BMC
Genomics

Genomic basis of symbiovar mimosae in *Rhizobium etli*

Marco A Rogel[1], Patricia Bustos[2], Rosa I Santamaría[2], Víctor González[2], David Romero[3], Miguel Ángel Cevallos[2], Luis Lozano[2], Jaime Castro-Mondragón[1], Julio Martínez-Romero[1], Ernesto Ormeño-Orrillo[1] and Esperanza Martínez-Romero[1*]

Abstract

Background: Symbiosis genes (*nod* and *nif*) involved in nodulation and nitrogen fixation in legumes are plasmid-borne in *Rhizobium*. Rhizobial symbiotic variants (symbiovars) with distinct host specificity would depend on the type of symbiosis plasmid. In *Rhizobium etli* or in *Rhizobium phaseoli,* symbiovar phaseoli strains have the capacity to form nodules in *Phaseolus vulgaris* while symbiovar mimosae confers a broad host range including different mimosa trees.

Results: We report on the genome of *R. etli* symbiovar mimosae strain Mim1 and its comparison to that from *R. etli* symbiovar phaseoli strain CFN42. Differences were found in plasmids especially in the symbiosis plasmid, not only in *nod* gene sequences but in *nod* gene content. Differences in Nod factors deduced from the presence of *nod* genes, in secretion systems or ACC-deaminase could help explain the distinct host specificity. Genes involved in *P. vulgaris* exudate uptake were not found in symbiovar mimosae but *hup* genes (involved in hydrogen uptake) were found. Plasmid pRetCFN42a was partially contained in Mim1 and a plasmid (pRetMim1c) was found only in Mim1. Chromids were well conserved.

Conclusions: The genomic differences between the two symbiovars, mimosae and phaseoli may explain different host specificity. With the genomic analysis presented, the term symbiovar is validated. Furthermore, our data support that the generalist symbiovar mimosae may be older than the specialist symbiovar phaseoli.

Keywords: Legume nodulation, Bacterial symbiosis, Nitrogen fixation, Host specificity

Background

Bacterial nitrogen fixation in legume nodules contributes to plant nutrition and allows plants to grow in nitrogen deficient soils. Genes for plant nodulation and nitrogen fixation are plasmid-borne in *Rhizobium* spp. (reviewed in [1]) and symbiovars define host specificity. There are over twenty different symbiovars reported not only in *Rhizobium* but also in *Bradyrhizobium* and in other genera of nodule forming bacteria [2-7]. The term symbiovar was proposed as a counterpart to the term pathovar in pathogenic bacteria [2]. A theoretical model proposes that a single species may exhibit alternative symbiovars depending on the presence of symbiotic plasmids or symbiotic islands [2]. The same symbiovar may be present in

distinct species as a consequence of the lateral transfer of symbiotic plasmids or islands. Symbiotic genes and other genes associated with niche adaptation may have evolutionary histories independent of the evolution of the chromosomal genes [8]. Two symbiovars are recognized in *Rhizobium etli*: phaseoli (conferring the ability to nodulate *Phaseolus vulgaris*) and mimosae (involved in nodulating mimosas and *P. vulgaris*, [9]). Symbiovar mimosae strains were isolated from *Mimosa affinis* in Morelos and have a broad host range, including plants of *M. affinis, Leucaena leucocephala, Calliandra grandiflora, Acaciella angustissima* as well as *P. vulgaris* [9]. Symbiovar mimosae was originally distinguished from sv. phaseoli by the sequence of a few symbiotic genes and by the organization of *nif* and common *nod* genes. Multiple copies of *nifH* genes and a *nodA* gene separated from *nodBC* found in sv. phaseoli and not in sv. mimosae served as a molecular basis to distinguish these

* Correspondence: esperanzaeriksson@yahoo.com.mx
[1]Ecological Genomics programs, Genomics Science Center, CCG, Cuernavaca, Morelos, Mexico
Full list of author information is available at the end of the article

symbiovars [10,11]. Phylogenies indicated that symbiovar mimosae and phaseoli *nifH* genes were related [9] and similar to the corresponding gene in sv. gallicum [12]. Different origins of replication were found in sv. phaseoli and sv. mimosae symbiotic plasmids and both symbiotic plasmids were compatible [9]. It has been proposed that sv. mimosae is older than sv. phaseoli and that the phaseoli symbiotic plasmid was selected by *P. vulgaris* [13]. *P. vulgaris* is a recent species (probably two million years old [14]), while mimosas seem to be older.

Mimosas are distributed worldwide with Brazil and Mexico as main diversification sites. Mimosas in South America are nodulated by β-Proteobacteria like *Burkholderia* or other β-Proteobacteria [15-19] while mimosas in Mexico are only exceptionally nodulated by *Burkholderia* (unpublished). Mimosas from Mexico and Brazil are phylogentically separated [20]. Additionally abiotic conditions like pH and soil nitrogen content may account for their differences in symbionts [21]. Native mimosas in India are nodulated by sinorhizobia [22], that we have also found in some mimosa nodules in Mexico (unpublished).

Based on multilocus enzyme electrophoresis analysis, *M. affinis* isolate Mim1 was found to group closely to *R. etli* sv. phaseoli CFN 42 [9]. Other *M. affinis* isolates such as Mim2 were separated from CFN42 (Figure two in [9]) and thus Mim2 has been recently reassigned to *Rhizobium phaseoli* [1,23]. Therefore we recognize now that symbiovar mimosae exists in *R. phaseoli* as well as in *R. etli*. It is the aim of this work to define the genomic differences between two *R. etli* strains (CFN42 and Mim1) representing the symbiovars phaseoli and mimosa, respectively.

Methods
Strains and growth conditions
Rhizobium strains were grown overnight at 28ºC in PY medium [24] after recovery from glycerol stocks at -80ºC. Bacterial strains to be inoculated on plants were grown on solid PY media and resuspended in water to an OD_{600} of 0.5. For PCR or DNA isolation, bacteria were grown in liquid PY cultures [24].

Plasmid profiles
Plasmid profiles were visualized on agarose gels according to the protocol described by Hynes and McGregor [25]. Plasmid patterns from *R. etli* CFN42 or *R. leucaenae* CFN299 were used as references.

Plant nodulation assays
L. leucocephala seeds were treated with concentrated sulfuric acid for 15 min, rinsed with water and surface disinfected with sodium hypochlorite as described [26]; the same procedure was used to disinfect *P. vulgaris*

seeds. Seeds were germinated in water-agar plates in the dark and transferred to flasks after 3 days. *L. leucocephala* plants were grown in vermiculite flasks with N free Fahraeus nutrient solution for 40 days and *P. vulgaris* plants in agar flasks with the same nutrient solution for 14 days.

