Article. Section Discoveries.

Recruitment of a lineage-specific virulence regulatory pathway promotes intracellular infection by a plant pathogen experimentally evolved into a legume symbiont

Delphine Capela[1]*, Marta Marchetti[1], Camille Clérissi[1,2,3], Anthony Perrier[1], Dorian Guetta[1], Carine Gris[1], Marc Valls[4], Alain Jauneau[5], Stéphane Cruveiller[6], Eduardo P. C. Rocha[2,3] and Catherine Masson-Boivin[1]*

[1] LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan cedex, France

[2] Microbial Evolutionary Genomics, Institut Pasteur, 25-28 rue Dr Roux, 75015 Paris, France

[3] CNRS, UMR3525, 25-28 rue Dr Roux, 75015 Paris, France

[4] Department of Genetics, University of Barcelona and Centre for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB), Edifici CRAG, Campus UAB, 08193 Bellaterra, Spain

[5] Fédération de Recherches Agrobiosciences, Interactions, Biodiversity, Plateforme d'Imagerie TRI, CNRS, UPS, 31326 Castanet-Tolosan cedex, France

[6] CNRS-UMR8030 and Commissariat à l'Energie Atomique et aux Energies alternatives CEA/DRF/IG/GEN LABGeM, 91057 Evry Cedex, France

*Corresponding authors

Catherine.Masson@inra.fr

Delphine.Capela@inra.fr

ABSTRACT

Ecological transitions between different lifestyles, such as pathogenicity, mutualism and saprophytism, have been very frequent in the course of microbial evolution, and often driven by horizontal gene transfer. Yet, how genomes achieve the ecological transition initiated by the transfer of complex biological traits remains poorly known. Here we used experimental evolution, genomics, transcriptomics and high-resolution phenotyping to analyze the evolution of the plant pathogen *Ralstonia solanacearum* into legume symbionts, following the transfer of a natural plasmid encoding the essential mutualistic genes. We show that a regulatory pathway of the recipient *R. solanacearum* genome involved in extracellular infection of natural hosts was reused to improve intracellular symbiosis with the *Mimosa pudica* legume. Optimization of intracellular infection capacity was gained through mutations affecting two components of a new regulatory pathway, the transcriptional regulator *efpR* and a region upstream from the RSc0965-0967 genes of unknown functions. Adaptive mutations caused the downregulation of *efpR* and the over-expression of a downstream regulatory module, the three unknown genes RSc3146-3148, two of which encoding proteins likely associated to the membrane. This over-expression led to important metabolic and transcriptomic changes and a drastic qualitative and quantitative improvement of nodule intracellular infection. In addition, these adaptive mutations decreased the virulence of the original pathogen. The complete *efpR*/RSc3146-3148 pathway could only be identified in the genomes of the pathogenic *R. solanacearum* species complex. Our findings illustrate how the rewiring of a genetic network regulating virulence allows a radically different type of symbiotic interaction and contributes to ecological transitions and trade-offs.

Key words:

Adaptive evolution, experimental evolution, horizontal gene transfer, rhizobium, symbiosis, pathogenicity

Abbreviations

BBH: Bidirectional Best Hit

EPS: exopolysaccharides

HGT: Horizontal Gene Transfer

HSI: Hue Saturation Intensity

MGE: Mobile Genetic Element

PBP: Periplasmic ligand-Binding Protein

qRT-PCR: quantitative Reverse Transcription Polymerase Chain Reaction

RSI: ExoR-ExoS-ChvI

SNP: Single Nucleotide Polymorphism

WGS: Whole Genome Sequencing

XRE: Xenobiotic Response Element

INTRODUCTION

Ecological transitions are a major source of biodiversity. Phylogenetic reconstructions indicate that transitions between very diverse lifestyles, such as parasitism or saprophytism towards mutualism, have been very frequent during the course of evolution (Sachs, et al. 2014). Facultative mutualistic or pathogenic endosymbionts, which alternate saprophytic and symbiotic lifestyles, often evolved through acquisition of mobile genetic elements (MGE) such as plasmids and genomic islands via horizontal gene transfer (HGT) (Sullivan and Ronson 1998; Ochman and Moran 2001; Dobrindt, et al. 2004). Acquisition of large MGE can provide complex novel traits in a single event of transfer (Frost, et al. 2005; Rankin, et al. 2011). The acquisition of such complex functions can itself open up new ecological niches (Wiedenbeck and Cohan 2011). Yet profound changes in lifestyle may require, beyond phenotypic expression of acquired traits, the recruitment of local regulatory and metabolic functions to fully thrive in the new environment (Pál, et al. 2005; Lercher and Pál 2008). Full integration of horizontally transferred genes in the new genetic background may take hundreds to millions of generations (Pál, et al. 2005; Lercher and Pál 2008).

Rhizobia, the nitrogen-fixing mutualistic symbionts of legumes, are an excellent biological system to investigate the molecular mechanisms underlying post-HGT adaptation. These bacteria are facultative endosymbionts that alternate between a saprophytic life in the soil and a symbiotic life when they encounter a compatible legume host (Masson-Boivin, et al. 2009). Rhizobia are phylogenetically diverse bacteria; they include hundreds of species in 14 saprophyte- and/or pathogen-containing genera of α- and β-proteobacteria (Masson-Boivin, et al. 2009; Remigi 2016; Remigi, et al. 2016). Most importantly, horizontal transfer of essential symbiotic genes, the *nod* and *nif* genes that are carried on plasmids or genomic islands, has played a prominent role in spreading symbiotic proficiency among diverse soil bacteria (Sullivan, et al. 1995; Nandasena, et al. 2007). The transition to a mutualistic lifestyle may require the neutralization of bacterial functions that impair the symbiotic process, such as type three secretion systems or peptidases (Marchetti, et al. 2010; Crook, et al. 2012; Price, et al. 2015). Mutualism may also require the recruitment of endogenous traits of the recipient genome (Amadou, et al. 2008; Tian, et al. 2012). These include, for example, specific exopolysaccharides that are recognized by the host plant for infection (Kawaharada, et al. 2015) or secretion systems that modulate host range (Nelson and Sadowsky 2015). Phylogenetic studies suggest that successful *nod-nif* transfers have been frequent within genera but rare between genera, family and beyond (Remigi, et al. 2016). Current rhizobial genera might thus have features (*e.g*. lineage specific exopolysaccharides), originally unrelated to symbiosis, that have been co-opted to achieve symbiosis with the legume hosts. Transfers between such pre-adapted genomes are more likely to be successful, because the trait may rapidly evolve to be efficiently expressed. In contrast, symbiotic transfers across wide phylogenetic distances may be unproductive unless the recipient

bacteria undergo further adaptations to go down the symbiotic route and adapt to the specific requirements, *e.g*. in terms of immunity and metabolism, of the new host.

To analyze post-HGT genomic adaptive changes, we previously launched the experimental evolution of the plant pathogen *Ralstonia solanacearum* into a legume symbiont, building on the natural evolutionary history of rhizobia. We transferred the symbiotic plasmid pRalta of the *Mimosa* symbiont *Cupriavidus taiwanensis* into *R. solanacearum* GMI1000 and further evolved this chimera (GMI1000 pRalta) using serial *ex planta-in planta* (*Mimosa pudica*) cycles (Marchetti, et al. 2010; Marchetti, et al. 2014) (supplementary fig. S1). The two β-proteobacterial protagonists of the experiment - *R. solanacearum* and *C. taiwanensis* - interact with plants in a drastically different way. *R. solanacearum* is a typical plant pathogen that penetrates plantlets and moves via the intercellular spaces to invade the vascular system, where excessive production of extracellular polysaccharides causes plant wilting (Genin 2010). In contrast, *C. taiwanensis* establishes mutualistic interactions with *Mimosa* spp. For this, it induces the formation of root nodules that it colonizes intracellularly. Each infected plant cell houses thousands of internalized bacteria that fix atmospheric nitrogen, to the benefit of the plant (Chen, et al. 2003; Marchetti, et al. 2011). The experimental setting, providing the chimera to repeatedly colonize the plant, together with the presence on pRalta of stress-responsive error-prone DNA polymerases that increase recipient genome evolvability (Remigi, et al. 2014), favored the emergence and fixation of subpopulations with improved nodulation and nodule infection capacities (Marchetti, et al. 2014). Although the original chimeric *Ralstonia* was still pathogenic to *A. thaliana* and unable to nodulate *M. pudica*, it progressively turned into a legume symbiont during the lab-evolution experiment conducted over *ca.* 400 generations (Marchetti, et al. 2014) (supplementary fig. S1). We previously demonstrated that the capacity to nodulate *M. pudica* and rudimentarily infect nodule cells was reached via inactivation of main regulatory or structural pathogenicity determinants of the recipient genome (Marchetti, et al. 2010; Guan, et al. 2013). Here we provide evidence for the recruitment of an endogenous virulence regulatory pathway specific to *R. solanacearum* to improve the symbiotic adaptation in *M. pudica*. Drastic improvement in nodule cell infection occurred via mutations in components of this pathway which led to the up-regulation of three target genes exhibiting regulatory and metabolic functions and crucial for symbiosis.

RESULTS

Evolution of symbiotic phenotypes

Previous phenotypic analysis showed that 2 out of 3 nodulating clones (CBM212 and CBM349), obtained as a result of a first selection cycle, were able to poorly infect *Mimosa* nodule cells, and that 5 of the 6 final clones derived from these two ancestors following 16 passages on plants, B16, C16,

G16, H16 and I16, had improved their ability to invade the plant cells while reducing the induction of plant defense reactions (Marchetti, et al. 2014) (supplementary fig. S1).

To determine when nodule cell infection was improved, we monitored changes in infection along the five lineages (B, C, G, H and I) by semi-quantitative cytological observations of nodules of *M. pudica* induced by individual intermediate clones. We measured the relative surface of the infected zone on sections of nodules collected 21 days post-inoculation, using the Hue, Saturation and Intensity (HSI) method that allows distinguishing infected from non-infected areas based on their color (Marchetti, et al. 2014). Significant up-shifts in the levels of intracellular infection were observed between the clone pairs B8-B9, C5-C6, G1-G2, H6-H7, and I1-I2 (fig. 1A and supplementary fig. S2). The surface of the infection zone, which roughly corresponds to the intracellular infection zone (Marchetti, et al. 2014), increased by 1.5-3 fold in these cases. The surface of the necrotic zones in nodules, which is an indication of the induction of plant defenses, significantly decreased in the clones improved for infection B9, C6, G2, H7, and I2 as compared to B8, C5, G1, H6, and I1, respectively (fig. 1B and supplementary fig. S2), indicating a simultaneous decrease in defense reactions to the improvement of intracellular infection.

Hence activation of a first level of intracellular infection in ancestral nodulating clones (Marchetti, et al. 2010) was followed by a second up-shift in infection capacity, *i.e.* an improvement of the quantity and quality of intracellular infection, that occurred in the B, C, G, H and I lines.

Identification of mutations that improve intracellular infection capacity

To identify the mutations associated with the phenotypic shifts, we re-sequenced the intermediate clones B8, B9, C5, C6, G1, G2, H6, H7, I1 and I2. Sequence data were mapped to the ancestral reference genome (6.37 Mb) based on the known genome sequences of *R. solanacearum* GMI1000 (Salanoubat, et al. 2002) and *C. taiwanensis* LMG19424 (Amadou, et al. 2008) and analyzed using the PALOMA bioinformatic pipeline (Vallenet, et al. 2006; Vallenet, et al. 2013). Supplementary table S1 lists all the SNPs (single nucleotide polymorphisms) and indels detected in each improved clone (n) relative to the n-1 clone in the same lineage.

In C, G, H and I lineages, a non-synonymous mutation affecting the *efpR* transcriptional regulator was identified. In clones G2, H7 and I2, *efpR* was altered by the same SNP leading to the E66K modification of the protein while another SNP leading to the P63S modification of the EfpR protein was found in clone C6, indicating a strong parallel evolution (table 1). To evaluate the possible role of *efpR* mutations on intracellular infection we reconstructed the *efpR*E66K mutation in clone G1 as well as in the ancestral chimeric *Ralstonia* background (GMI1000 pRalta) bearing the nodulation-conferring *hrpG* mutation by using the MuGENT method (Dalia, et al. 2014). The *efpR*E66K mutants (in G1 and GMI1000 pRalta *hrpG*) induced nodules clearly different from those induced by G1 or

GMI1000 pRalta *hrpG* and comparable to those induced by the clone G2 relative to the size of the intracellularly infected and necrotic areas (fig. 1, table 2).

We detected 30 point-mutations in clone B9 as compared to clone B8. Interestingly, none of them affected a gene mutated in the C, G, H or I lineages. We reconstructed in clone B8 and in the nodulating mutant chimera GMI1000 pRalta *hrpG* the non-synonymous or intergenic mutations detected in B9. Among those, B8 and GMI1000 pRalta *hrpG* carrying the G/A SNP mutation located in the RALSOc_1033-RSc0965 intergenic region (fig. 2), 115 nucleotides upstream from the start codon of the RSc0965 gene (B8 up^{115}-RSc0965 and GMI1000 pRalta *hrpG* up^{115}-RSc0965), induced nodules comparable to those induced by B9 (fig. 1, table 2). The inactivation of RSc0965, but not the inactivation of RALSOc_1033, abolished the phenotype of GMI1000 pRalta *hrpG* up^{115}-RSc0965, indicating that the RSc0965 gene is putatively involved in infection improvement (fig. 1).

