Funct Integr Genomics (2009) 9:263–270
DOI 10.1007/s10142-009-0109-z

SHORT COMMUNICATION

Novel genes related to nodulation, secretion systems, and surface structures revealed by a genome draft of *Rhizobium tropici* strain PRF 81

Fabiana G. S. Pinto · Ligia M. O. Chueire · Ana Tereza R. Vasconcelos · Marisa F. Nicolás · Luiz G. P. Almeida · Rangel C. Souza · Pâmela Menna · Fernando G. Barcellos · Manuel Megías · Mariangela Hungria

Received: 27 October 2008 / Revised: 9 December 2008 / Accepted: 30 December 2008 / Published online: 29 January 2009

Abstract *Rhizobium tropici* is representative of the diversity of tropical rhizobia, besides comprising strains very effective in fixing N_2 in symbiosis with the common bean (*Phaseolus vulgaris* L.). The genome of a Brazilian commercial inoculant *R. tropici* strain (PRF 81, =SEMIA 4088), estimated at 7.85 Mb, was analyzed through a total of 9,026 shotgun reads, assembled in 1,668 phrap contigs, and covering ≈30% of the genome. Annotation identified 2,135 coding DNA sequences (CDS), and only 57.2% have possible functions. The genome comprises a mosaic of genes, with CDS showing the highest similarities with 134 microorganisms, none of which represents more than 19% of the CDS with putative known functions. The high saprophytic capacity of PRF 81 may reside in a variety of genes related to transport, biodegradation of xenobiotics, defense, and secretion proteins, many of which were reported for the first time in the present study. Novelty was also found in nodulation (*nodG*, a double *nodIJ* system, *nodT*, *nolF*, *nolG*) and capsular polysaccharide genes, showing stronger similarities with *Sinorhizobium* (=*Ensifer*) than with the main symbionts of the common bean—*R. etli* and *R. leguminosarum*—suggesting that the original host of *R. tropici* might be another tropical legume or emphasizing the highly promiscuous nature of this rhizobial species.

F. G. S. Pinto · L. M. O. Chueire · P. Menna · F. G. Barcellos · M. Hungria (✉)
Embrapa Soja,
Cx. Postal 231, CEP 86001-970 Londrina, Paraná, Brazil
e-mail: hungria@cnpso.embrapa.br

F. G. S. Pinto · P. Menna · F. G. Barcellos · M. Hungria
Department of Microbiology, Universidade Estadual de Londrina,
Londrina, Brazil

F. G. S. Pinto · L. M. O. Chueire · A. T. R. Vasconcelos · P. Menna · M. Hungria
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-MCT),
Brasilia, Brazil

A. T. R. Vasconcelos · M. F. Nicolás · L. G. P. Almeida · R. C. Souza
Laboratório Nacional de Computação Científica,
Rua Getúlio Vargas 333,
CEP 25651-071 Petrópolis, Rio de Janeiro, Brazil

F. G. Barcellos
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES),
Brasilia, Brazil

M. Megías
Facultad de Farmacia, Univ. Sevilla,
Apdo. Postal 874, 41080 Sevilla, Spain

Present address:
F. G. S. Pinto
Universidade Estadual do Oeste do Paraná,
Campus Cascavel,
CEP 85819-110 Cascavel, Paraná, Brazil

Keywords Biological nitrogen fixation · Genome · Nodulation · *Phaseolus vulgaris* · *Rhizobium tropici* · Symbiosis

Introduction

The alpha- and beta-Proteobacteria classes include several species commonly denominated as "rhizobia", capable of nodulating and fixing N_2 with legumes. However, despite the importance of symbiotic diazotrophs to the global cycling of N, relatively few rhizobial strains have been completely sequenced so far—there are only nine genomes published on 10/2008 [*Mesorhizobium loti* MAFF303099, *Sinorhizobium* (=*Ensifer*) *meliloti* 1021, *Bradyrhizobium japonicum* USDA 110, *Rhizobium etli* biovar phaseoli CFN 42, *R. leguminosarum* biovar viciae 3841. *Bradyrhizobium* sp. ORS278 and BTAi1, *Azorhizobium caulinodans* ORS 571, and *Cupriavidus taiwanensis* LMG19424]. *Rhizobium tropici* is a poorly studied species, originally isolated from the common bean (*Phaseolus vulgaris* L.) in South America (Martínez-Romero et al. 1991). Its precise origin is still unknown, but Brazil is a strong candidate for its country of origin because the region is the source of the majority of the strains isolated so far (e.g., Mercante et al. 1998; Hungria et al. 2000; Andrade et al. 2002; Grange and Hungria 2004; Pinto et al. 2007). It has been hypothesized that *R. tropici* might be originally a symbiont of other indigenous host legumes of South America, e.g., *Mimosa* and *Gliricidia*, in Brazil and Mexico, respectively (Acosta-Durán and Martínez-Romero 2002; Menna et al. 2006), but its promiscuous nature has also been often emphasized (e.g., Michiels et al. 1998). Another interesting feature of *R. tropici* is its close phylogenetic relationship with the plant pathogen *Agrobacterium* (=*Rhizobium*) *tumefaciens* (Lloret and Martínez-Romero 2005). Finally, *R. tropici* is an important reservoir of strains, including CIAT 899, PRF 81, and H 12, which are effective in fixing N_2, competitive against indigenous rhizobial populations, genetically stable, and adapted to stressful tropical environments, all of which are important properties for their use in inoculants in tropical regions (Hungria et al. 2000, 2003). *R. tropici* is thus an interesting candidate for producing a draft genome, as demonstrated with *Rhizobium* sp. NGR234 (Viprey et al. 2000) and *B. japonicum* CPAC 15 (Godoy et al. 2008).

