RESEARCH ARTICLE

Towards a two-dimensional proteomic reference map of *Bradyrhizobium japonicum* CPAC 15: Spotlighting "hypothetical proteins"

Jesiane Stefânia da Silva Batista[1,2], *Adalgisa Ribeiro Torres*[1] *and Mariangela Hungria*[1]

[1] Embrapa Soja, Londrina, Paraná, Brazil
[2] Department of Microbiology, Universidade Estadual de Londrina, Londrina, Paraná, Brazil

The economic and ecological importance of the symbiosis of soybean with *Bradyrhizobium japonicum* strains is significant in several countries, particularly Brazil; however, up to now, only one complete and a draft genome for this species are available. In this study, we have obtained a proteomic reference map of *B. japonicum* strain CPAC 15 (= SEMIA 5079) – used in commercial inoculants for application to soybean crops in Brazil – grown under *in vitro* conditions. CPAC 15 belongs to the same serogroup as strain USDA 123, and both are known as the soybean bradyrhizobial strains with highest competitive and saprophytic known so far. To increase the precision of the proteomic map, we compared whole-cell 2-D protein gel-electrophoresis profiles of CPAC 15 and of two related strains. One-hundred and seventy representative spots, selected from the three profiles, were analyzed by MS. In total, 148 spots were successfully identified as cytoplasmic and periplasmic proteins belonging to diverse metabolic pathways, several of them related to the saprophytic and competitive abilities of CPAC 15. We attributed probable functions to 26 hypothetical proteins, including those involved in polyhydroxybutyrate metabolism, β-lactamase, stress responses and aromatic compound degradation, all with high probability of being related to the saprophytic ability of CPAC 15. In addition, by providing valuable information about expressed proteins in *B. japonicum in vitro*, our results emphasize the importance of accurate functional annotation of uncharacterized expressed proteins, improving considerably our understanding of the legume–rhizobia symbiosis.

Received: February 10, 2010
Revised: June 7, 2010
Accepted: June 14, 2010

Keywords:
Bradyrhizobium japonicum / Hypothetical proteins / Microbiology / Reference map

1 Introduction

The ability to fix atmospheric nitrogen (N_2) is widely distributed among prokaryotes, called diazotrophs, some of which, collectively called rhizobia, can induce the formation of specialized organs in the roots of host plants, mostly in the *Leguminosae* family, in which the biological N_2-fixation process takes place. This symbiotic interaction provides carbohydrate to the bacterium and nitrogen to the plant, resulting in high levels of protein in the leaves and grains [1].

In Brazil, the symbiosis in soybean [*Glycine max* (L.) Merr.] induced by inoculating seeds at sowing with elite strains of *Bradyrhizobium japonicum* and *B. elkanii* can fulfill most of the crop's need for N, resulting in estimated savings of more than US$ 6 billion *per* year to the country, in comparison with the cost of the fertilizer N that would be otherwise needed to achieve comparable yields [2]. *B. japonicum* CPAC 15 (= SEMIA 5079) is an outstanding strain that has been successfully and broadly used in Brazilian

Correspondence: Dr. Mariangela Hungria, Embrapa Soja, Cx. Postal 231, 86001-970, Londrina, Paraná, Brazil
E-mail: hungria@cnpso.embrapa.br
Fax: +55-43-33716100

Abreviations: ClpP, Clp protease; **EcfG,** sigma-E factor; **HmgA,** homogentisate 1,2-dioxygenase; **PBP,** periplasmic binding protein; **PHB,** polyhydroxybutyrate; **Pi,** inorganic phosphate

inoculants since 1992 [3]. CPAC 15 belongs to the same serogroup as USDA 123 and is now ubiquitous in Brazilian soils; together with USDA 123, the strain composes the soybean bradyrhizobia group with the highest competitive and saprophytic capacity known so far [2, 4]. A snapshot of the genome of CPAC 15 was recently obtained [5], and even with a coverage of only 13% several putative genes possibly related to competitiveness and saprophytic ability was revealed. In addition, the comparison of CPAC 15 with the only complete genome of *B. japonicum* (strain USDA 110) published so far – on COG and KEGG databases, tRNAs, transposases, G+C content – not only confirmed a successful coverage of the whole genome of CPAC 15, but also indicated that at least 35% of the putative genes of CPAC 15 shows higher similarity to microorganisms other than strain USDA 110 [5].

The establishment of a proteomic reference map can provide both valuable data for including protein-expression information into the genomic annotation process [6, 7] and represent a starting point for comparative analyses, allowing recognition of sets of proteins expressed under distinct conditions [8–10]. However, as pointed before [11], more reliable information can be obtained by the comparative analysis of more than one strain belonging to the same species. In the complete genome of *B. japonicum* strain USDA 110 obtained in 2002, 30.1% of the genes were assigned as hypothetical and 17.1% showed no similarity to any reported gene [12]. Later, the expression of some of the assigned related proteins was confirmed in transcriptomics and proteomic studies [10, 13–16]. Nevertheless, despite the economic importance of soybean, only one additional strain of *B. japonicum* has been partially sequenced so far: CPAC 15.

Although the protocol for bacteroid isolation from nodules has been established by Sarma and Emerich back in 2005, allowing the study of protein extracts from symbiotic growth conditions [10], proteomic studies with bacteria grown *in vitro* are also very important, as they can reveal important mechanisms related to the bacterium growth before establishing the symbiosis, *e.g.* the saprophytic capacity. Therefore, in this study, whole-cell 2-D electophoretic gels were generated for *B. japonicum* CPAC 15 grown under *in vitro* conditions and the profile was confirmed in two other putative variants of the same serogroup, strains S 370 and S 516, previously characterized by our group [3, 17]. The 2-D profiles were combined to provide reliable support for the proteomic reference map of this serogroup of *B. japonicum*.

2 Materials and methods

2.1 Bacterial strains and growth conditions

B. japonicum CPAC 15 is a putative natural variant of strain SEMIA 566, selected for superior symbiotic performance and used in commercial inoculants in Brazil since 1992; more information about this strain is given elsewhere [2]. Strains S 370 and S 516 are also putative variants of SEMIA 566, characterized by contrasting N_2-fixation properties; previous genetic characterizations of these two strains are available elsewhere [3, 17, 18]. The three strains are deposited at the ''Diazotrophic and Plant Growth Promoting Bacteria Culture Collection'' of Embrapa Soja (http://www.bmrc.lncc.br). The strains were precultured in 10 mL aliquots of tryptone–yeast extract medium, at 80 rpm and 28°C, in the dark. The precultures where then transferred to Erlenmeyer flasks containing tryptone–yeast medium and incubated under the same conditions until the exponential phase of growth was reached (optical density at 630 nm of 0.7–0.8).

2.2 Cell growth and preparation of whole-cell extracts of proteins

Cultures were centrifuged at $5000 \times g$, at 4°C and cells were cautiously washed with a solution containing 3 mM KCl, 1.5 mM KH_2PO_4, 68 mM NaCl and 9 mM NaH_2PO_4. Washed cells were resuspended in 600 μL of a buffer containing 10 mM Tris-HCl, pH 8.0, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT and 0.5 mM PMSF. Aliquots of 150 μL were stored in an ultrafreezer (−80°C) until the analyses.

For whole-cell protein extraction, aliquots were resuspended in lysis buffer containing 9.5 M urea, 2% CHAPS, 0.8% v/v Pharmalyte 3–10 and 1% DTT, and submitted to 30 cycles of freezing in liquid N_2 and thawing at 37°C. The lysates were separated from particulate material by centrifugation at $14\,000 \times g$ for 90 min, at 4°C. Total protein concentration was determined by Bradford's method [19].

2.3 2-D electrophoresis

For IEF, lysates were dissolved with DeStreak buffer (GE Healthcare) to a final concentration of 300 μg of protein and 2% v/v IPGphor in 250 μL of solution. IPG-strips (pH 3–10, 13 cm; GE Healthcare) were rehydrated with the protein solution and covered with cover fluid (GE Healthcare). Loaded strips were submitted to focalization in an Ettan IPGphor IEF system (GE Healthcare) for 1 h at 200 V, 1 h at 500 V, a gradient step to 1000 V for 1 h, a gradient step to 8000 V for 2 h 30 min and fixed at 8000 V for 1 h 30 min. The final Vh was fixed at 24 800. After focusing, strips were equilibrated first for 20 min in 5 mL of TE buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS and 0.2% v/v of a 1% solution of bromophenol blue) supplemented with 50 mg DTT and then in TE buffer with 175 mg iodoacetamine, also for 20 min.

2-D electrophoresis was performed on a 12% polyacrylamide gel in a Ruby SE 600 Vertical Electrophoresis System (GE Healthcare). The run was carried out for 30 min at 15 mA/gel and 240 min at 30 mA/gel, using the Low Molecular Weight Calibration Kit for SDS Electrophoresis (Amersham Biosciences) to provide standards. For each strain, the extraction and gel electrophoresis were run in triplicate. Gels were fixed overnight with an ethanol–acetic acid solution before being stained with Coomassie Blue PhastGel6^{TM} R-350 (GE Healthcare) and were scanned (ImageScanner LabScan v5.0) and analyzed with the ImageMaster 2D Platinum v 5.0 software (GE Healthcare) for spot selection.

2.4 Sample preparation for MS

Selected spots were excised and processed as described previously [20]. Digestions were done with trypsin (Gold Mass Spectrometry Grade, Promega, Madison, WI) at 37°C overnight.

2.5 MALDI-TOF/TOF-TOF analysis

Tryptic peptides (0.5 μL) were mixed with a saturated solution of CHCA in 50% ACN, 0.1% TFA. The mixture was spotted onto a MALDI sample plate and allowed to crystallize at room temperature. The same procedure was used for the standard peptide calibration mix (Bruker Daltonics).

For mass spectra acquisition, a MALDI-TOF-MS Autoflex spectrometer (Bruker Daltonics) was used and operated in the reflector for MALDI-TOF PMF in the fully automated mode, or manually in the LIFT mode for MALDI-TOF/TOF, using the FlexControl software.

2.6 Protein identification

PMFs and MS/MS ions were searched against the NCBI nr database using the MASCOT software (Matrix Science). We decided to use NCBI nr/Proteobacteria database and not the comparison with USDA 110 to increase the possibility of identifying new genes. For protein searches, monoisotopic masses were used, considering a peptide tolerance of 150 ppm and allowance of one missed cleavage.

The Decoy Score, given by MASCOT, was considered in the experiment, aiming at improving information about the false discovery rate; however, the Decoy scores were always insignificant. In addition, we have also analyzed the accuracy of theoretical and experimental values of molecular mass and p*I* to eliminate the false positives, despite knowing that, in some cases, post-translational modifications and truncated forms of the proteins may result in slight differences.

When MS/MS was carried out, a tolerance of 0.3 Da was acceptable. Carbamidomethylation of cysteine and oxidation of methionine were taken into consideration as fixed and variable modifications, respectively.

