



UNIVERSIDADE
ESTADUAL DE LONDRINA

LEANDRO PEREIRA DE GODOY

**PANORAMA GENÔMICO DA ESTIRPE SEMIA 5079 (=CPAC
15) DE *Bradyrhizobium japonicum*, RECOMENDADA
COMERCIALMENTE PARA A CULTURA DA SOJA (*Glycine
max* (L.) Merr.)**

LEANDRO PEREIRA DE GODOY

**PANORAMA GENÔMICO DA ESTIRPE SEMIA 5079 (=CPAC
15) DE *Bradyrhizobium japonicum*, RECOMENDADA
COMERCIALMENTE PARA A CULTURA DA SOJA (*Glycine
max* (L.) Merr.)**

Dissertação apresentada ao Programa de Pós-Graduação em Microbiologia da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre em Microbiologia.

Orientadora: Dra. Mariangela Hungria

Londrina
2007

LEANDRO PEREIRA DE GODOY

**PANORAMA GENÔMICO DA ESTIRPE SEMIA 5079 (=CPAC
15) DE *Bradyrhizobium japonicum*, RECOMENDADA
COMERCIALMENTE PARA A CULTURA DA SOJA (*Glycine
max* (L.) Merr.)**

BANCA EXAMINADORA

Dra. Mariângela Hungria da Cunha (orientadora)

Dr. Fernando Gomes Barcellos

Dra. Marisa Fabiana Nicolas

Londrina, 17 de agosto de 2007.

DEDICATÓRIA

Dedico aos meus pais, que no seu silêncio, nas suas orações e nas suas lições, sempre acompanharam meus passos, colocando-me a caminho, em busca de algo maior.

Amo vocês.

AGRADECIMENTOS

O trabalho aqui apresentado não é fruto do esforço e dedicação de apenas uma pessoa. São muitas as que, de uma ou outra forma, contribuíram para que os objetivos propostos fossem alcançados. Minha gratidão a todos que estiveram comigo nesta caminhada, agregando valores importantes à minha vida pessoal e profissional.

Aos meus pais, Antonio e Célia, por serem meu porto seguro, meu apoio e incentivo em todas as decisões que tomo, e os conseqüentes caminhos em que elas me levam.

Aos meus avós, Pedro, Elza e Vergínia (saudades vovó), pelas orações e lições de fé, amor e perseverança.

À minha maninha Lidiane, pelas suas preocupações e orações.

À minha orientadora, amiga, e exemplo de pessoa, Dra. Mariangela Hungria, pela oportunidade, confiança e profissionalismo demonstrados, com os quais muito aprendi.

À Marcela, que me ensinou muitas lições com seus gestos simples, carinhosos e inocentes.

Aos amigos de laboratório, Jesi, Pâmela, Alan, Fabinho, Rose, Maria, (um brinde a Bonito!), Fernando e Lígia (obrigado de coração), Renan Somália, Adalgisa, Susan, Luciana, Mirian, Daysi, Lucas, Fábio, Dr. Rubens, Renan, Nágila, Adrian, Eriana, Priscila, Fabiana, Rinaldo, Leni, Michele, que nas conversas, brincadeiras e discussões científicas, proporcionaram um ambiente de alegrias e de muitas histórias. Obrigado pelo ótimo convívio, pela amizade e o aprendizado.

A pessoal do Mestrado, principalmente Raquel, Alessandra, Nádia, Ariane, Marcelo, pela grande amizade conquistada, e por fazerem parte da turma que mais churrascos fez num mestrado, tornando nosso aprendizado mais eficiente.

Aos amigos do PROLOV, pelas suas histórias de lutas e superações, que me motivaram em muitos momentos Força amigos.

À amiga Angela, pelo incentivo inicial desta longa jornada.

À banca examinadora, Dr Fernando Gomes Barcellos, Dra Marisa Nicolás, e Dra Mariangela Hungria,, por sua amizade, disponibilidade, e por suas contribuições inestimáveis durante o desenvolvimento e a conclusão deste estudo.

À EMBRAPA soja pela estrutura que me foi concedida para a realização deste trabalho.

Aos funcionários e professores do curso de Pós-graduação em Microbiologia da Universidade Estadual de Londrina, pelo conhecimento transmitido e o apoio concedido.

Faço deste trabalho um louvor a Deus, que me guia pelos caminhos, me inspira e renova minhas forças.

Obrigado Senhor.

“Deus não escolhe os capacitados, mas capacita os escolhidos”.
“A maior recompensa do nosso trabalho não é o que nos pagam por ele,
mas aquilo em que ele nos transforma.”

(John Ruskin)

GODOY, Leandro Pereira de. **Panorama genômico da estirpe semia 5079 (=CPAC 15) DE *Bradyrhizobium japonicum*, recomendada comercialmente para a cultura da soja (*Glycine max* (L.) Merr.).** 2007. 66f. Dissertação (Mestrado em Microbiologia) – Universidade Estadual de Londrina, Londrina, 2007.

RESUMO

O processo de fixação biológica do nitrogênio atmosférico (FBN) com bactérias das espécies *Bradyrhizobium japonicum* e *B. elkanii* é fundamental para a viabilidade econômica da cultura da soja no Brasil. Com a expansão da cultura no País, no início dos anos 1960, também teve início um programa de seleção de estirpes adaptadas às condições ambientais e às cultivares nacionais e a estirpe SEMIA 5079 (=CPAC 15), pertencente à espécie *B. japonicum*, destacou-se por sua eficiência, competitividade e adaptação às condições ambientais estressantes, passando a ser utilizada em inoculantes comerciais desde 1992. Nesta última década, o seqüenciamento de alguns genomas de rizóbios permitiu obter grandes avanços no conhecimento da genética dessas bactérias, e este estudo teve por objetivo o seqüenciamento parcial do genoma da estirpe SEMIA 5079, visando a identificação de genes relacionados à capacidade saprofítica, à competitividade e à eficiência do processo de FBN. As leituras dos clones de bibliotecas do tipo “shotgun”, construídas a partir do DNA total da SEMIA 5079, resultaram em uma cobertura real de 13,17 % do genoma, com um total de 1.371 CDSs (“coding sequences”) anotadas, sendo 729 de função conhecida, 312 hipotéticas conservadas e 330 hipotéticas. A classificação funcional das CDSs pelo banco de dados COG (“Clusters of Orthologous Groups of proteins”) mostrou genes putativos em quase todas as classes. Em relação ao banco KEGG (“Kyoto Encyclopedia of Genes e Genomes”), houve similaridade das seqüências anotadas na SEMIA 5079 com outros 71 organismos, e a maior similaridade foi constatada com a estirpe USDA 110 de *B. japonicum*, e com as estirpes BTAi1 e ORS278 de *Bradyrhizobium* sp. Quatorze transposases foram encontradas, a maioria relacionada a seqüências de inserção, indicativo de plasticidade elevada do genoma da SEMIA 5079; além disso, 4,54% das CDSs foram identificadas como genes parálogos, também relacionados à adaptação ambiental. Várias CDSs foram anotadas como genes que codificam proteínas relacionadas à nodulação, como os genes *nodB*, *nodD2* e *nodW* e à FBN, como os genes *nifE*, *fixP*, *fixQ*, *fixR* e *fixO*, além de outras proteínas envolvidas no ciclo do N. No seqüenciamento parcial do genoma da estirpe SEMIA 5079, 5,76 % das CDSs válidas codificam enzimas que participam de quase todas as vias de biodegradação de xenobióticos, podendo representar um reservatório importante de genes com potencial biotecnológico.

Palavras-chave: Microbiologia. Microorganismos fixadores de nitrogênio. Bactérias fitopatogênicas

GODOY, Leandro Pereira de. **Panorama genômico da estirpe semia 5079 (=CPAC 15) DE *Bradyrhizobium japonicum*, recomendada comercialmente para a cultura da soja (*Glycine max* (L.) Merr.).** 2007. 66f. Dissertação (Mestrado em Microbiologia) – Universidade Estadual de Londrina, Londrina, 2007.