The objective of this study was to carry out a sequence level analysis of the genome of *R. tropici* strain PRF 81 (=SEMIA 4088), which is commercially recommended for the production of inoculants for common bean crops in Brazil. The study identified new genes that could account for the biological properties of this rhizobium strain.

Materials and methods

Bacterial strains, growth conditions, DNA extraction, and genomic size estimated by pulse field gel electrophoresis (PFGE)

Escherichia coli was grown at 37°C on Luria–Bertani medium supplemented with antibiotics, while *R. tropici* strain PRF 81 (=SEMIA 4088) was obtained from the "Diazotrophic and Plant Growth Promoting Bacteria Collection" of Embrapa Soja and grown on tryptone-yeast extract (TY) medium. For the PFGE analysis, PRF 81 was grown on TY for 3 days, and the DNA was prepared as described by Suzuki et al. (2001). PFGE was carried out using contour-clamped homogenous electric field (CHEF-DRIII; Bio-Rad), in 0.8% agarose gels prepared in 0.5× TBE with a constant voltage of 6 V cm^{-1} and pulses of 70 s for 18 h. The sizes of the fragments were estimated using the BioNumerics software (Applied Mathematics, Kortrijk, Belgium, version 4.6), based on the size of the chromosomal DNA of *Saccharomyces cerevisiae* (Bio-Rad).

Library construction and sequencing of shotgun clones

Shotgun libraries of the total DNA of PRF 81 were prepared as described before (Godoy et al. 2008) and involved DNA purification and random mechanical shearing by nebulization. Vector pUC18 was used for cloning and *E. coli* strain DH10B for transformation. Sequencing of shotgun clones was performed as described before (Godoy et al. 2008), on a MegaBACE 1000 capillary sequencer (Amersham Pharmacia Biotech).

Assemblage and annotation

Sequences were assembled and annotated using the software "System for Automated Bacterial Integrated Annotation" (SABIA; Almeida et al. 2004), developed to integrate public domain softwares. Programs used in the assemblage process included the Phred, Phrap, and Consed, with gene coding regions identified with "Glimmer". The sequences were compared with those deposited at the GenBank database using Basic Local Alignment Search Tool (BLAST, available at http://www.ncbi.nlm.nih.gov, Altschul et al. 1997) for nucleotides (BLASTN) and for proteins (BLASTP). In the software, metabolic pathways are predicted based on the Kyoto Encyclopedia of Genes and Genomes (KEGG, Kanehisa and Goto 2000), and protein sequences are compared by clusters of orthologous groups of proteins (COG, Tatusov et al. 2000), INTERPRO (families, domains and functional sites of proteins), PSORT (protein localization), TCDB (http://tcdb.org), and UniProt, for protein information. More details and references for all

programs cited above and integrated in the SABIA are given elsewhere (Almeida et al. 2004; Godoy et al. 2008).

The size of the each coding DNA sequence (CDS) without database homologues was established at ≈50 bp; this short size aimed at both capturing the information about an increased number of small proteins that has been reported, as well as information about partial proteins detected in the genome draft. In the case of shorter sequences, an observation was included in the CDS description.

All CDS were manually curated using a cutoff *E* value of $\approx 10^{-10}$ and ≈50% of identity (considering BLASTP, with the identity of proteins, based on the similarity of the amino acids). The following criteria were applied: CDS were target as hypothetical when no homolog could be detected; conserved hypothetical were those displaying strong similarity with several hypothetical or weak similarity to known genes; and CDS with high similarity with known genes were assigned the same name as the matching gene.

All supplementary information is available at http://www.bnf.lncc.br/webbie/final/main.html (or http://www.bnf.lncc.br, then *Rhizobium tropici* PRF 81, genome draft).

Results and discussion

Genome size

Digestion of intact genomic DNA of strain PRF 81 with the rare cutter enzyme *Swa*I generated fragments ranging from 2,200 to 225 kb, and the PFGE profiles indicated that the genome consists of a circular chromosome of 5,305 Mb (68% of the genome) and four plasmids of about of 1,700, 510, 185, and 150 kb. This estimate included the results of a previous report on the plasmid profile of PRF 81 (Pinto et al. 2007), as well as the knowledge that we have from other rhizobial genomes. In the two other *Rhizobium* genomic sequences published so far, *R. etli* and *R. leguminosarum*, the chromosomes represented similar percentages of the whole genome, of 67.1% and 65.2%, respectively; however, both had six plasmids (González et al. 2006; Young et al. 2006). The genome of *R. tropici* is larger than those of *R. etli* [6.53 Mb, U80928, (GenBank)], *Sinorhizobium meliloti* (=*Ensifer)* [6.69 Mb, chromosome AL591688 (EMBL), pSymA AE006469 (GenBank), pSymB (AL591985) (EMBL)], *M. loti* [7.6 Mb, AP002994 to AP003017 (EMBL)], *R. leguminosarum* [7.79 Mb, AM236080 to AM236086 (EMBL)], *Bradyrhizobium* sp. ORS278 [7.46 Mb, CU234118 (EMBL)], *C. taiwanensis* LMG19424 [6.48 Mb, NC 010530 and NC 010528], and *A. caulinodans* ORS571 [5.37 Mb, AP009384 (GenBank)], but smaller than those of *B. japonicum* strain USDA 110 [9 Mb, BA000040 (DDBJ/GenBank/EMBL)] and *Bradyrhizobium* sp. BTAi1 [8.26 Mb, CU234118 (EMBL)].