2.7 Protein characterization

A set of bioinformatics tools was used for an improved characterization of identified proteins. The proteins were fitted into COG (clusters of orthologous groups) categories according to their functional inference, using the program COGnitor (http://www.ncbi.nih.gov/COG) [21]. Software packages PSORT-B [22] and PSLpred [23] were used for prediction of subcellular localization, and SignalP [24] for the prediction of signal peptides.

Special attention was given to the peptides identified as ''hypothetical protein.'' For these proteins, we also used MicrobesOnline (http://www.microbesonline.org) [25], a suite of web-based comparative tools, and the Integrated Microbial Genomes system (http://img.jgi.doe.gov) [26]. STRING 8.1 (http://string-db.org/) [27] was used to predict physical and functional protein interactions.

3 Results and discussion

3.1 Comparison between predicted and experimental 2-D electrophoretic patterns of *B. japonicum*

In the complete genome of *B. japonicum* strain USDA 110 (9.1 Mb), 8317 putative genes were predicted, with an average G+C content of 64.1% [12]. In this study, a theoretical 2-D electrophoretic pattern was constructed using the JVirGel 2.0 software [28] based on the USDA 110 genome, excluding secreted and membrane proteins from the visualization. The spot distribution was confirmed with the experimental 2-D gels (Fig. 1). When plotted in relation to their predicted p*I* and MW values, the proteins of CPAC 15 were found to be more representative between p*I* 4–7 and 8–11 (data not shown). In the alkaline zone (p*I* 8–10), the experimental gel showed fewer protein spots in comparison to the virtual gel, confirming that only a portion of the theoretical proteome of an organism can be displayed on a 2-D gel (data not shown). Poor resolution in the alkaline region of 2-D gels has been reported by other authors [29, 30], and has been attributed to protein separation during IEF [31].

In this study, we used IPG strips of pH 3–10 for a broader overview of whole-cell proteins of *B. japonicum* grown under *in vitro* conditions. Despite the wide range of pH, spots displayed good separation resolution, with no further need for IPG strips for the purposes of this study. The three strains characterized (CPAC 15, S 370 and S 516) showed strong spot similarity, emphasizing the reliability of the protein expression profile obtained. However, variability

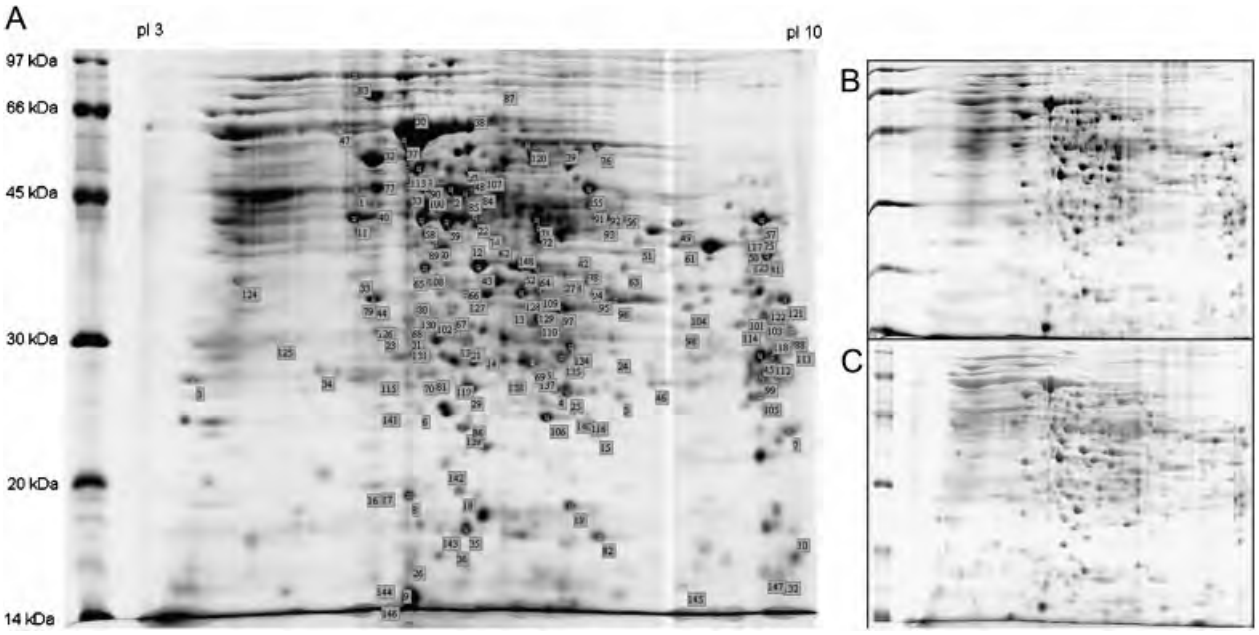

Figure 1. Protein reference map of (A) strain CPAC 15 of *B. japonicum*, (B) S 370 and (C) S 516. Letters correspond to identified proteins are listed in Supporting Information Table 1.

in relative volume of the same spots was observed between the strains, reflecting variations in expression level.

3.2 Spot identification and characterization

Well-defined spots were detected in all nine gels analyzed (Fig. 1). To assess the reproducibility of the proposed reference map, 170 correlated spots were selected from each strain and analyzed by MALDI-TOF-MS or, when necessary, by MALDI-TOF-TOF. Mass spectra of peptide fragments were compared with database entries and spectrometry data sets of 148 spots were submitted to PRIDE (http://ebi.ac.uk/pride/) with the experiment accession number 9769. Supporting Information Table 1 summarizes all identified proteins.

Using ImageMaster 2D Platinum version 5.0, the experimental and theoretical p*I* and MW values were compared. Scatter plots show that the experimental and theoretical values of p*I* and MW were consistent (Fig. 2). One possible reason for the lack of congruence for some proteins in Fig. 2 may rely on the difference between the strains, as USDA 110 was used for the calculation of theoretical p*I* and MW values and the experimental proteins were obtained from CPAC 15. However, the incongruence could also be attributed either to post-translational modifications or to artificial modifications occurring during the preparation of whole-cell lysates, as reported earlier by other authors [30, 32]. Finally, differences in the experimental and theoretical values for MW may be attributed to truncated forms of large proteins, which are generally poorly resolved on 2-D gels, *e.g.* the hybrid sensory histidine kinase protein.

3.3 Protein functional classification

Considering the classification in COG, proteins were distributed in 16 groups, with group E, related to amino acid transport and metabolism, being the most representative, corroborating the results of the genomic panorama of strain CPAC 15 [5]. Furthermore, nine proteins, all in the category of hypothetical function, were not classified in any group, and were assigned as ''not in COG,'' whereas eight other hypothetical proteins were classified in group R (general function prediction only). Discussion of specific COG groups is summarized in Table 1 and Supporting Information Table 1.

3.3.1 Information storage and processing

Highly expressed proteins, such as elongation factor Tu, are members of this class. We identified ribosomal proteins, both from the large (L9, L25 and L5) and from the small (S6 and S7) subunits (Supporting Information Table 1). Genes encoding proteins homologous to L25 (*rplY*) are found only in bacteria and have homology to the general stress protein CTC of *Bacillus subtilis*. Although morphologically normal, *Escherichia coli* mutants lacking the *rplY* gene exhibit slower growth rate in relation to the parental strain, due to a decreased protein-biosynthesizing capacity [33]. In *B. japonicum*, L25 protein possesses an unusual highly repetitive *C*-terminal poly-alanine sequence, a feature once believed to be specific to eukaryotes. In prokaryotes, it has also been observed in *Rhodopseudomonas palustris* [34].

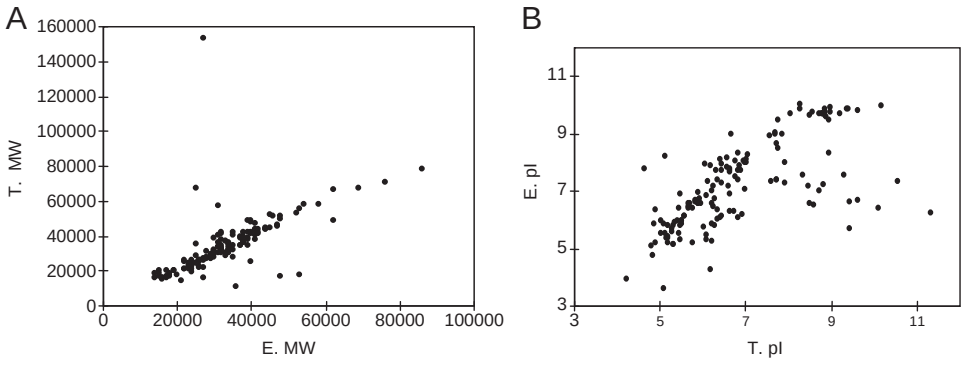

Figure 2. Scatter plots presenting the correlation of theoretical (T) and experimental (E) values of molecular weight (MW) and p*I* of identified proteins of *B. japonicum* whole cells proteome. (A) Molecular weight (MW); (B) p*I*.

We report the expression of a Lys-R family transcriptional regulatory protein (Supporting Information Table 1), which was described previously [9] as bacteroid specific. Proteins related to transcriptional processes are important for mediation of metabolic changes related to stress adaptation in free-living forms.

RNA polymerase σ factor (EcfG) is recognized as essential for bacterial growth at high temperatures in *E. coli* [35]. As recently demonstrated, in *B. japonicum* EcfG is strongly linked to another protein identified in this study, the two-component regulator PhyR [36]. Both regulators are part of the same signaling cascade, involved in responses of free-living forms to stressful conditions and required for efficient symbioses with host plants, by a mechanism still not identified.

3.3.2 Cellular processes and signaling

Strain CPAC 15 is characterized by a high persistence capacity in soils, even under the harsh environmental conditions – high daytime temperatures and long dry periods – of the Brazilian Cerrados [2]. Reference [5] pointed out that several putative genes are consistent with the adaptation capacity of CPAC 15 to a variety of soils, including genes related to signal transduction.

Two-component regulatory systems process various environmental signals, mediating rapid metabolic responses for adaptation to new conditions; they are usually represented by a sensor kinase and a response regulator, adapting to the new environment *via* protein phosphorylation [37]. One two-component regulator identified in our study, PhyR, is highly induced under desiccation stress [38] and was previously described as part of a signaling cascade with EcfG protein. In addition, a hybrid sensory histidine kinase with nitrogen specificity was detected.

Also important to stress adaptation are the proteins related to cell motility. The GTP-binding protein TypA is required for housekeeping functions, for survival under stress conditions and for the symbiosis with certain hosts [39]. Another of these functions is flagella-mediated cell motility. A complex formed between a fumarate reductase protein and a cytoplasmic protein FliG is crucial for both flagellar assembly and rotation function [40] and was expressed in free-living conditions in CPAC 15 (Supporting Information Table 1).

In *B. japonicum*, the mechanisms of transcriptional control of heat shock proteins are remarkably complex. One way is *via* the regulation by the protein HrcA (Supporting Information Table 1), which operates as a transcriptional repressor of the *grpE-dnaK-dnaJ* and *groELS* class I of heat shock operons. Chaperonine GroEL is also known to be involved in nitrogenase formation in *B. japonicum*, and mutants lacking *groEL* – although capable of nodulating – were ineffective in fixing N_2 [41]. Moreover, a correlation between the amount of GroEL synthesized and the rate of N_2 fixation and abundance of nitrogenase genes has been reported.

