Genetic variation in host-specific competitiveness of the symbiont Rhizobium leguminosarum symbiovar viciae
Stéphane Boivin, Frederic Mahé, Frederic Debellé, Marjorie Prevent, Mathilde Tancelin, Marc Tazun, Jerzy Wielbo, Sylvie Mazurier, Peter Young, Marc Lepetit

To cite this version:
Stéphane Boivin, Frederic Mahé, Frederic Debellé, Marjorie Prevent, Mathilde Tancelin, et al.. Genetic variation in host-specific competitiveness of the symbiont Rhizobium leguminosarum symbiovar viciae. Frontiers in Plant Science, Frontiers, 2021, 12, 10.22541/au.159237007.72934061 . hal-02967193

HAL Id: hal-02967193
https://hal.inrae.fr/hal-02967193
Submitted on 14 Oct 2020

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

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

Distributed under a Creative Commons Attribution 4.0 International License
Genetic variation in host-specific competitiveness of the symbiont *Rhizobium leguminosarum* symbiovar *viciae*

Stéphane Boivin¹, Frederic Mahé², Marjorie Prevent¹, Mathilde Tancelin¹, Marc Tauzin¹, Jerzy Wielbo³, Sylvie Mazurier⁴, Peter Young⁵ and Marc Lepetit¹

¹ Laboratoire des Symbioses Tropicales et Méditerranéennes (LSTM) INRA, IRD, CIRAD, University of Montpellier, Montpellier SupAgro, Montpellier, France

² Biologie et Génétique des Interactions Plante-Parasite (BGPI), University of Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

³ Department of Genetics and Microbiology, Maria Curie-Sklodowska University, Lublin, Poland

⁴ Agroecology, AgroSup Dijon, INRA, University of Bourgogne Franche-Comté, F-21000 Dijon, France

⁵ Department of Biology, University of York, York, United Kingdom

Corresponding author: Marc Lepetit

Running title: Rhizobial competition for nodule formation

Abstract: 212 words

Introduction: 1164 words

Material and Methods: 942 words

Results: 2617 words

Discussion: 1772 words
Abstract

Legumes in the tribe Fabaceae form nitrogen-fixing root nodules resulting from symbiotic interaction with the soil bacteria *Rhizobium leguminosarum* symbiovar *viciae* (*Rlv*). These bacteria are all potential symbionts of the Fabaceae hosts but display variable competitiveness to form root nodules (CFN) when co-inoculated in mixture. Because CFN and nitrogen-fixation capacity behave generally as genetically independent traits, the efficiency of symbiosis is often suboptimal when Fabaceae legumes are exposed to natural bacterial populations present in soil. A core collection of 32 bacteria was constituted based on the genomic comparison of a collection of 121 genome sequences representative of known worldwide diversity of the rhizobium symbiovar. A variable part of the *nodD* gene sequence was used as a DNA barcode to discriminate and quantify each of the 32 bacteria in a mixture. This core collection was co-inoculated on a panel of 9 genetically diverse *Pisum sativum*, *Vicia faba* and *Lens culinaris* cultivars. We estimated the relative CFN of the bacteria with the diverse hosts by DNA metabarcoding on the nodulated root systems. Comparative genomic analyses within the bacterial core collection allowed us to identify genes associated with host-dependent CFN. These results highlight the emergent properties of rhizobial populations and present a new strategy to identify genetic markers related to important symbiotic traits operating at this level.

Keywords: *Rhizobium leguminosarum* symbiovar *viciae*, competitiveness, core collection, pea, fababean, lentil, Fabaceae, DNA metabarcoding, symbiosis.
Introduction

Legumes can escape nitrogen-deficit conditions in soil by interacting with bacteria known as rhizobia to form nitrogen-fixing root nodules. Rhizobia in natural populations are often genetically diverse (Bailly et al., 2011). They are generally gathered in ‘symbiovars’ based on their host specificity (Triplett & Sadowsky, 1992). The Ability to Form root Nodules (AFN) in mono-inoculation with specific legume tribes defines this host specificity. For most of the legume-rhizobium symbioses, AFN is strongly associated with horizontally transferred symbiosis-related regions of the genome, present on specific plasmids or islands (Young, 2016). These regions carry the nod genes involved in the synthesis and secretion of lipo-chito-oligosaccharide Nod Factors (NFs) recognized by the plant. Therefore AFN is associated with the diversity of the nod genes rather than with that of the chromosome (Triplett & Sadowsky, 1992; Kumar et al., 2015). Rhizobium leguminosarum symbiovar viciae (Rlv) is the specific symbiont of the Fabaeae tribe that includes important crops such as pea (Pisum sativum L.), fababean (Vicia faba L.) and lentil (Lens culinaris L.). Our knowledge of Rlv genomic diversity has recently been improved by the release of 50 diverse genome sequences in Genbank (Boivin et al., 2020). Rlv is a complex species composed of at least 5 different genospecies that are not symbiovar-specific (Kumar et al., 2015; Boivin et al., 2020; Cavassim et al., 2020). The nodD gene encodes a transcriptional regulator of the bacterial symbiosis genes. It has frequently been used as a marker to discriminate R. leguminosarum symbiovars (Zézé, Mutch, & Young, 2001; Laguerre, Louvrier, Allard, & Amarger, 2003; Boivin et al., 2020). Although Rlv bacteria generally have the capacity to form root nodules with most Fabaeae legumes, competition occurs within Rlv mixtures and only bacteria displaying the best Competitiveness to Form root Nodules (CFN) finally occupy the roots of the host plants (for review see Boivin &
Lepetit, 2020). Therefore the *Rlv* populations associated with Fabaeae root nodules do not necessarily reflect the proportions of bacteria in the soil. CFN varies greatly depending on both the legume host and the *nod* alleles (Boivin et al., 2020). Generally, a poor association was found between CFN and the level of Symbiotic N\textsubscript{2} Fixation (SNF; Bourion et al., 2018; Boivin et al., 2020). This likely explains the frequent failure of inoculation strategies with highly effective *Rlv* strains, due to a higher competitiveness of indigenous ineffective bacteria as compared to inoculated strains (Fesenko, Provorov, Orlova, Orlov, & Simarov, 1995; McKenzie et al., 2001; Laguerre et al., 2003). Early partner choice, AFN and CFN are not the only mechanisms responsible for partner choice between symbiotic partners. As the young nodule becomes N\textsubscript{2}-fixing, local and systemic post-infection mechanisms related to the plant nitrogen demand pilot the development of the nascent nodule and may sanction inefficient bacteria (for review, see Boivin and Lepetit, 2020).