Assemblage and general genomic features

Approximately 4,800 clones from eight shotgun libraries were sequenced and resulted in 9,026 reads without vectors, and in the deposit of 10,046,403 bp, or 1.28 times the estimated genome, with 43.45% of the bases with a phred score of ≥20, and 34.84% with a score of ≥30, usual score values considered in the assemblage of bacterial genomes. The assemblage produced 2,789 contigs, with an average of 3.24 reads per contig. As in the "shotgun" library, some fragments represent identical sequences, and the actual coverage was of 1,668 phrap contigs, or 31% of the genome (supplementary information). After manual annotation and using the criteria described in the "Materials and methods" section, 2,135 CDS were confirmed and classified as follows: 57.2% with putative known functions, 30.4% as conserved hypothetical, and 12.4% as hypothetical genes (supplementary information). These numbers highlight our still poor knowledge of *R. tropici*; in comparison, in *R. etli* and *R. leguminosarum*, putative genes with known functions were reported at 71% and 70%, respectively (González et al. 2006; Young et al. 2006).

Mobile elements in PRF 81 were also particularly interesting; of the 17 putative transposases annotated, nine showed higher BLASTP similarity with bacteria other than rhizobia, including *Agrobacterium*, *Azoarcus*, *Brucella melitensis*, *Burkholderia fungorum*, and *Rhodospirillum rubrum* (supplementary information). The mobility shown by the transposases might help to explain the mosaic of genes found in the genome of PRF 81.

Functional classification by COG, Monica Riley and KEGG

Considering the matches with COG (a total of 1,561 CDS), 1,262 were classified into 20 COG categories (Tatusov et al. 2000), 208 showed only general predictions, and 91 had unknown functions (all categories and the numbers of CDS in each category are available as supplementary information). Higher percentages of CDS were present in the following COG categories: E-amino acid transport and metabolism (14.67%); G-carbohydrate transport and metabolism (9.67%); K-transcription (9.60%); C-energy production and conversion (6.41%); and P-inorganic ion transport and metabolism (5.63%). Considering the CDS classified in transport and metabolism categories, G, E, F, H, I, P, and Q, they represented 42.72% of the entire predicted genes, a strong indication of the free-living mode

and adaptation to a broad range of ecosystems varying in nutrient availability. It is noteworthy that the percentages of CDS in these categories, as well as in the categories of energy production and conversion (category C), were considerably higher than in other fast-growing rhizobia. Still thinking in energy saving, PRF 81 might also have the genes encoding the uptake-hydrogenase systems, as *hupK* and two copies of *hypF* were present in the genome, all of them showing high BLASTP similarity with *R. leguminosarum*. The distribution of CDS of PRF 81 in all COG categories (supplementary information) suggests that a representative coverage of the genome was obtained.

Most CDS were also categorized according to Riley's classification (Riley 1993): 1,132 CDS had matches with known genes, 671 were classified as conserved hypothetical, and 267 with hypothetical functions, a total of 2,070 CDS (supplementary information).

Of the CDS with matches in KEGG (Kanehisa and Goto 2000), a large number—65—fitted into biodegradation of xenobiotics, distributed in 11 subcategories. Of those CDS, 58% were related to the metabolism of benzoate, and there were also putative genes participating in the degradation of 1,2-dichloroethane, caprolactam, ethylbenzene, fluorine, nitrobenzene, styrene, tetrachloroethene, and gamma-hexachlorocyclohexane, which may confer metabolic versatility, favoring persistence in soils; additionally, they might indicate a potential use in the remediation of contaminated soils.

One major feature of the panoramic genome of PRF 81 is that it is represented by a mosaic of genes showing similarities with a variety of microorganisms and, as far as we are aware, with no parallel in other bacteria, what is highlighted by comparison using the KEGG database. Therefore, making a comparison of the genome of the PRF 81 with the functional classification of sequenced genomes available in KEGG, the 2,135 CDS of PRF 81 have shown the highest similarities with an impressive number of microorganisms—134 (supplementary information). In addition, in contrast with other sequenced genomes, where the great majority of the CDS show the highest similarity with one or two microorganisms, in the genome of PRF 81 the percentages of similarity with each microorganisms were not high, such that the six highest matches in KEGG were obtained with *R. leguminosarum* (29.22%), *R. etli* (19.81%), *A. tumefaciens* (5.76%), *S. meliloti* (4.87%), *M. loti* (4.87%), and *S. medicae* (3.93%). These percentages decreased when considering only the valid CDS (according to the definition in the "Materials and methods" section): 19.06%, 13.44%, 3.98%, 3.37%, 2.9%, and 2.62%, respectively.