ABSTRACT

The biological nitrogen fixation (BNF) process with bacteria belonging to the species *Bradyrhizobium japonicum* and *B. elkanii* is crucial for the economical viability of the soybean crop in Brazil. A strain selection program began with the crop expansion in the earlier 1960s, aiming at the identification of bacteria adapted to the local environmental conditions and to the Brazilian cultivars; strain SEMIA 5079 (= CPAC 15) of *B. japonicum* was then recognized by its efficiency, competitiveness and adaptability to stressful environmental conditions, therefore it has been employed in commercial inoculants since 1992. Few rhizobial genomes were completely sequenced in the last decade, and allowed an impressive increase in the genetic knowledge of these bacteria. In this context, this study aimed at the partial sequencing of the genome of strain SEMIA 5079, searching for genes related to the saprophytic capacity, competitiveness and efficiency of the BNF process. The readings of shotgun libraries built with SEMIA 5079 total DNA resulted in a real covering of 13.17% of the genome, with a total of 1.371 CDSs (coding sequences) annotated, 729 with known functions, 312 classified as conserved hypothetical and 330 as hypothetical. The functional classification of the CDSs in the database COG (Clusters of Orthologous Groups of proteins) identified putative genes in almost all classes. Furthermore, the comparison with strain USDA 110 of *B. japonicum*, sequenced in 2002 by Japanese researchers, has also demonstrated similar percentages of CDSs classified in the categories of COG, an important indication to support the strategy of partial genome sequencing of a bacterium. In relation to the KEGG (Kyoto Encyclopedia of Genes and Genomes) database, the CDSs of SEMIA 5079 have shown similarity with the sequences of 71 other organisms, but in a higher number with *B. japonicum* strain USDA 110, and with *Bradyrhizobium* sp. strains BTAi1 and ORS278. Fourteen transposases were found, most related to insertion sequences, indicating high plasticity of the genome of SEMIA 5079; furthermore, 4.54% of the CDSs were classified as paralog genes, also favoring environmental adaptation. Several CDSs were annotated as genes that codify proteins related to the nodulation, as *nodB*, *nodD2* and *nodW* genes, and to the BNF, as *nifE*, *fixP*, *fixQ*, and *fixR* genes, in addition to other proteins involved in the cycle of N. The partial sequencing of the genome of strain SEMIA 5079 has also identified 5.76% of CDSs participating in almost all pathways of xenobiotic biodegradation, and might represent a reservoir of important genes with biotechnological potential.

Keywords: Microbiology. Microorganisms fixation of nitrogen. Bacteria fitopatogenicas.

SUMÁRIO

1 REVISÃO DE LITERATURA	9
1.1 INTRODUÇÃO.....	9
1.2.A CULTURA DA SOJA.....	10
1.2.1 Importância da fixação biológica do N ₂ na cultura da soja.....	12
1.3 SIMBIOSE COM LEGUMINOSAS.....	12
1.4 CARACTERÍSTICAS DA ESTIRPE DE <i>BRADYRHIZOBIUM JAPONICUM</i> CPAC15.....	16
1.5 GENOMAS DE DIAZOTRÓFICOS SIMBIÓTICOS.....	18
2 OBJETIVO	23
2.1 OBJETIVO GERAL.....	23
2.2 OBJETIVO ESPECÍFICO.....	23
3 MATERIAL E MÉTODOS	24
3.1 ESTIRPE.....	24
3.2 EXTRAÇÃO DE DNA GENÔMICO.....	24
3.3 CONSTRUÇÃO DE BIBLIOTECAS “SHOTGUN”.....	25
3.4 SEQÜENCIAMENTO DOS CLONES.....	26
3.5 IDENTIFICAÇÃO E ANOTAÇÃO DOS GENES.....	27
3.6 ANOTAÇÃO MANUAL DOS GENES.....	28
4 RESULTADOS E DISCUSSÃO	29
5 PRINCIPAIS CONCLUSÕES	39
REFERÊNCIAS	40
LISTA DE QUADROS E FIGURAS	52

1 REVISÃO DE LITERATURA

1.1 INTRODUÇÃO

O nitrogênio (N), após o carbono (C), o oxigênio (O) e o hidrogênio (H) é o elemento mais importante na constituição da matéria orgânica, sendo o quarto elemento mais abundante nas plantas e representando cerca de 8% a 16% de sua composição. O N é constituinte essencial de aminoácidos, proteínas, bases nitrogenadas, ácidos nucleicos, hormônios, clorofila, entre outros (Morgante, 2003).

O nitrogênio (N) existe na natureza como N molecular (N_2), íons de nitrato (NO_3^-), amônia (NH_3), íons amônio (NH_4^+), ou incorporado em compostos orgânicos nitrogenados. Esse elemento participa da formação de moléculas fundamentais em diversos processos biológicos, tais como, produção de ácidos nucleicos e proteínas. Apesar de sua abundância na atmosfera na forma de N_2 (79%), o N constitui um dos principais fatores limitantes à produção agrícola mundial, pois nenhum animal ou planta é capaz de utilizá-lo diretamente, devido à tripla ligação que existe entre os dois átomos do N_2 , uma das mais fortes de que se tem conhecimento na natureza. (Hungria et al., 1994a).

As principais fontes fornecedoras do N necessário ao crescimento das plantas são o N do solo, proveniente, principalmente, da decomposição da matéria orgânica, o N fornecido por fertilizantes e o N fornecido pelo processo da fixação biológica do N_2 . Existe, também, uma pequena contribuição por processos não biológicos, com a reação de descargas elétricas com o N_2 , resultando em NO_3^- , que é adicionado ao solo e representa cerca de 10% das entradas positivas no balanço de N na Terra. O fornecimento, utilização e perdas do N formam um ciclo complexo, denominado “ciclo do N” (Hungria et al., 2001). Nos solos brasileiros, porém, o N presente na matéria orgânica do solo é limitado, resultando na necessidade de fornecimento desse nutriente, para que a produtividade das culturas não fique prejudicada (Hungria et al., 1994).

A síntese de fertilizantes nitrogenados iniciou na primeira década do século passado, quando Fritz Haber e Carl Bosch descobriram um processo industrial que possibilita a transformação do N_2 em amônia. As necessidades para essa síntese química são, principalmente, o hidrogênio (derivado de gás de petróleo), um catalisador contendo ferro, altas temperaturas (300°C a 600°C) e altas pressões (200 atm a 800 atm). Conseqüentemente,

o custo econômico para a síntese química de fertilizantes nitrogenados é elevado, resultante, principalmente, da necessidade de gasto de derivados de petróleo, portanto, energia não renovável. Calcula-se que, para a síntese de uma tonelada de amônia, sejam necessários seis barris de petróleo. Outro agravante na utilização dos fertilizantes nitrogenados reside na baixa eficiência de sua utilização pelas plantas, raramente ultrapassando 50%. Deve-se considerar, ainda, que o uso indiscriminado de fertilizantes nitrogenados resulta em poluição ambiental, pois a lixiviação do N e o escoamento desse nutriente pela superfície do solo resultam em acúmulo de formas nitrogenadas nas águas dos rios, lagos e lençóis subterrâneos, podendo atingir níveis tóxicos a peixes, animais e ao homem (Hungria et al., 2001). Práticas alternativas para diminuir esses custos precisam ser procuradas, principalmente no Brasil, que importa fertilizantes nitrogenados para satisfazer à demanda interna (Hungria et al., 1994).

Outra fonte de N é representada pela fixação biológica do N_2 (FBN), processo realizado por determinados procariontes, denominados organismos fixadores de nitrogênio. Esses organismos, também denominados diazotróficos, são procariotos providos de um complexo enzimático denominado dinitrogenase que catalisa a conversão, de N_2 a NH_3 .

Em termos globais, estima-se que a fixação biológica do N_2 contribua com 65% da entrada anual de N na Terra, enquanto que a produção industrial contribui com 24% e a fixação não-biológica com cerca de 10% (Hungria et al., 2001).

1.2 A CULTURA DA SOJA

A soja é considerada uma das culturas mais antigas do mundo, tendo provavelmente sido domesticada no século XI antes de Cristo, no centro e norte da China. A mais antiga referência escrita de seu cultivo se refere ao seu uso como material médico no herbário de Pen Ts'ao Kang Um, pelo imperador Shen Nung (2.828 a 2.383 a.C.) (Morse, 1950; Hymowitz, 1970; Hymowitz e Newell, 1980). O caminho da soja para a Índia, Nepal, Burma, Tailândia, Indochina, Coréia, Japão, Malásia, Indonésia e Filipinas ocorreu antes do primeiro milênio da era cristã e foi acompanhado de diversas alterações genéticas, afetando a simbiose (Hymowitz e Newell, 1980; Cregan et al., 2002; Hungria et al., 2005, 2006).