Genome sequencing, assembly and annotation
Genomic DNA from *R. etli* Mim1 was sheared to produce two paired-end libraries for 454 pyrosequencing, one with 3 Kb inserts and the other with 8 Kb inserts. An additional 3 Kb library was sequenced only at one end. The total amount of reads were 512,236 paired reads and 112, 079 single reads. Library construction and sequencing was done at Mogene LC (St. Louis, MO, USA). Additionally two paired-end libraries were constructed, one with 200 bp inserts and the second with 2 Kb inserts. Both libraries were sequenced by Illumina at BGI (Beijing, China). A total amount of 16, 000, 000 short-paired readings (50–60 bases) were assembled. To improve the scaffolding a BAC library was constructed in pIndogo BAC-5™ vector by BIOS&T (Quebec, Canada) using fragments from a partial genomic DNA *Hind*III digestion. 105 BAC-ends were sequenced with ABI3730xl sequencer by Sanger method. Additionally, three BACs were completely sequenced with the same technology at the Center for Genomic Sciences (Cuernavaca, México). Two of these BACs were selected by hybridizing with *nifH* and other pSym probes; they embraced half of the pSym plasmid sequence. The third BAC was from the chromosome. Final assembly of the symbiotic plasmid was obtained using sequence reads from the three sources of information: BAC sequences, Illumina, and 454 pyrosequencing.

Different assembly strategies were used with the following programs: Newbler 2.5.3 (ROCHE), Velvet 1.1.06 [27], Sspace-Basic 2.0 [28], and Consed v23 [29]. ORFs were predicted with Glimmer 3 [30], and annotations were done in Artemis 12.0 graphic display [31] using previous annotations made for *R. etli* CFN42 [32] and comparing with the non-redundant data base of the Genbank [33], Interpro database [34], and IS database (http://www-is.biotoul.fr).

Sequence analysis
Average nucleotide identity (ANI) was calculated using the JSpecies software [35]. The DNA conservation between two genomes or replicons was estimated by obtaining an alignment with NUCmer [36], run with default parameters, and dividing the summed lengths of all aligned regions by the length of the genome or replicon and expressing the value obtained as a percentage. Common and specific protein families between *R. etli* CFN42 and Mim1 were detected using MCL as described [37].

Genomic islands

Alien Hunter v1.7 software [38] was used to analyze the chromosome sequence of *R. etli* sv. mimosae Mim1 and *R. etli* CFN42. The minimum region length for HT detection was 5 kbs. The score thresholds were 12.92 and 14.96 for Mim1 and CFN42, respectively.

Phylogenetic analysis

Alignments were performed with MUSCLE [39] and manually verified. Maximum likelihood trees were generated with PhyML [40] with tree node support evaluated by bootstrap analysis based on 1000 pseudoreplicate datasets. Phylogenetic relationships were also assessed by Bayesian inference using MrBayes 3.1 [41]. Analyses were initiated with random starting trees, run for 2,000,000 generations and three separate analyses were executed. Markov chains were sampled every 100 generations. We discarded 25% of trees as "burn in".

Genome accession numbers

Sequences and annotations were deposited in the Genbank database under accession numbers CP005950 (chromosome), CP005951 (pRetMim1a), CP005952 (pRetMim1b), CP005953 (pRetMim1c), CP005954 (pRetMim1d), CP005955 (pRetMim1e) and CP005956 (pRetMim1f).

Results and discussion

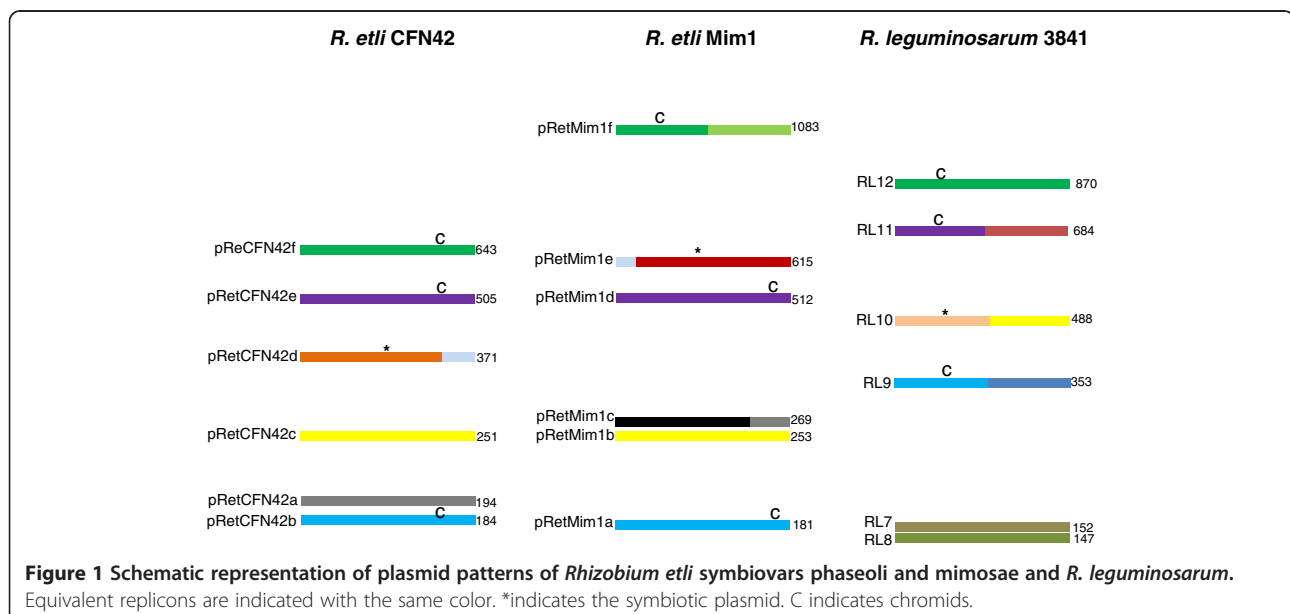### Genome of *R. etli* symbiovar mimosae strain Mim1

The final assembly of the *R. etli* sv. mimosae Mim1 genome rendered seven circular molecules: one chromosome and six plasmids at 150× coverage on average. The chromosome was 4.8 Mb in size while the plasmids ranged in size from 181 kb to 1.08 Mb (Figure 1). Average Nucleotide

Identity (ANI) and the percentage of conserved DNA between Mim1 and CFN42 were 98.6% and 82.4% respectively on a whole genome analysis, confirming that both strains belong to the same species. Lower ANI (less than 90%) was found between Mim1 and strains of other species such as *R. phaseoli* and *R. leguminosarum*.

There were more than twice as many unique genes in Mim1 than in CFN42, mainly in plasmids. The respective chromosomes of each strain had around 260 unique genes. In chromosomes, 35 genomic islands were identified only in Mim1 and 17 only in CFN42. (Figure 2); genes found in Mim1 genomic islands are shown in Additional file 1: Table S1. Examples of unique genes found in Mim1 and not in CFN42 are those encoding cytochrome oxidases, some chaperonins, dipeptide transporters, lactate dehydrogenase, a PHB depolymerase, a ferritin–like protein, exopolysaccharide biosynthesis genes, a type VI secretion system and menaquinone biosynthesis as well as many hypothetical genes.

The conserved genome in *R. etli* strains Mim1 and CFN42 includes not only the chromosome but two extrachromosomal replicons (pRetMim1a-pRetCFN42b and pRetMim1d-pRetCFN42e) that have been designated as chromids in CFN42 [42] and one plasmid (pRetMim1b-pRetCFN42c) (Figure 3). Each of the chromids had less than 20 unique genes and the chromid pairs had an ANI around 99% (Table 1). The small replicons pReCFN42a and pRetMim1c were partially conserved and large genomic differences were found in the symbiotic plasmids (Figures 1 and 4, Table 1).