Secretion systems

Several putative genes belonging to types I to IV secretion systems were annotated in the genome of PRF 81. Sec-proteins included the chaperone *secB* and the inner membrane proteins *secD* and *secY*, as well as the putative peptidase *lepB*. In relation to the type I system, putative ABC systems involved in macromolecules export identified in PRF 81 include the ATP-binding component responsible for resistance to the macrolide antibiotic tylosin, with similarity to TlrC protein of *Streptomyces fradiae* (Rosteck et al. 1991), and also *msbA*. Another interesting type I system found in PRF 81 is involved in (1-2)-β-glucan export, showing similarity with the NdvA protein of *S. meliloti* (Stanfield et al. 1988), for the first time reported in *R. tropici*. It has been demonstrated that *ndvA* is required for normal nodulation of alfalfa roots (Dickstein et al. 1988; Stanfield et al. 1988).

Considering the type III secretion systems (T3SS), only flagellum-related genes were identified. In relation to the T4SS systems, we have identified several genes encoding proteins associated with the conjugation-transfer machinery, showing similarity to *R. etli* and *S. meliloti*. Among the CDS of PRF 81, three had homology to the *virB* genes of *A. tumefaciens*—*trbD* (=virB3), *trbI* (=virB10) and *trbL* (=virB6)—and three were classified as *tra* genes, related to proteins of conjugative transfer: *traD*, *traR*, and *traG* (=*virD4*).

Surface structures

There is evidence that the production of a variety of symbiotically active polysaccharides may allow rhizobial strains to adapt to changing environmental conditions and to interact efficiently with legumes (Skorupska et al. 2006), but the only studies elucidating the structures of EPS, LPS, and (1-2)-β-D-glucan of *R. tropici* have been performed with strain CIAT 899 (Gil-Serrano et al. 1990, 1993, 1994).

Synthesis of repeating units of EPS, their modification, polymerization, and export to the cell surface are controlled by clusters of genes—*exo*/*exs*, *exp*, or *pss*—localized on rhizobial megaplasmids or on the chromosome (Skorupska et al. 2006), and nine of those genes were detected in PRF 81. In relation to the genes involved in the biosynthesis of nucleotide sugar precursors of EPS, most CDS of PRF 81, like those encoding for *exoB*, *exoU*, and endo-1,3-1,4-β-glycanase, showed higher BLASTP similarity to *S. meliloti* 1021 (Reinhold et al. 1994). Higher similarities with *S. meliloti* and *M. loti* were also detected in genes related to EPS assembly and export genes, such as *exoP* and *exoT*, as well as in enzymes modifying the unit with non-sugar moieties, such as *exoV*. Also in relation to the LPS, several putative genes, such as CDS RT02050 (an acyl-carrier protein), *msbA* (ABC transporter of lipid A), *lpcC* (a mannosyl transferase of lipopolysaccharide core biosynthesis), *rfaZ* (RT20472) (putative glycosyltransferase), and

RT13143 (biosynthesis of (1–2)-β-glucan), have shown high similarity with *S. meliloti* 1021.

Rhizobial surface polysaccharides (SPSs), including the capsular polysaccharides (KPS), exist in close proximity to plant-derived membranes throughout the infection process. It is known that SPSs are essential for bacterial survival and adaptation and as potential determinants of nodulation and/or host specificity (Le Quéré et al. 2006); they have been detected only in *S. meliloti* 1021 (Kereszt et al. 1998), in *S. fredii* strains USDA 257 (Forsberg and Reuhs 1997) and HH103 (Parada et al. 2006), and in *Rhizobium* sp. NGR234 (Le Quéré et al. 2006). In PRF 81, three CDS similar to SPSs were detected: RT11392, showing similarity with the capsular gene *rkpH* of *S. meliloti* 1021 (50%); *rkpI*, showing similarity to *S. meliloti* 1021 (48%) and *S. fredii* HH103 (46%); and RT10314, similar to *rkpJ*, responsible for the export of the capsular polysaccharide, also showing higher BLASTP similarity with *S. fredii* (55%) and *S. meliloti* 1021 (54%). No homology was found between the CDS encoding for KPS in PRF 81 and the sequenced genomes of *R. etli*, *R. leguminosarum*, *M. loti*, or *B. japonicum*. This is the first time that genes participating in the biosynthesis of KPS have been described in *R. tropici*, and it is noteworthy that the EPS and LPS of PRF 81 resemble those of *S. meliloti* (typical from temperate regions and capable of forming indeterminate nodules) and of *R. leguminosarum* and *M. loti*, than of other tropical rhizobia, such as the rhizobial strain NGR 234.

Nitrogen fixation and nodulation genes

Several CDS related to the process of biological N_2 fixation were detected in the draft genome of PRF 81, including genes related to the biosynthesis of the nitrogenase (*nifN*, *nifR nifS*, *nifU*, and *nifW*) and to the regulation of the expression of these genes (*nifA*, *ntrB*, *ntrC*, *ntrX*, and *ntrY*). We also identified six *nif* (*nifA*, *nifN*, *nifR*, *nifS*, *nifU*, and *nifW*) and three *fix* (*fixA, fixB*, and *fixR*) genes.

In *R. tropici*, *nod* genes have been studied using strain CIAT 899 as a model, and the gene organization and function on the operon *nodDABCSUIJHPQ* has been elucidated (Vargas et al. 1990; Sousa et al. 1993; Manyani et al. 2001). In the genome of PRF 81, we found 12 CDS related to nodulation with half of these not previously reported in *R. tropici* (Table 1). The *nodC* gene is responsible for the biosynthesis of the basic structure of the Nod factor, and in PRF 81, the highest BLASTP similarity was found with CIAT 899 (Folch-Mallol et al. 1996). Another gene resembling CIAT 899 is *nodU*, a 6-0-carbomoyl-transferase involved in the addition of the carbomoyl in the non-reductor extreme of the Nod factor (Waelkens et al. 1995) and detected in all rhizobia except for *R. leguminosarum* (Young et al. 2006). *nodU* is apparently related to host specificity, as mutants of CIAT 899 have decreased nodulation of *Leucaena* but not of the common bean (Waelkens et al. 1995).