Little is known about mechanisms controlling CFN. They might be related either to plant-microbe and/or to microbe-microbe interactions. Antibiosis and quorum-sensing mechanisms modulating the multiplication of free-living *Rlv* bacteria have been reported and are potentially involved in CFN (Robleto, Scupham, & Triplett, 1997; McAnulla, Edwards, Sanchez-Contreras, Sawers, & Downie, 2007; Naamala, Jaiswal, & Dakora, 2016). However, even if the preferential proliferation of certain rhizobium genotypes within host rhizospheres contributes to CFN, the plant-microbe interaction is probably a major driver of CFN (Moawad, Ellis, & Schmidt, 1984; Laguerre et al., 2003; for review see Boivin and Lepetit, 2020). Based on co-inoculation strategies with high densities of *Rlv* mixtures, recent studies showed that pea and fababean preferentially select different *Rlv* genotypes (Boivin et al., 2020). Candidate genes and/or genetic markers associated with pea/fababean CFN were identified. Most of
these genomic sequences belonged to Rlv plasmids, in agreement with the hypothesis that they are components of horizontally transferred symbiotic traits. They included some nodulation genes (nod genes) such as nodM, nodN, nodT and nodO. These genes were previously identified as highly polymorphic (Jorrin & Imperial, 2015), and suspected to be involved in rhizobial host specificity (Djordjevic, Schofield, & Rolfe, 1985; Surin & Downie, 1988; Lewis-Henderson & Djordjevic, 1991; Baev et al., 1992). Rhizobia produce a large diversity of NFs that bind to legume root LysM-RLK receptors (Oldroyd, Murray, Poole, & Downie, 2011). NFs are composed of a chitin-like N-acetyl glucosamine backbone with a fatty acyl chain at the non-reducing end, and carry various substitutions such as glycosylation, acetylation and/or sulfation on the backbone (Mergaert, Montagu, & Holsters, 1997). These modifications influence the binding between NFs and LysM-RLKs and modulate the establishment of the symbiosis (Dénarié, Debellé, & Rosenberg, 1992). For instance, the nodX gene has been reported as crucial for the specific partner choice between the Pisum sativum cultivar ‘Afghanistan’, carrying the SYM2 locus, and the Rlv strain TOM (Davis, Evans, & Johnston, 1988). The nodX gene encodes an acetyltransferase that modifies the NFs secreted by the bacteria, probably allowing them to bind with a specific LysM-RLK protein encoded within the SYM2 locus (Firmin, Wilson, Carlson, Davies, & Downie, 1993; Hogg, Davies, Wilson, Bisseling, & Downie, 2002; Sulima et al., 2017). This SYM2-nodX association represents a well-documented example of a mechanism restricting AFN. Although genetic association of nod gene diversity with CFN argues for the interaction of NFs with LysM-RLK having an important role in CFN, the underlying mechanisms are still unknown. Other mechanisms, related to plant recognition of bacterial surface polysaccharides or bacterial effectors, have been implicated in modulation of the legume-rhizobium interaction and therefore may
contribute to CFN (Janczarek, Rachwał, Marzec, Grządziel, & Palusińska-Szysz, 2015; Miwa & Okazaki, 2017).

In most published co-inoculation experimental strategies, bacteria were inoculated with a reference strain or with a limited number of strains (Triplett & Sadowsky, 1992; Laguerre et al., 2003; Bourion et al., 2018). Recently, a co-inoculation strategy with multiple Ensifer meliloti strains was applied on two Medicago truncatula genotypes to investigate the impact of the partner choice diversity on symbiotic traits (Epstein et al., 2018). However, phenotyping was done five weeks post-inoculation in this study, which did not allow a focus on early symbiotic traits such as CFN that may potentially be compensated by post-infection mechanisms. Taking the new opportunities offered by both NGS and DNA metabarcoding, we designed a strategy to estimate bacterial CFN in Fabae roots inoculated with Rlv populations, and to identify genes potentially associated with contrasted CFN phenotypes. We defined a core collection representative of the genomic diversity of the symbiovar viciae. We identified a nodD DNA barcode to discriminate and quantify each rhizobium of the core collection individually within a nodulated root system, using high throughput NGS. We inoculated the core collection in a mixture on 9 diverse genotypes of Pisum sativum, Vicia faba and Lens culinaris in order to compare their CFN with different Fabae hosts. We used saturating amounts of each strain to focus on plant-microbe interactions and to reduce impacts of differential bacterial growth. We characterized host-specific CFN profiles that varied among bacteria of the core collection. Using a comparative genomic approach we identified bacterial genes associated with host-specific CFN in the various hosts.
Material and methods

Bacterial collection, inoculation and plant growth conditions

Bacteria from different geographical origins isolated from *Pisum sativum, Vicia faba, Lens culinaris* or *Lathyrus pratensis* root nodules were obtained (Table S1). *Pisum sativum, Vicia faba* and *Lens culinaris* seeds were surface sterilized in 3% calcium-hypochlorite solution for 10 min, washed 5 times in sterilized water, and sown in 2L pots filled with sterilized perlite/sand (3/1). Bacterial strains were grown individually in YEM broth medium. For each culture, the number of colony forming units (CFU) was estimated by dilution plating on YEM medium. The identity of the bacteria was confirmed by PCR amplification and sequencing of the *nodD* gene. Then, bacteria were mixed together in equal amounts (10^7 CFU/mL/strain) to establish the complex inoculum. We inoculated different plant cultivars of *Pisum sativum* (cultivars ‘Kayanne’, ‘Isard’ and ‘Afghanistan’ pea), *Lens culinaris* (cultivars ‘Rosana’, ‘Anicia’ and ‘Flora’) and *Vicia faba* (cultivars ‘Diva’, ‘Organdi’ and ‘Tiffany’). Seeds were inoculated with the complex inoculum directly after sowing (3 pots of 4 seeds were used for each condition). Plants were grown under high-pressure sodium lamps with a mean photosynthetically active radiation of 250 µmol photons m^{-2}.s^{-1} (16h/8h 22/18°C day/night cycle). They were supplied with N-free nutrient solution (K_2HPO_4 0.8mM, MgSO_4 1 mM, K_2SO_4 0.6 mM, CaCl_2 2 mM, NaCl 0.2 mM adjusted to pH 6.5) twice a week. Plants were harvested 14 days after inoculation for *Pisum sativum* and *Lens culinaris*, and 21 days after inoculation for *Vicia faba*.

DNA extraction, PCR amplification and metabarcoding analysis

The nodulated root systems of the 4 plants in each pot were pooled together to form one DNA metabarcoding biological replicate. Three separate biological replicates
were used to determine each rhizobial CFN profile (12 plants). Whole nodulated root systems were ground in liquid N\textsubscript{2} and DNA extractions were made using the DNeasy Plant Mini Kit (www.qiagen.com). PCR amplifications of the nodD309 barcode sequences were performed using Phusion High-Fidelity DNA Polymerase (www.thermofisher.com) and specific primers and conditions (Table S2). Size of amplicons was checked on agarose gels before sequencing using Illumina MiSeq technology with a 2x250bp paired end protocol, performed at the Genotoul GeT-PlaGe facility (get.genotoul.fr). Single multiplexing was performed using a homemade 6 bp index, which was added to R784 during a second PCR with 12 cycles using specific primers (Table S2). The resulting PCR products were purified and loaded onto the Illumina MiSeq cartridge according to the manufacturer’s instructions. Paired Illumina MiSeq reads were assembled with vsearch v2.9.1 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) using the command fastq_mergepairs and the option fastq_allowmergestagger. Demultiplexing and primer clipping were performed with cutadapt v1.9 (Martin, 2011) forcing a full-length match for sample tags and allowing a 2/3rd-length partial match for forward and reverse primers. Only reads containing both primers were retained. For each trimmed read, the expected error was estimated with vsearch’s command fastq_filter and the option eeout. Each sample was then dereplicated (i.e. strictly identical reads were merged) using vsearch’s command derep_fulllength, and converted to FASTA format. To prepare for clustering, the samples were pooled and processed with another round of dereplication. Files containing expected error estimates were also dereplicated to retain only the lowest expected error for each unique sequence. To detect potential contaminants, the dereplicated data were further clustered with swarm v2.1.9 (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2015), and checked for chimeras using vsearch’s command
uchime_denovo (Edgar, Haas, Clemente, Quince, & Knight, 2011). As no significant contamination was detected, downstream analyses and results are based on unclustered data, only retaining reads strictly identical to the 32 expected *Rhizobium* nodD309 reference sequences (Table S2), yielding a total of 511294 reads for all replicates. For each plant cultivar, the read number of the four replicates has been averaged, yielding a mean of 14203 reads per condition (Table S3). We calculated the CFN index (CFNi) of a bacterium in a plant host as the percentage of the nodD309 sequences of the rhizobium of interest to the total number of nodD309 sequences generated by all rhizobia of the sample. The CFN indexes (%) of each rhizobium of the core collection in plant genotypes were the mean of 4 biological repeats (Table S3).

