30 °C with shaking. When exponential growth phase (OD_{600} = 0.3) and stationary phase (OD_{600} = 0.5) were reached, 1 ml of each culture was centrifuged at 10,000×*g* for 3 min. 80 μl of the supernatant of each culture and 80 μl of the medium supplemented or not with 10 μM luteolin (as negative controls) were dispensed into the wells of a microplate in triplicate, and then mixed with 160 μl CAS solution. After 3 min of incubation the absorbance at 670 nm (A_{670}) was measured by a microtiter plate reader (GDV, Model No. DV990BV5).

Motility assay

E. meliloti 3001 strain was grown in LB medium or in VMM supplemented with 0.2 % Na-succinate, in the presence or absence of 10 μM luteolin. Inocula were grown at 30 °C with shaking (100 rpm) until stationary phase was reached (OD_{600} = 1.0). Cells were diluted in

the original medium to reach OD_{600} = 0.5, then 5 μl were dipped in triplicate in the center of semisolid VMM supplemented with 0.2 % Na-succinate (0.2 % agar) and semisolid LB (0.2 % agar). Also plates with 10 μM luteolin were carried out. Semisolid agar plates were incubated at 30 °C for 2 days. The extent of swarming was determined by measuring the swarming ring diameter. Diameter was expressed as mean ± standard deviation of three independent experiments, each performed in triplicate. Significant differences between treated and untreated cultures were detected with *t* test ($P < 0.05$).

Biofilm quantification

The quantitative assessment of biofilm was evaluated by the crystal violet staining following the procedure for the microtiter plate assay used by O'Toole et al. (1999) with the following modifications. *E. meliloti* 3001 was grown to OD_{600} = 2.0 in LB and VMM containing 0.2 % Na-succinate media. Cells were then washed and resuspended to OD_{600} = 0.2 with or without 10 μM luteolin in the following desired media: VMM supplemented with 0.2 % D-glucose, VMM supplemented with 0.2 % D-fructose, VMM supplemented with 0.2 % D-ribose, VMM supplemented with 0.2 % Na-succinate, VMM supplemented with 0.2 % Na-piruvate, VMM supplemented with 0.2 % L-glutamine, VMM supplemented with 0.2 % L-histidine, and LB. One hundred μl of cultures per well were grown in microplates, statically incubated at 30 °C. At the final time point, the optical density (OD_{600}) of each well was read in a microtiter plate reader (GDV, Model No. DV990BV5). For each well, the parameter R_B was calculated as the ratio between the crystal violet absorbance and the culture absorbance. For each sample, R_B was expressed as mean ± standard deviation of two independent experiments with twelve replicates each time. Significant differences between treated and untreated cultures were determined with *t* tests ($P < 0.05$).

Production of extracellular enzymes

E. meliloti 3001 was grown on LB agar at 30 °C for 48 h. Then cells were picked up using a sterile cotton swab and suspended in VMM supplemented with 0.2 % Na-succinate until OD_{600} = 1.0. Two distinct aliquots of the prepared bacterial suspension, without or with 10 μM luteolin were incubated at 30 °C for 3 h. Five μl of the treated and untreated cellular suspensions were spotted onto the surface of every type of agarized media, dedicated to test each one of the following extracellular enzymatic activity. As well, plates with 10 μM luteolin were included.

Amylolitic activity was evaluated in 10 % Tryptone Soya Agar - TSA (Oxoid) or VMM supplemented with 0.2 % Na-succinate both added with 1 % starch, in triplicate. After 48 h of incubation at 30 °C, the plates were flooded with Lugol's iodine solution (1 % iodine and 1 % potassium iodine in distilled water). A pale yellow zone around the colony in the otherwise blue medium indicates starch degradation (Atlas and Park 1993).

Proteolytic activity was valued in 10 % TSA and VMM supplemented with 0.2 % Na-succinate both added with 1 % skim milk. After incubation for 48 h at 30 °C, a positive reaction was detected as a clear zone around the colony in the opaque medium.

Extracellular lipases were detected both in 10 % TSA and VMM supplemented with 0.2 % Na-succinate both supplemented with 1 % (*v*/*v*) tributyrin, and in 10 % TSA and VMM supplemented with 0.2 % Na-succinate both supplemented with 1 % (*v*/*v*) tween 80. After incubation for 48 h at 30 °C, a positive reaction was detected as a clear zone around colonies in opaque tributyrin agar media, or as a precipitate around the colony in the tween 80 agar media, respectively.

Phospholipase activity was detected in 10 % TSA and in VMM supplemented with 0.2 % Na-succinate both supplemented with 2 % egg yolk. The cleavage of the phosphate ester bonds formed water insoluble lipid. After incubation for 48 h at 30 °C, the enzyme activity was detected as a zone of opalescence in the medium surrounding the colonies.

Cellulolytic activity was checked in 10 % TSA and in VMM supplemented with 0.2 % Na-succinate both supplemented with 0.1 % carboxymethylcellulose (CMC). After incubation for 48 h at 30 °C, plates were stained with a solution 0.03 % Congo red (Strauss et al. 2001). A pink yellow zone around the colony in the otherwise red medium indicated CMC degradation.

For each enzymatic activity assay, four independent experiments were set up and each one tested in quadruplicate. The area of the halos surrounding the cellular stops was measured and expressed as mean ± standard deviation.

Indole-3-acetic acid (IAA) production

E. meliloti 3001 was grown on LB agar plates at 30 °C for 48 h. Single colonies were picked up using a sterile toothpick and suspended in 0.8 % NaCl until

OD_{600} = 1.0. An aliquot of 0.2 ml of this bacterial suspension was used to inoculate the following media: VMM supplemented with 0.2 % L-glucose, VMM supplemented with 0.2 % D-fructose, VMM supplemented with 0.2 % D-ribose, VMM supplemented with 0.2 % Na-succinate, VMM supplemented with 0.2 % Na-pyruvate, VMM supplemented with 0.2 % L-glutamine, VMM supplemented with 0.2 % L-histidine and LB. The media were supplemented or not with 500 μM L-tryptophan and 10 μM luteolin.

Cultures were then incubated at 30 °C under shaking (100 rpm) until OD_{600} = 1.0. Each culture was centrifuged at 5000×*g* for 15 min at 4 °C, and the supernatant was recovered for IAA detection. Briefly, 70 μl of the supernatant and 210 μl of Salkowsky reagent (1 ml 0.5 M $FeCl_3$ plus 50 ml 35 % perchloric acid) were dispensed in a well of a microplate (Batista and Hungria 2012; Gordon and Weber 1951). For each growth condition three independent experiments, with eight replicates per series, were done. After 30 min of incubation at room temperature in the dark, the OD_{530} was read in a GDV microplate reader (model DV990BV5). For each well, the ratio R_{IIA} between OD_{530} of spent medium and OD_{530} of the relative fresh medium was calculated. For each culture, the R_{IIA} was calculated as mean ± standard deviation. Significant differences between treated and untreated cultures were determined with *t* test ($P < 0.05$).

Results

Phenotype MicroArray (PM) analysis

PM system (Biolog) was used to characterize *E. meliloti* 3001 in presence of 10 μM luteolin on 190 different carbon sources (PM1-2), 95 nitrogen (PM3) sources and tolerances to different osmolytes and pH conditions (see ESM_1 in the Online Resources for a complete set of results). *E. meliloti* 3001 was found to be able to metabolize 95 out of 190 tested carbon sources and 74 out of 95 nitrogen sources in both the absence or presence of luteolin, indicating that luteolin did not affect carbon and nitrogen metabolism.

The effect of luteolin on the metabolic activity of *E. meliloti* 3001 under several osmolyte gradients and pH conditions was tested using PM9 and PM10 plates. The pH range where *E. meliloti* 3001 exhibited active metabolism was between 4.5 and 10, with an optimal pH value around 6.0-7.0. Luteolin did not affect the activity under the tested pH range, whereas it increased the osmotolerance of the strain to the sodium phosphate gradient ranging from 20 mM to 200 mM (Fig.1).

The *E. meliloti* 3001 strain was also analyzed for chemical sensitivity/resistance in presence or in absence of luteolin toward 240 compounds, each one presenting four different concentration levels, in PM11-20 plates. ΔIC50 parameter threshold, as defined in materials and methods section, was used to compare the PM profiles.

The comparison of chemical sensitivity profile of *E. meliloti* 3001 showed different phenotypic responses associated to the luteolin treatment on 20 % of the tested compounds (Online Resources ESM_2). Among the discriminating compounds, luteolin induced a higher metabolic activity for 82 % of them.

Enhanced luteolin-induced resistance was found for a broad set of chemicals such as antibiotics (18), toxic ions (10), respiration inhibitors (8), membrane damagers (6), DNA intercalants (3) and other antimicrobial agents (5) (Fig. 2). However, for other toxic compounds, including three antibiotics (cefotaxamine, phosphomycin, and sulfanilamide), L-glutamic acid g-hydroxamate, chloroxylenol (fungicide), ferric chloride and potassium tellurite (toxic ions), sanguinarine (membrane damagers) and hydroxylamine (mutagenic agent) luteolin induced a higher sensitivity (Table S2).

Carbon source utilization in relation to the of inoculum cellular density (ICD)

Utilization of carbon sources by *E. meliloti* 3001 was not influenced by luteolin treatement when PM plates were used. Nevertheless luteolin was previously reported to have an effect on growth rate *of R. meliloti* (Hartwig et al. 1991). Therefore, we decided to evaluated whether the cellular density of the inocula (ICD) could play a role in making luteolin active on the carbon source utilization by the bacterium. To test the effect of luteolin on the utilization of carbon sources, seven different substrates (three carbohydrates, two aminoacids and two organic acids) were selected among those on which the *E. meliloti* 3001 showed the highest activity, as resulted from PM analysis. Eleven concentrations of inocula, obtained by serial diluition (1:4) ranging from $9x10^8 \pm 2.0\times 10^7$ cells/ml to 180 ± 20 cells/ml (named D1-D11) were tested. Tested media were LB and VMM supplemented with 0.2 % different carbon sources (D-glucose,D-fructose, D-ribose, L-histidine, L-glutamine, Na-succinate and Na-pyruvate).

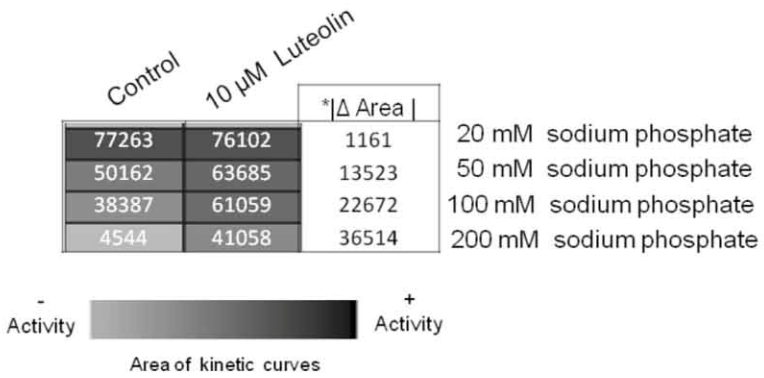


Fig. 1 Luteolin effect on metabolic activities of *E. meliloti* 3001 in the presence of a sodium phosphate gradient (20–200 mM). Area of the PM kinetic curves and the parameter Δarea are reported for each condition after 96 h of incubation. The intensity of color code used reflects the area values.*Δarea is defined as the difference of area under the PM kinetic curves for the luteolin treated *E. meliloti* 3001 strain compared to the untreated ones

Two-way ANOVA, performed on lag time results, showed a highly significant effect ($P < 0.01$) not only of ICD, as expected, but also of luteolin treatment for all tested media (Fig. 3). Lag time was significantly reduced by luteolin for all the ICDs in VMM medium supplemented with each carbon source tested (*t*-test, $P < 0.05$ and $P < 0.01$, respectively). Exceptions of D1 for L-glucose, L-ribose and L-succinate and D7, D8, D11 for L-histidine were observed. In LB medium, luteolin reduced (*t* test, $P < 0.05$) the latence time only at the high cellular density D2 and D3 of about 5 and 4 h, respectively.

Bacteria grown in VMM + 0.2 % L-glutamine and in VMM + 0.2 % L-histidine showed a reduction of lag time, in relation to luteolin treatment, of ca. 10 and 5 h, respectively, for all the ICDs tested (two-way ANOVA did not reveal a significant interaction between ICD and luteolin treament). On the contrary, the reduction of the lag time induced by luteolin was little at high ICD and became higher at low ICD for cultures grown in VMM supplemented with Na-succinate, Na-pyruvate, D-fructose, D-glucose and D-ribose (two-way ANOVA showed a highly significant interaction ($P < 0.01$) between ICD and luteolin treament). As an example, on L-fructose the lag time reduction was of 1.1 h when the culture had ICD equal to D1 and reached 36.2 h when ICD was equal to D9. On D10 and D11 the reduction of lag time induced by luteolin was not determined, but higher than 36.2 h.

Long-chain N-acyl homoserine lactones (AHLs) production

In order to evaluate the influence of luteolin on the long-chain AHLs overall production, exponential phase cultures of *E. meliloti* 3001 strain grown in the presence/absence of luteolin were analyzed using the *E. coli* JM109 pSB1142 biosensor as a bioluminescent reporter system for AHLs. As controls, supernatants obtained from un-inoculated medium and *E. meliloti* 3001 *sinIR* defective mutant, impaired in the synthesis and perception of AHLs, without and with 10 µM luteolin were used.

The presence of luteolin did not affect the long chain AHLs production in rich medium (LB) (data not shown). No significant changes in the bioluminescence produced by the *E. coli* biosensor were detected upon luteolin treatment compared to the untreated *E. meliloti* 3001, both during exponential and stationary growth phases (data not shown). Conversely, when minimal medium (VMM) was used, supplemented in separate experiments with seven carbon sources (D-glucose, D-ribose, D-fructose, L-histidine, Na-pyruvate, Na-succinate, L-glutamine), the presence of luteolin significantly reduced the overall AHLs production of *E. meliloti* 3001 (Fig. 4). The luteolin effect was observed during the exponential growth phase for all carbon sources (Fig. 4). The AHLs reduction levels detected in response to luteolin varied depending on carbon source (Fig. 4). The exponential culture exhibited a long chain AHL levels decrease of ca. 55 % on D-glucose, 38 % on D-ribose, 34 % on L-histidine, ca. 30 % on D-fructose, 28 % on Na-pyruvate, 20 % on Na-succinate and 14 % on L-glutamine in presence of luteolin.

Influence of luteolin on indole-3-acetic acid (IAA) production

The production of phytohormone IAA was evaluated in late exponential phase of *E. meliloti* 3001 grown both in

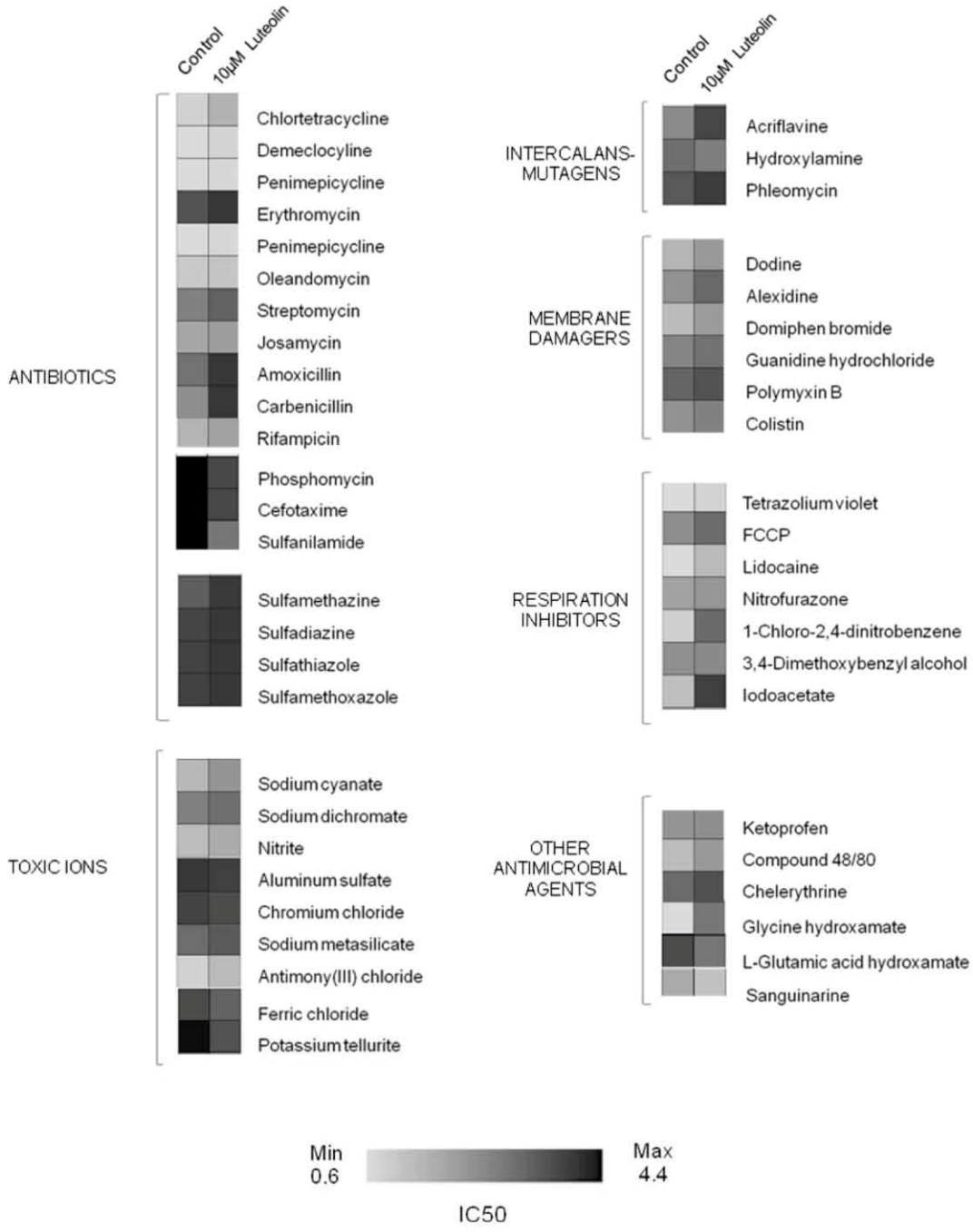


Fig. 2 Luteolin effect on resistance to chemicals of *E. meliloti* 3001. Chemicals identified by PM analysis toward which the presence of luteolin, after 96 h incubation time, induced an increase or a decrease of resistance ($\Delta IC50 \geq 0.3$) are reported. The color intensity of the code used reflects the IC50

LB and in VMM supplemented with several carbon sources. The most remarkable increase in IAA production was observed in VMM supplemented with D-ribose, Na-pyruvate and D-glucose, according to this order. The presence of luteolin only impacted the synthesis of IAA when grown in LB or VMM supplemented with D-fructose as a carbon source media (Fig. 5). The effect of luteolin on IAA production was not statistically significant in cultures grown in VMM supplemented with Na-succinate, D-ribose, D-glucose and Na-pyruvate. No IAA

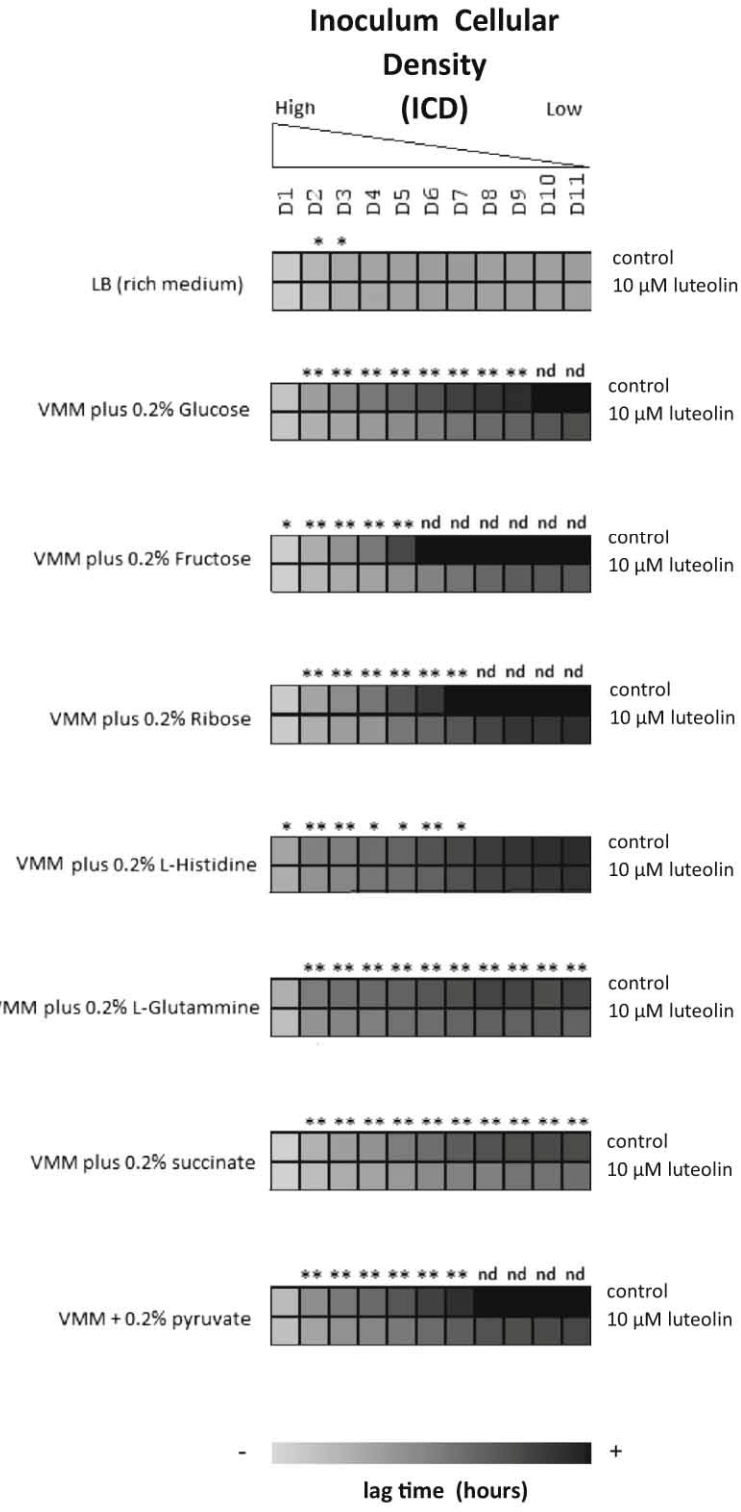


Fig. 3 Luteolin effect on the lag time of *E. meliloti* 3001 in relation to the inoculum cellular density (ICD). The growth of *E. meliloti* 3001 was evaluated using VMM minimal medium supplemented with different carbon sources (D-glucose, D-fructose, D-ribose, L-histidine, L-glutamine, Na-succinate and Na-pyruvate) and LB medium. The experiments were performed with luteolin and without luteolin (control). Cultures were incubated at 30 °C for 90 h. Lag time is reported in gray scale ranging from light gray to black. Black squares indicate that lag time was 90 h or higher (not determined). Lag time was reported as a mean of three independent replicates. Standard deviations were lower than 10 %. D1-D11 indicate the dilution series of inculum. The ICDs for which a significant reduction, *t* test, $P < 0.05$ or *t* test, $P < 0.01$, of lag time induced by luteolin was detected are marked with * or **, respectively. ICDs for which statistical analysis was not performed, since the lag time of untreated cultures (0 µM luteolin) was higher than 90 h, are marked as "nd" (not determined)

synthesis was detected in cultures grown in VMM supplemented with amino acid sources, such as L-histidine and L-glutamine, either in the presence or absence of luteolin.

Effect of luteolin on motility proficiency

In rhizobia, motility has been shown not to be essential for nodulation, but it allows the bacteria to find their specific host legume and establish symbiosis (Nievas et al. 2012). Bacterial motility is an energetically costly process, and therefore must be finely regulated (Nievas et al. 2012).

In order to assess the effect of luteolin on the motility of *E. meliloti* 3001, the growth on semisolid LB medium and VMM with Na-succinate as a carbon source was valued in the absence and presence of 10 μM luteolin. *E. meliloti* 3001 was motile in semisolid LB medium, resulting in the formation of a swarming halo of 3.4 cm average diameter. No significant changes in the swarming halo were found with and without luteolin in semisolid LB medium (data not shown). Whereas, *E. meliloti* 3001 in semisolid VMM medium exhibited a significantly different motile ability as compared to rich medium, resulting in an halo of just 1 cm average diameter ($P < 0.01$, *t*-test). Comparing the swarming behavior in VMM medium with and without luteolin, a significant reduction in the motility was found in the presence of luteolin, resulting in a swarm halo of 0.8 cm average size ($P < 0.01$,*t*-test).

Influence of luteolin on biofilm formation of *E. meliloti* 3001

The ability of rhizobia to establish a biofilm can be used as a strategy for survival under unfavorable conditions and for optimizing resource utilization in hostile environments. In certain species, the biofilm formation is an important feature for colonization and/or host invasion (Nievas et al. 2012; Perez-Montano et al. 2014). The biofilm formation assay performed on *E. meliloti* 3001 revealed that luteolin affected the biofilm formation on rich medium and minimal medium (Fig. 6). In the presence of luteolin a statistically significant increase (amounting to 68 %) of *E. meliloti* 3001 biofilm formation was observed on LB medium (Fig. 6a).

The 3001 strain exhibited a significantly reduced biofilm formation in presence of luteolin on VMM supplemented with D-glucose (26 %), D-ribose (ca.

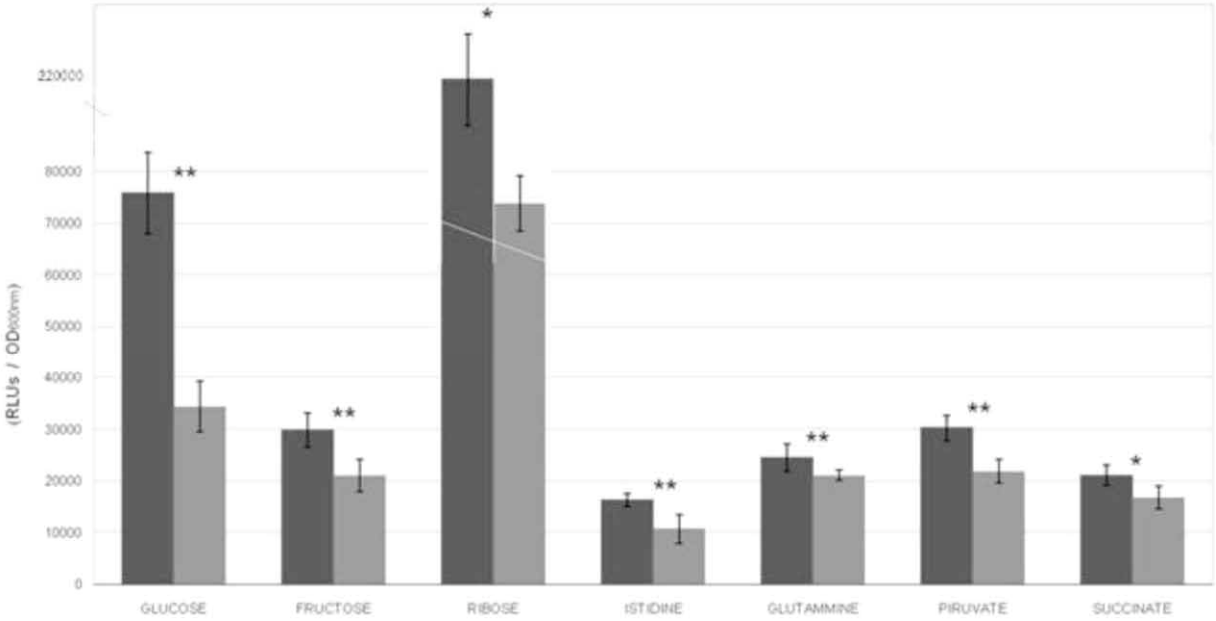


Fig. 4 Luteolin effect on the overall production of long-chain AHLs by *E. meliloti* 3001. The production of long-chain AHLs was evaluated for cultures grown in VMM minimal medium supplemented with different carbon sources (D-glucose,D-fructose, D-ribose, L-histidine, L-glutamine, Na-succinate and Na-pyruvate) at the exponential growth phase. The experiments were carried out in presence and in absence of luteolin. Long-chain AHLs were quantified using the *E. coli* biosensor strain JM109 pSB1142 as reporter system. Data refer to measured emitted bioluminescence (expressed as relative light units) normalized for biosensor optical density on each tested condition. Data are reported as a mean (± standard deviation of the mean) of four independent experiments, each one tested in quadruplicate. Dark grey bars correspond to control and light grey bars correspond to experiments performed in media supplemented with 10 μM luteolin. * or ** indicate a significant difference in presence of luteolin compared to the luteolin absence at the level *P*-value < 0.05 (*t* test) and *P*-value < 0.01 (*t* test), respectively

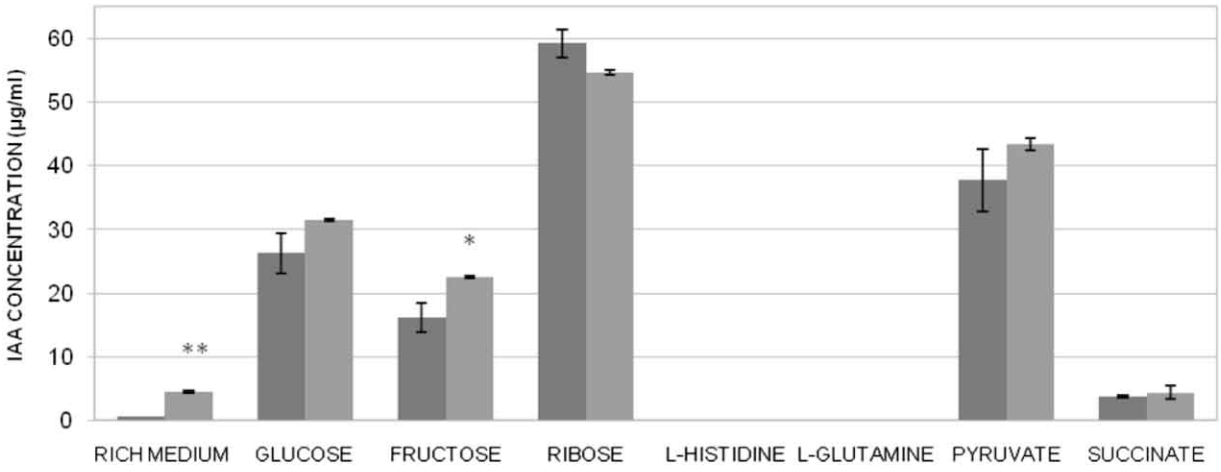


Fig. 5 Effect of luteolin on the IAA production by *E. meliloti* 3001. The IAA production was evaluated using LB and or VMM minimal medium supplemented with different carbon sources (D-glucose,D- fructose, D-ribose, L-histidine, L-glutamine, Na-succinate and Na-pyruvate); dark grey bars indicate the amount of IAA detected in control cultures, light grey bars indicate the amount of IAA detected in cultures with 10 µM luteolin. Means and standard deviations of three independent experiments, each one with eight replicates are reported for each condition. * or ** indicate statistically significant results in presence of luteolin compared to the luteolin absence at the level *P*-value < 0.05 (*t* test) and *P*-value < 0.01 (*t* test), respectively

36 %), Na-succinate (53 %) and L-histidine (ca 51 %) (Fig. 6b). No statistically significant differences were detected for biofilm formed by cultures without and with luteolin in VMM supplemented with D-fructose, Na-pyruvate and L-glutamine (Fig. 6b).

Siderophore production

E. meliloti is known to produce the siderophore rhizobactin (Smith et al. 1985), which can inhibit the growth of rhizobactin transport mutants in iron deficient conditions (diCenzo et al. 2014). The CAS assay was performed to value the effect of luteolin on siderophore production by *E. meliloti* 3001. The strain was grown in VMM medium without iron chloride to stimulate siderophore production, both in the absence and presence of luteolin. Supernatants of exponential and stationary cultures were mixed with CAS agar reagent and the A_{670nm} was measured. Two way ANOVA showed a significantly decrease of A_{670} ($P < 0.01$) induced by both growth phase and luteolin treatment. This result suggested that the siderophore production increased during growth as expected, but more interestingly that luteolin stimulated siderophore biosynthesis. In fact, although no differerence in A_{670} was observed in exponential growth cultures treated or not with luteolin, a significant decrese of A_{670} was observed in stationary phase cultures treated with luteolin in respect to not treated cultures (*t*-test, $P < 0.05$) (Fig. 7).

Extracellular enzyme production

Assays performed to evaluate extracellular enzymatic activity (protease, amylase, lipase, phospholipase C, carbossimethyl cellulose, glycanase) of *E. meliloti* 3001, grown with and without luteolin, revealed that *E. meliloti* 3001 was negative for the tested extracellular enzymes production in the tested conditions. The results obtained in presence of luteolin remained unchanged compared to those yielded in its absence, suggesting no luteolin influence on these cellular biosynthetic processes.

Discussion

The importance of flavonoids in modulating rhizobia activity has been brought to light by transcriptional studies, which allowed identifying genes whose expression is modulated by these molecules. However, the effective role of flavonoids on bacterial physiology cannot be fully described using just this kind of approach, because some of the genes found to be modulated by flavonoids have a still unknown function. In addition, these plant signal molecules could exert regulatory activity at post-transcriptional level. Studies aiming to assess the expression of small non-coding RNA (sRNA) in *E. meliloti* have pointed out that at least three small RNA transcripts involved in modulating gene expression are controlled by the flavonoid luteolin (del

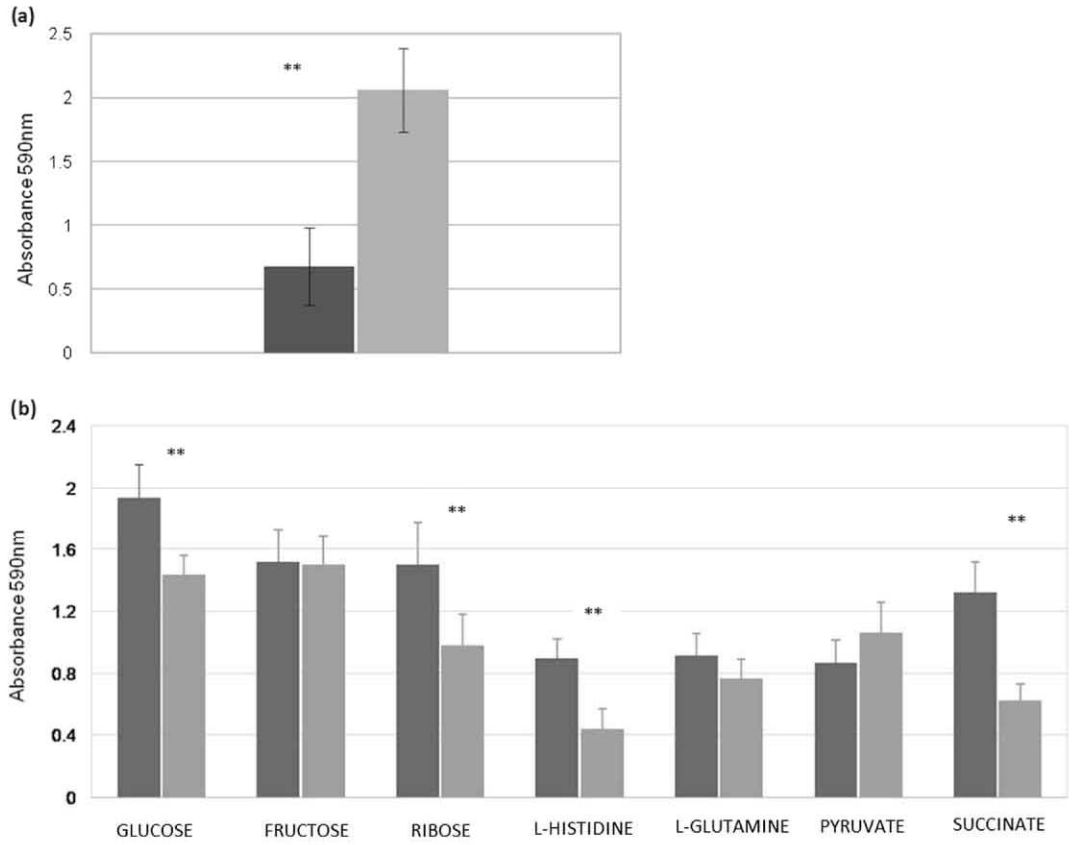


Fig. 6 Effect of luteolin on *E. meliloti* 3001 biofilm formation. The biofilm formation was evaluated on LB medium a) and VMM minimal medium supplemented with different carbon sources b), using crystal violet staining at 72 h after multi-well plate inoculation (see methods). Means and standard deviations of two independent experiments for each condition with twelve replicates each time are shown. Dark grey bars correspond to control and light grey bars correspond to experiments performed in presence of luteolin. * or ** indicate statistically significant differences in biofilm formation in presence of luteolin at the level of P-value < 0.05 (t test), and P-value < 0.01 (t test), respectively

Val et al. 2007). Therefore, we can expect that luteolin may have a pleiotropic effect on rhizobial phenotypes, possibly also unlinked to the Nod factors (NF) biosynthesis. To this aim, we performed an extensive analysis of a wide spectrum of phenotypes induced by luteolin in the *E. meliloti* 3001. PM analysis of *E. meliloti* 3001 carried out in 1437 different growth conditions showed that luteolin makes an important change in the sensitivity and osmotolerance profiles of *E. meliloti* 3001. In particular, the major effect of luteolin on the sensitivity profile of *E. meliloti* 3001 was an enhanced resistance towards a large set of antimicrobials and toxic compounds. One of the major system conferring multidrug resistance phenotypes in bacteria are efflux pumps (Eda et al. 2011). These systems have also been reported to play an important role in the establishment of the symbiosis between rhizobia and leguminous plants (Cosme et al. 2008; Eda et al. 2011). In *E. meliloti* 1021 14 different multidrug resistance efflux pumps (MDR), one pump belonging to the ATP-binding cassette family (ABC), three pumps belonging to the major facilitator superfamily (MFS), and ten pumps belonging to the resistance-nodulation-cell division family (RND), have been detected (Eda et al. 2011). Eda et al. (2011) hypothesized that the SmeAB RND type pump had a main role in antimicrobial resistance of *E. meliloti*. Nevertheless, the authors concluded that the 20 chemicals that they tested could have been unable to highlight the contributions of the other pumps (Eda et al. 2011). In our study the PM approach permitted to enlarge the numbers of tested substrates up to 240 and to identify a heterogeneity of substrates (several antibiotics, toxic ions, intercalating mutagens, membrane damagers, respirations inhibitors and other antimicrobial agents) to which the bacterium was made more tolerant by luteolin.

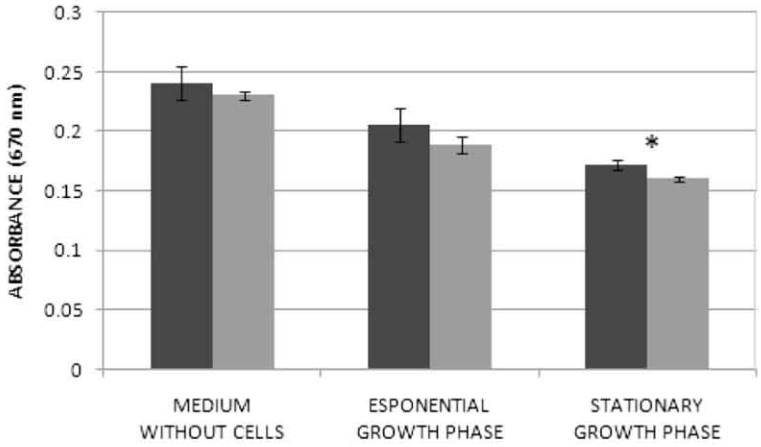


Fig. 7 Effect of luteolin on *E. meliloti* 3001 siderophore production. The siderophore production was valued by CAS assay on cultures grown in presence and absence of luteolin, in VMM minimal medium depleted of iron chloride and supplemented with Na-succinate. The CAS assay was performed on exponential growth phase and stationary growth phase, and on not inoculated media, as negative control. Means and standard deviations of three independent replicates are shown. Dark grey bars correspond to control and light grey bars correspond to experiments performed in presence of luteolin. * indicate statistically significant differences in A_{670} in presence of luteolin at the level of P-value < 0.05 (t test)

Transcriptional experiments performed until now to find luteolin responsive genes in *E. meliloti* did not point out any known efflux or resistance system other than ErmAB, which is a MFS type pump whose expression is induced by luteolin throught the ErmR regulator (Capela et al. 2005; Rossbach et al. 2014). Taking into account the molecular heterogeneity of the compounds towards which *E. meliloti* showed an increased resistance in presence of luteolin, our data support the hypothesis that more than one efflux pump or other resistance systems are activated by the luteolin. In other rhizobia, flavonoids can activate efflux pumps involved in resistance to rhizosperic compounds. Isoflavones genistein and daidzein induced resistance to the phytoalexin glyceollin in some soyabean nodulating rhizobia (Parniske et al. 1991). In *Rhizobium etli*, flavonoids induced genes *rmeA* and *rmeB*, which encode a MFS pump, that are required both for efficient nodulation in bean (Gonzalez-Pasayo and Martinez-Romero 2000) and for resistance to naringenin, coumaric acid or salicylic acid. Also, in *E. meliloti* 1021, the deletion of genes encoding the RND-type SmeAB pump resulted in increased susceptibility to antimicrobials and makes the bacterium defective in competing with the wild-type strain for nodulation (Eda et al. 2011). The resistance induced by luteolin pointed out by PM analysis could help *E. meliloti* to cope with toxic compounds released in the rhizosphere by the plant, to counteract pathogens, or to withstand oxidative burst within the plant during the infection (Santos et al. 2001), offering to the symbiotic bacterium an advantage to reach the roots and to establish an effective symbiosis.

The increased osmotolerance of *E. meliloti* 3001 to Na-phosphate found in the presence of luteolin could be rationalized considering the variation in soil pH and osmolytes to which microorganisms are subjected in rhizospheric conditions. The luteolin-induced osmotolerance could increase bacterial fitness and thus constitutes an advantage for its successful plant root colonization (Biondi et al. 2009). The two major osmoprotection systems characterized in *E. meliloti* are represented by betaine-glycine (Mandon et al. 2003; Le Rudulier et al. 1984; Pocard et al. 1997) and ectoine (Talibart et al. 1994). *In vitro* expression studies (Becker et al. 2014; Capela et al. 2005; Roux et al. 2014) revealed no evidences that the betaine system (*betICBA* operon) was induced by luteolin. However, it cannot be excluded that the osmoprotectant ectoine system is luteolin induced and therefore likely involved in the enhanced osmotolerance of *E. meliloti.*

PM analysis revealed that the capability of *E. meliloti* 3001 to use carbon and nitrogen sources is not influenced by luteolin in the tested conditions. Nevertheless, Hartwig and Phillips (1991) found that luteolin affected the ability of *E. meliloti* to use carbon sources. Therefore, we conducted experiments using VMM supplemented with different carbon sources and inoculum densities. We found that when the strain was grown in the sugars and carboxylic acid (D-glucose, D-ribose, D-fructose, Na-succinate or Na-pyruvate), luteolin reduced the lag time

of the cultures at decreasing bacterial density of the inoculum. In VMM supplemented with D-fructose, *E. meliloti* 3001 did not show metabolic activity over a time of 96 h when the inoculum was lower than about $7.5x10^5$ cells/ml. In presence of luteolin also at the lowest bacterial concentration tested (around 180 cell/ml) metabolic activity was shown (lag time was 66 h). The effect of luteolin on cultures grown in VMM supplemented with L-glutamine or L-histidine was different from that observed in VMM containing the sugars and the organic acids tested, in fact, the reduction of lag time induced by the luteolin remained of the same entity at the different cellular densities. These results pointed out that luteolin, through forcing small populations of the bacterium to be active, exerts a growth-promoting effect, as already shown by Hartwig and Phillips (Hartwig and Phillips 1991). Nevertheless, our study, extending the number of carbon source used and by including two amino acids, showed that the influence of luteolin on the activity of the *E. meliloti* 3001 is high in conditions of nitrogen limitation. This hypothesis is sustained by the evidence that nitrogen limitation is a factor controlling the establishment of the symbiosis (Dusha and Kondorosi 1993). Moreover, since rhizosphere soil contains organic acids, amino acids and other molecules, such as galactosides, which may serve as chemoattractants and nutrient sources for rhizobia (Dakora and Phillips 2002), we can speculate that the effect of luteoolin on carbon source utilization is related to the optimization of catabolic systems in the presence of plant. Indeed, in previous work, we showed that for some of the compounds tested on PM1 and PM2 plates, a high strain-related variation in their utilization is present in *E. meliloti* (Biondi et al. 2009), suggesting that the differential use of rhizosphere nutrient sources could be a fitness factor for rhizobial competition.