With regards the *nod* genes related to the RND efflux pump, the ATP-binding protein I (NodI) described in CIAT 899 (Folch-Mallol et al. 1996; Manyani et al. 2001) is 100% similar to two CDS of PRF 81, RT23334 and RT23345. However, two other contiguous CDS (RT20830 and RT20846) were found, possibly representing a second NodIJ system. Those CDS show high BLASTP similarity (83% and 88%, respectively) to two type I secretion proteins of *R. leguminosarum* bv. trifolii (AAL14907.1), and also with HndI and HndJ proteins of *Agrobacterium*. In *R. etli* CFN 42, mutants of *nodIJ* delay the exportation of Nod factors (Cárdenas et al. 1996), while in *S. fredii* HH103, in addition to the delay, the mutants also release different structures of Nod factors (M. Megías, unpublished). This is the first time that a second *nodIJ* transport system has been described in *R. tropici*. Interesting is also the CDS RT05132, which is similar to the RND efflux system outer membrane lipoprotein NodT of bacteria belonging to the genera such as *Burkholderia* and *Pseudomonas* but with no homologs in any rhizobia studies to date. The efflux transporter NolG of *S. meliloti* 1021 is 50% similar to RT22697 and has also never been reported in *R. tropici*. Functionality of these new genes should be investigated in *R. tropici*, as they could play a role in nodulation of different hosts or in the kinetics of nodulation.

In relation to the biosynthesis of Nod factors, *nodG* and *nodN* are responsible for modifications in the structure of the Nod factor and were present in PRF 81. RT11914 shows high BLASTP similarity (92%) with *nodG* of *R. leguminosarum*, encoding a 3-oxoacyl-(acyl) carrier protein reductase involved in the biosynthesis and transfer of common fatty acids (López-Lara and Geiger 2001); high homology was also found with other rhizobia (Table 1), but this putative gene has never been reported in *R. tropici*. *nodN* (RT05456) showed high BLASTP similarity with the protein of CIAT 899. In addition, RT14525 showed similarity with *noeJ*, responsible for the activity of mannose-1-phosphate guanylyl transferase of CIAT 899 (Nogales et al. 2002) and similar proteins in other rhizobia (Table 1). *nolF* showing low similarity to a gene encoding for a glucosamine synthetase described in *S. meliloti* 1021 (Galibert et al. 2001) has also never been reported in *R. tropici.*

R. tropici nodulates legumes of tropical origin such as the common bean and *Leucaena* spp., and it was amazing to find that the nodulation genes of PRF 81 and CIAT 899 resemble more closely those of *S. meliloti* and *R. leguminosarum*, symbionts of temperate legumes. We have expected the nodulation genes to show higher similarity with those of tropical rhizobia such as *Rhizobium* sp. NGR 234 and mainly of *R. etli*, considered as the main symbiont

Table 1 Nodulation genes detected in the genome draft of *Rhizobium tropici* PRF 81 and higher BLASTP similarity with other microorganisms

Gene	CDS[a]	Percentage of identity in BLASTP
nodC	RT23361	*Rhizobium tropici* CIAT 899 (100%), *Sinorhizobium* sp. BR 816 (66%), *Mesorhizobium loti* MAFF303099 (66%), *R. etli* CFN42 (60%), *Rhizobium* sp. N33 (58%), *Burkolderia tuberum* STM678 (57%)
nodG[b]	RT11914	*R. leguminosarum* LPR5045 (92%), *Agrobacterium tumefaciens* C58 (90%), *Sinorhizobium meliloti* 1021 (89%), *M. loti* MAFF303099 (88%), *Rhizobiun* sp. N33 (88%), *Mesorhizobium* sp. BNC1 (80%)
nodI	RT23345, RT23334	*R. tropici* CIAT 899 (100%), *M. loti* MAFF303099 (85%), *Rhizobium* sp. NGR234 (84%), *S. meliloti* 1021 (83%), *Sinorhizobium* sp. BR816 (79%), *Rhizobium* sp. N33 (79%)
nodJ[b] (=*hndI*)	RT20830	*R. leguminosarum* bv. trifolii (83%), *A. tumefaciens* C58 (70%), *S. meliloti* 1021 (67%)
nodJ[a] (=*hndJ*)	RT20846	*R. leguminosarum* bv. trifolii LPR5045 (88%), *A. tumefaciens* C58 (86%), *S. meliloti* 1021 (80%), *Azoarcus* sp. EbN1, *Bradyrhizobium japonicum* USDA 110, *Mesorhizobium* sp. BCN1 (71%)
nodN	RT05456	*R. tropici* CIAT 899 (92%), *A. tumefaciens* C58 (74%), *S. meliloti* 1021 (72%), *R. leguminosarum* bv. viceae (70%)
nodT[b]	RT05132	No homology with rhizobia; 33% with *Burkholderia cenocepacia* AU 1054 (33%), *Burkholderia cepacia* R18194 (31%), *Pseudomonas syringae* pv. syringae B728a (32%)
nodU	RT23354	*R. tropici* CIAT 899 (100%), *Sinorhizobium* sp. BR816 (82%), *Rhizobium* sp. NGR234 (77%), *B. japonicum* USDA 110 (73%), *S. fredii* USDA257 (73%), *Bradyrhizobium* sp. SNU001 and WM9 (72%), *R. galegae* (67%)
noeJ	RT14525	*R. tropici* CIAT 899 (96%), *Rhizobium* sp. NGR234 (66%), *Brucella abortus* biovar 1 9-941 (63%), *B. japonicum* USDA110 (59%)
nolF[b]	RT22693	*S. meliloti* 1021 (35%), *A. tumefaciens* C58 (34%)
nolG[a]	RT22697	*S. meliloti* 1021 (50%), *A. tumefaciens* C58 (46%), *Chromobacterium violaceum* ATCC 12472 (43%)