Mim1 and CFN42 chromosomes were syntenic, as were the chromid pairs pRetMim1a-pRetCFN42b and pRetMim1d-pRetCFN42e, and plasmids pRetMim1b-pRetCFN42c. In comparison to *R. leguminosarum* 3841, Mim1 plasmid
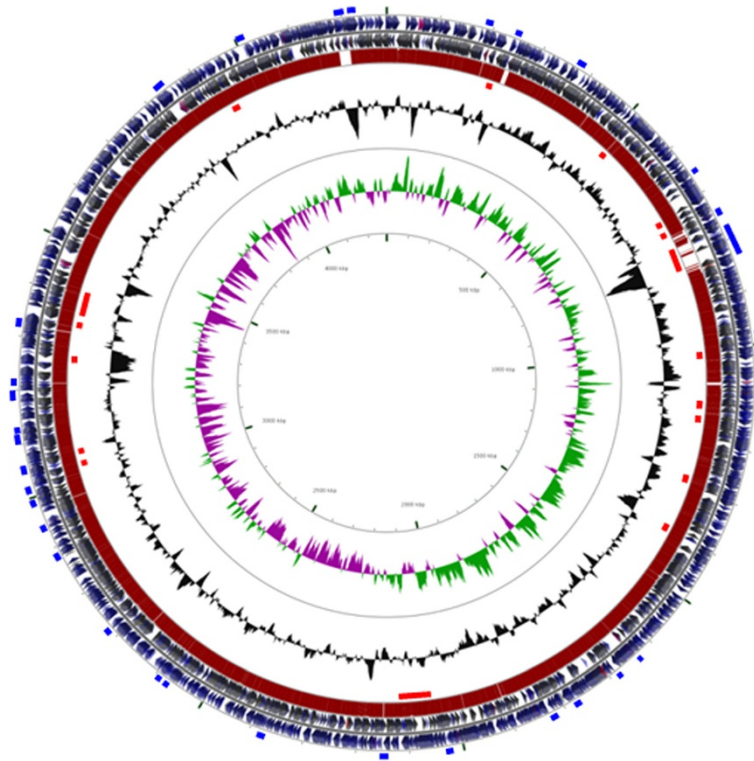

Figure 1 Schematic representation of plasmid patterns of *Rhizobium etli* symbiovars phaseoli and mimosae and *R. leguminosarum*.
Equivalent replicons are indicated with the same color. *indicates the symbiotic plasmid. C indicates chromids.

Figure 2 Representation of *R. etli* Mim1 chromosome. Circles from outermost to innermost indicate: genomic islands of Mim1 (in blue), ORFs in the leading strand, ORFs in the lagging strand, BLAST matches against CFN42 chromosome, genomic islands of CFN42 (in red), GC content, GC skew, coordinates.
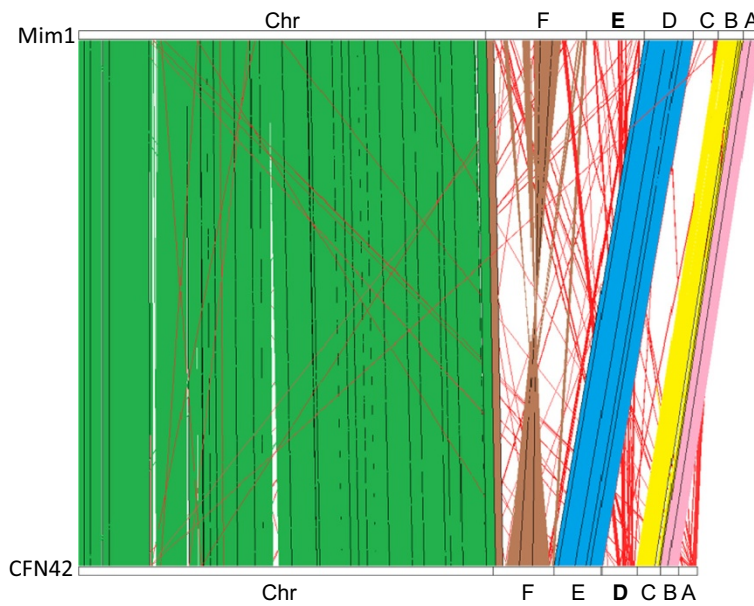

Figure 3 Graphic comparison of equivalences among the *R. etli* CFN42 and *R. etli* Mim1 genomes. Letters indicate the different extrachromosomal replicons found in both strains (see text). Letters in bold indicate the symbiotic plasmids.

Table 1 Average nucleotide identity (ANI) and conservation in percent between *R. etli* sv. mimosae Mim1 and *R. etli* sv. phaseoli CFN 42 or *R. leguminosarum* 3841 replicons

Mim1 replicons	ANI*/conservation§ to the corresponding replicons in	
	R. etli **CFN42**	*R. leguminosarum* **3841**
pRetMim1f	97.4/51.1 (pReCFN42f)	86.8/31.4 (pRL12)
pRetMim1e (pSym)	89.2/8.8 (ReCFN42d)	
pRetMim1d	98.9/97.5 (pReCFN42e)	88.2/61.2 (pRL11)
pRetMim1c	86.4/11.4 (pReCFN42a)	
pRetMim1b	99.1/92.5 (pReCFN42c)	87.2/55.7 (pRL10)
pRetMim1a	99.2/99.9 (pReCFN42b)	87.6/58.4 (pRL9)
RetMim1Ch	99.2/97.8 (ReCFN42Ch)	88.1/78.1 (RLChr)