These results establish clearly that mutations in *efpR* and in the intergenic region upstream from RSc0965 had a strong effect both on infection capacity and reduction of defense reactions. We then tested if the effect of these mutations depended on the genetic background given by the *hrpG* stop mutation. The reconstruction of the *efpR*E66K and the up^{115}-RSc0965 mutations in GMI1000 pRalta did not allow the original chimeric strain GMI1000 pRalta to nodulate *M. pudica* indicating that the *hrpG* stop mutation is epistatic to these mutations. Finally, we made co-inoculation experiments of *M. pudica* plants and confirmed that G1 *efpR*E66K and B8 up^{115}-RSc0965 are more fit *in planta* than G1 and B8, respectively (fig. 3). Yet these mutants were found less fit than G2 and B9, respectively (fig. 3), indicating that there are unknown additional adaptive mutations in G2 and B9.

Adaptive mutations affect the same regulatory pathway

EfpR belongs to the xenobiotic response element (XRE) family of transcriptional regulators (Guidot, et al. 2014; Perrier, et al. 2016). The two mutations, *efpR*P63S and *efpR*E66K, map in the HTH domain of the protein, in the turn region and the beginning of the recognition helix, respectively. This suggests that they affect the protein-DNA interaction. The up^{115}-RSc0965 mutation is located upstream from an operon with three genes, RSc0965, RSc0966 and RSc0967 (fig. 2), as shown by RT-PCR experiments (supplementary fig. S3). These genes encode proteins of unknown function and are located in the vicinity of genes encoding proteins with homology to bacteriophage genes, including a putative integrase (RSc0968), and an ^{Arg}tRNA (RS04383).

To evaluate whether the mutations in *efpR* or between RALSOc_1033 and the RSc0965-0967 operon affect the expression of these genes, we measured the expression of *efpR,* RALSOc_1033 and RSc0967 by qRT-PCR in different strains cultivated in minimal medium containing glutamate as sole carbon source, a medium that mimics the plant apoplast (Rahme, et al. 1992; Tang, et al. 2006). The up^{115}-RSc0965 mutation did not modify the expression of the RALSOc_1033 gene in the nodulating

mutant chimera GMI1000 pRalta *hrpG* (fig. S4A). Instead it triggered an increase in RSc0967 expression in GMI1000 pRalta *hrpG* and in the B lineage (fig. 4). The induction of RSc0967 was not observed in *efpR* mutants (Fig. 4). On the other hand the three mutations *efpR*E66K*, efpR*P63S and up^{115}-RSc0965 decreased *efpR* gene expression in GMI1000 pRalta *hrpG*. The expression of *efpR* was also drastically reduced in improved clones of the B, C, G, H and I lineages as compared to their respective n-1 clones, supporting the fact that the mutations *efpR*E66K, *efpR*P63S and up^{115}-RSc0965 reduce the expression of *efpR* in the evolved backgrounds (fig. 4). The decrease in *efpR* expression observed in GMI1000 pRalta *hrpG* up^{115}-RSc0965 mutant was abolished by the inactivation of RSc0965-0966 but not by the inactivation of RSc0967 or RALSOc_1033 (fig. S4B). Whether the repression of *efpR* in the up^{115}-RSc0965 mutants is linked to the over-expression of the RSc0965-0966 genes or involves a regulatory non-coding RNA located in this region was not investigated in this study.

Overall these results showed that the up^{115}-RSc0965, *efpR*E66K and *efpR*P63S mutations all lead to the down-regulation of *efpR*. This suggests that they all affect the same regulatory pathway.

EfpR controls the transcription of a wide set of genes through the repression of the intermediary RSc3146-3148 genes

In the *R. solanacearum* GMI1000 strain, EfpR was previously shown to be a global regulator controlling several virulence traits (Perrier, et al. 2016). In order to identify the functions controlled by EfpR possibly involved in the improvement of nodule infection ability of our evolved clones, we compared the transcriptome of the nodulating mutant chimera GMI1000 pRalta *hrpG* with that of its derivative *efpR*E66K mutant (GMI1000 pRalta *hrpG efpR*E66K) following cultivation in minimal medium with glutamate as sole carbon source. A total of 200 genes, almost exclusively belonging to the *R. solanacearum* recipient genome, were found more than 2-fold differentially expressed between the two strains (*P* value (*t*-test) <0.05), 124 genes being up-regulated and 76 being down-regulated in the mutant (table 3 and supplementary table S2). Only 44 of these genes were also found differentially expressed in the study of Perrier et al. suggesting that the genetic background deeply influences the *efpR* transcriptome. In our experiments, 52 genes involved in flagellar biosynthesis and motility, together with the *flhDC* associated regulatory genes (Tans-Kersten, et al. 2004), are strikingly induced in the *efpR* mutant while 17 genes involved in exopolysaccharides (EPS) biosynthesis, including their *xpsR* transcription regulator (Huang, et al. 1995), are strongly repressed. The expression of other virulence genes was found positively controlled by EfpR. These include genes encoding Hrp proteins of the type III secretion system (T3SS) and T3 effectors (table 3).

Around 30% of the putative EfpR targets show similarity to poorly characterized genes (table 3). Among the most differentially expressed genes we identified the RSc3146-RSc3147-RSc3148 genes

(fig. 2), which are repressed by EfpR. BlastP search in the UniProt database (Consortium 2015) and InterProScan analyses (Mitchell, et al. 2015) indicated that RSc3148, a 147-residue protein with two transmembrane domains, and RSc3146, a 133-residue putative cytoplasmic protein, displayed no sequence similarity with characterized proteins. RSc3147 possesses a clear signal peptide cleavage site predicted by SignalP (Petersen, et al. 2011) and presents similarities to extracellular or periplasmic ligand-binding proteins (PBP). Quantitative RT-PCR experiments from cultures in minimal medium containing glutamate confirmed the strong induction (700-fold on average) of RSc3148 in *efpR*E66K, *efpR*P63S, $efpR^{KO}$ and up^{115}-RSc0965 mutants as well as in the evolved clones with enhanced infection capacities (fig. 4).

We enquired on the role of the genes RSc3146-3148 in the regulatory network downstream of EfpR. Comparisons of the transcriptomes of the nodulating mutant chimera GMI1000 pRalta *hrpG* and its *efpR*E66K derivative either deleted for the three RSc3146-3148 genes or inactivated for the RSc3147 gene only (GMI1000 pRalta *hrpG efpR*E66K Δ(RSc3146-3148) or GMI1000 pRalta *hrpG efpR*E66K RSc3147::Spe) indicated that these genes are important intermediates of the regulatory pathway. Indeed 90% of gene expression changes observed in the *efpR* mutant, including changes in EPS, motility, and T3S functions, are lost in the *efpR* mutants inactivated for RSc3147 or deleted for the three RSc3146-3148 genes (table 3 and supplementary table S2). This suggests that transcriptomic modifications associated with the *efpR*E66K mutant are mainly due to the de-repression of the RSc3146-3148 genes (fig. 5).

EfpR represses catabolism through the repression of the intermediary RSc3146-3148 genes

EfpR was described to act as a catabolic repressor (Perrier, et al. 2016). To investigate the impact of the *efpR*E66K adaptive mutation on the metabolism of the nodulating mutant *Ralstonia* chimera we characterized the metabolic profile of GMI1000 pRalta *hrpG efpR*E66K mutant in comparison to GMI1000 pRalta *hrpG* using Biolog phenotype microarrays. Four microarrays dedicated to the catabolism of carbon, nitrogen, phosphorus and sulfur sources (PM1-PM4) were assessed in two independent experiments for each strain. We found that the *efpR*E66K mutation allowed a better utilization of 62 different substrates including 12 carbon, 7 nitrogen and 43 phosphorus sources (table 4). Among the most differentially catabolized substrates, three amino acids, L-histidine, L-alanine and L-serine, and one dicarboxylic acid, the alpha-keto glutarate, were better used as carbon sources in the *efpR* mutant (>3-fold). The utilization of phosphate from a large number of organic compounds was also improved in the *efpR* mutant. These organic compounds include hexose phosphate, amino acid phosphate but also important intracellular second messengers such as the cyclic nucleotides cCMP, cGMP and cUMP. Most notably utilization of phosphoenol pyruvate, the phosphoryl donor for the first enzyme of the phosphotranferase system (PTS), was strongly

enhanced (3.45 fold) in the *efpR* mutant. Interestingly, the 62 metabolic activities that were enhanced in the *efpR* mutant were lost either when the RSc3146-3148 genes were all deleted or when the RSc3147 gene only was inactivated (*i.e*, in strains GMI1000 pRalta *hrpG efpR*E66K Δ(RSc3146-3148) and GMI1000 pRalta *hrpG efpR*E66K RSc3147::Spe) (table 4, supplementary fig. S5). This result suggests that the products of this gene cluster play a role as metabolic activators repressed by EfpR.

RSc3146-3148 genes are essential for optimized infection

To investigate the contribution of RSc3146-3148 genes in *M. pudica* nodule infection, we quantified the infected and necrotic areas of nodule sections induced by the nodulating chimeric *Ralstonia* mutated in *efpR* and deleted for RSc3146-3148 or inactivated for RSc3147 (GMI1000 pRalta *hrpG efpR*E66K Δ(RSc3146-3148) and GMI1000 pRalta *hrpG efpR*E66K RSc3147::Spe). The deletion of RSc3146-3148 or inactivation of RSc3147 abolished the phenotypic changes induced by the *efpR*E66K mutation in GMI1000 pRalta *hrpG* (fig. 1, table 2). The sizes of the nodule infection and necrotic zones of the double mutants (*efpR*E66K Δ(RSc3146-3148) and *efpR*E66K RSc3147::Spe) of GMI1000 pRalta *hrpG* were similar to those measured for the GMI1000 pRalta *hrpG* strain. This demonstrated that the induction of the RSc3146-3148 genes is required for the *efpR* improved infection phenotype. Using a genomic fusion of the putative promoter region of RSc3148 with the fluorescent mCherry reporter gene, we confirmed that RSc3148 was expressed in invaded cells of nodules formed by B9 but not in nodules formed by B8 at 21 dpi (fig. 6). This fusion was also found expressed in B9 grown on rich medium (fig. 6), suggesting a constitutive high-level of expression of the RSc3146-3148 genes when *efpR* is down-regulated.

The RSc0965-0966/EfpR/RSc3146-3148 pathway is involved in virulence

Mutations inactivating *efpR* in the pathogenic *R. solanacearum* GMI1000 strain were previously shown to enhance bacterial fitness within vascular tissues on one hand and to decrease virulence on the other hand (Guidot, et al. 2014; Perrier, et al. 2016). To evaluate whether the adaptive mutations for symbiotic intracellular infection affected the capacity of *R. solanacearum* to trigger disease, we introduced the *efpR*E66K and up^{115}-RSc0965 mutations into the pathogenic wild-type chimeric strain GMI1000 pRalta and inoculated the generated mutants to the *R. solanacearum* host plant *Arabidopsis thaliana.* Disease symptoms were followed over three weeks using classical wilt symptom scoring and the Kaplan-Meier survival analyses (fig. 7). The GMI1000 pRalta up^{115}-RSc0965 mutant was avirulent, inducing almost no symptom. The GMI1000 pRalta *efpR*E66K mutant induced symptoms delayed by about 3 days as compared to the wild-type GMI1000 pRalta strain. To evaluate whether the increased expression of RSc3146-3148 was responsible for the effect of these mutations

on virulence, the GMI1000 pRalta *efpR*E66K and GMI1000 pRalta up^{115}-RSc0965 strains were deleted for the RSc3146-3148 genes or inactivated for RSc3147 and assessed on *A. thaliana*. The deletions of RSc3146-3148 or RSc3147 in the *efpR* mutant led to a full virulence phenotype (fig. 7A, table 2) while the deletions of RSc3146-3148 or RSc3147 in the up^{115}-RSc0965 mutant partially restored the wild-type virulence phenotype (fig. 7B, table 2). Thus, over-expression of the three RSc3146-3148 genes is partly responsible for the avirulent phenotype of the GMI1000 pRalta up^{115}-RSc0965 mutant and entirely responsible for the delayed pathogenicity phenotype of the GMI1000 pRalta *efpR*E66K mutant.

The RSc0965-0967/EfpR/RSc3146-3148 pathway is lineage specific

We showed that the three sets of virulence-associated genes studied in this work were important for the experimental evolution of symbiosis with legumes in *R. solanacearum*. We then studied their distribution among bacteria. We started by searching for homologs of *efpR*, RSc0965-RSc0967 and RSc3146-3148 in the genomes of α- and β-proteobacteria and their phages. We found many (432 genomes) homologs of *efpR*, but no occurrence of the other operons (with the exception of a pair of RSc0965-RSc0966 in two genomes of *Acetobacter* and a pair of RSc0966-RSc0967 in the genome of *Cupriavidus pinatubonensis* JMP134) outside the *Ralstonia* branch (supplementary table S3). Hence, RSc0965-RSc0967 and RSc3146-3148 seem to be strictly limited to *Ralstonia*.