In Proteobacteria, cell-density dependent signaling, known QS, is mediated by diffusible signal molecules, mainly represented by AHLs. A broad variety of important physiological traits related to the free living and symbiotic states are regulated by QS. Luteolin did not affect the production of long chain AHLs under high nutrient availability conditions (*e.g.* rich medium) during the exponential growth phase. More interestingly, results obtained in a nutritionally limited condition (*e.g.* minimal medium) clearly revealed that luteolin significantly reduced the overall long chain AHLs production by *E. meliloti* 3001. This luteolin effect was found during the exponential growth phase for all carbon sources used. *E. meliloti* and a few other bacterial species have been reported to produce uncommon AHLs with long acyl chains, containing more than 12 carbons (Gonzalez and Marketon 2003; Sanchez-Contreras et al. 2007; Steindler and Venturi 2007; Venturi 2006). It is known that plants produce compounds that mimic AHLs, which are able to interfere with bacterial quorum sensing genes (Gao et al. 2003; Hassan and Mathesius 2012; Teplitski et al. 2000). In strains of *S. fredii*, *R. etli*, and *R. sullae*, the production of long chain AHLs was found to be enhanced in presence of their respective *nod*-gene-inducing flavonoid (Perez-Montano et al. 2011). This result is in contrast with what was observed here for *E. meliloti* 3001 and it could suggest that the influence of flavonoid on rhizobial strain is host-specific. Nevertheless, since the three rhizobial strains were grown in yeast mannitol medium, we cannot exclude that the increase in AHLs induced by flavonoids could be dependent by the medium used. In fact, data obtained in this work suggest that the effect of the host flavonoid on the rhizobium depends on nutritional conditions. Indeed, in *Pseudomonas aeruginosa*, the flavanone naringenin reduces AHL production (Vandeputte et al. 2011), AHL production has been shown to be related to nodule number control in *Medicago truncatula* (Veliz-vallejos et al. 2014), and the mutant of *sinI* gene (encoding the AHL synthase) has delayed nodule formation (Gao et al. 2003).

The effect of luteolin on AHLs production in *E. meliloti* 3001 observed in minimal medium could be related to rhizosphere adaptation and possibly to the early phase of symbiotic interaction as well. We can speculate that a reduction in AHL production allows the presence of higher rhizobial cell number in the rhizosphere of the host plant compared to soil or to non-host plants, thereby increasing competitiveness for rhizosphere colonization and formation of nodule primordial. Moreover, in the same poor nutritional in vitro conditions the addition of luteolin caused an increase in IAA production, although to a different extent according to the carbon source. Auxin are known to be essential for nodule development and several flavonoids were shown to have a crucial role as signals in the initiation of nodule primordia, for their inhibitory activity on auxin transport (Zhang et al. 2009). Similarly, luteolin-induced IAA synthesis by rhizobial cell could be hypothesised to locally contribute to nodule formation.

The uptake of iron is another crucial aspect of rhizobia metabolism, because enzymes related to nitrogen fixation, such as nitrogenase and leghemoglobin, contain iron

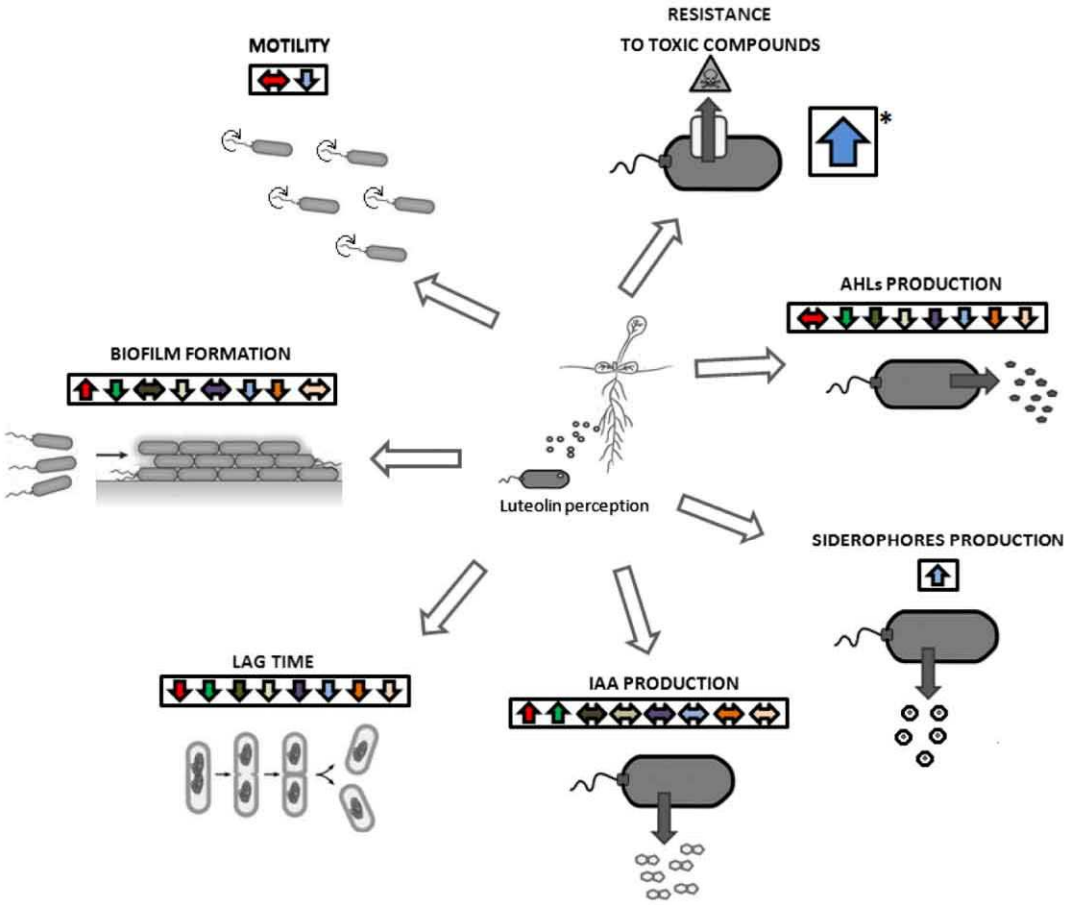


Fig. 8 Effect of luteolin on phenotypes of *E. meliloti* 3001. For each phenotype a set of arrows are reported representing the effect of luteolin in the different media tested. The direction of the arrows represents the influence of luteolin on the tested phenotypes: upward arrow=positive effect, downward arrow=negative effect, left right arrow=no effect. The arrow represents the medium used to test the phenotype and refers to (from left to right): LB, VMM+glucose, VMM+fructose, VMM+ribose, VMM+Na-pyruvate, VMM+Na-succinate,VMM+histidine, VMM+glutamine.The motility assay was performed on LB and VMM+Na succinate (from left to right). To test siderophore production VMM+Na succinate depleted of iron chloride was used.Resistance to toxic compounds (*) was valued towards 240 chemicals by Phenotype Microrarry in VMM+Na-succinate. The main effect was an increase in the resistance to toxic compounds (luteolin modified the sensitivity of *E. meliloti* 3001 towards 20 % of the tested chemicals, among these compounds 82 % was more tolerated by the strain in the presence of luteolin rather than in its absence

as cofactor (Lynch et al. 2001; Persmark et al. 1993). Ampe et al. (2003) found that luteolin significantly induces the expression of three genes involved in iron metabolism in *E. meliloti* 1021. One of these genes was SMa2939, encoding a probable siderophore biosynthesis protein. In order to verify the effective role of luteolin in regulating iron metabolism, we have performed phenotypic experiments to evaluate siderophores production in the *E. meliloti* 3001. Results showed that luteolin stimulated siderophore production in stationary phase cultures, grown in iron deficient medium, suggesting that the host plant might use the flavonoid signal to increase the uptake of iron by the bacterium.

It has been reported that in *E. meliloti* different types of transport systems play an important role in the plant host symbiosis and in rhizobia survival in bulk soil and rhizosphere (Cosme et al. 2008). Such transport systems are involved in the export of extracellular proteins, including hydrolytic enzymes such as carboxymethylcellulose (CMC) hydrolase, and they appear likely to be involved in the primary host infection process during rhizobium-legume symbiosis (Chen et al. 2004; Mateos et al. 2001).

E. meliloti 3001 in minimal medium exhibited a significantly lower motile ability as compared to rich medium. This result is consistent with the negative

regulation of *E. meliloti* motile behavior in response to limited nutrient conditions (Wei and Bauer 1998), and with the sharp down regulation of chemotaxic and flagellar biosynthesis genes expression of *E. meliloti* observed on minimal medium with respect to rich medium (Barnett et al. 2004). The luteolin effect here found on *E. meliloti* 3001 motility on rich medium might indicate a plant strategy to negatively regulate the motility of rhizobia. Such movement inhibition may promote the accumulation of rhizobial cells around the host roots, as highly localized bacterial "clouds" and simultaneously enhance the successful invasion of the roots.

Previous studies have pointed out that plant flavonoids influence biofilm production in rhizobia. In *E. fredii* the isoflavonoid genistein induces the transition from a monolayer-type biolfilm to a microcolony-type biofilm (Perez-Montano et al. 2014). Our finding suggests that the effect of flavonoids on biofilm production depended on the strain analyzed, and indicates that, as observed for other phenotypes tested, luteolin is able to influence biofilm formation in *E. meliloti* 3001 but its effect depends on the nutrient availability.

Overall, our data suggest that *E. meliloti* 3001 developed the ability to activate a complex network of responses following the perception of the plant signal luteolin (Fig. 8). Moreover, nutritional conditions to which the bacterium is exposed significantly impact the response to luteolin. In conclusion, our work has shown that luteolin triggers a pleiotropic response that is possibly unlinked to the nodulation factor biosynthesis and controls several aspects of bacterial physiology unexplored with molecular analysis conducted so far.

Acknowledgments We are grateful to Anke Becker and Matthew McIntosh (LOEWE Center for Synthetic Microbiology, SYNMIKRO, Philipps-Universität Marburg, Germany) for kindly providing the analyzed *E. meliloti* 3001 strain and its 3001 *sinIsinR* mutant, and to Miguel Camara (Faculty of Medicine & Health Sciences-Centre for Biomolecular Sciences, University of Nottingham, UK) for kindly providing the bioreporter *E. coli* strain JM109 pSB1142. Marco Rinaldo Oggioni is gratefully acknowledged for critical reading the manuscript.

References

Agati G, Tattini M (2010) Multiple functional roles of flavonoids in photoprotection. New Phytol 186:786–793

Alexander DB, Zuberer DA (1991) Use of chrome azurol-S reagents to evaluate siderophore production by rhizosphere bacteria. Biol Fertil Soils 12:39–45

Ampe F, Kiss E, Sabourdy F, Batut J (2003) Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. Genome Biol 4:R15

Ardissone S, Noel KD, Klement M, Broughton WJ, Deakin WJ (2011) Synthesis of the flavonoid-induced lipopolysaccharide of *Rhizobium* sp strain NGR234 requires rhamnosyl transferases encoded by genes *rgpF* and *wbgA*. Mol Plant Micr In 24:1513–1521

Arioli S, Guglielmetti S, Amalfitano S, Viti C, Marchi E, Decorosi F, Giovannetti L, Mora D (2014) Characterization of *tetA*-like gene encoding for a major facilitator superfamily efflux pump in *Streptococcus thermophilus*. FEMS Microbiol Lett 355:61–70

Atlas RM, Park LC (1993) Handbook of microbiological media, London

Barnett MJ, Tolman CJ, Fisher RF, Long SR (2004) A dual-genome symbiosis chip for coordinate study of signal exchange and development in a prokaryote-host interaction. Proc Nat Ac Sci USA 101:16636–16641

Batista JSD, Hungria M (2012) Proteomics reveals differential expression of proteins related to a variety of metabolic pathways by genistein-induced *Bradyrhizobium japonicum* strains. J Proteome 75:1211–1219

Becker A, Overloper A, Schluter JP, Reinkensmeier J, Robledo M, Giegerich R, Narberhaus F, Evguenieva-Hackenberg E, Narberhaus F (2014) Riboregulation in plant-associated alpha-proteobacteria. RNA Biol 11:550–562

Biondi EG, Tatti E, Comparini D, Giuntini E, Mocali S, Giovannetti L, Bazzicalupo M, Mengoni A, Viti C (2009) Metabolic capacity of *Sinorhizobium* (*Ensifer*) *meliloti* strains as determined by phenotype MicroArray analysis. Appl Environ Microbiol 75:5396–5404

Bochner BR, Gadzinski P, Panomitros E (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res 11:1246–1255

Broughton WJ, Hanin M, Relic B, Kopcinska J, Golinowski W, Simsek S, Ojanen-Reuhs T, Reuhs B, Marie C, Kobayashi H, Bordogna B, Le Quere A, Jabbouri S, Fellay R, Perret X, Deakin WJ (2006) Flavonoid-inducible modifications to rhamnan O antigens are necessary for *Rhizobium* sp strain NGR234-legume symbioses. J Bacteriol 188:3654–3663

Caetano-Anolles G, Cristestes DK, Bauer WD (1988) Chemotaxis of *Rhizobium meliloti* to the plant flavone luteolin requires functional nodulation genes. J Bacteriol 170:3164–3169

Capela D, Carrere S, Batut J (2005) Transcriptome-based identification of the *Sinorhizobium meliloti* NodD1 regulon. Appl Environ Microbiol 71:4910–4913

diCenzo GC, MacLean A, Milunovic B, Golding GB, Finan T (2014) Examination of prokaryotic multipartite genome evolution through experimental genome reduction. PLoS Genet 10: e1004742.

Chen PJ, Wei TC, Chang YT, Lin LP (2004) Purification and characterization of carboxymethyl cellulase from *Sinorhizobium fredii*. Bot Bull Acad Sinica 45:111–118

Cooper JE (2004) Multiple responses of rhizobia to flavonoids during legume root infection. Adv Bot Res 41:1–62

Cooper JE (2007) Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. J Appl Microbiol 103:1355–1365

Cosme AM, Becker A, Santos MR, Sharypova LA, Santos PM, Moreira LM (2008) The outer membrane protein TolC from

Sinorhizobium meliloti affects protein secretion, polysaccharide biosynthesis, antimicrobial resistance, and symbiosis. Mol Plant-Microbe Interact 21:947–957

Dakora F, Phillips DA (2002) Root exudates as mediators of mineral acquisition in low-nutrient environments. Plant Soil 245:35–47

Del Val C, Rivas E, Torres-Quesada O, Toro N, Jimenez-Zurdo JI (2007) Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. Mol Microbiol 66:1080–1091

Denarie J, Cullimore J (1993) Lipo-oligosaccharide nodulation factors - a minireview new class of signaling molecules mediating recognition and morphogenesis. Cell 74:951–954

Dusha I, Kondorosi A (1993) Genes at different regulatory levels are required for the ammonia control of nodulation in *Rhizobium meliloti*. Mol Gen Genet 240:435–444

Eda S, Mitsui H, Minamisawa K (2011) Involvement of the smeAB multidrug efflux pump in resistance to plant antimicrobials and contribution to nodulation competitiveness in *Sinorhizobium meliloti*. Appl Environ Microbiol 77:2855–2862

Fisher RF, Long SR (1993) Interactions of nodd at the nod box - nodd binds to 2 distinct sites on the same face of the helix and induces a bend in the Dna. J Mol Biol 233:336–348

Galan JE, Lara-Tejero M, Marlovits TC, Wagner S (2014) Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. Ann Rev Microbiol 68:415–438

Galardini M, Mengoni A, Brilli M, Pini F, Fioravanti A, Lucas S, Lapidus A, Cheng JF, Goodwin L, Pitluck S, Land M, Hauser L, Woyke T, Mikhailova N, Ivanova N, Daligault H, Bruce D, Detter C, Tapia R, Han C, Teshima H, Mocali S, Bazzicalupo M, Biondi EG (2011) Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti*. BMC Genomics 12:235

Gao M, Teplitski M, Robinson JB, Bauer WD (2003) Production of substances by *Medicago truncatula* that affect bacterial quorum sensing. Mol Plant-Microbe Interact 16:827–834

Gibson KE, Kobayashi H, Walker GC (2008) Molecular determinants of a symbiotic chronic infection. Annu Rev Genet 42: 413–441

Gonzalez JE, Marketon MM (2003) Quorum sensing in nitrogen-fixing rhizobia. Microbiol Mol Biol Rev 67:574–592

Gonzalez-Pasayo R, Martinez-Romero E (2000) Multiresistance genes of *Rhizobium etli* CFN42. Mol Plant-Microbe Interact 13:572–577

Gordon SA, Weber RP (1951) Colorimetric estimation of indole-acetic acid. Plant Physiol 26:192–195

Guerreiro N, Redmond JW, Rolfe BG, Djordjevic MA (1997) New rhizobium leguminosarum flavonoid-induced proteins revealed by proteome analysis of differentially displayed proteins. MPMI 10:506–516

Hartwig UA, Phillips DA (1991) Release and modification of *Nod*-gene-inducing flavonoids from alfalfa seeds. Plant Physiol 95:804–807

Hartwig UA, Joseph CM, Phillips DA (1991) Flavonoids released naturally from alfalfa seeds enhance growth-rate of *Rhizobium meliloti*. Plant Physiol 95:797–803

Hassan S, Mathesius U (2012) The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. J Exp Bot 63:3429–3444

Honma MA, Asomaning M, Ausubel FM (1990) *Rhizobium meliloti nodD* genes mediate host-specific activation of *nodABC*. J Bacteriol 172:901–911

Jimenez-Zurdo JI, Valverde C, Becker A (2013) Insights into the noncoding RNome of nitrogen-fixing endosymbiotic alphaproteobacteria. Mol Plant-Microbe Interact 26:160–167

Kapulnik Y, Cecillia MJ, Philipps DA (1987) Flavone limitations to root nodulation and symbiotic nitrogen fixation in alfalfa. Plant Physiol 84:1193–1196

Khatri B, Fielder M, Jones G, Newell W, Abu-Oun M, Wheeler PL (2013) High throughput phenotypic analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* strains' metabolism using biolog phenotype microarrays. PLoS One, 8:e52673.

Krol E, Becker A (2014) Rhizobial homologs of the fatty acid transporter FadL facilitate perception of long-chain acyl-homoserine lactone signals. PNAS 111(29):10702–10707

Lang K, Lindemann A, Hauser F, Gottfert M (2008) The genistein stimulon of *Bradyrhizobium japonicum*. Mol Gen Genomics 279:203–211

Le Quere AJL, Deakin WJ, Schmeisser C, Carlson RW, Streit WR, Broughton WJ, Forsberg LS (2006) Structural characterization of a K-antigen capsular polysaccharide essential for normal symbiotic infection in rhizobium sp NGR234 - deletion of the rkpMNO locus prevents synthesis of 5,7-diacetamido-3,5,7,9-tetradeoxy-non-2-ulosonic acid. J Biol Chem 281:28981–28992

Le Rudulier D, Storm AR, Dandekar AM, Smith LT, Valentine RC (1984) Molecular biology of osmoregulation. Science 224(4653):1064–1068

Llamas I, Keshavan N, Gonzalez JE (2004) Use of *Sinorhizobium meliloti* as an indicator for specific detection of long-chain N-acyl homoserine lactones. Appl Environ Microbiol 70:3715–3723

Long SR (2001) Genes and signals in the rhizobium-legume symbiosis. Plant Physiol 125:69–72

Lopez-Baena FJ, Vinardell JM, Perez-Montano F, Crespo-Rivas JC, Bellogin RA, Espuny MR, Ollero FJ (2008) Regulation and symbiotic significance of nodulation outer proteins secretion in *Sinorhizobium fredii* HH103. Microbiology-Sgm 154:1825–1836

Lynch D, O'Brien J, Welch T, Clarke P, Cuiv PO, Crosa JH, O'Connell M (2001) Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. J Bacteriol 183:2576–2585

Mandon K, Østerås M, Boncompagni E, Trinchant J, Spennato G, Poggi MC, Le Rudulier D (2003) Molecular biology of osmoregulation. Mol Plant-Microbe Interact 16:709–719

Mateos PF, Baker D, Petersen M, Velazquez E, Jimenez JI, Martinez-Molina E, Squartini A, Orgambide G, Hubbell DH, Dazzo FB (2001) Erosion of root epidermal cell walls by rhizobium polysaccharide-degrading enzymes as related to primary host infection in the rhizobium-legume symbiosis. Can J Microbiol 47:475–487

McIntosh M, Meyer S, Becker A (2009) Novel *Sinorhizobium meliloti* quorum sensing positive and negative regulatory feedback mechanisms respond to phosphate availability. Mol Microbiol 74:1238–1256

Mnasri B, Badri Y, Saidi S, de Lajudie P, Mhamdi R (2009) Symbiotic diversity of *Ensifer meliloti* strains recovered from

various legume species in Tunisia. Syst Appl Microbiol 32: 583–592

Nievas F, Bogino P, Sorroche F, Giordano W (2012) Detection, characterization, and biological effect of quorum-sensing signaling molecules in peanut-nodulating bradyrhizobia. Sensors 12:2851–2873

Nogales J, Bernabeu-Roda L, Cuellar V, Soto MJ (2012) ExpR is not required for swarming but promotes sliding in *Sinorhizobium meliloti*. J Bacteriol 194:2027–2035

O'Toole GA, Pratt LA, Watnick PI, Newman DK, Weaver VB, Kolter R (1999) Genetic approaches to study of biofilms. Meth Enzimol 310:91–109

Parniske M, Ahlborn B, Werner D (1991) Isoflavonoid-inducible resistance to the phytoalexin glyceollin in soybean rhizobia. J Bacteriol 173:3432–3439

Peck MC, Fisher RF, Long SR (2006) Diverse flavonoids stimulate NodD1 binding to *nod* gene promoters in *Sinorhizobium meliloti*. J Bacteriol 188:5417–5427

Peck MC, Fisher RF, Bliss R, Long SR (2013) Isolation and characterization of mutant *Sinorhizobium meliloti* NodD1 proteins with altered responses to luteolin. J Bacteriol 195: 3714–3723

Perez-Montano F, Guasch-Vidal B, Gonzalez-Barroso S, Lopez-Baena FJ, Cubo T, Ollero FJ, Gil-Serrano AM, Rodriguez-Carvajal MA, Bellogin RA, Espuny MR (2011) Nodulation-gene-inducing flavonoids increase overall production of autoinducers and expression of N-acyl homoserine lactone synthesis genes in rhizobia. Res Microbiol 162:715–723

Perez-Montano F, Jimenez-Guerrero I, Del CP, Baena-Ropero I, Lopez-Baena FJ, Ollero FJ, Bellogin R, Lloret J and Espuny R (2014) The symbiotic biofilm of *Sinorhizobium fredii* SMH12, necessary for successful colonization and symbiosis of *Glycine max* cv osumi, is regulated by quorum sensing systems and inducing flavonoids via NodD1. PLoS One 9: e105901.

Perret X, Staehelin C, Broughton WJ (2000) Molecular basis of symbiotic promiscuity. Microbiol Mol Biol Rev 64:180–201

Persmark M, Pittman P, Buyer JS, Schwyn B, Gill PR, Neilands JB (1993) Isolation and structure of rhizobactin-1021, a siderophore from the alfalfa symbiont rhizobium-meliloti 1021. J Am Chem Soc 115:3950–3956

Peters NK, Frost JW, Long SR (1986) A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science 233:977–980

Peters NK, Frost JW, Long SR (1988) Alfalfa root exudates and compounds which promote or inhibit induction of Rhizobium meliloti nodulation genes. Plant Physiol 88: 396–400

Pocard JA, Vincent N, Boncompagni E, Smith LT, Poggi MC, Le Rudulier D (1997) Molecular characterization of the bet genes encoding glycine betaine synthesis in *Sinorhizobium meliloti* 102F34. Microbiology 143:1369–1379

Rossbach S, Kunze K, Albert S, Zehner S, Gottfert M (2014) The *Sinorhizobium meliloti* EmrAB efflux system is regulated by flavonoids through a TetR-like regulator (EmrR). Mol Plant-Microbe Interact 27:379–387

Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, Cottret L, Carrere S, Sallet E, Courcelle E, Moreau S, Debelle F, Capela D, de Carvalho-Niebel F, Gouzy J, Bruand C, Gamas P (2014) An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J 77: 817–837

Sanchez-Contreras M, Bauer WD, Gao MS, Robinson JB, Downie JA (2007) Quorum-sensing regulation in rhizobia and its role in symbiotic interactions with legumes. Philos T Roy Soc B 362:1149–1163

Santos R, Herouart D, Sigaud S, Touati D, Puppo A (2001) Oxidative burst in alfalfa-*Sinorhizobium meliloti* symbiotic interaction. Mol Plant-Microbe Interact 14:86–89

Schluter JP, Reinkensmeier J, Barnett MJ, Lang C, Krol E, Giegerich R, Long SR, Becker A (2013) Global mapping of transcription start sites and promoter motifs in the symbiotic alfa-proteobacterium *Sinorhizobium meliloti* 1021. BMC Genomics 14:156

Schmeisser C, Liesegang H, Krysciak D, Bakkou N, Le Quere A, Wollherr A, Heinemeyer I, Morgenstern B, Pommerening-Roser A, Flores M, Palacios R, Brenner S, Gottschalk G, Schmitz RA, Broughton WJ, Perret X, Strittmatter AW, Streit WR (2009) *Rhizobium* sp strain NGR234 possesses a remarkable number of secretion systems. Appl Environ Microbiol 75:4035–4045

Shaw LJ, Morris P, Hooker JE (2006) Perception and modification of plant flavonoid signals by rhizosphere microorganisms. Environ Microbiol 8:1867–1880

Smith MJ, Shoolery JN, Schwyn B, Holden I, Neilands JB (1985) Rhizobactin, a structurally novel siderophore from *Rhizobium meliloti*. J Am Chem Soc 107:1739–1743

Spaink HP (1996) Regulation of plant morphogenesis by lipo-chitin oligosaccharides. Crit Rev Plant Sci 15:559–582

Steindler L, Venturi V (2007) Detection of quorum-sensing N-acyl homoserine lactone signal molecules by bacterial biosensors. FEMS Microbiol Lett 266:1–9

Strauss MLA, Jolly NP, Lambrechts MG, van Rensburg P (2001) Screening for the production of extracellular hydrolytic enzymes by non-*Saccharomyces* wine yeasts. J ApplMicrobiol 91:182–190

Talibart R, Jebbar M, Gouesbet G, Himdi-Kabbab S, Wroblewski H, Blanco C, Bernard (1994) Osmoadaptation in rhizobia: ectoine-induced salt tolerance. J Bacteriol 176:5210–5217

Teplitski M, Robinson JB, Bauer WD (2000) Plants secrete substances that mimic bacterial *N*-acyl homoserine lactone signal activities and affect population density-dependent behaviors in associated bacteria. Mol Plant-Microbe Interact 13:637–648

Tolin S, Arrigoni G, Moscatiello R, Masi A, Navazio L, Sablok G, Squartini A (2013) Quantitative analysis of the naringenin-inducible proteome in *Rhizobium leguminosarum* by isobaric tagging and mass spectrometry. Proteomics 13:1961–1972

Vandeputte OM, Kiendrebeogo M, Rasamiravaka T, Stévigny C, Duez P, Rajaonson S, Diallo B, Adeline Mol A, Baucher M, El Jaziri M (2011) The flavanone naringenin reduces the production of quorum sensing-controlled virulence factors in *Pseudomonas aeruginosa* PAOI. Microbiology 157: 2120–2132

Veliz-Vallejos DF, van Noorden GE, Yuan MQ, Mathesius U (2014) A *Sinorhizobium meliloti* specific N-acyl homoserine lactone quorum-sensing signal increases nodule numbers in *Medicago truncatula* independent of autoregulation. Front in Plant Sci 5

Venturi V (2006) Regulation of quorum sensing in *Pseudomonas*. FEMS Microbiol Rev 30:274–291

Vincent J M 1970 A manual for the practical study of root-nodule bacteria (IBP handbook no. 15). Blackwell Scientific Publications, Oxoford, UK.

Viti C, Decorosi F, Marchi E, Galardini M, Giovannetti L (2015) High-throughput phenomics. Methods Mol Biol 1231:99–123

Wei XM, Bauer WD (1998) Starvation-induced changes in motility, chemotaxis, and flagellation of *Rhizobium meliloti*. Appl Environ Microbiol 64:1708–1714

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 5:218–223

Winson MK, Swift S, Fish L, Throup JP, Jorgensen F, Chhabra SR, Bycroft BW, Williams P, Stewart GSAB (1998) Construction and analysis of luxCDABE-based plasmid sensors for investigating N-acyl homoserine lactone-mediated quorum sensing. FEMS Microbiol Lett 163:185–192

Young JPW, Johnston AWB (1989) The evolution of specificity in the legume *Rhizobium* symbiosis. Trends Ecol Evol 4:341–349

Zhang J, Subramanian S, Stacey G, Yu O (2009) Flavones and flavonols play distinct critical roles during nodulation of *Medicago truncatula* by *Sinorhizobium meliloti*. Plant J 57: 171–183

ADDITIONAL FILES

Additional file, ESM_1:

Phenotype MicroArray (PM) analysis on metabolic panels (PM1-2-3-9-10). Area of the PM kinetic curves obtained for each condition after 96 h incubation and the parameter Δarea (as defined in Materials and Methods) used to identify different responses in presence of luteolin.

Additional file, ESM_2:

Phenotype MicroArray (PM) analysis on chemical sensitivity panels (PM11-20). IC50 values obtained for each tested chemical after 96 h incubation in presence and absence of luteolin and the parameter ΔIC50 used to compare the phenotypic profiles.

Chapter 4

NodD-dependent and independent phenotypic responses triggered by the flavonoid luteolin in *Ensifer meliloti*

Submitted to *Research in Microbiology*

NodD- dependent and independent phenotypic responses triggered by the flavonoid luteolin in *Ensifer meliloti*

*Giulia Spini [a], Francesca Decorosi [a], Alice Checcucci [b], Alessio Mengoni [b], Luciana Giovannetti [a], Carlo Viti [a] **

[a] Department of Agrifood Production and Environmental Sciences (DISPAA) – Genexpress laboratory, University of Florence, Via della Lastruccia, 14 - 50019 Sesto Fiorentino (FI), Italy

[b] Department of Biology, University of Florence, Via Madonna del Piano, 6 - 50019 Sesto Fiorentino (FI), Florence, Italy

E-mail address:

giulia.spini@unifi.it

francesca.decorosi@unifi.it

alice.checcucci@unifi.it

alessio.mengoni@unifi.it

luciana.giovannetti@unifi.it

carlo.viti@unifi.it * Correspondence and reprints

Abstract

The early molecular signaling in the symbiosis between the nitrogen-fixing bacterium *Ensifer meliloti* and its host legumes is mediated by the plant flavonoid luteolin, which activates the transcriptional regulators NodD. NodD regulon has been deeply investigated, however a comprehensive scenario of *Ensifer meliloti* phenotypic responses induced by luteolin (dependent or independent from NodDs) remains to be elucidated. To investigate both the NodD-dependent and NodD-indipendent response to luteolin an extensive comparison of the phenotypes of both the wild type strain Rm1021 and the triple *nodD* mutant (A2012) in the presence of the luteolin was performed using Phenotype Microarray (PM) technology. PM results revealed that the utilization of some phosphorus sources were luteolin and NodDs dependent. Interestingly, NodDs-independent modulation of osmotolerance was found. Moreover, several resistance phenotypes to toxic compounds were induced by luteolin, both NodD- dependent and independent. The inactivation of *emrB* efflux pump gene, whose expression is induced by luteolin through ErmR regulator (NodD independent), resulted in an increased susceptibility to a range of toxic compounds and allowed to extend the number of compounds known to be substrate of the *emrB* efflux pump. Moreover, a lower ability to promote plant growth and an altered nodulation efficiency were found for the *emrB* mutant compared to the wild type. Overall, these findings suggest that luteolin plays a wider role in the symbiosis development than just the induction of Nod Factors biosynthesis.

Keywords**:** luteolin; *Ensifer meliloti*; rhizobium-legume symbiosis; NodDs regulators; Phenotype MicroArray (PM); antimicrobials resistance.

1. Introduction

The rhizobium-legumes symbiosis development requires a complex exchange of molecular signals between the two partners. The molecular dialogue involves, in the early stages, the perception of flavonoids present in root exudates, which subsequently activate the rhizobial transcriptional regulator NodD [1]. NodD proteins, which belong to the LysR family of transcriptional regulators, induce the expression of the nodulation genes (*nod genes)* [2]. The *nod* gene products are implicated in the biosynthesis of Nod factors (NFs) that trigger root infection and nodule organogenesis [3]. In some bacterial strains, alternative NodD activators recognizing different plant flavonoids provide an extended host range*.* Bradyrhizobium *japonicum* possesses an alternative two component regulatory pathway for activating its *nod* regulon, and the *nod* gene expression is fine-tuned by positive- and negative-control circuits in *Ensifer* (syn. *Sinorhizobium*) *meliloti* [4]. It is noteworthy that many reports showed that NodD proteins may control other symbiosis-related functions in rhizobia beyond the Nod factor biosynthesis. NodD regulator of *Rhizobium tropici* has been reported to play roles in swarming motility and IAA synthesis [5]. In *Sinorhizobium fredii* and *Bradrhizobium japonicum*, NodD activates a regulator of a type III secretion system (TTSS) that contributes to host range determination [6]. NodD also controls exopolysaccharide biosynthesis genes also [7], lipopolysaccharide modification and indole-3-acetic acid synthesis in *S. fredii* [8;9].
The sequenced genomes of rhizobia contain one to five copies of *nodD* gene. In the species that possess one copy only, such as *Rhizobium leguminosarum* bv*. trifolii*, a mutation usually results in the abolition of nodulation [10]. Whereas, in *Ensifer meliloti*, *R. leguminosarum* bv. *phaseoli* and *B. japonicum*, which have multiple *nodD* copies, the nodulation is not completely suppressed by mutations in a single *nodD* gene. The genome of *E. meliloti* harbors three functional copies of *nodD*, designed *nodD1*, *nodD2* and *nodD3 [11]*. These NodD paralogues are involved in the establishment of the rhizobium-legume symbiosis and in the control of host specificity [12;13]. The transcription factor NodD1, encoded by the constitutively expressed *nodD1* gene, is activated by the flavonoid luteolin. The regulator NodD1 is considered to play the main role in host nodulation by inducing the transcription of the *nodABC* operon involved in the Nod factors (NFs) biosynthesis [11]. *nodD2* has been reported to be luteolin unresponsive and to interfere with the activation of the *nodABC* operon in response to NodD1 leading to a negative effect on Nod factor production [7]. A suppressive role for the transcriptional regulator *nodD2* has been observed also in *Bradyrhizobium japonicum* and *Sinorhizobium fredii* strain NGR234 [14;15]. The transcriptional

regulator NodD3 does not require exogenous compounds to activate *nod* genes. Indeed, the expression of the *nodD3* gene in *E. meliloti* is subject to a complex regulation involving the regulatory protein SyrM, the flavonoid-activated regulator NodD1, and the nitrogen status of rhizobial cells. In turn, NodD3 can activate SyrM expression thus establishing a self-amplifying regulatory circuit [16]. Demont et al. [17] found that NodD3 of *E. meliloti* controls the production of variant acyl groups (18- to 26-carbon *N*-acyl groups with omega-1-OH modifications) that were present in Nod factor preparations. Additional regulators seem to operate for the nodulation gene control constituting further layers of regulation to the NodD-inducer circuit [18-20]. As an example, the transcription of *E. meliloti nodD1, nodD2* and several nod-boxes is negatively regulated by NolR, a repressor of the ArsR family that binds to conserved motifs [21]. The repertoire of NodD-regulated genes in *E. meliloti* was identified by bioinformatic/computational approaches (*e.g,* prediction of *nod* boxes, putative binding sites for NodD) and transcriptomic studies [22;23]. Notably, the NodD1 regulon of *E. meliloti* comprises a number of additional targets beyond *nod* genes. The expression of genes encoding the SyrM transcriptional regulator, the GroES/GroEL chaperones, three hypothetical proteins, conjugal transfer protein TraA, as well as an NTPase essential for the *E. meliloti* infective phenotype, and three genes involved in iron metabolism have been reported to be significantly induced by luteolin-activated NodD1 [24]. Furthermore, EmrAB efflux pump and its putative regulator EmrR have been also found to be regulated via luteolin [25]. We can consequently hypothesize that a number of additional pathways with respect to the only Nod factor biosynthesis are triggered by luteolin perception. In fact, we recently reported a large variety of phenotypes to be induced in *E. meliloti* by luteolin, as an enhanced resistance toward antibiotics and others antimicrobial agents, a reduction of AHLs and indole-3-acetic acid (IAA) production [26]. This work was then aimed to investigate the extent of substrate utilization phenotypes and toxic compounds tolerance in *E. meliloti*, triggered by the flavonoid luteolin perception that are dependent or independent from the NodD regulatory circuit using two mutants derived from *E. meliloti* Rm1021 [24]. The *E. meliloti* triple *nodD* A2012 mutant, which provided a genetic background devoid of all three NodD regulators, and the *E. meliloti emrB* mutant, to investigate the contribution of the *emrAB* efflux pump to the phenotypes not mediated by NodDs.

2. Materials and methods

2.1. Bacterial strains and growth conditions

The bacterial strains of *E. meliloti* used in this study are reported in Table 1. *E. meliloti* strains were grown at 30°C either in TY [27] or in Vincent minimal medium (VMM) [28] supplemented with 0.2% of Na-succinate as a carbon source. Antibiotics were used at the following final concentrations: neomycin, 100 µg/ml, and streptomycin 200 µg/ml. The stock solution of luteolin was prepared at a final concentration of 3.18 mM in a 9 mM NaOH dissolving solution. The working concentration of luteolin was 10 µM, known to be the required concentration *in vitro* for NodD induction [29][30]. Untreated cultures were supplemented with an equal volume of the 9 mM NaOH dissolving solution rather than luteolin.

2.2. Phenotype MicroArray (PM) and data analysis

The growth of *E. meliloti* strains was tested in 1,437 different culture conditions using PM metabolic (PM1, PM2, PM3, PM4) and chemical sensitivity panels (PM9-PM20). The tested conditions included carbon and nitrogen sources, several concentrations of ions and osmolytes, pH stress, and a wide variety of antibiotics, anti-metabolites, heavy metals and other inhibitors. PM11-PM20 allowed assaying for the sensitivity to 240 chemical agents at four concentrations. The complete list of the compounds assayed can be obtained at http://www.biolog.com/pdf/PM1-PM10.pdf. PM uses tetrazolium violet reduction as a reporter of active metabolism. The reduction of the dye causes the formation of a purple color that, recorded every 15 min, provides quantitative and kinetic information about the response of the cells in the PM plates [24]. *E. meliloti* strains were grown at 30°C on Biolog Universal Growth agar (BUG) (Biolog Inc, Hayward CA, US) for two days. Then colonies were picked up with a sterile cotton swab and suspended in 1x IF-0 (Biolog) until OD_{600}=0.1. Inoculation fluid for PM1 and PM2 was obtained diluting the cellular suspension (OD_{600}=0.1) 10 times in an appropriate volume of VMM supplemented with 1x Dye Mix A (Biolog). The inoculation fluid for PM3 was prepared diluting the cellular suspension (OD_{600}=0.1) 10 times in VMM without ammonium chloride, and supplemented with 0.2% Na-succinate as a carbon source and with 1x Dye Mix A (Biolog). The inoculation fluid for PM4 was prepared diluting the cellular suspension (OD_{600}=0.1) 10 times in IF-0 1x (Biolog) supplemented with 0.2% Na-succinate as a carbon source and with 1x Dye Mix A (Biolog). The inoculation fluid for PM9-10 was prepared diluting the cellular suspension (OD_{600}=0.1) 10 times in VMM supplemented with 0.2% Na-succinate as carbon source and with 1x Dye

Mix A (Biolog). The inoculation fluid for PM11-PM20 was prepared diluting the cellular suspension (OD_{600}=0.1) 13.64 times in VMM supplemented with 0.2% Na-succinate as a carbon source and with 1x Dye Mix A (Biolog). PM plates were inoculated with 100 µl per well. To test the influence of luteolin on the phenotype of the strains, luteolin was added to the inoculation fluids, at a final concentration of 10 µM, according to other *in vitro* induction assays [13;23;24]. All the PM experiments were performed in duplicate, as two independent experiments. PM panels were incubated statically at 30°C in an Omnilog Reader (Biolog) for 96 hours. The kinetic profiles for the *E. meliloti* wild type and mutant strains were analyzed by inspecting kinetic curves and compared using the Omnilog-PM software (release OM_PM_109M). In order to discard possible false errors, the set of criteria reported by Kathri et al. [31] were applied to the PM data analysis. The dedicated software DuctApe [32] was applied to the PM results to find the differences between the phenotype profiles of the *E. meliloti* strains in more detail. A single and concise parameter, Activity Index (AV), was calculated to rank and compare each kinetic curve, providing both qualitative and quantitative information about the ability to grow in a specific culture condition. The AV parameter was obtained through a k-means clustering (with k clusters) on five growth curve parameters (max, area, average height, lag time, and slope). Therefore, an AV value equal to zero indicates a curve with no metabolic activity, while higher AV values will be assigned to curves with increasing levels of metabolic activity. The difference of AV(ΔAV) was used as discriminating parameter for comparing the kinetic curves obtained in order to identify different responses on metabolic and chemical sensitivity panels (PM1-20). Any differences and thus the related compounds, lower than the standard threshold of the parameter ($\Delta AV \geq |2|$) was considered not significant and discarded.

2.3. Minimal inhibitory concentration assays

E. meliloti strains were subjected to an array of toxic compounds at different concentrations in VMM with 0.2% Na-succinate both in absence and in presence of 10 µM luteolin in order to establish the minimal inhibition concentration (MIC) of each chemical. The toxic chemicals tested were: polymyxin B, compound 48/80, chlorhexidine, 8-hydroxyquinoline, 5,7-dichloro- 8-hydroxy quinaldine, benzalkonium chloride, dequalinium. Toxic chemicals were prepared as solutions at 10 times the highest concentration desired in the working plate and sterilized by filtration. Subsequently, each solution was serially twofold diluted along twelve wells of a 96-well

plate, to obtain a 10X solution series (twelve concentration points). For each concentration series, a negative control without a toxic chemical was also included. Working solutions (1x) were freshly prepared by diluting the stock solutions (10x) with the appropriate medium (VMM+ 0.2% Na-succinate) into 96-wells standard microplates. Each well was inoculated with a bacterial culture in the late exponential phase to obtain an optical density at 600 nm (OD_{600}) of 0.1 (10^7 CFU ml^{-1}). For each chemical, cultures for MIC determination were set up both supplemented and not supplemented with 10 µM luteolin. The microplate was statically incubated at 30°C and growth data (OD_{590} determination) were collected after 24 h of incubation using a microtiter plate spectrophotometer (programmable MPT reader DV 990 BV5, GDV). Each experiment was performed in four replicates.

2.4. Nodulation assays

Seeds of *Medicago sativa* (cv. Pomposa) were sterilized in $HgCl_2$, repeatedly washed, and germinated in sterile plastic Petri dishes for 72 h in the dark and 48 h in the light in a growth chamber maintained at 26°C with a 16-h photoperiod (100 microeinstein m^{-2} s^{-1}). For *in vitro* assays, seedlings were transferred in Petri dishes containing Buffered Nod Medium [33] and 16 g/l of type A agar (Sigma-Aldrich). Plantlets were grown for an additional 3 to 5 days before inoculation with the *E. meliloti* strains. For nodulation assays strains were grown in liquid TY medium at 30°C for 48 h, then washed three times in 0.9% NaCl solution and resuspended to an OD_{600} =1.0 ($1{\cdot}10^7$cells/ml) (in single or a 1:1 ratio of the two strains for competition experiments). One hundred µl of the standardized bacterial suspension was spread over the seedling root (resulting about $4{\cdot}10^4$cells/cm^2). Plates were pierced to let the plant grow outside, and transferred in a near-vertical position to the growth chamber Each experiment was performed in eight replicates.

2.5 Plant symbiotic-related phenotypes

The effect of symbiotic interaction on plant growth was evaluated using conventional parameters, such as the number of nodules per plant, length of aerial part [34] and kinetics of nodule formation. Data were expressed as mean of eight replicates ± standard deviation. Statistically significant differences were detected by ANOVA analysis ($P<0.05$).

2.6. Estimation of bacterial loads in nodules

Bacteria within nodules were estimated by viable titers using standard cultivation method. The *emrB* mutant cells were discriminated from the wild type cells by selective plating on TY plates containing neomycin to which the *emrB* mutant is resistant. Single nodules of the same size (about 1 mm in length) were excised from plants after four weeks of growing, their surface was sterilized with 0.1% NaHClO for 30" then washed three times in sterile distilled water, crushed and resuspended in 100 µl of 0.9% NaCl solution. Aliquots of serial dilutions and the third wash water (as control) were then plated on TY+ neomycin plates, incubated at 30°C for 48 h and the numbers of CFU was determined. Data were expressed as mean of eight replicates ± standard deviation. Statistically significant differences between were detected by ANOVA analysis ($P<0.05$).