**Genome sequencing, genomic and association genetic analysis**

Bacterial genomes were sequenced by MicrobesNG (Birmingham, UK, [https://microbesng.uk/](https://microbesng.uk/)) on an Illumina HiSeq platform using a 2x250bp paired end protocol. Genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA). High-quality paired reads were assembled by the Galaxy/BBRIC pipeline ([https://bbric-pipelines.toulouse.inra.fr/](https://bbric-pipelines.toulouse.inra.fr/)) and genome annotations were performed using EuGene-PP ([Sallet et al, 2014](#)) and RAST (http://rast.nmpdr.org/). The pairwise ANI values were calculated using the JSpecies software ([http://jspecies.ribohost.com/jspeciesws](http://jspecies.ribohost.com/jspeciesws)). Heatmaps were built using the pheatmap R package (Kolde, 2015). Presence/absence of genes associated with CFN phenotypes were identified using the Bidirectional Best Hits (BBHs) tool available in RAST (http://rast.nmpdr.org/). The presence/absence of the specific genes identified was checked by a BLAST search in the genomes of the core collection (thresholds: nucleotide identity > 70%; sequence coverage > 70%). Kruskal-Wallis tests (eventually followed by a Dunn post-hoc test with correction for multiple testing) were performed
using R software (http://www.R-project.org/) to test whether the bacterial subgroups displayed equivalent CFN indexes (H₀ hypothesis), according to diverse criteria (genospecies/Nod types/groups with specific allele). The nucleotide sequences of the nodABCDEFIJKLMNOP genes were concatenated and aligned using ClustalOmega (www.ebi.ac.uk), and a Neighbor-Joining (NJ) phylogenetic tree was built using MEGA v7.0.26 (www.megasoftware.net). The comparisons between reference genomes, and between contigs and the genomic sequence of 3841, for genomic rearrangements/organization (Table S5 and Table S6) were performed using MAUVE software (Darling et al, 2004). Plasmid replicon classes (Rh groups) were identified by sequence similarity of the repA gene and presence of cognate repB and repC (Cavassim et al., 2020).

Results

Constitution of a bacterial core collection of Fabae rhizobial symbionts

At the beginning of this study we collected the 73 genome sequences of Fabae rhizobial symbiont available in GenBank. To maximize the diversity, the genomes of 48 additional rhizobia of more diverse geographical origins and/or carrying diverse sequences of the symbiotic marker nodD were sequenced (Table S1). They were isolated in previous studies from Fabae root nodules (Pisum sativum, Vicia faba, Lens culinaris and Lathyrus pratensis) in Bangladesh, China, Algeria, Poland, Germany, France, United Kingdom and Italy. As expected most of these bacteria were described as belonging to Rhizobium leguminosarum symbiovar viciae. Based on the genome comparison and using an Average Nucleotide Identity (ANI) threshold of 95%, we concluded that 117 bacteria belong the R. leguminosarum complex species
(ANI>90%). The *R. leguminosarum* complex species is composed of at least 7
genospecies (gs; Figure 1). Four strains were phylogenetically distant from the others
and therefore fell outside the *R. leguminosarum* species complex according the
criterion of ANI<90%, even though they shared the ability to form root nodules with
Fabae with the other bacteria of the collection. They belonged to other described
*Rhizobium* species: *R. binae* (BLR195), *R. pisi* (CZP3G4) and *R. anhuiense*
(CCBAU43229), except for SEF4G12 which belonged to an undescribed species close
to *R. acidisoli* FH23\textsuperscript{T} (ANI\textsubscript{m}=94.12%). The other 117 bacteria belonged to the *R.
leguminosarum* genospecies (ANI>90%). Most of them are included in the 7 *R.
leguminosarum* genospecies described previously (Kumar et al., 2015; Boivin et al.,
2020; Cavassim et al., 2020; gsA; gsB; gsC; gsD; gsE, gsF-1 and gsF-2/*R.
laguerreae*). A new *R. leguminosarum* genospecies, called gsG (Figure 1; Table S1),
was distinguished in accordance with the recently established taxonomic guidance (de
Lajudie et al., 2019). This genospecies was defined by grouping two different bacteria
isolated in China (CCBAU33195 and CCBAU11080), forming a discrete cluster
phylogenetically distant from all others (ANI<94%).

The 121 bacteria of the collection were also discriminated according to their
plasmid-borne symbiosis genes. A phylogenetic tree was constructed using the
sequence of the 11 conserved *nod* genes located on the symbiosis plasmid (Figure 2;
Table S1). In agreement with the *nod* gene phylogeny, the *Rlv* bacteria were
separated, independently of their genospecies, into 2 Nod types, named A and B, as
previously described (Boivin et al., 2020). These two types were subdivided into 10
Nod groups named A1 to A5 and B1 to B5 respectively. There was considerable
sequence variations within the Nod groups, except for the 22 strains belonging to the
B1 group, which had few differences. A total of 38 gs/Nod group combinations were
identified within the 121 bacteria, confirming the diversity of the association between
the symbiosis plasmid and the genospecies (Table S7). The number of bacteria per
combination is variable. For instance, 19 strains carried the combination gsE/A1
whereas only one strain had gsB/A2 (Table S7). The nodX gene located within the nod
gene cluster was present in 27/121 isolates (Figure 2).

We defined a DNA barcode to discriminate and quantify each bacteria in
mixture. The 16S or gyrB marker sequences, frequently used in DNA metarbarcoding
strategies, did not display sufficient intraspecific genetic variation among the 121 Rlv
bacteria (data not shown). Furthermore, they belonged to the chromosome and were
genetically unlinked to nodulation phenotypes. Alternatively, a 309bp fragment of the
nodD gene located on the symbiosis plasmid allowed 32 nodD309 alleles to be
discriminated among the 121 Rlv (Table S2), and covered all Nod groups (Table S7).
Based on our ability to discriminate the bacteria in a mixture, as well as criteria of
geographical origins, plant hosts used for isolation, diversity of gs/Nod group
combinations and presence/absence of nodX, 32 representative bacteria were
selected to constitute a fabeae symbiont core collection (Table 1; Figure S1; Table S2).
Most geographical origins (87%), plant hosts (100%) and gs/Nod group combinations
(67%), and bacteria with nodX (12/32), were represented (Table S1 and Table S7).
Nevertheless, it is noteworthy that, because of the limited variation of the barcode, we
could introduce only one strain representing the B1 group.

Competitiveness to Form root Nodules (CFN) varies across host plants

The 32 Rlv bacteria of the core collection formed the complex inoculum. We
combined saturating densities of each of the 32 bacteria (10^7 CFU/mL for each strain)
to minimize the putative impact of bacterial growth on nodulation success. We
inoculated different plant genotypes of *Pisum sativum*, *Lens culinaris* and *Vicia faba* with the core collection. They covered a large genetic diversity of these legume crops. Root systems were harvested once nodules had formed and the nodD309 DNA barcode was PCR-amplified from total DNA. The sequencing of the PCR product yielded numbers of nodD309 sequences, specific to each member of the core collection. Among the 32 bacteria, 6 were removed for the further analyses as they were not detected in any sample from any host (Table S8). We calculated the CFN index (CFNi) for each bacterial strain in each host plant. The CFNi of the 26 bacteria detected in metabarcoding ranged from 0% to 81.9% (Table S8). The mean CFNi on each plant genotype was always around 3.8%, and the mean CFNi per bacterium in all hosts ranged from 0.003% to 32.2%.

A hierarchical clustering, based on the CFNi of the 26 bacteria of the core collection detected in metabarcoding, separated plant hosts clearly into 4 groups with distinct competitiveness profiles (Figure 3). These 4 groups were composed respectively of the *Vicia faba* genotypes, the *Lens culinaris* genotypes, the *Pisum sativum* cultivars ‘Kayanne’ and ‘Isard’ (cultivated spring and winter peas), and the *Pisum sativum* cultivar ‘Afghanistan’ (Figure 3). For bacteria, the clustering was less marked than for plants. However, 4 clusters may be separated according to host-specific CFN profiles of bacteria (Figure 3; Table S9).