We then searched for the presence of orthologs, homologs, and pseudogenes in *Ralstonia* and the sister clade *Cupriavidus*. With very few exceptions, the orthologs of these genes were only found in the *R. solanacearum* complex (fig. 8) (Prior, et al. 2016). The genes *efpR,* RSc3147 and RSc3148 are ubiquitous in the complex, whereas RSc0965 has very few orthologs. The other genes have more scattered patterns of presence and absence. We identified two putative pseudogenizations of RSc3146 (fig. 8 and supplementary table S3). To test the hypothesis that several of these genes were present in the last common ancestor of the complex, we used Count to reconstruct by maximum likelihood the scenario of the patterns of presence (orthologs and pseudogenes) and absence of these genes in the tree. This analysis found very high posterior probabilities ($P > 0.99$) that at least *efpR* and RSc3146-3148 were present at the last common ancestor of the *R. solanacearum* complex. Altogether these results indicate that *efpR*, RSc0965-0967 and RSc3146-3148 are specific of the *Ralstonia* lineage.

DISCUSSION

Many pathogenic and mutualistic bacteria are non-obligate symbionts that have the ability to switch between alternative ecological niches with extremely different characteristics. Phenotypic plasticity, *i.e.* adaptation to the different niches outside and within the host, is ensured through differential

gene expression in response to specific environmental signals (Soupene, et al. 1995; Valls, et al. 2006). Sharp ecological transitions, such as those initiated by horizontal acquisition of complex plant-associated functions, need the rewiring of regulatory modules to co-opt native functions for the novel trait. Here we experimentally demonstrate the recruitment of a lineage-specific virulence pathway for legume infection during the lab-evolution of the plant pathogen *R. solanacearum* converted into a novel legume symbiont of *Mimosa pudica*.

Infection of nodule cells is a crucial step in the rhizobium-legume symbiosis. Intracellular accommodation shields internalized bacteria from plant defense reactions, while massive bacterial multiplication allows nitrogen fixation levels that sustain plant growth (Berrabah, et al. 2015). In this work, we have detailed how genomic changes associated with improvement of symbiosis in several parallel lineages revealed the recruitment of a *R. solanacearum* regulatory pathway which involves the region upstream the three RSc0965-0967 genes, the XRE-like transcriptional regulator EfpR, and three genes of unknown function, RSc3146-3148 (fig. 5). We showed that the point mutation located 115 nucleotides upstream from the RSc0965-0967 genes directly or indirectly negatively affects *efpR* expression. In turn, the repression of *efpR* directly or indirectly de-represses the expression of the RSc3146-3148 genes. EfpR was previously shown to be a metabolic repressor and a global regulator (Perrier, et al. 2016). Its pleiotropic effects turned out to be mainly due to the control of the RSc3146-3148 gene expression, suggesting that the latter genes act as regulatory intermediates (Fig. 5). RSc3147 encodes a putative periplasmic ligand-binding protein (PBP). Other PBP not linked with ABC transporters were shown to be involved in signal transduction by interacting with sensor proteins (Antoine, et al. 2005; Cockerell, et al. 2014; Sun, et al. 2015). The link between metabolic and gene expression changes induced by the over-expression of RSc3146-3148 genes, together with the environmental cues responsible for the induction of this pathway needs further investigations.

The adaptive mutations upstream RSc0965-0967 and within *efpR* leading to *efpR* down-regulation had a drastic positive effect on infection. This was manifested by a 2-3 fold increase in the size of the nodule infection zone, and a concomitant almost complete disappearance of plant defense reactions. These mutations are adaptive only in a *hrpG* context, *i.e.* when intracellular infection is activated, indicating that symbio-adaptive mutations with an intrinsic biological effect must arise in an appropriate order to be selected in this multifaceted environment and fixed over time. Interestingly *efpR* inactivation was recently shown to improve bacterial multiplication within *R. solanacearum* host plants (Guidot, et al. 2014), suggesting that both inter- and intracellular colonization are repressed by EfpR via unkown mechanisms. We found that the symbiotic effect of the adaptive mutations was entirely due to the induction of the RSc3146-3148 genes. However, because of the pleiotropic effect of the RSc3146-3148 over-expression, elucidating how the recruitment of these genes impacts on symbiosis is challenging. Transcriptome analyses showed that RSc3146-3148-dependent changes

induced by the symbio-adaptive mutations led to lower expression of exopolysaccharide (EPS) biosynthesis genes. Surface polysaccharide structure and synthesis are crucial in symbiosis since they are absolutely required for initiation and elongation of infection threads (Fraysse, et al. 2003; Downie 2010). Rhizobial mutants in EPS production induced non-invaded nodules with symptoms of plant defense reactions on legumes such as *Medicago* or *Lotus* (Jones, et al. 2007; Kawaharada, et al. 2015). Although their role in intracellular infection has not been proved so far, we hypothesize that exopolysaccharides of incompatible structure induce plant defense reactions or confer a higher sensitivity to host antimicrobial compounds. To support this hypothesis, a *Rhizobium leguminosarum pssM* mutant impaired in pyruvylation of EPS was reported to induce the formation of invaded nodules on peas in which symbiosomes rapidly underwent lysis and senesced prematurely (Ivashina, et al. 2010). Along the same line, *R. solanacearum* EPS was shown to induce defense responses in wilt-resistant plants (Milling, et al. 2011). Moreover over-expression of RSc3146-3148 had a huge impact on metabolism, leading to 62 metabolic (C, N or P) gains. Utilization of phosphoenol pyruvate as phosphoryl donor is one of the most enhanced metabolic properties, suggesting a link between these genes and the regulation of the phosphotransferase system. Recently, mutations in *ptsP* and *ptsO*, encoding the two first enzymes of the PTS^{Ntr} in *Sinorhizobium fredii*, were shown to affect symbiosis with soybeans, forming ineffective and poorly invaded nodules (Li, et al. 2016).

Despite the lack of structural and sequence homology, the RSc0965-0967/EfpR/RSc3146-3148 regulatory pathway exhibits functional analogies with the ExoR-ExoS-ChvI Invasion Switch of Rhizobiales (RSI) (Heavner, et al. 2015), an order containing all α-proteobacterial rhizobium species (Remigi, et al. 2016). In the *Medicago* symbiont *Sinorhizobium* (*Ensifer*) *meliloti*, the prototypical member of the Rhizobiales, the RSI regulates metabolism and antagonistically controls succinoglycan and flagellar biosynthesis genes, mediating the switch from flagellated free-living cells to succinoglycan-producing host-invading cells (Yao, et al. 2004; Bélanger, et al. 2009), a situation similar to the wild-type *R. solanacearum* and the evolved clones, where the *efpR* pathway controls metabolism and antagonistically regulates flagellum and EPS production. In addition, like the RSI Switch (Heavner, et al. 2015), the *efpR* pathway was found to be lineage-specific.

Trade-offs are central to adaptive radiation (Ferenci 2016). They are thought to be drawn up according to saving cellular resources allocated to one trait that may be to the cost of another trait. This was recently evidenced for virulence and metabolism in *R. solanacearum* (Peyraud, et al. 2016). We previously observed a clear trade-off between symbiosis and pathogenesis during the symbiotic evolution of *R. solanacearum*. The mutations driving symbiotic adaptation in the three nodulating ancestors affected virulence determinants in the original host. They included mutations either in the master virulence regulatory cascade *prhARIJ-hrpG* or the T3SS (Marchetti, et al. 2010; Guan, et al. 2013). Here we showed that mutations that improve intracellular infection affect the pathogenic

power of *R. solanacearum*, revealing a higher complexity of the known *R. solanacearum* virulence regulatory network (Genin and Denny 2012), as well as large negative correlations between symbiosis and pathogenesis at the molecular level. An *efpR* mutant of the original chimeric *Ralstonia* exhibited a delay in the appearance of disease symptoms on *Arabidopsis thaliana*. This phenotypic effect is in accordance to the fact that EfpR controls several virulence determinants, *e.g*. EPS that strongly contribute to plant colonization and death (Araud-Razou, et al. 1998), and T3 effectors required for optimal virulence and efficient multiplication in host plants (Cunnac, et al. 2004; Macho, et al. 2010). The point mutation upstream the RSc0965-0967 genes had an even stronger effect, as it conferred the chimeric strain a completely avirulent phenotype on *A. thaliana*, indicating that this mutation affects additional virulence genes. Induction of the RSc3146-3148 genes fully accounted for the effect of EfpR inactivation and only partly for the effect of the up^{115}-RSc0965 mutation. Hence the overexpression of the RSc3146-RSc3148 genes is required for symbiosis but impairs virulence. The co-option of this virulence pathway for another symbiotic interaction with concomitant loss of virulence reinforces the existence of trade-offs between the two processes.

Regulatory pathways are important determinants of the bacterial lifestyle, by ensuring the coordinated expression of appropriate functions at the right time and place. During evolution, adaptation to a new ecological niche can co-opt regulators to reshape the bacterial lifestyle, as predicted for the ExoR, ExoS, and ChvI components that control host invasion by pathogenic (*Agrobacterium/Brucella/Bartonella*) and mutualistic (*Sinorhizobium/Rhizobium*) bacterial species of the Rhizobiales lineage (Heavner, et al. 2015). The regulatory pathway that we described in this work is likely ancestral and specific to the phytopathogenic *R. solanacearum* complex. This highlights that acquisition of a complex trait may lead to the recruitment of regulatory networks to achieve a different purpose than the initial one, leading here to trade-offs between virulence and mutualism.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions

Bacterial strains and plasmids used in this work are listed in Table 2 and supplementary table S4. *R. solanacearum* strains were grown at 28°C either on rich BG medium supplemented with 28 mM glucose (Boucher, et al. 1985) or on MP minimal medium (Plener, et al. 2010) supplemented either with 20 mM glutamate for gene expression analyses or with glycerol 2% for bacterial transformation. Antibiotics were used at the following concentrations: trimethoprim 100 µg/ml, spectinomycin 40 µg/ml, kanamycin 50 µg/ml, gentamycin 10 µg/ml.

Construction of bacterial strains

For bacterial visualization *in planta*, reporter strains bearing the GFPuv or mCherry reporters integrated in the chromosome downstream from *glmS* were constructed using the pRC delivery system (Monteiro, et al. 2012). To this end, the reporter-promoter fusions from pRCG-P*ps*-GFP or pRCG-P*ps*-mcherry (Cruz, et al. 2014) were excised using *Avr*II and *Bgl*II and cloned into the same sites in pRCK-GWY (Monteiro, et al. 2012) to give rise to pRCK-P*ps*-GFP and pRCK-P*ps*-mCherry (supplementary table S4). For chromosomal integration, the plasmids were linearized with *Sca*I or *Sfi*I and introduced into *R. solanacearum* strains by natural transformation as described in (Cruz, et al. 2014).

The point mutations, *efpR*E66K, *efpR*P63S and up^{115}-RSc0965, were introduced in *R. solanacearum* genomes using the MuGENT technique described by (Dalia, et al. 2014). Briefly this technique is based on the natural competence of bacteria and the frequency of co-transformation of two unlinked DNA fragments, one fragment carrying an antibiotic resistance marker that can be selected and one PCR product carrying the point mutation to introduce (unselected). The selected DNA marker used was the *Sfi*I linearized pRCK-P*ps*-GFP plasmid which allows the chromosomal integration of the the kanamycin resistance gene and the constitutive P*psbA*-GFP fusion in the intergenic region downstream *glmS*. The unselected DNA fragment was obtained by PCR amplification of a 6 kb region surrounding the mutation position using genomic DNA of evolved cloned as template and high fidelity Phusion polymerase (ThermoFisher Scientific). The sequences of oligonucleotides used for amplifying unselected fragments are provided in supplementary table S5. PCR products were precipitated with 1/10 volume of sodium acetate 3 M and 2 volumes of ethanol and resuspended in 30 µl of water. Bacteria were transformed with 3 µg of PCR products and 300 ng of marked DNA. Transformants were selected on kanamycin and screened by PCR for the presence of the mutation using two pairs of primers specific to wild-type and mutated alleles respectively (table S5). Screening primers were designed with a mismatched nucleotide at the 3' penultimate position and either the wild-type or the mutated nucleotide at the 3' end.

To knock out the *efpR* gene in chimeric *Ralstonia* strains, a 6 kb PCR fragment containing the wild-type copy of *efpR* was amplified using oCBM2449-oCBM2450 pair of primers and cloned into the pBluescript SK vector. Then a *Sma*I-digested spectinomycin resistance cassette from pHP45Ω was inserted at the *Mcs*I blunt ended site of the *efpR* coding sequence. The resulting plasmid (pCBM182) was linearized with *Sca*I and introduced into *Ralstonia* strains GMI1000 pRalta and CBM1627 by natural transformation.

To inactivate the RSc0965, RSc0967, RALSOc_1033 and RSc3147 genes and to delete the three RSc3146-3147-3148 genes, upstream and downstream regions of these genes were amplified using the Phusion polymerase and the pairs of primers indicated in supplementary table S5. Fragments were A-tailed using 0.2 mM dATP and 5 units of Taq DNA polymerase (Gotaq, Promega) prior ligation

into the pGEM-T vector (Promega). For the RSc0965, RSc0967 and RALSOc_1033 mutants, DNA fragments corresponding to the downstream regions were digested from the pGEM-T plasmids by *Hin*dIII and *Bam*HI and cloned into the pGEM-T plasmids carrying the upstream fragments digested with *Hin*dIII and *Bam*HI. For the RSc3147 mutant, the DNA fragment corresponding to the downstream region was digested from the pGEM-T plasmid by *Hin*dIII and *Pst*I and cloned into the pGEM-T plasmids carrying the upstream fragment digested with *Hin*dIII and *Pst*I. Then a *Hin*dIII-digested spectinomycin resistance cassette from pHP45Ω was inserted in the resulting plasmids digested with *Hin*dIII. For the deletion of RSc3146-3148 genes, an *Eco*RI-digested spectinomycin resistance cassette from pHP45Ω together with the RSc3148 upstream region digested from the pGEM-T plasmid by *Sac*I and *Eco*RI were cloned into the pGEM-T plasmid carrying the RSc3146 downstream fragment digested by *Eco*RI and *Sac*I. All resulting plasmids were linearized by *Sca*I and introduced into chimeric *Ralstonia* strains by natural transformation.