3. Results and discussion

3.1 Detection of luteolin-responsive phenotypes, NodD-dependent and NodD-independent, in E. meliloti Rm1021

The wild type Rm1021 of *E. meliloti* and its triple *nodD* mutant A2012 were tested in presence and absence of the luteolin on a thousand of different growth conditions by the application of the high-throughput Phenotype MicroArray (PM) technology. The tested conditions included carbon, nitrogen, sulfur and phosphorous sources, several concentrations of ions and osmolytes, pH stresses, and a wide variety of antibiotics and toxic compounds. To identify the overall phenotypic responses triggered by the luteolin both in a NodDs dependent and independent manner these steps were followed: i) the comparison of the phenotypic profiles of Rm1021 in presence and absence of luteolin allowed to identify all the luteolin dependent phenotypes, detectable by PM system, ii) luteolin dependent phenotypes ruled by the NodDs regulation pathway were identified by the comparison of PM profiles of Rm1021 and A2012 strains both cultured in presence of luteolin, iii) the phenotypes induced by luteolin but not dependent by NodDs were defined “NodD independent phenotypes” and were supposed to be under the control of different regulation pathways. The whole metabolic and sensitivity pattern of *E. meliloti* Rm1021 in the presence and in absence of luteolin obtained by PM analysis is reported in Fig.1 (circular plot). The complete data set of the PM experiments can be found in Table S1. *E. meliloti* Rm1021 was found to be able to metabolize 44% of tested carbon sources, 88% of nitrogen sources, 80% of sulfur sources and 93% of phosphorous sources, in both the absence or presence of the luteolin. The profile of metabolic abilities obtained for the wild type strain Rm1021 revealed an overall phenotypic similarity for all the carbon, nitrogen, sulfur and phosphorous sources in presence as well as in absence of luteolin (Fig. 1). This finding indicated that plant inducer luteolin did not affect carbon, nitrogen as well as sulfur and phosphorous metabolism, as also pointed out in the *E. meliloti* 3001 strain [26]. Since rhizobia are exposed to fluctuating environmental conditions and must cope with several stresses in the rhizosphere, the effect of the luteolin on the metabolic activity of *E. meliloti* Rm1021 under several osmolyte gradients and pH conditions was evaluated [35]. The pH range where *E. meliloti* Rm1021 exhibited active metabolism was between 6 and 10, with an optimal pH value around 7.0. *E. meliloti* Rm1021 displayed an overlapping trend of tolerance under different osmolyte gradients and pH condition in the presence and in absence of luteolin [26]. An increased osmotolerance associated with luteolin presence was observed for the Rm1021 strain to 20% ethylene glycol (Fig. 1, Table S1). An analogous increased of osmotolerance,

triggered by luteolin, was observed for the *E. meliloti* 3001 strain [26], supporting the hypothesis of a luteolin involvement in mediating osmoprotection beside the standard stress tolerance systems. The uptake of osmoprotector peptides through ABC-transporters has been reported to confer osmoprotection in many rhizobia [36]. We can speculate that the observed osmotolerance of the strain grown in presence of luteolin might be due to the luteolin-mediated activation of some osmoprotective systems [26]. The *E. meliloti* Rm1021 strain was also analyzed for chemical sensitivity in the presence of luteolin to hundreds of antimicrobials and toxic compounds, each one presenting four different concentration levels. The chemical sensitivity profile of *E. meliloti* Rm1021 showed different phenotypic responses associated with the luteolin treatment on 22% of the tested compounds (Fig.1). They included substances belonging to antibiotics (17) toxic ions (6), chelators (4), membrane damagers (9), antimicrobial agents (7), oxidizing agents and respiration inhibitors (8), listed in Table 2. As also reported for other rhizospheric microorganisms, plant-derived flavonoids play multiple roles, depending on their structures, beside Nod Factors (NFs) biosynthesis, such as to inhibit several phytopathogens and evoke a strong chemoattractant response in rhizobia toward plant roots [37]. Moreover, efflux pumps and other resistance systems may be activated by plant flavonoids that in this way might offer a further level of protection and adaptation for symbionts in the soil [26]. The repertoire of adaptive changes mediated by luteolin pointed out PM analysis supported that the luteolin might exert an additional physiologically important role concerning oxidative and abiotic stress conditions as well as coping with antimicrobial and toxic compounds in the rhizosphere. Such pleiotropic luteolin effect constitutes an advantage in the selective rhizospheric soil to establish an effective symbiosis.

To detect phenotypic changes induced by the luteolin that are dependent on NodDs regulatory circuit, the extensive phenotypic analysis using PM, outlined above, was performed on the *E. meliloti* A2012 strain compared to the wild type Rm1021 upon the luteolin treatment. Comparison of the metabolic profile of the A2012 mutant with that of the wild type strain Rm1021 in the presence of luteolin revealed an overlapping trend for all the carbon, nitrogen, sulfur and phosphorous sources tested. Nevertheless, some different growth phenotypes associated with the triple *nodD* deletion were observed for some P sources (Fig.2). The A2012 mutant strain displayed a reduced activity on 2-deoxy-D-glucose-6-phosphate, revealing its tendency to metabolize this substrate with lower efficiency relative to the Rm1021 strain. Conversely, the A2012 mutant metabolized D-mannose-6-phosphate and phosphocreatine sources, on which the wild type Rm1021 did not display active metabolism (Table S1). Phosphorus is an essential nutrient, present at low concentration in most soils and free-living bacteria satisfy the

demand for this element by uptake of soluble inorganic and organic phosphate compounds. The scenario evolving from previous studies on rhizobia has suggested that phosphate uptake is due to at least two phosphate transport systems, which are differentially expressed under different growth conditions [38]. The involvement of cellular and extracellular phosphatases in response to phosphate deficiency has also been reported [39-42]. *E. meliloti* Rm1021 was found to be very efficient in utilizing unusual phosphorous compounds [33;38]. Gene expression profiles of the *E. meliloti* phosphate starvation response revealed eleven regulatory genes belonging to different families that might be responsible for secondary effects in the complex regulatory network activated under conditions of phosphate starvation [38]. The ChvI-ExoS two component system resulted one of them and its expression was recently shown to be linked to the NodD3-SyrM regulatory circuit [16]. Differences in transcriptional regulation, most likely NodD dependent, might explain the differences in P-sources utilization observed between the wild type Rm1021 and the A2012 mutant.

Under the osmolyte gradients and pH conditions tested using PM, the wild type Rm1021 and A2012 mutant strain displayed a similar overall profile (Fig. 2), indicating that *nodD* genes deletion did not affect the *E. meliloti* stress tolerance. The *E. meliloti* Rm1021 and A2012 strains were also analyzed for chemical sensitivity-resistance toward hundreds of antimicrobial and toxic compounds in presence of the inducing compound luteolin. A global overview of chemical sensitivity patterns, obtained for *E. meliloti* Rm1021 wild type and A2012 mutant strain is reported in Fig. 2. The comparison of chemical sensitivity profile of the strains highlighted different phenotypic responses associated with the deletion of the *nodD* genes on 16% (38/240) of the tested compounds (Table S1). The A2012 mutant strain compared to the wild type Rm1021 displayed a lower metabolic activity for 78% (30/38) of tested compounds. The A2012 mutant was found to be more sensitive for a broad set of chemicals as antibiotics (8), oxidizing agents (6), toxic anions (8), chelators (2), membrane damagers (6), and other antimicrobial agents (2) (Table 3). On the other hand, A2012 mutant showed a higher tolerance for 3,5-dinitrobenzene (respiration inhibitor), 4-aminopyridine (ion channel inhibitor), 1-hydroxy-piridine-2-thione (chelator) and three antibiotics (amoxicillin, chloramphenicol, cefmetazole). The increased susceptibility of A2012 suggests that NodD proteins contribute to mediate resistance toward antimicrobial and toxic compounds likely trough the regulation of gene encoding efflux pumps. This hypothesis is supported by the finding that genes encoding for MSF type pump were found to be luteolin-responsive and NodD1 dependent [24]. The activation of efflux pumps is a bacterial strategy to survive in the rhizosphere in which several antimicrobial compounds are released by the plant. This

strategy might confer an advantage to *E. meliloti* to gain an efficient nodulation of the host plant, as reported for *R. tropici* [5].

The comparison of luteolin dependent and luteolin-NodDs dependent phenotypes *of E. meliloti* Rm1021 showed that the phenotypes partially differ. An increased tolerance to the ethylene glycol osmolyte was found to be a luteolin dependent phenotype but not a luteolin-NodD dependent phenotype. Similarly, the resistance of *E. meliloti* Rm1021 to some toxic compounds was found to be luteolin dependent but not luteolin-Nods dependent, suggesting that luteolin modulates resistance in *E. meliloti* through a direct effect on resistance systems and/or indirectly through a regulation system that not involves NodD regulators. In order to verify these resistance phenotypes that are induced by luteolin but not dependent on NodD regulators, minimal inhibitory concentrations (MIC) experiments were performed. The MICs experiment were carried out on the wild type Rm1021 and A2012 mutant strains both in presence and in absence of luteolin (Table 4) The wild type and A2012 mutant strains, treated with luteolin compared to the untreated ones, were found more tolerant towards crystal violet, chlorhexidine, compound 48/80, polymyxin B, dequalinium chloride, benzalkonium chloride, 8 hydroxyquinoline and 5,7-dicloro-hydroxyquinaldine. Both strains showed an increased MIC to the toxic compounds tested upon luteolin treatment with the only exception of dequalinium chloride (Table 4). The obtained results indicate that the detected resistance phenotypes correlate with the luteolin presence but are independent from NodDs regulatory pathways. This observation suggested other targets, either regulators or cellular structures, modulated by the luteolin beyond NodD regulators. We can hypothesize the presence of an additional luteolin-dependent signal transduction pathway, independent from NodD. Indeed, flavonoids has been shown to induce a Ca^{2+} spikes, upstream to NodD activation in *Mesorhizobium loti* and *R. eguminosarum* bv. *viciae* [43-45], suggesting the presence of a perception of flavonoid not mediated by NodD. Indeed, similarly to other plant flavonoids, luteolin was found to induced in *E. meliloti* the expression of novels regulatory genes whose function of regulation has not been examined in detailed yet [24;25;46]. Moreover, at least three small non-coding RNA involved in the control of gene transcription were identified to be controlled by luteolin, suggesting a luteolin regulatory activity at the post transcriptional level [47-49]. Cannot be also excluded that the phenotype differences with and without luteolin were a consequence of the interaction between luteolin and possibly cellular structures, which indirectly could even turn out to hamper the gene expression.

3.2 Phenotypic and symbiotic profiling of E. meliloti emrB mutant strain

The high number of phenotypes related to the antimicrobial and toxic compounds resistance mediated by luteolin, strongly suggested the involvement of efflux pump as main mediators of such luteolin-induced changes. To shed some light on this hypothesis, the contribution of the EmrAB efflux system, reported to be induced by luteolin in *E. meliloti* througth ErmR regulator [25], was investigated. The susceptibility profile of the *emrB* mutant respect to the wild type was extensively characterized through Phenotype MicroArray (PM) high-throughput technology.

The comparison of chemical sensitivity profiles of the wild type Rm1021 and *emrB* mutant strains upon luteolin induction revealed that *emrB* gene inactivation increased the susceptibility to toxic compounds (29) and antibiotics (17) (Fig. 3)(Table S1). The *emrB* mutant strain displayed an enhanced susceptibility respect to the wild type toward compounds belong to classes already known to be efflux targets of transporters belonging to the MFS family [50;51], such as quaternary ammonium compounds-QACs (methyltrioctylammonium chloride, domiphen bromide, cetylpyridium chloride), bisguanides (alexidine, chlorexidine), bis-phenols and dyes (triclosan, umbelliferone, crystal violet, 2-phenylphenol, 2,4-dinitrophenol)(Table 5). In addition, the *emrB* mutant was more sensitive than the wild type Rm1021 to classes of compounds previously not associated with MFS transporters. These include chelators (EDTA, 1,10-phenanthroline, 5,7-dichloro-8-hydroxyquinaldine), inhibitors (b-chloro-L-alanine, guanazole, azathioprine) and antimicrobial agents (1-hydroxypyridine-2-thione, hydroguaiaretic acid, patulin, coumarin) (Table 5). A higher resistance for the mutant strain respect to the wild type was revealed in the presence of chlorambucil (chelator), hexachlorophene (respiration inhibitor) and 9-Aminoacridine (dye). The altered phenotypic response exhibited by the *emrB* mutant strain toward compounds outlined above, suggested that such compounds are likely preferred substrates transported by this efflux pump. Although these compounds have different structures, transporters of the MSF are known to be promiscuous in substrate recognition and efflux [52;53]. The MDR pumps ACrAB of *E. coli* and the MexAB of *P. aeruginosa* have demonstrated to play major roles in making these bacteria intrinsically resistance to most classes of antibiotics and compounds [54-56]. The increased susceptibility displayed by the strain lacking the *emrB* multidrug efflux pump gene indicates the involvement of the *emrAB* efflux system in the resistance of *E. meliloti* to a range to antimicrobial and toxic compounds. The extrusion of toxic compounds by MDR pumps might benefit rhizobia enabling bacterial cells to cope with the effects of naturally occurring chemicals in their environment [52;57-59]. This hypothesis is consistent with the observations reported for other rhizobia. In

Rhizobium etli, the flavonoids-inducible RmrAB pump was found to be involved in antimicrobial resistance [60] as well as the RND-type pump, BdeAB, of *B. japonicum* [61].

Several studies have pointed out the contribution of the flavonoids-inducible MDR efflux pumps to the extrusion-mediated resistance to antimicrobials and thereby to the colonization ability during symbiotic or virulent plant-bacteria interaction [62;63]. Consequently, to examine the involvement of *emrB* in nodulation and competition for symbiosis, the symbiotic phenotype of the *emrB* mutant respect to the wild type Rm1021 was evaluated both in single and competition assays. Nodulation assay was performed. *M. sativa* plants were inoculated with single and a mixture of both wild type and mutant strains. After 40 days of growth, the plants nodulated by the only wild type Rm1021 were 60% whereas those nodulated by the *emrB* mutant were 40%. Plants inoculated with the *emrB* mutant resulted significantly reduced in size compared to the plants inoculated with the wild type Rm1021 (Fig. 4a)(1-way ANOVA, $p<0.05$). Then, results indicated that the *emrB* mutant was able to induce nodule formation, but the number of nodules produced was significantly lower compared to the wild type strain (Fig. 4b)(1-way ANOVA, $p<0.05$). Additionally, *emrB* mutant formed root nodules more slowly than the wild type Rm1021 (Figure 5). These data suggested a lower ability of the *E. meliloti emrB* mutant to promote plant growth together with a reduced nodulation efficiency respect to the wild type strain. The defective mutualism displayed by the mutant indicates the importance of antimicrobial resistance mediated by efflux in the nodulation ability of the *E. meliloti*. This evidence supports what suggested by Gonzalez-Pasayo and coworker [60] that MDR pumps prevent the accumulation of toxic and plant-derived compounds within bacterial cells and thereby increase rhizobia fitness in the rhizosphere. Moreover, evidences that MDR efflux pumps contributed to establishing a successful interaction in plant-symbiotic rhizobia are also consistent with our observations reported above [59;61;62]. When the wild type Rm1021 and *emrB* mutant were co-inoculated in the same host plant, the defective nodulation behavior showed by the *emrB* mutant turned out to be compensated by the presence of the wild typethe mutant did not affect the behaviour of wild type. Indeed, the resulted symbiotic phenotype was more similar to that of the wild type Rm1021, for both number of nodules formed and plants size (Figs. 4a and 4b). Two main hypothesis can be formulated to explain this result: i) in the co-inoculated plants the mutant is completely overcome by the presence of the wild type that acts as the only competitor in the symbiotic process [64]; ii) the interaction between the two strains fits the mutant to the wild type in the establishment of the symbiosis [65]. To deepen and clarify the scenario of the interaction between mutant and wild type during the symbiotic process, the occupancy of the nodules by each strain was evaluated. Viable cells were

counted from each nodule formed by single and mixed inocula. Comparing the viable cells resulting from the plating of the *emrB* mutant nodules with that from the wild type, the mutant seemed able to colonize nodules in number significantly lower than wild type (Table 6)(1-way ANOVA, $p<0.05$). This result could suggest an altered efficiency in nodule colonization and in bacterial fitness for *emrB* mutant that is in agreement with the lower number of nodules and the smaller size of the plants inoculated with this strain (Figs. 4a and 4b). The results obtained by the plating of mixed inocula nodules didn't show a significantly difference in the presence of the two strains. Indeed, the nodule occupancy resulted of 51% for the wild type and 48% for the *emrB* mutant. This data may suggest that *emrB* mutant is equally competitive as the wild type 1021 strain in terms of nodule occupancy when both are present in the nodule. This data was in line with that observed by Capela et al. [24] and supported the hypothesis that the wild type strain able to efficiently colonized host plant and harboring a full-functioning network of efflux systems could help the less performance mutant to nodule colonization. Additionally, almost all the deletion mutants of *E. meliloti* in single multidrug resistance (MDR) systems investigated by Eda and coworkers were found competitive as the wild type strain in nodule occupancy [59]. Can be excluded that the *EmrAB* efflux pump could act as a system that can be shared by the rhizobia population, being not only a system that extrudes toxic compounds to protect the single cells but also a system that exports outside one or more endogenous cellular compounds required for the nodule induction. Indeed, it appears that apart from extruding antimicrobial compounds, efflux systems may also contribute to the export of intrinsic, potentially beneficial, molecules that can help rhizobia to survive the root hair microenvironment and indispensable in plant-microorganisms interactions [66;67].

4. Conclusions

Luteolin is well known as an inducer of the NodD transcriptional factor in *E. meliloti* thereby triggering the biosynthesis of Nod factor and a successful nodulation of the host plant. Nevertheless, this study has shown that luteolin can elicit additional responses suggesting another major role in *E. meliloti* Rm1021, which consists in modulating osmolytes and chemical resistance in accordance with what reported for *E. meliloti* 3001 [21]. The detected resistance phenotypes were found to be NodD dependent or NodD independent suggesting that i) NodD is involved in the regulation of resistance systems ii) the flavonoid luteolin has other regulatory targets, different from NodD, which confer to

E. meliloti the ability to tolerate toxic compounds and osmolytes. EmrR has been identified as a transcriptional regulator ruled by luteolin that acts as a repressor of the EmrAB efflux pumps. This study has clarified that EmrAB could act as an efflux pump protecting the bacterium from the effect of toxic chemicals, but it is likely that it can export bacterial compounds that have a role in the symbiosis establishment.

Conflict of interest

The authors declare no conflict of interest

Acknowledgements

We are grateful to Delphine Capela (Laboratoire des Interactions Plantes-Microorganismes, UMR INRA-CNRS Castanet-Tolosan Cedex, France) for kindly providing the analyzed A2012 and *emrB* mutant strains of *E. meliloti*.

References

1. J.E.Cooper, Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue, Journal of Applied Microbiology 103 (2007) 1355-1365.

2. M.C.Peck, R.F.Fisher, and S.R.Long, Diverse flavonoids stimulate NodD1 binding to nod gene promoters in *Sinorhizobium meliloti*, J. Bacteriol. 188 (2006) 5417-5427.

3. G.E.D.Oldroyd, Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants, Nature Reviews Microbiology 11 (2013) 252-263.

4. J.Stougaard, Regulators and regulation of legume root nodule development, Plant Physiology 124 (2000) 531-540.

5. P.del Cerro, A.A.P.Rolla-Santos, D.F.Gomes, B.B.Marks, F.Perez-Montano, M.A.Rodriguez-Carvajal, A.S.Nakatani, A.Gil-Serrano, M.Megias, F.J.Ollero, and M.Hungria, Regulatory nodD1 and nodD2 genes of *Rhizobium tropic*i strain CIAT899 and their roles in the early stages of molecular signaling and host-legume nodulation, BMC Genomics 16 (2015).

6. V.Viprey, A.Del Greco, W.Golinowski, W.J.Broughton, and X.Perret, Symbiotic implications of type III protein secretion machinery in Rhizobium, Molecular Microbiology 28 (1998) 1381-1389.

7. D.Machado and H.B.Krishnan, nodD alleles of *Sinorhizobium fredii* USDA191 differentially influence soybean nodulation, nodC expression, and production of exopolysaccharides, Current Microbiology 47 (2003) 134-137.

8. H.Kobayashi, Y.Naciri-Graven, W.J.Broughton, and X.Perret, Flavonoids induce temporal shifts in gene-expression of nod-box controlled loci in *Rhizobium* sp NGR234, Molecular Microbiology 51 (2004) 335-347.

9. M.Theunis, H.Kobayashi, W.J.Broughton, and E.Prinsen, Flavonoids, NodD1, NodD2, and nod-box NB15 modulate expression of the y4wEFG locus that is required for indole-3-acetic acid synthesis in *Rhizobium sp* strain NGR234, Molecular Plant-Microbe Interactions 17 (2004) 1153-1161.

10. M.A.Djordjevic, P.R.Schofield, and B.G.Rolfe, Tn5 mutagenesis of *Rhizobium trifolii* host-specific nodulation genes result in mutants with altered host-range ability, Molecular & General Genetics 200 (1985) 463-471.

11. Z.Gyorgypal, N.Iyer, and A.Kondorosi, 3 Regulatory Nodd Alleles of Diverged Flavonoid-Specificity Are Involved in Host-Dependent Nodulation by Rhizobium-Meliloti, Molecular & General Genetics 212 (1988) 85-92.

12. M.A.Honma and F.M.Ausubel, *Rhizobium meliloti* has three functional copies of the nodD symbiotic regulatory gene, Proc. Natl. Acad. Sci. U S. A 84 (1987) 8558-8562.

13. M.A.Honma, M.Asomaning, and F.M.Ausubel, *Rhizobium meliloti nodD* genes mediate host-specific activation of *nodABC*, Journal of Bacteriology 172 (1990) 901-911.

14. J.Loh and G.Stacey, Nodulation gene regulation in *Bradyrhizobium japonicum:* A unique integration of global regulatory circuits, Applied and Environmental Microbiology 69 (2003) 10-17.

15. R.Fellay, M.Hanin, G.Montorzi, J.Frey, C.Freiberg, W.Golinowski, C.Staehelin, W.J.Broughton, and S.Jabbouri, NodD2 of *Rhizobium* sp NGR234 is involved in the repression of the *nodABC* operon, Molecular Microbiology 27 (1998) 1039-1050.

16. M.J.Barnett and S.R.Long, The *Sinorhizobium meliloti* SyrM Regulon: effects on global gene expression are mediated by SyrA and NodD3, Journal of Bacteriology 197 (2015) 1792-1806.

17. N.Demont, M.Ardourel, F.Maillet, D.Prome, M.Ferro, J.C.Prome, and J.Denarie, The Rhizobium-Meliloti Regulatory Nodd3 and Syrm Genes Control the Synthesis of A Particular Class of Nodulation Factors N-Acylated by (Omega-1)-Hydroxylated Fatty-Acids, Embo Journal 13 (1994) 2139-2149.

18. G.Stacey, S.Luka, J.Sanjuan, Z.Banfalvi, A.J.Nieuwkoop, J.Y.Chun, L.S.Forsberg, and R.Carlson, Nodz, A Unique Host-Specific Nodulation Gene, Is Involved in the Fucosylation of the Lipooligosaccharide Nodulation Signal of *Bradyrhizobium japonicum*, Journal of Bacteriology 176 (1994) 620-633.

19. J.Sanjuan, P.Grob, M.Gottfert, H.Hennecke, and G.Stacey, NodW Is Essential for Full Expression of the Common Nodulation Genes in *Bradyrhizobium japonicum*, Molecular Plant-Microbe Interactions 7 (1994) 364-369.

20. T.C.Dockendorff, J.Sanjuan, P.Grob, and G.Stacey, NolA represses Nod gene-expression in *Bradyrhizobium japonicum*, Molecular Plant-Microbe Interactions 7 (1994) 596-602.

21. M.Cren, A.Kondorosi, and E.Kondorosi, NolR controls expression of the *Rhizobium meliloti* nodulation genes involved in the core Nod Factor synthesis, Molecular Microbiology 15 (1995) 733-747.

22. S.Tolin, G.Arrigoni, R.Moscatiello, A.Masi, L.Navazio, G.Sablok, and A.Squartini, Quantitative analysis of the naringenin-inducible proteome in *Rhizobium leguminosarum* by isobaric tagging and mass spectrometry, Proteomics 13 (2013) 1961-1972.

23. J.P.Schluter, J.Reinkensmeier, M.J.Barnett, C.Lang, E.Krol, R.Giegerich, S.R.Long, and A.Becker, Global mapping of transcription start sites and promoter

motifs in the symbiotic a-proteobacterium *Sinorhizobium meliloti* 1021, BMC Genomics 14 (2013).

24. D.Capela, S.Carrere, and J.Batut, Transcriptome-based identification of the *Sinorhizobium meliloti* NodD1 regulon, Appl. Environ. Microbiol. 71 (2005) 4910-4913.

25. S.Rossbach, K.Kunze, S.Albert, S.Zehner, and M.Gottfert, The *Sinorhizobium meliloti* EmrAB efflux system is regulated by flavonoids through a TetR-like regulator (EmrR), Mol. Plant Microbe Interact. 27 (2014) 379-387.

26. G.Spini, F.Decorosi, M.Cerboneschi, S.Tegli, L.Giovannetti, and C.Viti, Effect of the plant flavonoid luteolin on *Ensifer meliloti* 3001 phenotypic responses, Plant and Soil In Press (2015).

27. Beringer JE, R factor transfer in *Rhizobium leguminosarum*, Journal of General Microbiology (1974).

28. Vincent JM, A manual for the practical study of the root-nodule bacteria, journal of basic microbiology (1972).

29. Y.Kapulnik and D.A.Phillips, Effect of Flavones on Nod Genes Expression and Improved Nodulation in Rhizobium Alfalfa Interaction, Phytoparasitica 15 (1987) 144.

30. N.K.Peters and S.R.Long, Alfalfa Root Exudates and Compounds Which Promote Or Inhibit Induction of *Rhizobium meliloti* Nodulation Genes, Plant Physiology 88 (1988) 396-400.

31. B.Khatri, M.Fielder, G.Jones, W.Newell, M.bu-Oun, and P.R.Wheeler, High Throughput Phenotypic Analysis of Mycobacterium tuberculosis and Mycobacterium bovis Strains' Metabolism Using Biolog Phenotype Microarrays, PLoS One 8 (2013).

32. M.Galardini, A.Mengoni, E.G.Biondi, R.Semeraro, A.Florio, M.Bazzicalupo, A.Benedetti, and S.Mocali, DuctApe: a suite for the analysis and correlation of genomic and OmniLog Phenotype Microarray data, Genomics 103 (2014) 1-10.

33. E.G.Biondi, E.Tatti, D.Comparini, E.Giuntini, S.Mocali, L.Giovannetti, M.Bazzicalupo, A.Mengoni, and C.Viti, Metabolic capacity of *Sinorhizobium (Ensifer) meliloti* strains as determined by Phenotype MicroArray analysis, Appl. Environ. Microbiol. 75 (2009) 5396-5404.

34. F.Pini, G.Spini, M.Galardini, M.Bazzicalupo, A.Benedetti, M.Chiancianesi, A.Florio, A.Lagomarsino, M.Migliore, S.Mocali, and A.Mengoni, Molecular phylogeny of the nickel-resistance gene nreB and functional role in the nickel sensitive symbiotic nitrogen fixing bacterium *Sinorhizobium meliloti*, Plant and Soil 377 (2014) 189-201.

35. N.K.Peters, J.W.Frost, and S.R.Long, A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes, Science 233 (1986) 977-980.

36. E.T.Kiers, R.A.Rousseau, S.A.West, and R.F.Denison, Host sanctions and the legume-rhizobium mutualism, Nature 425 (2003) 78-81.

37. G.Caetanoanolles, D.K.Cristestes, and W.D.Bauer, Chemotaxis of *Rhizobium Meliloti* to the plant flavone luteolin requires functional nodulation genes, Journal of Bacteriology 170 (1988) 3164-3169.

38. Schachtman DP, Phosphorus uptake by plants: from soil to cell, Plant Physiology 116 (1998) 447-453.

39. E.Krol and A.Becker, Global transcriptional analysis of the phosphate starvation response in *Sinorhizobium meliloti* strains 1021 and 2011, Molecular Genetics and Genomics 272 (2004) 1-17.

40. O.Geiger, V.Rohrs, B.Weissenmayer, T.M.Finan, and J.E.Thomas-Oates, The regulator gene phoB mediates phosphate stress-controlled synthesis of the membrane lipid diacylglyceryl-N,N,N-trimethylhomoserine in *Rhizobium (Sinorhizobium) meliloti*, Molecular Microbiology 32 (1999) 63-73.

41. J.C.Tarafdar and A.Jungk, Phosphatase-Activity in the Rhizosphere and Its Relation to the Depletion of Soil Organic Phosphorus, Biology and Fertility of Soils 3 (1987) 199-204.

42. J.C.Tarafdar, A.V.Rao, and K.Bala, Production of Phosphatates by Fungi Isolated from Desert Soils, Folia Microbiologica 33 (1988) 453-457.

43. R.Moscatiello, S.Alberghini, A.Squartini, P.Mariani, and L.Navazio, Evidence for calcium-mediated perception of plant symbiotic signals in aequorin-expressing Mesorhizobium loti, Bmc Microbiology 9 (2009).

44. R.Moscatiello, A.Squartini, P.Mariani, and L.Navazio, Flavonoid-induced calcium signalling in *Rhizobium leguminosarum* bv. *viciae*, New Phytologist 188 (2010) 814-823.

45. G.Arrigoni, S.Tolin, R.Moscatiello, A.Masi, L.Navazio, and A.Squartini, Calcium-Dependent Regulation of Genes for Plant Nodulation in Rhizobium leguminosarum Detected by iTRAQ Quantitative Proteomic Analysis, Journal of Proteome Research 12 (2013) 5323-5330.

46. F.Ampe, E.Kiss, F.Sabourdy, and J.Batut, Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis, Genome Biol. 4 (2003) R15.

47. C.del Val, E.Rivas, O.Torres-Quesada, N.Toro, and J.I.Jimenez-Zurdo, Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics, Molecular Microbiology 66 (2007) 1080-1091.

48. B.Roux, N.Rodde, M.F.Jardinaud, T.Timmers, L.Sauviac, L.Cottret, S.Carrere, E.Sallet, E.Courcelle, S.Moreau, F.Debelle, D.Capela, F.de Carvalho-Niebel, J.Gouzy, C.Bruand, and P.Gamas, An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing, Plant Journal 77 (2014) 817-837.

49. J.I.Jimenez-Zurdo, C.Valverde, and A.Becker, Insights into the Noncoding RNome of Nitrogen-Fixing Endosymbiotic alpha-Proteobacteria, Molecular Plant-Microbe Interactions 26 (2013) 160-167.

50. T.G.Littlejohn, I.T.Paulsen, M.T.Gillespie, J.M.Tennent, M.Midgley, I.G.Jones, A.S.Purewal, and R.A.Skurray, Substrate specificity and energetics of antiseptic and disinfectant resistance in *Staphylococcus aureus*, FEMS Microbiol. Lett. 74 (1992) 259-265.

51. B.A.Mitchell, M.H.Brown, and R.A.Skurray, QacA multidrug efflux pump from *Staphylococcus aureus:* Comparative analysis of resistance to diamidines, biguanidines, and guanylhydrazones, Antimicrobial Agents and Chemotherapy 42 (1998) 475-477.

52. P.Van Dillewijn, J.Sanjuan, J.Olivares, and M.J.Soto, The tep1 gene of *Sinorhizobium meliloti* coding for a putative transmembrane efflux protein and N-acetyl glucosamine affect nod gene expression and nodulation of alfalfa plants, Bmc Microbiology 9 (2009).

53. O.Lewinson, J.Adler, N.Sigal, and E.Bibi, Promiscuity in multidrug recognition and transport: the bacterial MFS Mdr transporters, Molecular Microbiology 61 (2006) 277-284.

54. M.C.Sulavik, C.Houseweart, C.Cramer, N.Jiwani, N.Murgolo, J.Greene, B.DiDomenico, K.J.Shaw, G.H.Miller, R.Hare, and G.Shimer, Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes, Antimicrobial Agents and Chemotherapy 45 (2001) 1126-1136.

55. X.Z.Li and H.Nikaido, Efflux-Mediated Drug Resistance in Bacteria An Update, Drugs 69 (2009) 1555-1623.

56. X.Z.Li, H.Nikaido, and K.Poole, Role of Mexa-Mexb-Oprm in Antibiotic Efflux in Pseudomonas-Aeruginosa, Antimicrobial Agents and Chemotherapy 39 (1995) 1948-1953.

57. J.L.Martinez, M.B.Sanchez, L.Martinez-Solano, A.Hernandez, L.Garmendia, A.Fajardo, and C.varez-Ortega, Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems, FEMS Microbiol. Rev. 33 (2009) 430-449.

58. L.J.Piddock, Multidrug-resistance efflux pumps - not just for resistance, Nat. Rev. Microbiol. 4 (2006) 629-636.

59. S.Eda, H.Mitsui, and K.Minamisawa, Involvement of the SmeAB multidrug efflux pump in resistance to plant antimicrobials and contribution to nodulation competitiveness in *Sinorhizobium meliloti*, Appl. Environ. Microbiol. 77 (2011) 2855-2862.

60. R.Gonzalez-Pasayo and E.Martinez-Romero, Multiresistance genes of *Rhizobium etli* CFN42, Molecular Plant-Microbe Interactions 13 (2000) 572-577.

61. A.Lindemann, M.Koch, G.Pessi, A.J.Muller, S.Balsiger, H.Hennecke, and H.M.Fischer, Host-specific symbiotic requirement of BdeAB, a RegR-controlled RND-type efflux system in *Bradyrhizobium japonicum*, Fems Microbiology Letters 312 (2010) 184-191.

62. J.D.Palumbo, C.I.Kado, and D.A.Phillips, An isoflavonoid-inducible efflux pump in Agrobacterium tumefaciens is involved in competitive colonization of roots, Journal of Bacteriology 180 (1998) 3107-3113.

63. A.Burse, H.Weingart, and M.S.Ullrich, The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, *Erwinia amylovora*, Molecular Plant-Microbe Interactions 17 (2004) 43-54.

64. B.B.Somasegaran P, Single strain versus multi-strains inoculation- Effect of soil mineral N availability on rhizobial strain effectiveness and competition for nodulation on Chickpea, Soybean and dry Bean, Applied and Environmental Microbiology 56 (1990) 3298-3303.

65. R.F.Denison, Legume sanctions and the evolution of symbiotic cooperation by rhizobia, American Naturalist 156 (2000) 567-576.

66. T.Cai, W.T.Cai, J.Zhang, H.M.Zheng, A.M.Tsou, L.Xiao, Z.T.Zhong, and J.Zhu, Host legume-exuded antimetabolites optimize the symbiotic rhizosphere, Molecular Microbiology 73 (2009) 507-517.

67. T.R.de Kievit, M.D.Parkins, R.J.Gillis, R.Srikumar, H.Ceri, K.Poole, B.H.Iglewski, and D.G.Storey, Multidrug efflux pumps: expression patterns and contribution to antibiotic resistance in *Pseudomonas aeruginosa* biofilms, Antimicrob. Agents Chemother. 45 (2001) 1761-1770.

68. H.M.Meade, S.R.Long, G.B.Ruvkun, S.E.Brown, and F.M.Ausubel, Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis, J. Bacteriol. 149 (1982) 114-122.

Fig. 1. Circular plot representing the whole metabolic and chemical sensitivity profile of *E. meliloti* Rm1021 strain in presence and in absence of *nodDs* inducer luteolin. The activity index (AV) calculated for the wild-type strain Rm1021 for each PM condition is reported as color stripes going from red (AV = 0) to green (AV = 5) (inner ring). The external ring reports the phenotypic differences detected for the wild-type Rm1021 in presence of luteolin (AV difference ≥threshold); purple stripes indicate an higher activity for the strain treated with luteolin compared to the untreated one; orange stripes indicate a lower activity (more details on the calculation of the AV can be found in "Material and methods" section).

Fig. 2. Circular plot representing the whole metabolic and chemical sensitivity profile of *E. meliloti* A2012 mutant in comparison with the wild-type Rm1021. The activity index (AV) calculated for the wild-type strain Rm1021 for each PM tested condition is reported as color stripes going from red (AV = 0) to green (AV = 5) (inner ring). The external ring reports the phenotypic differences detected for the A2012 mutant compared to the wild-type Rm1021 (AV difference ≥threshold); purple stripes indicate an higher activity with respect to the selected wild-type 1021 strain; orange stripes indicate a lower activity (more details on the calculation of the AV can be found in " Material and methods" section).

Fig. 3. **Circular plot** representing the whole chemical sensitivity profile of *E. meliloti emrB* mutant in comparison with the wild-type Rm1021 upon luteolin induction. The activity index (AV) calculated for the wild-type Rm1021 strain for each PM chemicals is reported as color stripes going from red (AV = 0) to green (AV = 5) (inner ring). The external ring reports the phenotypic differences detected for the *emrB* mutant compared to the wild-typeRm1021 (AV difference ≥threshold); purple stripes indicate an higher activity with respect to the selected wild-type strain; orange stripes indicate a lower activity (more details on the calculation of the AV can be found in "Material and methods" section).

Fig. 4. Symbiotic phenotypes of the *E. meliloti* Rm1021 wild-type, *emrB* mutant and co-inoculated strains; **a)** Mean length of aerial parts of *M. sativa* plants inoculated with the indicated strains is reported. **b)** Mean of nodules number, four weeks after

inoculation of *M. sativa* plants with the indicated strains, is reported. Error bars indicate standard deviation from eight replicates for each condition. Different letters above bars indicate significant different means after 1-way ANOVA (P<0.05).

Fig. 5. Nodulation kinetic. Rate of root nodules induced by the *E. meliloti* strain Rm1021 (●) and the *emrB* mutant (■). Means of nodules ± standard deviation of eight independent experiments for each conditions are reported. * statistically significant differences between data at the same sampling time at the level of P-value<0.05 (one-way ANOVA).

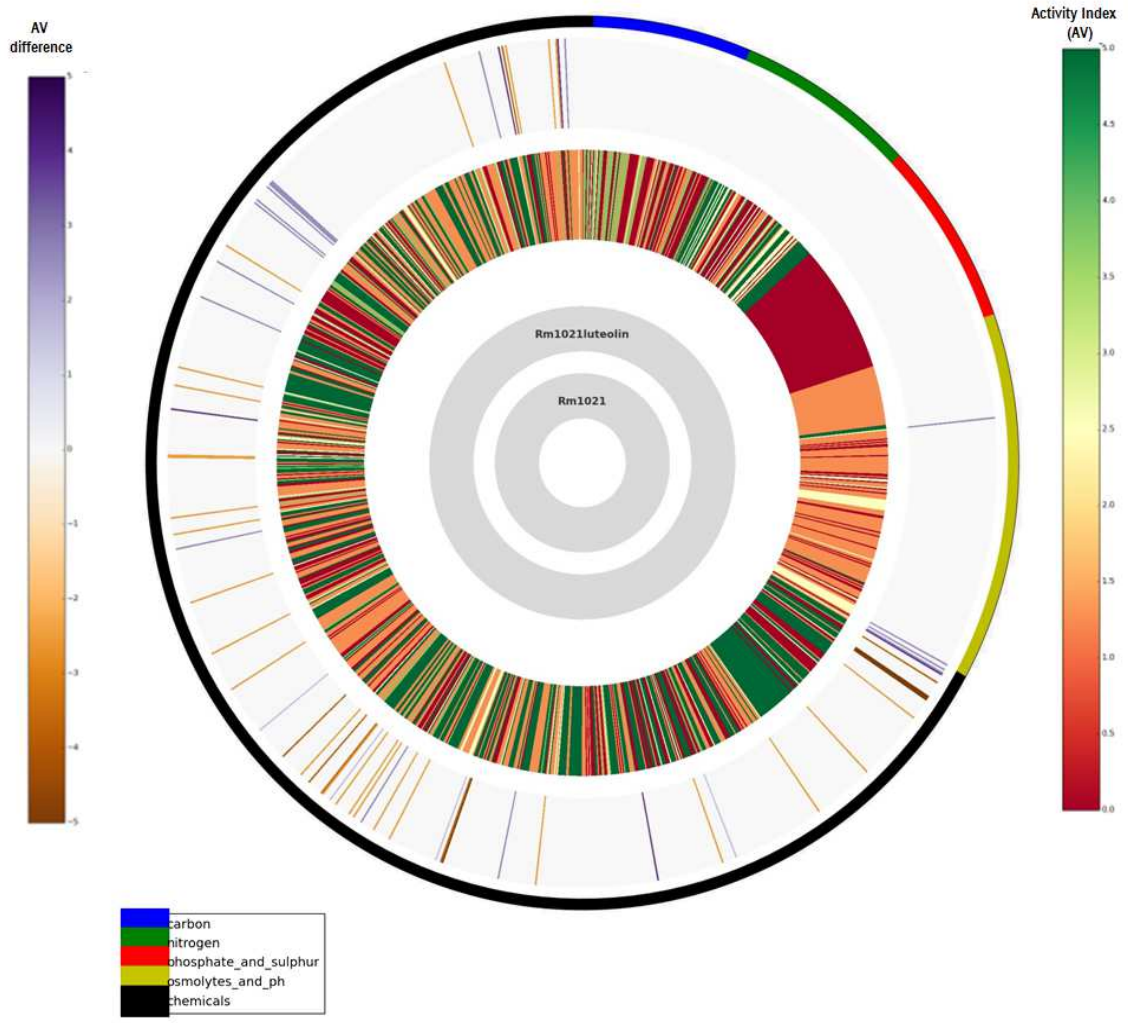

AV
difference
Activity Index
(AV)
Rm1021luteolin
Rm1021
carbon
nitrogen
phosphate_and_sulphur
osmolytes_and_ph
chemicals

Figure 1

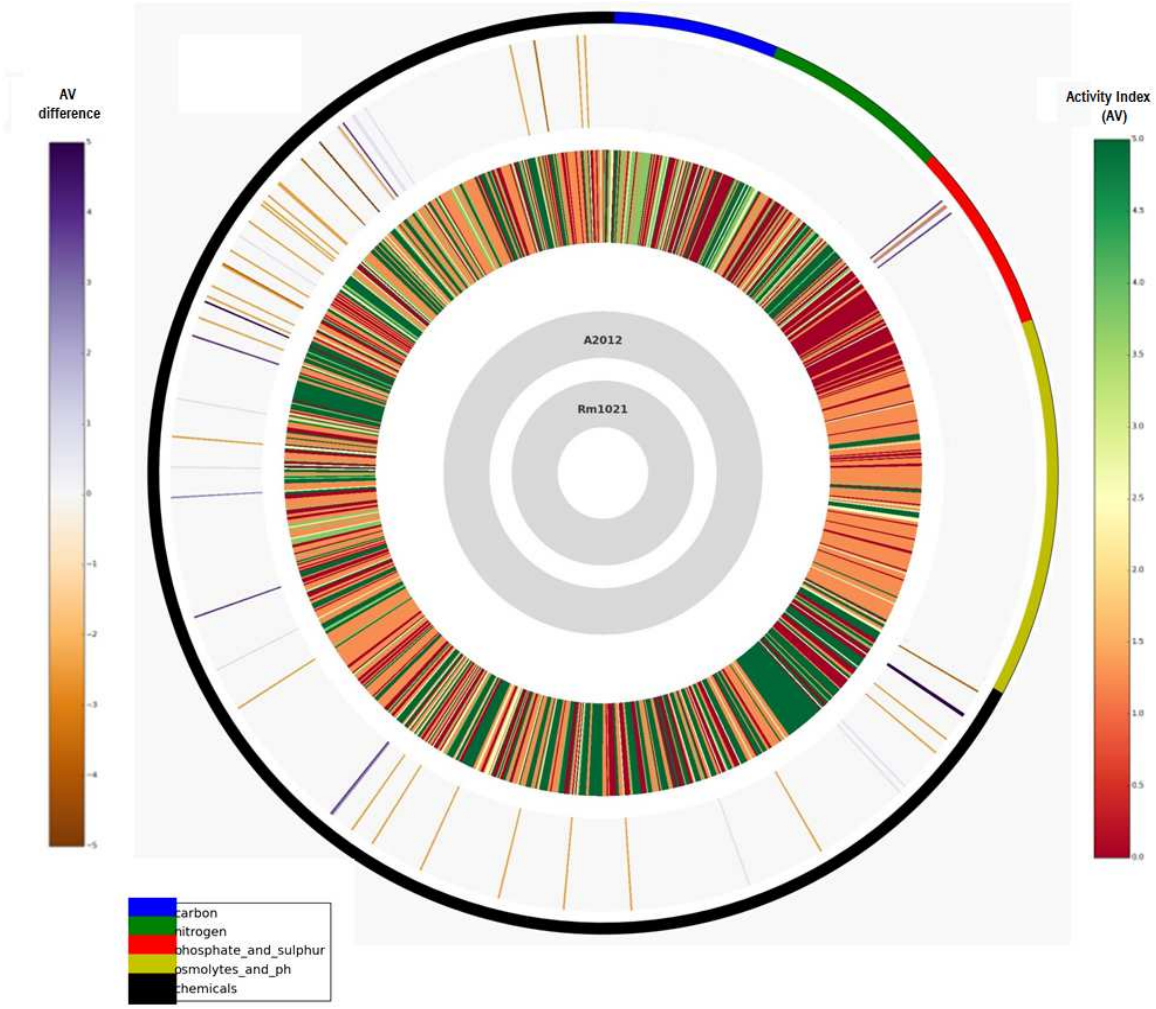

AV
difference
Activity Index
(AV)
A2012
Rm1021
carbon
nitrogen
phosphate_and_sulphur
osmolytes_and_ph
chemicals