[a] Coding DNA sequence(s)
[b] Described for the first time in *R. tropici*

of the common bean in the Mesoamerican center of genetic diversification (Segovia et al. 1993; Grange et al. 2007). This finding also raises the issue that although described as a common bean symbiont (Martínez-Romero et al. 1991), the lack of similarity of nodulation genes of *R. tropici* PRF 81 with *R. etli*—the predominant species in the both the Mesoamerican and the Andean centers of genetic origin (e.g., Segovia et al. 1993; Aguilar et al. 2004; Grange et al. 2007)—adds more evidences to the hypothesis that *R. tropici* might originally be the symbiont of other host legumes, e.g., *Mimosa* or *Gliricidia* that later adapted to nodulate the common bean (Menna et al. 2006; Pinto et al. 2007). However, the results might also emphasize the highly promiscuous nature of *R. tropici* (e.g., Michiels et al. 1998). Finally, we should mention that the genomes of two other symbionts of the common bean, *R. gallicum* and *R. giardinii*, have not been studied yet; however, the agronomic importance of *R. giardinii* is strict, as the species is non-effective in fixing N_2 with the common bean (Amarger et al. 1997), and *R. gallicum* is rarely observed in surveys performed in Brazil (e.g., Mostasso et al. 2002; Grange and Hungria 2004; Pinto et al. 2007).

Final comments

R. tropici is predominant in the nodules of field-grown common bean plants in Brazil, when both Mesoamerican and Andean cultivars are used as trap hosts (e.g., Hungria et al. 2000, 2003; Mostasso et al. 2002). This trait has been attributed to the intrinsic properties of the species, particularly tolerance of acidic and high-temperature conditions (Martínez-Romero et al. 1991; Graham et al. 1994; Hungria et al. 2000, 2003), resulting in stronger competitiveness under the often stressful tropical conditions in Brazilian soils (Mercante et al. 1998; Hungria et al. 2000, 2003; Mostasso et al. 2002). Another major attribute is the higher symbiotic stability of *R. tropici* in comparison to other common bean rhizobial species, probably due to the presence of a unique copy of the *nifH* gene (Martínez et al. 1985; Quinto et al. 1985; Martínez-Romero et al. 1991; Geniaux et al. 1993). Based on these properties, a strain selection program was established in Brazil in 1994 and PRF 81 has been identified as an elite strain and is now used in commercial inoculants for the common bean crop (Hungria et al. 2000, 2003). PRF 81 is thus a good representative of both the biotechnological potential and also the rhizobial diversity still to be revealed in the tropics. The approach used in this study has resulted in an interesting panoramic view of the genome of PRF 81, at a cost affordable to many laboratories. Most importantly, a broad range of novel genes have been described.

A particularly interesting aspect is the observation that the genome of PRF 81 is comprised of genes showing similarity with more than a hundred microorganisms (considering the functional classification in the KEGG database), none of which, surprisingly, represents more than

19% of the CDS with putative known functions. This mosaic of genes seems to be built to confer saprophytic competence and versatility. Interesting also is the similarity to genes related to nodulation and to superficial structures of *Sinorhizobium* rather than with the major symbiont of the common bean in the genetic centers of origin of the legume, *R. etli*, or with another common bean symbiont, *R. leguminosarum*. These findings might be indicative of two concepts: the first one that *R. tropici* could the symbiont of another indigenous legume that adapted to nodulate the common bean, and the second one reinforcing the very promiscuous nature of this rhizobial species, being capable of nodulating a wide range of legumes.

Acknowledgments Reagents were financed by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, project # 505499/2004-5). The authors thank Daisy R. Binde and Jesiane S.S. Batista (fellows from CNPq # 505946/2004-1) for the laboratory assistance in the sequencing project, and Dr. Fabio O. Pedrosa and Dr. Emanuel M. Souza for scientific advice. F.G.S. Pinto received a Ph.D. fellow from CNPq. L.M.O. Chueire acknowledges a technology fellowship and A.T.R. Vasconcelos and M. Hungria research fellowships from CNPq. F.G. Barcellos is a pos-doc fellow from CAPES. M. Megías acknowledges MEC for the project # AGL2006-13758-C05-01-AGR.