**NodX is required but not sufficient for nodulation of *Pisum sativum* cv. ‘Afghanistan’**

The nodulation profile of the *Pisum sativum* cultivar ‘Afghanistan’ was highly divergent compared to the other pea cultivars (Figure 3). The data confirmed that the bacterial *nodX* gene is required for nodulation with this plant genotype. All competitive
strains with this pea genotype, representing 89% of the Rlv barcode of the root (BLR195, CCBAU11080 and CCBAU10279), carried the nodX gene (Figure 3). These bacteria were not competitive with other pea cultivars. However, when the whole set of bacteria was considered, neither the presence of nodX, nor the Nod type, nor the genospecies were solely associated with nodulation success on this host (Table S10 and Table S4). Although the 4 most competitive isolates had the nodX gene, 8 other isolates of the core collection, belonging to Nod type A or B, also had the nodX gene, but were not competitive with this cultivar (Table 1 and Table S4). A significant association with CFN was identified only with bacteria having both the nodX gene and Nod type B (Table S4). Mono-inoculation experiments with several strains carrying nodX indicated that only strains of Nod type B carrying this gene were able to form nodules with the pea cultivar ‘Afghanistan’, whereas the strains of Nod type A were unable (Table S11). We concluded that the nodX gene was necessary but not sufficient to confer the ability to nodulate this host genotype and genetic determinants related to Nod type B were also required. Nevertheless, there was a wide range of CFN within the Nod type B strains carrying nodX and thus able to nodulate the cultivar ‘Afghanistan’. We concluded that genetic factors independent of NodX, present in Nod type B, were responsible for CFN variations in the context of host-specific ability to form nodule with this cultivar.

**Host-specific CFN factors are associated with Rlv plasmids and/or genospecies**

Beyond the particular case of pea cultivar ‘Afghanistan’, the three other groups of pea, lentil and fababean cultivars displayed contrasted CFN profiles. The Rlv strain 3841 was predominant in the three Vicia faba genotypes (from 70% to 82% of the total Rlv reads in fababean roots), but not at the same level in pea or lentil genotypes. The Rlv strain GLR2 was found in nodulated roots of the three Lens culinaris genotypes,
but not in other plant species (Figure 3 and Table S8). Behaviors did not depend only on the plant species, and the CFN profiles also varied between genotypes within the three plant species. For example, *Rlv* strains 3841 and L113 were differentially competitive with the pea cultivars ‘Kayanne’ and ‘Isard’ (Figure 3). Concerning the CFN in *Vicia faba* genotypes, the predominance of 3841 confirmed our previous study indicating that the Nod type B (and especially the Nod group B1) is strongly associated with CFN on *Vicia faba*. Unfortunately, as 3841 is the only strain of the core collection belonging to the Nod group B1, it was not possible to extend the comparison to other B1 strains in order, potentially, to associate other genetic variation with CFN. Consequently, further investigations were restricted to pea cultivar ‘Kayanne’ and ‘Isard’ as well as to the three lentil cultivars.

Many bacteria of the core collection displayed a wide range of CFN variation with these hosts. We investigated CFN associations with the Nod type, the Nod group, or the genospecies of bacteria at two levels of plant diversity: the plant cultivar (Table S10) and the plant species (Table S12). Because the core collection had a limited number of bacteria for each category (Nod type, Nod group, and genospecies), and it was tested with a limited number of plant genotypes, the power of the statistical analysis was limited. However, the data showed global effects of bacterial genotype variation on CFN with lentil and/or pea at the level of the cultivar (Table S10). The Nod type was associated with the CFN phenotype in the two pea cultivars ‘Isard’ and ‘Kayanne’, but not in lentil. An association with the Nod group was found with ‘Isard’. The genospecies was significantly associated with the CFN phenotype in the *Pisum sativum* cultivar ‘Isard’ as well as in lentil cultivars ‘Anicia’ and ‘Flora’.

Investigations at the level of the plant species, combining the data obtained on the various cultivars, were more powerful. At this level, it was possible to test more
efficiently the global effect of the bacterial genotype variation, but also to identify by pairwise testing which category of the core collection may explain the variation (Table S12). Nod type A bacteria generally displayed generally higher CFN in the *Pisum sativum* cultivars ‘Kayanne’ and ‘Isard’. Bacteria of the core collection belonging to the Nod groups A1 and A4 had a higher mean CFN than those belonging to Nod groups B3, B4 and B5 (Table S12). The only exception is 3841, the unique representative of the B1 Nod group in the core collection, which displayed high CFN with pea. This particular result was somehow in contradiction with our previous observations indicating that natural isolates of this Nod group are generally good competitor with fababean but poorly competitive with pea (Boivin et al., 2020). Although CFNi of bacteria of the core collection belonging to various Nod groups varied widely in lentil, we got no evidence indicating that the Nod type or the Nod group could predict competitiveness with this plant host.

We found associations of genospecies with CFN in pea (particularly in the ‘Isard’ cultivar) and lentil. Some of them are common to both host species. For instance, bacteria of the core collection belonging to gsB often displayed high competitiveness in pea and lentil whereas the bacteria of gsF-1 always performed poorly with these hosts. Nevertheless, there were also host-specific examples. For example, bacteria belonging to gsF-2 were in the medium range of competitiveness in lentil but poor competitors in pea. Despite these global trends, associations were never systematic and exceptions were observed. Genomic investigations will be required to identify which parts of the *Rlv* genomes are responsible for these CFN variations.

**Contrasted CFN profiles are associated with the presence of specific bacterial genes**
Comparative genomics identified bacterial genetic factors associated with the contrasted nodulation profiles. For each plant genotype, we compared the 4-5 most competitive bacteria to the 4-5 least competitive to identify genes specifically present or absent. Then, we tested whether, in the whole set of 26 bacteria that were detected in the root systems, the presence/absence of the identified genes was significantly associated with the host-specific nodulation phenotype (Table S4). Using this strategy, the number of specific genes identified varied from 2 to 13 depending on the plant species and cultivar (Table S4).

In the particular case of the *Pisum sativum* cultivar ‘Afghanistan’, only a fraction of strains were able to form nodules with this host. Our first aim was to identify the genetic factors that, together with NodX, allow this host-specific nodulation. We restricted the first genomic comparison to the 8 bacteria of the core collection carrying the *nodX* gene and either Nod type B (these strains could nodulate this host) or Nod type A (these could not). Seventeen genes associated with nodulation of this host were identified by this comparison (Table S13 and Table S14). Our second aim was to identify genetic factors specifically associated with CFN in this particular host by comparing only bacteria able to form nodules with this host, i.e., the 5 bacteria of Nod type B with the *nodX* gene. These displayed contrasted CFN with their host (see above), and genomic comparisons identified a further 6 genes (Table 2; Table S4 and Table S14).

Similar genome comparisons were made for bacteria that displayed contrasted CFN in association with the *Pisum sativum* cultivars ‘Kayanne’ and ‘Isard’, as well as with lentil cultivars ‘Anicia’ and ‘Rosana’. This analysis yielded 36 genes potentially involved in the competitive success to form nodules of the core collection bacteria in mixture with pea (except cv ‘Afghanistan’) and lentil hosts (Table 2 and Table S14).
Homologs of these genes were sought in the 6 fully-assembled *Rlv* genomes described in Genbank: not only the 3841 reference strain but also Vaf10, Vaf108, BIHB1217, UPM791 and TOM. The replication and maintenance of plasmids in *R. leguminosarum* are controlled by the *repABC* genes, and the sequences of these genes fall into distinct clusters that allows ‘Rh’ incompatibility groups to be defined (Cavassim et al., 2020). The ‘Rh’ types of the replicons of the *Rlv* strains were characterized, in order to define series of homologous replicons (Table 2, Table S5, Table S6 and Table S14). However, we found plenty of evidence for variations and frequent rearrangements between replicons in these genomes. Most of the Vaf10, Vaf108, BIHB1217, UPM791 and TOM plasmids shared homologies with more than one 3841 plasmid, revealing that replicons (and particularly the distribution of sequences among replicons) vary greatly among the various *Rlv* bacteria (Table S5). For example, in strain BIHB1217, plasmid pPR4 (Rh08) shared sequence homologies with the pRL7 (Rh12), pRL8 (Rh13) and pRL10 (Rh03) plasmids of strain 3841. In the Vaf-108 genome, the chromosome shared homologies with both the 3841 chromosome and the pRL10 plasmid, indicating that rearrangements were probably not restricted to plasmids. This fluidity of the accessory genome was very apparent when we examined the location of the genes that showed significant associations with CFN or AFN (Table 2, Table S14). For example, the genes PI9 and PI10 are always adjacent in the fully assembled genomes, but they are on an Rh01 plasmid in Vaf-108, Rh03 in 3841, and Rh08 in Vaf-10, BIHB1217 and UPM791. In the genomes of BIHB1217 and UPM791, PI12 is on Rh02 and PI13 is on Rh04a, but in 3841 these genes are both on Rh05, just 2.5kb apart. The genes PAFN8-10 and PAFN17 are all on Rh12 in 3841, Vaf-10 and UPM791, but on Rh01 in Vaf-108, and Rh06 in BIHB1217. In TOM, PAFN8-10 are on Rh06, but PAFN17 is on Rh03. These examples
of genome rearrangement, and numerous others, indicated that the replicon location of a gene could not be reliably inferred from its location in another strain.