Figure 2

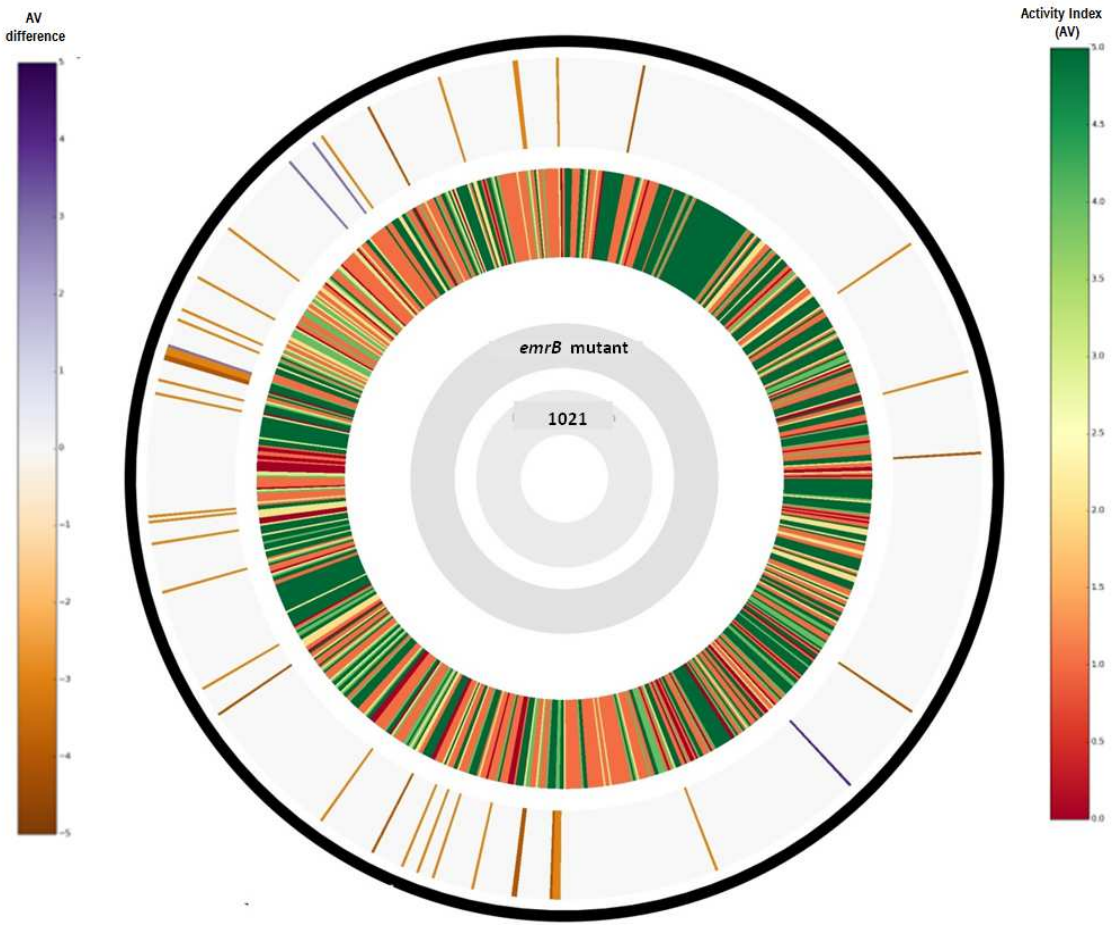


Figure 3

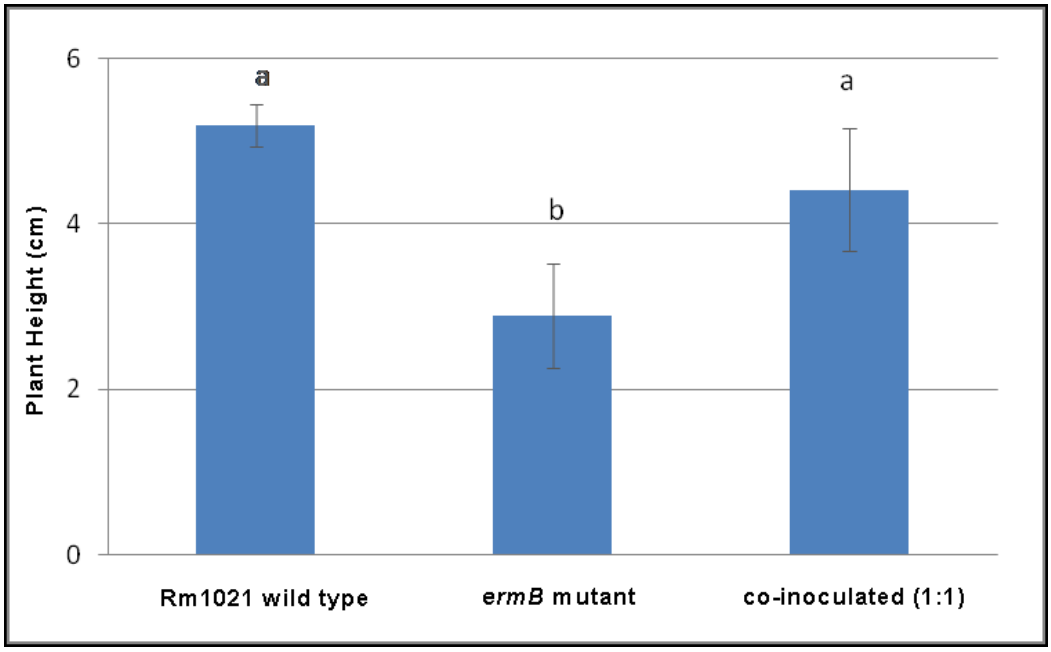


Figure 4a

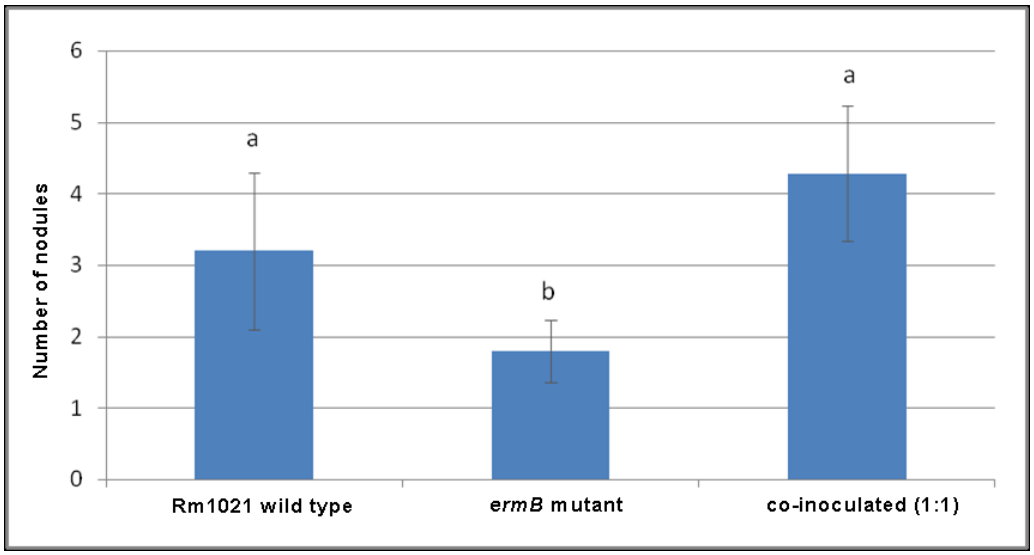


Figure 4b

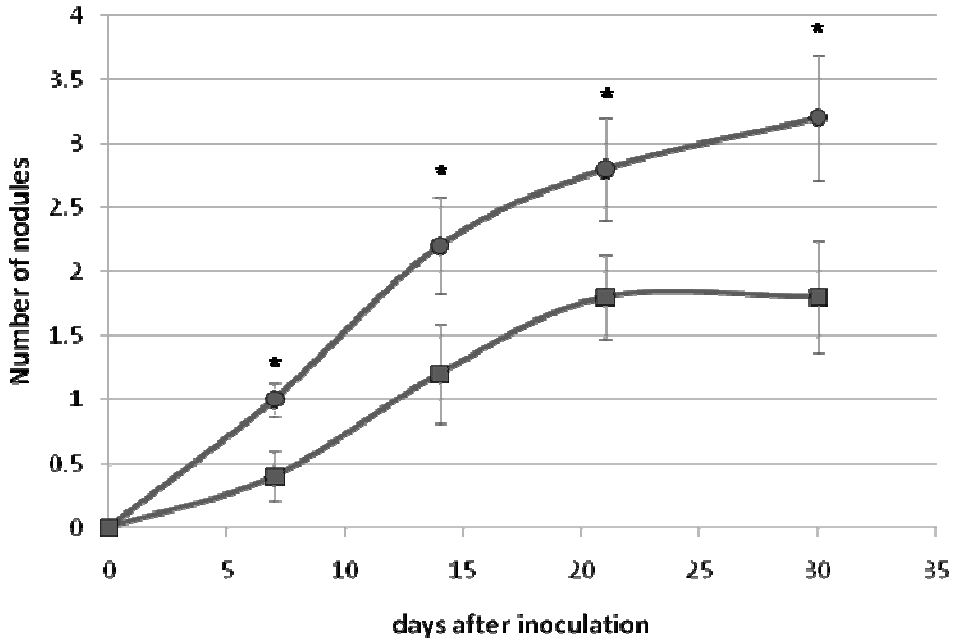


Figure 5

Table 1. Strains used in this work.

Strains	Description	Reference
E. meliloti Rm1021	wild-type Str^r derivative of SU47	[68]
E. meliloti A2012	Rm1021 *nodD1 nodD2 nodD3*	[13]
E. meliloti emrB	Rm1021 *SMc03167* Neo^r	[24]

Table 2. Compounds identified through PM analysis towards which the presence of luteolin induced an increase or a decrease of resistance phenotype in *E. meliloti* Rm1021 strain.

Compound	Class	Compound	Class
Amoxicillin	Antibiotics	Promethazine	Fungicide
Chlortetracycline	Antibiotics	Dequalinium	Ion channel inhibitor
Bleomycin	Antibiotics	4-Aminopyridine	Ion channel inhibitor
Tetracycline	Antibiotics	Benzethonium chloride	Membrane damage
L-Aspartic acid hydroxamate	Antibiotics	Procaine	Membrane damage
Cefmetazole	Antibiotics	Methyltrioctylammonium chloride	Membrane damage
Norfloxacine	Antibiotics	Polymyxin B	Membrane damage
L-Glutamic acid hydroxamate	Antibiotics	Colistin	Membrane damage
Pentachlorophenol	Antibiotics	Alexidine	Membrane damage
Thiamphenicol	Antibiotics	Niaproof	Membrane damage
Josamycin	Antibiotics	Domiphen bromide	Membrane damage, fungicide
Chlorexidine	Antibiotics	Dodine	Membrane damage, fungicide
Disulfiram	Antibiotics	2-Nitroimidazole	Oxidizing agent
Dihydrostreptomycin	Antibiotics	3,4-Dimethoxybenzyl alcohol	Oxidizing agent
Thiamphenicol	Antibiotics	Tinidazole	Oxidizing agent
Azathioprine	Antibiotics	Chelerithrine	QACs
3,5-Diamino-1,2,4-triazole	Antibiotics	Cmidazole	Oxidizing agent
Chlerythrine	Ant microbial agent	Plumbagin	Plant-derived oxidizing agent
Compound 48/80	Ant microbial agent	3,5-Dinitrobenzene	Respiration inhibitors
5-Fluoro-5-deoxyuridine	Ant microbial agent	2,4-Dinitrophenol	Respiration inhibitors
Myricetin	Ant microbial agent	Cristal viole:	Respiration Inhibitors
Coumarin	Ant microbial agent	Sodium bromate	Toxic Arion
Hydroxylamine	Ant microbial agent	Potassium chromate	Toxic Arion
Phenyl-methylsulforyl-fluoride	Ant microbial agent	Sodium orthovanadate	Toxic Arion
1-Hydroxypyridine-2-thione	Chelator	Sodium metasilicate	Toxic Arion
5,7-Dichloro-8-hydroxyquinoline	Chelator	Sodium arsenate	Toxic Arion
EDTA	Chelator	Sodium dichromate	Toxic Arion
8-hydroxyquinoline	Chelator		

Table 3. Compounds identified through PM analysis towards which A2012 mutant strain displayed a higher sensitivity or a higher resistance than wild-type Rm1021 upon luteolin induction.

Compound	Class
Capreomycin	Antibiotics
Cefoxitin	Antibiotics
Thiamphenicol	Antibiotics
Azathioprine	Antibiotics
Sulfisoxazole	Antibiotics
3,5- Diamino-1,2,4-triazole	Antibiotics
Carbenicillin	Antibiotics
Dihydrostreptomycin	Antibiotics
Amoxicillin	Antibiotics
Cefmetazole	Antibiotics
Chloramphenicol	Antibiotics
Patulin	Fungicide
Chlorhexidine diacetate	Membrane damage
Methyltrioctylammonium chloride	Membrane damage
Colistin	Membrane damage
Alexidine	Membrane damage
Niaproof	Membrane damage
Myricetin	Mutagenic agent
Furaltadone	Oxidizing agent
2-Nitroimidazole	Oxidizing agent
Tinidazole	Oxidizing agent
Plumbagin	Oxidizing agent
3,4-dimethoxybenzyl alcohol	Oxidizing agent
Ornidazole	Oxidizing agent
3,5-dinitrobenzene	Inhibitors
4-aminopyridine	Inhibitors
1-hydroxy-piridine-2-thione	Chelator
Sodium bromate	Toxic anion
Potassium chromate	Toxic anion
Sodium orthovanadate	Toxic anion
Sodium metasilicate	Toxic anion
Sodium arsenate	Toxic anion
Sodium dichromate	Toxic anion
Sodium metavanadate	Toxic anion

Table 4. MICs of toxic compounds detected for the wild type1021 and the triple *nodDs* mutant A2012 in absence and presence of 10 µM luteolin.

Compound	Description	MIC (µg/ml)[a]			
		Wild-type 1021		A2012	
		0 µM luteolin	10 µM luteolin	0 µM luteolin	10 µM luteolin
Polimixin B	Cationic surfactant	3.5	5	2.8	4
Compound 8/80	Phosphodiesterase inhibitor	10	20	8	20
Chlorhexidine	Membrane damage	2.5	5	2.5	5
8-hydroxyquinoline	Chelator	50	200	50	100
5,7-dichloro-8-hidroxy-quinaldine	Chelator	85	100	76	100
Benzalkonium chloride	Cationic membrane detergent	9	11.5	6	10
Dequalinium chloride	Ion channel inhibitor	14	25	15	15

[a] **MICs** were determined for the *E. meliloti* Rm1021 and A2012 mutant strains in VMM +Na-succinate both in absence and in presence of 10 µM luteolin. Each MIc determination was repeat four times.

Table 5. Compounds identified through PM analysis towards which *emrB* mutant displayed a higher sensitivity or a higher resistance than wild-type Rm1021 in presence of luteolin.

Compound	Class	Compound	Class
Plumbagin	Plant antimicrobial	Sorbic acid	Respiration inhibitor
Chromium (III) chloride	Toxic cation	Pentachlorophenol	Respiration inhibitor
Alexidine	Bisguanidines	3,5-Dinitrobenzoic acid	Respiration inhibitor
Chlorehexidine	Bisguanidines	Hexachlorophene	Respiration inhibitor
Umbelliferone	Dyes	Potassium tellurite	Toxic anion
Crystal violet	Dyes	Lithium chloride	Toxic cation
2,4-Dinitrophenol	Phenols	Sodium metasilicate	Toxic anion
2- Phenylphenol	Phenols	Chloramphenicol	Antibiotics
Myricetin	Poliphenols	5-Nitro-2-furaldehyde semicarbazone	Antibiotics
Domiphen bromide	QACs	Norfloxacin	Antibiotics
Cetylpyridinium chloride	QACs	L-Glutamic acid g-monohydroxamate	Antibiotics
Methyltrioctylammoniu m chloride	QACs	Sulfachloropyridazine	Antibiotics
Lincomycin	Antibiotics	Sulfamonomethoxine	Antibiotics
Chloramphenicol	Antibiotics	Thiamphenicol	Antibiotics
Cephalothin	Antibiotics	Sulfisoxazole	Antibiotics
Penicillin G	Antibiotics	Fusaric acid	Chelator
Sulfadiazine	Antibiotics	5,7-Dichloro-8-hydroxyquinoline	Chelator
Phleomycin	Antibiotics	Chlorambucil	Chelator
Cefotaxime	Antibiotics	1-Hydroxypyridine-2-thione	Fungicide
Streptomycin	Antibiotics	Nordihydroguaiaretic acid	Fungicide
Azathioprine	Antibiotics	Patulin	Fungicide
Coumarin	Antimicrobial	9-Aminoacridine	Dye
Orphenadrine	Antimicrobial	Triclosan	Phenols (-bis)
Amitriptyline	Membrane damage	Chelerithrine	QACs

Table 6. Bacteria loads inside the nodules of single strain inoculum.

	[a]Viable cells/nodule
Rm1021	1332 (± 472)
***emrB* mutant**	731 (± 262) *

[a] Values as determined by vital count are the mean ± standard deviation of eight independent experiments. * statistically significant differences at the level of P-value<0.05 (one-way ANOVA) compared with wild-type Rm1021

ADDITIONAL FILE

Additional file, Table S1:

Data set of the PM analysis on *E. meliloti* strains obtained on metabolic plates (PM1-PM10) and on chemical sensitivity plates (PM11-PM20). The AV value and the AV difference for each one of the *E. meliloti* strains in each tested growth condition are reported.

Chapter 5

Role of the LuxR-like transcriptional regulator SMc00658 in *Ensifer meliloti*

Role of the LuxR-like transcriptional regulator SMc00658 in *Ensifer meliloti*

in collaboration with the Professor Becker's team of the LOEWE Center for Synthetic Microbiology (SYNMIKRO) at the University of Marburg (Germany)

Introduction

The *SMc00658* gene of *E. meliloti* is predicted to encode a LuxR-like transcriptional regulator involved in the cell-density dependent intercellular signaling, known as *Quorum Sensing* (QS). The QS system is reported to play a significant role in regulation of a variety of genes responsible for physiological traits related both to the free-living and symbiotic state [1;2]. The QS system of *E. meliloti* is composed by a LuxI-family synthase responsible for synthesizing the signal molecule (autoinducer) that then interacts at a quorum concentration with the cognate LuxR-family transcription factors affecting the expression of target genes [3]. The SMc00658 LuxR-type protein contains two functional domains, a DNA-binding domain and a signal binding domain. The DNA-binding domain recognizes and binds to a DNA conserved site called *lux* box. The signal binding domain recognizes a ligand that is usually an acyl homoserine lactone (AHL) [2;4-7]. Unlike common response regulators of QS, SMc00658 is devoid of an AHL LuxI synthase associated with it in the genome of *E. meliloti* and therefore is defined as orphan LuxR regulator [6;8]. LuxR orphans are shown to be responsive to exogenous AHLs produced by neighboring cells as well as to endogenously produced AHLs [9]. It is now also evident that some LuxR orphan proteins have evolved the ability to respond to other molecular signals different from AHLs. Recently, a group of LuxR orphans that lost the capacity to bind AHLs and respond to plant secondary metabolites, as the flavonoids have been discovered in plant associated bacteria (PAB)[10-12]. The QS signaling is not restricted to bacterial cell-to-cell communication, but also allows an interkingdom signaling between microorganisms and their hosts [12].

The results obtained from the experiments reported in the previous chapters pointed out that the luteolin affects the biosynthesis of the AHL signal molecules that are mediators of the *E. meliloti* QS. Therefore, the research activity was focused on the analysis of the *SMc00658* gene encoding a transcriptional regulator of QS in *E. meliloti* [13] to provide insight into its regulatory role. The gene expression profile of the *SMc00658* mutant strain, available at the bacterial strains repository of the LOEWE Center for Synthetic Microbiology, was investigated compared to its wild-type Rm2011 through a transcriptomic approach.

Materials and methods

Strains, media, and growth conditions

The strains of *E. meliloti* used for microarray experiments were the wild-type Rm2011 [14] and its SMc00658 mutant strain, kindly provided by Professor Becker (LOEWE Center for Synthetic Microbiology of the University of Marburg, Germany). *Ensifer meliloti* strains were grown in VMM at 30°C. VMM medium was composed of 14.7mM K_2HPO_4, 11.5 mM KH_2PO_4, 0.46mM $CaCl_2$, 0.037mM $FeCl_3$, 1mM $MgSO_4$, 15.7 mM NH_4Cl, 10 mM Na-succinate, 4.1µM biotin, 48.5 µM H_3BO_3, 10 µM $MnSO_4$, 1 µM $ZnSO_4$, 0.5 µM $CuSO_4$, 0.27 µM $CoCl_2$, and 0.5 µM $NaMoO_4$; pH 7. Four biological replicates of the wild-type strain Rm2011 or SMc00658 mutant strain were grown in 50 ml of liquid medium in 250-ml Erlenmeyer flasks and shaken at 150 rpm until mid-exponential phase was reached. Total RNA isolated from these cultures was used for comparison of gene expression between mutant and wild-type in VMM media.

RNA isolation

Cells from cultures of the *E. meliloti* Rm2011 wild-type and SMc00658 mutant (50 ml at OD_{600} = 0.8) were harvested by centrifugation (10,000 × *g*, 1 min, 20°C) for comparison of gene expression. Cell pellets were immediately frozen in liquid nitrogen and stored at –80°C until RNA extraction. Total RNA was purified using the RNeasy Mini Kit (Qiagen, Hilden, Germany). Cells were resuspended in 10 mM Tris-HCl (pH 8.0) and disrupted in RLT buffer provided with the kit in Fast Protein tubes (Q BIOgene, Carlsbad, CA, U.S.A.) using the Ribolyser (Hybaid, Heidelberg, Germany) (30 s, level 6.5) prior to spin column purification according to the RNeasy mini kit RNA purification protocol. Then RNA was extracted with phenol-chloroform and ethanol precipitated. DNA was removed from RNA preparations by DNase I using Qiagen columns (clean-up procedure). The integrity of all RNA samples was assessed visually using 1% agarose gel electrophoresis and ethidium bromide staining, and then the quantity and purity of the RNA were measured using NanoDrop ND- 1000 UV-VIS Spectrophotometer version 3.2.1. All RNA samples were prepared from four independent biological replicates per strain.

Labeling of cDNA probes, hybridization, and image acquisition

For microarray experiment of the Rm2011 wild-type and SMc00658 mutant comparison, four slide hybridizations were performed in parallel using the labelled cDNA synthesized from four independent RNA preparations obtained from four independent bacterial cultures.
Fluorescent-labeled cDNA with Cy3- and Cy5- probes were prepared according to DeRisi et al.[15] (http://www.microarrays.org/protocols.html) starting from 10 to 30 µg of total RNA. Amino-allyl modified first strand cDNA was synthesized by reverse transcription using random hexamer primers (Qiagen-Operon, Hilden, Germany), SuperScriptII reverse transcriptase (Stratagene, La Jolla, CA) and a dNTP + aa-dUTP mixture (dTTP:aa-dUTP = 1:4) (dNTPs: Peqlab, Erlangen, Germany; aa-dUTP: Sigma-Aldrich, Taufkirchen, Germany). After hydrolysis and cleanup using Microcon-30 filters (Millipore, Eschborn, Germany), the N-hydroxysuccinimidyl ester dyes (Cy3- and Cy5-NHS esters; Amersham Biosciences, Freiburg, Germany) were coupled to the amino-allyl-labeled first-strand cDNA. Uncoupled dye was removed using QiaQuick PCR Purification columns (Qiagen, Hilden, Germany). Pre-hybridization of microarrays was carried out for 45 min at 42°C in Easyhyb hybridization solution (Roche Diagnostics, Mannheim, Germany) supplemented with sonicated salmon sperm DNA at 5 µg/ml. Following pre-hybridization microarrays were washed in MilliQ water (21°C, 1 min), dunked in ethanol (21°C, 10 s), and centrifuged (185 × *g*, 3 min, 20°C). Hybridization of the fluorescent-labeled cDNA was performed at 42°C for 16 h in Easyhyb hybridization solution (Roche Diagnostics) supplemented with sonicated salmon sperm DNA at 50 µg/ml in a final volume of 65 µl under a cover slip. Before applying the fluorescent-labeled cDNA to the microarray, it was denatured for 5 min at 65°C. Microarrays were washed once in 2× SSC, 0.2% SDS (5 min, 42°C), twice in 0.2× SSC, 0.1% SDS (2 min, 21°C), and twice in 0.2× SSC (2 min, 21°C). Following the washes, slides were dried by centrifugation (3min, 185 × *g*, 20°C) and scanned with a pixel size of 10 µm using the ScanArray 4000 microarray scanner (Perkin-Elmer, Boston).

Content and layout of Sm14kOligo microarrays

The Sm14k microarrays, described by Ruberg et al. [16], representing all currently predicted 6,207 protein-coding genes were used for genome-wide gene expression analysis of *Ensifer meliloti* strains. Sm6kPCR microarrays contain 6,046 internal open reading frame (ORF)-specific DNA fragments of 80 to 350 bp, 161 70-mer oligonucleotides as ORF-specific probes and 3 alien DNA fragments (Spot Report Alien PCR product #1, Stratagene 252551; Spot Report Alien PCR product #2, Stratagene 252552; Spot Report Alien PCR product #3, Stratagene 252553; Stratagene, La Jolla, CA, USA) that can serve as probes for spiking

controls. Each probe was spotted in triplicate. DNA fragments were generated by two rounds of PCR amplification. In the first round, ORF-specific primers carrying standard primer sequences at their 5'ends were used. Then, re-amplification using standard primers directed against 5'- Extensions of the ORF-specific primers was carried out to generate PCR fragments sets for the production of microarray. PCR fragments (200–300ng/l) and oligonucleotides (50 M) in 1.5 M betaine, 3× SSC (1× SSC is 0.15 M sodium chloride, 0.015 M sodium citrate) [17] were printed onto 3-aminopropyltrimethoxysilane coated SA-1 glass slides (Asper Biotech, Tartu, Estonia) using the MicroGrid II 600 spotter (BioRobotics, Cambridge, UK) equipped with 48 SMP3 stealth pins (TeleChem International, Sunnyvale, CA, USA). DNA was cross-linked to the surface by incubation of the slides for 3 h at 80 °C. Unbound DNA was removed by two washes in 0.1% (w/v) SDS for 2 min at 42 °C and two washes in distilled water for 2 min at 20 °C. After denaturation at 100 °C for 3 min in distilled water, slides were dunked into ethanol and dried by centrifugation (3 min, 185 × g, 20 °C).

Data analysis

Mean signal and mean local background intensities were obtained for each spot of the microarray images using the ImaGene 8.0 software for spot detection, image segmentation, and signal quantification (Biodiscovery Inc., Los Angeles). Spots were flagged as "empty" if $R \leq 1.5$ in both channels, where R = (signal mean – background mean)/background standard deviation. The remaining spots were considered for further analysis. The $\log_2$ value of the ratio of intensities (fold change in gene expression) was calculated for each spot using the following formula:

$$M\text{i} = \log_2(R\text{i}/G\text{i})$$

where Ri = Ich1(i) –Bgch1(i)i and Gi = Ich2(i) – Bgch2(i), Ich1(i) or Ich2(i) is the intensity of a spot in channel 1 or channel 2 and Bgch1(i)i or Bgch2(i) i is the background intensity of a spot in channel 1 or channel 2, respectively. The mean intensity was calculated for each spot, Ai = $\log_2(R\text{i}G\text{i})$*0.5 (Dudoit et al. 2002). A normalization method based on local regression that account for intensity and spatial dependence in dye biases was applied. Within a print tip group, normalization was performed according to Yang et al. [18]:

$$M\text{i} = \log_2(R\text{i}/G\text{i}) \rightarrow \log_2(R\text{i}/G\text{i}) - \text{cj}(A) = \log_2(R\text{i}/[k\text{j}(A)G\text{i}])$$

where cj(A) is the lowest fit to the MA plot for the jth grid only (i.e., for the jth print tip group), j = 1, …, J, and J denotes the number of print tips. A floor value of 20 was introduced before normalization. Genes significantly up or down regulated were identified by t statistics [19].

Genes were regarded as differentially expressed if $P \leq 0.05$ and $M \geq 1.00$ or $M \leq -1.00$. LOWESS Normalization and *t* statistics were carried out using the *EMMA 2.8.2* microarray data analysis software developed at the Bioinformatics Resource Facility, Center for Biotechnology, Bielefeld University) (http://www.genetik.uni-bielefeld.de/ EMMA/) [20].

Results

Transcriptome analysis of the E. meliloti strains

The expression profile of the *E. meliloti* SMc00658 mutant strain lacking the transcriptional regulator SMc00658 was analyzed compared to that of its wild-type Rm2011 using a transcriptomic approach to identify the spectrum of differentially expressed genes and gain more information about the functions controlled by the SMc00658 factor.

Analyses of the microarray data identified a total number of 86 significantly differentially expressed genes more than 2-fold (M value≥ 1.0 and $P \leq 0.05$) in the SMc00658 mutant respect to the wild-type, 60 of which up-regulated and 26 down-regulated (Table 5.1). The evidence that 60 genes were found up-regulated rather than down-regulated in the mutant might suggest that the SMc00658 factor primarily functions as a negative regulator of transcription in *E. meliloti*. The majority of differentially expressed genes (61/86) belonged to the chromosome and therefore mainly related to processes of the bacterial central metabolism [21]. Conversely, only 7 genes were located on the pSymA plasmid and 18 genes on the pSymB plasmid of *E. meliloti* (Table 5.1). pSymB was reported to encode exopolysaccharides biosynthesis and many ABC transporters [22] whereas most of the genes required for nodulation and nitrogen-fixation (*nod*, *nif*, and *fix* genes) were present on pSymA [21].

Table 5.1. Differentially expressed genes number in transcriptome profiling experiments of the *E. meliloti* SMc00658 mutant in comparison with the wild-type Rm2011 and relative number of annotated genes in COGs categories

	Differentially Expressed Genes (DEGs)			
	Total	Up-regulated	Down-regulated	Annotated in COGs Categories
SMc00658 mutant	86	60	26	59
Chromosome	-	45	16	-
pSymA	-	4	3	-
pSymB	-	11	7	-

According to the COG functional classification, the main classes of genes differentially expressed were represented by translation, cell wall/membrane biogenesis and inorganic ions transport and metabolism (Figure 5.1).

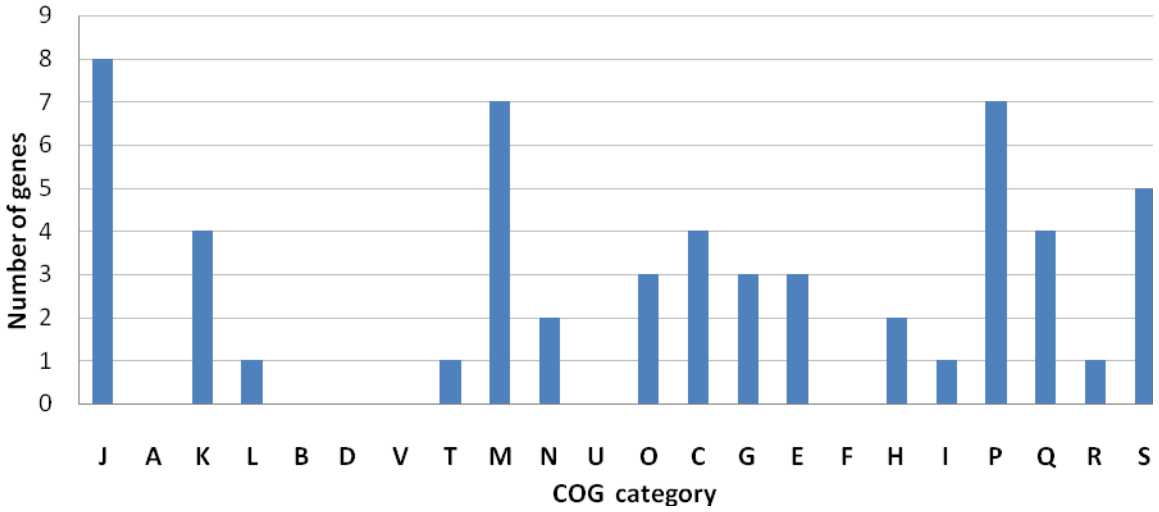


Figure 5.1. Number of differentially expressed genes in the SMc00658 mutant strain compared to the wild-type Rm2011. Genes are grouped by functional classification according to the COG annotation *of the E. meliloti* Rm2011 genome. *Code J*, Translation, *code A*, RNA processing and modification; *code K*, Transcription, *code L*, Replication, recombination and repair, *code B*, Chromatin structure and dynamics, *code D*, Cell cycle control, cell division, chromosome partitioning, *code V*, Defense mechanisms, *code T*, Signal transduction mechanisms, *code M*, Cell wall/membrane biogenesis, *code N*, Cell motility, *code U*, Intracellular trafficking and secretion , *code O*, Posttranslational modification, protein turnover, chaperones; *code C*, Energy production and conversion, *code G*, Carbohydrate transport and metabolism, *code E*, Amino acid transport and metabolism, *code F*, Nucleotide transport and metabolism; *code H*, Coenzyme transport and metabolism; *code I*, Lipid transport and metabolism, *code P*, Inorganic ion transport and metabolism; *code Q*, Secondary metabolites biosynthesis, transport and catabolism, *code R*, General function prediction only, respiration; *code S*, Function unknown.

The analysis of the entire transcriptome data has revealed that the major processes modulated by the SMc00658 regulator are associated with the transcription/translation, iron uptake, metabolism and uptake of phosphorous compounds, flagellar assembly and motility as well as oxidative stress response. The major sigma factor of *E. meliloti*, RpoE1, was found to be transcriptionally up-regulated in the SMc00658 mutant (Table 5.2). RpoE1 was reported to be active upon entry in stationary phase, and it mainly controlled the expression of chaperones and proteases involved in folding and degradation of cytoplasmic and secreted proteins, respectively [23]. The up-regulation of the transcriptional sigma factor supports the hypothesis of a manly negative regulatory role of the *SMc00658* gene product. Accordingly to the transcription up-regulation, the translation machinery was largely increased in the SMc00658 mutant, since ribosomal protein-encoding genes (n=8), as well as the *infC* gene required for translation initiation were found to be up-regulated (Table 5.2).

Table 5.2. Transcriptome profile of the *E. meliloti* SMc00658 mutant in comparison with the wild-type Rm2011

Genes up-regulated more than 2-fold (M value≥ 1.0) in the SMc00658 mutant

Gene	Gene product description	M value *
SMa0011	SelA selenocysteine synthase	1.47**
SMa0745	groES2 chaperonin	1.05
SMa1081	Hypothetical protein	1.08
SMa2111	Hypothetical protein	2.02
SMb20066	Hypothetical protein	1.10
SMb20089	Hypothetical protein	1.04
SMb20910	Hypothetical protein	1.26
SMb21483	Hypothetical protein	1.16
SMb21681	Hypothetical protein	3.26
SMc00198	Hypothetical protein	1.78
SMb20207	*pqqD*- pyrroloquinoline quinone synthesis protein	1.12
SMb20230	smc22-r protein	1.31
SMb20894	*gguA*- sugar uptake ABC transporter	0.98
SMb20932	*exsH* - endo-beta-glycanase, C-terminal secretion signal protein	1.20
SMb21069	Hypothetical protein	1.20
SMb21314	*expE* - secreted calcium-binding protein	1.22
SMc00286	Hemolysin-type calcium binding protein	0.98
SMc00362	*infC* - Translation initiation factor IF-3 protein	1.60
SMc00432	*iolB* Myo-inositol catabolism protein	1.10
SMc00604	*ropB* Outer membrane protein	0.99
SMc00722	Hypothetical protein	1.14
SMc00732	Hypothetical protein	1.30
SMc00809	Hypothetical signal peptide protein	1.68
SMc00885	Hypothetical transmembrane signal peptide protein	2.34
SMc00912	co-chaperonine *groES*	1.30
SMc01111	Apolipoprotein N-acyltransferase (acid-inducible gene) transmembrane	1.22
SMc01173	Hypothetical protein	1.04
SMc01182	Hypothetical protein	1.85
SMc01302	*rplP* 50S Ribosomal protein L16	1.27
SMc01303	*rpsC* 30S Ribosomal protein S3	1.20
SMc01310	*rpsJ* Ribosomal protein S10	1.22
SMc01314	*rpsL* 30S Ribosomal protein S12	1.57
SMc01418	Hypothetical signal peptide protein	3.85
SMc01419	*rpoE1* RNA polymerase sigma factor	2.43
SMc01557	Hypothetical signal peptide protein	1.24
SMc01804	r*plM* 50S Ribosomal protein L13	1.36
SMc02051	Hypothetical protein	1.06
SMb21174	*phoT* phosphate uptake ABC transporter permease	1.01
SMb21177	*phoC* phosphate uptake ABC transporter	1.39
SMc02144	*pstC* Phosphate transport system permease	1.16
SMc02145	Hypothetical signal peptide protein	1.07
SMc02146	Phosphate- binding periplsmatic protein	2.21
SMc02156	Hypothetical protein	4.17
SMc02389	Hypothetical protein	1.00
SMc02405	*smpB* SsrA-binding protein	1.01
SMc02407	Hypothetical protein	1.05

SMc02755	*ahcY* Adenosylhomocysteinase hydrolase	1.07
SMc02868	Multidrug efflux system	1.03
SMc02898	*kdsB* 3-deoxy-manno-octulosonate cytidylyltransferase protein	1.34
SMc03037	*flaA* Flagellin A	0.98
SMc03050	*flaF* Flagellin synthesis regulator protein	1.25
SMc03100	Hypothetical protein	1.03
SMc03124	Periplasmatic binding ABC transporter	1.76
SMc03208	hmgA Homogentisate 1,2-dioxygenase	1.58
SMc03770	*rplU* 50S Ribosomal protein L21	1.29
SMc03780	Hypothetical protein	1.34
SMc04239	Hypothetical protein	1.16
SMc04291	L-sorbosone dehydrogenase (SNDH)	1.05
SMc04316	*afuA* iron transport ABC permease	1.34

* Average M value (=$\log_2$FC) calculated from microarray analysis were reported

** Based on the M value, significantly differentially expressed genes ($P \leq 0.05$) were divided as induced ($M \geq 1$) or repressed ($M \leq 1$) in the SMc00658 mutant

The induction of the chaperone *groES (SMa0745)* and co-chaperonine *groES (SMc00912)* genes observed in the SM0c0658 mutant, may also correlate with an increased protein synthesis. The products of the *SMa0745* and *SMc00912* genes are known as "heat shock proteins" since involved in the bacterial response to stress such as a heat shock and are essential genes for the bacterial growth [24]. The *SMa0745* and *SMc00912* genes, which encode for molecular chaperons required to help refolding of denatured proteins, are also involved in protein secretion and in facilitating the correct folding of *de novo* synthesized proteins both during and after translation [25]. Two fully functional copies of *groES* chaperone are reported for *E. meliloti*: the chromosomal copy is required for activation of nodulation (*nod)* genes through NodD regulator and are induced by the plant flavonoid luteolin [26]; the second copy borne on the plasmid pSymA appears sufficient for all other growth functions [27]. The *afuA* gene (*SMc04316*) encoded an ABC-type transport system involved in the iron uptake resulted induced in the SMc00658 mutant (Table 5.2). The uptake of iron is a crucial aspect of rhizobial metabolism because enzymes related to nitrogen-fixation such as nitrogenase and leghemoglobin contain iron as cofactor [28]. Genes related to the uptake and metabolism of phosphorous compounds were induced in the SMc00658 mutant (Table 5.2). The *phoC* and *phoT* genes were known to be member of *Pho* regulon of *E. meliloti*. A similar pattern of up-regulation was observed for *pstC*, *SMc02145*, *SMc02146* and *pstS* genes, members of the *pstSCAB* operon that may encode an ABC-type phosphate uptake system. In *E. coli*, the *pstSCAB* operon was involved in the inhibition of the sensor kinase PhoR, which was acting as activator of the transcriptional regulator PhoB in the presence of high external phosphate concentrations [29]. A similar regulatory mechanism may operate in *E. meliloti*, although this assumption remains to be demonstrated experimentally.

E. meliloti SMc00658 mutant revealed some differentially expressed genes that were implicated in oxidative stress and adaptation. The *katA* gene, encoding a monofunctional catalase enzyme and the oxidoreductase transmembrane protein (*SMc00374*) were repressed (Table 5.3) [30]. The KatA protein of *E. meliloti* was responsible for detoxifying "Reactive Oxygen Species" (ROS), thereby limiting their concentrations within the bacterial cell to cope the subsequence stress [31]. As with many host-microbe interactions, the rhizobium-legume symbiosis can be associated with a host-generated release of ROS (O_2^-, H_2O_2 and HO^-) [32] that play a role in limiting rhizobial invasion. The *katA* gene was found induced in bacteroids and proposed to have a role in responding to oxidative stress [33]. Thus the repressed pattern found for *katA* in the mutant suggested that SMc00658 regulator might play a role in *E. meliloti* adaptation and protection to oxidative stress. Seven genes most likely involved in cell wall/membrane biosynthesis were differentially expressed. Genes associated to the biosynthesis of lipopolysaccharide component (LPS) were found among them. *SMc01790* and *SMc02640* genes encoding a glycosyltransferase (*lpsB*) and an UDP-glucuronic acid epimerase (*lpsL*), respectively, and participating in the biosynthesis of the LPS core were observed induced in the *SMc00658* mutant (Table 5.2). The surface polysaccharide LPS plays a role in promoting the rhizobial adaptation and persistance within rhizosphere as well as in promoting the invasion of plant roots [34]. *E. meliloti lpsB* mutant, due to a dramatically altered LPS core, failed to effectively colonized host plant roots and was proposed to contributed to the structural integrity of LPS [35]. Moreover, defects in LPS can sensitize rhizobial cells to membrane-disrupting agents and antimicrobial peptides. Thus LPS layer may provide a protective barrier against environmental stress and host defense responses. There are indications that rhizobial LPS may also play an active role by suppressing the release of ROS [36].

Table 5.3. Transcriptome profile of the *E. meliloti* SMc00658 mutant in comparison with the wild-type Rm2011

Genes down-regulated more than 2-fold (M value ≤ 1.0) in the SMc00658 mutant

Gene	Gene product description	M value *
SMa0121	Hypothetical protein	-2.42 **
SMa0725	Hypothetical protein	-1.15
SMb21584	Hypothetical protein	-1.54
SMc00022	Hypothetical protein	-1.24
SMc02689	Aldehyde dehydrogenase	-2.32
SMb20020	*pdh* -pyruvate dehydrogenase	-1.13
SMb20204	*pqqA* - pyrroloquinoline quinone synthesis protein	-2.07
SMb20209	Hypothetical protein	-1.25
SMb21130	Sulfate uptake ABC transporter	-1.65
SMb21491	Hypothetical exported protein	-1.12
SMc00229	Hypothetical transmembrane protein	-1.08
SMc00349	lepA- GTP-binding protein membrane	-1.31
SMb20268	Proline racemase	-1.56

SMc00374	Oxidoreductase transmembrane protein	-1.03
SMc00714	1-acyl-sn-glycerol-3-phosphateacyltransferase (PLSC)	-1.16
SMc00819	*katA* - Catalase hydroperoxidase HPII(III) protein	-1.120
SMc01227	*nerA* - Glycerol trinitrate (GTN) reductase	-1.17
SMc01666	*mdeA*-Methionine gamma-lyase	-1.24
SMc01790	Glycosyltransferase	-1.23
SMc02640	*lpsL*- UDP-Glucuronic acid epimerase	-1.25
SMc01369	*rpmG* - 50S Ribosomal protein L33	-1.16
SMc01609	*ribH2* riboflavin synthase s	-1.47
SMc03948	TRm1b transposase for insertion sequence element	-1.02
SMc00794	Two component response regulator (TCS)	-1.48
SMc01403	Transcriptional regulator protein	-1.24
SMa0850	SyrM transcriptional regulator	-1.16

* Average M value (=$\log_2$FC) calculated from microarray analysis were reported

** Based on the M value, significantly differentially expressed genes (P ≤ 0.05) were divided as induced (M ≥ 1) or repressed (M ≤ 1) in the SMc00658 mutant

Some genes involved in energy and central metabolism were differentially regulated. Genes encoding pyruvate dehydrogenase enzymes taking part in the pathways required to generate acetyl-coA (*pdhAbB*, *SMc02689* aldehyde dhydrogenase, *pqqA*) were repressed in the absence of SMc00658 regulator under free-living conditions (Table 5.3). Conversely, Cabanes and coworkers [37] provided evidence that *pdh* genes expression of *E. meliloti* was induced during symbiosis, compared with free-living conditions. The *iolB* gene encoding a protein involved in the catabolism of myo-inositol and a gene for L-sorbosone dehydrogenase (SNDH) were found induced suggesting that these compounds may be used as a carbon source by *E. meliloti*. Flagellar genes involved in the control of motility were up-regulated in the *SMc00658* mutant. These genes included the structural flagellin gene *flaA* (*SMc03037*) for assembling a functional flagellar filament and the transcriptional regulator *flaF* (*SMc03050)* of the flagellin biosynthesis. Mutational analysis in *R. lupini* and *E. meliloti* revealed that flagellin A is the essential subunit required for assembling a functional flagellar filament [38]. The motile ability of rhizobia facilitates survival and optimized resource utilization in hostile environments. Furthermore, motility allows the rhizobial cells to find their specific host legume and has been shown to be an important phenotypic trait related to the bacterial survival as well as to the symbiosis [39]. Such movement ability, mediated by cell-surface organelles, such as flagella, may promote the accumulation of rhizobial cells around the host roots, as highly localized bacterial “clouds” and simultaneously enhance the successful invasion of the roots. The main genes involved in nodulation (*nod* operon) and in nitrogen fixation process (*fix* and *nif* operons) were not found to be differentially expressed in the mutant compared to the wild-type. However, the SyrM transcriptional regulator was down-regulated in the *SMc00658* mutant. SyrM has been related to the nodulation and constituted a self-amplifying circuit with the regulator NodD3 for *nod* genes activation [40].