*Average nucleotide identity was calculated using all portions of the replicons that could be aligned with the nucmer program. These regions included both genic and intergenic regions.
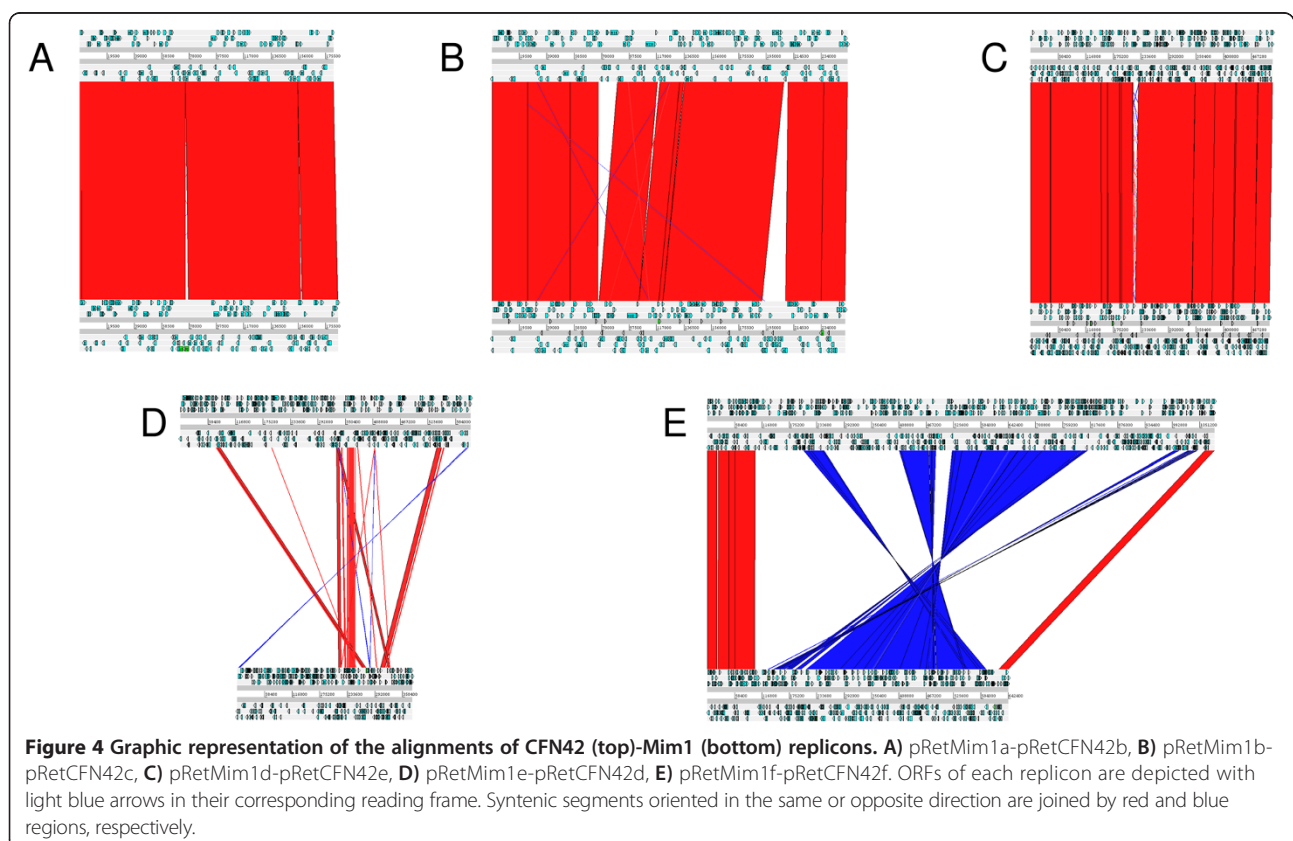§Percentage of the Mim1 replicon involved in the ANI calculation is expressed here as the conservation value.

equivalences were similar to those reported previously for *R. etli* CFN42 [8]. ANI values were lower with *R. leguminosarum* 3841 that among the *R. etli* strains (Table 1).

All extrachromosomal replicons in Mim1 belong to the *repABC* plasmid family [43]. The protein products of the *repABC* operons of the replicon homologous pairs (pRetMim1a-pRetCFN42b, pRetMim1d-pRetCFN42e and pRetMim1b-pRetCFN42c) were almost identical with a sequence identity greater than 97.51%, strongly suggesting that members of each replicon pair belong to the same incompatibility group.

pRetCFN42f-pRetMim1f comparison

The largest extrachromosomal replicon in Mim1 (pRetMim1f) was only partially conserved in the putative chromid pRetCFN42f (Figures 1 and 4). pRetMim1f and pRetCFN42f possess two *repABC* operons: *repABC1* and *repABC2*. The sequence identity between the two *repABC* operons in Mim1 is low. The degree of sequence identities between the corresponding *repABC* genes of pRetCFN42f and pRetMim1f is large enough to suggest that both plasmids share the same incompatibility group and evolutionary origin. However only 51% of the pRetMim1f replicon

Figure 4 Graphic representation of the alignments of CFN42 (top)-Mim1 (bottom) replicons. A) pRetMim1a-pRetCFN42b, **B)** pRetMim1b-pRetCFN42c, **C)** pRetMim1d-pRetCFN42e, **D)** pRetMim1e-pRetCFN42d, **E)** pRetMim1f-pRetCFN42f. ORFs of each replicon are depicted with light blue arrows in their corresponding reading frame. Syntenic segments oriented in the same or opposite direction are joined by red and blue regions, respectively.

is conserved in CFN42 while 86% of pRetCFN42f was found in pRetMim1f (Figures 1 and 4).

Type VI secretion system genes were only found in the megaplasmid pRetMim1f and the conserved *impB* component is phylogenetically related to the corresponding genes in *Rhizobium* sp. KIM5 (corresponding to PEL1 lineage, [44]) and *R. leguminosarum* strain 3841. Mim1 pRetMim1f has a duplicated citrate synthase gene as in sv. tropici symbiotic plasmids. The plasmid duplicated citrate synthase gene in sv. tropici is required for eliciting a normal number of nodules and is regulated differently than the copy in the chromosome [45,46]. A plasmid borne citrate synthase was not found in CFN42. Phylogenetic analysis of citrate synthase genes showed that the Mim1 chromosomal gene product is identical to that of CFN42, while the gene in pRetMim1f has a novel sequence only distantly related to a plasmid copy of *R. gallicum* R602 (55% identity).

Genes such as *raiI* and *raiR* were found in the conserved region of pRetCFN42f and pRetMim1f. RaiI produces homoserine lactones and RaiR is the transcriptional regulator. The *rai* system in *R. phaseoli* sv. phaseoli CNPAF 512 affects nodule number but not nitrogen fixation in *P. vulgaris* [47]. This system also controls growth in *R. phaseoli* [48] and if this is the case in *R. etli*, this would explain its conservation in both symbiovars.

A transcriptomic study compared the genome expression of *R. phaseoli* Ch24-10 in maize and *P. vulgaris* rhizospheres [49]. Over 50% of the extrachromosomal genes highly expressed in *P. vulgaris* but not in maize roots were found in a Ch24-10 replicon equivalent to pRetCFN42f. It seems that genes in this replicon are involved in plant specific interactions.

Symbiosis plasmid gene comparison

Large differences were observed between the symbiotic plasmids of CFN42 and Mim1 (Figure 4). Around 10% and 15% of the symbiotic plasmids of Mim1 and CFN42 had conserved syntenic regions with an average nucleotide identity of 89.2% (Table 1).

Differences in symbiosis genes in CFN42 and Mim1 genomes are shown in Table 2. The Nod factor from *R. etli* sv. phaseoli strain CFN42 is a pentamer of *N*-acetylglucosamine with an acetyl fucose at the reducing end and methyl and carbamoyl groups at the non reducing end [50]. The heterologous expression in *Azorhizobium caulinodans* of Nod modification genes showed that fucosylated Nod factors were the most suitable to induce *P. vulgaris* nodulation [51]. In symbiovar mimosae no genes related to Nod factor fucosylation (*nodZ*) were observed (Table 2), in their place, *nodHPQ* genes that modify the Nod factor with sulfate, were found. Such genes are present in sv. tropici strains that are also *Leucaena* symbionts [10]. Like *R. etli* strain CFN42, Mim1 may produce nodulation factors bearing carbamoyl groups at their non reducing end residues but the position of these

Table 2 Relevant symbiotic plasmid differences between *R. etli* sv. mimosae Mim1 and *R. etli* sv. phaseoli CFN 42

Gene*	symbiovar	
	mimosae	phaseoli
nodHPQ	+	-
nodZ	-	+
nolL	-	+
nodO	+	-
nolO	-	+
nodU	+	-
fixKL	+	-
fxkR	+	-
acdS	+	-
teu	-	+
hup-hyp	+	-

*Functions of each gene are explained in the text.

decorations must differ because their pSyms encoded distinct carbamoyl transferases, NolO in CFN42 and NodU in Mim1. Carbamoylation at the C-6 position introduced by NodU maybe promotes *Leucaena* nodulation [52,53]. Both Mim1 and CFN42 symbiotic plasmids carry *nodS* involved in methylation at the non reducing end, a decoration that is essential for bean and *Leucaena* nodulation in *R. tropici* CIAT 899 [52]. A *nodO* homologue, 70% identical to that of *Rhizobium* sp. BR816, was found only in the sv. mimosae pSym. It has been shown that heterologous expression of *nodO* can improve nodulation of *L. leucocephala* by different rhizobia and can even extend the host range [54,55].