Discussion

A metabarcoding approach to measure competitiveness to form nodules

Selection pressures related to plant-microbe, microbe-microbe and/or microbe-environment interactions are likely important drivers of rhizobium populations in interaction with their hosts. Previous studies estimated the CFN of \( Rlv \) isolates individually by co-inoculating a host plant with both a strain of interest and a reference strain (Bourion et al, 2018; Boivin et al., 2020). Selecting antibiotic resistant mutants (Amarger 1981) or introducing fluorescent markers on plasmids has greatly improved the recognition of bacteria (Melkonian et al 2014, Westhoek et al, 2017) but could modify CFN of the bacteria of interest. Nevertheless, these techniques minimize the effect of multiple interactions. Construction of bacterial populations in which each individual can be quantified allows the study of complex interactions that cannot be revealed when single individuals are considered separately (Epstein et al., 2018; Paredes et al., 2018; Carlström et al., 2019; Boivin and Lepetit, 2020). The use of the DNA metabarcoding strategy allowed evaluating CFN after co-inoculation of roots with large populations of diverse unmodified rhizobia and therefore solved this problem. However bacterial competitiveness to form root nodules has generally been estimated by the number of nodules formed with the bacteria of interest as compared to total nodule number on the plant root system. This is consistent with the general hypothesis that CFN is a important symbiotic trait for early partner choice but not for ‘post-infection nodule expansion’ that mainly involves other preference mechanisms controlled by the
plant (Boivin & Lepetit, 2020). Evaluating CFN by DNA metabarcoding approach quantifies bacteria interacting with the plant and therefore may be biased as result of differential nodule growth or differential levels of bacteroid endoreduplication of the various bacteria (Kazmierczak et al., 2017). To circumvent this potential bias, we applied DNA metabarcoding on roots at an early stage of interaction before these processes being prominent.

**A representative core collection of Fabeae symbiont**

Sequencing 48 additional genomes of bacteria isolated from Fabeae root nodules of various geographical origins allowed us to characterize and extend our current knowledge of the Fabeae symbiovar diversity, augmenting the 73 *Rlv* genome sequences available in Genbank at the beginning of the study. It allowed new *Rlv* genospecies to be defined based on established criteria for bacteria in general and *Rhizobium* in particular (Jain, Rodriguez-R, Phillippy, Konstantinidis, & Aluru, 2018; de Lajudie et al., 2019). All the bacteria shared closely related *nod* gene sequences typical of the symbiovar *viciae*. However, several isolates were genetically distant from the main genospecies in the *R. leguminosarum* species complex (ANI<90%), raising the question of the appropriate boundary of this species complex, which has not yet been clearly defined. A set of 32 bacteria representative of the known genomic diversity of Fabeae symbiont was selected to study host-specific dynamics of symbiotic interaction. Individuals within the bacterial mixture have been discriminated and quantified by a DNA barcode located in the *nodD* gene of the symbiosis plasmid. This *nodD* metabarcoding enabled us to explore much of the intraspecific variability of the symbiovar. Nevertheless, for some sequence variants, the barcode has a limit. For example, *Rlv* strains carrying the Nod type B1 shared an identical barcode, so only one could be included in the competition study, but the strains displayed divergent
genome sequences that may potentially be associated with variation of the bacterial phenotype. The design and the high-throughput sequencing of a longer barcode will improve the resolution of the technique in the near future.

Co-inoculation of the core collection of 32 bacteria, in mixture, on the *Pisum sativum*, *Vicia faba* and *Lens culinaris* genotypes revealed host-specific CFN profiles. They varied according to the bacterium, the plant species and the plant genotype, and were probably related to the capacities of the various plant-microbe partnerships. These specificities were unlikely to be due to typical microbe-microbe interactions, generally resulting in differential bacterial multiplication and therefore limiting interaction with the plant, because bacteria were amplified separately and were present at high density in the inoculum (>10^7 CFU/mL). This study allowed us to identify the most competitive bacteria of the *Rlv* core collection with *Vicia faba*, *Pisum sativum* and *Lens culinaris* genotypes. Nevertheless, this study was done on plants cultivated in the same standard substrate and the question of the effect of soil environment on CFN has not been addressed here.

The *nodX* gene is necessary but not sufficient to confer host-specific nodulation of *Rlv* with *Pisum sativum* cultivar ‘Afghanistan’

Our data confirmed earlier work indicating that only *Rlv* bacteria carrying the *nodX* gene can nodulate the pea cultivar ‘Afghanistan’, resulting in a highly specific association profile (Davis et al., 1988; Firmin et al., 1993). This study shed light on unexpected complexities of the NodX/Sym2 interaction. The *nodX* gene was surprisingly found in a large portion of the *Rlv* genomes (22%), including those of both Nod types A and B. The corresponding bacteria were isolated from different Fabaeae host plants and have multiple geographical origins around the world, including some
locations where relatives of cv. ‘Afghanistan’ are probably not present. The nodX gene was not sufficient to allow Rlv to form root nodules with this cultivar, and we obtained evidence indicating that other genetic determinants, present only in the Nod type B strains, were required. We suggest that the host-specific nodulation of cv. ‘Afghanistan’ or relatives may be not the unique function of NodX. Another unknown function may explain the conservation of this gene within natural Rlv populations unable to associate with such hosts. Interestingly, this gene has been also found in R. leguminosarum symbiovar trifolii strains, also unable to nodulate P. sativum cv. ‘Afghanistan’ (Ovtsyna et al., 1999). The control of host-specific nodulation might be more complex, and the function of NodX deserves to be revisited. Nevertheless, NodX alone did not fully explain the specificity of the nodulation profile of cv. ‘Afghanistan’, co-inoculated with the core collection. The 5 bacteria of Nod type B that had the nodX gene, though they all had a good capacity to form root nodules with cv. ‘Afghanistan’ in mono-inoculation assays, displayed a wide range of contrasted CFN phenotypes with this host when co-inoculated in mixture. A striking example is the TOM strain, the well-studied reference symbiont of cv. ‘Afghanistan’, which is outcompeted by more competitive Nod type B bacteria carrying nodX gene. We identified 6 genomic regions associated with variation of CFN with this host in Nod type B strains with NodX.

**Host-specific CFN is associated with different genomic regions depending on the host plant and the genetic background of bacteria**

We confirmed that pea and fababean differentially select rhizobia depending on the variation of the Nod type/group (Boivin et al., 2020). Host-specific CFN was associated with the Nod types A and B in Pisum sativum (except cv. ‘Afghanistan’) and Vicia faba cultivars respectively. In contrast, the CFN with lentil was strongly associated with the genospecies, but poorly with the Nod type or Nod group,
suggesting that different genetic determinants, possibly located mainly on the chromosome, controlled CFN with this host. Despite these global trends, the genetic control of CFN is likely to be complex as, for all species, several data suggested effects of both Nod groups and genospecies, possibly in interaction. An unexpected result was the observation of a high CFN with pea of 3841, the unique bacterium of Nod group B1 present in the core collection. Our previous investigation on natural isolates using CFN measurements with a reference strain indicated that bacteria of this group B1 were generally competitive with fababean but poorly competitive with pea (Boivin et al., 2020). This different behavior of the strain 3841 deserves further investigation with particular emphasis to the association with the pea CFN phenotype. Because the strain 3841, initially isolated from pea root nodules, was cultured in the laboratories for many years, it cannot be ruled out that this particular phenotype may be related to spontaneous unknown mutations in the laboratory, in absence of natural selection.