Two other genes encoding regulatory proteins were down-regulated in the *SMc00658* mutant. *SMc1403* codifying a transcriptional regulator and *SMc00794* codifying the RsiB2 response regulator of a two-component regulatory system [41]. *SMc1403* showed homology with proteins belonging to the Lrp/AsnC family of global or specific transcriptional regulator [42]. Members of this family influence cellular metabolism and are widely distributed in numerous prokaryotes, including bacteria and *Archaea.* The best-characterized member of the Lrp/AsnC family was *E. coli* Lrp [43]. The genes found to be regulated by members of Lrp/AsnC family are involved in transport, degradation and biosynthesis of amino acids, as well as a small number of proteins involved in the production of pilum, porins, sugar transporters and nucleotide hydrogenases. RsiB2 has been reported in *E. meliloti* to positively regulate the RpoE2-dependent response, blocking the anti-sigma factors [44].
About the 20% of the total number of induced genes and repressed genes in the *SMc00658* mutant were functionally classified as hypothetical proteins with unassigned cellular functions (Table 5.2), thus remaining an open question to be explored.

Concluding remarks and future works

The transcriptional profiling of the *E. meliloti* SMc00658 mutant compared to the wild-type Rm2011 allowed to elucidate the repertoire of genes differentially expressed and therefore modulated by the LuxR-like transcriptional regulator SMc00658. The majority of the differentially regulated genes (70%) were up-regulated in the mutant respect to the wild-type, indicating that the regulatory role of SMc00658 has a mainly negative effect on the expression of the target genes. Accordingly, the inactivation of the SMc00658 regulator resulted in the up-regulation of the major transcriptional σ-factor RpoE1 in *E. meliloti*, in the up-regulation of the translation machinery and of the protein-folding chaperone GroEL. Moreover, the SMc00658 regulator turned out to control genes involved in symbiosis-related process such as iron uptake, motility, LPS biosynthesis and also genes for adaptation and to cope with oxidative stress. The portion of genes with unassigned cellular functions among induced or repressed genes revealed that other genetic pathways under the control of the SMc00658 regulator remain to be explored. All the identified genes as well as genes with still unknown function are good candidates for further molecular and phenotypic analysis of the SMc00658 regulon in *E. meliloti*. Future analysis would involve the quantitative assessment of the expression of promising identified genes by real-time PCR. The SMc00658 mutant would be subject of an extensive phenotypic characterization respect to

the wild-type using the Phenotype Microarray (PM) high throughput technology. Moreover, motility assays, *in vitro* nodulation experiments, assays under several stress and nutrients-limiting conditions would be performed to further investigate these phenotypic traits. All the planned experiments would be carried out in the absence as well as in presence of the plant signal luteolin to evaluate whether the SMc00658 regulator of *E. meliloti* is luteolin responsive.

References

1. I. Llamas, N. Keshavan, and J.E. Gonzalez, Use of *Sinorhizobium meliloti* as an indicator for specific detection of long-chain N-acyl homoserine lactones (2004), Applied Environmental Microbiology 70 3715-3723.

2. M. Sanchez-Contreras, W.D. Bauer, M.S. Gao, J.B. Robinson, and J.A. Downie, Quorum sensing regulation in rhizobia and its role in symbiotic interactions with legumes (2007), Philosophical Transactions of the Royal Society B-Biological Sciences 362 1149-1163.

3. C. Fuqua, M.R. Parsek, and E.P. Greenberg, Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing (2001), Annual Review of Genetics 35 439-468.

4. J.E. Gonzalez and M.M. Marketon, Quorum sensing in nitrogen-fixing rhizobia (2003), Microbiology and Molecular Biology Reviews 67 574.

5. L. Steindler and V. Venturi, Detection of quorum-sensing N-acyl homoserine lactone signal molecules by bacterial biosensors (2007), FEMS Microbiology Letters 266 1-9.

6. S. Subramoni and V. Venturi, LuxR-family 'solos': bachelor sensors/regulators of signaling molecules (2009), Microbiology-Sgm 155 1377-1385.

7. J.H. Devine, G.S. Shadel, and T.O. Baldwin, Identification of the operator of the *Lux* regulon from the *Vibrio fischeri* strain ATCC7744 (1989), Proceedings of the National Academy of Sciences of the United States of America 86 5688-5692.

8. A.V. Patankar and J.E. Gonzalez, An orphan LuxR homolog of *Sinorhizobium meliloti* affects stress adaptation and competition for nodulation (2009), Applied and Environmental Microbiology 75 946-955.

9. B.M.M. hmer, Cell-to-cell signaling in *Escherichia coli* and *Salmonella enterica* (2004), Molecular Microbiology 52 933-945.

10. V. Venturi and C. Fuqua, Chemical signaling between plants and plant-pathogenic bacteria (2013), Annual Review of Phytopathology, 51 17-37.

11. J.F. Gonzalez and V. Venturi, A novel widespread interkingdom signaling circuit (2013), Trends in Plant Science 18 167-174.

12. D.T. Hughes and V. Sperandio, Inter-kingdom signaling: communication between bacteria and their hosts (2008), Nature Reviews Microbiology 6 111-120.

13. J.E. Gonzalez and M.M. Marketon, Quorum sensing in nitrogen-fixing rhizobia (2003), Microbiology and Molecular Biology Reviews 67 574.

14. H.M. Meade and E.R. Signer, Genetic mapping of *Rhizobium meliloti* (1977), Proc. Natl. Acad. Sci. U S. A 74 2076-2078.

15. J.L. Derisi, V.R. Iyer, and P.O. Brown, Exploring the metabolic and genetic control of gene expression on a genomic scale (1997), Science 278 680-686.

16. S. Ruberg, Z.X. Tian, E. Krol, B. Linke, F. Meyer, Y.P. Wang, A. Puhler, S. Weidner, and A. Becker, Construction and validation of a *Sinorhizobium meliloti* whole genome DNA microarray: genome-wide profiling of osmoadaptive gene expression (2003), Journal of Biotechnology 106 255-268.

17. F. Diehl, S. Grahlmann, M. Beier, and J.D. Hoheisel, Manufacturing DNA microarrays of high spot homogeneity and reduced background signal (2001), Nucleic Acids Research 29

18. Y.H. Yang, S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai, and T.P. Speed, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation (2002), Nucleic Acids Research 30

19. S. Dudoit, Y.H. Yang, M.J. Callow, and T.P. Speed, Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments (2002), Statistica Sinica 12 111-139.

20. M. Dondrup, A. Goesmann, D. Bartels, J. Kalinowski et al., EMMA: a platform for consistent storage and efficient analysis of microarray data (2003), Journal of Biotechnology 106 135-146.

21. D. Capela, F. Barloy-Hubler, J. Gouzy, G. Bothe, F. Ampe, J. Batut, et al., Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021 (2001), Proceedings of the National Academy of Sciences of the United States of America 98 9877-9882.

22. T.M. Finan, S. Weidner, K. Wong, J. Buhrmester, P. Chain, F.J. Vorholter et al., The complete sequence of the 1,683-kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti* (2001), Proceedings of the National Academy of Sciences of the United States of America 98 9889-9894.

23. M.A. Wagner, D. Zahrl, G. Rieser, and G. Koraimann, Growth phase- and cell division-dependent activation and inactivation of the sigma(32) regulon in Escherichia *coli* (2009), Journal of Bacteriology 191 1695-1702.

24. C. Georgopoulos, The emergence of the chaperone machines (1992), Trends in Biochemical Sciences 17 295-299.

25. P.A. Lund, Microbial molecular chaperones (2001), Advances in Microbial Physiology, Vol 44 93-140.

26. D. Capela, S. Carrere, and J. Batut, Transcriptome-based identification of the *Sinorhizobium meliloti* NodD1 regulon (2005), Applied Environmental Microbiology 71 4910-4913.

27. J. Ogawa and S.R. Long, The *Rhizobium meliloti* GroEL Locus is required for regulation of early nod genes by the transcription activator NodD (1995), Genes & Development 9 714-729.

28. D. Lynch, J. O'Brien, T. Welch, P. Clarke, P.O. Cuiv, J.H. Crosa, and M. O'Connell, Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti* (2001), Journal of Bacteriology 183 2576-2585.

29. Wanner BL, Phosphorus assimilation and control of the phosphate regulon. In: *Escherichia coli* and *Salmonella enterica:* cellular and molecular biology (1996), pp. 1357-1381.

30. D. Herouart, S. Sigaud, S. Moreau, P Frendo, D. Touati, and A. Puppo, Cloning and characterization of the *katA* gene of *Rhizobium meliloti* encoding a hydrogen peroxide-inducible catalase (1996), Journal of Bacteriology 178 6802-6809.

31. S. Ardissone, P. Frendo, E. Laurenti, W. Jantschko, C. Obinger, A. Puppo, and R.P.Ferrari, Purification and physical-chemical characterization of the three hydroperoxidases from the symbiotic bacterium *Sinorhizobium meliloti* (2004), Biochemistry 43 (2004) 12692-12699.

32. C. Nathan and M.U. Shiloh, Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens (2000), Proceedings of the National Academy of Sciences of the United States of America 97 8841-8848.

33. M. Djordjevic, A global analysis of protein expression profiles in *Sinorhizobium meliloti*: discovery of new genes for nodule occupancy and stress adaptation (2003), Molecular Plant Microbe Interactions.

34. A. Becker, N. Fraysse, and L. Sharypova, Recent advances in studies on structure and symbiosis-related function of rhizobial K-antigens and lipopolysaccharides (2005), Molecular Plant-Microbe Interactions 18 899-905.

35. G.R.O. Campbell, B.L. Reuhs, and G.C. Walker, Chronic intracellular infection of alfalfa nodules by *Sinorhizobium meliloti* requires correct lipopolysaccharide core (2002), Proceedings of the National Academy of Sciences of the United States of America 99 3938-3943.

36. H. Scheidle, A. Gross, and K. Niehaus, The Lipid A substructure of the *Sinorhizobium meliloti* lipopolysaccharides is sufficient to suppress the oxidative burst in host plants (2005), New Phytologist 165 559-565.

37. D. Cabanes, P. Boistard, and J. Batut, Symbiotic induction of pyruvate dehydrogenase genes from *Sinorhizobium meliloti* (2000), Molecular Plant-Microbe Interactions 13 483-493.

38. B. Scharf, H. Schuster-Wolff-Buhring, R. Rachel, and R. Schmitt, Mutational analysis of the *Rhizobium lupini* H13-3 and *Sinorhizobium meliloti* flagellin genes: Importance

of flagellin A for flagellar filament structure and transcriptional regulation (2001), Journal of Bacteriology 183 5334-5342.

39. F. Nievas, P. Bogino, F. Sorroche, and W. Giordano, Detection, characterization, and biological effect of quorum sensing signaling molecules in peanut-nodulating *Bradyrhizobia* (2012), Sensors 12 2851-2873.

40. M.J. Barnett and S.R. Long, The *Sinorhizobium meliloti* SyrM Regulon: effects on global gene expression are mediated by SyrA and NodD3 (2015), Journal of Bacteriology 197 1792-1806.

41. B. Gourion, A. Francez-Charlot, and J.A. Vorholt, PhyR is involved in the general stress response of *Methylobacterium extorquens* AM1 (2008), Journal of Bacteriology 190 1027-1035.

42. A. Napoli, J. Van der Oost, C.W. Sensen, R.L. Charlebois, M. Rossi, and M. Ciaramella, An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter (1999), Journal of Bacteriology 181 1474-1480.

43. T.H. Tani, A. Khodursky, R.M. Blumenthal, P.O. Brown, and R.G.Matthews, Adaptation to famine: A family of stationary-phase genes revealed by microarray analysis (2002), Proceedings of the National Academy of Sciences of the United States of America 99 13471-13476.

44. J.M.Calvo and R.G.Matthews, The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli* (1995), Microbiological Reviews 59 323.

Chapter 6

Concluding remarks

The study of the nitrogen-fixing symbiosis between *Ensifer meliloti* and leguminous plants is an important contribution of the microbiology to agricultural applications aimed to improve the yield of legume crops and to find innovative approaches to increase their environmental sustainability. Therefore, an extensive investigation of the responses induced by the plant flavonoid luteolin in the symbiont *E. meliloti* was provide in the present work.

The high-throughput Phenotype MicroArray (PM) approach had a key role in the characterization of the metabolic and chemical sensitivity responses mediated by luteolin in *E. meliloti*. For the first time, the analysis revealed that the plant signal luteolin makes a significance change in the sensitivity and osmotolerance profile of *E. meliloti*. The major effect of luteolin was an enhanced resistance phenotype to osmolytes and to a broad set of antimicrobials and toxic compounds. The luteolin-mediated resistance, firstly reported in this work, indicates that the compound exerts an additional physiologically important role concerning oxidative and abiotic stress conditions. Furthermore, luteolin allows *E. meliloti* coping with antimicrobial and toxic compounds in the rhizosphere. All the above mentioned luteolin effects confer an advantage for *E. meliloti* to establish an effective symbiosis in the selective rhizospheric environment.

The extensive characterization of the *E. meliloti* strain defective in the NodD regulator, which is reported to be the luteolin-sensor, elucidated that the resistance to osmolytes is a luteolin dependent phenotype but it not luteolin-NodD dependent. Therefore, luteolin stimulates osmotolerance in *E. meliloti* through the activation of a regulation system that does not involve the NodD regulator. The resistance phenotypes related to the antimicrobial and toxic compounds induced by luteolin turned out to be dependent and independent from the NodD regulatory circuit. Consequently, the NodD factor is involved not only in the activation of nodulation genes but also in the regulation of systems that contribute to mediate resistance. Moreover, the flavonoid luteolin has other regulatory targets, beyond NodD, which confer to *E. meliloti* the ability to tolerate toxic compounds and osmolytes. Results from this study provide evidences that the EmrAB efflux pump acts as further luteolin-mediator to the chemical resistance not mediated by NodD. The reduced nodulation efficiency displayed by the *emrB* mutant revealed also the involvement of the efflux-mediated resistance in establishing a successful symbiosis interaction. The phenotypic assays, performed to evaluate the luteolin effect on a set of symbiosis- related phenotypes, point out that the luteolin exerts a growth-promoting effect on low cellular densities of *E. meliloti.* A significant reduction of the QS signals production and lower levels of cellular motility were observed in presence of luteolin, revealing that the plant signal promotes the accumulation of a higher rhizobial cell

number in the rhizosphere of the host plant compared to the bulk soil. Such mechanism increases the competitiveness of *E. meliloti* during roots colonization. The host plant can use the flavonoid luteolin to stimulates siderophores production by rhizobial cells and then the bacterial uptake of iron, which is crucial cofactor for the enzymes related to nitrogen-fixation.

Overall, obtained data show that the plant signal luteolin triggers a pleiotropic response in *E. meliloti* strongly related to the nutritional conditions to which the bacterium is exposed, extending the current scenario. Luteolin turned out to affect a broad spectrum of *E. meliloti* phenotypes and to control several aspect of bacterial physiology, unknown until now.

The acquired information about the phenotypic traits enhanced and modulated by the luteolin may be exploited to improve the fitness of the *E. meliloti* strains in the rhizosphere as well as the efficiency and the competitiveness for the plant host nodulation

Chapter 7

Publications resulting from collaborations during the PhD period

Standards in Genomic Sciences (2013) 9:325-333 DOI:10.4056/sigs.3797438

Permanent draft genome sequences of the symbiotic nitrogen fixing *Ensifer meliloti* strains BO21CC and AK58

Marco Galardini[1], Marco Bazzicalupo[1], Emanuele Biondi[2], Eveline Brambilla[3], Matteo Brilli[4], David Bruce[5], Patrick Chain[5], Amy Chen[6], Hajnalka Daligault[6], Karen Walston Davenport[5], Shweta Deshpande[6], John C. Detter[5], Lynne A. Goodwin[5], Cliff Han[5], James Han[6], Marcel Huntemann[6], Natalia Ivanova[6], Hans-Peter Klenk[3], Nikos C. Kyrpides[6], Victor Markowitz[6], Kostas Mavrommatis[6], Stefano Mocali[7], Matt Nolan[6], Ioanna Pagani[6], Amrita Pati[6], Francesco Pini[2], Sam Pitluck[6], Giulia Spini[1], Ernest Szeto[6], Hazuki Teshima[5], Tanja Woyke[6], Alessio Mengoni[1,*]

[1] Department of Biology, University of Firenze, via Madonna del Piano 6, I-50019, Sesto Fiorentino, Italy
[2] Interdisciplinary Research Institute - CNRS, Villenenuve d'Ascq, France
[3] Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[4] Edmund Mach Foundation, San Michele all'Adige, Italy
[5] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
[6] DOE Joint Genome Institute, Walnut Creek, California, USA
[7] Consiglio per la Ricerca e la Sperimentazione in Agricoltura - Centro di Ricerca per l'Agropedologia e la Pedologia, Firenze, Italy

*Corresponding author: Alessio Mengoni (alessio.mengoni@unifi.it)

Keywords: Aerobic, motile, Gram-negative, mesophilic, chemoorganotrophic, chemoautotrophic, soil, plant symbiont, biological nitrogen fixation, *Ensifer* (*Sinorhizobium*) *meliloti*, legume yield

Ensifer (syn. *Sinorhizobium*) *meliloti* is an important symbiotic bacterial species that fixes nitrogen. Strains BO21CC and AK58 were previously investigated for their substrate utilization and their plant-growth promoting abilities showing interesting features. Here, we describe the complete genome sequence and annotation of these strains. BO21CC and AK58 genomes are 6,985,065 and 6,974,333 bp long with 6,746 and 6,992 genes predicted, respectively.

Introduction

Strains AK58 and BO21CC belong to the species *Ensifer* (syn. *Sinorhizobium*) *meliloti* (*Alphaproteobacteria*, *Rhizobiales*, *Rhizobiaceae*, *Sinorhizobium*/*Ensifer* group) [1,2], an important symbiotic nitrogen fixing bacterial species that associates with roots of leguminous plants of several genera, mainly from *Melilotus*, *Medicago* and *Trigonella* [3]. These strains have been originally isolated from *Medicago* spp. during a long course experiment (BO21CC) and from plants collected in the north Aral sea region (Kazakhstan) (AK58). Previous analyses conducted by comparative genomic hybridization (CGH), nodulation tests and Phenotype Microarray™(Biolog, Inc.) showed that AK58 (= DSM 23808) and BO21CC (= DSM 23809) are highly diverse in both genomic and phenotypic properties. In particular, they show different symbiotic phenotypes with respect to the crop legume *Medicago sativa* L [4,5]. In a previous collaboration with DOE-JGI, the genomes of strains AK83 (= DSM 23913) and BL225C (= DSM 23914) were also sequenced, allowing the identification of putative genetic determinants for their different symbiotic phenotypes [6]. Consequently, interest in strains AK58 and BO21CC arose, sincegenomic analysis of these strains would foster a greater understanding of the *E. meliloti* pangenome [7], and facilitate deeper investigation of the genomic determinants responsible for differences in symbiotic performances between *E. meliloti* strains found in nature. These research goals may lead to improved strain selection and better inoculants of the legume crop *M. sativa*.

Classification and features

Representative genomic 16S rRNA sequences of strains AK58 and BO21CC were compared with those present in the Ribosomal Database by using Match Sequence module of Ribosomal Database Project [8]. Representative genomic 16S rRNA sequences of closer phylogenetic relatives of the genus *Ensifer*/*Sinorhizobium* and of *Rhizobiales* family (as outgroup) were then selected from IMG-ER database [Table 1], [16]. All strains from the genus *Ensifer*/*Sinorhizobium* form a close cluster, including strains AK58 and BO21CC, thus confirming the affiliation of these two strains within the species. Figure 1 shows the phylogenetic neighborhood of *E. meliloti* AK58 and BO21CC in a 16S rRNA based tree.

E. meliloti AK58 and BO21CC show different symbiotic phenotypes with respect to the host plant *Medicago sativa*, as well as differences in substrates utilization [5]. Moreover *E. meliloti* AK58 and BO21CC present differences in cell morphology also, with AK58 being smaller than BO21CC and the other *E. meliloti* strains for which genome sequencing is available (Figure 2). Interestingly, BO21CC is also showing cells with a ratio between cell axes nearer 1 (more rounded cells), when compared with AK58 and with the other *E. meliloti* strains (Figure 2).

Genome sequencing information

Genome project history

AK58 and BO21CC strains were selected for sequencing on the basis of the Community Sequencing Program 2010 of DOE Joint Genome Institute (JGI) in relation to the project entitled "Complete genome sequencing of *Sinorhizobium meliloti* AK58 and BO21CC strains: Improving alfalfa performances through the exploitation of *Sinorhizobium* genomic data". The overall rationale for their genome sequencing was related to the identification of genomic determinants of different symbiotic performances between *S. meliloti* strains. The genome project is deposited in the Genomes On Line Database [21] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE-JGI. A summary of the project information is shown in Table 2.

Growth conditions and DNA isolation

E. meliloti strains AK58 and BO21CC (DSM23808 and DSM23809, respectively) were grown in DSMZ medium 98 (Rhizobium medium) [22] at 28°C. DNA was isolated from 0.5-1 g of cell paste using Jetflex Genomic DNA Purification kit (GENOMED 600100) following the standard protocol as recommended by the manufacturer with modification st/LALMP [23] for strain AK58 and additional 5 µl proteinase K incubation at 58° for 1 hour for strain BO21CC, respectively. DNA will be available on request through the DNA Bank Network [24].

Genome sequencing and assembly

The draft genomes were generated at the DOE Joint Genome Institute (JGI) using Illumina data [25]. For BO21CC genome, we constructed and sequenced an Illumina short-insert paired-end library with an average insert size of 270 bp which generated 76,033,356 reads and an Illumina long-insert paired-end library with an average insert size of 9,141.74 ± 1,934.63 bp which generated 4,563,348 reads totaling 6,463 Mbp of Illumina data. For AK58, a combination of Illumina [25] and 454 technologies [26] was used. For the AK58 genome we constructed and sequenced an Illumina GAii shotgun library which generated 80,296,956 reads totaling 6,102.6 Mb, a 454 Titanium standard library which generated 0 reads and 1 paired end 454 library with an average insert size of 10 kb which generated 326,569 reads totaling 96 Mb of 454 data. All general aspects of library construction and sequencing performed at the JGI can be found at [27]. The initial draft assemblies contained 194 contigs in 16 scaffold(s) for BO21CC, and 311 contigs in 5 scaffolds for AK58.

For BO21CC the initial draft data was assembled with Allpaths and the consensus was computationally shredded into 10 Kbp overlapping fake reads (shreds). The Illumina draft data was also assembled with Velvet, version 1.1.05 [28], and the consensus sequences were computationally shredded into 1.5 Kbp overlapping fake reads (shreds). The Illumina draft data was assembled again with Velvet using the shreds from the first Velvet assembly to guide the next assembly. The consensus from the second Velvet assembly was shredded into 1.5 Kbp overlapping fake reads. The fake reads from the Allpaths assembly and both Velvet assemblies and a subset of the Illumina CLIP paired-end reads were assembled using parallel phrap, version 4.24 (High Performance Software, LLC). Possible misassemblies were corrected with manual editing in Consed [29-31].

Table 1. Classification and general features of *E. meliloti* AK58 and BO21CC according to the MIGS recommendations [9] and the Names for Life database [10]

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Bacteria*	TAS [11]
		Phylum *Proteobacteria*	TAS [12]
		Class *Alphaproteobacteria*	TAS [12]
		Order *Rhizobiales*	TAS [12]
		Family *Rhizobiaceae*	TAS [12]
		Genus *Ensifer*	TAS [2,12]
		Species *Ensifer meliloti*	TAS [13]
		Strain BO21CC	TAS [4,5]
		Strain AK58	TAS [4,5]
	Gram stain	negative	TAS [12]
	Cell shape	rods	TAS [12]
	Motility	Motile	TAS [12]
	Sporulation	non-sporulating	TAS [12]
	Temperature range	mesophile, 20-37°C	TAS [12]
	Optimum temperature	25-30°C	TAS [12]
	Salinity	Tolerate 1.0% NaCl	TAS [12]
MIGS-22	Oxygen requirement	Aerobe	TAS [12]
	Carbon source	carbohydrates and salts of organic acids	TAS [12]
	Energy metabolism	chemoorganotroph	TAS [12]
MIGS-6	Habitat	Soil, root nodules of legumes	TAS [3,12]
MIGS-15	Biotic relationship	free living, symbiont	TAS [12]
MIGS-14	Pathogenicity	not reported	
	Biosafety level	1	TAS [14]
MIGS-23.1	Isolation	BO21CC: root nodules of *Medicago sativa* cv. 'Oneida' AK58: root nodules of *Medicago falcata*	TAS [4]
MIGS-4	Geographic location	BO21CC: Lodi, Italy AK58: Kazakhstan,	TAS [4]
MIGS-5	Sample collection time	BO21CC: 1997 AK58: 2001	NAS
MIGS-4.1	Latitude	BO21CC: 45.31 AK58: 58.75	NAS
MIGS-4.2	Longitude	BO21CC: 9.50 AK58: 48.98	NAS
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	BO21CC: 70 m AK58: 305 m	NAS

Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [15].

Gap closure was accomplished using repeat resolution software (Wei Gu, unpublished), and sequencing of bridging PCR fragments with Sanger and/or PacBio (unpublished, Cliff Han) technologies. For improved high quality draft and noncontiguous finished projects, one round of manual/wet lab finishing may have been completed. Primer walks, shatter libraries, and/or subsequent PCR reads may also be included for a finished project. A total of 128 additional sequencing reactions and 126 PCR PacBio consensus sequences were completed to close gaps and to raise the quality of the final sequence. The total ("estimated size" for unfinished) size of the BO21CC genome is 7.1 Mb and the final assembly is based on 6,463 Mbp of Illumina draft data, which provides an average 910 × coverage of the genome.

For AK58, the 454 Titanium standard data and the 454 paired end data were assembled together with Newbler, version 2.6 (20110517_1502). The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data was assembled with Velvet, version 1.1.05 [28], and the consensus sequence was computationally shredded into 1.5 kb overlapping fake reads (shreds). We integrated the 454 Newbler consensus shreds, the Illumina Velvet consensus shreds and the read pairs in the 454 paired end library using parallel phrap, version SPS - 4.24 (High Performance Software, LLC). The software Consed [29-31] was used in the following finishing process. Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible misassemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher [32], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR (J-F Cheng, unpublished) primer walks. A total of 0 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The estimated genome size of AK58 is 7 Mb and the final assembly is based on 61.5 Mb of 454 draft data which provides an average 8.8 × coverage of the genome and 420 Mb of Illumina draft data which provides an average 60 × coverage of the genome.

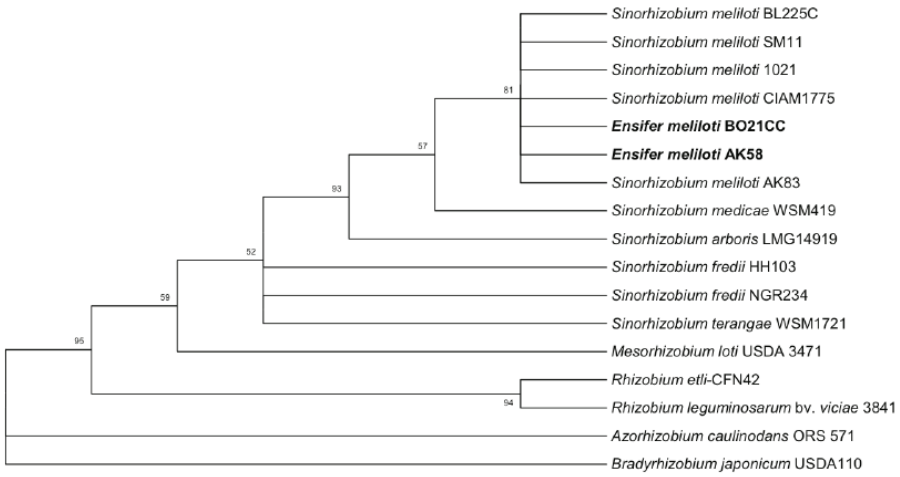


Figure 1. Phylogenetic consensus tree showing the position of *E. meliloti* AK58 and BO21CC strains in the *Ensifer/Sinorhizobium* genus. The phylogenetic tree was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model [17], chosen as model with the lowest BIC scores (Bayesian Information Criterion) after running a Maximum Likelihood fits of 24 different nucleotide substitution models (Model Test). The bootstrap consensus tree inferred from 500 replicates [18] is taken to represent the phylogenetic pattern of the taxa analyzed [18]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree with the highest log likelihood (-3411.7124) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (G, parameter = 0.3439). A total of 1,284 nt positions were present in the final dataset. Model test and Maximum Likelihood inference were conducted in MEGA5 [19]. In bold *E. meliloti* AK58 and BO21CC strains.

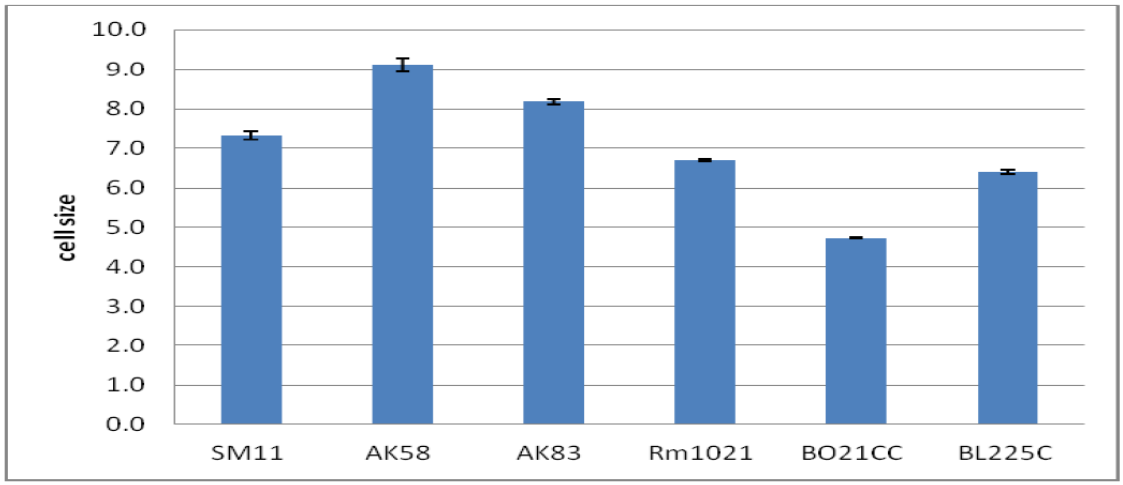


Figure 2. Cell morphology and cell size analysis of *E. meliloti* strains. Cell size analysis with Pixcavator IA 5.1.0.0 software [20] of logarithmically grown cultures (OD_{600}=0.6) in TY medium of AK58, BO21CC, plus other completely sequenced *E. meliloti* strains is reported. Cell size is expressed as cell area in μm^2, while roundness is the ratio between the two main axes of the cell. Standard errors after more than 300 individual observations are reported. Different letters indicate significant differences (P<0.05) after 1-way ANOVA.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-Quality Draft
MIGS-28	Libraries used	Two genomic libraries: one 454 PE library (9 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	60 × (AK58) 910 × (BO21CC) Illumina; 8.8 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3, Velvet version 1.0.13, phrap version, 1.080812, Allpaths version 39750,
MIGS-32	Gene calling method	Prodigal
	GenBank Date of Release	Pending
	GOLD ID	BO21CC: Gi07569 AK58: Gi07577
	NCBI project ID	BO21CC: 375171 AK58: 928722
	Database: IMG	BO21CC: 9144 AK58: 7327
MIGS-13	Source material identifier	BO21CC: DSM23809 AK58: DSM23808
	Project relevance	CSP2010, biotechnological, biodiversity

Genome annotation

Genes were identified using Prodigal [33] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [34]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [16].

Genome properties

The High-Quality draft assemblies of the genomes consist of 41 scaffolds for BO21CC and 9 scaffolds for AK58 representing overall 6,985,065 and 6,974,333 bp, respectively. The overall G+C content was 62.12% and 62.04% for BO21CC and AK58, respectively (Table 3a and Table 3b). Of the 6,746 and 6,992 genes predicted, 5,357 and 5,549 were protein-coding genes, and 105 and 79 RNAs were present in BO21CC and AK58, respectively. The large majority of the protein-coding genes (79.32% and 78.03%, BO21CC and AK58, respectively) were assigned a putative function as COGs. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3a. Genome Statistics for strain BO21CC

Attribute	Value	% of Total
Genome size (bp)	6,985,065	100.00%
DNA coding region (bp)	6,011,953	86.07%
DNA G+C content (bp)	4,339,356	62.12%
Number of scaffolds	41	
Total genes	6,746	100.00%
RNA genes	105	1.72%
rRNA operons	3	
tRNA genes	58	0.86%
Protein-coding genes		
Genes with function prediction (proteins)	5,357	79.41%
Genes in paralog clusters	3,275	48.55%
Genes assigned to COGs	5,351	79.32%
Genes assigned Pfam domains	5,318	78.83%
Genes with signal peptides	1,427	21.15%
Genes with transmembrane helices	1,521	22.55%

Table 3b. Genome statistics for strain AK58

Attribute	Value	%age
Genome size (bp)	6,974,333	100.00%
DNA coding region (bp)	5,914,246	84.80%
DNA G+C content (bp)	4,315,694	62.04%
Number of scaffolds	9	
Total genes	6,992	100.00%
RNA genes	79	1.13%
rRNA operons	1*	
tRNA genes	49	0.70%
Protein-coding genes	6,934	98.87%
Genes with function prediction (proteins)	5,459	77.84%
Genes in paralog clusters	2,912	41.52%
Genes assigned to COGs	5,472	78.03%
Genes assigned Pfam domains	5,420	77.29%
Genes with signal peptides	1,432	20.42% %
Genes with transmembrane helices	1,465	20.89%

*only one rRNA operon appears to be complete.

Table 4. Number of genes associated with the general COG functional categories

	BO21CC		AK58		
Code	**Value**	**%age**	**Value**	**% age**	**Description**
E	637	10.69	685	11.20	Amino acid transport and metabolism
G	604	10.14	596	9.75	Carbohydrate transport and metabolism
D	45	0.76	53	0.87	Cell cycle control, cell division, chromosome partitioning
N	69	1.16	68	1.11	Cell motility
M	305	5.12	298	4.87	Cell wall/membrane biogenesis
B	1	0.02	3	0.05	Chromatin structure and dynamics
H	202	3.39	205	3.35	Coenzyme transport and metabolism
V	64	1.17	62	1.01	Defense mechanisms
C	365	6.13	356	5.82	Energy production and conversion
W	1	0.02	1	0.02	Extracellular structures
S	608	10.20	617	10.09	Function unknown
R	730	12.25	767	12.54	General function prediction only
P	320	5.17	294	4.81	Inorganic ion transport and metabolism
U	104	1.75	102	1.67	Intracellular trafficking and secretion, and vesicular transport
I	210	3.52	217	3.55	Lipid transport and metabolism
F	107	1.80	114	1.86	Nucleotide transport and metabolism
O	185	3.10	189	3.09	Posttranslational modification, protein turnover, chaperones
L	273	4.58	327	5.35	Replication, recombination and repair
Q	163	2.74	159	2.60	Secondary metabolites biosynthesis, transport and catabolism
T	247	4.14	249	4.07	Signal transduction mechanisms
K	524	8.79	551	9.01	Transcription
J	195	3.27	201	3.29	Translation, ribosomal structure and biogenesis
-	1395	20.68	1541	21.97	Not in COGs

Acknowledgements

We are grateful to Dr. M.L. Roumiantseva and Dr. B. Simarov (Research Institute for Agricultural Microbiology, St-Petersburg-Puskin, Russia) for original isolation and the permission to use strain AK58 in this work. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy Under Contract No. DE-AC02-05CH11231.

References

1. Martens M, Delaere M, Coopman R, De Vos P, Gillis M, Willems A. Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* 2007; **57**:489-503. PubMed http://dx.doi.org/10.1099/ijs.0.64344-0

2. Young JM. The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen et al. 1988, and *Sinorhizobium morelense* Wang et al. 2002 is a later synonym of *Ensifer adhaerens* Casida 1982. Is the combination "Sinorhizobium adhaerens" (Casida 1982) Willems et al. 2003 legitimate? Request for an Opinion. *Int J Syst Evol Microbiol* 2003; **53**:2107-2110. PubMed http://dx.doi.org/10.1099/ijs.0.02665-0

3. Sprent JI. Nodulation in legumes. London: Royal Botanic Gardens, Kew.; 2001.

4. Giuntini E, Mengoni A, De Filippo C, Cavalieri D, Aubin-Horth N, Landry CR, Becker A, Bazzicalupo M. Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. *BMC Genomics* 2005; **6**:158. PubMed http://dx.doi.org/10.1186/1471-2164-6-158

5. Biondi EG, Tatti E, Comparini D, Giuntini E, Mocali S, Giovannetti L, Bazzicalupo M, Mengoni A, Viti C. Metabolic capacity of

Sinorhizobium (*Ensifer*) *meliloti* strains as determined by phenotype microarray analysis. *Appl Environ Microbiol* 2009; **75**:5396-5404. PubMed http://dx.doi.org/10.1128/AEM.00196-09

6. Galardini M, Mengoni A, Brilli M, Pini F, Fioravanti A, Lucas S, Lapidus A, Cheng JF, Goodwin L, Pitluck S, *et al*. Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti*. *BMC Genomics* 2011; **12**:235. PubMed http://dx.doi.org/10.1186/1471-2164-12-235

7. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. *Curr Opin Genet Dev* 2005; **15**:589-594. PubMed http://dx.doi.org/10.1016/j.gde.2005.09.006

8. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM and others. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucl. Acids Res. 2009;37(suppl_1):D141-145.

9. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

10. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.

11. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

12. Brenner DJ, Krieg NR, Staley JT. Bergeys' Manual of Systematic Bacteriology. Volume 2 The Proteobacteria. Part C The Alpha-, Beta-, Delta-, and Epsilonproteobacteria. Garrity GM, editor: Springer; 2005.

13. De Lajudie P, Willems A, Pot B, Dewettinck D, Maestrojuan G, Neyra M, Collins MD, Dreyfus B, Kersters K, Gillis M. Polyphasic Taxonomy of Rhizobia: Emendation of the Genus *Sinorhizobium* and Description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *Int J Syst Bacteriol* 1994; **44**:715-733. http://dx.doi.org/10.1099/00207713-44-4-715

14. BAuA. Classification of *Bacteria* and *Archaea* in risk groups. *TRBA* 2010; **466**:80.

15. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

16. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. PubMed http://dx.doi.org/10.1093/bioinformatics/btp393

17. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol* 1992; **9**:678-687. PubMed

18. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 1985; **39**:783-791. http://dx.doi.org/10.2307/2408678

19. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011; **28**:2731-2739. PubMed http://dx.doi.org/10.1093/molbev/msr121

20. Pixcavator IA. 5.1.0.0 Intelligent Perception <http://inperc.com >.

21. Pagani I, Liolios K, Jansson J, Chen IMA, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012; **40**(D1):D571-D579. PubMed http://dx.doi.org/10.1093/nar/gkr1100

22. List of growth media used at DSMZ <http://www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html>.

23. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009; **462**:1056-1060. PubMed http://dx.doi.org/10.1038/nature08656

24. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network. *The Start from a German Initiative Biopreservation and Biobanking* 2011; **9**:51-55. http://dx.doi.org/10.1089/bio.2010.0029

25. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. PubMed http://dx.doi.org/10.1517/14622416.5.4.433

26. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, *et al*. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; **437**:376-380. PubMed

27. DOE joint Genome Institute. The Regents of the University of California <http://www.jgi.doe.gov/>.

28. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. PubMed http://dx.doi.org/10.1101/gr.074492.107

29. Ewing B, Green P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res* 1998; **8**:186-194. PubMed http://dx.doi.org/10.1101/gr.8.3.175

30. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res* 1998; **8**:175-185. PubMed http://dx.doi.org/10.1101/gr.8.3.175

31. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**:195-202. PubMed http://dx.doi.org/10.1101/gr.8.3.195

32. Han C. P. C. Finishing Repetitive Regions Automatically with Dupfinisher. 2006; Las Vegas, Nevada, USA. CSREA Press.

33. Hyatt D, Chen GL, LoCascio P, Land M, Larimer F, Hauser L. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. PubMed http://dx.doi.org/10.1186/1471-2105-11-119

34. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. PubMed http://dx.doi.org/10.1038/nmeth.1457

Plant Soil (2014) 377:189–201
DOI 10.1007/s11104-013-1979-3

REGULAR ARTICLE

Molecular phylogeny of the nickel-resistance gene *nreB* and functional role in the nickel sensitive symbiotic nitrogen fixing bacterium *Sinorhizobium meliloti*

Francesco Pini · Giulia Spini · Marco Galardini · Marco Bazzicalupo · Anna Benedetti · Manuela Chiancianesi · Alessandro Florio · Alessandra Lagomarsino · Melania Migliore · Stefano Mocali · Alessio Mengoni

Received: 25 June 2013 / Accepted: 11 November 2013 / Published online: 17 December 2013

Abstract
Aims Heavy-metal tolerance is a widespread phenotype in bacteria, particularly occurring in strains isolated from heavy-metal contaminated sites. Concerning nickel tolerance, the *nre* system is one of the most common. An ortholog of the *nreB* gene is present in the alfalfa symbiont *Sinorhizobium meliloti* also, which stir the attention on its functional role in such Ni-sensitive species and on the evolutionary relationships with Ni-resistant strain orthologs.
Methods Phylogenetic reconstruction and comparative genomics were performed to analyze the phylogenetic relationships of *nreB* orthologs, as well aa *nreB* deletion mutant *S. meliloti* strain was constructed and subjected to phenotypic analysis.
Results Phylogenetic analysis of *nreB* genes indicated horizontal gene transfer events, possibly mediated via mobile genetic elements. Phenotype Microarray, biochemical and symbiotic analyses of the deletion mutant strain (Δ*nreB*) showed that in *S. meliloti nreB* is involved in the tolerance to several stresses other than Ni (mainly urea and copper), possibly partially mediated through the modulation of urease and hydrogenase activities.
Conclusions Obtained results allowed us to speculate that *nreB* is a highly mobile gene cassette, spread in the bacterial phylogenetic tree via many HGT events, which could have been recruited to confer nickel-tolerance in strains thriving in contaminated environments, by small changes linked to its basic functions (e.g. modulation of urease and hydrogenase activity).

Keywords Nickel tolerance · *nreB* · Horizontal gene transfer · Phenotype Microarray · *Sinorhizobium meliloti*

Responsible editor: Henk Schat.

Electronic supplementary material The online version of this article (doi:10.1007/s11104-013-1979-3) contains supplementary material, which is available to authorized users.

F. Pini · G. Spini · M. Galardini · M. Bazzicalupo · M. Chiancianesi · A. Mengoni (✉)
Department of Biology, University of Firenze, via Madonna del Piano 6, 50015 Sesto Fiorentino, Italy
e-mail: alessio.mengoni@unifi.it

F. Pini
Interdisciplinary Research Institute - IRI CNRS USR3078, Parc de la Haute Borne 50 avenue de Halley, 59658 Villeneuve d'Ascq Cedex, France

A. Benedetti · A. Florio · M. Migliore
Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per lo Studio delle Relazioni Fra Pianta e Suolo (CAR-RPS), Via della Navicella, 2/4, 00184 Roma, Italy

A. Lagomarsino · S. Mocali
Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per l'Agrobiologia e la Pedologia (CRA-ABP), P.za D'Azeglio, 30, 50121 Firenze, Italy

Introduction

Heavy-metal resistance is a widespread phenotype in both plant and microbial world; it is found in species

growing in either metal-polluted soils (as for instance mining and industrial areas) and in heavy-metal rich soils due to their peculiar geology (e.g. serpentine outcrops). Heavy-metal resistance is interesting both for its possible application in bioremediation practices and for its peculiar evolutionary features. In fact, heavy-metal tolerance appear rapidly several times in unrelated organisms thriving in contaminated areas, suggesting the presence of pre-adapted physiological (molecular) mechanisms. In particular for bacteria, heavy-metal resistance strains belong to multiple phylogenetic lineages and several types of genes conferring the heavy-metal resistance phenotype have been found. Indeed, several of these genetic determinants belong to common metabolic pathways (e.g. glutathione synthesis) or may have orthologs present in nonresistant strains, which may contribute to some additional cellular function (Mengoni et al. 2010).

Metal resistance presence in plant and soil bacteria has then stirred the attention of several investigators (for a review see (Mengoni et al. 2010)). In particular, metal resistance bacterial strains were initially found in metal contaminated sites, and their genetic determinants were identified (e.g. cnr, nre, czc systems) (Mergeay et al. 2003). Then, some of these genetic determinants were found in metal-resistant bacteria isolated from metal-rich serpentine soils, suggesting an evolutionary route of these genetic determinants from serpentine strains to strains inhabiting metal-polluted sites (Mengoni et al. 2010).

However, homologs to metal resistance determinants have been also found in many nonresistant bacteria (Mengoni et al. 2010), but their functional role in those strains has not been determined.