Stress-induced ATP-dependent Clp protease (ClpP) participates in diverse cell processes, depending on conditions, including motility, biofilm formation, sporulation and saprophytic competence and, in this study, we demonstrated constitutive expression in *B. japonicum* CPAC 15 (Supporting Information Table 1). In *Listeria monocytogenes*, ClpP is known to be involved in rapid adaptative responses of intracellular pathogens during the infection process [42]. All of these functions are well characterized in pathogenic bacteria and ClpP appears to be vital both under optimal and under stress conditions. The absence of *clpP* affected the growth rate of *B. subtilis* cells in the presence of salt, alcohol and heat [43].

Rhizobial extracellular polysaccharides are required for root infection as well as to provide protection in terrestrial environments. Transcriptional analyses have shown that several genes involved in exopolysaccharide biosynthesis are induced under desiccation [38], and one example is *exoN* gene, identified in this study (Supporting Information Table 1). This gene encodes a cytoplasmic UTP-glucose-1-phosphate uridylyltransferase that acts in the production of succinoglycan, an important rhizobial extracellular polysaccharide. Succinoglycan signals from *Sinorhizobium meliloti* on *Medicago truncatula* roots were reported to induce formation of infection threads [44].

Genes conferring resistance to antibiotics are widespread in prokaryotic genomes. β-Lactamase genes have been identified in several members of the order *Rhizobiales*, but their functions remain uncertain since their products

Table 1. Hypothetical proteins identified

Spot ID	NCBI Identifier	Protein description	T. mass	E. mass	T. p*I*	E. p*I*	COG	Cellular location	SignalP	Function prediction
120	gi|27382548	Bll7437[a)]	17 407	53 000	7.93	7.27	I	Cytoplasmic[b)]	0.000	Acyl dehydratase (COG); MaoC like dehydratase (Pfam); oxidoreductase activity (GO); transcriptional regulation of fatty acid biosynthesis
121	gi|27378272	Blr3161[a)]	35 632	34 000	9.39	9.88	S	Periplasmic[b)]	1.000	Uncharacterized protein conserved in bacteria (COG); PBP superfamily (CDD); putative tricarboxylic transport membrane protein (Keeg); *N*-terminal membrane domain, probable signal peptide related (TMHMM)
122	gi|27379700	Bll4589[a)]	29 016	33 000	9.20	9.70	R	Periplasmic[b)]	1.000	Invasion associated locus protein B (COG, Pfam, NCBI Blast)
123	gi|27379032	Blr3921[a)]	34 559	37 000	8.8	9.71	MG	Cytoplasmic	0.000	Nucleoside-diphosphate-sugar epimerases (COG, Pfam); UDP-glucose 4-epimerase (NCBI Blast)
124	gi|27382639	Blr7528[a)]	28 047	35 000	4.22	3.95	S	Outer membrane[b)]	0.000	Uncharacterized protein conserved in bacteria (COG, Pfam, KEEG); putative periplasmic ligand-binding sensor protein (NCBI Blast); potentially secreted *via* a nonclassical pathway (SecretomeP)
125	gi|27379478	Bll4367[a)] [47][c)]	35 149	34 000	6.17	4.26	R	Cytoplasmic[b)]	1.000	Zn-dependent hydrolases (COG); Metallo-β-lactamase superfamily (Pfam); twin-arginine translocation pathway signal sequence (TatP)
126	gi|148240988	BBta_p0149 (*Bradyrhizobium* sp. BTAi1)	32 597	32 000	6.02	5.75	R	Cytoplasmic[b)]	0.000	Predicted Rossmann fold nucleotide-binding protein (COG); Possible lysine decarboxylase (Pfam, CDD); 3-isopropylmalate dehydrogenase (NCBI Blast)
127	gi|27377849	Bll2738[a)]	14 096	21 000	11.33	6.25		Cytoplasmic[b)]	0.001	No related data
128	gi|27375956	Blr0845[a)] [10, 47]	34 939	32 000	8.46	7.19		Periplasmic[b)]	1.000	Uncharacterized protein conserved in bacteria (COG);PBP superfamily (CDD)
129	gi|27377621	Blr2510[a)]	33 487	31 000	6.12	7.31	R	Cytoplasmic[b)]	0.000	Conserved predicted metalloprotease (COG, Pfam, NCBI Blast); protein association with moeB (STRING)
130	gi|27379198	Blr4087[a)] [47]	32 180	30 000	5.41	5.95	S	Cytoplasmic	0.002	Uncharacterized stress-induced protein (COG); YicC-like family (Pfam)
131	gi|27375728	Blr0617[a)]	26 660	28 000	6.29	5.82	M	Periplasmic[b)]	0.996	Outer membrane lipoprotein-sorting protein (COG, Pfam, CDD); lipoprotein signal peptide (LipoP)
132	gi|27377585	Blr2474[a)] [16]	17 060	15 000	9.62	9.83	M	Cytoplasmic[b)]	1.000	Identical to Bll5191, protein containing fasciclin-like repeats (COG. Pfam, CDD); homology with other symbiotically induced proteins (NCBI Blast); *N*-terminal membrane domain, probable signal peptide related (TMHMM)
133	gi|27377976	Blr2865[a)] [47, 63]	28 090	28 000	6.25	6.46	J	Cytoplasmic[b)]	0.014	Methionyl-tRNA synthetase (COGnitor)
134	gi|27377483	Blr2372[a)]	27 323	27 000	6.3	7.74	H	Cytoplasmic	0.000	Ubiquinone biosynthesis *O*-methyltransferase (CDD)
135	gi|27379679	Blr4568[a)]	27 070	27 000	6.83	7.70	R	Cytoplasmic[b)]	0.000	Homolog of lactam utilization proteins of the LamB/YcsF family (COG, Pfam); UPF0271 protein (NCBI Blast)
136	gi|27381044	Blr5933[a)]	25 547	26 000	6.60	7.19	Q	Cytoplasmic[b)]	0.001	E.C. 3.-.-.- ; fumarylacetoacetate (FAA) hydrolase, cathecol pathway (COG, Pfam), potentially secreted by a nonclassical pathway (SecretomeP)

Table 1. Continued

Spot ID	NCBI Identifier	Protein description	T. mass	E. mass	T. p*I*	E. p*I*	COG	Cellular location	SignalP	Function prediction
137	gil27377378	Blr2267[a)]	29 017	25 000	10.54	7.34	T	Cytoplasmic[b)]	0.000	Putative gualylate cyclase, family III protein, (COG, CDD)
138	gil27377598	Blr2487[a)] [63]	24 450	25 000	6.22	6.98	Q	Cytoplasmic[b)]	0.000	EC:3.1.1.45, carboxymethylenebutenolidase (NCBI Blast); γ-hexachlorocyclohexane, fluorobenzoate and 1,4-dichlorobenzene degradation (KEEG); dienelactone hydrolase family (COG, Pfam); membrane associated (TMHMM)
139	gil27375338	Blr0227[a)] [10, 63]	22 619	24 000	5.17	5.35	S	Periplasmic[b)]	0.000	Polyhydroxyalkanoate synthesis repressor PhaR-like (NCBI Blast, Pfam); association with several proteins related to PHB metabolism (STRING); uncharacterized protein conserved in bacteria (COG); potentially secreted by a nonclassical pathway (SecretomeP)
140	gil27378905	Bll3794[a)] [47]	22 406	23 000	6.85	7.91	R	Cytoplasmic[b)]	0.003	Predicted Rossmann fold nucleotide-binding protein (COG); possible lysine decarboxylase (Pfam)
141	gil27377608	Bll2497[a)]	25 498	22 000	8.60	6.50	O	Periplasmic[b)]	0.930	Protein-disulfide isomerase(COG); thioredoxin like domain (CDD); *N*-terminal membrane domain, probable signal peptide related (TMHMM)
142	gil27381760	Bll6649[a)]	18 033	20 000	6.74	6.28		Periplasmic[b)]	1.000	PRC-barrel like (Pfam, NCBI Blast); two membrane domain (TMHMM)
143	gil27379818	Bll4707[a)]	19 884	17 000	6.62	6.29		Cytoplasmic[b)]	1.000	Polyketide cyclase/dehydrase and lipid transport (Pfam, CDD); *N*-terminal membrane domain, probable signal peptide related (TMHMM)
144	gil27376280	Bll1169[a)]	18 872	15 000	6.10	5.33	R	Cytoplasmic[b)]	1.000	Invasion-associated locus B (IalB) protein (COG, Pfam, CDD, NCBI Blast); membrane-associated domain (TMHMM)
145	gil27380662	Bll5551[a)]	15 694	14 000	7.85	8.98	R	Cytoplasmic[b)]	0.000	Predicted signal transduction protein containing cAMP-binding and CBS domains (COG, Pfam, NCBI Blast)
146	gil27376280	Bll1169[a)]	18 872	14 000	6.10	5.45		Cytoplasmic[b)]	1.000	Invasion-associated locus B (IalB) protein (COG, Pfam, CDD, NCBI Blast); membrane-associated domain (TMHMM)
147	gil27382480	Bll7369[a)]	16 350	27 000	8.73	6.99	S	Periplasmic[b)]	0.000	GatB/Yqey domain protein (NCBI Blast, Pfam); Uncharacterized conserved protein (COG)
148	gil27383064	Bll7953[a)] [10]	36 214	38 000	6.27	6.74	EM	Cytoplasmic	0.000	E.C. 4.2.1.52, dihydrodipicolinate synthase (COG, KEEG, NCBI Blast, CDD)

Matched peptides masses and MS/MS combined results are available in PRIDE (http://ebi.ac.uk/pride/) under the experiment accession number 9769. With a combination of data based on sequence, structure, the presence of conserved domains and motifs, protein–protein interactions, physicochemical properties and cellular location; a functional prediction was done.

a) Best-hit with *B. japonicum* strain USDA 110. When best hit corresponded to other species, the name of the species is shown in brackets.

b) Unable to predict in PSORT or contrasting results.

c) Reference of the proteomic study which has described the protein.

exhibited no or low affinities with β-lactam substrates [45]. We report here the expression of a β-lactamase precursor, the sequence of which shows similarity with class D β-lactamase OXA-5 (Supporting Information Table 1).

3.3.3 Energy production and carbohydrate metabolism

It is remarkable that rhizobia show a variety of metabolic traits, consistent with their adaptation to various soil environments and ability to function within root nodules. The tricarboxylic acid (TCA) cycle is key, since C_4 dicarboxylates represent the main carbon source for *B. japonicum* bacteroids [46] and a set of proteins in our data set is consistent with signal transduction of histidine kinase regulating C4-dicarboxylate transport, malate, pyruvate and succinate dehydrogenases (Supporting Information Table 1). The TCA cycle is also important for producing precursors for the biosynthesis of amino acids, purines, pyrimidines and vitamins.