The strategy of comparing genomes between competitive and uncompetitive strains allowed us to identify, for each plant genotype, regions of the bacterial genomes associated with host-specific CFN. The assignment of the identified genomic regions to a particular replicon deserves further study. As the knowledge of R. leguminosarum genomes is expanding, increasing evidence indicates that plasmid number, size and composition vary greatly in this bacterial species (Mazurier et al., 1997; Laguerre et al., 2003; Kumar et al., 2015; Cavassim et al., 2020). Different bacteria belonging to the same symbiovar share homologous sequences, organized differently in various diverse replicons. There is evidence that genetic rearrangements may occur even between plasmids and chromosome. There is an emerging picture of a high fluidity between R. leguminosarum replicons that argues against attempts to assign genomic regions to replicons by homology with a reference genome. Interestingly earlier works
have already suggested that large genetic rearrangements between replicons could modify the genetic environment around the *nod* gene cluster, and create and/or disrupt genetic links (Mazurier & Laguerre, 1997; Zhang, Kosier, & Priefer, 2001). This fluidity of the accessory genome is essential for the gene association approach that we have taken in this study, which requires a certain degree of independence between loci in order to associate phenotypes with individual genes rather than with whole replicons or large blocks of co-inherited genes.

As comparative studies yielded multiple potential genomic regions associated with CFN, it is tempting to hypothesize that multiple genes control this trait. A significant fraction of the protein sequences associated with CFN has known functions, whereas others are only hypothetical (Table 2 and Table S14). Known proteins were involved in amino acid modification (amidinotransferase, homoserine O-acetyltransferase), amino acid transport (L-proline/glycine betaine transporter), nucleic acid repair/modification (DNA/RNA helicase, excinuclease), bacteroid aerotolerance (bat operon), rhizobial nod factor secretion (NodT), or toxin/antitoxin systems (VapC, RelE/StbE). Interestingly, the *nodT* gene has already been associated with CFN in pea and fababean (Boivin et al., 2020). Members of Vap-type toxin/antitoxin system and genes involved in amino acid modifications (aminotransferase) were also associated with CFN in pea and fababean (Boivin et al., 2020) but, as they belong to multigenic families, we cannot confirm that genes identified in both studies are true orthologs. Further investigations are needed to specify this point. Globally, it would be too speculative at this stage to associate these specific functions with CFN. Although the identified genes were statistically associated with the host-specific CFN phenotypes, these associations may be indirect and related to neighboring genetic sequences. Reverse genetics studies with different combinations of alleles will be required to
validate their potential biological role in host-specific CFN in *Rlv*. Nevertheless, these
sequences are valuable markers to select competitive *Rlv* strains with pea and/or lentil
and they will allow the identification of genes controlling CFN. This knowledge opens
new perspectives to select bacteria and plants, by genetic association, for new
inoculation strategies that will ultimately improve the agro-ecological services of
Fabae legume holobionts.
Acknowledgements

We acknowledge Markus Braun (University of Heidelberg), Xinhua Sui and Changfu Tian (China Agricultural University, Beijing), Anna Skorupska and Andrzej Mazur (University of Lublin), and Juan Imperial (Instituto de Ciencias Agrarias, Madrid) for providing some of the rhizobial strains included in this study as well as the Genotoul GeT-PlaGe platform of Toulouse for the Illumina MiSeq sequencing. This work was supported by the GRaSP grant of the Agence Nationale pour la Recherche (ANR).
References


Melkonian, R., Moulin, L., Bena, G., Tisseyre, P., Chaintreuil, C., Heulin, K., ... Laguerre, G. (2014). The geographical patterns of symbiont diversity in the invasive legume Mimosa pudica can be explained by the competitiveness of its symbionts and by the host genotype. Environmental Microbiology, 16(7), 2099–2111. doi: 10.1111/1462-2920.12286


**Data Accessibility**

Rhizobial genome sequences: GenBank assembly accessions (see Table S1)

**Author Contributions**

ML and SB designed the research. ML, SB, PY, JW and SM provided the rhizobial strains. SB, MP, MatT, FM and MarT performed research. FM performed analysis of metabarcoding data. SB, ML, FM and PY analyzed data. SB and ML wrote the paper with the contribution of PY.
### Table 1: *Rhizobium leguminosarum* symbiovar *viciae* bacteria included in the core collection

<table>
<thead>
<tr>
<th>Strains</th>
<th>Isolated from</th>
<th>Locations</th>
<th>Genospecies</th>
<th>Nod groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>7841</td>
<td><em>Vicia faba</em></td>
<td>United Kingdom</td>
<td><em>gsB</em></td>
<td>B1</td>
</tr>
<tr>
<td>BLR195</td>
<td><em>Lens culinaris</em></td>
<td>Bangladesh</td>
<td><em>Rhizobium binae</em></td>
<td>B4†</td>
</tr>
<tr>
<td>CCBAU03058</td>
<td><em>Vicia faba</em></td>
<td>China</td>
<td><em>gsB</em></td>
<td>B5†</td>
</tr>
<tr>
<td>CCBAU10279</td>
<td><em>Vicia faba</em></td>
<td>China</td>
<td><em>gsF-2 / Rhizobium laguerreae</em></td>
<td>B5†</td>
</tr>
<tr>
<td>CCBAU11080</td>
<td><em>Vicia faba</em></td>
<td>China</td>
<td><em>gsG</em></td>
<td>B5†</td>
</tr>
<tr>
<td>CCBAU43229</td>
<td><em>Vicia faba</em></td>
<td>China</td>
<td><em>Rhizobium anhuiense</em></td>
<td>B3</td>
</tr>
<tr>
<td>CCBAU83258</td>
<td><em>Vicia faba</em></td>
<td>China</td>
<td><em>gsF-1</em></td>
<td>B5†</td>
</tr>
<tr>
<td>CZP1G9</td>
<td><em>Pisum sativum</em></td>
<td>Czech Republic</td>
<td><em>gsE</em></td>
<td>A4</td>
</tr>
<tr>
<td>CZP3G4</td>
<td><em>Pisum sativum</em></td>
<td>Czech Republic</td>
<td><em>Rhizobium pisi</em></td>
<td>A1†</td>
</tr>
<tr>
<td>CZP3H6</td>
<td><em>Pisum sativum</em></td>
<td>Czech Republic</td>
<td><em>gsC</em></td>
<td>A1</td>
</tr>
<tr>
<td>FRP3E11</td>
<td><em>Pisum sativum</em></td>
<td>France</td>
<td><em>gsE</em></td>
<td>B4</td>
</tr>
<tr>
<td>FRP3G5</td>
<td><em>Pisum sativum</em></td>
<td>France</td>
<td><em>gsE</em></td>
<td>B3</td>
</tr>
<tr>
<td>FRP4D11</td>
<td><em>Pisum sativum</em></td>
<td>France</td>
<td><em>gsE</em></td>
<td>A3†</td>
</tr>
<tr>
<td>GB29</td>
<td><em>Pisum sativum</em></td>
<td>Poland</td>
<td><em>gsC</em></td>
<td>A1</td>
</tr>
<tr>
<td>GB51</td>
<td><em>Pisum sativum</em></td>
<td>Poland</td>
<td><em>gsE</em></td>
<td>A1†</td>
</tr>
<tr>
<td>GD25</td>
<td><em>Pisum sativum</em></td>
<td>Poland</td>
<td><em>gsE</em></td>
<td>A4†</td>
</tr>
<tr>
<td>GLR17</td>
<td><em>Lens culinaris</em></td>
<td>Germany</td>
<td><em>gsF-2 / Rhizobium laguerreae</em></td>
<td>B4</td>
</tr>
<tr>
<td>GLR2</td>
<td><em>Lens culinaris</em></td>
<td>Germany</td>
<td><em>gsF-2 / Rhizobium laguerreae</em></td>
<td>B4</td>
</tr>
<tr>
<td>HP3</td>
<td><em>Pisum sativum</em></td>
<td>Algeria</td>
<td><em>gsF-2 / Rhizobium laguerreae</em></td>
<td>B2</td>
</tr>
<tr>
<td>IAU611</td>
<td><em>Pisum sativum</em></td>
<td>France</td>
<td><em>gsE</em></td>
<td>A4</td>
</tr>
<tr>
<td>L113</td>
<td><em>Lens culinaris</em></td>
<td>France</td>
<td><em>gsE</em></td>
<td>A1†</td>
</tr>
<tr>
<td>L125</td>
<td><em>Lens culinaris</em></td>
<td>France</td>
<td><em>gsE</em></td>
<td>A1†</td>
</tr>
<tr>
<td>L411</td>
<td><em>Lens culinaris</em></td>
<td>France</td>
<td><em>gsF-1</em></td>
<td>B4</td>
</tr>
<tr>
<td>Lp8</td>
<td><em>Lathyrus pratensis</em></td>
<td>United Kingdom</td>
<td><em>gsC</em></td>
<td>A4</td>
</tr>
<tr>
<td>OS25</td>
<td><em>Pisum sativum</em></td>
<td>Algeria</td>
<td><em>gsF-2 / Rhizobium laguerreae</em></td>
<td>B3</td>
</tr>
<tr>
<td>P1NP2H</td>
<td><em>Pisum sativum</em></td>
<td>France</td>
<td><em>gsB</em></td>
<td>A4</td>
</tr>
<tr>
<td>P1NP2K</td>
<td><em>Pisum sativum</em></td>
<td>France</td>
<td><em>gsB</em></td>
<td>A3</td>
</tr>
<tr>
<td>P221</td>
<td><em>Pisum sativum</em></td>
<td>France</td>
<td><em>gsE</em></td>
<td>A1†</td>
</tr>
<tr>
<td>SEF5C12</td>
<td><em>Vicia faba</em></td>
<td>Sweden</td>
<td><em>gsC</em></td>
<td>A1</td>
</tr>
<tr>
<td>SL16</td>
<td><em>Lens culinaris</em></td>
<td>Algeria</td>
<td><em>gsF-2 / Rhizobium laguerreae</em></td>
<td>B3</td>
</tr>
<tr>
<td>TOM</td>
<td><em>Pisum sativum</em></td>
<td>Turkey</td>
<td><em>gsF-1</em></td>
<td>B5†</td>
</tr>
<tr>
<td>UPM1134</td>
<td><em>Pisum sativum</em></td>
<td>Italy</td>
<td><em>gsC</em></td>
<td>A5</td>
</tr>
</tbody>
</table>