Concerning nickel, several nickel-resistant bacteria have been isolated from heavy-metal-contaminated sites and, at least, ten different molecular (genetic) mechanism of resistance have been found (Mengoni et al. 2010). Two well-studied examples include *Cupriavidus metallidurans* CH34 and *Cupriavidus metallidurans* 31A (previously named *Achromobacter xylosoxidans* 31A) (Grass et al. 2001). In particular, in *C. metallidurans* 31A, the *nreB* gene only is required for nickel resistance. This gene encodes for a Ni/H^+ antiporter belonging to the DHA3 family of the Major Facilitator Superfamily (MFS), the largest group of secondary active transportes, which includes 58 different families and is ubiquitously found in all the three domains of of life, viz. Bacteria, Eukarya and Archaea (Law et al. 2008). Homologs of *nreB* have been found in *Bradyrhizobium* strains isolated from nodules of the New Caledonian endemic legume *Serianthes calycina* growing in nickel-rich soils (Chaintreuil et al. 2007). These strains were able to grow in the presence of 15 mM $NiCl_2$ thanks to the *cnr* operon, while the *nreB* gene seems to be not directly correlated with nickel resistance (Chaintreuil et al. 2007).

Sinorhizobium meliloti is one of the most investigated model bacterial species for symbiotic interaction with plants. Recently, during a comparative genome analysis of three strains (Rm1021, AK83, BL225C), the gene SMa1641, encoding for a Ni^+/H^+ antiporter homolog to *nreB* from *C. metallidurans* 31A, was found to be present in all sequenced genomes (Galardini et al. 2011), thus being part of the so-called "core genome" of *S. meliloti* species. Such evidence has also been confirmed by an additional PCR screening of a collection of 148 *S. meliloti* strains coming from Italy, Tunisia, Iran and Kazakhstan which found *nreB* ortholog present in all *S. meliloti* strains (unpublished data). Interestingly, the gene SMa1641 is harbored by the symbiotic-required megaplasmid pSymA allowing us to hypothesize that SMa1641 could be involved in some core biological process, other than Ni tolerance, and that it could be involved in the symbiotic abilities of *S. meliloti*. Indeed, a correct balance of metals could be crucial in the symbiosome development. From one side bacteria need trace metals for the synthesis of key enzymes in nitrogen fixation, from the other side the excess of these metals could be toxic. Under this rational a $P_{1B}P_{1B-5}$ ATPase (Nia, Nickel/Iron ATPase, encoded by SMa1163) has been discovered recently, which prevents an accumulation of iron over the toxic level in the symbiosome (Zielazinski et al. 2013). Indeed, symbiotic conditions are particularly sensible to trace metal homeostasis, because of the lack of extracellular polymeric substances, which play a significant role in metal homeostasis under free living conditions (Hou et al. 2013). *S. meliloti* contains another putative nickel homeostasis mechanism, *dme*RF (SMc04167-SMc04618), located on the chromosome. This system has been demonstrated in *Rhizobium leguminosarum* bv. *viciae* to be involved in Co(II) and Ni(II) tolerance (Rubio-Sanz et al. 2013). Additionally, in *S. meliloti* a P-ATPase (SMa1163) involved in Ni and Fe transport during symbiosis has been identified (Zielazinski et al. 2013). Consequently the function of *nreB* in *S. meliloti* is particularly intriguing, in relation also to its genomic localization on the symbiotic megaplasmid pSymA.

To shed light on the phylogenetic relationships and functions of *nreB* orthologs we used a dual approach based on: i) phylogenetic reconstruction of *nreB* orthologs genealogy, taxonomic occurrence and genomic location by probing the genome sequence databases; ii) characterization of SMa1641 functions in the model strain *S. meliloti* Rm1021 by using a deletion mutant for this gene.

Materials and methods

Phylogenetic analysis

Phylogenetic analyses were performed using the Phylogeny.fr web server (Dereeper et al. 2008, 2010). Sequences were aligned with MUSCLE (v3.7) configured for highest accuracy (MUSCLE with default settings). After alignment, ambiguous regions (i.e. containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) using the following parameters: minimum length of a block after gap cleaning: 10; no gap positions were allowed in the final alignment; all segments with contiguous nonconserved positions bigger than 8 were rejected; minimum number of sequences for a flank position: 85 %. The phylogenetic trees were reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0 aLRT). The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.007) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma = 0.948). Reliability for internal branch was assessed using the aLRT test (SH-Like). Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3). Codon Adaptation Index was computed through the E-CAI server (Puigbo et al. 2008).

Bacterial strains, plasmids, and growth conditions

Bacterial strains and plasmids used in this study are listed in Table 1 *Escherichia coli* strains were grown in liquid or solid Luria-Bertani (LB) broth (Sigma Aldritch) (Sambrook et al. 1989) at 37 °C, supplemented with opportune antibiotics: kanamycin (50 μg/ml in broth and agar), tetracycline (10 μg/ml in broth and agar). *S. meliloti* strains were grown in TY medium (Beringer 1974) supplemented with kanamycin (200 μg/ml in broth and agar), streptomycin (500 μg/ml in broth and agar), tetracycline (1 μg/ml in liquid broth, 2 μg/ml in agar), when necessary. For mutants, the counter-selection was performed in 10 % sucrose added to agar plates.

For conjugation, recipients *S. meliloti* Rm1021 was grown overnight in TY medium, donor *E. coli* S17-1 containing the plasmid was grown overnight in LB medium supplemented with opportune antibiotic. About 1×10^9 *S. meliloti* and 0.5×10^9 *E. coli* cells were used for each mating. Cells of both donor and recipient were washed twice with 0.85 % NaCl. Then *S. meliloti* and *E. coli* cells were mixed and resuspended in a final volume of 0.1 ml of 0.85 % NaCl. Mating cells were transferred to TY plates and incubated at 30 °C for 24 h. For selection of transconjugants, aliquots from serial dilutions were plated on selective and non-selective medium and incubated at 30 °C for 3 days. For genes deletion, two fragments of about 1000-bp long (P1-P2 and P3-P4) on either side of *nreB* (SMa01641) were amplified by PCR using specific oligonucleotides containing restriction enzymes sites (see Table 1) for directional forced cloning with a Tc cassette (Pini et al. 2013). Regions of homology were amplified by PCR using the standard conditions. Products were then gel-purified (Qiagen, Valencia, California, United States) cloned in pGEM-T easy and sequenced. Fragments were excised from pGEM-T easy gel purified and ligated with a Tc cassette into pNPTS138. Ligations were transformed into DH5α and positive colonies selected by blue/white screening.

Deletion plasmid was transformed into *E. coli* S17-1 and then transferred by conjugation into Rm1021, first integrants were selected by plating on TY containing kanamycin and tetracycline. Colonies with the integrated deletion plasmid were inoculated into liquid TY medium with tetracycline and grown for 12–16 h. Serial dilutions were then plated on TY plates containing tetracycline and sucrose. Colonies were screened for tetracycline resistance and for resistance to the activity of *sacB* gene (loss of the plasmid), to identify deletion strains (Skerker and Laub 2004). Gene deletion was verified by PCRs using primers Full_Sma1641_flanking_Fw and Full_Sma1641_flanking_Rv (see Table 1). A control strain of Rm1021 carrying a tetracycline resistance cassette was also prepared by conjugating *E. coli* S17-1 cells harboring pMR20 vector (Roberts et al. 1996) to obtain *S. meliloti* Rm1021 pMR20 strain (BM623) and

Table 1 Strains, plasmids and primers used in this work

Strains	Strain, plasmid or primer name	Description	Resistance	Source
S. meliloti	Rm1021	SU47 *str-21*	Sm	Galibert et al. 2001
	BM589	Rm1021 Δ*nreB* ::tc	Sm, Tc	This work
	BM623	Rm1021 pMR20	Sm, Tc	This work
E. coli	DH5α	F, *supE44*, *lacU169*, *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1* (80*lacZ*M15)	-	Hanahan 1983
	S17-1	*recA*, *pro*, *hsdR*, *RP4-2-Tc::Mu-km::Tn7*	–	Simon et al. 1983
Plasmids				
General purpose vector	pNTPS138	Suicide vector, *oriT*, *sacB*	Km	Gift from D. Alley
	pMR20	Low copy number vector	Tc	Roberts et al. 1996
Deletion plasmid	p*ΔnreB*	pNPTS138-Tc deletion cassette for *nreB*	Km, Tc	This work
Primers				
	P1_Sma1641_HindIII	AAGCTTAAGAGCGGCCTCAG		This work
	P2_Sma1641_EcoRI	GAATTCGCGGTTGGCAAGGA		This work
	P3_Sma1641_EcoRI	GAATTCATTGGCCGACCGAG		This work
	P4_Sma1641_BamHI	GGATCCGCCGCTGCTGCTAA		This work
	Full_Sma1641_flanking_Fw	CCAGCAGCTCCATGCTGTTG		This work
	Full_Sma1641_flanking_Fw	TCTCCATTCGCGTCCGGAAG		This work
	rplM qPCR_Fw	AAGCGGCCTTCGATGATCTG		This work
	rplM qPCR_Rv	CTCCACCGGCAGAAGTACAC		This work
	UreC qPCR_Fw	AGATTCATCGGAAACGCATC		This work
	UreC qPCR_Rv	CAGATCGAAGAGGCGTTGAT		This work

used for phenotype microarray experiments and for testing the sensitivity to Ni(II) and Cu(II).

Phenotype microarray analysis and growth curves

The mutant strain was assayed on PM (Biolog Inc., Hayward, CA) microplates PM9, PM10 and PM13B testing 288 different conditions. All procedures were performed as indicated by the manufacturer. Inoculants were prepared as described by Viti and colleagues (Viti et al. 2007), for the microplate PM13B was used the inoculation fluid IF-10b. All PM microplates were incubated at 30 °C in an OmniLog reader, and changes of color in the wells were monitored automatically every 15 min. Readings were recorded for 72 h, and data (from duplicate tests for each strains) were analyzed using OmniLog PM software (release OM_PM_109M) and subsequently analyzed following the same approach by Peleg and colleagues (Peleg et al. 2012). Growth curves were measured in liquid TY medium in presence of different $NiCl_2$ and $CuCl_2$. After a pre-inoculum of 24 h at 30 °C under continuous agitation, cultures were diluted to OD_{600nm}=0.1 and $NiCl_2$ and $CuCl_2$ solutions were added to final concentrations of 0.5 mM, 1.0 mM, 1.5 mM and 2.0 mM. Then, cultures were let grown for 48 h under continuous agitation at 30 °C and growth was measured by OD_{600nm} readings.

Nodulation of *M. sativa*

Seedlings of *M. sativa* cv. '*Pomposa*' were germinated for 48 h in a dark at room temperature and transferred directly to sterile plastic petri dishes containing buffered Nod medium (Ehrhardt et al. 1992) supplemented with 1 μM AVG (ethylene biosynthesis inhibitor aminoethoxy-vinyl-glycine) in 1.5 % (wt/vol) Noble agar (Sigma). Plantlets were grown for an additional 4 days before inoculation with 100 μl of *S. meliloti* strains at OD_{600nm} of 0,5. Thirteen plates for each strain were sealed with Parafilm, with a hole to let the plant grow outside, and transferred in a near-vertical position to a growth chamber maintained at 26 °C with a 16-h photoperiod (100 microeinstein m^{-2} s^{-1}). The plants' height was scored after 30 days.

Urease activity measurement and quantitative PCR on *ureC* gene

Urease activity was assayed on liquid cultures by using a modified protocol from Kandeler and Gerber (Kandeler and Gerber 1988). In this work 300 μl of cellular suspension (approx. 10^8 cells) in BNM liquid media were incubated with 2,5 ml of urea 0, 5,10, 30, 50, 80, 100, 150, 250, 500 mM for 2 h at 30 °C. Experimental data were analyzed using the Michaelis–Menten model. Km, Vmax and the specificity constant (Ka) were calculated by the non-linear regression using Graph-Pad PRISM as described in (Moscatelli et al. 2012). Total RNA was extracted with RNeasyMini Kit (Qiagen) and retrotranscribed with SuperScriptIII Reverse Transcriptase (Invitrogen) as described in (Vanucci et al. 2005). Quantitative PCR was performed by using primers on *ureC* gene and *rplM* (Table 1) as reference gene with SYBR Green Mix (Sigma-Aldrich) primers (10 μM each), reverse transcriptase products (50 ng), and water to 25 μl total volume. Cycling conditions were as follows: 2′ 94 °C, followed by 15″ 94 °C, 15″ 62 °C, 30″ 72 °C repeated for 40 cycles. A melting curve was performed after cycling. REST2009 software (Pfaffl et al. 2002) was used to analyze relative expression data with the Delta-DeltaCt method.

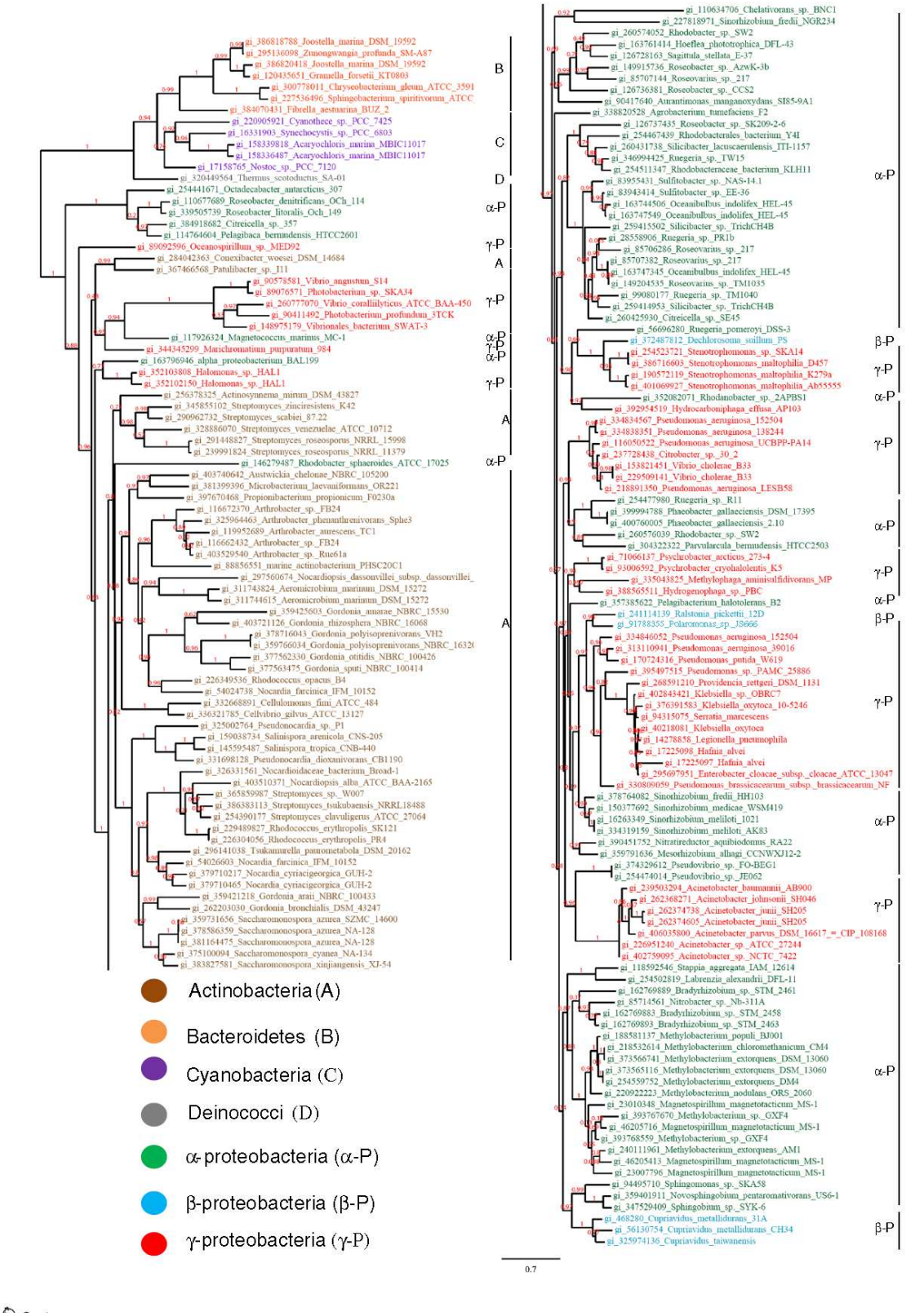
Results

Molecular phylogeny of *nreB* genes

Using the protein sequence of *C. metallidurans* 31A as query on the protein sequence GenBank database 191 ortholog sequences were retrieved. These sequences were used to build a phylogenetic tree (Fig. 1). Ortholog sequences of *nreB* from *Cupriavidus metallidurans* 31A were retrieved in 7 different groups (phyla or classes), but most of the sequence falls within the *Actinobacteria* (27.23 % of total) and in *Alphaproteobacteria* (38.74 % of total) and *Gammaproteobacteria* (24.61 % of total).

Notably, the topology of the tree based on NreB did not match with bacterial phylogeny, which suggest that *nreB* orthologs were subjected to horizontal gene transfer (HGT) events. In fact, NreB orthologs from the same taxonomic group were splitted into several different clusters, intermingled with other clusters composed by NreB orthologs from different taxa. This situation was particularly evident for *Proteobacteria* orthologs, where members from the same class clusterized in different positions, clearly indicating the presence of past horizontal gene transfer events. Looking more closely to the orthologs of the betaproteobacterium *C. metallidurans* (strains 31A and CH34), these were affiliated with the ortholog from *C. taiwanensis*, a beta-rhizobium originally isolated from *Mimosa* root nodules (Amadou et al. 2008). Then, *Cupriavidus* orthologs clusterized with orthologs from *Alphaproteobacteria* (*Bradyrhizobium* and *Methylobacterium*, mainly). Orthologs from *Sinorhizobium meliloti* formed a subcluster within a group of *Gammaproteobacteria*, among which representatives from *Serratia marcenscens*, *Hafnia alvei* and *Klebsiella oxytoca* (which include heavy-metal tolerant strains) were present (Marrero et al. 2007; Park et al. 2004, 2008; Stoppel et al. 1995). Traces of past HGT events were also present inside *Actinobacteria*, i.e. in the different groupings of *Gordonia* and *Streptomyces* orthologs. Evidences for HGT of *nreB* genes were also supported when looking at G + C content of *nreB* gene, by comparison with the G + C content of surrounding genomic regions and of total genome of strains (Table 2). In most cases *nreB* genes had higher (even more than 2 percentage points) GC% than total genome (only the ortholog of *Sphingobium* sp. SYK-6 had a slightly lower GC% value), strongly suggesting that *nreB* orthologs were acquired from other species.

The potential mechanisms responsible for HGT events could be clearly appreciated when looking at the genomic context of *nreB* orthologs in completely sequenced genomes (Fig. 2). In fact, inside *Proteobacteria*, the presence of transposases (as insertion sequences, IS, signatures) flanking *nreB* was found (Fig. 2b). Moreover, in 14 out of 52 genome analyzed, *nreB* is harbored by plasmids (or megaplasmids), which may favor HGT events (Wiedenbeck and Cohan 2011). Indeed in *Ralstonia pickettii* 12D, *nreB* is located on a plasmid (pRp12D02) rich in sequences related to mobile elements (5.4 % of the plasmid gene content). Moreover this plasmid contains several genes involved in heavy metal resistance (19.6 % of the genes), some of them being highly similar to those of metal-resistant *C. metallidurans* strains. From these findings, observing the position of *R. pickettii nreB* gene in the phylogenetic tree (grouped with gammaproteobacteria) and considering the multiple environments colonized by this species (see Fig. 2), we can speculate that *R. pickettii* 12D could be a hot-spot or "hub" for HGT events (Fondi et al. 2010). Interestingly,


Actinobacteria (A)
Bacteroidetes (B)
Cyanobacteria (C)
Deinococci (D)
α-proteobacteria (α-P)
β-proteobacteria (β-P)
γ-proteobacteria (γ-P)
0.7

◀ **Fig. 1** Phylogenetic tree of NreB proteins found in GenBank database. The tree was built using the maximum likelihood method and the Approximate Likelihood-Ratio Test (aLRT) was used as statistical tests for branch support. On the right of the strains are indicated the groups to which they belong: Actinobateria (**a**), Bacteroidetes (**b**), Cyanobacteria (**c**), Deinococci (**d**) α-proteobacteria (α-P), β-proteobacteria (β-P) and γ-proteobacteria (γ-P)

nreB is often associated with a gene encoding for a putative transcriptional regulator, a homologue of *nreA* in *Proteobacteria* and an *arsR* family trascriptional regulator in *Actinobacteria* (Fig. 2b), which suggest that the entire gene cassette encoding for the NreB hydrogenase Ni^{2+}/H^{+} antiporter and its transcriptional regulator could have been transferred often. However, in several strains no transcriptional regulator is present in proximity of *nreB*, suggesting a different transcriptional regulation and puzzling over its role in cellular functions other than nickel resistance.

Finally, by mining the GOLD database (Bernal et al. 2001), considering the isolation sites of the respective bacterial strains for which genome sequences are available, 8 out of 52 strains only have been isolated from polluted sites, the remaining being strains present in water, in soil or in association (as pathogens or symbionts/commensals) with animals and plants. Interestingly, no relationships is present between the environmental distribution of strains (Fig. 2c), the cluster topology of NreB (Fig. 2a) and the genomic context (Fig. 2b), thus leaving open the question about the dynamics of HGT of *nreB* in the different bacterial strains.

Genomic context and functional role of *nreB* in the nonresistant bacterium *Sinorhizobium meliloti*

Since NreB of *Cupriavidus metallidurans* 31A and of the other *Cupriavidus* species, form a cluster with members of nonresistant *Alphaproteobacteria* (see Fig. 1), we focused our attention on the model symbiotic alpharhizobium *S. meliloti*, thanks to the availability of the genome sequences of several strains (Galardini et al. 2011) and the amenability of the model strain *S. meliloti* Rm1021 to genetic manipulation (see for instance (Ferri et al. 2010)). In *S. meliloti* Rm1021 *nreB* gene is located on the megaplasmid pSymA, which harbor several of the genes essential for establishing the symbiosis with leguminous plants. As previously shown (Fig. 2b), in *S. meliloti* Rm1021 *nreB* is flanked by a homologue of *nreA*, a putative regulator. A PCR amplification confirmed the presence of *nreA* flanking *nreB* in all 148

Table 2 Difference in G + C content between *nreB* and the genomes of selected alpha-proteobacterial strains

gi	Strain name	GC% total genome	GC% *nreB* gene	GC% ± 10kbp
56696280	*Ruegeria pomeroyi* DSS-3	64.1	69.64	65.5
99080177	*Ruegeria sp.* TM1040	60.1	64.1	62.26
304322322	*Parvularcula bermudensis* HTCC2503	60.7	66.9	64.89
347529409	*Sphingobium sp.* SYK-6	65.6	65.02	65.9
240111961	*Methylobacterium extorquens* AM1	68.5	70.1	67.03
220922223	*Methylobacterium nodulans* ORS 2060	68.5	70.56	65.9
188581137	*Methylobacterium populi* BJ001	69.4	72.51	72.51
218532614	*Methylobacterium chloromethanicum* CM4	68.1	71.31	69.85
254559752	*Methylobacterium extorquens* DM4	68.5	71.69	68.48
357385622	*Pelagibacterium halotolerans* B2	61.4	61.74	61.17
150377692	*Sinorhizobium medicae* WSM419	61.2	64.78	61.4
16263349	*Sinorhizobium meliloti* 1021	62.2	63.34	59.37
334319159	*Sinorhizobium meliloti* AK83	61.9	63.42	59.17
110677689	*Roseobacter denitrificans* OCh 114	58.9	61.43	59.27
339505739	*Roseobacter litoralis* Och 149	57.2	59.57	58.9
117926324	*Magnetococcus sp.* MC-1	54.2	58.21	53.26
146279487	*Rhodobacter sphaeroides* ATCC 17025	68.2	69.45	67.92

The G + C percentage (GC%) of total genome, nreB orthologs and of the 10 kbp surrounding nreB genes are reported

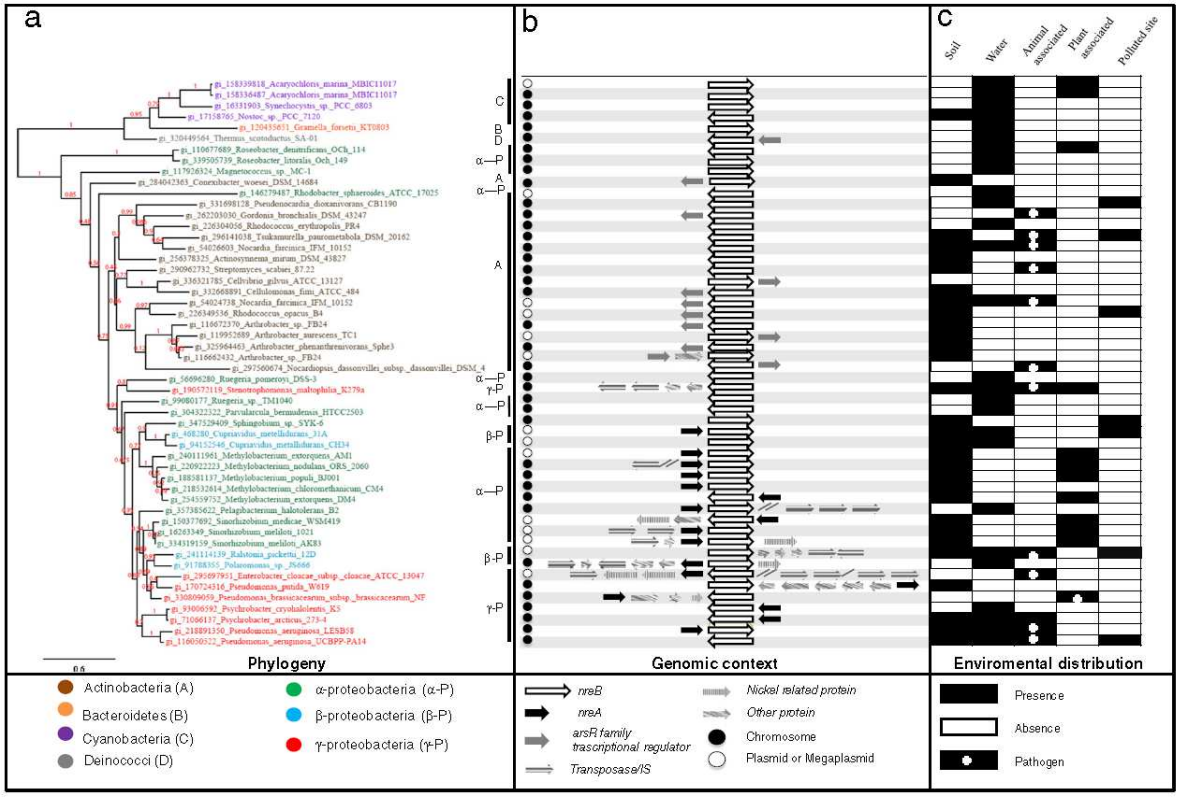


Fig. 2 Genomic context of NreB in completely sequences bacterial genomes. **a** A phylogenetic tree (maximum likelihood method), **b** the genomic context and **c** the environmental distribution (using GOLD database) of strains are shown. Details for the phylogenetic tree are the same as in Fig. 1

strains (coming from Italy, Germany, Tunisia, Iran and Kazakhstan) of our collection (Table S1).

These data may suggest a relatively ancient introgression of *nreB* gene in *S. meliloti* pangenome. Codon adaptation index (CAI) was then calculated, as a metrics of introgression time, by comparing the codon usage table of *nreB* with that of total *S. meliloti* genome. Obtained results showed an eCAI ($p<0.05$)=0.781, suggesting that the HGT event was indeed not recent (for a comparison on CAIs on other *S. meliloti* genes, see (Castillo-Ramirez et al. 2009), and may predate the origin of genus *Sinorhizobium*, as also suggested by the dendrogram (Fig. 1).

In order to elucidate the function of *nreB* gene (SMa1641) we constructed a deletion (Δ*nreB*) mutant strain (BM589). The BM589 mutant strain was then tested for its capabilities to grow on nickel in comparison with respect to the control strain (BM623), which carries the tetracycline resistance cassette in a low copy number vector (pMR20). As shown in Fig. 3 the mutant is slightly more sensitive to $NiCl_2$, indeed confirming earlier reports on *C. metallidurans* 31A on the role of NreB in determining nickel tolerance through the efflux of Ni^{2+} mediated by the proton motive force (Grass et al. 2001).

To investigate if *nreB* ortholog in *S. meliloti* is involved in some additional cellular functions, other than nickel tolerance, the Phenotype Microarray (PM) system was used on PM9, PM10 and PM13B plates, for a total of 288 different growth conditions tested, including tolerances to 96 different osmolytes and 96 different pH conditions and resistance to several compounds. (see Fig. S1 in the supplemental material for a complete set of results). Metabolically active conditions were scored using a threshold calculation based on growth curve data (see Materials and Methods).

Within these 288 different conditions, five resulted different between BM623 and BM589 (Δ*nreB*) (Fig. 4).

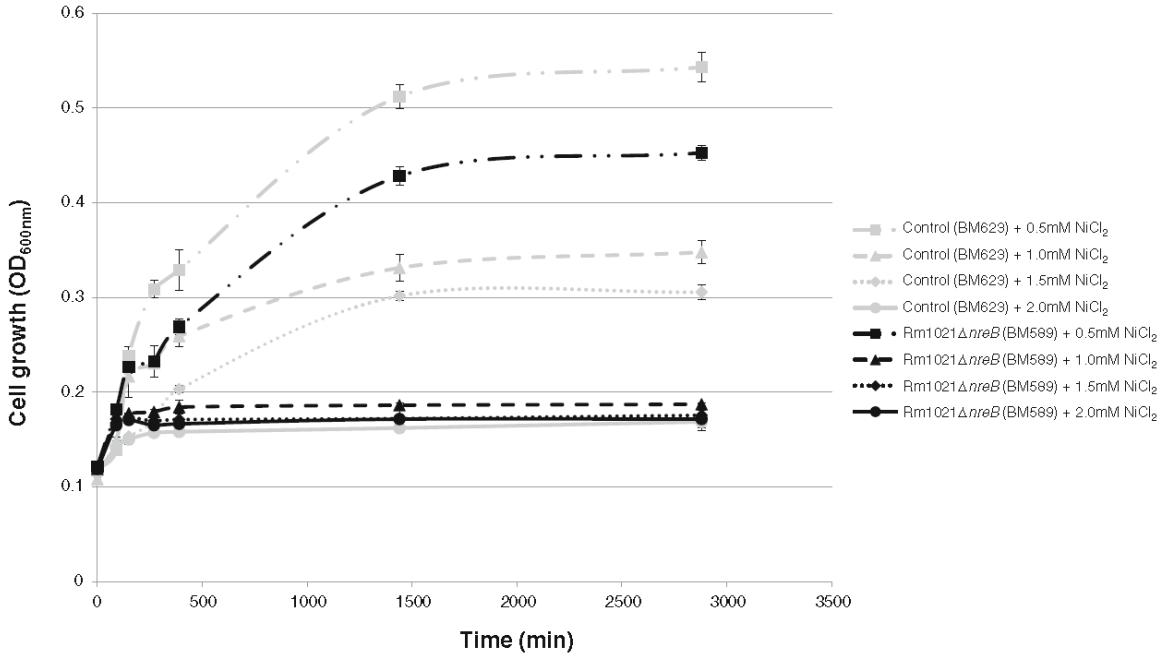


Fig. 3 Growth curves of BM623 (Rm1021 pMR20) (*grey line*) and BM589 (Rm1021Δ*nreB*) (*black line*) in presence of various concentrations of $NiCl_2$: 0.5 mM $NiCl_2$ *dash-dot-dot lines with squares*, 1.0 mM $NiCl_2$ *dashed lines with triangles*, 1.5 mM $NiCl_2$ *dotted lines with diamonds* and 2.0 mM of $NiCl_2$ *solid lines with circles*

Strikingly, nickel tolerance was only slightly affected. This result is in agreement with the presence of other genes, such as for instance SMa1163 (Zielazinski et al. 2013) and *dmeRF* (Rubio-Sanz et al. 2013) involved innickel homeostasis in *S. meliloti*. Interestingly, the mutant strain showed to be more sensitive to copper chloride and more tolerant to urea osmotic effects. Moreover, the mutant displayed higher activity than the control strain at low pH (4.5) in presence of L-threonine. The reduced tolerance to $CuCl_2$ (Fig. 4) was indeed confirmed also by a growth curve in liquid TY medium in presence of various concentrations of $CuCl_2$ (Fig. S2). Additionally, the deletion mutant showed an enhanced resistance to urea osmotic effects (Fig. 4). A test of urease activity was consequently performed on bacterial cultures of the two strains (Δ*nreB* and Rm1021), showing, in agreement with the above-mentioned hypothesis, a significant increase of urease catalytic properties (K_a, $p<0,05$) under culture conditions for the mutant (V_{max} 1,405±0,082; K_m 141,9±18,67; K_a*10^3 10,07±0,67) compared to wild type strain (V_{max} 1,412±0,187; K_m 227,5±58,08; K_a*10^3 6,6±0,78) as shown in Table 3. Interestingly, no difference was detected for *ureC* transcript by quantitative PCR (data not shown), indicating that post-transcriptional events (e.g. Ni^{2+} availability) are possibly the cause of the increased urease activity.

Finally, because SMa1641 (*nreB*) is located on pSymA megaplasmid, which harbors several functions related to symbiosis functions (e.g. nodulation and nitrogen fixation genes) and an ortholog is present in the beta-rhizobium *C. taiwanensis* (Klonowska et al. 2012) also, we investigated the symbiotic phenotype of BM589 (Δ*nreB*) compared to the wild type strain Rm1021, by using alfalfa (*Medicago sativa*) as host plant. Results showed (Fig. 5) that the mutant strain is more efficient in establishing the symbiosis than the wild type strain, considering the increase in plant growth as parameter (Δ*nreB*=8.49±3.49 cm, Rm1021=5.79± 2.45 cm), determining a 1.7 fold increase in shoot length.

Discussion

The gene *nreB* encodes for a Ni^{2+}/H^+ antiporter (hydrogenase). This gene was initially discovered in *C. metallidurans* 31A, where the function of this antiporter

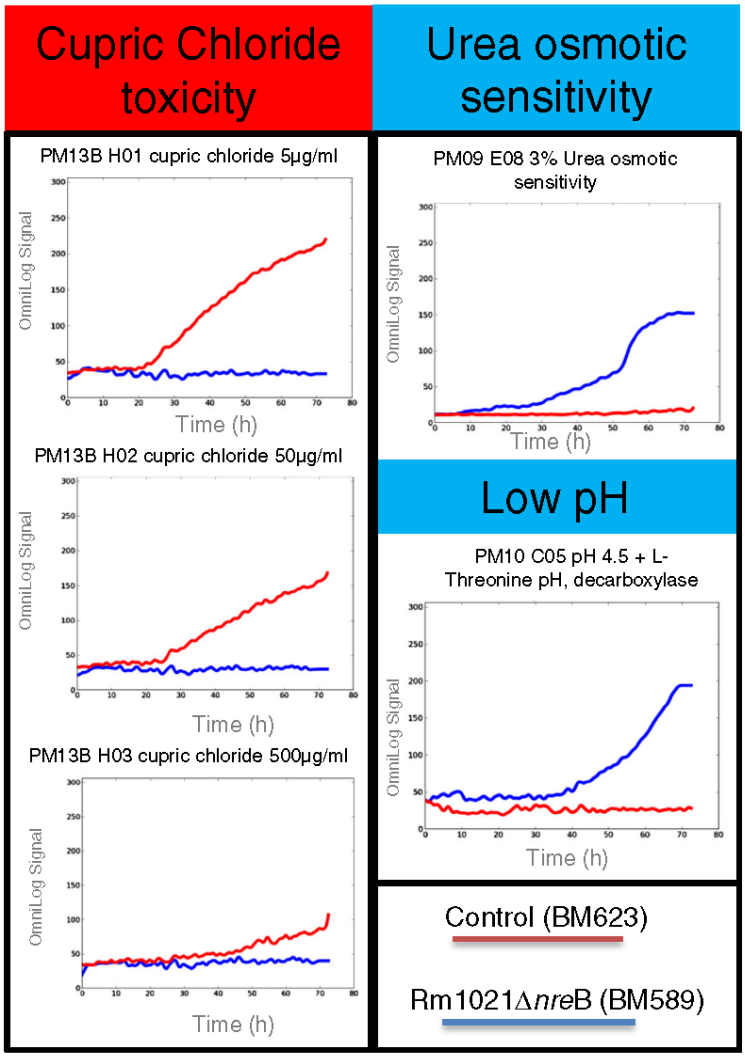


Fig. 4 Phenotype Microarray results comparing metabolic abilities of BM589 (Rm1021 Δ*nreB*) and BM623 (Rm1021 pMR20) strains with respect to heavy-metals pH and osmolytes

was related to the nickel resistance displayed by such bacterial strain (Grass et al. 2001). Recently, *nreB* orthologs have been found in plant-symbiotic bacteria of the group of *Bradyrhizobium* isolated from leguminous plants growing in the serpentine outcrops of New Caledonia (Chaintreuil et al. 2007). However, for these strains the presence of *nreB* gene seemed to be not correlated with nickel resistance, opening the question over the evolutionary pathways of *nreB* gene in bacterial taxa and of its possible functions, unrelated to nickel tolerance. To answer these questions we performed a detailed phylogenetic analysis of NreB orthologs present in protein databases and of the genomic context of their corresponding genes in completely sequenced bacteria. Moreover, we used a nickel sensitive model strain (*S. meliloti* Rm1021), which presents *nreB* gene in its core genome, to assign a putative role of NreB in nickel-sensitive species. Obtained phylogenetic results strongly supported the hypothesis that *nreB* underwent several past events of horizontal gene transfer (HGT), as it has been documented for other metal-tolerance systems

Table 3 Urease activity of Rm1021 and Rm1021Δ*nreB*. Values indicate the V_{max}, the K_m and the K_a

	Rm1021Δ*nreB*	Rm1021
V_{max}	1,405±0,082	1,412±0,187
K_m	141,90±18,67	227,50±58,08
K_a	10,07±0,67	6,60±0,78

Plant Soil (2014) 377:189–201

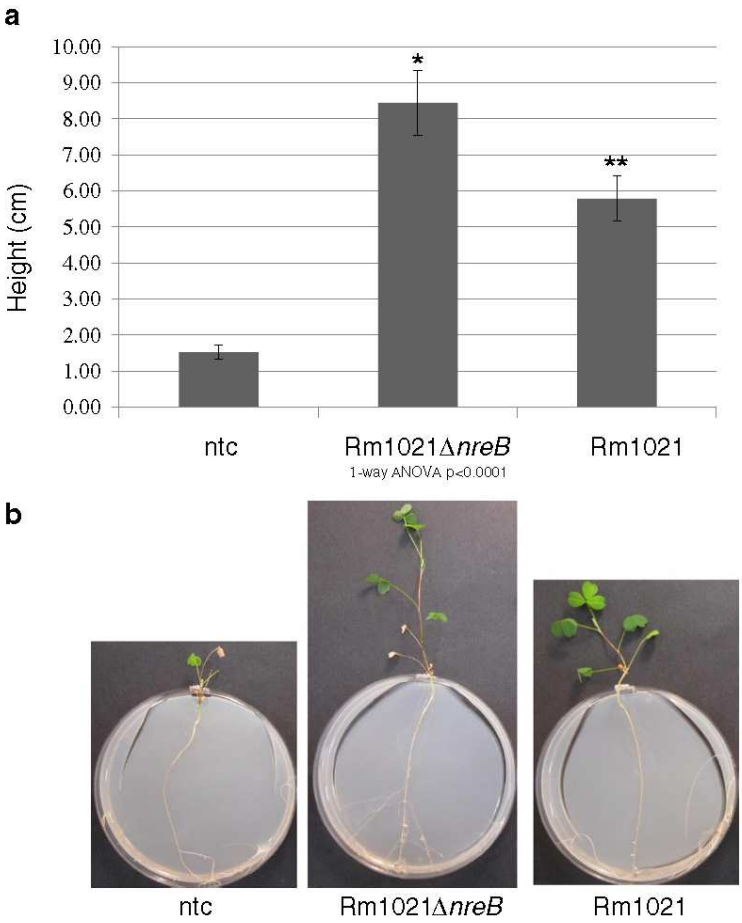


Fig. 5 Symbiotic phenotypes of Rm1021Δ*nreB* and control (Rm1021) strains. **a** Mean length of aerial parts of plants inoculated with Rm1021 and the deletion mutant (Δ*nreB*) is reported. *Error bars* indicate standard deviation from 13 plants. Different *asterisks* indicate significant different means after 1-way ANOVA ($p<0.0001$). ntc, uninoculated control plants (negative control), **b** images of plants in the different conditions

too. One of the most recent examples is the *mer* operon, conferring resistance to mercury, in *Acinetobacter* species (Fondi et al. 2010), where a movement of the operon in different plasmid backgrounds was detected. Other recent examples are those of ATPases, as P_{IB} ATP-ases (Nongkhlaw et al. 2012), which are known to transport heavy metals such as Cu, Zn, Pb and Cd in bacteria. These ATPases are encoded mostly chromosomally, but the presence of PIB-type ATPase genes on mobile genetic elements (i.e. plasmids and transposons) was detected also (see for instance Mergeay et al. 2003). It is still an open question the evolutionary pathways of HGT of *nreB*, since no clear relationships with the environmental distribution of strains were detected, then hampering possible speculation on the role of ubiquitary bacteria as "drivers" of HGT, as it was recently suggested for water environment in relation to antibiotic resistance (Fondi and Fani 2010). However, it should be noticed that most of the 12 plant-associated bacteria reported in Fig. 2b, are also ubiquitous to other environment (soil or water). Anyway, the presence of *nreB* in the genome of all *S. meliloti* tested and the CAI similar to the rest of the genome strongly suggest that the introgression of *nreB* into the *S. meliloti* core genome may predate the origin of *Sinorhizobium* genus, suggesting an involvement of NreB in some relatively basic cellular functions. Here, results obtained with the mutant strain have shown that indeed *nreB* gene is involved in intriguing cellular phenotypes. In particular an effect on copper tolerance was observed, since the *nreB* deletion mutant showed reduced tolerance to $CuCl_2$ which could imply the possibility that copper also could be exported through the cell by NreB of *S. meliloti*. However, it should be mentioned that to the best of our knowledge no data have previously been reported on an association of *nreB* with copper tolerance, deserving future molecular investigation on the interaction of Cu^{2+} with NreB and on NreB involvement

in copper homeostasis. Interestingly, in *R. pickettii* 12D *nreB* is located on a plasmid (pRp12D02) enriched in heavy metal tolerance genes (19.6 % of the total genes). In particular 66.6 % of these genes (36 out of 54) are related to copper homeostasis, supporting the hypothesis of additional roles for *nreB* genes in relation to trace metals tolerance. Most of the other phenotypes observed in the deletion mutant (as tolerance to urea and plant symbiotic performance) were mainly possibly related to an increase Ni^{2+} availability, since Ni(II) is a cofactor of urease and hydrogenases . Actually, NreB function as Ni^{2+}/H^{+} antiporter could be linked to tolerance to high urea concentrations, through the function of the enzyme urease, which consists of an apoenzyme containing a Ni^{2+} ion as cofactor (Mobley et al. 1995). We can consequently hypothesize that the deletion of *nreB* could produce an increased cytoplasmic Ni^{2+} concentration, then bringing to higher urease activity, which, in turn, allows an increased urea degradation (then a higher urea tolerance) and an increased production of NH_3. The higher ammonia levels could then result in a more efficient transfer of ammonia to plants, then increasing plant growth promoting abilities of the *nreB* mutant strain with respect to the wild type strain. Moreover, similarly to the detected increased urease activity, a possible higher activity of [NiFe] hydrogenases could be present. Increased hydrogenase expression in rhizobia (*Rhizobium leguminosarum* biovar *viciae*) has in fact been correlated with increased symbiotic efficiency and promotion of plant growth (Ureta et al. 2005).

In conclusion, *nreB* could have evolved as player for nickel-tolerance in heavy-metal tolerant bacterial strains by exaptation of the preexisting NreB functions, possibly linked to homeostasis of urea, hydrogenase activities and, maybe, Cu(II) tolerance. However, biochemical studies on NreB of *S. meliloti* in comparison with the ortholog of *C. metallidurans* CH34 are needed to determine the specificity of metal transport and possible functional differences among *nreB* orthologs.