Mim1 *nod* gene phylogenies are congruent, resembling the corresponding genes from sv. giardinii (not shown) while Mim1 *nifH* genes resemble those from sv. phaseoli. Different NodDs in phaseoli and mimosae symbiovars may reflect their affinities for the different flavonoids exuded by the different host plants. Mim1 *nodH* gene (encoding the sulfotransferase involved in the synthesis of the Nod factor) resembles the corresponding gene in *Rhizobium* sp. IE4771 that represents a novel genomic lineage related to *R. etli* and *R. phaseoli* [1,44].

acdS gene coding ACC-deaminase was found in the symbiotic plasmid of Mim1 but not in CFN42. ACC-deaminase decreases the amount of ACC that is a precursor of ethylene that may diminish nodule number. A heterologous ACC-deaminase in *Rhizobium* sp TAL 1145 enhanced nodulation in *Leucaena* [56].

In CFN42, the *fixGHIS-fixNOQP* genes required for biosynthesis of the symbiotic terminal oxidase are present in the pSym and also in pRetCFN42f [57]. The regulatory genes *fixK* and *fixL* are adjacent to this reiteration in pRetCFN42f while a *fixK* pseudogene is found in the pSym [58]. In Mim1, we found that the symbiotic terminal oxidase genes are also reiterated in pRetMim1f but, in contrast to

CFN42, the sv mimosae pSym carried complete *fixK* and *fixL* genes. The *fixGHIS-fixNOQP-fixKL* region shared by pSym and pRetMim1f plasmids in Mim1 is 95% identical while the reiterated regions in CFN42 are only 87% identical. The recently described *fxkR* gene in pRetCFN42f [59], coding for a response regulator that acts in conjunction with FixL and FixK, is present in the sv. mimosae pSym. There is a reiteration of this gene in pRetMim1f as well.

Genes involved in *P. vulgaris* exudate uptake (*teu* genes, [60]) are found in the symbiotic plasmids in symbiovar phaseoli strains *R. etli* CFN42 and *R. phaseoli* CIAT652 but they were not found in the Mim1 genome. *Rhizobium* mutants in *teu* genes had reduced nodulation competitiveness in *P. vulgaris* [60].

The symbiosis plasmid of Mim1 has genes for a type III secretion system (T3SS) that are more closely related to those found in *Mesorhizobium* and *Sinorhizobium* strains than to those encoded in the CFN42 pSym (not shown). This difference may contribute to the disparate host ranges displayed by sv. phaseoli and mimosae strains considering the function of rhizobial T3SS in specificity [61].

A cluster of *hup-hyp* genes encoding the components of an uptake hydrogenase (Hup) was found in the pSym of Mim1 but not in CFN42. The Mim1 products showed high identities (>70%) to their counterparts coded in *R. leguminosarum* and *R. tropici* [10,62]. *R. tropici* CIAT 899 lacks several *hup* genes and displays a Hup minus phenotype. In Mim1, all genes except *hupE* are present. HupE is an uptake transporter for nickel [63], a metal required for Hup function. Since another nickel transporter is encoded elsewhere in Mim1 symbiotic plasmid, the Hup system could be functional.

Mim1 symbiotic plasmid has a *repABC* origin of replication as well as a *repC* gene that are not phylogenetically related to the *R. etli* CFN42 corresponding genes. pRetMim1e *repC* resembles those from *R. gallicum* and *Rhizobium* sp. sv. giardinii IE4771 (corresponding to PEL1 lineage). Mim1 *repABC* from the pSym resembles the corresponding genes in *R. endophyticum* CCGE502 but an extensive plasmid conservation was not observed. CCGE502 is Nod⁻ and does not have a symbiotic plasmid [64].

The analysis of insertion sequences in *R. etli* CFN42 suggested that the symbiotic plasmid did not significantly share IS sequences with the rest of the genome [65]. This was interpreted as evidence that the pSym was a recent acquisition in this bacterium. The analysis of IS sequences in Mim1 indicated that the symbiotic plasmid had the largest number of IS sequences, some of them shared with the chromosome, pRetMim1f and pRetMim1b, this may perhaps indicate that this symbiotic plasmid has an older history with the *R. etli* genomic background than the phaseoli plasmid. Mim1 has a large number of IS66 that are common in rhizobia.

Genome similarities were found among the different Mim1 replicons. pRetMim1e (pSym) has similar sequences to pRetMim1c and the same is observed among pRetMim1e
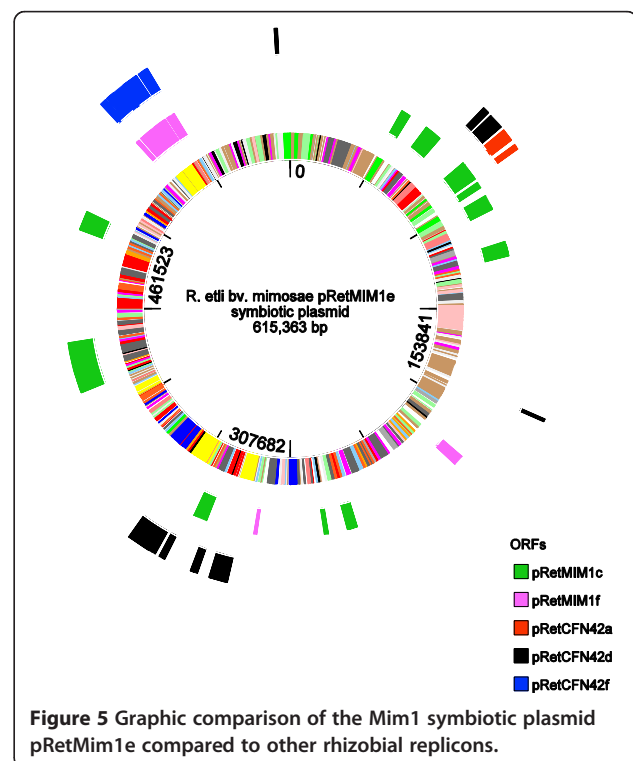
(pSym) and pRetMim1f (Figure 5). The similarities of Mim1 symbiotic plasmid and the putative chromid pRetMim1f could support that symbiovar mimosae symbiotic plasmid is ancestral in *R. etli.* In contrast the phaseoli symbiotic plasmid was found dissimilar to the rest of the genomic background in *R. etli*, except to pReCFN42a [32,66].

In *R. etli* CFN42, the symbiotic plasmid and pRetCFN42a have common sequences such as *tra* and *vir* genes and repeated sequences that mediate the natural cointegration of both plasmids for the conjugative transfer of CFN42 pSym [67]. In contrast to pReCFN42a, pRetMim1c from Mim1 does not seem to participate in the transfer of the Mim1 symbiotic plasmid, which we have been unable to transfer to other bacterial hosts (Marco A. Rogel, unpublished observations).