**Note:** † indicates bacteria carrying the nodX gene within the nod gene cluster.
Table 2: Putative functions and replicon location of genes associated with pea/lentil AFN or CFN identified by the comparative genomic analyses.

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Comparison</th>
<th>Putative function</th>
<th>3841</th>
<th>Vaf10</th>
<th>Vaf108</th>
<th>BIHB1217</th>
<th>UPM791</th>
<th>TOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK1 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>NA</td>
<td>Rh08</td>
<td>Rh08</td>
<td>Rh08</td>
<td>Rh08</td>
<td>NA</td>
</tr>
<tr>
<td>PK2 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>NA</td>
<td>NA</td>
<td>Rh02</td>
<td>Rh02</td>
<td>Rh02</td>
<td>NA</td>
</tr>
<tr>
<td>PI1 AFN</td>
<td></td>
<td>putative transmembrane</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>PI2 AFN</td>
<td></td>
<td>putative transmembrane</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>PI3 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>PI4 AFN</td>
<td></td>
<td>RelE/SbE replicon stabilization toxin</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>PI5 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PI6 AFN</td>
<td></td>
<td>amido transferase</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>PI7 AFN</td>
<td></td>
<td>amido transferase</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PI8 AFN</td>
<td></td>
<td>miscellaneous</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PI9 AFN</td>
<td></td>
<td>putative transcriptional regulator</td>
<td>Rh03</td>
<td>Rh08</td>
<td>Rh01</td>
<td>Rh08</td>
<td>Rh08</td>
<td>NA</td>
</tr>
<tr>
<td>PI10 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh03</td>
<td>Rh08</td>
<td>Rh01</td>
<td>Rh08</td>
<td>Rh08</td>
<td>NA</td>
</tr>
<tr>
<td>PI11 AFN</td>
<td></td>
<td>uracil/thymine DNA glycolase</td>
<td>Rh02</td>
<td>NA</td>
<td>NA</td>
<td>Rh02</td>
<td>Rh02</td>
<td>NA</td>
</tr>
<tr>
<td>PI12 AFN</td>
<td></td>
<td>L-2-hydroxyglutarate oxidase</td>
<td>Rh05</td>
<td>NA</td>
<td>NA</td>
<td>Rh04a</td>
<td>Rh04a</td>
<td>NA</td>
</tr>
<tr>
<td>PI13 AFN</td>
<td></td>
<td>L-proline/lactate/L-proline transporter ProP</td>
<td>Rh05</td>
<td>NA</td>
<td>NA</td>
<td>Rh01</td>
<td>Rh01</td>
<td>NA</td>
</tr>
<tr>
<td>PA1 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PA2 AFN</td>
<td></td>
<td>Delta-9 fatty acid desaturase</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PA3 AFN</td>
<td></td>
<td>Putative membrane-bound ClpP-class protease</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>PA4 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh02</td>
<td>NA</td>
<td>Rh02</td>
<td>Rh02</td>
<td>Rh02</td>
<td>NA</td>
</tr>
<tr>
<td>PA5 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PA6 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>NA</td>
<td>Rh08</td>
<td>Rh17</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LA1 AFN</td>
<td></td>
<td>TPR domain protein in aerotolerance operon</td>
<td>chr</td>
<td>Rh12</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LA2 AFN</td>
<td></td>
<td>BatA</td>
<td>chr</td>
<td>Rh12</td>
<td>Rh01</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LA3 AFN</td>
<td></td>
<td>Possible Neurospora U precursor</td>
<td>chr</td>
<td>Rh12</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LA4 AFN</td>
<td></td>
<td>Eno1-acyl-carrier-protein reductase</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LA5 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh01</td>
<td>NA</td>
<td>Rh01</td>
<td>NA</td>
<td>Rh01</td>
<td>NA</td>
</tr>
<tr>
<td>LA6 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>LR1 AFN</td>
<td></td>
<td>Cell filamentation protein flc</td>
<td>chr</td>
<td>Rh08</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
</tr>
<tr>
<td>LR2 AFN</td>
<td></td>
<td>VirC toxin protein antagonist</td>
<td>Rh01</td>
<td>NA</td>
<td>Rh01</td>
<td>NA</td>
<td>NA</td>
<td>Rh01</td>
</tr>
<tr>
<td>LF1 AFN</td>
<td></td>
<td>Polyhydroxyalkanoic acid synthase</td>
<td>Rh03</td>
<td>Rh17</td>
<td>Rh17</td>
<td>Rh08</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LF2 AFN</td>
<td></td>
<td>Acetyl-coenzyme A synthetase</td>
<td>Rh03</td>
<td>Rh01</td>
<td>NA</td>
<td>Rh08</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LF3 AFN</td>
<td></td>
<td>Homoserine O-acetyltransferase</td>
<td>Rh03</td>
<td>Rh01</td>
<td>NA</td>
<td>Rh08</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LF4 AFN</td>
<td></td>
<td>putative dehalogenase-hydrolase</td>
<td>Rh03</td>
<td>Rh01</td>
<td>NA</td>
<td>Rh08</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LF5 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh03</td>
<td>Rh01</td>
<td>Rh01</td>
<td>Rh08</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LF6 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh03</td>
<td>Rh01</td>
<td>Rh01</td>
<td>Rh08</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>LF7 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh03</td>
<td>Rh01</td>
<td>Rh01</td>
<td>Rh08</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN1 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>NA</td>
<td>Rh01</td>
<td>NA</td>
<td>NA</td>
<td>Na</td>
<td>Rh01</td>
</tr>
<tr>
<td>PAFN2 AFN</td>
<td></td>
<td>MFS permease</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>chr</td>
</tr>
<tr>
<td>PAFN3 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN4 AFN</td>
<td></td>
<td>Beta-galactosidase</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN5 AFN</td>
<td></td>
<td>Beta-galactosidase</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN6 AFN</td>
<td></td>
<td>CobN component of cobalt chelatase</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN7 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>NA</td>
<td>Rh12</td>
<td>NA</td>
<td>Rh06</td>
<td>NA</td>
<td>Rh06</td>
</tr>
<tr>
<td>PAFN8 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh12</td>
<td>Rh12</td>
<td>Rh01</td>
<td>Rh06</td>
<td>Rh12</td>
<td>Rh06</td>
</tr>
<tr>
<td>PAFN9 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh12</td>
<td>Rh12</td>
<td>Rh01</td>
<td>Rh06</td>
<td>Rh12</td>
<td>Rh06</td>
</tr>
<tr>
<td>PAFN10 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>Rh12</td>
<td>Rh12</td>
<td>Rh01</td>
<td>Rh06</td>
<td>Rh12</td>
<td>Rh06</td>
</tr>
<tr>
<td>PAFN11 AFN</td>
<td></td>
<td>Probably methylase/helicase</td>
<td>Rh12</td>
<td>Rh12</td>
<td>Rh01</td>
<td>Rh06</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>PAFN12 AFN</td>
<td></td>
<td>nodT RND efflux system</td>
<td>Rh03</td>
<td>Rh01</td>
<td>Rh08</td>
<td>Rh08</td>
<td>NA</td>
<td>Na</td>
</tr>
<tr>
<td>PAFN13 AFN</td>
<td></td>
<td>Alpha-aspartyl dipetidase Peptidase E</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN14 AFN</td>
<td></td>
<td>putative transmembrane protein</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
<td>chr</td>
</tr>
<tr>
<td>PAFN15 AFN</td>
<td></td>
<td>PE-PQRS FAMILY PROTEIN</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN16 AFN</td>
<td></td>
<td>hypothetical protein</td>
<td>chr</td>
<td>NA</td>
<td>chr</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PAFN17 AFN</td>
<td></td>
<td>BII066 protein</td>
<td>Rh12</td>
<td>Rh12</td>
<td>Rh01</td>
<td>Rh06</td>
<td>Rh12</td>
<td>Rh03</td>
</tr>
</tbody>
</table>