We have identified the α and β subunits of the heterodimer electron transfer flavoproteins *etfL* and *etfS* (Supporting Information Table 1). These proteins are constitutively expressed during aerobic growth, acting as specific electron acceptors for various dehydrogenases. As expected, members of the other set of electron transfer flavoproteins (*fixA* and *fixB*) were not present in our data set, since they are expressed only in bacteroids, under microaerobic conditions [47].

3.3.4 Amino acid, nucleotide and fatty acid metabolism

The metabolism of amino acids and nucleotides by rhizobia may play important roles at various stages of the symbiotic interaction, since auxotrophic mutants can be impaired in nodulation and/or N_2 fixation abilities. In some cases, such failure cannot be recovered merely by supplementation with the nutrient related to the missing gene, indicating the importance not only of individual molecules but also of biosynthetic pathways as a whole [48].

Arginine biosynthesis genes seem to be differently required according to the rhizobial species, which may be related to its ability to adjust to the environmental changes that occur during nodulation [49]. We identified the expression of two genes related to arginine in CPAC 15, *argG* and *argB* (Supporting Information Table 1).

Expression of *aatA* was detected in CPAC 15 (Supporting Information Table 1), and aspartate aminotransferase is a central transaminase in most organisms, acting in the reversible conversion of aspartate plus α-ketoglutarate to glutamate plus oxalacetate. This reaction represents an important link between the metabolisms of C and N, playing a significant role in the metabolic interconversions that occur both in the plant and in the rhizobia. Mutants lacking *aatA* gene were Fix^- in *Rhizobium leguminosarum* [50] and in *S. meliloti* [51]. Similar results were obtained with *B. japonicum* by other authors [48], although – in contrast with the results of [52] who verified the occurrence of N_2 fixation – at a low level.

Nucleotide requirement of the bacteroids is low and downregulated, whereas high levels are observed in free-living rhizobia [53], and in our sample set we identified proteins related both with purine and with pyrimidine biosynthesis.

As summarized in Supporting Information Table 1, proteins related to lipid metabolism (COG group I) represent an important functional class in free living of *B. japonicum* strain CPAC 15. The majority of these proteins are elements of fatty-acid biosynthesis pathways, but they are also involved in the metabolism of isoprenoids (*dxr* and *atoB*), benzoate and butanoate (*hbdA*).

3.3.5 Coenzyme, inorganic ion and secondary compound metabolism

Vitamin limitation influences rhizobial growth in the rhizosphere, which can impair nodule formation [54]. Proteins related to folate biosynthesis were among the most representative, as observed by other authors [47] in *B. japonicum* growing *in vitro*. In our study, we identified a bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase encoded by the *folD* gene. We also identified a thiamine biosynthesis protein (ThiC) and a cobalt insertion protein related to cobalamin biosynthesis (Supporting Information Table 1).

Inorganic phosphate (Pi) has several key roles in cells. In some Gram-negative bacteria, the transport and metabolism of Pi and other P-containing compounds is regulated at the transcriptional level by a two-component system. Gene *pstS* encodes a periplasmic Pi-binding component of the high-affinity ABC-type phosphate-uptake system and its expression was detected in CPAC 15 (Supporting Information Table 1). Because of its high affinity and unusually high speed, this system is rapidly induced when cells are starved for phosphate; however, nodulation was strongly negatively affected when *S. meliloti* lacked this system [55].

Organisms living under aerobic conditions generate ROS resulting from the partial reduction of molecular oxygen. As ROS can adversely affect lipids, nucleic acids and proteins, bacteria must activate responses to oxidative stress. Superoxide dismutases represent the first line of defense against ROS, and are especially important during desiccation stress [38]. *B. japonicum* USDA 110 contains four putative carbonic anhydrases, zinc-regulated metalloenzymes that catalyze the reverse hydration of CO_2 to bicarbonate and are involved in both N metabolism and cyanate hydrolysis; cyanate is a toxic compound generated from metabolites such as urea and carbamoyl phosphate. One of these putative genes encoding

a β-type carbonic anhydrase is induced under chemoautotrophic conditions in *B. japonicum* [56]. The confirmation of the expression of this putative protein in CPAC 15 supports the xenobiotic degradation ability of *B. japonicum*, as well as shows potential as a biotechnological tool, *e.g.* for the degradation of cyanate [57].

Aromatic compounds, important sources of N and C for some soil microorganisms, accumulate in soil as a result of degradation of plant-derived compounds (lignin in particular), or as pollutants. Many aromatic compounds can be degraded by bacteria *via* catechol pathways. As previously reported, and in contrast to other rhizobial genera, these genes seem to be constitutively expressed in *Bradyrhizobium* [58]. In this study, we identified the expression of 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (Supporting Information Table 1), an enzyme in the catechol pathway capable of degrading several aromatic compounds.

Nitrogen limitation, or growth on less commonly available nitrogenous compounds, requires the synthesis of specific enzymes and one example is homogentisate 1,2-dioxygenase (HmgA, Supporting Information Table 1), which is involved in the metabolism of both tyrosine and phenylalanine, and is known to be upregulated under starvation conditions in *S. meliloti* [59].

3.3.6 Other proteins with general function prediction

Competence damage-associated protein CinA, specially required for the process of natural cell transformation and probably related to the molybdopterin-biosynthesis enzyme, is another gene product reported to be specifically expressed in bacteroids [9] that we detected in free-living conditions (Supporting Information Table 1).

Tryptophan repressor-binding protein A, WrbA, is a flavodoxin-like protein that regulates the synthesis of this amino acid, especially during the stationary phase of growth (Supporting Information Table 1). This protein was also found to be downregulated in *S. medicae* under low pH conditions [60].

3.4 Cellular localization and signal peptide prediction

In proteomic studies, the protein extraction procedure and the solution agents are adopted in order to obtain specific fractions of proteins and, in our study, allowed the detection of cytoplasmic and periplasmic proteins, since *B. japonicum* is a Gram-negative bacterium. As already mentioned, two different programs were used for subcellular location: PSORT-B and PSLpred. For several proteins, PSORT-B failed to reveal a cellular location, whereas PSLpred effectively provided the cellular location and a confidence score in most cases.

The proteins were mostly located in the cytoplasm. Based on PSLpred data, hypothetical protein Blr7528 was presumably located on the outer membrane (Table 1); however, this protein possesses high similarity with a putative periplasmic ligand-binding sensor protein of *R. palustris* and the confidence index of PSLpred was low. Still, in the PSLpred data, four other spots were defined as inner-membrane proteins. For prediction of transmembrane domains, we used TMHMM software. Only one spot, corresponding to a sensor protein component of the C4-dicarboxylate-transport system had proper transmembrane domains detected. It is important to consider that subcellular locations must be carefully analyzed because the presence of hydrophobic helices and cores may lead to questionable predictions, such as one of the ''inner-membrane proteins'' actually being the 30S ribosomal protein S7.

Signal peptides are those composed of short amino acid sequences that direct protein translocation to the proper cellular or extracellular location [24]. In Gram-negative bacteria, proteins that pass through the inner membrane to the periplasmic space or outer membrane have their signal peptides removed by specialized signal peptidases; therefore, these proteins contain a cleavage site at the *C*-terminal end. Although the PSORT report gives this information, we have also used the SignalP program, which additionally gives a score for each signal peptide. Some periplasmic proteins, such as succinate dehydrogenase flavoprotein subunit A and an immunogenic protein precursor, exhibited a negative classification, implying that these proteins are membrane associated or secreted by a nonclassical or leaderless pathway.

Five proteins identified as hypothetical had positive score results for the presence of signal peptide (Table 1); however, based on the subcelullar location prediction tools they are probably located in the cytoplasm.

3.5 How hypothetical are the hypothetical proteins?

A relevant portion of the annotated sequences has been classified in ''hypothetical,'' ''conserved hypothetical'' or ''unknown function protein'' categories in several genomes. These denominations are used when the existence of a gene is supported only by prediction of gene-finding software, and when they do not show significant homology to any characterized gene. However, the detection of proteins in a 2-D gel implies their removal from the ''hypothetical'' denomination [61]. As highlighted by other authors [62], the lack of homology of these sequences to those of known proteins, combined with their detection, suggests that they may be interesting subjects for further study, possibly providing helpful, new information and deeper understanding of these organisms.

In total, we have identified 29 spots corresponding to 28 different hypothetical proteins and, as expected, most fit into

COG functional groups R (general function prediction only) and S (function unknown) (Table 1). Functional assignment of these proteins was performed by the utilization of several bioinformatics tools, based on the study by other authors [61].

Blr2865 was classified in group J, which corresponds to translation, ribosomal structure and biogenesis, whose expression was previously reported by the specific expression of this protein in *B. japonicum* under acidic conditions [63]. A signal transduction-related protein, Blr2267, was characterized as a putative adenylate cyclase, class III protein, cytoplasm located and with an internal helix, showing a probable membrane-adjacent domain.

Three out of the four proteins functionally attributed to cell wall/membrane/envelope biogenesis (group M) were classified as hypothetical. Blr3921 product is a nucleoside diphosphate sugar epimerase, probably a UDP-glucose-4-epimerase, which operates in carbohydrate metabolism and in biosynthesis of membrane and extracellular components (exopolysaccharides and lipopolysaccharides). Blr0617 is a periplasmic-space protein, bounded to the outer membrane and related to the shuttle of lipoproteins, similar to lipoprotein localization factors (Lol). Blr2474 was found to be downregulated in bacteroids, in comparison with aerobically grown cells [64]. This fasciclin-like protein was very similar to the reported symbiotically induced *S. meliloti* protein Nex18 and was identified in the secretome of *B. japonicum* USDA 110 [16, 65]. In fact, we further characterized this protein (with the TMHMM 1.0 program) as cytoplasm located with an *N*-terminal transmembrane domain that corresponds to a predicted signal peptide. In complete genome of *B. japonicum*, there are two identical copies of this protein, the other one being Bll5191.

Bll2497 was identified as a disulfite isomerase periplasmic protein with a twin-arginine translocation signal domain, possessing sequence similarity to DstA. This protein is implicated in folding and stability of exported proteins and cell–surface structures, including flagellae and is known for its strong oxidizing activity [66]; none of the annotated genes described in [12], of the *B. japonicum* USDA 110 genome, was assigned as *dstA*. The protein is well characterized in *E. coli* and its absence affects several outer membrane components, leading to a pleiotropic effect [67].

The expression of the hypothetical protein Bll7953 was reported previously [10], which belongs to both E (amino acid transport and metabolis) and M (cell wall, membrane and envelope biogenesis) COG groups. This cytoplasmic protein is a dihydrodipicolinate synthase involved in aspartate–lysine conversion. The verification that the concentration of lysine in *B. japonicum* bacteroids is much higher than in the nodule cytosol, in addition to the substantial number of enzymes involved in lysine biosynthesis expressed in this condition, may help to clarify the process of exchange between the host plant and the bacteroids. Blr2372 has an UbiG-conserved domain, which represents a putative *O*-methyltransferase implicated in the ubiquinone biosynthesis pathway in bacteria.