Plant-interaction genes not in the symbiosis plasmid

Some genes involved in plant interactions were found conserved in both symbiovars, such as those encoding Rmr extrusion pumps that may be involved in eliminating plant produced phytoalexins. *R. etli* mutants in these genes had reduced nodulation [68]. Those genes are encoded in chromid pRetCFN42b in CFN42 and in the corresponding replicon pRetMim1a. *rmrA* gene had 97% identity and *rmrB* and *rmrR* genes 98% in CFN42 and Mim1 genomes. Homologous genes were found being expressed in different plant rhizospheres [1,49,69].

Even though sv. mimosae strains are capable of forming nodules in *Leucaena*, we did not find genes resembling *mid* or *pyd* genes involved in the catabolism of the toxic

Figure 5 Graphic comparison of the Mim1 symbiotic plasmid pRetMim1e compared to other rhizobial replicons.

amino acid derivative mimosine found in *Leucaena* plants. Such genes were described from a *Rhizobium* sp. strain TAL 1145 (related to *R. tropici*) that was isolated from *Leucaena* plants [70]. *Mimosa pigra* has a much lower level of mimosine that *Leucaena* plants [71] and data for other mimosa species is not available.

Symbiovar phaseoli is prevalent in different *Phaseolus vulgaris* nodule bacteria

R. etli sv. mimosae strains have a broader host range than symbiovar phaseoli strains. In particular, *L. leucocephala* plants served as a host to distinguish symbiovar mimosae strains. Thirty-six *P. vulgaris* nodulating bacteria (with *R. etli*-like 16S rRNA gene sequences) obtained from the rain forest of Los Tuxtlas in Mexico corresponded to symbiovar phaseoli on the basis of *nodA* gene organization [10] and for being unable to nodulate *Leucaena* plants. We found that some strains that were previously classified as *R. etli* such as CIAT652, Ch24-10, CNPAF512, 8C-3 and Brasil5 now assigned to *R. phaseoli* [1], as well as others like Kim5, GR56 and CIAT 894 now assigned to recently named novel lineages [44] corresponded to sv. phaseoli when we analyzed their genomes. Symbiotic plasmids are highly conserved in symbiovar phaseoli [66] perhaps from being recently evolved [13]. Considering that the majority of *P. vulgaris* nodule isolates tested corresponded to sv. phaseoli we may conclude that this symbiovar is better adapted to its host, thus phaseoli seems to be a specialist symbiovar having a narrow range not including mimosa plants. The phaseoli symbiovar is found in several *Rhizobium* species or lineages (*R. gallicum, R. giardinii, R, phaseoli, R. etli* and *Rhizobium* sp. corresponding to PEL1 lineage). The widespread of this symbiovar may be in relation to its host historic worldwide distribution and to the transferability of the symbiotic plasmid, seemingly an epidemic plasmid. Besides having the phaseoli symbiovar, the *Rhizobium* species mentioned above have additional generalist symbiovars: gallicum or giardinii or mimosae Having alternative symbiovars with different host ranges may be advantageous in rhizobia, as it expands their legume niches and allows them to avoid the specialist-generalist dilemma.

Conclusions

The term symbiovar is validated with genomic analyses that show that a common genomic background may harbor different symbiotic plasmids determining host specificity. However, besides differences in the symbiotic plasmids there were differences in other ERs and in the chromosomes in the two strains analyzed, CFN42 and Mim1. In Mim1, Nod factors with sulfate modifications, secretion systems or ACC-deaminase may help explain the extended host range of symbiovar mimosae. In CFN42, *teu* genes that participate in exudate uptake [60] and

genes involved in Nod factor fucosylation (*nodZ*) may contribute to *P. vulgaris* host specialization. The discussion that mimosae is older than phaseoli may apply to gallicum and giardinii, thus we propose that gallicum and giardinii are older than phaseoli.

Additional file

Additional file 1: Table S1. Genes found in different genomic islands exclusive of Mim1.

Authors' contributions
EMR initiated and designed the study. MAR performed experiments. VG, DR, MAC contributed reagents and analyzed data. EOO, PB, RIS, JMR, JCM, LL performed data analysis. EMR, EOO, MAR wrote the paper. All authors read and approved the manuscript.

Author details
[1]Ecological Genomics programs, Genomics Science Center, CCG, Cuernavaca, Morelos, Mexico. [2]Evolutionary Genomics programs, Genomics Science Center, CCG, Cuernavaca, Morelos, Mexico. [3]Genome Engineering programs, Genomics Science Center, CCG, Cuernavaca, Morelos, Mexico.

References
1. López-Guerrero MG, Ormeño-Orrillo E, Velázquez E, Rogel MA, Acosta JL, Gónzalez V, Martínez J, Martínez-Romero E: *Rhizobium etli* taxonomy revised with novel genomic data and analyses. *Syst Appl Microbiol* 2012, **35**:353–358.
2. Rogel MA, Ormeño-Orrillo E, Martinez Romero E: **Symbiovars in rhizobia reflect bacterial adaptation to legumes.** *Syst Appl Microbiol* 2011, **34**:96–104.
3. Cobo-Díaz JF, Martínez-Hidalgo P, Fernández-González AJ, Martínez-Molina E, Toro N, Velázquez E, Fernández-López M: **The endemic *Genista versicolor* from Sierra Nevada National Park in Spain is nodulated by putative new *Bradyrhizobium* species and a novel symbiovar (sierranevadense).** *Syst Appl Microbiol* 2013, **37**:177–185.
4. Gubry-Rangin C, Béna G, Cleyet-Marel JC, Brunel B: **Definition and evolution of a new symbiovar, sv. riguloides, among *Ensifer meliloti* efficiently nodulating *Medicago* species.** *Syst Appl Microbiol* 2013, **36**:490–496.
5. Ramírez-Bahena MH, Chahboune R, Velázquez E, Gómez-Moriano A, Mora E, Peix A, Toro M: *Centrosema* is a promiscuous legume nodulated by several new putative species and symbiovars of *Bradyrhizobium* in various American countries. *Syst Appl Microbiol* 2013, **36**:392–400.
6. Laranjo M, Young JP, Oliveira S: **Multilocus sequence analysis reveals multiple symbiovars within *Mesorhizobium* species.** *Syst Appl Microbiol* 2012, **35**:359–367.
7. Ramírez-Bahena MH, Hernández M, Peix A, Velázquez E, León-Barrios M: **Mesorhizobial strains nodulating *Anagyris latifolia* and *Lotus berthelotii* in Tamadaya ravine (Tenerife, Canary Islands) are two symbiovars of the same species, *Mesorhizobium tamadayense* sp. nov.** *Syst Appl Microbiol* 2012, **35**:334–341.
8. Crossman LC, Castillo-Ramírez S, McAnnula C, Lozano L, Vernikos GS, Acosta JL, Ghazoui ZF, Hernández-González I, Meakin G, Walker AW, Hynes MF, Young JP, Downie JA, Romero D, Johnston AW, Dávila G, Parkhill J, González V: **A common genomic framework for a diverse assembly of plasmids in the symbiotic nitrogen fixing bacteria.** *PLoS One* 2008, **3**:e2567.