No Brasil, a soja foi provavelmente introduzida em 1882, na Bahia, pelo agrônomo Gustavo D'utra, como *Soja hispida* (*Glycine soja*) e *Soja ochroleuca* (D'utra, 1882). Dez anos depois, em um relatório da Estação Agronômica de Campinas, constava que

S. hispida também havia sido plantada em novembro de 1891 e, ainda em Campinas, novos experimentos foram conduzidos em 1898 e 1899, com análises químicas de grãos e folhas (D'utra, 1899). Contudo, somente a partir da década de 40 a soja apareceu, pela primeira vez, nas estatísticas do Rio Grande do Sul e, no Paraná, o cultivo se estabeleceu somente nos anos 1960, mas a grande expansão ocorreu de fato nos anos 1970, com o estabelecimento na Região dos Cerrados (Vargas e Hungria, 1997).

O teor protéico elevado dos grãos de soja, de cerca de 40%, faz com que a cultura tenha uma grande importância econômica e social para o Brasil participando, de modo crescente, direta ou indiretamente, na dieta do brasileiro, além de representar a nossa principal fonte de exportação agrícola. Embora a soja seja uma cultura exótica introduzida no País há pouco mais de um século, hoje está estabelecida em quase todo o território nacional, inclusive na região Norte e Nordeste, tornando o Brasil o segundo maior produtor mundial de grãos dessa leguminosa. Durante esse período de expansão a área cultivada com essa leguminosa incrementou de 702.000 ha em 1940/41 para 20.7 milhões de ha na safra de 2006/07 ocupando, atualmente, 45% de toda a área cultivada no país. Apesar de ter ocorrido uma diminuição de 6,9% (1,5 milhão de hectares) em relação à safra passada, que foi de 22,2 milhões de hectares, houve um incremento na produtividade média do País de 2.403 kg ha⁻¹ para 2.802 kg ha⁻¹, ou seja, de 16,6%, elevando a produção de 53,4 milhões de toneladas para 58,0 milhões de toneladas, ou seja, 8,5%. Desse modo, para salientar o progresso genético, pode-se citar que o rendimento de 1.250 kg ha⁻¹ obtido na safra 1977/1978, incrementou para 2.802 kg ha⁻¹ em 2006/2007, portanto um aumento de mais de quatro vezes no rendimento médio nacional. (CONAB, 2007).

Estimativas apontam que o custo de produção da soja nacional é cerca de 40% inferior ao dos Estados Unidos da América (E.U.A.), e com uma produtividade igual ou superior. O menor custo deve-se, especialmente, à prática no Brasil, da inoculação das sementes com rizóbios comprovadamente eficientes e competitivos na FBN com a soja. A simbiose entre rizóbios selecionados e plantas leguminosas através da inoculação constitui uma prática agrônômica de suprimento de N para cultivos com baixo custo, segura e eficiente (Hungria et al., 2006).

1.2.1 A importância da fixação biológica do N₂ na cultura da soja

A importância econômica da FBN para a cultura da soja fica evidenciada quando se considera alguns dos valores atuais da produção nacional. Para cada 1.000 kg de grãos de soja produzidos, com um teor médio de 40% de proteína, são necessários cerca de 80 kg de N ha⁻¹ (65 kg alocados nas sementes e 15 kg de N nas folhas, caule e raízes). Considerando-se o rendimento médio da cultura em 2006/2007, de 2.802 kg ha⁻¹, a planta necessitaria, aproximadamente, 220 kg de N ha⁻¹. Os solos brasileiros são pobres em N, normalmente sendo capazes de fornecer somente 10 a 30 kg de N ha⁻¹, portanto, as plantas necessitariam cerca de 200 kg de N ha⁻¹. Se forem utilizados fertilizantes nitrogenados, as perdas de N são elevadas, por lixiviação, desnitrificação, raramente excedendo 60% do aplicado, desse modo, na verdade seriam necessários 330 kg de N ha⁻¹, ou 730 kg de uréia, a forma mais utilizada, que contém 45% de N (CONAB, 2007; Hungria et al., 2006).

Estimativas de FBN indicam que, no Brasil, 69% a 94% do N necessário à soja é fornecido pelo processo biológico, contribuindo com taxas de 109 a, aproximadamente, 300 kg de N ha⁻¹ (Hungria et al., 2006). Já no caso do maior produtor mundial, os E.U.A., o baixo custo dos fertilizantes nitrogenados resultou em poucos investimentos na pesquisa em FBN, conseqüentemente, a contribuição do processo biológico foi estimada em apenas cerca de 50% (van Kessel e Hartley, 2000). Em contraste, no Brasil, o processo biológico contribui, em grande parte, para que o custo da produção brasileira seja inferior, em cerca de 40%, ao dos E.U.A. (CONAB, 2004).

1.3 SIMBIOSE COM LEGUMINOSAS

Os organismos diazotróficos, possuem diferentes tipos de associações com diversas plantas. Os de vida livre são capazes de fixar N₂ independente de um hospedeiro superior, utilizando energia disponível no ambiente. Essas bactérias podem estar presentes na região rizosférica (bactérias associativas), ou mesmo dentro dos tecidos das raízes, colmos e folhas das plantas (bactérias endofíticas) (Postgate, 1982). As bactérias da ordem Rhizobiales formam estruturas altamente específicas nas raízes de diversas espécies de plantas da família

leguminosae, onde ocorre a conversão do N_2 a NH_3 , que é, então, incorporada em diversas formas de N orgânico. Durante a reação de redução do N_2 , a dinitrogenase é auxiliada por uma terceira molécula transportadora de elétrons, a ferridoxina, a qual, na sua forma reduzida, transfere um elétron para a unidade Fe-proteína que, então reduzida, doa o elétron recebido para a MoFe-proteína, a qual acumula os elétrons até que ocorram oito transferências concentrando oito elétrons, os quais são necessários para que a redução do N_2 à NH_3 ocorra (Hungria et al., 1997; Morgante, 2003). A reação pode ser expressa como: $N_2 + 8 e^- + 8 H^+ + 16 ATP \rightarrow 2 NH_3 + H_2 + 16 ADP + 16 Pi$.

A formação de nódulos é um processo complexo, que ocorre em várias etapas, envolvendo mudanças fisiológicas e morfológicas, tanto na célula hospedeira, como na bactéria. As mudanças na bactéria visam, principalmente, o recebimento de fontes de carbono da planta hospedeira, para prover o ATP e poder redutor necessários para o processo de FBN. As mudanças, tanto morfológicas como bioquímicas ocorridas na planta hospedeira, visam assimilar a NH_3 produzida pelas bactérias (Hungria et al., 1994).

Para que ocorra a formação dos nódulos, ambos, bactéria simbiótica e planta hospedeira, desenvolveram um complexo sistema de interação, mantendo uma constante comunicação molecular. A grande complexidade do processo está nas numerosas trocas de sinais moleculares entre a bactéria e a planta durante o processo de infecção, o desenvolvimento e o funcionamento dos nódulos com, aproximadamente, 100 genes da bactéria e da planta hospedeira envolvidos neste processo (Vance, 2001). As alterações na bactéria visam, especialmente, o recebimento de fontes de carbono da planta hospedeira, para prover o ATP e o poder redutor necessários para o processo da fixação biológica (Vance, 2001). Esse sistema, denominado simbiose, faz com que bactérias simbióticas vivendo saprofiticamente no solo percebam sinais químicos sintetizados pela planta hospedeira, geralmente flavonóides, os quais fazem com que as bactérias sejam atraídas em direção às raízes da planta, por quimiotactismo positivo (Drozdowicz, 1997).