Two hypothetical proteins were predicted to be involved in secondary metabolite pathways involved in the degradation of xenobiotics. Blr5933 is a member of the fumarylacetoacetate hydrolase family and potentially involved in the catecol pathway, possessing a functional analogy to *hmgB* gene. As already commented, another protein belonging to this pathway, HmgA, was identified in CPAC 15; however, there are no annotated sequences of *hmgBC* in *B. japonicum*. The other protein, Blr2487, is a carboxymethylenebutenolidase, related to degradation of xenobiotic compounds such as the organic insecticide γ-hexachlorocyclohexane, fluorobenzoate and 1,4-dichlorobenzene. This protein was also found to be upregulated under acidic conditions [63]. Strain USDA 110 possesses two other copies of putative carboxymethylenebutenolidases (Bll0837 and Bll6841).

Five of the hypothetical proteins were allocated in the R group, belonging to the proteins that have only general functions predicted (Table 1). Bll4367, a metal-dependent hydrolase of the β-lactamase superfamily with a twin arginine translocation (Tat) signal sequence, was also transcriptionally identified and upregulated in response to iron limitation in *B. japonicum* [68]. This pathway represents one mechanism of translocation of β-lactams to the periplasm of Gram-negative bacteria. Bll5551 has a predicted CBS domain, characteristic of signal transduction proteins.

The predicted metalloprotease Blr2510 (Table 1) belongs to the same chromosomal cluster of molybdopterin biosynthesis protein B (*moeB*), also identified in our study (Supporting Information Table 1). Using STRING 8.1 software, we found that the physical association between *moeB* and Blr2510 homologues is conserved among strains of *Nitrobacter* and *R. palustris* (data not shown). Furthermore, analyzing the homology of both sequences in different organisms, we found that ''hypothetical protein'' Blr2510 is more conserved than MoeB (data not shown).

As most of the hypothetical sequences were not assigned into a COG functional group, we combined tools to predict their functionality. Blr3161 was found to belong to the periplasmic binding protein (PBP) superfamily, characterized by their high affinity and specificity and to have an ABC-transporter auxiliary domain. This protein was also shown to be differentially expressed depending on oxygen pressure, and to belong to the RegR regulon, the same regulatory system of *fixR-nifA* [69]. Blr7528, Bll7369 and Blr4087 are all uncharacterized proteins (Table 1) that have homologues in other genomes.

Invasion-associated locus B (LalB) protein is the major virulence factor of *Bartonella bacilliformis*, implicated in the invasive phenotype of this erythrocyte parasite. Despite this specific function, several species have putative *IalB* genes, *e.g. B. japonicum* Bll4589 and Bll1169, both identified in our

study. The similarities between bacterial plant and animal pathogens with symbionts have been emphasized with the advent of genomics and the identification of extensive gene synteny [70].

The expression of the hypothetical protein Blr0227 has also been demonstrated in bacteroids [10] and specifically in acid conditions *in vitro* [63]. The sequence possesses a polyhydroxybutyrate (PHB) accumulation regulatory domain with a high degree of similarity with PhaR protein (polyhydroxyalkanoate synthesis repressor) and, in STRING data, it is correlated with some other PHB-related proteins. In *B. japonicum* and other rhizobia, much of the C is directed to the synthesis of storage compounds, especially PHB and glycogen. This system is not fully understood, but it may be relevant for nodule-formation and bacteroid-differentiation processes [71].

Detection of the periplasmic protein Bll6649 was also previously reported in transcriptomic studies in *B. japonicum*, first, in mature nodules of soybean [64] and more recently as belonging to the PhyR-EcfG regulon, a novel signaling cascade involved in general stress response and symbiotic efficiency [36].

4 Concluding remarks

Despite the economic and ecological importance of symbiotic diazotrophic bacteria, when compared with pathogenic bacteria, very few rhizobial genomes are available so far. Furthermore, to assess biological implications of a sequenced genome, accurate identification of the protein-coding genes is required [7]. Proteomic techniques are thus used to analyze the global content of proteins of a sample and can be consider as a new set of information that complements genomic studies, being extremely valuable for the genome-annotation process [6, 7, 72].

References [12] and [5] have both highlighted a large number of putative genes with still unknown functions in the genome of *B. japonicum*. The results from our study provide important new information about the proteome of *B. japonicum* strain CPAC 15, greatly contributing to understanding the main molecular forces contributing to the soil persistence and tolerance of harsh environmental conditions of this highly competitive strain established in more than 22 million hectares in Brazil and also in soils of other South American countries [2, 17, 73]. However, it should be mentioned that in our study the cells were grown in a culture medium under laboratory conditions; therefore, the next step of our investigation should be the confirmation of proteins related to the saprophytic capacity, competitiveness and tolerance to environmental stress in a soil environment.

Matched peptide masses and MS/MS combined results are available in PRIDE (http://ebi.acuk/pride/) under the experiment accession number 9769.

The work was partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil)/ MCT/MAPA (577933/2008), CPNq-Universal (470162/2009) and CNPq-GenoSoja (552735/2007-8). MALDITOF was acquired with resources from Fundação Araucária, in a common project coordinated by Federal University of Paraná. Authors thank Dr. Luciano Huergo, Emanuel M. Souza and Fábio Pedrosa for help and suggestions in the proteome analysis. J. S. S. Batista received a Ph.D. fellowship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil) and A. Torres a fellowship from CNPq. M. Hungria is also a research fellow from CNPq. The authors thank Ligia M. O. Chueire for help in several steps of this work and Dr. Allan R. J. Eaglesham for suggestions in the manuscript.

The authors have declared no conflict of interest.

5 References

[1] Lei, Z., Elmer, A. M., Watson, B. S., Dixon, R. A. *et al.*, A Two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume *Medicago truncatula*. *Mol. Cell. Proteomics* 2005, *4*, 1812–1825.

[2] Hungria, M., Campo, R. J., Mendes, I. C., Graham, P. H., Contribution of biological nitrogen fixation to the N nutrition of grain crops in the tropics: the success of soybean (*Glycine max* (L.) Merr.) in South America. in: Singh, R. P., Shankar, N., Jaiwal, P. K. (Eds.), *Nitrogen Nutrition and Sustainable Plant Productivity*, Studium Press, LLC, Houston 2006, pp. 43–93.

[3] Hungria, M., Boddey, L. H., Santos, M. A., Vargas, M. A. T., Nitrogen fixation capacity and nodule occupancy by *Bradyrhizobium japonicum* and *B. elkanii* strains. *Biol. Fertil. Soils* 1998, *27*, 393–399.

[4] Ferreira, M. C., Hungria, M., Recovery of soybean inoculant strains from uncropped soils in Brazil. *Field Crops Res.* 2002, *79*, 139–152.

[5] Godoy, L. P., Vasconcelos, A. T. R., Chueire, L. M. O., Souza, R. C. *et al.*, Genomic panorama of *Bradyrhizobium japonicum* CPAC 15, a commercial inoculant strain largely established in Brazillian soils and belonging to the same serogroup as USDA 123. *Soil Biol. Biochem.* 2008, *40*, 2743–2753.

[6] Gupta, N., Tanner, S., Jaitly, N., Adkins, J. N. *et al.*, Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res.* 2007, *17*, 1362–1377.

[7] Ansong, C., Purvine, S. O., Adkins, J. N., Lipton, M. S., Smith, R. D., Proteogenomics: needs and roles to be filled by proteomics in genome annotation. *Brief. Funct. Genomics Proteomics* 2008, *7*, 50–62.

[8] Münchbach, M., Dainese, P., Staudenmann, W., Naberhaus, F., James, P., Proteome analysis of heat shock protein

expression in *Bradyrhizobium japonicum*. *Eur. J. Biochem.* 1999, *264*, 39–48.

[9] Hoa, L. T., Nomura, M., Tajima, S., Characterization of bacteroid proteins in soybean nodules formed with *Bradyrhizobium japonicum* USDA 110. *Microbes Environ.* 2004, *19*, 71–75.

[10] Sarma, A. D., Emerich, D. W., Global protein expression pattern of *Bradyrhizobium japonicum* bacteroids: a prelude to functional proteomics. *Proteomics* 2005, *5*, 4170–4184.

[11] Mathesius, U., Imin, N., Chen, H., Djordjevic, M. A. *et al.*, Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. *Proteomics* 2002, *2*, 1288–1303.

[12] Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K. *et al.*, Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum*. *DNA Res.* 2002, *9*, 189–197.

[13] Süß, C., Hempel, J., Zehner, S., Krause, A. *et al.*, Identification of genistein-inducible and type III-secreted proteins of *Bradyrhizobium japonicum*. *J. Biotechnol.* 2006, *126*, 69–77.

[14] Lang, K., Lindemann, A., Hauser, F., Göttfert, M., The genistein stimulon of *Bradyrhizobium japonicum*. *Mol. Genet. Genomics* 2008, *279*, 203–211.

[15] Wei, M., Yokoyama, T., Minamisawa, K., Mitsui, H. *et al.*, Soybean seed extracts preferentially express genomic loci of *Bradyrhizobium japonicum* in the initial interaction with soybean, *Glycine max* (L.) Merr. *DNA Res.* 2008, *15*, 201–214.

[16] Hempel, J., Zehner, S., Göttfert, M., Patschkowski, T., Analysis of the secretome of the soybean *Bradyrhizobium japonicum*. *J. Biotechnol.* 2009, *140*, 51–58.

[17] Barcellos, F. G., Menna, P., Batista, J. S. S., Hungria, M., Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium* (*Ensifer*) *fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil. *Appl. Environ. Microbiol.* 2007, *73*, 2635–2643.

[18] Barcellos, F. G., Batista, J. S. S., Menna, P., Hungria, M., Genetic differences between *Bradyrhizobium japonicum* variant strains contrasting in N2-fixation efficiency revealed by representational difference analysis. *Arch. Microbiol.* 2009, *191*, 113–122.

[19] Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* 1976, *72*, 248–254.

[20] Chaves, D. F. S., Souza, E. M., Monteiro, R. A., Pedrosa, F. O., A two-dimensional electrophoretic profile of the proteins secreted by *Herbaspirillum seropedicae* strain Z78. *J. Proteomics* 2009, *73*, 50–56.

[21] Tatusov, R. L., Galperin, M., Natale, D. A., Koonin, E. V., The COG database: a tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000, *28*, 33–36.

[22] Gardy, J. L., Laird, M. R., Chen, F., Rey, S. *et al.*, PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* 2005, *21*, 617–623.

[23] Bhasin, M., Garg, A., Raghava, G. P. S., PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics* 2005, *21*, 2522–2524.

[24] Bendtsen, J. D., Nielsen, H., von Heijne, G., Brunak, S., Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 2004, *340*, 783–795.

[25] Alm, E. J., Huang, K. H., Price, M. N., Koche, R. P. *et al.*, The MicrobesOnline Web site for comparative genomics. *Genome Res.* 2005, *15*, 1015–1022.

[26] Markowitz, V. M., Korzeniewski, F., Palaniappan, K., Szeto, E. *et al.,* The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* 2006, *34*, D344–D348.

[27] Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S. *et al.*, STRING 8 – a global view on protein and their function interactions in 630 organisms. *Nucleic Acids Res.* 2009, *37*, D412–D416.