To check the expression of the genes RSc3146-3148 during symbiosis, the region upstream the RSc3148 gene was amplified by PCR by using the oCBM2745 and oCBM2746 primers, digested with *Avr*II and *Acc*65I and cloned into the pRCK-P*ps*-mCherry vector digested with *Avr*II and *Acc*65I to replace the *psbA* promoter region. The resulting plasmid was linearized with *Sca*I and introduced into *Ralstonia* evolved clones B8 and B9 by natural transformation.

Plant assays and cytological analysis

Mimosa pudica seedlings from Australian origin (B&T World Seed, Paguignan, France) were cultivated as previously described (Marchetti, et al. 2014).

For relative fitness measurements, two strains carrying either differential antibiotic resistance genes (spectinomycin or kanamycin) or differential constitutive fluorophores (GFPuv or mCherry) were co-inoculated in equivalent proportion (10^6 bacteria of each strain per tube of plants). Nodules from 20 plants were harvested 21 days after inoculation, pooled, surface sterilized and crushed. Dilutions of nodule crushes were spread on selective solid medium either for antibiotic screening or for fluorescent colony counting under a fluorescent stereo zoom microscope V16 (Zeiss).

For cytology, nodules were harvested at 21 dpi and cut into 55 µm sections with a Leica VT1000S vibratome apparatus. Longitudinal sections were observed using an inverted microscope (DMIRBE, Leica) and images acquired using a CCD camera (Color Coolview, Photonic Science, Milham, UK). For semi-quantitative infection analyses, one representative section of each nodule was quantified by measuring the infection and necrotic zones using the Image-Pro Plus software based on the HIS method (Media Cybernetics MD, USA) as previously described (Marchetti, et al. 2014; Marchetti, et al. 2016). Infection and necrosis data were obtained from 20 to 56 nodules from 1 to 3 independent

experiments for each strain. In each experiment nodules were harvested from 20 plants. Raw data are provided in supplementary Table S6.

Genome resequencing and detection of mutations

Evolved clones immediately before and after infection phenotypic shifts were re-sequenced either by C.E.A/IG/Genoscope using the Illumina GA2X technology (clone B8) or by the GeT platform (http://get.genotoul.fr/) using the Illumina HiSeq2000 technology (clones B9, C5, C6, G1, G2, H6, H7, I1, I2). High throughput sequencing data were analyzed as described previously (Remigi, et al. 2014) using the PALOMA bioinformatics pipeline implemented in the Microscope platform (Vallenet, et al. 2013). The complete list of events generated for all the clones from this study are available on the Microscope platform (https://www.genoscope.cns.fr/agc/microscope/expdata/evoProject.php, SYMPA tag). To filter sequencing errors, the following criteria were applied: minimum coverage on a position was 10 and minimum strand ratio was 0.25. SNPs/indels having a score <0.4 (score of an event is defined by *s= 0.5 x allele ratio + 0.5 x technical score* where technical score is a function of quality bases and strand ratio) and an allele ratio (mutated reads on the position/total reads on the position) <0.61 were removed. Also SNPs/indels present in more than 30% of all resequenced evolved clones were considered as false positive (Remigi, et al. 2014). The final list of mutations identified with high confidence (*i.e.* coverage>10, strand ratio>0.25, score>0.4 and allele ratio>0.61) is displayed in supplementary table S1.

Gene expression analyses

Bacteria cultivated in minimal medium were harvested at mid-exponential phase by adding 1/10 volume of cold phenol/ethanol (5/95) mixture. Bacteria were centrifuged 10 min at 4°C at 4000 rpm, pellets were frozen in liquid nitrogen and stored at -80°C until RNA extraction. Pellets were resuspended in 1 mg/ml of lysozyme in Tris/EDTA (10/1, pH8) buffer and incubated for 5 minutes at room temperature prior RNA extraction. RNAs were extracted using the mirVana kit (Ambion), treated with Turbo DNase for 30 min at 37°C (Ambion) to remove residual genomic DNA, and then purified with the Sureprep kit (Fisher BioReagents). RNA concentration was determined using a Nanodrop ND 1000 spectrophotometer (ThermoFisher Scientific). RNA quality was assessed using a 2100 Bioanalyzer instrument (Agilent Technologies).

For transcriptomic analyses, samples, whose concentration was below 250 ng/µl, were concentrated using the RNA clean and concentrator kit (Zymo Research). Hybridization of NimbleGen microarrays (72 K) were performed according to manufacturer instructions. Briefly, 2 µg of RNAs were reverse transcribed using the DoubleStrand cDNA Synthesis kit (Invitrogen). One microgram of double stranded cDNAs was labeled with the One-Color DNA Labeling kit (Roche NimbleGen). Two

micrograms of obtained labeled DNA was dried and resuspended in Sample Tracking Controls (Roche NimbleGen). Hybridization was performed using a NimbleGen Hybridization System for 17 hours at 42°C. Arrays were washed with the Wash Buffer Kit (Roche NimbleGen), dried, then automatically scanned with a NimbleGen MS200 Microarray Scanner and analyzed using the DEVA software to extract gene expression data. Raw data were normalized by the mean of all signals, and the median signal of all probes targeting the same gene was used to quantify gene expression. Experiments were performed in triplicate.

For quantitative Reverse Transcription-Polymerase Chain Reactions (RT-PCR), 1 µg of RNA were reverse transcribed using the Transcriptor Reverse Transcriptase kit (Roche, Life technologies) and random hexamers as primers. Real-time PCRs were run on a LightCycler system (Roche, Life technologies) using the FastStart DNA MasterPLUS SYBRGreen 1 kit (Roche, Life technologies) according to manufacturer's instructions.

Co-transcription of the RSc0965, RSc0966 and RSc0967 genes was assessed by RT-PCR. The oCBM2707-oCBM2751 pair of primers was used to amplify the 2.3 kb region between RSc0965 and RSc0967 on cDNAs obtained from the RCM1919 strain cultivated in rich or minimal medium. The CBM1627 strain, in which the genes are not expressed, was tested as negative control (supplementary fig. S3). Forty PCR cycles were performed on 2 µl of cDNAs using a Gotaq DNA polymerase (Promega).

All oligonucleotide sequences are provided in supplementary table S5.

Phenotype microarrays

Metabolic profiles were determined using 4 Biolog phenotype microarrays, PM1 and PM2 for carbon sources, PM3 for nitrogen sources, and PM4 for phosphorus and sulfur sources according to the manufacturer's instructions (Biolog, Inc., Hayward, CA, USA). Glucose was used as carbon source at 0.5% final concentration with PM3 and PM4 microarrays. Bacteria were incubated at 28°C for 96 hours in an OmniLog reader. Two independent experiments were performed for each strain. Data were analyzed using the parametric v1.3 Biolog software. Areas under the curves were used to score the metabolic activities in various conditions.

Pathogenicity assays on *Arabidopsis* plants

Susceptible *Arabidopsis thaliana* Col-0 plants were grown and inoculated after root clipping as previously described (Deslandes, et al. 1998). Disease symptoms were followed for three weeks and scored according to the following disease index scale: healthy plants (0), 1 to 25% leaves wilted (1), 26 to 50% leaves wilted (2), 51 to 75% leaves wilted (3), and more than 75% leaves wilted (4). Each virulence assay was repeated between three to eight times independently with 16 plants per strain

and per assay. For Kaplan-Meier survival analyses, data were transformed into a binary index: 0 for <75% of wilted leaves and 1 for more or equal to 75% wilted leaves. To compare the disease development of two given strains, we used the Gehan–Breslow–Wilcoxon method to compute the *P*-values as described (Remigi, et al. 2011). Statistical analyses were done with Prism version 5.00 (GraphPad software).

Public genome data

Genomes of *Ralstonia* and *Cupriavidus* larger than 5 Mb were downloaded from GenBank RefSeq and the MicroScope platform (http://www.genoscope.cns.fr/agc/microscope/home/) as available in September 2015. We computed for each genome the L90 value (defined as the smallest number of contigs whose length sum accounts for 90% of the genome) and only retained for further analysis those with a value smaller than 150. Genomes of α- and β-proteobacteria larger than 1 Mb and genomes of phages were downloaded from GenBank RefSeq as available in February 2013.

Core-genomes of *Ralstonia* and *Cupriavidus* strains

Core-genomes were computed using bi-directional best hits (hereafter named BBH), using end-gap free Needleman-Wunsch alignments, between the proteome of *Ralstonia solanacearum* GMI1000 as a pivot and each of the other proteomes as previously described (Rocha 2006). Hits with less than 40% similarity in amino acid sequence or more than a third of difference in protein length were discarded. The lists of putative orthologs were then filtered using positional information. Putative positional orthologs were defined as BBH adjacent to at least two other pairs of BBH within a neighbourhood of ten genes (five up- and down-stream). Lastly, the core-genome of the 44 strains was defined as the intersection of pairwise lists of positional orthologs. Accession numbers of the 44 *Ralstonia* and *Cupriavidus* genomes used are provided in supplementary table S7.

Multiple alignments and phylogenetic analyses

Alignments were performed on amino acids using Muscle v3.8.31 (Edgar 2004), and back-translated to DNA. Different models of protein or DNA evolution and associated parameters were investigated using IQ-TREE v1.3.8 (Nguyen, et al. 2015) on the concatenated alignment of 1136 core-genes. The best parameters were identified based on the Bayesian information criterion (BIC). Maximum likelihood trees were then computed with IQ-TREE v1.3.8 using the best model (JTT+I+G4+F), and validated via the ultrafast bootstrap procedure with 1000 replicates (Minh, et al. 2013).

Orthologs and pseudogenes of the *efpR* pathway

Putative orthologs of the 7 genes were identified as BBH (as defined above). We identified positional orthologs for the RSc3146-3148 genes in *Ralstonia* and *Cupriavidus* strains, but not in α- and β-proteobacteria and their phages because the wider phylogenetic distance renders such analyses less robust. Nevertheless, the RSc3146-3148 and RSc0965-0967 groups of genes were considered present only if at least 2 out of the 3 genes of the group were found in these strains within a neighbourhood of ten genes. We also searched the nucleotide sequences of *Ralstonia* and *Cupriavidus* genomes for pseudogenes and orthologs that might not be correctly annotated. For this, we took the protein in the reference genome and searched for this query in the genomes using LAST (Sheetlin, et al. 2014) with a score penalty of 15 for frameshifts. We discarded hits with e-values below 10^{-5}, with less than 40% similarity in amino acid sequence, or aligning less than 30% of the query protein. Hits lacking at least two other pairs of BBH or LAST hits within a neighbourhood of ten genes (five up- and down-stream) were discarded. Hits with frameshifts, stop codons or aligning less than 70% of the query protein were defined as pseudogenes. Hits lacking stops and frameshifts and aligning more than 70% of the query protein were defined as putative orthologs.

Evolutionary analysis of the *efpR* pathway

Count (Csurös 2010) was used to study the variation in gene content of both the genus *Ralstonia* and *Cupriavidus*. This program computed the scenario of gene gain and loss of the 7 genes from the *efpR* pathways along the phylogeny. Different models of gene content evolution were tested and the best parameters (4 category rate variation for loss and length) were identified based on the Akaike information criterion (AIC). Maximum likelihood was then used to compute posterior probabilities for gene content at inner nodes.

Figures

Fig. 1. Quantification of the infected and necrotic areas of nodules induced by evolved clones and mutants.

Distribution of the percentage of infected area (*A*) and necrotic area (*B*) per nodule section recovered at 21 dpi. The central rectangle spans the first quartile to the third quartile (that is, the interquartile range or IQR), the bold segment inside the rectangle shows the median, unfilled circles indicate outliers, whiskers above and below the box show either the locations of the minimum and maximum in the absence of suspected outlying data or 1.5 × IQR if an outlier is present. Dark colors represent ancestors and light colors represent the corresponding derivatives clones or mutants. Spe, interruption of the gene by a spectinomycin resistance cassette. * Significantly different from the respective ancestral strain, either CBM212, CBM349 or GMI1000 pRalta *hrpG*, # significantly different

from the (n-1) evolved clone (P<0.05, multiple comparison test after Kruskal–Wallis). n, number of nodules analyzed. Nodules were obtained from 1 to 3 independent experiments. For each experiment nodules were harvested from 20 plants. Raw data are provided in the supplementary table S6. Values for CBM212 and CBM349 were from (Marchetti, et al. 2016).

Fig.2. Genomic organization of three loci containing the RALSOc_1033-RSc0967, *efpR* and RSc03146-3148 genes. The point mutation G/A located 115 bp upstream from RSc0965 is positioned. The putative functions of the corresponding proteins are indicated below the genes.

Fig. 3. Relative *in planta* fitness of evolved clones and reconstructed mutants. *M. pudica* plantlets were co-inoculated with pairs of strains at a 1:1 ratio. Nodules were harvested at 21 dpi and the number of nodule bacteria was evaluated by plating, using either antibiotic resistance or fluorescent markers as screening method. The percentages of bacteria in nodule populations were normalized by the inoculum ratio. Results are from 3 to 4 independent experiments. * The percentage of the two strains in nodule populations was significantly different (P<0.01, Student t-test).