References

Acosta-Durán C, Martínez-Romero E (2002) Diversity of rhizobia from nodules of the leguminous tree *Gliricidia sepium*, a natural host of *Rhizobium tropici*. Arch Microbiol 178:161–164

Aguilar OM, Riva O, Peltzer E (2004) Analysis of *Rhizobium etli* and its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. Proc Natl Acad Sci U S A 101:13548–13553

Almeida LG, Paixão R, Souza RC, Costa GC, Barrientos JA, Santos MT, Almeida DF, Vasconcelos ATR (2004) A system for automated bacterial (genome) integrated annotation—SABIA. Bioinformatics 20:2832–2833

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Amarger N, Macheret V, Laguerre G (1997) *Rhizobium gallicum* sp. nov. Int J Syst Bacteriol 47:996–1006

Andrade DS, Murphy PJ, Giller KE (2002) The diversity of *Phaseolus*-nodulating rhizobial populations is altered by liming of acid soils planted with *Phaseolus vulgaris* L in Brazil. Appl Env Microbiol 68:4025–4034

Cárdenas L, Domínguez J, Santana O, Quinto C (1996) The role of the *nodI* and *nodJ* genes in the transport of Nod metabolites in *Rhizobium etli*. Gene 173:183–187

Dickstein R, Bisseling T, Reinhold VN, Ausubel FM (1988) Expression of nodule-specific genes in alfalfa root nodules blocked at an early stage of development. Genes Dev 2:677–687

Folch-Mallol JL, Marroqui S, Sousa C, Manyani H, López-Lara IM, van der Drift KM, Haverkamp J, Quinto C, Gil-Serrano A, Thomas-Oates J, Spaink HP, Megías M (1996) Characterization of *Rhizobium tropici* CIAT899 nodulation factors: the role of *nodH* and *nodPQ* genes in their sulfation. Mol Plant-Microb Interact 9:151–163

Forsberg LS, Reuhs BL (1997) Structural characterization of the K antigens from *Rhizobium fredii* USDA257: evidence for a common structural motif, with strain-specific variation, in the capsular polysaccharides of *Rhizobium* spp. J Bacteriol 179:5366–5371

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293:668–672

Geniaux E, Laguerre G, Amarger N (1993) Comparison of geographically distant population of *Rhizobium* isolated from root nodules of *Phaseolus vulgaris*. Mol Ecol 2:295–302

Gil-Serrano A, Sanchez del Junco A, Tejero-Mateo P, Megias M, Caviedes MA (1990) Structure of the extracellular polysaccharide secreted by *Rhizobium leguminosarum* var phaseoli CIAT 899. Carbohydr Res 204:103–107

Gil-Serrano A, Franco G, González-Jiménez I, Tejero P, Molina J, Dobado JA, Romero MJ, Megías M (1993) The structure and molecular mechanics calculations of the cyclic-(1-2)-β-D-glucan secreted by *Rhizobium tropici* CIAT899. J Mol Struct 301:211–226

Gil-Serrano AM, González-Jiménez I, Tejero-Mateo P, Megias M, Romero-Vazquez MJ (1994) Analysis of the lipid moiety of lipopolysaccharide from *Rhizobium tropici* CIAT899: identification of 29-hydroxytriacontanoic acid. J Bacteriol 176:2454–2457

Godoy LP, Vasconcelos ATR, Chueire LMO, Souza RC, Nicolás MF, Barcellos FG, Hungria M (2008) Genomic panorama of *Bradyrhizobium japonicum* CPAC 15, a commercial inoculant strain largely established in Brazilian soils and belonging to the same serogroup as USDA 123. Soil Biol Biochem 40:2743–2753

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

Graham PH, Draeger KJ, Ferrey ML, Conroy MJ, Hammer BE, Martínez E, Aarons SR, Quinto C (1994) Acid pH tolerance in strains of *Rhizobium* and *Bradyrhizobium*, and initial studies on the basis for acid tolerance of *Rhizobium tropici* UMR1899. Can J Microbiol 40:198–207

Grange L, Hungria M (2004) Genetic diversity of indigenous common bean (*Phaseolus vulgaris*) rhizobia in two Brazilian ecosystems. Soil Biol Biochem 36:1389–1398

Grange L, Hungria M, Graham PH, Martínez-Romero E (2007) New insights into the origins and evolution of rhizobia that nodulate common bean (*Phaseolus vulgaris*) in Brazil. Soil Biol Biochem 39:867–876

Hungria M, Andrade DS, Chueire LMO, Probanza A, Guitierrez-Manero FJ, Megías M (2000) Isolation and characterization of new efficient and competitive bean (*Phaseolus vulgaris* L) rhizobia from Brazil. Soil Biol Biochem 21:1515–1528

Hungria M, Campo RJ, Mendes IC (2003) Benefits of inoculation of common bean (*Phaseolus vulgaris*) crop with efficient and competitive *Rhizobium tropici* strains. Biol Fertil Soils 39:88–93

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:29–34

Kereszt A, Kiss E, Reuhs BL, Carlson RW, Kondorosi A, Putnoky P (1998) Novel *rkp* gene clusters of *Sinorhizobium meliloti*

involved in capsular polysaccharide production and invasion of the symbiotic nodule: the *rkpK* gene encodes a UDP-glucose dehydrogenase. J Bacteriol 80:5426–5431

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

Lloret L, Martínez-Romero E (2005) Evolución y filogenia de *Rhizobium*. Rev Latinoam Microbiol 47:43–60

López-Lara IM, Geiger O (2001) The nodulation protein NodG shows the enzymatic activity of an 3-oxoacyl-acyl carrier protein reductase. Mol Plant-Microb Interact 14:349–357