[28] Hiller, K., Grote, A., Maneck, M., Münch, R., Jahn, D., JVirGel 2.0: computational prediction of proteomes separated via two-dimensional gel electrophoresis under consideration of membrane and secreted proteins. *Bioinformatics* 2006, *22*, 2441–2443.

[29] Eymann, C., Dreisbach, A., Albrecht, D., Bernhardt, J. *et al.*, A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* 2004, *4*, 2849–2876.

[30] Jing, H. B., Yuan, J., Wang, J., Yuan, Y. *et al.*, Proteome analysis of *Streptococcus suis* serotype 2. *Proteomics* 2008, *8*, 333–349.

[31] Pennington, K., McGregor, E., Beasley, C. L., Everall, I. *et al.*, Optimization of the first dimension for separation by two-dimensional gel electrophoresis of basic proteins from human brain tissue. *Proteomics* 2004, *4*, 27–30.

[32] Kohler, C., Wolf, S., Albrecht, D., Fuchs, S. *et al.*, Proteome analyses of *Staphylococcus aureus* in growing and non-growing cells: a physiological approach. *Int. J. Med. Microbiol.* 2005, *295*, 547–565.

[33] Korepanov, A. P., Gongadze, G. M., Garber, M. B., Court, D. L., Bubunenko, M. G., Importance of the 5 S rRNA-binding ribosomal proteins for cell viability and translation in *Escherichia coli*. *J. Mol. Biol.* 2007, *366*, 1199–1208.

[34] Strader, M. B., Verberkmoes, N. C., Tabb, D. L., Connelly, H. M. *et al.*, Characterization of the 70S ribosome from *Rhodopseudomonas palustris* using an integrated ''top-down'' and ''bottom-up'' mass spectrometric approach. *J. Proteome Res.* 2004, *3*, 965–978.

[35] Hiratsu, K., Amemura, M., Nashimoto, H., Shinagawa, H., Makino, K., The *rpoE* gene of *Escherichia coli*, which encodes sigma E, is essential for bacterial growth at high temperature. *J. Bacteriol.* 1995, *177*, 2918–2922.

[36] Gourion, B., Sulser, S., Frunzke, J., Francez-Charlot, A. *et al.*, The PhyR-σEcfG signalling cascade is involved in stress response and symbiotic efficiency in *Bradyrhizobium japonicum*. *Mol. Microbiol.* 2009, *73*, 291–305.

[37] Alex, L. A., Simon, M. I., Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. *Trends Genet.* 1994, *10*, 133–138.

[38] Cytryn, E. J., Sangurdekar, D. P., Streeter, J. G., Franck, W. L. *et al.*, Transcriptional and physiological responses of *Bradyrhizobium japonicum* to desiccation-induced stress. *J. Bacteriol.* 2007, *189*, 6751–6762.

[39] Kiss, E., Huguet, T., Poinsot, V., Batut, J., The *typA* gene is required for stress adaptation as well as for symbiosis of *Sinorhizobium meliloti* 1021 with certain *Medicago truncatula* lines. *Mol. Plant Microbe Interact.* 2004, *17*, 235–244.

[40] Cohen-Ben-Lulu, G. N., Francis, N. R., Shimoni, E., Noy, D. *et al.*, The bacterial flagellar switch complex getting more complex. *EMBO J.* 2008, *27*, 1134–1144.

[41] Fischer, H. M., Schneider, K., Babst, M., Hennecke, H., GroEL chaperonins are required for the formation of a functional nitrogenase in *Bradyrhizobium japonicum* . *Arch. Microbiol.* 1999, *171*, 279–289.

[42] Gaillot, O., Bregenholt, S., Jaubert, E., Di Santo, J. P., Berche, P., Stress-induced ClpP serine protease of *Listeria monocytogenes* is essencial for induction of listeriolysin O-dependent protective immunity. *Infect. Immun.* 2001, *69*, 4938–4943.

[43] Gerth, U., Krüger, E., Derré, I., Msadek, T., Hecker, M., Stress induction of the *Bacillus subtilis clpP* gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance. *Mol. Microbiol.* 1998, *28*, 787–802.

[44] Jones, K. M., Walker, G. C., Responses of the model legume *Medicago truncatula* to the rhizobial exopolysaccharide succinoglycan. *Plant Signal. Behav.* 2008, *3*, 888–890.

[45] Stoczko, M., Frère, J., Rossolini, G. M., Docquier, J., Postgenomic scan of metallo-β-lactmase homologues in Rhizobacteria: identification and characterization of BJP-1, a subclass B3 ortholog from *Bradyrhizobium japonicum*. *Antimicrob. Agents Chemother.* 2006, *50*, 1973–1981.

[46] Dunn, M. F., Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia. *FEMS Microbiol. Rev.* 1998, *22*, 105–123.

[47] Sarma, A. D., Emerich, D. W., A comparative proteomic evaluation of culture grown *vs* nodule isolated *Bradyrhizobium japonicum*. *Proteomics* 2006, *6*, 3008–3028.

[48] Yadav, A. S., Auxotrophy in rhizobia revisited. *Indian J. Microbiol.* 2007, *47*, 279–288.

[49] Mishima, E., Hosokawa, A., Imaizumi-Anraku, H., Saito, K. *et al.*, Requirement for *Mesorhizobium loti* ornithine transcarbamoylase for successful symbiosis with *Lotus japonicus* as revealed by an unexpected long-range genome deletion. *Plant Cell Physiol.* 2008, *49*, 301–313.

[50] Lodwig, E. M., Hosie, A. H., Bourdès, A., Findlay, K. *et al.*, Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* 2003, *422*, 722–726.

[51] Rastogi, V. K., Watson, R. J., Aspartate aminotransferase activity is required for aspartate catabolism and symbiotic nitrogen fixation in *Rhizobium meliloti*. *J. Bacteriol.* 1991, *173*, 2879–2887.

[52] Salminen, S. O., Streeter, J. G., Factors contributing to the accumulation of glutamate in *Bradyrhizobium japonicum* bacteroids under microaerobic conditions. *J. Gen. Microbiol.* 1990, *136*, 2119–2126.

[53] Becker, A., Bergès, H., Krol, E., Bruand, C. *et al.*, Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol. Plant Microbe Interact.* 2004, *17*, 292–303.

[54] Streit, W. R., Joseph, C. M., Philips, D. A., Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by *Rhizobium meliloti* 1021. *Mol. Plant Microbe Interact.* 1996, *9*, 330–338.

[55] Pobigaylo, N., Szymczak, S., Nattkemper, T. W., Becker, A., Identification of genes Relevant to symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants. *Mol. Plant Microbe Interact.* 2008, *21*, 219–231.

[56] Franck, W. L., Chang, W., Qiu, J., Sugawara, M. *et al.*, Whole-genome transcriptional profiling of *Bradyrhizobium japonicum* during chemoautotrophic growth. *J. Bacteriol.* 2008, *190*, 6697–6705.

[57] Kuiper, I., Lagendijk, E. L., Bloemberg, G. V., Lugtenberg, B. J. J., Rhizoremediation: a beneficial plant-microbe interaction. *Mol. Plant Microbe Interact.* 2004, *17*, 6–15.

[58] Latha, S., Mahadevan, A., Role of rhizobia in the degradation of aromatic substances. *World J. Microbiol. Biotechnol.* 1997, *13*, 601–607.

[59] Capela, D., Filipe, C., Bobik, C., Batut, J., Bruand, C., *Sinorhizobium meliloti* differentiation during symbiosis with alfalfa: a transcriptomic dissection. *Mol. Plant Microbe Interact.* 2006, *19*, 363–372.

[60] Reeve, W. G., Tiwari, R. P., Guerreiro, N., Stubbs, J. *et al.*, Probing for pH-regulated proteins in *Sinorhizobium medicae* using proteomic analysis. *J. Mol. Microbiol. Biotechnol.* 2004, *7*, 140–147.

[61] Lubec, G., Afjehi-Sadat, L., Yang, J., John, J. P. P., Searching for hypothetical proteins: theory and practice based upon original data and literature. *Prog. Neurobiol.* 2005, *77*, 90–127.

[62] Elias, D. A., Monroe, M. E., Marshall, M. J., Romine, M. F. *et al.*, Global detection and characterization of hypothetical proteins in *Shewanella oneidensis* MR-1 using LC-MS based proteomics. *Proteomics* 2005, *5*, 3120–3130.

[63] Puranamaneewiwat, N., Tajima, S., Niamsup, H., Proteomic analysis of *Bradyrhizobium japonicum* USDA110 in acidic condition. *Chiang Mai J. Sci.* 2006, *33*, 335–345.

[64] Pessi, G., Ahrens, C. H., Rehrauer, H., Lindemann, A. *et al.*, Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. *Mol. Plant Microbe Interact.* 2007, *20*, 1353–1363.

[65] Rosander, A., Frykberg, L., Ausmees, N., Müller, P., Identification of extracytoplasmic proteins in *Bradyrhizobium japonicum* using phage display. *Mol. Plant Microbe Interact.* 2003, *16*, 727–737.

[66] Zapun, A., Bardwell, J. C. A., Creighton, T. E., The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation *in vivo*. *Biochemistry* 1993, *32*, 5083–5092.

[67] Fabianek, R. A., Hennecke, H., Thöny-Meyer, L., Periplasmic protein thiol:sulfide oxidoreductases of *Escherichia coli*. *FEMS Microbiol. Rev.* 2000, *24*, 303–316.

[68] Yang, J., Sangwan, I., Lindemann, A., Hauser, F. *et al.*, *Bradyrhizobium japonicum* senses iron through the status of haem to regulate iron homeostasis and metabolism. *Mol. Microbiol.* 2006, *60*, 427–437.

[69] Lindemann, A., Moser, A., Pessi, G., Hauser, F. *et al.*, New target genes controlled by the *Bradyrhizobium japonicum* two-component regulatory system RegSR. *J. Bacteriol.* 2007, *189*, 8928–8943.

[70] Kambara, K., Ardissone, S., Kobayashi, H., Saad, M. M. *et al.*, Rhizobia utilize pathogen-like effector proteins during symbiosis. *Mol. Microbiol.* 2009, *71*, 92–106.

[71] Lodwig, E. M., Leonard, M., Marroqui, S., Wheeler, T. R., Findlay, K., Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. *Mol. Plant Microbe Interact.* 2005, *18*, 67–74.

[72] Becker, A., Barnett, M. J., Capela, D., Dondrup, M. *et al.*, A portal for rhizobial genomes: *Rhizo*GATE integrates a *Sinorhizobium meliloti* genome annotation update with postgenome data. *J. Biotechnol.* 2009, *140*, 45–50.

[73] Batista, J. S. S., Hungria, M., Barcellos, F. G., Ferreira, M. C., Mendes, I. C., Variability in *Bradyrhizobium japonicum* and *B.* elkanii seven years after introduction of both the exotic microsymbiont and the soybean host in a Cerrados soil. *Microbial Ecol.* 2007, *53*, 270–284.

[74] Demotte, N., Ahrens, C. H., Knief, C., Oelir, E. *et al.*, An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. *Proteomics* 2010, *10*, 1391–1400.