Fig. 4. Expression of the *efpR*, RSc03148 and RSc0967 genes in GMI1000 pRalta mutants and evolved clones. Strains were cultivated in minimal medium containing 20 mM glutamate as sole carbon source until mid-exponential phase. The expression of *efpR*, RSc3148 and RSc0967 genes were quantified by qRT-PCR. Raw expression levels were normalized by the *rplM* expression and ratios of gene expression in mutants versus wild-type strains were calculated. Data are presented as mean ratios +/- standard deviation and were obtained from three independent experiments.

Fig. 5. Model representing the EfpR/RSc3146-3148 regulatory cascade. The *efpR* expression may be regulated by an upstream yet unidentified component which is affected by the up^{115}-RSc0965 mutation. The EfpR regulator directly or indirectly controls 200 genes, 180 of which are dependent upon the expression of RSc3146-3148 genes. The RSc3146-3148 proteins may in turn directly or indirectly control the expression of several genes among them those involved in metabolism, EPS production, motility and chemotaxis. These transcriptomic and/or metabolic modifications triggered both a decrease in the pathogenicity of *the* chimeric *Ralstonia* on its original host *A. thaliana* and a better intracellular infection capacity of *M. pudica* nodules. EPS, exopolysaccharides. T3E, Type III effectors. TTSS, Type Three Secretion System. OM, outer membrane. IM, inner membrane.

Fig. 6. Expression of RSc3148 in *M. pudica* nodules. Expression of the RSc3148-mCherry transcriptional fusion was analyzed in B8 (*A*, *C*, *E*) and B9 (*B*, *D*, *F*) either grown on BG glucose plates

(*A*, *B*) or on *M. pudica* nodule sections of 21 dpi (*C*-*F*) under light (*C*, *D*) or fluorescence microscopy (*E*, *F*). The fluorescence in *E* corresponds to the necrotic zone of the nodule (arrow) and not to the intracellular infected zone (asterisks).

Fig. 7. Disease symptoms induced by wild-type and mutant strains of the chimeric *Ralstonia* on *A. thaliana.* Kaplan-Meyer survival analysis of *A. thaliana* plants inoculated by the wild-type GMI1000 pRalta chimeric *Ralstonia* and its *efpR*E66K (*A*) or up^{115}-RSc0965 (*B*) derivative mutants. * The survival curves are significantly different from the wild-type curve, # the survival curves are significantly different from the *efpR*E66K or up^{115}-RSc0965 mutant curves (P<0.001, Gehan-Breslow-Wilcoxon test). Values were obtained from 3 to 8 independent series of 16 plants.

Fig. 8. Taxonomic distribution of the RSc0965-0967, *efpR* and RSc3146-3148 genes. The tree represents the phylogenetic association between *Ralstonia* and *Cupriavidus*. Numbers are ultra fast bootstrap values (%) reflecting clade support (only values <100 are shown). Dark green, light green, orange and empty circles indicate BBH orthologs, orthologs identified using LAST, pseudogenes and absence of genes in the corresponding genome, respectively. From top to bottom, boxes indicate lineages recently proposed as three different species within the *Ralstonia* genus: *R. solanacearum*, *R. syzygii*, and *R. pseudosolanacearum* (Prior, et al. 2016).

Tables.

Table 1. Adaptive mutations for improved infection.

Clone	Absolute Position	Gene ID	Description	Distance to start codon	Mutation	Protein modification
B9	1013285	RSc0965[a]	Protein of unknown function	115	G/A	
C6	1155492	RSc1097	Transcription regulator EfpR	-	C/T	P63S
G2	1155501	RSc1097	Transcription regulator EfpR	-	G/A	E66K
H7	1155501	RSc1097	Transcription regulator EfpR	-	G/A	E66K
I2	1155501	RSc1097	Transcription regulator EfpR	-	G/A	E66K

[a] mutation located in the intergenic region upstream from the indicated gene.

Table 2. Description and phenotypes of the main chimeric *Ralstonia* strains used in this study

Strain	Description and origin	Nodulation on *M. pudica*	Infection on *M. pudica*	Virulence on *Arabidopsis*
Original chimeric Ralstonia				
GMI1000 pRalta	Original chimera, ancestor of the evolution experiment	-	-	+
Evolved clones				
CBM212	Nodulating chimera, GMI1000 pRalta-derived, selection cycle, bears an *hrpG* mutation	+	+/-	-
B8	CBM212-derived, lineage B, cycle 8	+	+/-	-
B9	CBM212-derived, lineage B, cycle 9	+	+	-
CBM349	Nodulating chimera, GMI1000 pRalta-derived, selection cycle, bears an *hrpG* mutation	+	+/-	-
G1	CBM349-derived, lineage G, cycle 1	+	+/-	-
G2	CBM349-derived, lineage G, cycle 2	+	+	-

Reconstructed mutants				
B8 up^{115}-RSc0965	up^{115}-RSc0965 mutant of B8	+	+	-
G1 *efpR*E66K	*efpR*E66K mutant of G1	+	+	-
GMI1000 pRalta *hrpG*	*hrpG* mutant of GMI1000 pRalta, referred as the nodulating mutant chimera	+	+/-	-
GMI1000 pRalta *hrpG efpR*E66K	*efpR*E66K mutant of GMI1000 pRalta *hrpG*	+	+	-
GMI1000 pRalta *hrpG* up^{115}-RSc0965	up^{115}-RSc0965 mutant of GMI1000 pRalta *hrpG*	+	+	-
GMI1000 pRalta *hrpG efpR*E66K Δ(RSc3146-3148)	RSc3146-3148 deletion mutant of GMI1000 pRalta *hrpG efpR*E66K	+	+/-	-
GMI1000 pRalta *hrpG efpR*E66K RSc3147::Spe	RSc3147 mutant of GMI1000 pRalta *hrpG efpR*E66K	+	+/-	-
GMI1000 pRalta *efpR*E66K	efpRE66K mutant of GMI1000 pRalta	-	-	delayed
GMI1000 pRalta *efpR*E66K Δ(RSc3146-3148)	RSc3146-3148 deletion mutant of GMI1000 pRalta *efpR*E66K	-	-	+
GMI1000 pRalta *efpR*E66K RSc3147::Spe	RSc3147 mutant of GMI1000 pRalta *efpR*E66K	-	-	+
GMI1000 pRalta up^{115}-RSc0965	up^{115}-RSc0965 mutant of GMI1000 pRalta	-	-	-
GMI1000 pRalta up^{115}-RSc0965 Δ(RSc3146-3148)	RSc3146-3148 deletion mutant of GMI1000 pRalta up^{115}-RSc0965	-	-	delayed
GMI1000 pRalta up^{115}-RSc0965 RSc3147::Spe	RSc3147 mutant of GMI1000 pRalta up^{115}-RSc0965	-	-	delayed

Table 3. Genes directly or indirectly regulated by EfpR.

Functional categories	**No of genes**	**Genes**	**Description**	**EfpR control[a]**	**RSc3146-3148 dependency[b]**
Motility-chemotaxis	52	RSp0339-0352, RSp0374-0379, RSp0382-0385, RSp0388-0395, RSp1390-1393	Flagellar biosynthesis proteins	-	+
		RSp1401-1403, RSp1405-1411	Chemotaxis regulatory and flagellar motor proteins	-	+
		RSp1412-RSp1413	Transcriptional activators FlhC, FlhD	-	+
		RSc1156, RSc1894, RSc3412, RSp1209	Methyl-accepting chemotaxis transducer proteins	-	+
Siderophore/ hemin uptake and degradation	21	RSc1804-1813	Siderophore uptake and degradation proteins	-	+
		RSp0247	Ferric uptake transcription regulator Fur2	-	+
		RSc0800, RSp0414, RSp0416-0418	Ferric siderophore receptor and biosynthesis proteins	-	+
		RSp0422, RSp0424	Siderophore biosynthesis proteins	-	+
		RSc1966, RSp0243-0244	Putative hemin transport proteins	-	+
Transcription regulation	3	RSc0407	Putative sigma-54 modulation protein	-	+
		RSp0598	Putative response regulator	-	+
		RSp0849	RNA polymerase sigma factor, PrhI	-	+
Other functions negatively controlled by EfpR and dependent on RSc3146-3148	36	RSc0022	Carbon starvation A transmembrane protein	-	+
		RSc0349	Putative bacterioferritin-associated ferredoxin protein	-	+
		RSc0801	Putative lipoprotein transmembrane	-	+
		RSc0926	D-amino acid dehydrogenase small subunit	-	+
		RSc1600, RSc1603	Pyruvate dehydrogenase subunits	-	+
		RSc1305	Ferredoxin--NADP reductase (FNR) (Protein X)	-	+
		RSc1964-1965	Biopolymer transport transmembrane proteins	-	+
		RSc2464	ATP-dependent protease subunit	-	+
		RSc2465	ATP-dependent Clp protease adaptor protein ClpS	-	+
		RSc2805	Ribonucleotide-diphosphate reductase subunit alpha	-	+
		RSp0242	Glycogen synthase	-	+
		RSp0412	Putative fosmidomycin resistance protein	-	+
		RSp0904	Putative dehydrogenase oxidoreductase protein	-	+
		RSp0906	Putative signal peptide protein	-	+
		RSp1592	Leu/Ile/Val-binding transmembrane protein	-	+
		RSp1047	Transporter transmembrane protein	-	+
		RSc0023, RSc0415-0416, RSc0592, RSc1470, RSc1962,RSc2894, RSc3386, RSp0240, RSp0248, RSp0413, RSp0728-0730, RSp0903, RSp1415, pRALTA_0197-0198	Hypothetical proteins	-	+
EPS biosynthesis	17	RSp1003	Transcription regulator XpsR	+	+
		RSp1004, RSp1006-1020	EPS biosynthesis and export proteins	+	+
Virulence	12	RSc1475, RSc2775, RSp0213, RSp0304, RSp0572, RSp0876, RSp0877	Type III effector proteins *(ripM, ripW, ripBJ, ripD*, *ripAX2*, *ripAB* and *ripX,* respectively)	+	+

		RSp0854-RSp0856, RSp0865	Hrp proteins	+	+
		RSp1555	Type III secretion translocon protein RipF1_1	+	+
Other functions positively controlled by EfpR and dependent on RSc3146-3148	39	RSc1704	ISRSO18-transposase protein	+	+
		RSc1817	Putative siderophore biosynthesis protein	+	+
		RSc1818	Methionyl-tRNA synthetase	+	+
		RSp1417	Transmembrane efflux protein	+	+
		RSp1418	Hydrolase protein	+	+
		RSp1419	Ralfuranone biosynthesis protein, RalA	+	+
		RSc2108	Porin signal peptide protein	+	+
		RSc3002	30S ribosomal protein S5	+	+
		RSp0173	Putative cysteine desulfurase	+	+
		RSp0277	Periplasmic alpha,alpha-trehalase protein	+	+
		RSp0603	Putative serine protease protein	+	+
		RSp0841	Putative lipoprotein	+	+
		RSp0983	Transcription regulator protein	+	+
		RSp1167	Oxidoreductase	+	+
		RSc0025, RSc0117, RSc0582, RSc0616-0617, RSc0894, RSc1705, RSc1728, RSc1816, RSc1819, RSc1908-1909, RSc1942-1943, RSc2024-2026, RSc2107, RSc2388, RSc3197, RSc3292, RSp0172, RSp0261, RSp0587, RSp0957	Hypothetical proteins	+	+
Other functions controlled by EfpR and independent from RSc3146-3148	20	RSc2148	Peptidoglycan LysM-family protein	-	-
		RSc2662	Mechanosensitive ion channel	-	-
		RSc0815, RSc3146-3148, RSp0206-0208, RSp0516-0518	Hypothetical proteins	-	-
		RSc3005, RSc3010	30S ribosomal protein subunits	+	-
		RSp1581	Catalase hydroperoxidase HpiI oxidoreductase	+	-
		pRALTA_0250	Putative glycosyl hydrolase	+	-
		pRALTA_0429	Transposase fragment, IS1111	+	-
		RSc1944, RSp0602, RSp1580	Hypothetical proteins	+	-

[a]+, genes positively controlled by EfpR. -, genes negatively controlled by EfpR.

[b]+, genes whose expression is dependent on RSc3146-3148. -, genes whose expression is independent on RSc3146-3148.