9. Wang ET, Rogel MA, García-de los Santos A, Martínez-Romero J, Cevallos MA, Martínez-Romero E: *Rhizobium etli* bv. mimosae, a novel biovar isolated from *Mimosa affinis*. *Int J Syst Bacteriol* 1999, **49**:1479–1491.

10. Ormeño-Orrillo E, Rogel-Hernández MA, Lloret L, López-López A, Martínez J, Barois I, Martínez-Romero E: Change in land use alters the diversity and composition of *Bradyrhizobium* communities and led to the introduction of *Rhizobium etli* into the tropical rain forest of Los Tuxtlas (Mexico). *Microb Ecol* 2012, **63**:822–834.

11. Vázquez M, Dávalos A, de las Peñas A, Sánchez F, Quinto C: Novel organization of the common nodulation genes in *Rhizobium leguminosarum* bv. phaseoli strains. *J Bacteriol* 1991, **173**:1250–1258.

12. Silva C, Vinuesa P, Eguiarte LE, Souza V, Martínez-Romero E: Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes. *Mol Ecol* 2005, **14**:4033–4050.

13. Martínez-Romero E: Coevolution in *Rhizobium*-legume symbiosis? *DNA Cell Biol* 2009, **28**:361–370.

14. Delgado-Salinas A, Bibler R, Lavin M: Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Syst Bot* 2006, **31**:779–791.

15. Dos Reis FBJ, Simon MF, Gross E, Boddey RM, Elliott GN, Neto NE, Loureiro MF, de Queiroz LP, Scotti MR, Chen WM, Norén A, Rubio MC, de Faria SM, Bontemps C, Goi SR, Young JP, Sprent JI, James EK: Nodulation and nitrogen fixation by *Mimosa* spp. in the Cerrado and Caatinga biomes of Brazil. *New Phytol* 2010, **186**:934–946.

16. Chen WM, de Faria SM, James EK, Elliott GN, Lin KY, Chou JH, Sheu SY, Cnockaert M, Sprent JI, Vandamme P: *Burkholderia nodosa* sp. nov., isolated from root nodules of the woody Brazilian legumes *Mimosa bimucronata* and *Mimosa scabrella*. *Int J Syst Evol Microbiol* 2007, **57**:1055–1059.

17. Chen WM, de Faria SM, Chou JH, James EK, Elliott GN, Sprent JI, Bontemps C, Young JP, Vandamme P: *Burkholderia sabiae* sp. nov., isolated from root nodules of *Mimosa caesalpiniifolia*. *Int J Syst Evol Microbiol* 2008, **58**:2174–2179.

18. Sheu SY, Chou JH, Bontemps C, Elliott GN, Gross E, James EK, Sprent JI, Young JP, Chen WM: *Burkholderia symbiotica* sp. nov., isolated from root nodules of *Mimosa* spp. native to north-east Brazil. *Int J Syst Evol Microbiol* 2012, **62**:2272–2278.

19. Taulé C, Zabaleta M, Mareque C, Platero R, Sanjurjo L, Sicardi M, Frioni L, Battistoni F, Fabiano E: New Betaproteobacterial *Rhizobium* strains able to efficiently nodulate *Parapiptadenia rigida* (Benth.) Brenan. *Appl Environ Microbiol* 2012, **78**:1692.

20. Simon MF, Grether R, de Queiroz LP, Särkinen TE, Dutra VF, Hughes CE: The evolutionary history of *Mimosa* (Leguminosae): toward a phylogeny of the sensitive plants. *Am J Bot* 2011, **98**:1201–1221.

21. Elliott GN, Chou JH, Chen WM, Bloemberg GV, Bontemps C, Martínez-Romero E, Velázquez E, Young JP, Sprent JI, James EK: *Burkholderia* spp. are the most competitive symbionts of *Mimosa*, particularly under N-limited conditions. *Environ Microbiol* 2009, **11**:762–778.

22. Gehlot HS, Tak N, Kaushik M, Mitra S, Chen W-M, Poweleit N, Panwar D, Poonar N, Parihar R, Tak A, Sankhla IS, Ojha A, Rao SR, Simon MF, Dos Reis FB, Perigolo N, Tripathi A, Sprent JI, Young JPW, James EK, Gyaneshwar P: An invasive *Mimosa* in India does not adopt the symbionts of its native relatives. *Ann Bot* 2013. in press.

23. Robledo M, Velázquez E, Ramírez-Bahena MH, García-Fraile P, Pérez-Alonso A, Rivas R, Martínez-Molina E, Mateos PF: The *celC* gene, a new phylogenetic marker useful for taxonomic studies in *Rhizobium*. *Syst Appl Microbiol* 2011, **34**:393–399.

24. Toledo I, Lloret L, Martínez-Romero E: *Sinorhizobium americanum* sp. nov., a new *Sinorhizobium* species nodulating native *Acacia* spp. in Mexico. *Syst Appl Microbiol* 2003, **26**:54–64. Erratum in: S*yst Appl Microbiol* 2003, 26:319.

25. Hynes MF, McGregor NF: Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. *Mol Microbiol* 1990, **4**:567–574.

26. Martinez-Romero E, Rosenblueth M: Increased bean (*Phaseolus vulgaris* L.) nodulation competitiveness of genetically modified *Rhizobium* strains. *Appl Environ Microbiol* 1990, **56**:2384–2388.

27. Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008, **18**:821–829.

28. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 2011, **27**:578–579.

29. Gordon D, Abajian C, Green P: Consed: a graphical tool for sequence finishing. *Genome Res* 1998, **8**:195–202.

30. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL: Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 1999, **27**:4636–4641.

31. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B: Artemis: sequence visualization and annotation. *Bioinformatics* 2000, **16**:944–945.

32. González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G: The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci U S A* 2006, **103**:3834–3839.

33. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: GenBank. *Nucleic Acids Res* 2008, **36**:D25–D30.

34. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MDR, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJA, Zdobnov EM: The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 2001, **29**:37–40.

35. Richter M, Rossello-Mora R: Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009, **106**:19126–19131.

36. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: Versatile and open software for comparing large genomes. *Genome Biol* 2004, **5**:R12.

37. Enright AJ, Van Dongen S, Ouzounis CA: An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 2002, **30**:1575–1584.

38. Vernikos GS, Parkhill J: Interpolated variable order motifs for detection of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 2006, **22**:2196–2203.

39. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004, **32**:1792–1797.

40. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O: New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010, **59**:307–321.

41. Ronquist F, Huelsenbeck JP: MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, **19**:1572–1574.

42. Harrison PW, Lower RP, Kim NK, Young JP: Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol* 2010, **18**:141–148.

43. Cevallos MA, Cervantes-Rivera R, Gutiérrez-Ríos RM: The *repABC* plasmid family. *Plasmid* 2008, **60**:19–37.

44. Ribeiro RA, Ormeño-Orrillo E, Fuzinatto Dall'Agnol R, Graham PH, Martínez-Romero E, Hungria M: Novel *Rhizobium* lineages isolated from root nodules of common bean (*Phaseolus vulgaris* L.) in Andean and Mesoamerican areas. *Res Microbiol* 2013. in press.

45. Pardo MA, Lagunez J, Miranda J, Martínez E: Nodulating ability of *Rhizobium tropici* is conditioned by a plasmid-encoded citrate synthase. *Mol Microbiol* 1994, **11**:315–321.