PROTEOMICS

Supporting Information for Proteomics

DOI 10.1002/pmic.201000092

Jesiane Stefânia da Silva Batista, Adalgisa Ribeiro Torres and,

Mariangela Hungria

Towards a two-dimensional proteomic reference map of *Bradyrhizobium japonicum* CPAC 15: Spotlighting "hypothetical proteins"

Table 1. Identified proteins of *B. japonicum* CPAC15 whole-cell extracts

Spot ID	NCBI ID	Gene	Protein description	T. mass	E. mass	T. p*I*	E. p*I*	Cellular location	SignalP
Information storage and processing									
J – Translation, ribosomal structure and biogenesis									
1	gil27380513	*tuf*	Elongation factor Tu (EC = 3.6.5.3)[a] [47][b]	43569	44000	5.78	5.21	Cytoplasmic	0.000
2	gil27380513	*tuf*	Elongation factor Tu (EC = 3.6.5.3)[a] [47]	43569	44000	5.78	6.41	Cytoplasmic	0.000
3	gil27379187	*rplI*	50S ribosomal protein L9[a] [10]	21886	26000	5.08	3.6	Cytoplasmic	0.000
4	gil27382552	*rplY*	50S ribosomal protein L25/general stress protein Ctc[a]	24606	25000	9.30	7.57	Cytoplasmic	0.000
5	gil27379969	*frr*	Ribosome recycling factor[a]	20948	24000	7.04	8.25	Cytoplasmic	0.000
6	gil27352649	*efp*	Elongation factor P[a]	19439	24000	5.02	5.95	Cytoplasmic	0.000
7	gil27380499	*rplE*	50S ribosomal protein L5[a] [15]	21018	22000	9.37	9.84	Cytoplasmic[c]	0.000
8	gil27379190	*rpsF*	30S ribosomal protein S6[a] [10, 16]	18616	18000	5.46	5.80	Cytoplasmic	0.000
9	gil27375821		Translation initiation factor IF-3[a]	20459	15000	9.42	5.71	Cytoplasmic	0.000
10	gil146192859	*rpsG*	30S ribosomal protein S7 (*Bradyrhizobium* sp. ORS278)	16390	17000	10.17	9.97	Inner membrane[c]	0.000
K – Transcription									
11	gil27380487	*rpoA*	DNA-directed RNA polymerase subunit α(EC = 2.7.7.6)[a] [63]	38035	41000	4.9	5.21	Cytoplasmic	0.000
12	gil27375786	*hrcA*	Heat-inducible transcription repressor[a]	39142	37000	5.83	6.61	Cytoplasmic	0.000
13	gil27382492		Transcriptional regulatory protein, Lys-R family[a] [9, 10]	33633	31000	6.26	7.15	Cytoplasmic	0.000
14	gil146339278		Putative transcription regulator (AraC-type) (*Bradyrhizobium* sp. ORS278)	35038	25000	10.09	6.40	Cytoplasmic	0.000
15	gil27380527	*nusG*	Transcription antitermination protein[a] [15]	20718	22000	7.93	8.01	Cytoplasmic	0.000
16	gil27382908		RNA polymerase σ factor[a]	19977	19000	5.30	5.16	Cytoplasmic	0.340
17	gil27382908		RNA polymerase σ factor[a]	19977	19000	5.30	5.26	Cytoplasmic	0.340
18	gil27382489	*greA*	Transcription elongation factor[a]	17169	18000	5.67	6.45	Cytoplasmic[c]	0.023
L – Replication, recombination and repair									
19	gil27379809	*ssb*	Single-strand DNA-binding protein[a]	17569	18000	6.62	7.72	Cytoplasmic[c]	0.000
20	gil91977491	*yqgF*	Holliday junction resolvase (*R. palustris* BisB5)	16992	48000	9.42	6.60	Cytoplasmic[c]	0.000
Cellular processes and signaling									
V – Defense mechanisms									
21(581)	gil27380471		Putative β-lactamase precursor (EC = 3.5.2.6)[a]	31482	28000	8.5	6.58	Cytoplasmic[c]	1.000
T – Signal transduction mechanisms									
22	gil27375441		Two-component response regulator[a]	25338	40000	9.62	6.69	Cytoplasmic	0.000
23	gil27382906		Two-component response regulator PhyR[a] [10, 47]	29606	29000	5.04	5.52	Cytoplasmic	0.000
24	gil13473203		Hybrid sensory histidine kinase (*Mesorhizobium loti* MAFF303099)	153629	27000	5.11	8.19	Inner membrane	0.000
25	gil27375497		Two-component response regulator[a] [47, 63]	22739	24000	6.45	7.71	Cytoplasmic	0.000
26	gil27379399		Two-component response regulator[a]	16184	16000	5.29	5.70	Cytoplasmic[c]	0.000
M – Cell wall/membrane/envelope biogenesis									
27	gil27376610	*exoN*	UTP-glucose-1-phosphate uridylyltransferase (EC = 2.7.7.9)[a] [47, 63]	32134	34000	6.62	7.67	Cytoplasmic	0.246
28	gil27376610	*exoN*	UTP-glucose-1-phosphate uridylyltransferase(EC = 2.7.7.9)[a]	32134	34000	6.62	7.75	Cytoplasmic	0.246

Table 1. Continued

Spot ID	NCBI ID	Gene	Protein description	T. mass	E. mass	T. p*I*	E. p*I*	Cellular location	SignalP
N – Cell motility									
29	gil27380055	*clpP*	ATP-dependent ClpP proteolytic subunit[a)]	23552	24000	5.70	6.56	Cytoplasmic	0.000
30	gil27375651	*typA*	GTP-binding tyrosin phosphorylated protein[a)] [10]	67131	69000	5.49	6.90	Cytoplasmic	0.000
31	gil27382111	*fliG*	Flagellar motor switch protein G[a)]	40482	31000	4.88	5.86	Cytoplasmic	0.014
O – Post-translational modification, protein turnover and chaperones									
32	gil27380737	*groEL*	Chaperonin GroEL[a)], [10, 16, 47, 63]	57716	54000	5.45	5.54	Cytoplasmic	0.000
33	gil27375705	*trxA*	Thioredoxin[a)]	33507	34000	5.19	5.22	Cytoplasmic	0.001
34	gil27375787	*grpE*	Heat shock protein[a)]	21642	27000	4.84	4.75	Cytoplasmic[c)]	0.000
35	gil27379801	*ppiB*	Peptidyl prolyl *cis-trans* isomerase[a)]	16995	17000	5.67	6.54	Periplasmic[c)]	0.000
36	gil27382155	*msrB*	Methionine sulfoxide reductase B (EC = 1.8.4.12)[a)]	15745	16000	5.66	6.38	Cytoplasmic	0.000
37	gil27377170	*groEL*	Chaperonin GroEL[a)] [9, 10]	57749	58000	5.19	5.82	Cytoplasmic	0.000
Metabolism									
C (10) – Energy production and conversion									
38	gil27375625	*sdhA*	Succinate dehydrogenase flavoprotein subunit (EC = 1.3.5.1)[a)] [10, 47]	66903	62000	5.91	6.71	Periplasmic	0.000
39	gil27375553	*atpA*	F0F1 ATP synthase subunit α (EC = 3.6.3.14)[a)] [10, 16, 47]	55301	53000	6.88	7.72	Cytoplasmic[c)]	0.000
40	gil27378006	*paaE*	Putative ferredoxin reductase electron transfer component protein[a)]	41516	42000	5.19	5.40	Cytoplasmic	0.081
41	gil27381248		Quinone oxidoreductase (EC = 1.6.5.2)[a)]	34356	37000	8.84	9.86	Cytoplasmic[c)]	0.000
42	gil218516225		Pyruvate dehydrogenase (acetyl-transferring) protein, α subunit (*R. etli* 8C-3)	10793	36000	4.64	7.78	Cytoplasmic	0.000
43	gil27375567	*mdh*	Malate dehydrogenase (EC = 1.1.1.37)[a)] [10, 16]	34275	35000	5.88	6.96	Cytoplasmic	0.011
44	gil27376489	*etfL*	Electron transfer flavoprotein α subunit[a)] [47, 63]	32186	32000	5.14	5.38	Cytoplasmic[c)]	0.002
45	gil27376488	*etfS*	Electron transfer flavoprotein β subunit[a)] [9, 16]	26495	27000	8.86	9.73	Cytoplasmic[c)]	0.000
46	gil146339324		Two-component C4-dicarboxylate transport system, sensor protein (EC = 3.1.2.4.3) (*Bradyrhizobium* sp. ORS278)	67379	25000	7.74	8.63	Inner membrane	0.000
47	gil27379893	*pdhB*	Pyruvate dehydrogenase subunit β (EC = 1.2.4.1)[a)]	48906	62000	4.81	5.11	Cytoplasmic	0.000
G – Carbohydrate transport and metabolism									
48	gil27382210		ABC transporter substrate-binding protein[a)] [16]	51924	45000	6.97	7.08	Periplasmic[c)]	1.000
49	gil27375844		ABC transporter glycerol-3-phosphate-binding protein[a)] [10, 16]	48280	40000	7.68	8.96	Periplasmic[c)]	1.000
50	gil27376634	*gapA*	Glyceraldehyde-3-phosphate dehydrogenase[a)] [16, 63]	35826	38000	7.77	9.45	Cytoplasmic[c)]	0.000
51	gil27376634	*gapA*	Glyceraldehyde-3-phosphate dehydrogenase[a)] [16, 63]	35826	38000	7.77	8.51	Cytoplasmic	0.000
52	gil27378854		Periplasmic mannitol-binding protein[a)] [16, 63]	42147	35000	8.83	7.20	Periplasmic[c)]	0.996
53	gil27379905	*eno*	Enolase[a)]	45314	47000	5.08	5.86	Cytoplasmic	0.000
E – Amino acid transport and metabolism									
54	gil27375633	*argG*	Argininosuccinate synthase(EC = 6.3.4.5)[a)] [10, 47, 63]	49371	48000	5.50	5.97	Cytoplasmic[c)]	0.014
55	gil27382527	*aatA*	Aspartate aminotransferase (EC = 2.6.1.1)[a)] [10, 47]	44485	44000	6.04	7.95	Cytoplasmic[c)]	0.000
56	gil27381148	*sgaA*	Serine-glyoxylate aminotransferase (EC = 2.6.1.45)[a)]	47033	41000	8.96	8.31	Cytoplasmic	0.000
57	gil27380786		ABC transporter substrate-binding protein[a)] [16]	40020	41000	8.90	9.46	Periplasmic	1.000
58	gil27378033		ABC transporter amino acid-binding protein[a)] [16, 47, 63]	48454	40000	6.35	6.03	Periplasmic	0.126
59	gil27378033		ABC transporter amino acid-binding protein[a)] [16, 47, 63]	48454	39000	6.35	6.33	Periplasmic	0.126