Table 4. Carbon, nitrogen and phosphorus sources better metabolized in the *efpR* mutant

Metabolic sources	Chemical	Ratio of AUC[a] *hrpG efpR*E66K /*hrpG*[b]	Ratio of AUC[a] *hrpG efpR*E66K Δ(RSc3146-3148) /*hrpG*[b]	Ratio of AUC[a] *hrpG efpR*E66K RSc3147::Spe /*hrpG*[b]
Carbon sources				
-*Carboxylic acids*	D-Saccharic acid	2.32	0.77	0.59
	Succinic acid	2.54	0.78	0.68
	a-Ketoglutaric acid	3.00	0.79	0.78
	Citric acid	2.59	0.90	0.55
	Bromosuccinic acid	2.48	0.50	0.45
	L-Galactonic acid-g-Lactone	2.53	0.64	0.82
	D-Galacturonic acid	2.52	0.62	0.73
	Quinic acid	2.25	0.79	0.85
-*Amino acids*	L-Proline	2.16	1.03	0.94
	L-Serine	3.33	0.99	0.88
	L-Alanine	3.72	0.83	0.73
	L-Histidine	3.86	0.62	0.55
Nitrogen sources				
-*Amino acids*	Glycine	2.02	0.97	0.76
	L-Phenylalanine	2.48	1.16	1.04
	L-Pyroglutamic acid	2.33	0.83	0.87
-*Dipeptides*	Ala-Gln	2.15	0.49	0.91
	Ala-Glu	3.05	0.70	0.87
-*Other*	Formamide	2.72	1.10	0.95
	Adenine	2.40	0.57	0.89
Phosphorus sources				
-*Hexose phosphate*	D-Glucose-1-Phosphate	2.66	0.88	0.68
	D-Glucose-6-Phosphate	2.95	0.91	0.80
	D-Mannose-1-Phosphate	2.40	1.02	0.83
	D-Mannose-6-Phosphate	2.47	0.90	0.68
	2-Deoxy-D-Glucose 6-Phosphate	2.46	0.99	0.83
	D-Glucosamine-6-Phosphate	2.41	0.88	0.76
	6-Phospho-Gluconic acid	2.45	0.81	0.65
-*Amino acid phosphate*	Phospho-L-Arginine	2.40	0.84	0.70
	O-Phospho-L-Serine	2.52	1.07	0.70
	O-Phospho-L-Threonine	2.51	1.02	0.73
	O-Phospho-D-Tyrosine	2.80	0.97	0.68
-*Nucleotides*	Adenosine 2`-Monophosphate	2.40	1.04	0.71

	Adenosine 3`-Monophosphate	2.56	1.09	0.86
	Adenosine 5`-Monophosphate	2.39	1.15	0.93
	Guanosine 2`-Monophosphate	2.35	0.86	0.72
	Guanosine 3`-Monophosphate	2.54	0.95	0.84
	Guanosine 5`-Monophosphate	2.32	0.89	0.78
	Guanosine 2`,3`-Cyclic Monophosphate	2.23	1.00	0.70
	Cytidine 2`-Monophosphate	2.02	0.78	0.55
	Cytidine 3`-Monophosphate	2.33	0.92	0.76
	Cytidine 5`-Monophosphate	2.24	0.93	0.88
	Cytidine 2`,3`-Cyclic Monophosphate	2.21	0.96	0.87
	Uridine 2`-Monophosphate	2.15	0.81	0.64
	Uridine 3`-Monophosphate	2.77	0.96	0.91
	Uridine 5`-Monophosphate	2.37	0.84	0.82
	Uridine 2`,3`-Cyclic Monophosphate	2.43	0.96	0.92
	Thymidine 3`-Monophosphate	2.47	0.95	0.87
	Thymidine 5`-Monophosphate	2.34	0.90	0.77
-Other organic phosphate	Phosphoenol Pyruvate	3.45	1.23	0.79
	D,L-a-Glycerol Phosphate	2.28	0.89	0.69
	b-Glycerol Phosphate	2.16	0.88	0.75
	D-2-Phospho-Glyceric acid	2.61	1.00	0.86
	D-3-Phospho-Glyceric acid	2.40	0.90	0.69
	Phospho-Glycolic acid	2.94	1.03	0.78
	Cysteamine-S-Phosphate	2.65	0.94	0.77
	Phosphocreatine	2.80	0.95	0.72
	Phosphoryl Choline	2.10	0.66	0.57
	O-Phosphoryl-Ethanolamine	2.72	0.97	0.81
	Inositol Hexaphosphate	2.70	0.93	0.82
-Inorganic phosphate	Phosphate	2.34	1.08	0.85
	Tripolyphosphate	3.10	1.19	0.78
	Thiophosphate	2.50	1.01	0.81
	Dithiophosphate	2.35	1.04	0.79

[a]AUC, areas under the curve generated by an Omnilog PM system measuring cellular respiration along 96 hours. [b]Both strains are in the GMI1000 pRalta background. Values are the mean of two independent experiments.

Supplementary Materials

Fig. S1. The evolution experiment**.**

Fig. S2. Evolution of symbiotic phenotypes along CBM212-derived (*A*, *C*) and CBM349-derived (*B*, *D*) lineages.

Fig. S3. Co-transcription of the RSc0965-0966-0967 genes.

Fig. S4. Expression of RALSOc_1033 (*A*) and *efpR* (*B*) in GMI1000 pRalta and derivatives mutants.

Fig. S5. Graphic representation of metabolic gains detected in the *efpR*E66K mutant but not in the *efpR*E66K Δ(RSc3146-3148) and *efpR*E66K RSc3147::Spe strains.

Table S1: Candidate point mutations associated with infection shifts.

Table S2. Nimblegen microarray raw data.

Table S3. Strains harboring orthologous copies of RSc0965-0967, *efpR* and RSc3146-3148.

Table S4. Bacterial strains and plasmids used in this study.

Table S5. PCR primers used in this study.

Table S6. Raw data of the quantification of infected and necrotic areas of nodules induced by evolved clones and mutants.

Table S7. Genomic datasets of the 44 strains from *Ralstonia* and *Cupriavidus* genera.

Acknowledgements

We are grateful to Lidwine Trouilh for helping in NimbleGen microarray hybridizations and Loic Escoriza for mutant construction. JPC and CC were supported by the Initiative d'Excellence IDEX UNITI Actions Thématiques Stratégiques program (RHIZOWHEAT 2014) and by the French National Research Agency (ANR-12-ADAP-0014-01). This work was supported by funds from the French National Institute for Agricultural Research (Plant Health and the Environment Division), the French National Research Agency (ANR-12-ADAP-0014-01).

The complete collections of events generated for all the clones from this study are available on the Microscope platform (https://www.genoscope.cns.fr/agc/microscope/expdata/evoProject.php, SYMPA tag).

References

Amadou C, Pascal G, Mangenot S, Glew M, Bontemps C, Capela D, Carrère S, Cruveiller S, Dossat C, Lajus A, et al. 2008. Genome sequence of the beta-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res*. 18:1472-1483.

Antoine R, Huvent I, Chemlal K, Deray I, Raze D, Locht C, Jacob-Dubuisson F. 2005. The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction. *J Mol Biol.* 351:799-809.

Araud-Razou I, Vasse J, Montrozier H, Etchebar C, Trigalet A. 1998. Detection and visualization of the major acidic exopolysaccharide of *Ralstonia solanacearum* and its role in tomato root infection and vascular colonization. *Eur J Plant Pathol*. 104:795-809.

Berrabah F, Ratet P, Gourion B. 2015. Multiple steps control immunity during the intracellular accommodation of rhizobia. *J Exp Bot. 66*:1977-1985.

Boucher CA, Barberis PA, Trigalet AP, Demery DA. 1985. Transposon mutagenesis of *Pseudomonas solanacearum*: isolation of Tn5-induced avirulent mutants. *J Gen Microbiol*. 131:2449-2457.

Bélanger L, Dimmick KA, Fleming JS, Charles TC. 2009. Null mutations in *Sinorhizobium meliloti exoS* and *chvI* demonstrate the importance of this two-component regulatory system for symbiosis. *Mol Microbiol.* 74:1223-1237.

Chen WM, James EK, Prescott AR, Kierans M, Sprent JI. 2003. Nodulation of *Mimosa spp*. by the beta-proteobacterium *Ralstonia taiwanensis*. *Mol Plant Microbe Interact*. 16:1051-1061.

Chen WM, Laevens S, Lee TM, Coenye T, De Vos P, Mergeay M, Vandamme P. 2001. *Ralstonia taiwanensis* sp nov., isolated from root nodules of *Mimosa* species and sputum of a cystic fibrosis patient. *Int J Syst Evol Microbiol*. 51:1729-1735.

Cockerell SR, Rutkovsky AC, Zayner JP, Cooper RE, Porter LR, Pendergraft SS, Parker ZM, McGinnis MW, Karatan E. 2014. *Vibrio cholerae* NspS, a homologue of ABC-type periplasmic solute binding proteins, facilitates transduction of polyamine signals independent of their transport. *Microbiology* 160:832-843.

Consortium U. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204-212.

Crook MB, Lindsay DP, Biggs MB, Bentley JS, Price JC, Clement SC, Clement MJ, Long SR, Griffitts JS. 2012. Rhizobial plasmids that cause impaired symbiotic nitrogen fixation and enhanced host invasion. *Mol Plant Microbe Interact*. 25:1026-1033.

Cruz AP, Ferreira V, Pianzzola MJ, Siri MI, Coll NS, Valls M. 2014. A novel, sensitive method to evaluate potato germplasm for bacterial wilt resistance using a luminescent *Ralstonia solanacearum* reporter strain. *Mol Plant Microbe Interact*. 27:277-285.

Csurös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26:1910-1912.

Cunnac S, Occhialini A, Barberis P, Boucher C, Genin S. 2004. Inventory and functional analysis of the large Hrp regulon in Ralstonia solanacearum: identification of novel effector proteins translocated to plant host cells through the type III secretion system. *Mol Microbiol*. 53:115-128.

Dalia AB, McDonough E, Camilli A. 2014. Multiplex genome editing by natural transformation. *Proc Natl Acad Sci USA.* 111:8937-8942.

Deslandes L, Pileur F, Liaubet L, Camut S, Can C, Williams K, Holub E, Beynon J, Arlat M, Marco Y. 1998. Genetic characterization of RRS1, a recessive locus in *Arabidopsis thaliana* that confers resistance to the bacterial soilborne pathogen *Ralstonia solanacearum*. *Mol Plant Microbe Interact.* 11:659-667.

Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol*. 2:414-424.

Downie JA. 2010. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol Rev*. 34:150-170.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

Ferenci T. 2016. Trade-off mechanisms shaping the diversity of bacteria. *Trends Microbiol*. 24:209-223.

Fraysse N, Couderc F, Poinsot V. 2003. Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. *E J Biochem*. 270:1365-1380.

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol*. 3:722-732.

Genin S. 2010. Molecular traits controlling host range and adaptation to plants in *Ralstonia solanacearum. New Phytol*. 187:920-928.

Genin S, Denny TP. 2012. Pathogenomics of the *Ralstonia solanacearum* species complex. *Annu Rev Phytopathol.* 50:67-89.

Guan SH, Gris C, Cruveiller S, Pouzet C, Tasse L, Leru A, Maillard A, Médigue C, Batut J, Masson-Boivin C, et al. 2013. Experimental evolution of nodule intracellular infection in legume symbionts. *ISME J.* 7:1367-1377.

Guidot A, Jiang W, Ferdy J-B, Thebaud C, Barberis P, Gouzy J, Genin S. 2014. Multihost experimental evolution of the pathogen *Ralstonia solanacearum* unveils genes involved in adaptation to plants. *Mol Biol Evol*. 31:2913-2928.

Heavner ME, Qiu W-G, Cheng H-P. 2015. Phylogenetic co-occurrence of ExoR, ExoS, and ChvI, components of the RSI bacterial invasion switch, suggests a key adaptive mechanism regulating the transition between free-living and host-invading phases in Rhizobiales. *PLoS One* 10.

Huang JZ, Carney BF, Denny TP, Weissinger AK, Schell MA. 1995. A complex network regulates expression of EPS and other virulence genes of *Pseudomonas solanacearum*. *J Bacteriol*. 177:1259-1267.

Ivashina TV, Fedorova EE, Ashina NP, Kalinchuk NA, Druzhinina TN, Shashkov AS, Shibaev VN, Ksenzenko VN. 2010. Mutation in the *pssM* gene encoding ketal pyruvate transferase leads to disruption of *Rhizobium leguminosarum* bv. viciae-*Pisum sativum* symbiosis. *J Appl Microbiol*. 109:731-742.

Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC. 2007. How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat Rev Microbiol*. 5:619-633.

Kawaharada Y, Kelly S, Nielsen MW, Hjuler CT, Gysel K, Muszynski A, Carlson RW, Thygesen MB, Sandal N, Asmussen MH, et al. 2015. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523:308-312.

Lercher MJ, Pál C. 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol Biol Evol*. 25:559-567.

Li YZ, Wang D, Feng XY, Jiao J, Chen WX, Tian CF. 2016. Genetic analysis reveals the essential role of nitrogen phosphotransferase system components in *Sinorhizobium fredii* CCBAU 45436 symbioses with soybean and pigeonpea plants. *Appl Environ Microbiol*. 82:1305-1315.

Macho AP, Guidot A, Barberis P, Beuzon CR, Genin S. 2010. A competitive index assay identifies several *Ralstonia solanacearum* Type III effector mutant strains with reduced fitness in host plants. *Mol Plant Microbe Interact*. 23:1197-1205.

Marchetti M, Capela D, Glew M, Cruveiller S, Chane-Woon-Ming B, Gris C, Timmers T, Poinsot V, Gilbert LB, Heeb P, et al. 2010. Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol*. 8.

Marchetti M, Catrice O, Batut J, Masson-Boivin C. 2011. *Cupriavidus taiwanensis* bacteroids in *Mimosa pudica* indeterminate nodules are not terminally differentiated. *Appl Environ Microbiol.* 77:2161-2164.

Marchetti M, Jauneau A, Capela D, Remigi P, Gris C, Batut J, Masson-Boivin C. 2014. Shaping bacterial symbiosis with legumes by experimental evolution. *Mol Plant Microbe Interact*. 27:956-964.

Marchetti M, Clerissi C, Yousfi Y, Gris C, Bouchez O, Rocha E, Cruveiller S, Jauneau A, Capela D, Masson-Boivin C. 2016. Experimental evolution of rhizobia may lead to either extra- or intracellular symbiotic adaptation depending on the selection regime. *Mol Ecol*. 26:1818-1831.

Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol*. 17:458-466.

Milling A, Babujee L, Allen C. 2011. *Ralstonia solanacearum* extracellular polysaccharide is a specific elicitor of defense responses in wilt-resistant tomato plants. *PLoS One* 6:e15853.

Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 30:1188-1195.

Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, et al. 2015. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*. 43:D213-221.

Monteiro F, Solé M, van Dijk I, Valls M. 2012. A chromosomal insertion toolbox for promoter probing, mutant complementation, and pathogenicity studies in *Ralstonia solanacearum*. *Mol Plant Microbe Interact*. 25:557-568.

Nandasena KG, O'Hara GW, Tiwari RP, Sezmis E, Howieson JG. 2007. *In situ* lateral transfer of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. *Environ Microbiol*. 9:2496-2511.

Nelson MS, Sadowsky MJ. 2015. Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. *Front Plant Sci*. 6:491.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32:268-274.

Ochman H, Moran N. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292:1096-1099.

Pál C, Papp B, Lercher MJ. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet.* 37:1372-1375.

Perrier A, Peyraud R, Rengel D, Barlet X, Lucasson E, Gouzy J, Peeters N, Genin S, Guidot A. 2016. Enhanced in planta fitness through adaptive mutations in EfpR, a dual regulator of virulence and metabolic functions in the plant pathogen *Ralstonia solanacearum*. *PLoS Pathog.* 12:e1006044.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 8:785-786.

Peyraud R, Cottret L, Marmiesse L, Gouzy J, Genin S. 2016. A resource allocation trade-off between virulence and proliferation drives metabolic versatility in the plant pathogen *Ralstonia solanacearum*. *PLoS Pathog*. 12:e1005939.

Plener L, Manfredi P, Valls M, Genin S. 2010. PrhG, a transcriptional regulator responding to growth conditions, is involved in the control of the Type III Secretion System regulon in *Ralstonia solanacearum*. *J Bacteriol*. 192:1011-1019.

Price PA, Tanner HR, Dillon BA, Shabab M, Walker GC, Griffitts JS. 2015. Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility. *Proc Natl Acad Sci USA.* 112:15244-15249.

Prior P, Ailloud F, Dalsing BL, Remenant B, Sanchez B, Allen C. 2016. Genomic and proteomic evidence supporting the division of the plant pathogen *Ralstonia solanacearum* into three species. *BMC Genomics* 17.

Rahme LG, Mindrinos MN, Panopoulos NJ. 1992. Plant and environmental sensory signals control the expression of *hrp* genes in *Pseudomonas syringae* pv. *phaseolicola*. *J. Bacteriol*. 174:3499-3507.

Rankin DJ, Rocha EP, Brown SP. 2011. What traits are carried on mobile genetic elements, and why? Heredity (Edinb) 106:1-10.

Remigi P, Anisimova M, Guidot A, Genin S, Peeters N. 2011. Functional diversification of the GALA type III effector family contributes to *Ralstonia solanacearum* adaptation on different plant hosts. *New Phytol*. 192:976-987.

Remigi P, Capela D, Clerissi C, Tasse L, Torchet R, Bouchez O, Batut J, Cruveiller S, Rocha EPC, Masson-Boivin C. 2014. Transient hypermutagenesis accelerates the evolution of legume endosymbionts following horizontal gene transfer. *PLoS Biol*. 12.

Remigi P, Zhu J, Young JPW, Masson-Boivin C. 2016. Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. *Trends Microbiol*. 24:63-75.

Rocha EPC. 2006. Inference and analysis of the relative stability of bacterial chromosomes. *Mol Biol Evol.* 23:513-522.

Sachs JL, Skophammer RG, Bansal N, Stajich JE. 2014. Evolutionary origins and diversification of proteobacterial mutualists. *Proc Biol Sci*. 281:9.

Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, et al. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415:497-502.

Sheetlin SL, Park Y, Frith MC, Spouge JL. 2014. Frameshift alignment: statistics and post-genomic applications. *Bioinformatics* 30:3575-3582.

Soupène E, Foussard M, Boistard P, Truchet G, Batut J. 1995. Oxygen as a key developmental regulator of *Rhizobium meliloti* N_2-fixation gene expression within the alfalfa root nodule. *Proc Natl Acad Sci USA.* 92:3759-3763.

Sullivan J, Ronson C. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci USA.* 95:5145-5149.

Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. 1995. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci USA.* 92:8985-8989.

Sun Z, Chen Y, Yang C, Yang S, Gu Y, Jiang W. 2015. A novel three-component system-based regulatory model for D-xylose sensing and transport in *Clostridium beijerinckii*. *Mol Microbiol.* 95:576-589.

Tang X, Xiao Y, Zhou JM. 2006. Regulation of the type III secretion system in phytopathogenic bacteria. *Mol Plant Microbe Interact.* 19:1159-1166.

Tans-Kersten J, Brown D, Allen C. 2004. Swimming motility, a virulence trait of *Ralstonia solanacearum*, is regulated by FlhDC and the plant host environment. *Mol Plant Microbe Interact*. 17:686-695.

Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, Wang S, Wang J, Gilbert LB, Li YR, et al. 2012. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc Natl Acad Sci USA.* 109:8629-8634.

Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fevre F, Longin C, Mornico D, Roche D, et al. 2013. MicroScope-an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res*. 41:E636-E647.

Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Médigue C. 2006. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res*. 34:53-65.

Valls M, Genin S, Boucher C. 2006. Integrated regulation of the type III secretion system and other virulence determinants in *Ralstonia solanacearum*. *PLoS Pathog.* 2:798-807.

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.

Yao SY, Luo L, Har KJ, Becker A, Ruberg S, Yu GQ, Zhu JB, Cheng HP. 2004. *Sinorhizobium meliloti* ExoR and ExoS proteins regulate both succinoglycan and flagellum production. *J Bacteriol*. 186:6042-6049.

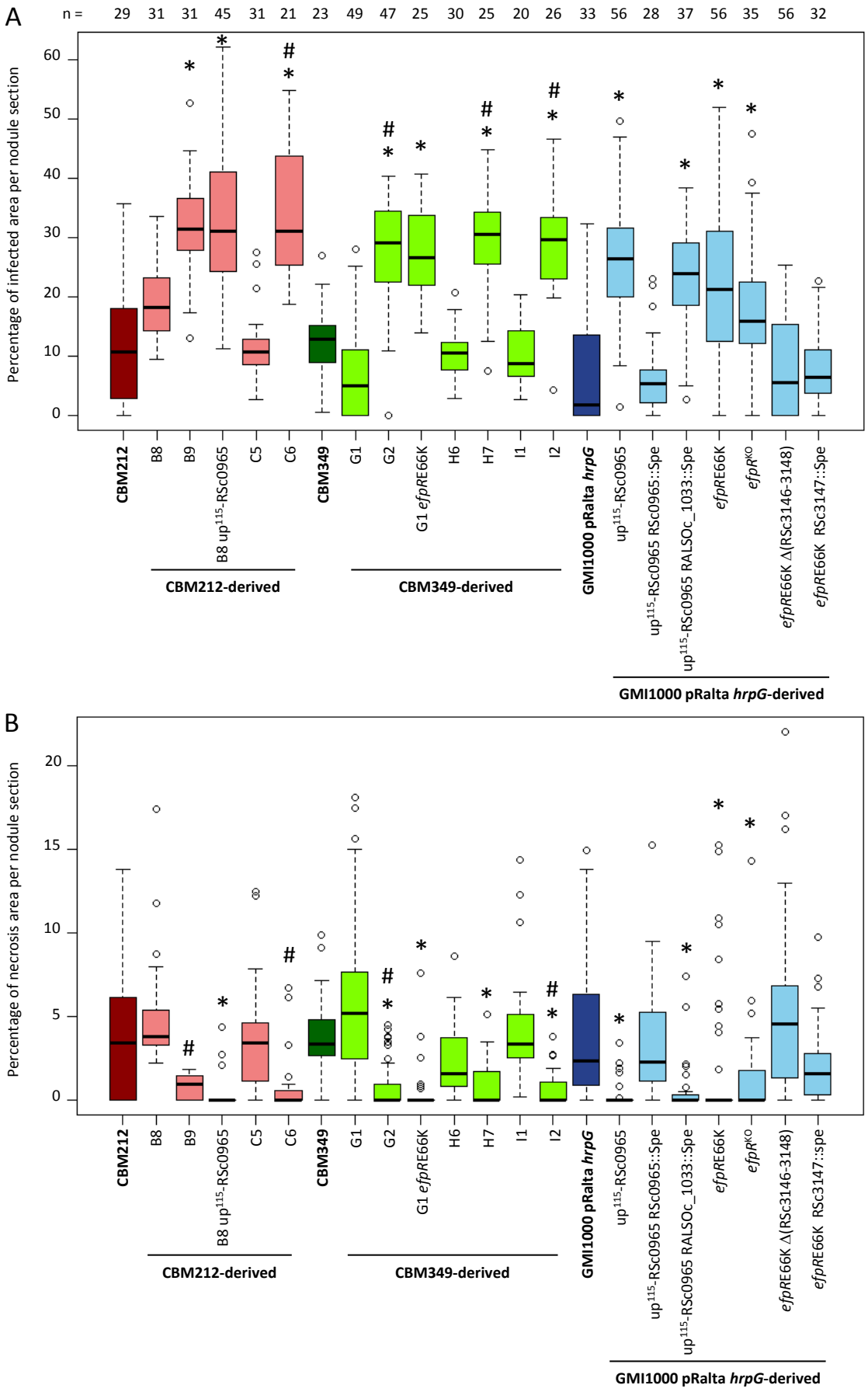
A
n = 29 31 31 45 31 21 23 49 47 25 30 25 20 26 33 56 28 37 56 35 56 32
Percentage of infected area per nodule section
0 10 20 30 40 50 60
CBM212
B8
B9
B8 up115-RSc0965
C5
C6
CBM349
G1
G2
G1 efpRE66K
H6
H7
I1
I2
GMI1000 pRalta hrpG
up115-RSc0965
up115-RSc0965 RSc0965::Spe
up115-RSc0965 RALSOc_1033::Spe
efpRE66K
efpKO
efpRE66K Δ(RSc3146-3148)
efpRE66K RSc3147::Spe
CBM212-derived
CBM349-derived
GMI1000 pRalta hrpG-derived
B
Percentage of necrosis area per nodule section
0 5 10 15 20
CBM212
B8
B9
B8 up115-RSc0965
C5
C6
CBM349
G1
G2
G1 efpRE66K
H6
H7
I1
I2
GMI1000 pRalta hrpG
up115-RSc0965
up115-RSc0965 RSc0965::Spe
up115-RSc0965 RALSOc_1033::Spe
efpRE66K
efpKO
efpRE66K Δ(RSc3146-3148)
efpRE66K RSc3147::Spe
CBM212-derived
CBM349-derived
GMI1000 pRalta hrpG-derived

Fig. 1

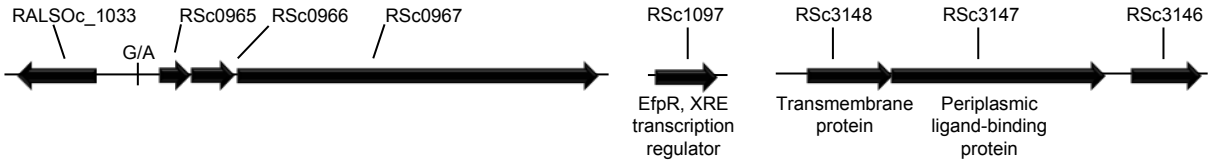
RALSOc_1033
G/A
RSc0965
RSc0966
RSc0967
RSc1097
EfpR, XRE transcription regulator
RSc3148
Transmembrane protein
RSc3147
Periplasmic ligand-binding protein
RSc3146

Fig. 2

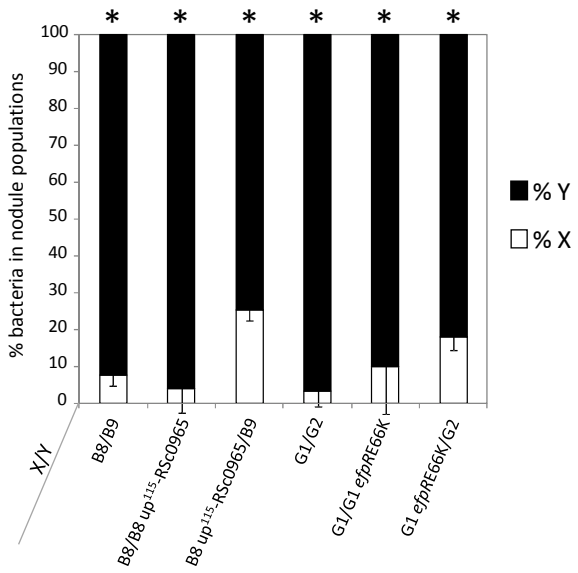
*
*
*
*
*
*
100
90
80
70
60
50
40
30
20
10
0
% bacteria in nodule populations
% Y
% X
X/Y
B8/B9
B8/B8 up115-RSc0965
B8 up115-RSc0965/B9
G1/G2
G1/G1 efpRE66K
G1 efpRE66K/G2

Fig. 3.