O estabelecimento da simbiose rizóbio-leguminosa inicia-se com a excreção, pela planta hospedeira, de compostos de ação quimiotática sobre o rizóbio (Hungria et al., 1997), facilitando a colonização rizosférica e estimulando o crescimento da bactéria (Hungria e Stacey, 1997). Geralmente, estes compostos excretados são açúcares, aminoácidos e ácidos dicarboxílicos que promovem a adesão dos rizóbios aos pêlos radiculares das plantas, num processo relativamente estável e irreversível, que ocorre em duas etapas: Na primeira etapa, as bactérias isoladas se aderem às que estão presas aos pêlos radiculares (Hungria et al.,

1997) e, simultaneamente a essa adesão, tem início uma troca de sinais moleculares entre o microssimbionte e a planta hospedeira, que libera compostos fenólicos, geralmente flavonóides, e também quaternários de amônia denominados coletivamente de betaínas, que são responsáveis pela indução de transcrição de genes bacterianos, essenciais à nodulação (genes *nod*, *nol* e *noe*) (Brenner e Winans, 2005). Na segunda etapa de troca de sinais moleculares, os rizóbios produzem sinais específicos para cada planta hospedeira, por meio de genes de nodulação, conduzindo à produção dos fatores de nodulação, ou fatores Nod, que são oligossacarídeos lipoquitínicos (LCOs) que induzem modificações radiculares no estágio de pré-infecção, como: formação de raízes curtas e grossas, deformação, encurvamento e aumento no número de pêlos radiculares, seguido por invaginação da parede celular e formação de um cordão de infecção no interior do pêlo radicular. Associado com a infecção ocorre a indução de divisão celular nas células do córtex e a formação do primórdio nodular, cujo desenvolvimento dá origem ao nódulo (Pichon et al., 1992). A partir do estabelecimento do nódulo radicular, as bactérias que se encontram dentro das células radiculares hospedeiras param de se multiplicar, aumentam de tamanho e sofrem várias alterações bioquímicas para se diferenciarem em bacterióides, formas capazes de fixar N₂. Quando o nódulo está formado são sintetizadas enzimas relacionadas com a quebra da tripla ligação do N e com a assimilação do N₂ fixado (Schultze e Kondorosi, 1998).

A enzima-chave do processo de FBN é a dinitrogenase, que é formada por duas unidades protéicas, a ferro-proteína (Fe-proteína) e a molibdênio-ferro-proteína (MoFe-proteína), ambas capazes de transportar elétrons. A síntese da nitrogenase é codificada pelos genes estruturais *nifD*, *K* e *H*. Sua expressão gênica está associada à síntese das duas metaloproteínas que compõem a dinitrogenase, sendo a Fe-proteína codificada pelo gene *nifH* e a MoFe-proteína codificada pelos genes *nifD* e *K* (Halbleib e Ludden, 2000).

Em *B. japonicum*, os genes envolvidos na FBN estão organizados em uma região no DNA cromossomal denominada de “ilha simbiótica” de, aproximadamente, 410 kb, a qual contém os genes de nodulação e fixação de nitrogênio (Göttfert et al., 2001; Kaneko et al., 2002).

Os genes “essenciais” à infecção e formação do nódulo podem ser divididos em duas classes. Uma classe inclui os genes envolvidos na composição da superfície celular da bactéria, tais como os genes determinantes da síntese de exopolissacarídeos (genes *exo*), lipopolissacarídeos (genes *Ips*) e 1,2-b-glucanos (genes *ndv*), ligados à aderência aos pêlos radiculares e endocitose nas células do córtex.

A segunda classe dos “genes essenciais” consiste dos genes da nodulação-

“*nod/ nol / noe*”, que resultam em alterações no processo de infecção, quando mutados. A inativação dos genes desta classe pode resultar em diversos fenótipos na planta, como ausência de nodulação (Nod⁻), nodulação tardia, porém efetiva (Nod^d Fix⁺), ou uma mudança na faixa de hospedeiros.

Alguns genes *nod* apresentam um grau elevado de homologia, podendo ser trocados entre diferentes espécies e biovars de rizóbios, sem que percam suas propriedades funcionais. Os genes *nod* incluídos nesta categoria são designados como “genes *nod* comuns” (Djordjevic et al., 1985; Kondorosi et al., 1989). Por outro lado, outros genes *nod* estão envolvidos na nodulação de um hospedeiro específico e são, por isso, chamados de “genes *nod* específicos do hospedeiro”- *hsn* (Kondorosi et al., 1989).

Alguns genes de nodulação são “induzíveis”, por exemplo, tais como os genes *nodA*, *nodB* e *nodC*, os quais não são expressos constitutivamente, precisando, de um indutor (Mulligan e Long, 1985). Muitos genes estão em “clusters”, como é o exemplo do operon *nodABCSTUIJnolMNOnodZ*, encontrado em *B. japonicum*, com tamanho de 12 kb que contém 12 genes individuais (Surin e Downie, 1988; Schwedock e Long, 1989; Baev et al., 1991; Rivilla e Downie, 1994).

Em muitos rizóbios, os flavonóides também induzem a secreção de proteínas extracelulares através do sistema tipo III de secreção (TTSS). Essas proteínas secretadas, chamadas Nops (nodulation outer proteins), têm um papel na nodulação específica em uma variedade de leguminosas. Em torno de vinte proteínas de membrana são utilizadas pela bactéria para este sistema de secreção. A maquinaria do TTSS é bem conservada entre diferentes bactérias, sendo identificados genes que codificam para este sistema em *Bradyrhizobium japonicum* USDA110. Alguns destes genes são necessários para o aparato da biosíntese do sistema TTSS e outros codificam proteínas secretoras, tais como NopB (anteriormente NolB), Nopx (anteriormente Nolx), NopL (anteriormente y4xL), NopP (anteriormente y4yP), e NopA. O cluster *tts* inclui o locus *nolXWBTUV*, embora *B japonicum* USDA 110 não tenha o gene *nopX*. O gene *TtsI* está envolvido na regulação transcricional do cluster *tts* e em mutantes para este gene, diferenças fenotípicas na simbiose foram observadas dependendo da planta hospedeira. A anulação da secreção do fator Nop culminou com a diminuição do número de nódulos formados com *Macropitilium atropurpureum* e da nodulação com a variedade Williams de soja americana (Krause et al., 2002).

1.4 CARACTERÍSTICAS DA ESTIRPE DE *BRADYRHIZOBIUM JAPONICUM* CPAC 15 (=SEMIA 5079) MICROSSIMBIONTE DA SOJA

As estirpes de *B. japonicum* e *B. elkanii* selecionadas pela pesquisa brasileira nesses últimos 40 anos e utilizadas nos inoculantes comerciais têm conseguido suprir as necessidades de N da cultura da soja. Atualmente quatro estirpes são recomendadas para o uso em inoculantes comerciais, a SEMIA 587 (recomendada no período de 1968 a 1975 e novamente desde 1979) e a SEMIA 5019 (=29w), recomendadas desde 1979. Essas duas estirpes foram de grande importância para o estabelecimento da soja na Região dos Cerrados, na década de 1980 (Peres e Vidor, 1980; Vargas e Hungria, 1997; Hungria et al., 2005; Hungria et al., 2006).

A estirpe CPAC 15 (= SEMIA 5079, =566a, =DF 24) pertence ao mesmo sorogrupo da SEMIA 566. Em relação ao histórico da SEMIA 566, muitos inoculantes americanos empregados na década de 1960 não se mostraram eficientes e, então, foi realizado o isolamento da estirpe "adaptada" SEMIA 566 (= BR 40) (SEMIA, "Seção de Microbiologia Agrícola", nomenclatura utilizada para a Coleção Nacional de Rizóbios, da FEPAGRO - Fundação Estadual de Pesquisa Agropecuária, Porto Alegre, RS, em conjunto com o MIRCEN, Microbiological Resources Centre Network), que pode ser considerada como uma das primeiras "estirpes brasileiras".

A SEMIA 566 foi isolada em 1966, de um nódulo da cultivar Hardee, em vaso de Leonard que havia recebido inoculante americano distribuído pela firma Dixie Inoc, provavelmente um inoculante da Nitragin (Milwaukee, E.U.A.), quando se buscava superar problemas de nodulação a campo com essa cultivar. Essa estirpe era capaz de nodular efetivamente as cultivares utilizadas na época, sendo recomendada para o uso em inoculantes comerciais no período de 1966 a 1978. A CPAC 15 é uma variante natural da SEMIA 566, obtida a partir do reisolamento, em solos do DF, de estirpes vários anos após a última inoculação, buscando aquelas adaptadas, mas com elevada capacidade de FBN e competitividade (Vargas et al., 1992; Hungria e Vargas, 2000; Vargas e Hungria, 2000; Hungria et al., 2006).