Manyani H, Sousa C, Soria Díaz ME, Gil-Serrano A, Megías M (2001) Regulation of nod factor sulphation genes in *Rhizobium tropici* CIAT899. Can J Microbiol 47:574–579

Martínez E, Pardo MA, Palacios R, Cevallos MA (1985) Reiteration of nitrogen fixation gene sequences and specificity of *Rhizobium* in nodulation and nitrogen fixation in *Phaseolus vulgaris*. J Gen Microbiol 131:1779–1786

Martínez-Romero E, Segovia L, Mercante FM, Franco AA, Graham P, Pardo MA (1991) *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L beans and *Leucaena* sp trees. Int J Syst Bacteriol 41:417–426

Menna P, Hungria M, Barcellos FG, Bangel EV, Hess PN, Martínez-Romero E (2006) Molecular phylogeny based on the 16S rRNA gene of elite rhizobial strains used in Brazilian commercial inoculants. Syst Appl Microbiol 29:315–332

Mercante FM, Cunha CO, Straliotto R, Ribeiro-Juinior WQ, Vanderleyden J, Franco AA (1998) *Leucaena leucocephala* as a trap-host for *Rhizobium tropici* strain from the Brazilian Cerrado region. Rev Microbiol 29:49–58

Michiels J, Dombrecht B, Vermeiren N, Xi C, Luyten E, Vanderleyden J (1998) *Phaseolus vulgaris* is a non-selective host for nodulation. FEMS Microbiol Ecol 26:193–205

Mostasso L, Mostasso FL, Dias BG, Vargas MAT, Hungria M (2002) Selection of bean (*Phaseolus vulgaris* L) rhizobial strains for the Brazilian Cerrados. Field Crops Res 73:261–272

Nogales J, Campos R, BenAbdelkhalek H, Olivares J, Carmen L, Sanjuan J (2002) *Rhizobium tropici* genes involved in free-living salt tolerance are required for the establishment of efficient nitrogen-fixing symbiosis with *Phaseolus vulgaris*. Mol Plant-Microb Interact 15:225–232

Parada M, Vinardell JM, Ollero FJ, Hidalgo A, Gutiérrez R, Buendía-Clavería AM, Lei W, Margaret I, López-Baena FJ, Gil-Serrano AM, Rodríguez-Carvajal MA, Moreno J, Ruiz-Sainz JE (2006) *Sinorhizobium fredii* HH103 mutants affected in capsular polysaccharide (KPS) are impaired for nodulation with soybean and *Cajanus cajan*. Mol Plant-Microb Interact 19:43–52

Pinto FGS, Hungria M, Mercante FM (2007) Polyphasic characterization of Brazilian *Rhizobium tropici* strains effective in fixing N_2 with common bean (*Phaseolus vulgaris* L). Soil Biol Biochem 39:1851–1864

Quinto C, Flores M, Leemans J, Cevallos MA, Pardo MA, Azpiroz R, Girard ML, Calva E, Palacios R (1985) Nitrogenase reductase: a functional multigene family in *Rhizobium phaseoli*. Proc Natl Acad Sci U S A 82:1170–1174

Reinhold BB, Chan SY, Reuber TL, Marra A, Walker GC, Reinhold VN (1994) Detailed structural characterization of succinoglycan, the major exopolysaccharide of *Rhizobium meliloti* Rm1021. J Bacteriol 176:1997–2002

Riley M (1993) Functions of the gene products of *Escherichia coli*. Microbiol Rev 57:862–952

Rosteck PR Jr, Reynolds PA, Hershberger CL (1991) Homology between proteins controlling *Streptomyces fradiae* tylosin resistance and ATP-binding transport. Gene 102:27–32

Segovia L, Young JPW, Martínez-Romero E (1993) Reclassification of American *Rhizobium leguminosarum* biovar phaseoli type I strains as *Rhizobium etli* sp nov. Int J Syst Bacteriol 43:374–377

Skorupska A, Janczarek M, Marczak M, Mazur A, Król J (2006) Rhizobial exopolysaccharides: genetic control and symbiotic functions. Microb Cell Fact 5:7

Sousa C, Folch JL, Boloix P, Megías M, Nava N, Quinto C (1993) A *Rhizobium tropici* DNA region carrying the amino-terminal half of a *nodD* gene and a *nod-box*-like sequence confers host-range extension. Mol Microbiol 9:1157–1168

Stanfield SW, Ielpi L, O'Brochta D, Helinski DR, Ditta GS (1988) The *ndvA* gene product of *Rhizobium meliloti* is required for beta-(1→2)glucan production and has homology to the ATP-binding export protein HlyB. J Bacteriol 170:3523–3530

Suzuki K, Iwata K, Yoshida K (2001) Genome analysis of Agrobacterium tumefaciens: construction of physical maps for linear and circular chromosomal DNAs, determination of copy number ratio and mapping of chromosomal virulence genes. DNA Res 8:141–152

Tatusov R, Galperin M, Natale D, Koonin E (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36

Vargas C, Martinez LJ, Megias M, Quinto C (1990) Identification and cloning of nodulation genes and host specificity determinants of the broad host-range *Rhizobium leguminosarum* biovar phaseoli strain CIAT899. Mol Microbiol 4:1899–1910

Viprey V, Rosenthal A, Broughton WK, Perret X (2000) Genetic snapshots of the *Rhizobium* species NGR234 genome. Gen Biol 1:1–17

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

Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson ARJ, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhil J (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol 7:R34