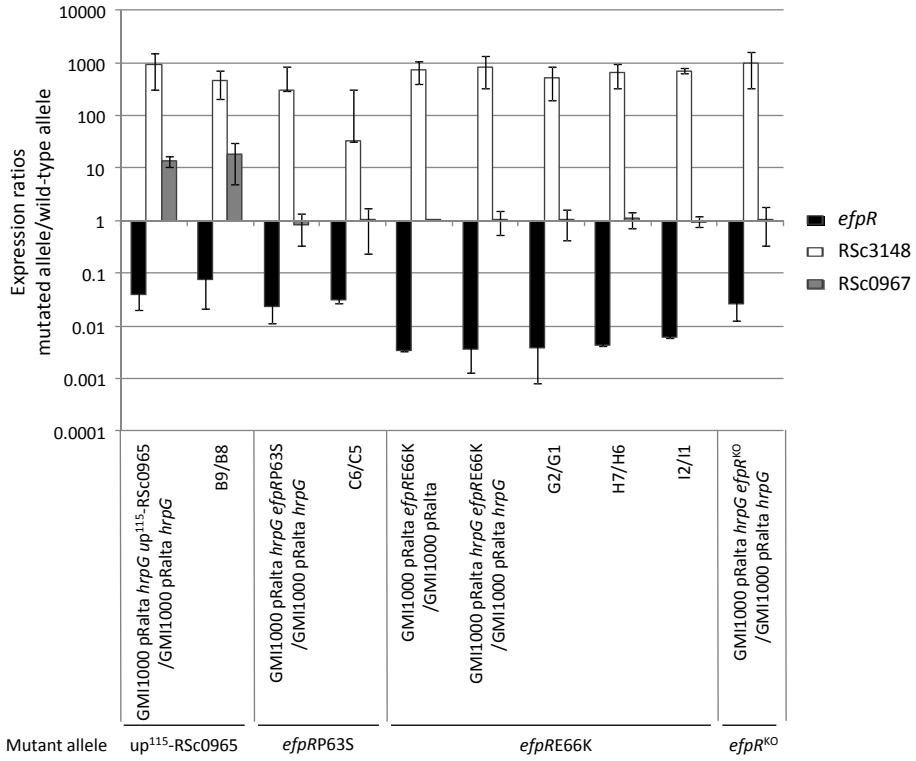
10000
1000
100
10
1
0.1
0.01
0.001
0.0001
Expression ratios
mutated allele/wild-type allele
efpR
RSc3148
RSc0967
GMI1000 pRalta hrpG up115-RSc0965 /GMI1000 pRalta hrpG
B9/B8
GMI1000 pRalta hrpG efpRP63S /GMI1000 pRalta hrpG
C6/C5
GMI1000 pRalta efpRE66K /GMI1000 pRalta
GMI1000 pRalta hrpG efpRE66K /GMI1000 pRalta hrpG
G2/G1
H7/H6
I2/I1
GMI1000 pRalta hrpG efpRKO /GMI1000 pRalta hrpG
Mutant allele
up115-RSc0965
efpRP63S
efpRE66K
efpRKO

Fig. 4

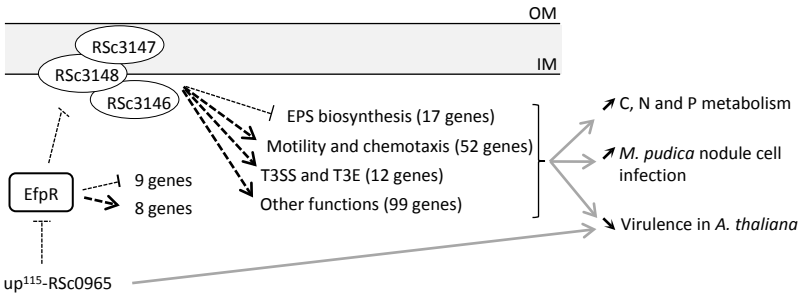
OM
IM
RSc3147
RSc3148
RSc3146
EPS biosynthesis (17 genes)
Motility and chemotaxis (52 genes)
T3SS and T3E (12 genes)
Other functions (99 genes)
EfpR
9 genes
8 genes
up115-RSc0965
C, N and P metabolism
M. pudica nodule cell infection
Virulence in A. thaliana

Fig. 5

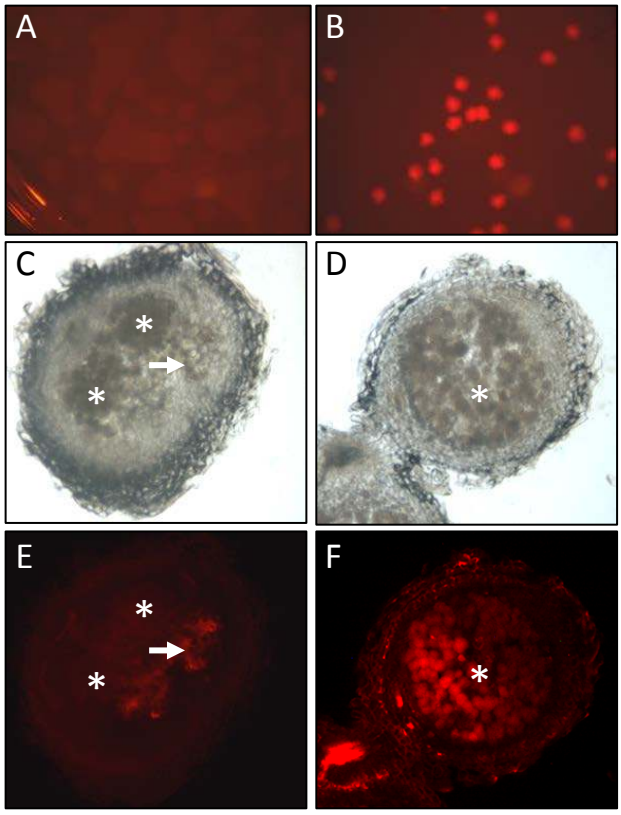

Fig.6

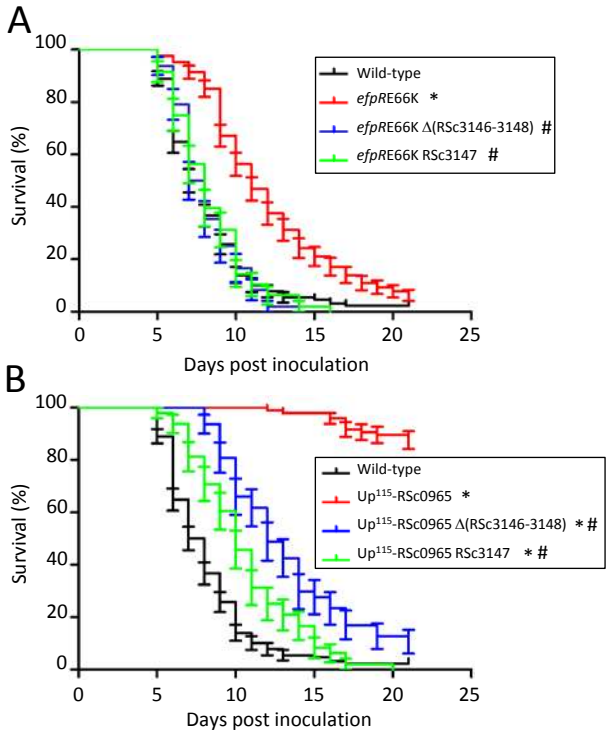
A
100
80
60
40
20
0
Survival (%)
0
5
10
15
20
25
Days post inoculation
Wild-type
efpRE66K *
efpRE66K Δ(RSc3146-3148) #
efpRE66K RSc3147 #
B
100
80
60
40
20
0
Survival (%)
0
5
10
15
20
25
Days post inoculation
Wild-type
Up115-RSc0965 *
Up115-RSc0965 Δ(RSc3146-3148) * #
Up115-RSc0965 RSc3147 * #

Fig. 7

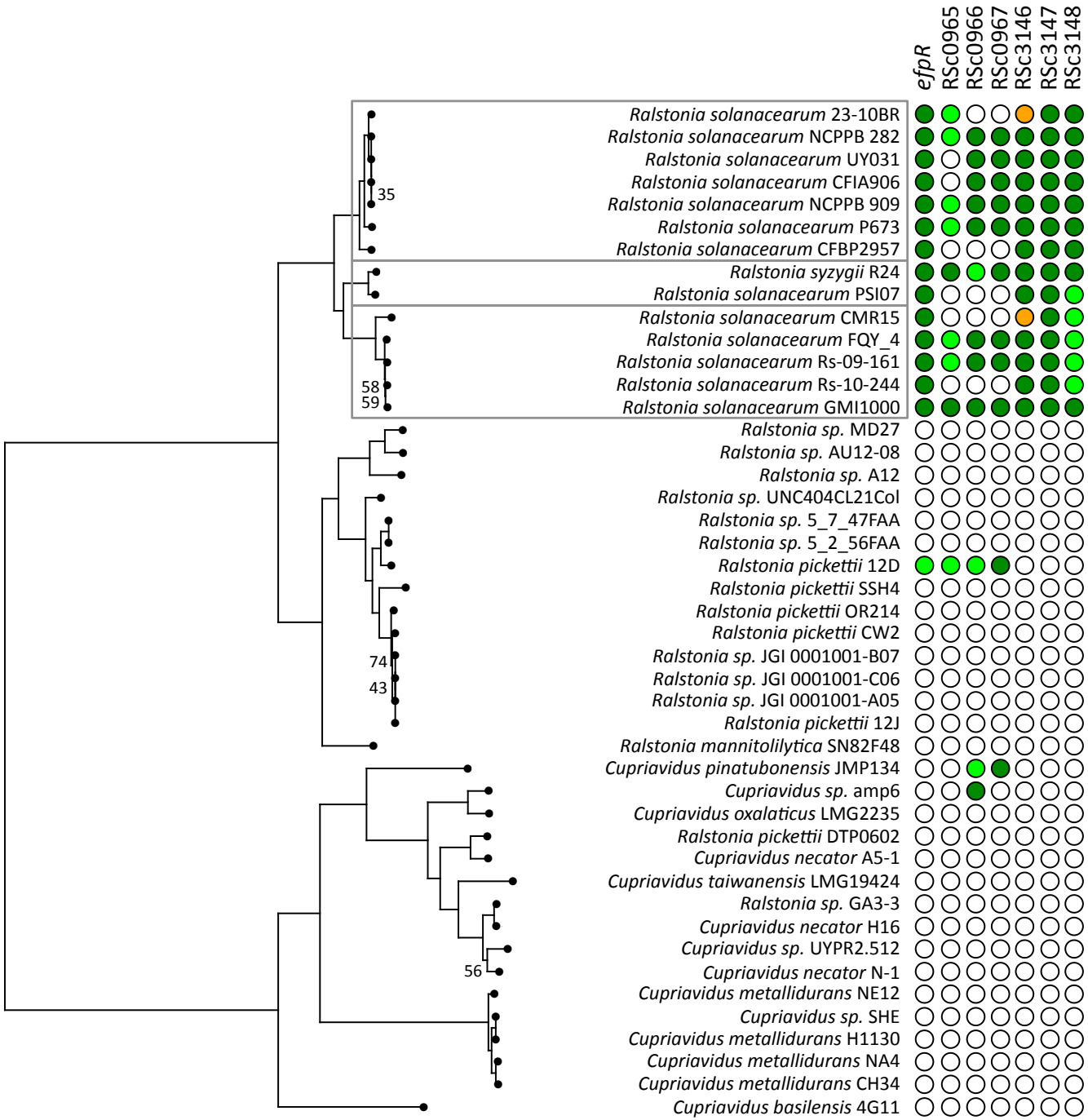
efpR
RSc0965
RSc0966
RSc0967
RSc3146
RSc3147
RSc3148
Ralstonia solanacearum 23-10BR
Ralstonia solanacearum NCPPB 282
Ralstonia solanacearum UY031
Ralstonia solanacearum CFIA906
Ralstonia solanacearum NCPPB 909
Ralstonia solanacearum P673
Ralstonia solanacearum CFBP2957
Ralstonia syzygii R24
Ralstonia solanacearum PSI07
Ralstonia solanacearum CMR15
Ralstonia solanacearum FQY_4
Ralstonia solanacearum Rs-09-161
Ralstonia solanacearum Rs-10-244
Ralstonia solanacearum GMI1000
Ralstonia sp. MD27
Ralstonia sp. AU12-08
Ralstonia sp. A12
Ralstonia sp. UNC404CL21Col
Ralstonia sp. 5_7_47FAA
Ralstonia sp. 5_2_56FAA
Ralstonia pickettii 12D
Ralstonia pickettii SSH4
Ralstonia pickettii OR214
Ralstonia pickettii CW2
Ralstonia sp. JGI 0001001-B07
Ralstonia sp. JGI 0001001-C06
Ralstonia sp. JGI 0001001-A05
Ralstonia pickettii 12J
Ralstonia mannitolilytica SN82F48
Cupriavidus pinatubonensis JMP134
Cupriavidus sp. amp6
Cupriavidus oxalaticus LMG2235
Ralstonia pickettii DTP0602
Cupriavidus necator A5-1
Cupriavidus taiwanensis LMG19424
Ralstonia sp. GA3-3
Cupriavidus necator H16
Cupriavidus sp. UYPR2.512
Cupriavidus necator N-1
Cupriavidus metallidurans NE12
Cupriavidus sp. SHE
Cupriavidus metallidurans H1130
Cupriavidus metallidurans NA4
Cupriavidus metallidurans CH34
Cupriavidus basilensis 4G11
35
58
59
74
43
56

Fig. 8