A estirpe SEMIA 566, bem como sua variante CPAC 15 são, sem dúvida, as mais competitivas estabelecidas em solos brasileiros. Essa estirpe pertencente ao mesmo sorogrupo da USDA 123 (=311b123, =TAL 376, =ACCC15036) (Mendes et al., 2004), que foi isolada em 1960 de um nódulo de soja em Iowa, nos E.U.A., sendo considerada indígena

naquele país (Keyser e Cregan, 1987). Schmidt et al. (1986) propuseram o termo sorogrupo 123 porque, na verdade, ele consiste de três grupos sorologicamente relacionados, o 123, o 127 e o 129. As estirpes relacionadas sorologicamente ao sorogrupo 123, têm sido relatadas como sendo as mais competitivas da região do meio-oeste dos E.U.A. ocupando, caracteristicamente, de 60% a 80% dos nódulos formados (Ham et al., 1971; Cregan et al., 1989; Weber et al., 1989), sendo muito difícil introduzir novas estirpes em solos com estirpes desse sorogrupo (Ham et al., 1971). Além disso, esse sorogrupo domina a nodulação da soja nos E.U.A., independente da prática da inoculação e da cultivar utilizada (Ham et al., 1971) e a competitividade não parece estar relacionada à colonização superior dessa estirpe na rizosfera (Moawad et al., 1984). Também há relatos de ocorrência desse sorogrupo no Canadá (Semu e Hume, 1979) e na Coreia (Kang et al., 1991). Contudo, é importante salientar que a SEMIA 566 difere da USDA 123 quanto aos perfis de proteína e LPS. (M. Hungria, dados não publicados).

No Brasil, há relatos de que, nos primeiros dois anos, a SEMIA 566 e a CPAC 15 apresentam baixa competência saprofítica, contudo, logo depois se estabelecem fortemente (Freire et al., 1983; Vargas e Hungria, 1997; Mendes et al., 2004). Relatos semelhantes foram feitos para a USDA 123 nos E.U.A., sugerindo que a estirpe deva residir no solo por um determinado período antes de dominar os nódulos (Streeter, 1994). No Brasil, estirpes desse sorogrupo foram, inclusive, recuperadas em sítios distantes, no RS, DF e AM, mesmo em áreas nunca cultivadas anteriormente, confirmando a sua competência saprofítica e o transporte dessa estirpe, provavelmente por correntes de vento, ou por implementos agrícolas (Vargas e Hungria, 1997; Ferreira e Hungria, 2002). Um exemplo dessa capacidade foi, recentemente, mostrado por Mendes et al. (2004), em um ensaio conduzido por sete anos na Região dos Cerrados. No quarto e sexto anos desse experimento, como consequência da dispersão, estirpes do sorogrupo SEMIA 566 foram encontradas em todas as parcelas, mesmo naquelas que haviam sido inoculadas com as estirpes SEMIA 587, SEMIA 5080 e SEMIA 5019. Esse sorogrupo dominou a nodulação ocupando, em média, mais de 50% dos nódulos dos tratamentos onde nunca havia sido incluída no inoculante (Mendes et al., 2004).

Vários estudos sobre variabilidade morfológica, fisiológica, genética e simbiótica resultantes da adaptação de estirpes do sorogrupo SEMIA 566 a solos brasileiros foram conduzidos por nosso grupo de pesquisa nestes últimos anos, permitindo o acúmulo de um grande conhecimento sobre essa bactéria (Hungria e Vargas, 1996; Nishi et al., 1996; Boddey e Hungria, 1997; Hungria et al., 1998, 1999; Hungria e Vargas, 2000; Vargas e Hungria, 2000).

1.5 GENOMAS DE BACTÉRIAS DIAZOTRÓFICAS SIMBIÓTICAS

Nos últimos sete anos, enormes avanços no conhecimento de genética de bactérias vêm sendo obtidos, sem sombra de dúvida, pelo seqüenciamento de genomas. No ano de 2000 foram depositados, no banco de genes do “National Center for Biotechnology Information” (NCBI), cinco genomas de bactérias: *Ureaplasma urealyticum*, *Borrelia burgdorferi*, *Chlamydia muridarum*, *Chlamydomophila pneumoniae* e *Corynebacterium stricutum* (NCBI, 2007a). Em julho de 2007 já estão depositados os genomas de 566 organismos, sendo 540 de bactérias, (com mais de uma estirpe para algumas espécies), e 406 genomas bacterianos incompletos. Contudo, a maioria desses genomas é referente a bactérias patogênicas ao homem, ou extremofílicas (NCBI, 2007a).

Os genomas de bactérias são comparativamente pequenos em relação ao genoma dos organismos eucariontes, variando de 580.074 pares de bases (pb), em *Mycoplasma genitalium*, a 9.105.828 pb em *B. japonicum* (NCBI, 2007a). O banco de genes do NCBI conta, ainda, com seqüências parciais de genomas e plasmídeos de bactérias e o seqüenciamento de outros genomas que estão em progresso (NCBI, 2007a). Além disso, outra estratégia vem sendo utilizada, com o seqüenciamento parcial de genomas, permitindo obter valiosas informações. Esse foi o caso, por exemplo, da determinação de cerca de 15 % do genoma de *Rhizobium* NGR 234 (Viprey et al., 2000).

Até o presente momento, foram depositados apenas sete genomas completos de rizóbios: *Mesorhizobium loti* estirpe MAFF303099, incluindo dois plasmídeos (Kaneko et al., 2000a, 2000b), *Sinorhizobium meliloti* estirpe 1021, também incluindo dois plasmídeos (Capela et al., 2001; Galibert et al., 2001), *Bradyrhizobium japonicum* estirpe USDA 110 (Kaneko et al., 2002), *Rhizobium etli* CFN42 (González et al., 2006), incluindo seis plasmídeos, *Rhizobium leguminosarum* biovar *viciae* (Rlv) estirpe 3841, também incluindo seis plasmídeos (Young et al., 2006) e *Bradyrhizobium* estirpe BTAi1 com um plasmídeo e a estirpe ORS278 (Giraud et al., 2007). Os plasmídeos simbióticos de *Rhizobium* NGR 234 (pNGR234a) (Freiberg et al., 1997) e um pequeno plasmídeo críptico de *S. meliloti* (pRm1132f) (Barran et al., 2001) também estão depositados. Finalmente, há indicação do seqüenciamento incompleto de *Mesorhizobium* sp. estirpe BNC1, com 3 plasmídeos (dados não publicados) (NCBI, 2007b).

Nos gêneros *Rhizobium* e *Sinorhizobium*, o plasmídeo *Sym* contém a maioria dos genes responsáveis pela nodulação e fixação do N₂. O seqüenciamento do

plasmídeo simbiótico de *Rhizobium* sp. NGR234 identificou a presença de 416 genes funcionais, incluindo 26 genes para nodulação e 26 genes para a fixação do nitrogênio (Freiberg et al., 1997).

A seqüência completa de *Mesorhizobium loti* MAFF303099 mostrou que seu genoma consiste de um único cromossomo, de 7 Mb, e dois plasmídeos, pMLa, de 351,9 kb e pMLb, de 208,3 kb. O cromossomo contém 6.752 genes codificadores de proteínas, além dos genes para RNAs ribossômicos e transportadores. Dos genes funcionais, 54% mostraram similaridade com genes de funções conhecidas. Um segmento de DNA de 611 Kb foi identificado como uma possível ilha simbiótica, contendo 30 genes para fixação de N₂ e 24 genes para nodulação. Os genomas dos plasmídeos pMLa e pMLb contêm, respectivamente, 320 e 209 genes potenciais para diferentes funções biológicas. Contudo, apenas um único gene plasmidial para nodulação foi identificado (Kaneko et al., 2000).

O genoma de *S. meliloti* estirpe 1021 também foi completamente seqüenciado, consistindo de um cromossomo de 3,7 Mb e dois plasmídeos, *pSymA* de 1,4 MB e *pSymB* de 1,7 Mb. Foram anotados 6.204 genes funcionais, sendo que 59,7% podem ter suas funções determinadas por comparações em bancos de dados. O cromossomo de *S. meliloti* é típico de bactérias aeróbicas heterotróficas. O plasmídeo *pSymB* contém genes que codificam as enzimas que permitem a bactéria metabolizar uma variedade de pequenos compostos encontrados no solo ou na rizosfera da planta, além de permitir a síntese de polissacarídeos envolvidos na colonização do hospedeiro. O plasmídeo *pSymA* contém a maioria dos genes relacionados à nodulação e FBN, conferindo a capacidade para a bactéria colonizar ambientes de baixa tensão de oxigênio e metabolizar uma variedade de formas químicas de nitrogênio, inclusive a forma molecular N₂ (Galibert et al., 2001).