Note: Replicons are classified as chromosome ‘chr’ or ‘rh’ repABC type plasmid (i.e. based on sequence homology to repABC genes as described by Cavassim et al., 2020). NA indicates no significant homologous sequence. Gene identifiers refered to the host: pea cv. ‘Kayanne’ (PK), pea cv. ‘Isard’ (PI), pea cv. ‘Afghanistan’ (PA), lentil cv. ‘Anicia’ (LA), lentil cv. ‘Rosana’ (LR) and lentil cv. ‘Flora’ (LF). More details are provided in Table S14.
Figure Legends:

Figure 1: Genomic diversity of the 121 fabae symbiont genomes selected for this study. Hierarchical clustering and heat map were based on the Average Nucleotide Identity (ANI) values between each couple of the 124 bacterial genomes. Rlv genospecies classification (gs) has been based on an ANI threshold of 95%. Star gathered Sm1021 and isolates phylogenetically distantly related to Rlv (ANI<90%). Reference bacteria Sinorhizobium mellioti Sm1021, Rhizobium leguminosarum symbiovar trifoli WSM1689, Rhizobium leguminosarum symbiovar phaseoli Rlp4292 were also included in the comparison. Additional informations are provided in the Table S1.

Figure 2: Nod gene cluster diversity of the 121 fabae symbiont genomes selected for this study. Phylogenetic tree was based on the nodABCDEFIJLMN concatenated gene sequences of the 123 genomes. Boxes defined Nod groups. The 48 new fabae symbiont genomes are indicated in italic. The 32 bacteria of the core collection are indicated in red. Stars indicate bacteria carrying the nodX gene within the nod gene cluster. Rhizobium leguminosarum symbiovar trifoli WSM1689 and Sinorhizobium mellioti 1021 have been used as outgroups.

Figure 1: Genomic diversity of the 121 fabae symbiont genomes selected for this study. Hierarchical clustering and heat map were based on the Average Nucleotide Identity (ANI) values between each couple of the 124 bacterial genomes. Rlv genospecies classification (gs) has been based on an ANI threshold of 95%. Star gathered Sm1021 and isolates phylogenetically distantly related to Rlv (ANI<90%). Reference bacteria *Sinorhizobium melloti* Sm1021, *Rhizobium leguminosarum symbiovar trifolii* WSM1689, *Rhizobium leguminosarum symbiovar phaseoli* Rlp4292 were also included in the comparison. Additional informations are provided in the Table S1.
Figure 2: Nod gene cluster diversity of the 121 fabae symbiont genomes selected for this study. Phylogenetic tree was based on the nodABCDEFJLMN concatenated gene sequences of the 123 genomes. Boxes defined Nod groups. The 48 new fabae symbiont genomes are indicated in italic. The 32 bacteria of the core collection are indicated in red. Stars indicate bacteria carrying the nodX gene within the nod gene cluster. *Rhizobium leguminosarum* symbiovar *trifolii* WSM1689 and *Sinorhizobium meliloti* 1021 have been used as outgroups.
Figure 3: Competitiveness to Form root Nodule of rhizobia with Fabaeae plant species/genotypes. Hierarchical clustering and heat map based on the CFNI for each condition, including the 20 most competitive rhizobia within the core collection. Stars indicate bacteria carrying the nodX gene within the nod gene cluster. FO: Vicia faba cultivar 'Organdi'; FD: Vicia faba cultivar 'Diva'; FT: Vicia faba cultivar 'Tiffany'; PK: Pisum sativum cultivar 'Kayanne'; P1: Pisum sativum cultivar 'Isard'; PJ: Pisum sativum cultivar 'Afghanistan J11357'; LF: Lens culinaris cultivar 'Flora'; LR: Lens culinaris cultivar 'Rosana'; LA: Lens culinaris cultivar 'Anicia'.

810
Supplemental Tables and Figures

**Figure S1**: NodD309 barcode diversity among the 121 *Rhizobium leguminosarum* symbiovar *viciae* bacteria of this study.

**Table S1**: Genomic data of the sequenced *Rhizobium leguminosarum* symbiovar *viciae* bacteria.

**Table S2**: NodD309 sequences for the 32 *Rlv* bacteria included in the core collection, primers used and PCR conditions.

**Table S3**: Number of reads in each sample after quality filtering.

**Table S4**: Presence/absence of genomic regions associated with CFN phenotypes in pea, faba bean and lentil.

**Table S5**: Global rearrangements of plasmids/contigs between reference strains, in comparison with 3841.

**Table S6**: Genomic organization of contigs, containing genes associated to CFN, compared to 3841.

**Table S7**: Genospecies and Nod group combinations among the bacterial collection.

**Table S8**: Competitiveness to Form Nodule indexes (CFNi) for each of the plant/bacteria couples.

**Table S9**: Clusters of Rlv bacteria from the core collection, depending on their CFN profiles in the Figure 3.

**Table S10**: Effect of variation of *Rlv* genospecies, Nod types and Nod groups in the core collection on the CFN in the pea and lentil cultivars.

**Table S11**: Mono-inoculation of *Rlv* strains included in the core collection with the *Pisum sativum* cultivar ‘Afghanistan’.

**Table S12**: Effect of the plant species on the CFN of the *Rlv* bacteria from the core collection, as function of the Nod type, the Nod group, and the genospecies.

**Table S13**: Presence/absence of genomic regions associated with AFN in *Pisum sativum* cultivar ‘Afghanistan’

**Table S14**: Locations of genes significantly associated with AFN and CFN from the comparative genomic analyses.