References

Amadou C, Pascal G, Mangenot S, Glew M, Bontemps C, Capela D, Carrere S, Cruveiller S, Dossat C, Lajus A, Marchetti M, Poinsot V, Rouy Z, Servin B, Saad M, Schenowitz C, Barbe V, Batut J, Medigue C, Masson-Boivin C (2008) Genome sequence of the beta-rhizobium Cupriavidus taiwanensis and comparative genomics of rhizobia. Genome Res 18: 1472–1483

Beringer JE (1974) R factor transfer in Rhizobium leguminosarum. J Gen Microbiol 84:188–198

Bernal A, Ear U, Kyrpides N (2001) Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucleic Acids Res 29:126–127

Castillo-Ramirez S, Vazquez-Castellanos JF, Gonzalez V, Cevallos MA (2009) Horizontal gene transfer and diverse functional constrains within a common replication-partitioning system in Alphaproteobacteria: the repABC operon. BMC Genomics 10:536

Chaintreuil C, Rigault F, Moulin L, Jaffre T, Fardoux J, Giraud E, Dreyfus B, Bailly X (2007) Nickel resistance determinants in bradyrhizobium strains from nodules of the endemic New Caledonia legume Serianthes calycina. Appl Environ Microbiol 73:8018–8022

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36: W465–W469

Dereeper A, Audic S, Claverie JM, Blanc G (2010) Blast-explorer helps you building datasets for phylogenetic analysis. BMC Evol Biol 10:8

Ehrhardt DW, Atkinson EM, Long SR (1992) Depolarization of alfalfa root hair membrane potential by Rhizobium meliloti Nod factors. Science 256:998–1000

Ferri L, Gori A, Biondi EG, Mengoni A, Bazzicalupo M (2010) Plasmid electroporation of Sinorhizobium strains: the role of the restriction gene hsdR in type strain Rm1021. Plasmid 63: 128–135

Fondi M, Fani R (2010) The horizontal flow of the plasmid resistome: clues from inter-generic similarity networks. Environ Microbiol 12:3228–3242

Fondi M, Bacci G, Brilli M, Papaleo MC, Mengoni A, Vaneechoutte M, Dijkshoorn L, Fani R (2010) Exploring the evolutionary dynamics of plasmids: the Acinetobacter pan-plasmidome. BMC Evol Biol 10

Galardini M, Mengoni A, Brilli M, Pini F, Fioravanti A, Lucas S, Lapidus A, Cheng JF, Goodwin L, Pitluck S, Land M, Hauser L, Woyke T, Mikhailova N, Ivanova N, Daligault H, Bruce D, Detter C, Tapia R, Han C, Teshima H, Mocali S, Bazzicalupo M, Biondi EG (2011) Exploring the symbiotic pangenome of the nitrogen-fixing bacterium Sinorhizobium meliloti. BMC Genomics 12:235

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J (2001) The composite genome of the legume symbiont Sinorhizobium meliloti. Science 293:668–672

Grass G, Fan B, Rosen BP, Lemke K, Schlegel HG, Rensing C (2001) NreB from *Achromobacter xylosoxidans* 31A is a nickel-induced transporter conferring nickel resistance. J Bacteriol 183:2803–2807

Hanahan D (1983) Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166:557–580

Hou W, Ma Z, Sun L, Han M, Lu J, Li Z, Mohamad OA, Wei G (2013) Extracellular polymeric substances from copper-tolerance *Sinorhizobium meliloti* immobilize Cu. J Hazard Mater 261C:614–620

Kandeler E, Gerber H (1988) Short-term assay of soil urease activity using colorimetric determination of ammonium. Biol Fertil Soils 6:68–72

Klonowska A, Chaintreuil C, Tisseyre P, Miche L, Melkonian R, Ducousso M, Laguerre G, Brunel B, Moulin L (2012) Biodiversity of Mimosa pudica rhizobial symbionts (*Cupriavidus taiwanensis*, *Rhizobium mesoamericanum*) in New Caledonia and their adaptation to heavy metal-rich soils. FEMS Microbiol Ecol 81:618–635

Law CJ, Maloney PC, Wang DN (2008) Ins and outs of major facilitator superfamily antiporters. Annu Rev Microbiol 62: 289–305

Marrero J, Auling G, Coto O, Nies DH (2007) High-level resistance to cobalt and nickel but probably no transenvelope efflux: metal resistance in the cuban Serratia marcescens strain C-1. Microb Ecol 53:123–133

Mengoni A, Schat H, Vangronsveld J (2010) Plants as extreme environments? Ni-resistant bacteria and Ni-hyperaccumulators of serpentine flora. Plant Soil 331:5–16

Mergeay M, Monchy S, Vallaeys T, Auquier V, Benotmane A, Bertin F, Taghavi S, Dunn J, van der Lelie D, Wattiez R (2003) *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. FEMS Microbiol Rev 27:385–410

Mobley HL, Island MD, Hausinger RP (1995) Molecular biology of microbial ureases. Microbiol Rev 59:451–480

Moscatelli MC, Lagomarsino A, Garzillo AMV, Pignataro A, Grego S (2012) β-Glucosidase kinetic parameters as indicators of soil quality under conventional and organic cropping systems applying two analytical approaches. Ecol Indic 13:322–327

Nongkhlaw M, Kumar R, Acharya C, Joshi SR (2012) Occurrence of horizontal gene transfer of P(IB)-type ATPase genes among bacteria isolated from the uranium rich deposit of Domiasiat in North East India. PLoS One 7:e48199

Park JE, Schlegel HG, Rhie HG, Lee HS (2004) Nucleotide sequence and expression of the ncr nickel and cobalt resistance in Hafnia alvei 5–5. Int Microbiol 7:27–34

Park JS, Lee SJ, Rhie HG, Lee HS (2008) Characterization of a chromosomal nickel resistance determinant from *Klebsiella oxytoca* CCUG 15788. J Microbiol Biotechnol 18:1040–1043

Peleg AY, de Breij A, Adams MD, Cerqueira GM, Mocali S, Galardini M, Nibbering PH, Earl AM, Ward DV, Paterson DL, Seifert H, Dijkshoorn L (2012) The success of acinetobacter species; genetic, metabolic and virulence attributes. PLoS One 7:e46984

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30:e36

Pini F, Frage B, Ferri L, De Nisco NJ, Mohapatra SS, Taddei L, Fioravanti A, Dewitte F, Galardini M, Brilli M, Villeret V, Bazzicalupo M, Mengoni A, Walker GC, Becker A, Biondi EG (2013) The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. Mol Microbiol Jul 30

Puigbo P, Bravo IG, Garcia-Vallve S (2008) E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). BMC Bioinforma 9:65

Roberts RC, Toochinda C, Avedissian M, Baldini RL, Gomes SL, Shapiro L (1996) Identification of a *Caulobacter crescentus* operon encoding *hrcA*, involved in negatively regulating heat-inducible transcription, and the chaperone gene *grpE*. J Bacteriol 178:1829–1841

Rubio-Sanz L, Prieto RI, Imperial J, Palacios JM, Brito B (2013) Functional and expression analysis of the metal-inducible dmeRF System from *Rhizobium leguminosarum* bv. *viciae*. Appl Environ Microbiol 79:6414–6422

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor

Simon R, Priefer U, Pühler A (1983) A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram-negative bacteria. Bio/Technology 1: 784–791

Skerker JM, Laub MT (2004) Cell-cycle progression and the generation of asymmetry in Caulobacter crescentus. Nat Rev Microbiol 2:325–337

Stoppel RD, Meyer M, Schlegel HG (1995) The nickel resistance determinant cloned from the enterobacterium Klebsiella oxytoca: conjugational transfer, expression, regulation and DNA homologies to various nickel-resistant bacteria. Biometals 8:70–79

Ureta AC, Imperial J, Ruiz-Argueso T, Palacios JM (2005) Rhizobium leguminosarum biovar viciae symbiotic hydrogenase activity and processing are limited by the level of nickel in agricultural soils. Appl Environ Microbiol 71: 7603–7606

Vanucci S, Minerdi D, Kadomatsu K, Mengoni A, Bazzicalupo M (2005) Putative midkine family protein up-regulation in Patella caerulea (Mollusca, Gastropoda) exposed to sublethal concentrations of cadmium. Aquat Toxicol (Amsterdam, Netherlands) 75:374–379

Viti C, Decorosi F, Tatti E, Giovannetti L (2007) Characterization of chromate-resistant and -reducing bacteria by traditional means and by a high-throughput phenomic technique for bioremediation purposes. Biotechnol Prog 23:553–559

Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev 35:957–976

Zielazinski EL, Gonzalez-Guerrero M, Subramanian P, Stemmler TL, Arguello JM, Rosenzweig AC (2013) *Sinorhizobium meliloti* Nia is a P-ATPase expressed in the nodule during plant symbiosis and is involved in Ni and Fe transport. Metallomics. doi:10.1039/C3MT00195D

ADDITIONAL FILES

Additional file, Table S1:

Collection of *S. meliloti* strains over which *nreB* gene has been detected.

Additional file, Fig. S1:

Results of Phenotype Microarray: a) PM09, b) PM10 and c) PM13B. Rm1021 pMR20 (red line) Rm1021Δ*nreB* (blue line). Conditions present in each well are reported on the following websites: http://www.biolog.com/pdf/pm_lit/PM1-PM10.pdf (for PM09 and PM 10); http://www.biolog.com/pdf/pm_lit/PM11-PM20.pdf (for PM13B).

Additional file, Fig. S2:

Sensitivity to $CuCl_2$ of BM623 (Rm1021 pMR20) (grey line) and BM589 (Rm1021Δ*nreB*) (black line). 0.5 mM $NiCl_2$dash-dot-dot lines with squares, 1.0 mM $NiCl_2$ dashed lines with triangles, 1.5 mM $NiCl_2$ dotted lines with diamonds and 2.0 mM of $NiCl_2$ solid lines with circles.

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Evolution of Intra-specific Regulatory Networks in a Multipartite Bacterial Genome

Marco Galardini[1¤], Matteo Brilli[2], Giulia Spini[3], Matteo Rossi[1], Bianca Roncaglia[1], Alessia Bani[1], Manuela Chiancianesi[1], Marco Moretto[4], Kristof Engelen[4], Giovanni Bacci[1,5], Francesco Pini[6], Emanuele G. Biondi[6], Marco Bazzicalupo[1], Alessio Mengoni[1]*

1 Department of Biology, University of Florence, Florence, Italy, **2** Department of Genomics and Biology of Fruit Crops, Research and Innovation Centre, Fondazione Edmund Mach (FEM), San Michele all'Adige, Italy, **3** Dipartimento di Biotecnologie Agrarie, Sezione di Microbiologia, University of Florence, Florence, Italy, **4** Department of Computational Biology, Research and Innovation Centre, Fondazione Edmund Mach (FEM), San Michele all'Adige, Italy, **5** Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per lo Studio delle Relazioni tra Pianta e Suolo (CRA-RPS), Rome, Italy, **6** Interdisciplinary Research Institute USR3078, CNRS-Universit Lille Nord de France, Villeneuve d'Ascq, France

¤ Current address: EMBL-EBI, Wellcome Trust Genome Campus, Cambridge, United Kingdom
* alessio.mengoni@unifi.it

OPEN ACCESS

Citation: Galardini M, Brilli M, Spini G, Rossi M, Roncaglia B, Bani A, et al. (2015) Evolution of Intra-specific Regulatory Networks in a Multipartite Bacterial Genome. PLoS Comput Biol 11(9): e1004478. doi:10.1371/journal.pcbi.1004478

Editor: Feilim Mac Gabhann, Johns Hopkins University, UNITED STATES

Received: November 24, 2014

Accepted: July 24, 2015

Published: September 4, 2015

Data Availability Statement: Methods are available as a git repository (regtools: https://github.com/combogenomics/regtools/tree/paper). Data (input genomes, regulators PSSM and predictions) are available as a git repository (rhizoreg: https://github.com/combogenomics/rhizoreg/tree/paper).

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Reconstruction of the regulatory network is an important step in understanding how organisms control the expression of gene products and therefore phenotypes. Recent studies have pointed out the importance of regulatory network plasticity in bacterial adaptation and evolution. The evolution of such networks within and outside the species boundary is however still obscure. *Sinorhizobium meliloti* is an ideal species for such study, having three large replicons, many genomes available and a significant knowledge of its transcription factors (TF). Each replicon has a specific functional and evolutionary mark; which might also emerge from the analysis of their regulatory signatures. Here we have studied the plasticity of the regulatory network within and outside the *S. meliloti* species, looking for the presence of 41 TFs binding motifs in 51 strains and 5 related rhizobial species. We have detected a preference of several TFs for one of the three replicons, and the function of regulated genes was found to be in accordance with the overall replicon functional signature: house-keeping functions for the chromosome, metabolism for the chromid, symbiosis for the megaplasmid. This therefore suggests a replicon-specific wiring of the regulatory network in the *S. meliloti* species. At the same time a significant part of the predicted regulatory network is shared between the chromosome and the chromid, thus adding an additional layer by which the chromid integrates itself in the core genome. Furthermore, the regulatory network distance was found to be correlated with both promoter regions and accessory genome evolution inside the species, indicating that both pangenome compartments are involved in the regulatory network evolution. We also observed that genes which are not included in the species regulatory network are more likely to belong to the accessory genome, indicating that regulatory interactions should also be considered to predict gene conservation in bacterial pangenomes.

Author Summary

The influence of transcriptional regulatory networks on the evolution of bacterial pangenomes has not yet been elucidated, even though the role of transcriptional regulation is widely recognized. Using the model symbiont *Sinorhizobium meliloti* we have predicted the regulatory targets of 41 transcription factors in 51 strains and 5 other rhizobial species, showing a correlation between regulon diversity and pangenome evolution, through upstream sequence diversity and accessory genome composition. We have also shown that genes not wired to the regulatory network are more likely to belong to the accessory genome, thus suggesting that inclusion in the regulatory circuits may be an indicator of gene conservation. We have also highlighted a series of transcription factors that preferentially regulate genes belonging to one of the three replicons of this species, indicating the presence of replicon-specific regulatory modules, with peculiar functional signatures. At the same time the chromid shares a significant part of the regulatory network with the chromosome, indicating an additional way by which this replicon integrates itself in the pangenome.

Introduction

Regulation of gene expression is recognized as a key component in the cellular response to the environment. This is especially true in the microbial world, for two reasons: bacterial cells are often under severe energy constraints, the most important being protein translation [1] and they usually face a vast range of environmental and physiological conditions; being able to efficiently and readily react to ever changing conditions can most certainly give a selective advantage over competitors and give rise to specific regulatory networks.

Transcription is mainly regulated by proteins, called transcription factors (TF), which usually contain a protein domain capable of binding to specific DNA sequences, called TF binding sites (TFBS). Depending on the position of the TFBS with respect to the transcriptional start site of the regulated gene, the TF can act either as a transcriptional activator or a repressor, mostly because of its interaction with the RNA polymerase and sigma factors [2, 3]. The binding of the TF to its cognate TFBS is based on non-covalent interactions whose strength is indicated by the so-called affinity constant. Since TFBS can have variations around a preferred sequence, the affinity of a TF for its TFBSs covers a continuous range of values; however, since the TF binding strength appears to follow a sigmoid behaviour, it is possible to distinguish between 'weak' and 'strong' TFBSs [4].

As opposed to eukaryotic species, prokaryotic TFBSs are usually distinguishable from the 'background DNA', and they tend to have a simpler structure and a close proximity to the transcription start site [5]. The application of information theory concepts to TFBS identification and analysis, revealed that specificity of the TF for a certain TFBS depends on the length, variability and composition of the TFBS itself with respect to the overall genomic background (i.e. the sequence composition). Intuitively, the minimum information content able to provide specific recognition of the TFBS by the TF mostly depends on the genome size and its composition; increasing the size of the genome clearly increases the number of putatively non-functional TFBSs, and when the TFBS bases composition is close to the background DNA composition it may be impossible to discern a true functional TFBS from the surrounding DNA. Transcription factors recognizing TFBS characterized by low information content usually control the transcription of many genes across the genome; alternative sigma factors usually belong to this class, and their TFBSs also show larger variability between species [5]. Gene targets of these TFs are harder to reliably predict, for the presence of many non-functional sites along the

genome. The high gene density of bacterial genomes and its organization in operons results in specific expression or repression of whole functional pathways in response to stimuli. Furthermore, the presence of several TFBSs in the upstream region of a gene can result in a complex transcriptional response that recall the behaviour of logic gates [6].

Prediction of TFBSs in a genome usually relies on the availability of a position specific scoring matrix (PSSM) storing the frequency of each nucleotide at each position of a TFBS. PSSM modelling the variability of a TFBS can be built by identifying enriched DNA patterns in promoter regions of genes that are known to be under the control of the TF under analysis, better if guided by other assays, like the binding of the TF to synthetic nucleotides. Several algorithms have been developed to use such PSSM to search for TFBSs in nucleotide sequences, such as the MEME suite [7], RSAT [8–10] and the Bio.motif package [11]. A recent alternative method relies on the construction of a hidden markov model (HMM) from an alignment of nucleotide sequences, which can then be used to scan a query nucleotide sequence [12–14]. Since all these methods and their implementations have different weaknesses, it has been advised to use their combination to run predictions [15].

Regulatory networks evolve rapidly, making the comparisons between distant organisms difficult [16–19]. At broad phylogenetic distances, it has been shown that the conservation of a TF is lower than its targets [16]. Additionally, species with similar lifestyles tend to show conservation of regulatory network motifs, despite significant variability in the gene composition of the network, suggesting an evolutionary pressure towards the emergence of certain regulatory logics [16].

The fluidity of most transcriptional regulatory connections is well known and documented, not only at large phylogenetic distances, but also at the level of intra-species comparisons too [20–23]. Experiments have shown that Bacteria have high tolerance towards changes in the regulatory circuitry, making them potentially able to exploit even radical changes to the regulatory network, without extensive changes in phenotypes [24]. However, this is strongly dependent on which regulatory interaction undergoes changes, since there are also examples where a single change determines an observable difference in phenotype [25, 26]. Bacteria have therefore a mixture of robust and fragile edges in their regulatory networks and evolution can play with them at different extent to explore: i) the function of new genes, by integrating them in the old gene regulatory network, and ii) if genes that are part of the gene regulatory network can be removed without harm to the physiology of the cell. The extent of variability and evolution of the regulatory network inside a species is, however, still poorly understood.

The aim of this study is a comparative genomics analysis of regulatory networks, to understand the impact of regulatory network variability on pangenome evolution. We decided to use the *Sinorhizobium meliloti* species, the nitrogen-fixing symbiont of plants from the genus *Medicago*. *S. meliloti* has been deeply investigated as a model for symbiotic interaction and an extensive knowledge on its TFs is present in the literature [27, 28]. This species presents a marked genomic difference with respect to other well-know bacterial model species, such as *Escherichia coli*, since *S. meliloti* genome comprises three replicons of comparable size: a chromosome, a chromid [29] and a megaplasmid, characterized by functionally and evolutionary distinct signatures [30, 31]. This arrangement raises the question of how TF targets are distributed over the replicons. Recent reports have shown that there are only two genes essential for growth in minimal media and soil encoded in the *S. meliloti* chromid [32], even though the chromid harbours many genes shared by all sequenced strains of *S. meliloti* species. Moreover, *S. meliloti* has several genomes sequenced to date [23, 30, 33–39] and the potential for biotechnological and agricultural applications, which could benefit from this analysis. At the comparative genomics level, different strains show quite a high level of variation. Indeed, the pangenome (the collection of all genes from different strains [40]) of this species has an abundant fraction of genes common to all

members of the species (termed core genome, as opposed to the strain-exclusive and/or partially shared fraction, called accessory genome) of around 5000 gene families; approximately 40% of the genome belongs to the accessory fraction [31, 35]. A preliminary analysis revealed that some of the TFs of the core genome also control genes of the accessory genome [23]. This allowed to propose that, when comparing the same regulon in different strains, we can define a *panregulon*, including a set of core (shared) target genes and an accessory (variable) regulon fraction [23]. It should be noticed that while the core regulon is necessarily formed by genes belonging to the core genome, the opposite can also be true (i.e. that a gene belonging to the core genome belongs to the accessory regulon). However, the dynamics of the panregulon in relation to the evolutionary rules controlling the variability of the accessory regulon fraction are still not understood.

We have therefore constructed the regulatory network of the *S. meliloti* species, using the PSSMs of 41 TFs collected from the literature and public databases. We have applied a combination of TFBS prediction methods, combining their output with information about the core and accessory gene families. We have also predicted the presence of the same TFBSs in five other closely related rhizobial species (termed 'outgroups': *Rhizobium leguminosarum* bv. *viciae*, *Rhizobium etli*, *Mesorhizobium loti*, *Sinorhizobium fredii* and *Sinorhizobium medicae*). This regulatory network has been used to highlight the different behaviours that are present within and between species. Our predictions and other comparative genomics observations are publicly available (https://github.com/combogenomics/rhizoreg/).

Results

General features of the predicted regulatory network of *S. meliloti*

Based on COG annotations, all the 51 *S. meliloti* strains analysed in this study, have been found to encode a similar number of predicted TFs (an average of 522); a similar number has been also found in the five outgroups (an average of 533). This is in accordance with previous reports correlating genome size with the number of TFs [41]. Rhizobia belonging to the *Alphaproteobacteria* class (alpha-rhizobia), which are known to have larger genomes compared to other bacteria from the same class [42], have then one of the largest collection of TFs in the known bacterial kingdom. As the accessory genome accounts for about 40% of the proteome size [31, 35], it is reasonable to expect that a similar proportion of TFs will belong to the accessory genome. Indeed, about 70% of the TFs encoded in the *S. meliloti* pangenome belong to the core genome, while the remaining TFs are present in 1–3 genomes only; this orthologous genes distribution is similar to the one observed for the whole pangenome [43] (S1 Fig). However, most of the 41 TFs analyzed in this study were found to belong to the core genome (37), with the only notable exception represented by RhrA, the activator of the rhizobactin regulon, which is absent in 35% of the strains under study, confirming previous analysis [23, 44, 45]. More interestingly, recent reports have demonstrated how the presence of the rhizobactin operon confers competitive advantage over other *S. meliloti* strains in iron limited environments [32]; we could therefore speculate that a significant fraction of the *S. meliloti* strains have a competitive disadvantage in environments with limitation in iron bioavailability. Surprisingly, an ortholog of FixJ (the component of the global two-component system FixJL, which turns on nitrogen-fixation genes in microaerobiosis during symbiosis) was not predicted in two *S. meliloti* strains (A0643DD and C0438LL); the absence of the gene was further confirmed by PCR. Even though such an important regulator has been found to be absent in these two strains, another gene with similar domains (orthologous group SinMel7252, containing gene SMa1686 from the reference strain Rm1021) was found to belong to the core genome. SMa1686 was shown to be regulated by RirA [46], but to the best of our knowledge no indications of its relationships with microaerophilic growth conditions and symbiosis are present.

Consequently, we cannot a priori exclude that the regulatory functions of FixJ may be carried on by homologs (as for instance orthologs of SMa1686) in strains A0643DD and C0438LL. Indeed, previous works have indicated that several target genes of FixJ lack a direct symbiotic function, suggesting the presence of functional redundancy in the genome [47].

Sixteen TFs were absent in at least one of the outgroups. Of these, 6 are encoded by pSymA, the symbiotic megaplasmid, including two copies of NodD, FixJ, RctR, SyrM and RhrA (S1 Fig). Such difference between intraspecific and interspecific TF gene content may anticipate a similar difference at the downstream regulatory network, for the absence of cross-regulatory links.

To minimize the number of false positives in our predictions, we selected PSSMs with relatively high information content (over the reference strain minimum information content, see Materials and Methods) A wide range of information gain for PSSMs was observed; of the starting 83 TFBSs retrieved from literature and databases, 41 have been found to have enough information content to reliably predict their TFBSs (Fig 1a, S1 Table). For FixJ, two separate motifs acting together have been described [48], one above and one slightly below the threshold: both motifs have been used.

We have applied a novel TFBS prediction approach to overcome common problems associated with the prediction algorithms and to maximize accuracy and sensitivity [3], including operon predictions to recover most of the downstream regulated genes (see Materials and methods). The predictions accuracy was determined with a comparison with the downstream regulons reported in the literature, when available (Fig 1b and 1c); the average accuracy of the predictions was found to be around 55%, with a tendency to positively correlate with the motif information gain (S2 Fig). This behaviour may be explained by the fact that most regulons have been defined on the basis of gene expression data and therefore contain both direct and indirect targets of the TF; our strategy is then not able to recover the indirect targets which might explain the relatively low accuracy. An example of a known regulatory interaction predicted by our approach is rem (SMc03046), a putative transcriptional regulator involved in the control of motility in *S. meliloti* Rm1021 [49], which was predicted to be under the control of MucR in our analysis (S1 Material).

To provide additional validation to our predictions, we used a compendium of *S. meliloti* gene expression data from the Colombos database [50] (see Materials and Methods). The full compendium contained 424 conditions and was used to calculate average correlation coefficients among the genes of i) the same predicted regulons, ii) the regulons reported in the literature and iii) random groups of genes sampled from the genome (Fig 1d and S2 Material). We have selected the conditions maximising the average correlation for a group of genes using a genetic algorithm (see Materials and Methods). Correlations for our predictions were not significantly different from the experimentally defined regulons; genes belonging to predicted regulons had a slight tendency to be higher than the random regulons, but if this difference was not significant ($p = 0.09$). We further experimentally confirmed some of the predictions on a subset of predicted promoters of the NodD regulon (S2 Table).

Predicted TFBSs in upstream regions against TFBSs predicted in coding regions were considered as signal to noise ratio (upstream hits on total hits) to measure the predictions quality (Fig 1e); for more than 70% of the analysed TF the observed ratio was above 50%, with a very poor correlation with the motif information content.

Taken together these results show that our predictions are of fairly good quality.

Little variability in the number of genes under the control of each TF was observed among different strains (Fig 2 and Table 1). Each TF was predicted to control the transcription of 12 genes on average, with RirA showing the largest regulon (with an average of 71.6 genes) and SyrM the smallest one (with an average of 1.1 genes). TFs with lower information content TFBSs showed a tendency to control a larger number of genes (S2 Fig), which confirms the

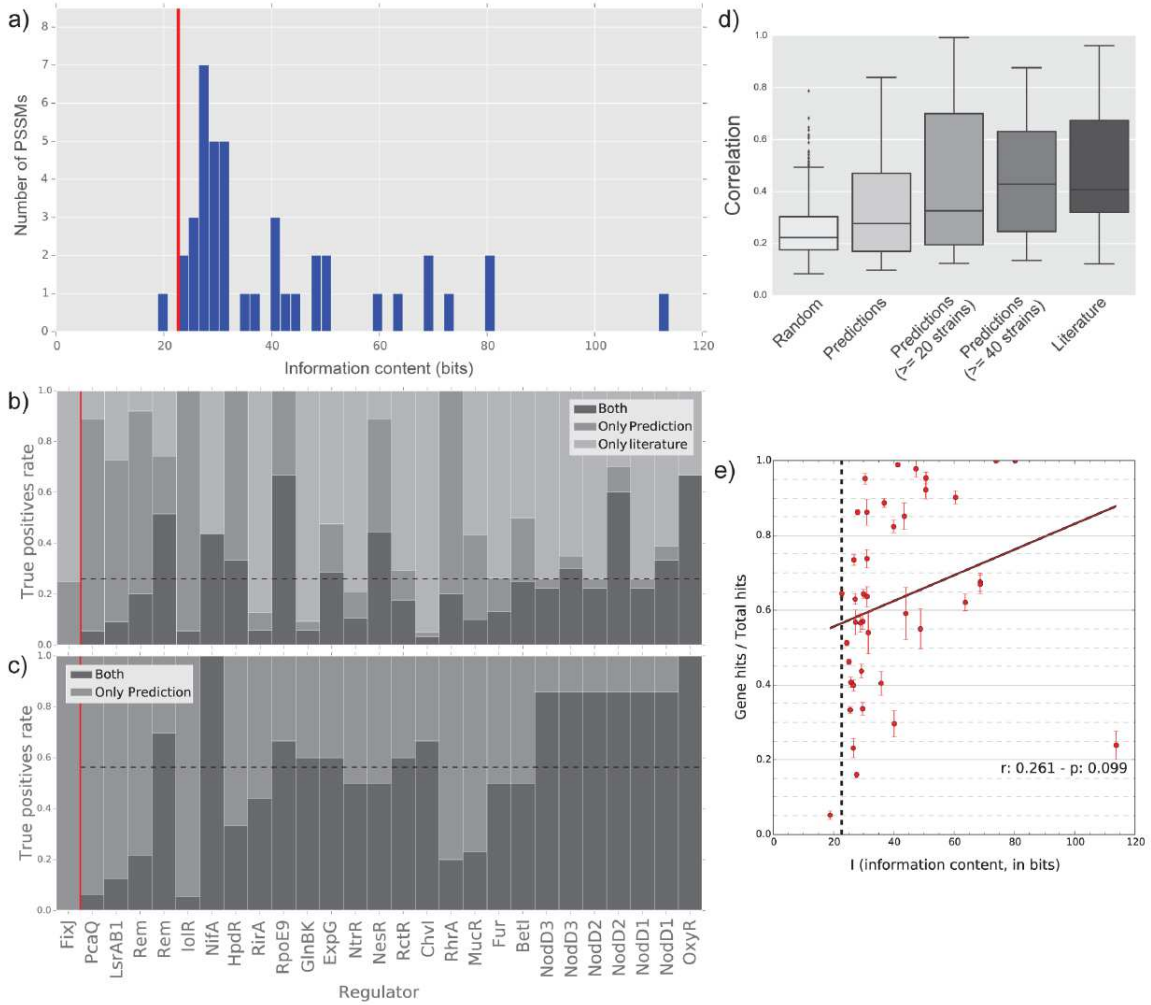


Fig 1. General characteristics of the presented TF predictions and quality control. a) Information content frequencies for the 41 analysed TFs: vertical line indicates the minimum information content, as measured for *S. meliloti* strain Rm1021; b-c) comparison between TFBS predictions and the reported experimental results in strain Rm1021: the dashed horizontal line indicates the mean value for the TFs with information content higher than the minimum value; d) correlations with the COLOMBOS expression compendium for *S. meliloti* Rm1021; e) correlation between the TFs information content and the signal-to-noise ratio, measured as the proportion of prediction in genes upstream regions over the total number of predictions: vertical bars indicate the error level measured in all the strains.

doi:10.1371/journal.pcbi.1004478.g001

influence of the information content on motif recognition. The predicted regulons were found to have comparable sizes in the outgroups; therefore the regulon is conserved in size between different species; this might be the result of the conservation across the species of the TFBS or of more general energy constraints on transcription/translation.

PLOS | COMPUTATIONAL BIOLOGY

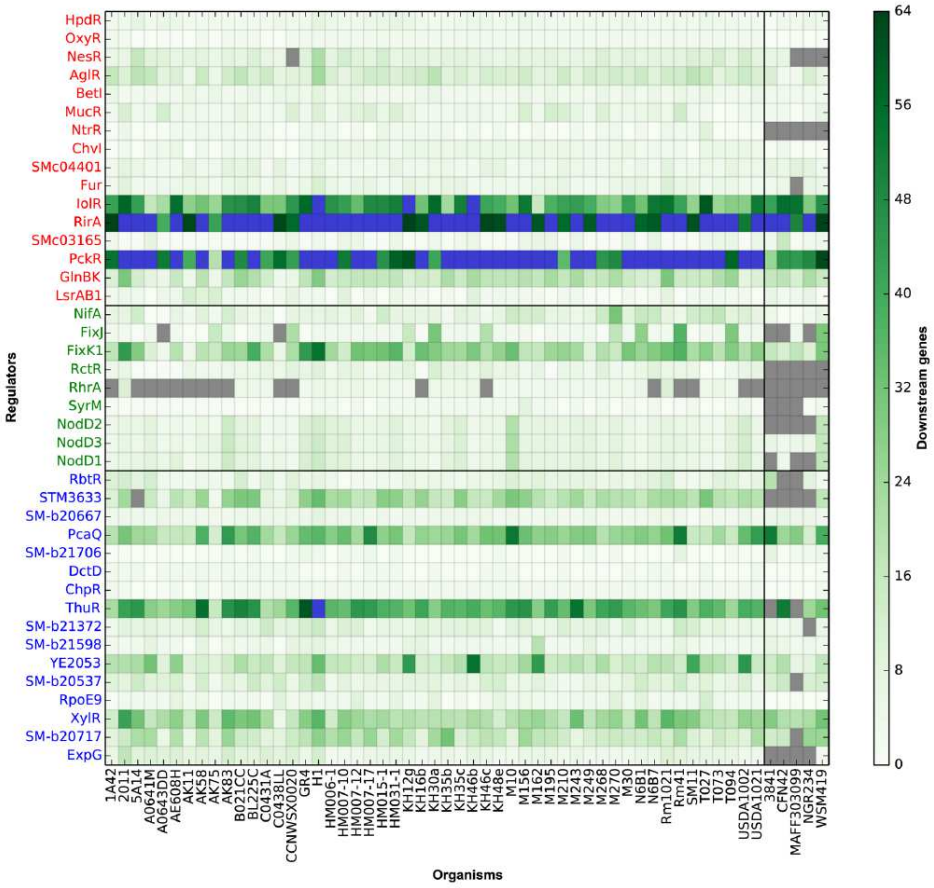


Fig 2. Variability in regulon size. Color intensity indicates the number of downstream regulated genes in each strain; gray squares indicate the TF absence in the genome of that particular strain. Blue squares indicate that there are more than 64 genes predicted to be under the control of the TF. TFs are colored according to the replicon they belong to: red for chromosome, green for the pSymA megaplasmid and blue for the pSymB chromid.

doi:10.1371/journal.pcbi.1004478.g002

Besides similar regulon sizes, we found that an average 40% of genes belonging to a regulon belong to the accessory genome (Table 2); this implies that although variable, each TF recruits a similar number of genes under its control, at least in the species analysed here. Obviously, the variability of the regulons is related with both the variability in upstream regions of core genes and the presence of genes from the accessory genome (whose presence varies across and between the species) in the regulons.

Predictions for TFs with low information content TFBSs showed a very poor accuracy and precision when compared to experimental data found in the literature; an efficient search strategy for such TFBSs using PSSM has still to be developed. However, from an evolutionary point of view, since those TFs are predicted to bind rather aspecifically to many sites along the genome, this would result in even a larger divergence of regulons between strains, as recently reported in comparison among species [51].

PLOS | COMPUTATIONAL BIOLOGY

Table 1. Regulon downstream genes.

		S. meliloti		Outgroups	
Regulator	*Replicon* [a]	Mean regulon size	*MAD* [b]	Mean regulon size	*MAD* [b]
HpdR	Chromosome	3.10	1.0	2.4	1.0
OxyR	Chromosome	1.71	0.0	0.0	0.0
NesR	Chromosome	8.24	1.0	6.0	NA
AgiR	Chromosome	11.69	2.0	6.4	4.0
BetI	Chromosome	3.22	0.0	2.2	0.0
MucR	Chromosome	5.20	1.0	2.4	0.0
NtrR	Chromosome	1.57	2.0	NA	NA
ChvI	Chromosome	3.53	1.0	0.6	0.0
SMc04401	Chromosome	4.04	1.0	6.4	9.0
Fur	Chromosome	4.45	2.0	4.5	1.0
IolR	Chromosome	41.80	10.0	45.4	7.0
RirA	Chromosome	71.55	8.0	78.2	4.0
SMc03165	Chromosome	1.69	0.0	4.4	1.0
PckR	Chromosome	69.80	13.0	45.0	3.0
GlnBK	Chromosome	15.35	4.0	16.2	2.0
LsrAB1	Chromosome	2.86	2.0	3.4	3.0
NifA	pSymA	8.02	2.0	2.6	3.0
FixJ	pSymA	5.86	1.0	16.5	NA
FixK1	pSymA	24.27	6.0	16.8	5.0
RctR	pSymA	4.59	2.0	NA	NA
RhrA	pSymA	3.79	1.0	NA	NA
SyrM	pSymA	1.10	0.0	1.0	NA
NodD1	pSymA	6.77	2.0	10.0	NA
NodD2	pSymA	6.41	2.0	17.0	NA
NodD3	pSymA	6.71	2.0	6.4	2.0
RbtR	pSymB	5.67	2.0	9.67	17.0
STM3633	pSymB	19.72	4.0	17.0	NA
SM-b20667	pSymB	3.18	1.0	4.0	0.0
PcaQ	pSymB	27.82	5.0	31.2	10.0
SM-b21706	pSymB	0.80	0.0	3.0	2.0
DctD	pSymB	1.24	1.0	0.0	0.0
ChpR	pSymB	1.88	0.0	0.4	0.0
ThuR	pSymB	37.63	7.0	36.0	28.0
SM-b21372	pSymB	7.33	2.0	3.5	0.0
SM-b21598	pSymB	3.51	1.0	3.4	1.0
YE2053	pSymB	19.47	4.0	13.0	6.0
RpoE9	pSymB	3.35	1.0	0.0	0.0
XylR	pSymB	22.61	5.0	24.2	2.0
SM-b20717	pSymB	15.24	4.0	19.25	11.0
SM-b20537	pSymB	7.77	3.0	12.0	1.0
ExpG	pSymB	6.0	2.0	2.0	NA

Regulatory network general statistics over the strains used in this study.

[a] Position according to the Rm1021 reference strain;

[b] Mean Absolute Deviation;

NA: not defined.

doi:10.1371/journal.pcbi.1004478.t001

Table 2. Regulon conservation.

Regulator	Replicon	*S. meliloti*	*Outgroups* [a]
HpdR	Chromosome	0.56	0.95
OxyR	Chromosome	1.00	1.00
NesR	Chromosome	0.57	0.52
AglR	Chromosome	0.57	0.48
BetI	Chromosome	0.33	0.50
MucR	Chromosome	0.89	0.71
NtrR	Chromosome	0.56	NA
ChvI	Chromosome	0.98	0.86
SMc04401	Chromosome	0.56	0.74
Fur	Chromosome	0.49	0.73
IolR	Chromosome	0.59	0.52
RirA	Chromosome	0.58	0.56
SMc03165	Chromosome	0.68	0.65
PckR	Chromosome	0.60	0.56
GlnBK	Chromosome	0.58	0.63
LsrAB1	Chromosome	0.56	0.71
NifA	pSymA	0.57	0.74
FixJ	pSymA	0.56	0.63
FixK1	pSymA	0.56	0.63
RctR	pSymA	0.57	NA
RhrA	pSymA	0.56	NA
SyrM	pSymA	0.57	0.00
NodD1	pSymA	0.56	0.65
NodD2	pSymA	0.56	0.66
NodD3	pSymA	0.57	0.69
RbtR	pSymB	0.57	0.46
STM3633	pSymB	0.57	0.63
SM-b20667	pSymB	0.57	0.53
PcaQ	pSymB	0.57	0.51
SM-b21706	pSymB	0.96	0.75
DctD	pSymB	1.00	1.00
ChpR	pSymB	0.57	1.00
ThuR	pSymB	0.58	0.47
SM-b21372	pSymB	0.57	0.51
SM-b21598	pSymB	0.57	0.70
YE2053	pSymB	0.58	0.50
RpoE9	pSymB	0.99	0.60
XylR	pSymB	0.57	0.54
SM-b20717	pSymB	0.56	0.52
SM-b20537	pSymB	0.56	0.50
ExpG	pSymB	0.56	0.45

Regulatory network conservation in *S. meliloti* and near rhizobial species. For each regulator the number of conserved downstream genes over the average regulon size is reported.

[a] *S. meliloti* strain Rm1021 is also considered.

NA: not defined.

doi:10.1371/journal.pcbi.1004478.t002

Upstream sequences and accessory genome changes are correlated with regulon diversity

To clarify if the patterns of variability of the regulatory network are related to the phylogenetic distance among strains a comparison between divergence of panregulons and divergence of pangenomes was performed.

Following the pangenome analysis, we calculate three sets of distance matrices among the genomes under analysis (see Materials and Methods): the first was obtained from the alignment of core genes (hereinafter the *core distance*), the second from alignments of the upstream regions of the core genes (the *upstream distance*), and the third is instead based on the presence/absence profiles of accessory genes (*gene content distance*). The three distances were then compared with the *regulatory network distance* of the corresponding strains/species, which was calculated with the same metric defined by Babu and collaborators [16]. Intuitively, the divergence in upstream regions should be paralleled by divergence in the regulatory network, since the former will at some point determine a loss/gain of TFBSs affecting the structure of the regulatory network. Similarly, a larger difference in gene content should also be mirrored by a higher variability in the regulatory network, since new genes may be recruited in the regulatory network and/or TFs may be lost/gained. On the other hand, we don't expect to observe a strong correlation between core and regulatory network distances; this is also due to the lower divergence at the coding level between strains, implying that regulon diversity inside a species could be driven by gene content variability and upstream sequences variability.

These hypotheses on patterns of correlations between pangenome differences and regulatory divergence were confirmed at the species level (Fig 3a and 3b). The comparison between *S. meliloti* strains showed that the regulatory network distance is correlated with both the upstream distance and with gene content distance. The core distance showed no significant correlation with the regulatory network distance (Fig 3b). When considering the outgroup species, all three distances were found to be similarly correlated with the regulatory network distance (Fig 3c). Since the divergence in coding sequences cannot directly influence transcriptional regulation (with the exception of non-synonymous mutations in the DNA binding domain of a TF), we propose that the most likely explanation of the observed correlations is the overall genome divergence between species, which is ultimately reflected by a higher divergence at the regulatory network level. This is also confirmed by the high correlation coefficients among the three distances. We then concluded that the patterns of regulatory network variation are paralleled, at the species level, by changes in promoter sequences and by the variation in the accessory genome composition, at least in *S. meliloti*. These two fractions of the pangenome could then be used as *bona fide* predictors of the extent of rewiring in regulatory networks. However, from these data we cannot confirm a direct causative explanation for the observed regulatory network variation, as this analysis has been focused on the whole pangenome. The striking difference between the slow rate of coding sequence evolution versus the much larger difference in the regulatory networks is however worth noting.

Evolutionary dynamics of regulatory networks

Regulatory network evolutionary dynamics showed interesting differences within and between species. Each observed regulatory interaction in the two datasets (*S. meliloti* and the outgroups) and its state across all strains was used to build a hidden markov model to infer the preferred state transitions in our predictions (see Materials and methods), that corresponds to the ways the gene regulatory network can grow and shrink. The possible states of a target gene depend on the presence of the TF, the target gene itself and the upstream TFBS. Therefore, each target gene can be found in one of six different states (Fig 4a). The "plugged" state being the only

PLOS | COMPUTATIONAL BIOLOGY

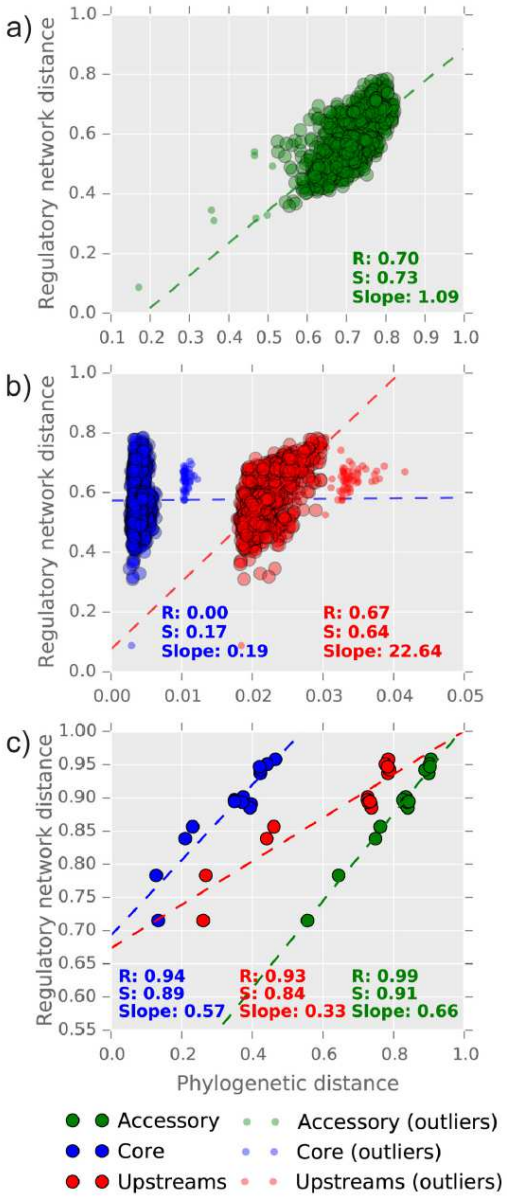


Fig 3. Correlations between pangenome diversity and regulatory network distances. R and S indicate the Pearson's and Spearman's correlation coefficients between the regulatory network and each pangenome partition distances (see Materials and Methods for the definition of the distances metrics used here). Outliers have been defined using a *Z*-score threshold of 3.5 on the mean absolute deviation of the distances. a) correlations within the *S. meliloti* species for the accessory genome; b) correlations within the *S. meliloti* species for coding and upstream regions; and c) correlation between the outgroups.

doi:10.1371/journal.pcbi.1004478.g003

PLOS | COMPUTATIONAL BIOLOGY

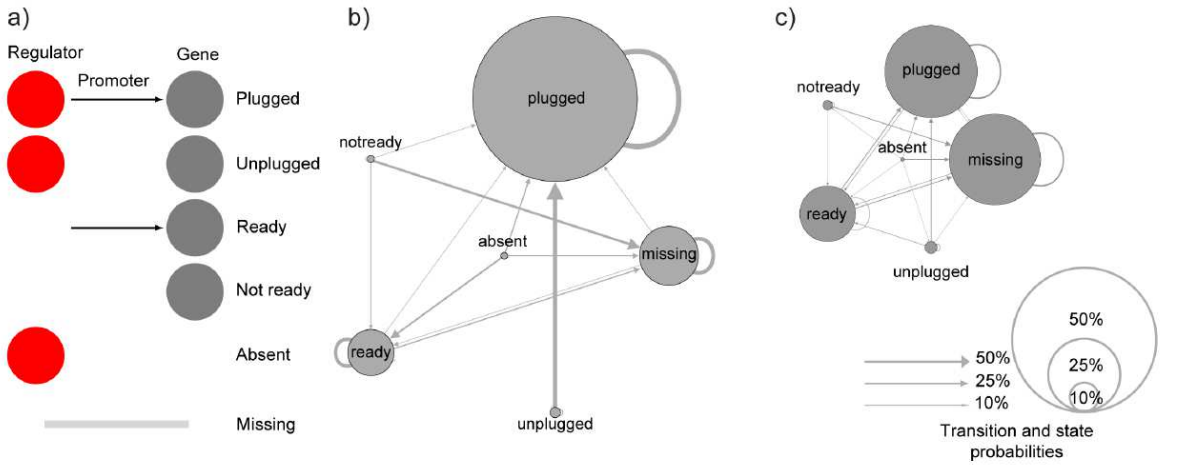


Fig 4. Regulatory network dynamics. a) Graphical representation of the six states in which each regulatory link (a gene found with a TFBS in at least one genome) can be found in the *S. meliloti* species and between the outgroup species; b) states probabilities and states transitions probabilities inside the *S. meliloti* species: nodes and edges sizes are proportional to the probability in the model. For each state, the sum of transition probabilities is one; transition probabilities below 0.1 are not shown; c) states probabilities and states transitions probabilities between the outgroup species.

doi:10.1371/journal.pcbi.1004478.g004

functional one, which corresponds to a target gene with a TFBS in its promoter region when the TF is present in the genome. The other five are non-functional states but may represent transitory states during the evolution of gene regulatory networks. Each of these states lack: i) the TFBS ("unplugged"), ii) the TF ("ready"), iii) both the TF and the TFBS ("not ready"), iv) the regulated gene ("absent") or v) both the TF and the gene itself ("missing"). This HMM can be used to estimate the probability for state transitions, that is the probability of observing a change from one state to another between two strains. This results in a model that is able to provide a general description of the evolution of regulatory networks within and between bacterial species. Since the models is based on observed states in the available strains, we consider it as a "snapshot" of the regulatory network evolution, and not an equilibrium model.