No gênero *Bradyrhizobium*, os genes envolvidos no evento de nodulação (genes *nod*, *nol* e *noe*), fixação do nitrogênio (genes *fix*) e síntese de nitrogenase (genes *nif*) estão localizados no cromossomo. A seqüência completa do genoma de *B. japonicum* estirpe USDA110 confirmou que seu genoma consiste de um único cromossomo de 9 Mb, sem a presença de plasmídeos. O cromossomo contém 8.317 genes codificadores de proteínas, além dos genes para RNAs ribossômicos e transportadores. Dos genes funcionais, 52% puderam ter suas funções determinadas por comparações em bancos de dados, 34% dos genes apresentaram similaridade com *M. loti* e *S. meliloti* e 23% são únicos desta espécie. (Kaneko et al., 2002) Um segmento de 681 kb foi identificado como uma possível ilha simbiótica, incluindo a região cromossômica de 410 kb onde estão agrupados os genes de nodulação e fixação. Foram atribuídos 655 genes codificantes de proteína para essa região, sendo que 301

deles estão envolvidos na fixação simbiótica de N₂. Foram anotados 167 genes para transposases, dos quais 100 foram localizados na ilha simbiótica. Foi observado, em 14 posições diferentes do genoma, um segmento de DNA de 4 a 97 kb inserido nos genes de tRNA, gerando duplicação parcial nos genes alvos do tRNA. Estas observações sugerem uma plasticidade elevada no genoma de *B japonicum*, a qual, provavelmente, ocorre devido ao rearranjo complexo do genoma, incluindo transferência horizontal e inserção de vários elementos de DNA e recombinação homóloga (Kaneko et al., 2002).

Ainda no gênero *Bradyrhizobium*, os genomas de duas estirpes fotossintéticas, ORS278 e BTAi1, também seqüenciados por inteiro, não apresentaram genes para nodulação, *nodABC*, indicando que este processo seria realizado por outros genes, envolvidos na simbiose planta-hospedeiro, e em interações patogênicas, como genes de síntese e degradação de fitormônios, modificações no LPS e biossíntese de exopolissacarídeos, em ambas as estirpes. Na estirpe BTAi1, o genoma seqüenciado apresentou um cromossomo de, aproximadamente, 8,6Mb, com 7.729 CDSs, sendo 60,1% dos genes com função conhecida, 31,8% conservados hipotéticos e 8,1% hipotéticos, além de um plasmídeo de 800 kb. Já na estirpe ORS278, foi encontrado um cromossomo com 7,4 Mb, contendo 6.752 CDSs, com 59% dos genes com função conhecida, 27,7% hipotéticos conservados, 13,3% hipotéticos e nenhum plasmídeo. A presença de inúmeros elementos genéticos móveis, indica plasticidade elevada destes genomas. O genoma de ORS278 e BTAi1 contém, respectivamente, 21 e 29 possíveis ilhas genômicas adquiridas por transferência horizontal (HAIs). Estas ilhas podem conferir vantagens funcionais e de adaptação, para que estas bactérias ocupem nichos simbióticos, ou de vida livre (Giraud et al., 2007).

O genoma de *Rhizobium leguminosarum* bv. *viciae*, com um total de 7,75MB, contém 65% do seu tamanho em um cromossomo circular de 5MB, e 2,5MB, estão divididos em seis plasmídeos, pRL12, pRL11, pRL10, pRL9, pRL8 e pRL7. Nesses plasmídeos, com exceção do pRL7, também estão contidos genes considerados essenciais, como de transcrição, biossíntese de ribossomos, chaperoninas e divisão celular. No caso do plasmídeo pRL7, mais de 80% dos seus genes são exógenos ou de função desconhecida e 28% dos genes, são transposases. Foi observada uma homologia com *Bradyrhizobium japonicum* em 264 genes, incluindo, entre eles, genes relacionados à FBN. Muitos destes genes como, *nifHKDN*, *nifAB*, *fixABCX* e *fixNOQPGHIS*, estão situados no pRL10, caracterizando-o como plasmídeo simbiótico. (Young et al., 2006)

A seqüência genômica completa de *Rhizobium etli* bv. *phasoli* CFN42

consiste de um cromossomo circular de 4,3M e mais seis plasmídeos, perfazendo um total de 6,5MB. Mais de 4.200 CDSs (71%), pertencem a alguma classe funcional, sendo 23% de CDSs hipotéticas conservadas e 6% hipotéticas. O seqüenciamento do plasmídeo p42d (*pSym*) (González, et al., 2003), identificou a presença da maioria dos genes relacionados à FBN, entre estes existem genes que codificam proteínas que participam da biosíntese e modificação da fucose e manose (*nodL*, *noeL*, *nolK*, *noeK*, *noeJ*, e *nodN*), transportador (*nodT*), reguladores *nodVW* e *nfeD*, o supressor do *nodVW*, e dois componentes reguladores, *nwsAB*. Em *S. fredii*, *S. melioli*, *M. loti* e *B. japonicum*, estes genes não são essenciais para simbiose, e estão envolvidos na competição e eficiência de nodulação nas plantas hospedeiras, mas ainda não foi estabelecida sua função em *R. etli*. Outros 27 genes foram encontrados em diferentes regiões genômicas, relacionados à nodulação (González et al., 2006).

Não restam dúvidas sobre a importância das informações obtidas pelo seqüenciamento de genomas de microrganismos. Mas conforme já comentado, no caso das bactérias, a maioria dos genomas já seqüenciados é de patogênicas ou extremofílicas. É importante, portanto, levantar mais informações sobre bactérias de importância agrícola.

O Brasil tem se destacado em pesquisas na área agrícola, razão pela qual esta dissertação propõe o seqüenciamento parcial de uma estirpe comercial, *Bradyrhizobium japonicum*, bactéria envolvida no processo de FBN com uma das leguminosas mais importantes para o País, a soja. Realizar o seqüenciamento completo de um genoma da bactéria, porém, é um processo caro e demorado, que está muito além da capacidade da maioria dos laboratórios. Além disso, no seqüenciamento completo, a fase de fechamento do genoma, em geral, consome de 50% a 60% do tempo do projeto, com custos elevados (Simpson, 2001). Ainda que técnicas inovadoras possam reduzir o tempo de fechamento de um genoma, como por exemplo, o método PACE (PCR-assisted contig extension), desenvolvido durante o fechamento de *Chromobacterium violaceum* (Carraro et al., 2004), o dinheiro e tempo gastos no processo ainda são muito elevados.

Uma estratégia alternativa, de seqüenciamento parcial de genomas, vem sendo investigada há alguns anos. Já foi demonstrado que informações importantes e representativas da estrutura e conteúdo de genomas podem ser obtidas a partir do seqüenciamento parcial e ao acaso de bibliotecas do tipo “shotgun”, feitas a partir do DNA total dos organismos, incluindo, por exemplo, fungos (Avalos et al., 1996) e bactérias (Kim et al., 1993), entre outros.

Essa estratégia, de seqüenciamento parcial por “shotgun”, foi utilizada na estirpe de *Rhizobium* ANU265, que corresponde a uma derivativa da estirpe NGR 234, mas

curada do plasmídeo simbiótico (Viprey et al., 2000). A estirpe NGR 234 é reconhecida por seu amplo espectro de nodulação, incluindo várias plantas hospedeiras. O DNA genômico da estirpe ANU265 foi utilizado para construir bibliotecas M13 com insertos de 0,9 a 1,5 kb. Foram seqüenciados 2.856 clones ao acaso, dos quais 80% (2.275) produziram seqüências de alta qualidade (PHRED igual ou superior a 20), com comprimento médio de 253 pb. Desse modo, mais de 575 kb do total de nucleotídeos foram obtidos, correspondendo a, aproximadamente, 10% do genoma da ANU265. A busca de homologia em bancos de dados resultou na identificação de 1.130 seqüências codificando proteínas putativas e a análise dessas seqüências, em comparação com outros genomas, sugere que esse número foi suficiente para representar o genoma dessa bactéria. Além disso, diversas seqüências novas foram identificadas, indicando que informações relevantes podem ser obtidas, a um custo acessível, pelo seqüenciamento parcial do genoma (Viprey et al., 2000), que também vem sendo chamado, por nosso grupo de pesquisa, de panorama genômico. Dentre os genes simbióticos de interesse identificados no trabalho de Viprey et al. (2000), pode-se citar o loci envolvido na biossíntese de exopolissacarídeos (EPS), genes de síntese de polissacarídeos capsulares (K-antígenos), lipopolissacarídeos, genes de fixação como *fixN*, o locus *nwsA*, que codifica uma quinase putativa que parece estar envolvida na expressão de genes de nodulação em *Bradyrhizobium*, entre outros.