46. Hernández-Lucas I, Pardo MA, Segovia L, Miranda J, Martínez-Romero E: *Rhizobium tropici* chromosomal citrate synthase gene. *Appl Environ Microbiol* 1995, **61**:3992–3997.

47. Rosemeyer V, Michiels J, Verreth C, Vanderleyden J: *luxI*- and *luxR*-homologous genes of *Rhizobium etli* CNPAF512 contribute to synthesis of autoinducer molecules and nodulation of *Phaseolus vulgaris*. *J Bacteriol* 1998, **180**:815–821.

48. Daniels R, De Vos DE, Desair J, Raedschelders G, Luyten E, Rosemeyer V, Verreth C, Schoeters E, Vanderleyden J, Michiels J: The *cin* quorum sensing locus of *Rhizobium etli* CNPAF512 affects growth and symbiotic nitrogen fixation. *J Biol Chem* 2002, **277**:462–468.

49. López-Guerrero MG, Ormeño-Orrillo E, Acosta JL, Mendoza-Vargas A, Rogel MA, Ramírez MA, Rosenblueth M, Martínez-Romero J, Martínez-Romero E: Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. *Plasmid* 2012, **68**:149–158.

50. Poupot R, Martinez-Romero E, Gautier N, Promé JC: Wild type *Rhizobium etli*, a bean symbiont, produces acetyl-fucosylated, N-methylated, and carbamoylated nodulation factors. *J Biol Chem* 1995, **270**:6050–6055.

51. Laeremans T, Snoeck C, Mariën J, Verreth C, Martínez-Romero E, Promé JC, Vanderleyden J: *Phaseolus vulgaris* recognizes *Azorhizobium caulinodans*

Nod factors with a variety of chemical substituents. *Mol Plant Microbe Interact* 1999, **12**:820–824.

52. Waelkens F, Voets T, Vlassak K, Vanderleyden J, van Rhijn P: The *nodS* gene of *Rhizobium tropici* strain CIAT899 is necessary for nodulation on *Phaseolus vulgaris* and on *Leucaena leucocephala*. *Mol Plant Microbe Interact* 1995, **8**:147–154.

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

54. van Rhijn P, Luyten E, Vlassak K, Vanderleyden J: Isolation and characterization of a pSym locus of *Rhizobium* sp. BR816 that extends nodulation ability of narrow host range *Phaseolus vulgaris* symbionts to *Leucaena leucocephala*. *Mol Plant Microbe Interact* 1996, **9**:74–77.

55. Vlassak KM, Luyten E, Verreth C, van Rhijn P, Bisseling T, Vanderleyden J: The *Rhizobium* sp. BR816 *nodO* gene can function as a determinant for nodulation of *Leucaena leucocephala*, *Phaseolus vulgaris*, and *Trifolium repens* by a diversity of *Rhizobium* spp. *Mol Plant Microbe Interact* 1998, **11**:383–392.

56. Tittabutr P, Awaya JD, Li QX, Borthakur D: The cloned 1-aminocyclopropane-1-carboxylate (ACC) deaminase gene from *Sinorhizobium* sp. strain BL3 in *Rhizobium* sp. strain TAL1145 promotes nodulation and growth of *Leucaena leucocephala*. *Syst Appl Microbiol* 2008, **31**:141–150.

57. González V, Bustos P, Ramírez-Romero MA, Medrano-Soto A, Salgado H, Hernández-González I, Hernández-Celis JC, Quintero V, Moreno-Hagelsieb G, Girard L, Rodríguez O, Flores M, Cevallos MA, Collado-Vides J, Romero D, Dávila G: The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biol* 2003, **4**:R36.

58. Girard L, Brom S, Dávalos A, López O, Soberón M, Romero D: Differential regulation of *fixN*-reiterated genes in *Rhizobium etli* by a novel *fixL–fixK* cascade. *Mol Plant Microbe Interact* 2000, **12**:1283–1292.

59. Zamorano-Sánchez D, Reyes-González A, Gómez-Hernández N, Rivera P, Georgellis D, Girard L: FxkR provides the missing link in the *fixL-fixK* signal transduction cascade in *Rhizobium etli* CFN42. *Mol Plant Microbe Interact* 2012, **25**:1506–1517.

60. Rosenblueth M, Hynes MF, Martínez-Romero E: *Rhizobium tropici teu* genes involved in specific uptake of *Phaseolus vulgaris* bean-exudate compounds. *Mol Gen Genet* 1998, **258**:587–598.

61. Fauvart M, Michiels J: Rhizobial secreted proteins as determinants of host specificity in the rhizobium-legume symbiosis. *FEMS Microbiol Lett* 2008, **285**:1–9.

62. Baginsky C, Brito B, Imperial J, Palacios J-M, Ruiz-Argüeso T: Diversity and evolution of hydrogenase systems in rhizobia. *Appl Environ Microbiol* 2002, **68**:4915–4924.

63. Brito B, Prieto RI, Cabrera E, Mandrand-Berthelot MA, Imperial J, Ruiz-Argueso T, Palacios JM: *Rhizobium leguminosarum hupE* encodes a nickel transporter required for hydrogenase activity. *J Bacteriol* 2010, **192**:925–935.

64. López-López A, Rogel MA, Ormeño-Orrillo E, Martínez-Romero J, Martínez-Romero E: *Phaseolus vulgaris* seed-borne endophytic community with novel bacterial species such as *Rhizobium endophyticum* sp. nov. *Syst Appl Microbiol* 2010, **33**:322–327.

65. Lozano L, Hernández-González I, Bustos P, Santamaría RI, Souza V, Young JP, Dávila G, González V: Evolutionary dynamics of insertion sequences in relation to the evolutionary histories of the chromosome and symbiotic plasmid genes of *Rhizobium etli* populations. *Appl Environ Microbiol* 2010, **76**:6504–6513.

66. González V, Acosta JL, Santamaría RI, Bustos P, Fernández JL, Hernández González IL, Díaz R, Flores M, Palacios R, Mora J, Dávila G: Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. *Appl Environ Microbiol* 2010, **76**:1604–1614.

67. Brom S, Girard L, Tun-Garrido C, García-de los Santos A, Bustos P, González V, Romero D: Transfer of the symbiotic plasmid of *Rhizobium etli* CFN42 requires cointegration with p42a, which may be mediated by site-specific recombination. *J Bacteriol* 2004, **186**:7538–7548.

68. González-Pasayo R, Martínez-Romero E: Multiresistance genes of *Rhizobium etli* CFN42. *Mol Plant Microbe Interact* 2000, **13**:572–577.

69. Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS: Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol* 2011, **12**:R106.

70. Borthakur D, Soedarjo M: Isolation and Characterization of a DNA Fragment Containing Genes for Mimosine Degradation from *Rhizobium* sp. Strain TAL1145. In *Highlights of Nitrogen Fixation Research*. Edited by Martinez E, Hernandez G. New York: Kluwer/Plenum; 1999:91–95.

71. Soedarjo M, Borthakur D: Mimosine, a toxin produced by the tree-legume *Leucaena* provides a nodulation competition advantage to mimosine-degrading *Rhizobium* strains. *Soil Biol Biochem* 1998, **30**:1605–1613.