According to the model, the most represented state in the *S. meliloti* regulatory network is the "plugged" one, indicating conservation of regulatory interactions at the species level (Fig 4b and S3 Table). More interestingly, the model predicts that the "unplugged" genes are mostly seen recruited by the regulatory network and that the regulatory link is then maintained with high probability. Very little probability was given to the "plugged" to "missing" and "plugged" to "absent" transitions, indicating that genes belonging to the gene regulatory network are rarely removed from the genome. On the other hand, genes with no TFBS and its cognate TF are more frequently found to undergo loss ("not ready" to "missing"), suggesting that regulatory interactions are important for gene conservation at the species level. When considering a wider phylogenetic level (the outgroups), the broader variability in TF gene targets resulted in the "plugged" and "missing" state as equally probable, indicating that regulons might evolve by adding and removing new elements to a conserved kernel of gene targets (Fig 4c and S3 Table). This is also reflected in a smaller probability that a target gene i) remains in the "plugged" state when compared to the *S. meliloti* species level, and ii) that it acquires a TFBS. On the other hand, the same probability as within the *S. meliloti* species was observed for the transition "not ready" to "missing", which seems to confirm the importance of regulatory features in

explaining the accessory genome fraction evolution. Consequently, a different evolutionary dynamics of regulatory circuitry changes seems to be present in relation to the taxonomic ranks; at the species level, robust networks are formed and they tend to include new genes from the species pangenome, which then may be conserved. On the contrary, when comparing wider taxonomic ranges, regulatory networks are less conserved and genes are apparently included in each species' genome directly with their regulatory features (in a sort of plug-and-play model).

Replicon-specific regulation and cross-regulation

Transcription factors with replicon preference were found to have functional signatures in accordance with the functions encoded in the three main replicons of *S. meliloti*. This aspect has been evaluated by mapping each draft genome on the *S. meliloti* replicons (see Materials and methods) and considering the presence of each gene in the replicons for each of the 51 strains analysed here. Using a clustering approach on normalized gene hits on each replicon we have found that 19 TFs preferentially regulate genes belonging to one of the three replicons: five to the chromosome (NtrR, OxyR, NesR, ChvI and SMc03165), six to the pSymB chromid (SM-b21706, SM-b20667, ChpR, RbtR, SM-b21598 and SM-b21372) and eight to the symbiotic megaplasmid pSymA (SyrM, NodD3, RhrA, NodD1, NodD2, FixJ, FixK1 and NifA) (Fig 5a); these TFs are also encoded by the same replicon.

The six TFs encoded by the pSymB chromid (whose regulon is also preferentially located on pSymB) appear to mostly regulate the transport and metabolism of various carbon and nitrogen sources, including ribitol (RbtR), tagatose, sorbitol and mannitol (SM-b21372), ribose (SM-

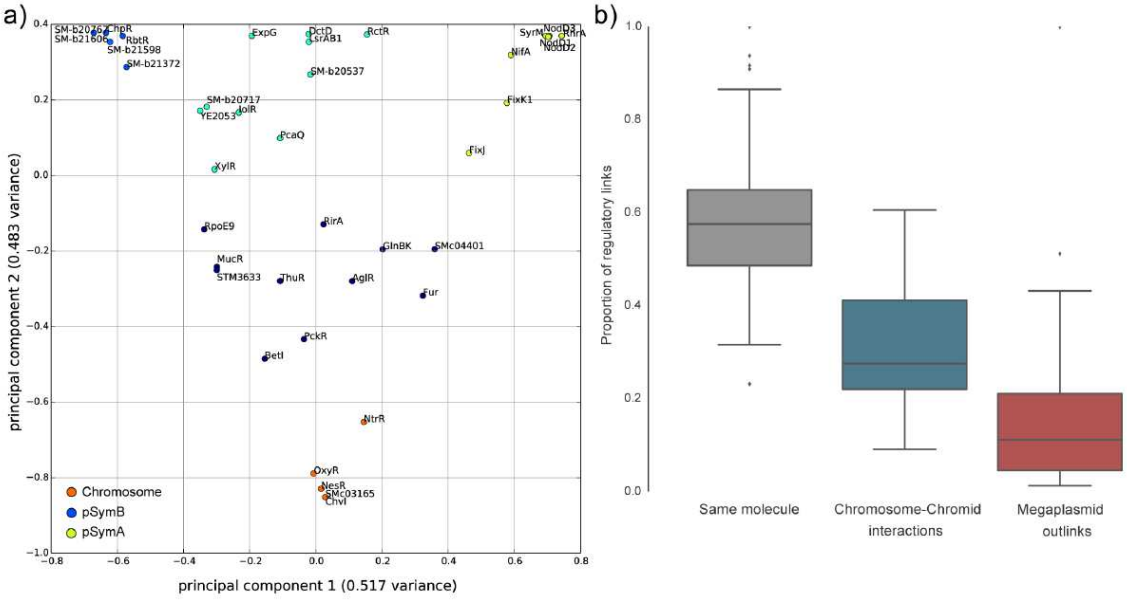


Fig 5. TFs preferentially associated with a replicon. a) K-means clustering of the normalized proportion of genes regulated in each of the three main replicons of *S. meliloti*, visualized in a two-dimensional PCA. The dark blue and cyan clusters contain TFs with no clear replicon preference; b) Variability in the number of regulatory links in the same replicon and between replicons. All differences are significant (t-test p-value < 0.05).

doi:10.1371/journal.pcbi.1004478.g005

b21598), lactose (SM-b21706) and tartrate, succinate, butyrate and pyruvate (SM-b20667). The eight TFs present in the symbiotic megaplasmid pSymA (with regulons preferentially located on pSymA) were found to be involved in the regulation of key symbiotic processes, including nitrogenase synthesis and functioning through micro-aerophilia (FixJ, FixK1 and NifA), nod-factors biosynthesis (SyrM, NodD1, NodD2 and NodD3), and iron scavenging (RhrA).

A functional enrichment analysis using COG annotations (S3 Fig) on genes belonging to the regulons of the replicon-biased TFs confirmed this general observation: no functional category was enriched in the chromosome. The G category (*carbohydrate metabolism and transport*) was enriched in genes regulated by pSymB encoded TFs, in agreement with the role of chromid pSymB in providing metabolic versatility to *S. meliloti*. The C (*energy production and conversion*), U (*intracellular trafficing and secretion*) and T (*Signal Transduction*) categories were enriched in genes under the control of pSymA-harboured TFs, which show some relationship with the establishment on the plant symbiosis. This analysis allowed us to depict a scenario where a significant part of the regulatory network is replicon-specific, with a tendency to maintain the functional signature of the host replicon, thus confirming earlier reports on the evolutionary independence of chromids and megaplasmids in *S. meliloti* [29, 31, 32].

Interestingly, a fraction of TFs have target genes which span over different replicons, and show a preference for cross regulation between the chromosome and the chromid (Fig 5b). The presence of cross-replicon regulons, may indeed allow a stabilization of genomic structure, genetically and metabolically connecting chromosome encoded functions with those present in the other two *S. meliloti* replicons. In the evolutionary model of the chromid [29, 31, 32], its stabilization within the host genome is related to the acquisition of essential (core) genes in a previously introgressed megaplasmid which gained niche-specific genes. Here, we found that for TFs encoded on the chromosome (as AglR, GlnBK, IolR, BetI, LsrAB, MucR, PckR, RirA, NesR) a variable number of target genes are present on pSymB (S1 Material). The preference for cross-regulation between the chromosome and the chromid, as opposed to the megaplasmid uncovers an additional mechanism by which a chromid integrates itself in bacterial pangenomes.

Discussion

Regulatory networks are key components of cell's response to environmental and physiological changes. In the past years, several works have highlighted a high transcriptomic variability in strains or individuals from the same species [52, 53], in addition to genomic variation. Consequently, regulatory network variation might have profound impact on local adaptation and fitness of organisms. Recent studies have confirmed that bacterial regulatory networks are able to tolerate the addition of new genes [24], which in turn can serve as raw material for selection to operate. Using our original combined search strategy, we indeed found variability in regulon composition within the *S. meliloti* species, which in fact accounted on average on 40% of the regulon of each strain. On the other hand the regulon size was found to be conserved even outside the species boundary. This could suggest that even though the genes under the control of a TF vary between strains, there is a general constraint on the size of the transcriptional response. Whether this is due to energy constraints or being simply an effect due to the genome base composition is yet to be clarified.

We found that the regulatory network distance (as defined in [16]) correlates with the upstream distance and also with the gene content distance. This correlations may suggest that regulatory network composition is influenced by both promoter variability and accessory genome variability. Indeed, we may speculate that the sequence divergence in upstream regions can result in the appearance or disappearance of TFBSs, thus changing the regulatory network content. Moreover, gene content dynamics may also have a strong impact on the regulatory

network, with the introduction of new gene cassettes containing TFBS recognized by resident TFs. We can consequently hypothesize that the evolution of bacterial regulatory networks, as that of the pangenome, may be influenced by mechanisms of gene acquisitions, such as lateral gene transfer, and it's not only linked to mutations in upstream regions.

The observed changes in the regulatory network also show interesting features with respect to pangenome composition. Indeed, even if a significant difference in the state transitions of regulatory links inside and outside the species boundary has been shown, for genes that lack both a TFBS and their cognate TF, we have observed a similar tendency to disappear from the pangenome. This observation may suggest that the dynamics governing pangenome evolution within a species could depend in part on a 'gene fitness' related to being wired into the regulatory network. We can then propose that regulatory networks have an important role in shaping the bacterial gene content and can contribute to gene fitness, which in turn may be linked to environmental adaptation.

Moreover, the preference of nineteen TFs for target genes on one of the three replicons of *S. meliloti* indicates that in multipartite bacterial genomes, similarly to replicon-dependent patterns of evolution in gene and functions content [31], a replicon-specific transcriptional regulation is to be expected. At the same time, a significant number of cross-links between the chromosome and the chromid suggest for the first time an additional mechanism by which new replicons can be integrated into a bacterial pangenome.

Materials and Methods

Genome sequences

The 51 genomic sequences belonging to *Sinorhizobium meliloti* and the five genomic sequences from closely related symbiotic species are listed in S4 Table.

Orthology

The orthology relationships inside the 51 *S. meliloti* strains has been computed using the Blast-BBH algorithm implemented in the DuctApe suite (version 0.13.0) [54], using default parameters. The same analysis has been conducted on the five closely related species with the addition of the Rm1021 reference strain, using the BLOSUM62 scoring matrix to account for their greater sequence diversity.

Regulators estimation

The number of regulators present in each genome has been estimated using COG annotations. The similarity of each protein against the COG database has been measured with a rpsblast scan [55], using an E-value threshold of 1e-10. Each protein mapped to the COG category K (Transcription) has been considered as a putative regulator.

Confirmation of the absence of the *fixJ* gene

To confirm the absence of the *fixJ* gene in strains A0643DD and C0438LL, PCR primers amplifying a large portion (from nucleotide position 32 nt to 595 out of 615 nt total) of the coding sequence of *fixJ* gene have been designed on the basis based on the ortholog sequence in strain BL225C (SinmeB_6173) with Primer3Plus (fw: 5′-ACGAAGAGCCGGTCAGGAAGTCGCTG GCATTCATGCTG-3′; rv 5-CGGCGAGAGCCATGCGAACGAGATGGGGGAGGCTC-3) [56]. PCR has been performed with the Maxima Hot Start Green Master Mix (Thermo Fisher) in 20 microL total volume by using 10 ng of DNA, purified from liquid culture with FAST DNA Kit (QBiogene) and 10 pmols of each primer. Cycling conditions were as follows: 5′

94°C, followed by 30" 94°C, 30" 55°C, 1′ 72°C repeated for 35 cycles. PCR products were resolved after agarose gel electrophoresis (1.5 w/v) in TAE buffer with ethidium bromide (10 microg/ml) as staining agent.

Regulatory motifs collection

The 83 regulators whose PSSM has been extracted from the various sources are listed in S1 Table. For those PSSMs retrieved from the literature, we collected the upstream regions of the regulated genes and (when available), the consensus binding sites from bibliographical records; the upstream regions have then been analysed with the *meme* program [7](version 4.9.0), using the model that retrieved the PSMM with higher similarity to literature. Twenty-two motif files have been generated using the information retrieved from the RhizoRegNet database [27]. Fifteen motif files have been generated using the information retrieved from the RegTransBase database [57]. For the 5 regulators having more than one predicted motif, for instance those having a variable length (FixJ, RpoD, RpoE2, RpoH1 and RpoH2), one motif file for each motif length has been generated. All the retrieved PSSMs have been converted to HMM models using the *hmmbuild* program from the HMMer suite [12–14](version 3.1b1), using the alignments present in the MEME motif file. It has been previously shown that in bacterial genomes TFBS can be reliably distinguished from background DNA only if their information content is higher than the minimum information content for the target genome, which depends on the genome size and composition [5] (this simplification of course ignores other factors such accessibility or proximity of the RNA polymerase). The information gain of the TFBS with respect to the genome is calculated using the Kullback-Leibler divergence between the corresponding nucleotide frequencies [58], and it has been shown to correlate with the motif length and base composition of the motif with respect to the surrounding genome sequence. TF motifs with sufficient information content also tend to show less variability in their regulon composition between species [51]; by focusing our analysis on such TFs we ensured a more precise analysis. The information content of each motif has been calculated as suggested by Wunderlich et al [5], using the Rm1021 reference genome for the calculation of the minimum information content; given the dependence of this variable on genome size and the fact that all the *S. meliloti* strains have similar genome size, there has been no need to calculate a strain specific threshold. PSSMs whose information content was found to be lower the minimum information content have been discarded with exception of FixJ, which has two distinct PSSM, one of which is above the threshold. In the presence of more than one source for a regulator (literature, RhizoRegNet or RegTransBase), the PSSM having the highest information content has been considered in the final analysis.

Search of regulatory motifs occurrences

For each genome, background k-mers frequencies have been calculated using the *fasta-get-markov* program from the MEME suite (version 4.9.0) [7], using 3 as the maximum value for k. Each regulatory motif has been searched inside each genomic sequence using four scanning algorithms. The *mast* program from the MEME suite (version 4.9.0) [7] has been used with an E-value threshold of 100 and the use of a genome-specific background file. The *matrix-scan* program from the RSAT suite [8–10] has been used with a P-value threshold of 0.001, the background file and a pseudocount of 0.01, as suggested by Nishida et al. [59]. The *Bio.motifs* package from the Biopython library (version 1.62b) [11] has been used with a false negative rate threshold of 0.05 and a pseudocount of 0.01, as suggested by Nishida et al. [59]. The *nHMMer* program from the HMMer suite (version 3.1b1) [12–14] has been used with an E-value threshold of 100 and with all the heuristic filters turned off. Each regulatory motif hit has been parsed, separating the hits being present in the upstream region of a gene from the others. The

upstream region has been defined as the intergenic region (not overlapping any coding sequence) in front of the first codon with a maximum size of 600 bp. In the case of a palindrome motif, the motif orientation has been ignored.

The distributions of the raw scores has been tested using a normality test, as implemented in the SciPy library (version 0.13.3) [60][61]. The score threshold has been determined through the calculation of the raw scores quartiles (Q1 and Q3) and defining the score threshold (τ_S in Eq 1) in order to consider only the upper outliers [62].

$$\tau_s = Q3 + (1.5(Q3 - Q1)). \quad (1)$$

For the Biopython method the bit score has been used, while for the RSAT, HMMer and MEME methods the negative base 10 logarithm of the E-value has been considered. The regulatory motifs predicted by at least three methods have been considered for further analysis.

Validation of the predictions

The compendium of gene expression data for *S. meliloti* str. Rm1021 from the Colombos database [50] was used to calculate correlation coefficients among genes in the regulons reported in the literature, our predictions and random sets of genes. Random regulons were produced by random sampling groups of genes of size 5, 10 and 15, for which 500 sets were produced. Correlation was quantified by the squared uncentered correlation coefficient, which was calculated using Matlab, as the square of 1 – *cos distance*. Values plotted in Fig 1d are averages over the entire set of genes under analysis. We have implemented a strategy allowing to select the conditions maximizing the average squared correlation within a group of genes, since many of the conditions of the compendium are likely not related to our predictions. Selection of the conditions was performed using the genetic algorithm implemented in the GA Matlab function, with default tolerances (TolCon = 10^{-6}, TolFun = 10^{-6}). We let the algorithm select the conditions minimizing $\frac{1}{R^2}$ where R is the uncentered correlation averaged over all pairwise comparisons made within the group of genes under analysis. Since we noticed that correlations are strongly and inversely correlated with the number N of included conditions, especially when $N \leq 20$, we discarded all cases where the number of conditions was less than 20 (final $N = 950$). All conditions containing missing data in at least one of the genes under analysis were discarded before starting the procedure. For some of the known and predicted regulons, correlations were not calculated as the available number of conditions after removing missing data was less than 30 before the optimization.

Experimental confirmation of promoters

Upstream sequences from selected putative target genes of NodD regulon were analysed (see S2 Table). Sequences (approximately 400 nt upstream the translation start site of the gene) were amplified from crude lisates of *S. meliloti* strains with AccuPrime *Pfx* DNA Polymerase (Thermo Fisher) and cloned into pTO2 vector (which carries GFPuv as reporter gene [63]) by using *SalI* and *KnpI* restriction sites. Recombinant clones of *E. coli* S17-1 strain were selected by gentamycin resistance and verified by sequencing of inserted fragments. Positive clones were used for transferring recombinant pOT2 vectors to *S. meliloti* Rm1021 by bi-parental conjugation by using previously described protocols [64][65]. *S. meliloti* Rm1021 recombinant strains were then tested for GFP fluorescence after incubation of a 5 ml culture grown at the mid-exponential phase with 1 microM luteolin (Sigma-Aldrich) in liquid TY medium at 30°C for 3h. GFP fluorecence was measured on a Infine200 Pro plate reader (Tecan). Measures were taken in triplicate and normalized to cell growth estimates as absorbance to 600nm.

Operon prediction

The operons belonging to the 56 genomes of this study have been predicted using the Operon Prediction Software (OPS, version 1.2) [66], using a beta threshold of 0.7 and a probability threshold of 0.5. The number and length of the predicted operons in each strain are listed in S5 Table.

Replicon mapping

Each contig of the 44 *S. meliloti* draft genomes has been mapped to the seven complete genomes using CONTIGuator (version 2.7.3) [67], using a 15% coverage threshold and considering blast hits over 1000 bp in length. A contig has been considered mapped to a replicon when it has been found mapped to the replicon in at least five complete genomes, or when it has been mapped to the replicon in at least one complete genome and to no replicon in the others. Knowing that very few portions of the *S. meliloti* genome are shuffled between replicons [31], we assessed the quality of this mapping procedure by checking whether the *S. meliloti* orthologs were found to be mapped to more than one replicon; for each orthologous group the genes not mapped to any replicon have been removed, and the relative abundance of the most representative mapped replicon has been computed. A relative abundance of 1 means that the orthologs have all been mapped to the same replicon in all the strains. The vast majority of the orthologous groups was found to map to a single replicon (S4 Fig).

The number of average gene hits has been divided for each replicon (either from a complete genome or a draft genome) and normalized by the number of genes belonging to each replicon in the Rm1021 reference strain. Regulators with preferential regulatory hits in a specific replicon have been highlighted performing a k-means clustering (k = 5, selected using an elbow test [68]) and plotted using the two principal components of the proportion of hits in each replicon, using the scikits-learn package (version 0.14.1) [69]. Only the three main replicons (chromosome, pSymB and pSymA) have been considered. COG categories enrichments have been tested using a Fisher's exact test, as implemented in the DendroPy package [70].

Phylogenetic distance

Phylogenetic distance inside the *S. meliloti* pangenome and the pangenome of the five related species has been computed as described in a previous work [31]. The pangenome has been divided in three fractions, allowing the use of three distinct phylogenetic distances. The "core" distance has been calculated through the alignment of all the nucleotide sequences of each core gene, discarding those genes where at least one sequence was 60bp shorter or longer with respect to the other sequences. The "upstream" distance has been calculated through the alignment of the core genes upstream regions, discarding sequences below 5bp in length. The alignments have been calculated using MUSCLE (version 3.8.31) [71] and the bayesian tree has been inferred using MrBayes (version 3.2.0) [72]. The distance matrix for both distance categories has been computed from the phylogenetic tree using the textitBio.Phylo package inside the Biopython library (version 1.62b) [73]. The "accessory" distance has been calculated through the construction of a presence/absence binary matrix for all the accessory genome OGs; the distance between each strain has been then calculated using the Jaccard distance measure, as implemented in the SciPy library (version 0.13.3) [61].

Regulatory network distance

The distance between each strain inside the *S. meliloti* and the other five related species regulatory network has been computed using the distance in the presence/absence of regulatory interactions as suggested in the work of Babu and collaborators [16]. The distance between strain A

and B is computed using Eq 2.

$$D_{AB} = 1 - \frac{core_{AB}}{total_{AB}}, \quad (2)$$

where $core_{AB}$ and $total_{AB}$ represent the number of conserved and total regulatory interactions, respectively.

Pearson and Spearman correlation coefficients between the pangenome and the regulatory network distance have been calculated using the implementations of the SciPy library (version 0.13.3) [61], removing the outliers using a Z-score threshold of 3.5 on the mean absolute deviation of the distances.

Regulatory network transistions

The state transitions of the regulatory network has been inferred by encoding them in a hidden markov model. Each one of the regulatory links observed in at least one strain has been tested for their state in each organism, following the labelling of Fig 4a. Specifically, each regulatory link in the network of each organism could belong to one of the following categories:

- **Plugged:** regulator, gene and TFBS present
- **Unplugged:** regulator and gene present, TFBS absent
- **Ready:** gene and TFBS present, regulator absent
- **Not ready:** gene present, regulator and TFBS absent
- **Absent:** regulator present, gene and TFBS absent
- **Missing:** regulator, gene and TFBS absent

The hidden markov model has been constructed using the Baum-Welch algorithm [74], as implemented in the GHMM python library. For each observed regulatory link in the regulatory network, the observed transition between each permutation of pairs of strains has been used to train the HMM and then compute the states and transitions probabilities. The transition probability has been defined for each state as the probability of observing the transition between two strains. Since each state has different transition probabilities and their sum is one for each state, we do not observe symmetrical probabilities.

Results analysis and visualization

Regulatory motifs data has been analysed and visualized using the NumPy [75] and matplotlib [76] libraries inside the iPython environment [77]. Regulatory networks have been built using the networkx library [78] and visualized using Gephi [79].

Data and methods availability

Genomic sequences, regulatory motif files and search and analysis scripts are available as separate git repositories. The rhizoreg repository (https://github.com/combogenomics/rhizoreg/), contains the input data; the regtools repository (https://github.com/combogenomics/regtools/) contains the main scripts used to conduct the analysis.

Supporting Information

S1 Material. Inter and intra-regulation in the 51 S. meliloti strains.
(ZIP)

S2 Material. Single regulons correlations with the COLOMBOS expression compendium.
(ZIP)

S1 Table. Sources and information content of the TF PSSM of this study.
(XLS)

S2 Table. Experimental validation of NodD targets.
(CSV)

S3 Table. State transitions probability for the regulatory networks.
(CSV)

S4 Table. Genomic sequences used in this study.
(CSV)

S5 Table. Predicted operons statistics.
(CSV)

S1 Fig. Total TFs encoded in the pangenome. a) TFs frequency (expressed as the number of strains having the TF encoded in their genome) in S. meliloti and the other rhizobial genomes; b) TF presence/absence matrix in the strains analysed in this study: red indicates the TF absence. TFs are colored according to the replicon they belong to: red for chromosome, green for the pSymA megaplasmid and blue for the pSymB chromid.
(TIF)

S2 Fig. Correlation between predictions quality and TF information content. Vertical dashed line indicates the minimum information content for S. meliloti strain Rm1021. a) Correlation between predictions true positive rate and information content; b) Correlation between the number of predicted regulated genes and information content.
(TIF)

S3 Fig. COG categories enrichment in the replicons. For each replicon, the proportion of regulated downstream genes belonging to each category is compared with the genes belonging to other replicons. Purple categories indicate a statistically significant enrichment.
(TIF)

S4 Fig. Replicon mapping quality control. For each orthologous group in the S. meliloti pangenome, the abundance of the most mapped replicons has been computed as a proxy for the consistency of the replicon mappings.
(TIF)

Author Contributions

Conceived and designed the experiments: MG MBr EGB MBa AM. Performed the experiments: MG MBr MM KE GS MR BR AB MC AM. Analyzed the data: MG MBr KE MM GB FP MBa AM. Contributed reagents/materials/analysis tools: MG MBr MM KE. Wrote the paper: MG FP MBr AM.

References

1. Depardieu F, Podglajen I, Leclercq R, Collatz E, Courvalin P (2007) Modes and modulations of antibiotic resistance gene expression. Clinical microbiology reviews 20: 79–114. doi: 10.1128/CMR.00015-06 PMID: 17223624
2. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annual Reviews in Microbiology 57: 441–466. doi: 10.1146/annurev.micro.57.030502.090913

3. van Hijum SA, Medema MH, Kuipers OP (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. Microbiology and Molecular Biology Reviews 73: 481–509. doi: 10.1128/MMBR.00037-08 PMID: 19721087
4. Lässig M (2007) From biophysics to evolutionary genetics: statistical aspects of gene regulation. BMC bioinformatics 8: S7. doi: 10.1186/1471-2105-8-S6-S7 PMID: 17903288
5. Wunderlich Z, Mirny LA (2009) Different gene regulation strategies revealed by analysis of binding motifs. Trends in genetics 25: 434–440. doi: 10.1016/j.tig.2009.08.003 PMID: 19815308
6. Hunziker A, Tuboly C, Horváth P, Krishna S, Semsey S (2010) Genetic flexibility of regulatory networks. Proceedings of the National Academy of Sciences 107: 12998–13003. doi: 10.1073/pnas.0915003107
7. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) Meme suite: tools for motif discovery and searching. Nucleic acids research 37: W202–W208. doi: 10.1093/nar/gkp335 PMID: 19458158
8. Van Helden J (2003) Regulatory sequence analysis tools. Nucleic acids research 31: 3593–3596. doi: 10.1093/nar/gkg567 PMID: 12824373
9. Thomas-Chollier M, Sand O, Turatsinze JV, Defrance M, Vervisch E, et al. (2008) Rsat: regulatory sequence analysis tools. Nucleic acids research 36: W119–W127. doi: 10.1093/nar/gkn304 PMID: 18495751
10. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, et al. (2011) Rsat 2011: regulatory sequence analysis tools. Nucleic acids research 39: W86–W91. doi: 10.1093/nar/gkr377 PMID: 21715389
11. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, et al. (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 25: 1422–1423. doi: 10.1093/bioinformatics/btp163 PMID: 19304878
12. Eddy SR, et al. (2009) A new generation of homology search tools based on probabilistic inference. In: Genome Inform. World Scientific, volume 23, pp. 205–211.
13. Johnson LS, Eddy S, Portugaly E (2010) Hidden markov model speed heuristic and iterative hmm search procedure. BMC bioinformatics 11: 431. doi: 10.1186/1471-2105-11-431 PMID: 20718988
14. Eddy SR (2011) Accelerated profile hmm searches. PLoS computational biology 7: e1002195. doi: 10.1371/journal.pcbi.1002195 PMID: 22039361
15. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, et al. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431: 99–104. doi: 10.1038/nature02800 PMID: 15343339
16. Babu M, Teichmann SA, Aravind L (2006) Evolutionary dynamics of prokaryotic transcriptional regulatory networks. Journal of molecular biology 358: 614–633. doi: 10.1016/j.jmb.2006.02.019
17. Gelfand MS (2006) Evolution of transcriptional regulatory networks in microbial genomes. Current opinion in structural biology 16: 420–429. doi: 10.1016/j.sbi.2006.04.001 PMID: 16650982
18. Janga SC, Collado-Vides J (2007) Structure and evolution of gene regulatory networks in microbial genomes. Research in microbiology 158: 787–794. doi: 10.1016/j.resmic.2007.09.001 PMID: 17996425
19. Lozada-Chavez I, Janga SC, Collado-Vides J (2006) Bacterial regulatory networks are extremely flexible in evolution. Nucleic acids research 34: 3434–3445. doi: 10.1093/nar/gkl423 PMID: 16840530
20. Hendriksen WT, Silva N, Bootsma HJ, Blue CE, Paterson GK, et al. (2007) Regulation of gene expression in streptococcus pneumoniae by response regulator 09 is strain dependent. Journal of bacteriology 189: 1382–1389. doi: 10.1128/JB.01144-06 PMID: 17085554
21. Brilli M, Fondi M, Fani R, Mengoni A, Ferri L, et al. (2010) The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. BMC systems biology 4: 52. doi: 10.1186/1752-0509-4-52 PMID: 20426835
22. Frandi A, Mengoni A, Brilli M (2010) Comparative genomics of virr regulons in clostridium perfringens strains. BMC microbiology 10: 65. doi: 10.1186/1471-2180-10-65 PMID: 20184757
23. Galardini M, Mengoni A, Brilli M, Pini F, Fioravanti A, et al. (2011) Exploring the symbiotic pangenome of the nitrogen-fixing bacterium sinorhizobium meliloti. BMC genomics 12: 235. doi: 10.1186/1471-2164-12-235 PMID: 21569405
24. Isalan M, Lemerle C, Michalodimitrakis K, Horn C, Beltrao P, et al. (2008) Evolvability and hierarchy in rewired bacterial gene networks. Nature 452: 840–845. doi: 10.1038/nature06847 PMID: 18421347
25. Somvanshi VS, Sloup RE, Crawford JM, Martin AR, Heidt AJ, et al. (2012) A single promoter inversion switches photorhabdus between pathogenic and mutualistic states. Science 337: 88–93. doi: 10.1126/science.1216641 PMID: 22767929
26. Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental escherichia coli population. Nature 489: 513–518. doi: 10.1038/nature11514 PMID: 22992527

27. Krol E, Blom J, Winnebald J, Berhörster A, Barnett MJ, et al. (2011) Rhizoregneta database of rhizobial transcription factors and regulatory networks. Journal of biotechnology 155: 127–134. doi: 10.1016/j.jbiotec.2010.11.004 PMID: 21087643

28. Schlüter JP, Reinkensmeier J, Barnett MJ, Lang C, Krol E, et al. (2013) Global mapping of transcription start sites and promoter motifs in the symbiotic α-proteobacterium sinorhizobium meliloti 1021. BMC genomics 14: 156. doi: 10.1186/1471-2164-14-156 PMID: 23497287

29. Harrison PW, Lower RP, Kim NK, Young JPW (2010) Introducing the bacterial chromid: not a chromosome, not a plasmid. Trends in microbiology 18: 141–148. doi: 10.1016/j.tim.2009.12.010 PMID: 20080407

30. Galibert F, Finan TM, Long SR, Pühler A, Abola P, et al. (2001) The composite genome of the legume symbiont sinorhizobium meliloti. Science 293: 668–672. doi: 10.1126/science.1060966 PMID: 11474104

31. Galardini M, Pini F, Bazzicalupo M, Biondi EG, Mengoni A (2013) Replicon-dependent bacterial genome evolution: the case of sinorhizobium meliloti. Genome biology and evolution 5: 542–558. doi: 10.1093/gbe/evt027 PMID: 23431003

32. diCenzo GC, MacLean AM, Milunovic B, Golding GB, Finan TM (2014) Examination of prokaryotic multipartite genome evolution through experimental genome reduction. PLoS genetics 10: e1004742. doi: 10.1371/journal.pgen.1004742 PMID: 25340565

33. Schneiker-Bekel S, Wibberg D, Bekel T, Blom J, Linke B, et al. (2011) The complete genome sequence of the dominant *Sinorhizobium meliloti* field isolate sm11 extends the *S. meliloti* pan-genome. Journal of biotechnology 155: 20–33. doi: 10.1016/j.jbiotec.2010.12.018 PMID: 21396969

34. Li Z, Ma Z, Hao X, Wei G (2012) Draft genome sequence of sinorhizobium meliloti ccnwsx0020, a nitrogen-fixing symbiont with copper tolerance capability isolated from lead-zinc mine tailings. Journal of bacteriology 194: 1267–1268. doi: 10.1128/JB.06682-11 PMID: 22328762

35. Sugawara M, Epstein B, Badgley B, Unno T, Xu L, et al. (2013) Comparative genomics of the core and accessory genomes of 48 sinorhizobium strains comprising five genospecies. Genome biology 14: R17. doi: 10.1186/gb-2013-14-2-r17 PMID: 23425606

36. Weidner S, Baumgarth B, Göttfert M, Jaenicke S, Pühler A, et al. (2013) Genome sequence of sinorhizobium meliloti rm41. Genome announcements 1: e00013–12. doi: 10.1128/genomeA.00013-12 PMID: 23405285

37. Martínez-Abarca F, Martínez-Rodríguez L, López-Contreras JA, Jiménez-Zurdo JI, Toro N (2013) Complete genome sequence of the alfalfa symbiont sinorhizobium/ensifer meliloti strain gr4. Genome announcements 1: e00174–12. doi: 10.1128/genomeA.00174-12 PMID: 23409262

38. Sallet E, Roux B, Sauviac L, Carrère S, Faraut T, et al. (2013) Next-generation annotation of prokaryotic genomes with eugene-p: application to sinorhizobium meliloti 2011. DNA research 20: 339–354. doi: 10.1093/dnares/dst014 PMID: 23599422

39. Galardini M, Bazzicalupo M, Mengoni A, Biondi E, Brambilla E, et al. (2013) Permanent draft genome sequences of the symbiotic nitrogen fixing ensifer meliloti strains bo21cc and ak58. Standards in Genomic Sciences 9. doi: 10.4056/sigs.3797438 PMID: 24976889

40. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. Current opinion in genetics & development 15: 589–594. doi: 10.1016/j.gde.2005.09.006

41. Charoensawan V, Wilson D, Teichmann SA (2010) Genomic repertoires of dna-binding transcription factors across the tree of life. Nucleic acids research 38: 7364–7377. doi: 10.1093/nar/gkq617 PMID: 20675356

42. Pini F, Galardini M, Bazzicalupo M, Mengoni A (2011) Plant-bacteria association and symbiosis: are there common genomic traits in alphaproteobacteria? Genes 2: 1017–1032. doi: 10.3390/genes2041017 PMID: 24710303

43. Lobkovsky AE, Wolf YI, Koonin EV (2013) Gene frequency distributions reject a neutral model of genome evolution. Genome biology and evolution 5: 233–242. doi: 10.1093/gbe/evt002 PMID: 23315380

44. Lynch D, O'Brien J, Welch T, Clarke P, ÓCuiv P, et al. (2001) Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by sinorhizobium meliloti. Journal of bacteriology 183: 2576–2585. doi: 10.1128/JB.183.8.2576-2585.2001 PMID: 11274118

45. Gill P Jr, Barton L, Scoble M, Neilands J (1991) A high-affinity iron transport system of rhizobium meliloti may be required for efficient nitrogen fixation in planta. Plant and Soil 130: 211–217. doi: 10.1007/BF00011875

46. Chao TC, Buhrmester J, Hansmeier N, Pühler A, Weidner S (2005) Role of the regulatory gene rirA in the transcriptional response of sinorhizobium meliloti to iron limitation. Applied and environmental microbiology 71: 5969–5982. doi: 10.1128/AEM.71.10.5969-5982.2005 PMID: 16204511

47. Bobik C, Meilhoc E, Batut J (2006) Fixj: a major regulator of the oxygen limitation response and late symbiotic functions of sinorhizobium meliloti. Journal of bacteriology 188: 4890–4902. doi: 10.1128/JB.00251-06 PMID: 16788198

48. Ferrières L, Kahn D (2002) Two distinct classes of fixj binding sites defined by in vitro selection. FEBS letters 517: 185–189. doi: 10.1016/S0014-5793(02)02618-2 PMID: 12062434

49. Hoang HH, Gurich N, González JE (2008) Regulation of motility by the expr/sin quorum-sensing system in sinorhizobium meliloti. Journal of bacteriology 190: 861–871. doi: 10.1128/JB.01310-07 PMID: 18024512

50. Meysman P, Sonego P, Bianco L, Fu Q, Ledezma-Tejeida D, et al. (2014) Colombos v2. 0: an ever expanding collection of bacterial expression compendia. Nucleic acids research 42: D649–D653. doi: 10.1093/nar/gkt1086 PMID: 24214998

51. Quinn HJ, Cameron AD, Dorman CJ (2014) Bacterial regulon evolution: Distinct responses and roles for the identical ompr proteins of salmonella typhimurium and escherichia coli in the acid stress response. PLoS genetics 10: e1004215. doi: 10.1371/journal.pgen.1004215 PMID: 24603618

52. Cavalieri D, Townsend JP, Hartl DL (2000) Manifold anomalies in gene expression in a vineyard isolate of saccharomyces cerevisiae revealed by dna microarray analysis. Proceedings of the National Academy of Sciences 97: 12369–12374. doi: 10.1073/pnas.210395297

53. Kvitek DJ, Will JL, Gasch AP (2008) Variations in stress sensitivity and genomic expression in diverse s. cerevisiae isolates. PLoS genetics 4: e1000223. doi: 10.1371/journal.pgen.1000223 PMID: 18927628

54. Galardini M, Mengoni A, Biondi EG, Semeraro R, Florio A, et al. (2014) Ductape: A suite for the analysis and correlation of genomic and omnilog phenotype microarray data. Genomics 103: 1–10. doi: 10.1016/j.ygeno.2013.11.005 PMID: 24316132

55. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215: 403–410. doi: 10.1016/S0022-2836(05)80360-2 PMID: 2231712

56. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, et al. (2007) Primer3plus, an enhanced web interface to primer3. Nucleic acids research 35: W71–W74. doi: 10.1093/nar/gkm306 PMID: 17485472

57. Kazakov AE, Cipriano MJ, Novichkov PS, Minovitsky S, Vinogradov DV, et al. (2007) Regtransbasea database of regulatory sequences and interactions in a wide range of prokaryotic genomes. Nucleic acids research 35: D407–D412. doi: 10.1093/nar/gkl865 PMID: 17142223

58. Berg OG, von Hippel PH (1987) Selection of dna binding sites by regulatory proteins: Statistical-mechanical theory and application to operators and promoters. Journal of molecular biology 193: 723–743. doi: 10.1016/0022-2836(87)90354-8 PMID: 3612791

59. Nishida K, Frith MC, Nakai K (2009) Pseudocounts for transcription factor binding sites. Nucleic acids research 37: 939–944. doi: 10.1093/nar/gkn1019 PMID: 19106141

60. d'Agostino RB (1971) An omnibus test of normality for moderate and large size samples. Biometrika 58: 341–348. doi: 10.1093/biomet/58.2.341

61. Jones E, Oliphant T, Peterson P (2001) Scipy: Open source scientific tools for python. http://www.scipy.org/.

62. Hojo T, Pearson K (1931) Distribution of the median, quartiles and interquartile distance in samples from a normal population. Biometrika 23: 315–363. doi: 10.2307/2332422

63. Karunakaran R, Mauchline T, Hosie AH, Poole PS (2005) A family of promoter probe vectors incorporating autofluorescent and chromogenic reporter proteins for studying gene expression in gram-negative bacteria. Microbiology 151: 3249–3256. doi: 10.1099/mic.0.28311-0 PMID: 16207908

64. Pini F, Spini G, Galardini M, Bazzicalupo M, Benedetti A, et al. (2014) Molecular phylogeny of the nickel-resistance gene nreb and functional role in the nickel sensitive symbiotic nitrogen fixing bacterium sinorhizobium meliloti. Plant and Soil 377: 189–201. doi: 10.1007/s11104-013-1979-3

65. Pini F, Frage B, Ferri L, De Nisco NJ, Mohapatra SS, et al. (2013) The divj, cbra and plec system controls divk phosphorylation and symbiosis in sinorhizobium meliloti. Molecular microbiology 90: 54–71. doi: 10.1111/mmi.12347 PMID: 23909720

66. Westover BP, Buhler JD, Sonnenburg JL, Gordon JI (2005) Operon prediction without a training set. Bioinformatics 21: 880–888. doi: 10.1093/bioinformatics/bti123 PMID: 15539453

67. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A, et al. (2011) Contiguator: a bacterial genomes finishing tool for structural insights on draft genomes. Source code for biology and medicine 6. doi: 10.1186/1751-0473-6-11 PMID: 21693004

68. Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. Journal of the American statistical association 58: 236–244. doi: 10.1080/01621459.1963.10500845

69. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. (2011) Scikit-learn: Machine learning in python. The Journal of Machine Learning Research 12: 2825–2830.

70. Sukumaran J, Holder MT (2010) Dendropy: a python library for phylogenetic computing. Bioinformatics 26: 1569–1571. doi: 10.1093/bioinformatics/btq228 PMID: 20421198

71. Edgar RC (2004) Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32: 1792–1797. doi: 10.1093/nar/gkh340 PMID: 15034147

72. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Systematic biology 61: 539–542. doi: 10.1093/sysbio/sys029 PMID: 22357727

73. Talevich E, Invergo BM, Cock PJ, Chapman BA (2012) Bio. phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in biopython. BMC bioinformatics 13: 209. doi: 10.1186/1471-2105-13-209 PMID: 22909249

74. Jelinek F, Bahl L, Mercer R (1975) Design of a linguistic statistical decoder for the recognition of continuous speech. Information Theory, IEEE Transactions on 21: 250–256. doi: 10.1109/TIT.1975.1055384

75. Van Der Walt S, Colbert SC, Varoquaux G (2011) The numpy array: a structure for efficient numerical computation. Computing in Science & Engineering 13: 22–30. doi: 10.1109/MCSE.2011.37

76. Hunter JD (2007) Matplotlib: A 2d graphics environment. Computing in Science & Engineering: 90–95. doi: 10.1109/MCSE.2007.55

77. Perez F, Granger BE (2007) Ipython: a system for interactive scientific computing. Computing in Science & Engineering 9: 21–29. doi: 10.1109/MCSE.2007.53

78. The networkx python library. URL http://networkx.github.io/.

79. Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. In: ICWSM.

ADDITIONAL FILES

Additional file, Material S1:
Inter and intra-regulation in the 51 *S. meliloti* srains. (ZIP)

Additional file, Material S2:
Single regulons correlations with the COLOMBOS expression compendium. (ZIP)

Additional file, Table S1:
Sources and information content of the TF PSSM of this study.

Additional file, Table S2:
Experimental validation of NodD targets

Additional file, Table S3:
State transitions probability for the regulatory networks.

Additional file, Table S4:
Genomic sequences used in this study.

Additional file, Table S5:
Predicted operons statistics.

Additional file, Fig. S1:
Total TFs encoded in the pangenome.

a) TFs frequency (expressed as the number of strains having the TF encoded in their genome) in *S. meliloti* and the other rhizobial genomes; b) TF presence/absence matrix in the strains analyzed in this study: red indicates the TF absence. TFs are colored according to the replicon they belong to: red for chromosome, green for the pSymA megaplasmid and blue for the pSymB chromid.

Additional file, Fig. S2:

Correlation between predictions quality and TF information content.

Vertical dashed line indicates the minimum information content for *S. meliloti* strain Rm1021. a) Correlation between predictions true positive rate and information content; b) Correlation between the number of predicted regulated genes and information content.

Additional file, Fig. S3:

COG categories enrichment in the replicons.

For each replicon, the proportion of regulated downstream genes belonging to each category is compared with the genes belonging to other replicons. Purple categories indicate a statistically significant enrichment.

Additional file, Fig. S4:

Replicon mapping quality control.

For each orthologous group in the *S. meliloti* pangenome, the abundance of the most mapped replicons has been computed as a proxy for the consistency of the replicon mappings.