Com o seqüenciamento completo do genoma de *Bradyrhizobium japonicum* USDA110, duas abordagens podem ser feitas a partir do seqüenciamento parcial do genoma de outra estirpe dessa espécie: uma refere-se à busca de homologia e a outra à genômica comparativa, através de programas especiais, tais como o GINGA (Paschoal et al., 2006).

2 OBJETIVOS

2.1 OBJETIVO GERAL

Realizar a cobertura de 10 a 15% do genoma da estirpe CPAC 15 (= SEMIA 5079) de *Bradyrhizobium japonicum*, procurando genes relacionados à capacidade saprofítica, à competitividade e à eficiência do processo de fixação biológica do nitrogênio.

2.2 OBJETIVOS ESPECÍFICOS

- 1) Realizar o seqüenciamento parcial (10 a 15%) do genoma da estirpe CPAC 15 (= SEMIA 5079) de *Bradyrhizobium japonicum*, utilizada em inoculantes comerciais para a cultura da soja no Brasil.
- 2) Identificar e anotar os genes seqüenciados da estirpe CPAC 15 de *B. japonicum*.

3 MATERIAL E MÉTODOS

3.1 ESTIRPE DE *BRADYRHIZOBIUM JAPONICUM*

Foi utilizada a estirpe de *B. japonicum* CPAC 15 (=SEMIA 5079), microssimbionte da soja. A estirpe é utilizada em inoculantes comerciais no Brasil desde 1992, e não oferece risco para o meio ambiente e saúde humana, encaixando-se nos padrões de biossegurança (NB-1). A estirpe consta do banco de germoplasma de rizóbios da Embrapa Soja.

3.2 EXTRAÇÃO DO DNA GENÔMICO

Inicialmente inoculou-se uma cultura de *B. japonicum* CPAC 15 em 25 ml de meio YM líquido, e a cultura (pré-inóculo) foi mantida sob agitação a 28°C por aproximadamente 72 horas. Após este período inoculou-se um ml do pré-inóculo em 500 ml de meio YM líquido e a cultura foi mantida sob agitação a 28°C por aproximadamente 72 horas (até a fase exponencial de crescimento). Para a extração do DNA genômico procedeu-se à centrifugação da cultura a 5.000 rpm por 10 min e a ressuspensão do precipitado celular em 50 mL de TES (50 mM Tris-HCl, pH 8,0; 20 mM EDTA, pH 8,0; 200 mM NaCl). A suspensão celular foi, então, centrifugada a 5.000 rpm por 10 min e ressuspensa em 10 mL de TES. Foram adicionados 100 uL de lisozima (200 ug mL⁻¹) e 30 ul de RNase (1 mg mL⁻¹) e, em seguida, incubou-se a suspensão celular a uma temperatura de 30°C por 4 h. A seguir, foram adicionados 10 uL de proteinase K (20 ug uL⁻¹) e a mistura foi incubada durante a noite à 37°C. Um volume igual (10 mL) de fenol equilibrado pH 8,0 foi adicionado ao tubo, que foi gentilmente invertido por 30 min. A mistura DNA/fenol foi, então, centrifugada a 10.000 rpm por 15 min em temperatura ambiente. O sobrenadante (fase aquosa) foi transferido para um outro tubo e a este adicionou-se um volume igual de fenol e a mistura foi homogenizada, gentilmente, por 20 min. A seguir, adicionaram-se 5 mL de clorofórmio e homogenizou-se gentilmente por inversão por 10 min. A mistura DNA/fenol/clorofórmio foi,

então, centrifugada por 15 min à temperatura ambiente. A fase aquosa superior foi transferida para um novo tubo e adicionou-se igual volume de isopropanol. A mistura DNA/isopropanol foi homogenizada gentilmente por inversão por 5 min e centrifugada a 5.000 rpm por 15 min. O isopropanol foi removido e ao pélete de DNA foi adicionado 1 mL de etanol 70%, seguido por centrifugação a 5.000 rpm por 5 min. O etanol 70% foi, então, removido e o DNA foi seco em temperatura ambiente por 20 min. O DNA foi ressuspensão em 50 uL de TE (10 mM Tris, pH 8,0; 0,1 mM EDTA) e incubado à 4°C por um período de aproximadamente 12 horas. A quantidade e a qualidade de DNA foram verificadas por eletroforese em gel de agarose 1% usando marcadores padrões de DNA como parâmetros para a quantificação de DNA. O DNA de alto peso molecular e de boa qualidade foi utilizado para a construção das bibliotecas do tipo “shotgun”.

3.3 CONSTRUÇÃO DE BIBLIOTECAS “SHOTGUN”

O preparo das bibliotecas “shotgun” envolveu a purificação do DNA genômico, conforme descrito no item 3.2, e a fragmentação desse DNA ao acaso, por meio mecânico (nebulização - 2,0 kgf.cm⁻²). Após a fragmentação, as extremidades do DNA foram reparadas com o uso das enzimas T4 DNA polimerase e Klenow de *Escherichia coli*.

Para a posterior ligação com o vetor pUC 18, procedeu-se à fosforilação destes fragmentos, com o uso da enzima “polinucleotídeo kinase” (PNK) (New England Biolabs) enquanto que, para a ligação com o vetor TOPO, procedeu-se à defosforilação (Fleischmann et al., 1995).

Fragmentos com tamanho entre 1,5-3,0 kb foram separados em gel 1% de agarose “low melting”, grau analítico (Promega).

Após a recuperação destes fragmentos a partir do gel de agarose com “kit” de extração de DNA de agarose da Qiagen (QIAquick gel extraction Kit), procedeu-se à ligação ao vetor pUC18, usando a enzima T4 DNA ligase. O vetor foi previamente digerido com a enzima de restrição *Sma*I (Amersham Pharmacia Biotech), e defosforilado com a enzima BAP (Bacterial Alkaline Phosphatase) (Promega). O DNA genômico fragmentado foi também clonado no vetor pCR4Blunt-TOPO, com o uso do “kit” “TOPO Shotgun Subcloning Kit” (Invitrogen), segundo as instruções do fabricante (Fleischmann et al., 1995).

O DNA clonado foi utilizado para transformação de células de *Escherichia*

coli TOP10 (Invitrogen), através da técnica de eletroporação (Sambrook et al., 1989). As células foram plaqueadas em meio LB contendo ampicilina ($100 \mu\text{g mL}^{-1}$, Sigma), IPTG/X-Gal e X-Gal (Invitrogen) e crescidas durante a noite a 37°C . Os clones recombinantes produzidos foram, então, transferidos para placas do tipo ELISA de 96 poços, contendo meio líquido “Circle Grow” (BIO 101), suplementado com ampicilina ($100 \mu\text{g mL}^{-1}$) e glicerol a 80% (Sigma, 99.5%, para biologia molecular) e armazenados a -70°C .