Table 1. Continued

Spot ID	NCBI ID	Gene	Protein description	T. mass	E. mass	T. p*I*	E. p*I*	Cellular location	SignalP
60	gil27380061		L-Asparaginase (EC = 3.5.1.1)[a] [9, 10, 47]	39 549	38 000	6.93	6.16	Periplasmic[c]	1.000
61	gil27379564	*cysK*	Cysteine synthase A (EC = 2.5.1.47)[a]	37 243	38 000	7.57	8.93	Cytoplasmic[c]	0.000
62	gil27376169		ABC transporter substrate-binding protein[a] [16, 47]	39 717	37 000	6.62	6.91	Periplasmic	1.000
63	gil27380183	*dapA*	Dihydrodipicolinate synthase (DHDPS) (EC = 4.2.1.52)[a]	32 077	35 000	6.84	8.33	Cytoplasmic[c]	0.015
64	gil27378020		Amino acid binding protein[a] [9, 10]	36 096	34 000	7.59	7.34	Periplasmic	1.000
65	gil27379557		ABC transporter amino acid-binding protein[a] [9, 10, 16, 47, 74]	36 860	33 000	6.21	5.83	Periplasmic	1.000
66	gil27379557		ABC transporter amino acid-binding protein[a] [9, 10, 16, 47, 74]	36 860	33 000	6.21	6.57	Periplasmic	1.000
67	gil27375765	*hisF*	Imidazole glycerol phosphate synthase subunit[a] [47]	27 763	30 000	5.44	6.40	Cytoplasmic	0.000
68	gil27383212	*argB*	Acetylglutamate kinase (EC = 2.7.2.8)[a] [47]	31 294	30 000	5.33	5.82	Cytoplasmic[c]	0.000
69	gil27378036		Amino acid ABC transporter ATP-binding protein[a]	26 116	26 000	6.45	7.30	Cytoplasmic/ inner membrane[c]	0.000
70	gil27375606	*leuD*	Isopropylmalate isomerase small subunit (EC = 4.2.1.33)[a] [47]	22 781	26 000	5.52	5.97	Cytoplasmic[c]	0.000
71	gil27381064		ABC transporter substrate-binding protein[a] [16, 47]	41 827	39 000	7.71	7.39	Periplasmic[c]	1.000
72	gil27381064		ABC transporter substrate-binding protein[a] [16, 47]	41 827	38 000	7.71	7.39	Periplasmic[c]	1.000
73	gil27379995		ABC transporter substrate-binding protein[a] [47]	37 695		8.72	9.68	Periplasmic	1.000
74	gil27381681		Dehydrogenase[a]	38 493	39 000	5.82	6.66	Cytoplasmic	0.000
75	gil27379793	*cysK*	Cysteine synthase[a]	34 228	39 000	8.78	9.67	Cytoplasmic[c]	0.000
F – Nucleotide transport and metabolism									
76	gil27379083	*guaB*	Inositol-5-monophosphate dehydrogenase (EC = 1.1.1.205)[a]	52 908	52 000	6.96	8.04	Cytoplasmic	0.000
77	gil27380211		Dihydroorotase (EC = 3.5.2.3)[a] [47]	46 672	47 000	5.12	5.54	Cytoplasmic[c]	0.000
78	gil148258057	*purK*	Phosphoribosylaminoimidazole carboxylase ATPase subunit (EC = 4.1.1.21) (*Bradyrhizobium* sp. BTAi1)	40 130	35 000	6.17	7.91	Cytoplasmic[c]	0.035
79	gil27376071		5′-Methylthioadenosine phosphorylase (EC = 2.4.2.28)[a] [47]	31 783	32 000	6.21	5.24	Cytoplasmic	0.000
80	gil27375923	*purC*	Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC = 6.3.2.6)[a] [47]	33 736	32 000	5.35	5.91	Cytoplasmic	0.000
81	gil27379970	*pyrH*	Uridylate kinase[a] [47]	26 222	26 000	6.44	6.15	Cytoplasmic	0.000
82	gil27379230	*ndk*	Nucleoside diphosphate kinase[a] [47]	15 050	16 000	6.75	8.03	Cytoplasmic	0.000
83	gil27375890	*pnpA*	Polynucleotide phosphorylase/polyadenylase[a] [47, 63]	78 307	86 000	5.47	5.31	Cytoplasmic	0.000
H – Coenzyme transport and metabolism									
84	gil27381056	*metK*	*S*-adenosylmethionine synthetase (EC = 2.5.1.6)[a] [9, 10, 16, 47]	43 613	42 000	5.88	6.58	Cytoplasmic	0.000
85	gil27381056	*metK*	*S*-adenosylmethionine synthetase (EC = 2.5.1.6)[a] [9, 10, 16, 47]	43 613	41 000	5.88	6.58	Cytoplasmic	0.000
86	gil27377622	*moeB*	Molybdopterin biosynthesis protein B[a] [10]	25 995	22 000	5.97	6.59	Cytoplasmic	0.000
87	gil27381770	*thiC*	Thiamine biosynthesis protein[a]	70 400	76 000	5.89	6.69	Cytoplasmic	0.222
88	gil27375660	*folD*	Bifunctional methylenetetrahydrofolate dehydrogenase/ cyclohydrolase[a]	30 659	31 000	8.97	9.94	Cytoplasmic	0.000
89	gil27375291	*cobS*	Cobalt insertion protein (EC:6.6.1.2)[a] [10, 47]	42 351	40 000	6.42	6.05	Cytoplasmic[c]	0.003
I – Lipid metabolism									
90	gil27378919		3-Oxoacyl-(acyl carrier protein) synthase II (EC = 2.3.1.41)[a] [47]	44 874	45 000	5.56	6.12	Cytoplasmic	0.288
91	gil27382511		Acetyl-CoA acetyltransferase (EC = 2.3.1.9)[a]	41 982	42 000	6.56	7.84	Cytoplasmic	0.000
92	gil27379966	*dxr*	1-Deoxy-D-xylulose 5-phosphate reductoisomerase (EC = 1.1.1.267)[a]	42 957	41 000	6.41	8.13	Cytoplasmic	0.570
93	gil27375337	*atoB*	Acetyl-CoA acetyltransferase (EC = 2.3.1.9)[a] [10, 47]	40 964	40 000	7.01	8.08	Cytoplasmic	0.007
94	gil27381334	*hbdA*	3-Hydroxybutyryl-CoA dehydrogenase (EC = 1.1.1.157)[a] [47]	30 302	33 000	6.44	7.94	Cytoplasmic	0.209

Table 1. Continued

Spot ID	NCBI ID	Gene	Protein description	T. mass	E. mass	T. p*I*	E. p*I*	Cellular location	SignalP
95	gil27375337	*atoB*	Acetyl-CoA acetyltransferase (EC = 2.3.1.9)[a] [10, 47]	40 964	32 000	7.01	8.01	Cytoplasmic	0.007
96	gil27382511		Acetyl-CoA acetyltransferase(EC = 2.3.1.9)[a]	41 982	32 000	6.56	8.15	Cytoplasmic	0.000
97	gil27379051		Putative propionyl-CoA carboxylase β chain (EC = 6.4.1.3)[a]	57 105	31 000	8.35	7.55	Cytoplasmic[c]	0.000
98	gil146337774		Putative enoyl-CoA hydratase(EC = 5.3.3.8/4.2.1.17) (*Bradyrhizobium* sp. ORS278)	29 088	30 000	6.67	8.98	Cytoplasmic	0.000
99	gil27382204	*pcaI*	3-Oxoadipate CoA-transferase subunit A[a] [47]	25 523	26 000	8.56	9.75	Cytoplasmic	0.000
100	gil27376270		Acetyl-CoA acetyltransferase (EC = 2.3.1.9)[a]	42 174	46 000	5.54	6.02	Cytoplasmic	0.000
101	gil27380444		Putative dehydrogenase[a]	38 314	32 000	9.52	9.67	Cytoplasmic[c]	0.000
102	gil27379193	*fabD*	Acyl-carrier-protein *S*-malonyltransferase[a] [47]	32 462	32 000	5.57	6.13	Cytoplasmic[c]	0.054
103	gil27378851		Dehydrogenase[a]	30 708	32 000	8.50	9.66	Cytoplasmic[c]	0.322
P – Inorganic ion transport and metabolism									
104	gil27376202	*pstS*	ABC transporter phosphate-binding protein[a] [9, 10, 16, 63]	36 308	31 000	7.68	9.05	Periplasmic	1.000
105	gil27379976		Putative β-type carbonic anhydrase (EC = 4.2.1.1)[a]	25 959	24 000	8.26	9.85	Cytoplasmic[c]	0.997
106	gil27382885	*sodF*	Superoxide dismutase (EC = 1.15.1.1)[a] [10, 47, 74]	24 326	23 000	6.75	7.49	Cytoplasmic[c]	0.000
Q – Secondary metabolites biosynthesis, transport and catabolism									
107	gil27375454	*hmgA*	HmgA (EC = 1.13.11.5)[a]	51 303	46 000	6.08	6.82	Cytoplasmic[c]	0.000
108	gil27383002		2-Hydroxyhepta-2,4-diene-1,7-dioate isomerase[a] [10]	30 183	34 000	6.84	6.05	Cytoplasmic[c]	0.000
109	gil27379071		Short-chain dehydrogenase[a]	32 145	32 000	6.35	7.37	Cytoplasmic	0.049
110	gil27383002		2-Hydroxyhepta-2,4-diene-1,7-dioate isomerase[a] [10]	30 183	30 000	6.84	7.38	Cytoplasmic[c]	0.000
111	gil27375248		Oxidoreductase[a]	26 976	29 000	8.27	10.00	Cytoplasmic[c]	0.994
112	gil27382999		Putative 3-oxoacyl-(acyl-carrier-protein) reductase (EC = 1.1.1.100)[a]	27 406	29 000	8.99	9.76	Cytoplasmic	0.015
Poorly characterized									
R – General function prediction only									
113	gil27376279	*tldD*	TldD protein – predicted Zn-dependent proteases and their inactivated homologs[a] [10, 47]	51 101	48 000	5.51	5.86	Cytoplasmic[c]	0.000
114	gil27376412		Immunogenic protein precursor[a]	39 117	30 000	8.87	9.59	Periplasmic[c]	0.000
115	gil27380722	*cinA*	Competence-damage-associated protein[a] [9, 10]	26 451	26 000	5.29	5.57	Cytoplasmic	0.000
116	gil27382679	*wrbA*	TrpR binding protein[a]	20 861	23 000	6.97	8.01	Cytoplasmic[c]	0.009
117	gil27377891		NAD-dependent alcohol dehydrogenase (EC = 1.1.1.1)[a]	37 812	39 000	8.05	9.67	Cytoplasmic	0.000
118	gil27377720		Oxidoreductase[a]	29 963	30 000	8.84	9.66	Cytoplasmic[c]	0.000
119	gil27377543		CbbY/CbbZ/GpH/YieH family protein[a]	25 723	27 000	4.90	6.33	Cytoplasmic[c]	0.000

Matched peptides masses and MS/MS combined results are available in PRIDE (http://ebi.ac.uk/pride/) under the experiment accession number 9769. T., theoretical; E., experimental.
a) Best-hit with *B. japonicum* strain USDA 110. When best-hit corresponded to other species, the name of the species is shown in brackets.
b) Reference of the proteomic study which has also described the protein.
c) Unable to predict in PSORT or contrasting results.