3.4 SEQÜENCIAMENTO DOS CLONES

As colônias individuais das bibliotecas obtidas foram inoculadas em microplacas com 96 poços contendo meio “Terrific Grow” (contendo 12 g L^{-1} de bactotripton, 24 g L^{-1} de extrato de levedura e 4 mL L^{-1} de glicerol, Sigma) que, depois de autoclavado, recebeu ampicilina ($100 \mu\text{g mL}^{-1}$) previamente esterilizada por filtração (filtro $0,2 \mu\text{m}$, Millipore) e então, crescidas com agitação de 300 rpm, por 16 horas, a 37°C . O DNA foi extraído pelo método usual de lise alcalina (Sambrook et al., 1989), com uma modificação no final do procedimento, que é o da passagem do sobrenadante por filtros múltiplos (MultiScreen, Millipore) antes da precipitação do DNA (Vasconcelos et al., 2003). O DNA purificado foi ressuspensão em H_2O milliQ e verificado em gel de agarose (Invitrogen) a 0,8%, conforme descrito por Sambrook et al. (1989). O DNA foi precipitado, procedendo-se ao seqüenciamento utilizando o “kit” “DYEnamicTM ET dye terminator cycle sequencing (MegaBACETM)” (Amersham Pharmacia Biotech). As reações de PCR (“polymerase chain reaction”, reação em cadeia da polimerase) foram realizadas com os oligonucleotídeos “Universal” (‘GAC GTT GTA AAA CGA CGG CCA GT’) e “Reverso” (‘TTT CAC ACA GGA AAC AGC TAT GAC’) (Invitrogen ou RW genes), para se conseguir a amplificação dos genes. Os produtos da reação foram analisados em um seqüenciador automático (MegaBace1000, Amersham) pelo método dos terminadores fluorescentes, onde o último nucleotídeo incorporado na extremidade de alongamento da cadeia de DNA é marcado com uma molécula fluorescente, que é excitada pelo laser do aparelho, emitindo um comprimento de onda característico daquele fluorocromo, que por sua vez é específico para uma base nitrogenada. Esse comprimento de onda específico gerado pela excitação da molécula, passa por um conjunto de filtros onde, então, é direcionado para uma câmara específica, sensibilizando-a e gravando a informação da determinada base que foi lida. As seqüências

obtidas foram enviadas para a homepage do laboratório de bioinformática do LNCC, “<http://www.bnf.lncc.br>”, e submetidas a um conjunto de programas de alinhamento de seqüências e construção de possíveis "contigs" (Phred, Phrap e Consed).

3.5 IDENTIFICAÇÃO E ANOTAÇÃO DOS GENES

As seqüências obtidas foram submetidas e armazenadas no Laboratório de Bioinformática do Laboratório Nacional de Computação Científica (LNCC), que foi o responsável pela montagem do genoma. Os eletroferogramas foram transformados em seqüências de bases utilizando o programa “phred”, que também atribui valores de qualidade a cada base e identifica as seqüências repetitivas e aquelas pertencentes a vetores. A montagem das seqüências foi feita utilizando o programa “phrap” (Ewing e Green, 1998; Ewing et al., 1998). A visualização e edição de montagens das seqüências foi realizada usando os programas “Consed” (Gordon et al., 1998) e “phrapview”.

As CDSs “Coding Sequences”, (ORFs “open reading frames”), foram preditas utilizando ferramentas de bioinformática. Essas potenciais regiões codificadoras foram preditas através dos programas “Glimmer” (Delcher et al., 1999). A busca de homologia de cada CDS foi através da comparação com o banco de dados de seqüências do GenBank usando o algoritmo BLAST (Altschul et al., 1990). . (Os programas KEGG (Kyoto Encyclopedia of Genes e Genomes) (Kanehisa e Goto, 2000), COG (Clusters of Orthologous Groups of proteins) (Tatusov et al., 2000), INTERPRO (uma fonte de busca de famílias, domínios e sítios funcionais de proteínas) (Apweiler et al., 2000; Mulder et al., 2003), PSORT (desenvolvido para localização das proteínas) (Nakai e Kanehisa, 1991) e Uniprot (Wu et al., 2006), foram utilizados para a identificação, predição e localização celular das proteínas. As leituras de bases de boa qualidade definiram as “ORFs” de produto conhecido, hipotéticas e conservadas. Essa abordagem foi, recentemente, utilizada com bastante sucesso no estudo do genoma de *C. violaceum* (Vasconcelos et al., 2003) e de duas espécies de *Mycoplasma* (Vasconcelos et al., 2004).

Toda a anotação foi feita com base em metodologia desenvolvida através do “software” para montagem e anotação automática de Genomas: **SABIÁ** (System of Automated Bacterial Integrated Annotation), que integra a montagem e a anotação de genomas (Almeida et al., 2004). A seguir, cada CDS foi conferida manualmente, uniformizada quanto às informações e validada, ou não.

3.6 ANOTAÇÃO MANUAL DOS GENES

Como um processo decorrente e padrão, após o término da anotação automática, utilizou-se o critério de anotação manual. As ferramentas de bioinformática utilizadas estavam todas interligadas pelo *software* SABIÁ. Observado o alinhamento com os bancos de dados, a primeira observação era em relação aos “códon” iniciadores, procurando os que melhor definiam a CDS. A seguir, era realizada a observação nos bancos de dados do COG (genes ortólogos) e do KEGG (metabolismo). A consulta a vários bancos para a identificação dos genes [GenelD (NCBI), GI (NCBI) e UniProtKB/Swiss-Prot (The UniProt Consortium, 2007), bem como bancos específicos de *E. coli*, a bactéria modelo dos genomas de procariotos, incluindo o banco Colibri (Institute Pasteur), Regulon DB (Universidade Autônoma do México) e Wisconsin (com o Institute Pasteur) era realizada nesta fase. Ainda com base no banco de dados KEGG, são fornecidas as informações de similaridade com proteínas e ácidos nucleicos. Informações dos bancos InterPro e GO (Gene Ontology) também estão disponibilizadas na interface de anotação.

Uma das consultas mais importantes na etapa da anotação foi a da similaridade de proteínas e de bases de nucleotídeos do GeneBank, no NCBI, pelo conjunto de algoritmos BLAST. Nessa etapa são verificadas as maiores similaridades com outros organismos, bem como se os genes putativos estão completos, se estes genes são conservados ou hipotéticos, entre outros. Outro banco de grande importância é o UniProtKB/Swiss-Prot, que apresenta uma detalhada descrição das famílias das proteínas, vias metabólicas, funções, organismos relacionados, topologia da seqüência e demais informações relevantes para a classificação das proteínas. Finalmente, a localização celular da proteína a ser predita é inferida pelo banco de dados PSORT ou pelo Uniprot-Swiss-Prot. Famílias de genes ortólogos foram definidas utilizando valores mínimos de "E-value" de 10^{-10} , com cobertura superior a 40%, e 50% de identidade.

4 RESULTADOS E DISCUSSÃO

A estirpe SEMIA 566 de *B. japonicum* pertence ao mesmo sorogrupo da norte-americana USDA 123 e foi amplamente utilizada em inoculantes comerciais brasileiros por 13 anos, desde 1966 (Hungria & Vargas, 2000; Hungria et al., 2006). Posteriormente, a estirpe CPAC 15 — uma variante natural da SEMIA 566 — foi identificada e, por sua superioridade na capacidade de fixação de N₂, em comparação com as demais estirpes comerciais, passou a ser empregada massivamente em inoculantes comerciais desde 1992 (Peres et al., 1993; Vargas et al., 1993; Hungria & Vargas, 2000). A variante natural foi isolada após um longo período de adaptação às condições estressantes dos Cerrados, utilizando soja como planta-isca e, portanto, além da superioridade na capacidade de fixação de N₂ em comparação com a SEMIA 566, a CPAC 15 também é "adaptada" aos solos ácidos tropicais (Hungria & Vargas, 2000). Diversas diferenças morfológicas, fisiológicas e genéticas entre a CPAC 15 e a parental SEMIA 566 já foram relatadas (Hungria et al., 1996, 1998; Nishi et al., 1996; Boddey & Hungria, 1997) e poderiam estar relacionadas à capacidade simbiótica superior e/ou à capacidade saprofítica da estirpe variante.

O sorogrupo SEMIA 566 tem implicações ecológicas importantes nos solos brasileiros, uma vez que é dominante em praticamente todos os levantamentos realizados no Brasil até o presente momento — e o País tem, atualmente, 22 milhões de ha cultivados com essa leguminosa; além disso, o sorogrupo também está estabelecido em áreas que nunca foram inoculadas anteriormente, inclusive nos Cerrados e na Amazônia, onde provavelmente chegou por dispersão (Freire, 1977; Freire & Vidor, 1981; Freire et al., 1983; Vargas et al., 1993; Vargas & Hungria, 1997; Ferreira & Hungria, 2002; Mendes et al., 2004; Ferreira & Hungria, 2002; Mendes et al., 2004). Os problemas de competitividade elevada relacionados ao sorogrupo SEMIA 566/USDA 123 também afetam outros países, como os E.U.A. (Damirgi et al. 1967; Ham et al., 1971; Kvien et al. 1981; Cregan et al., 1989; Weber et al., 1989; Ham et al., 1971; Reyes & Schmidt, 1979), o Canadá (Semu & Hume 1979; Streeter, 1994) e a Coréia (Kang et al., 1991), ressaltando a importância de obter um maior conhecimento de estirpes desse sorogrupo. Por fim, outro fator que reforça a importância de seqüenciar o genoma da CPAC 15 reside nos relatos recentes da existência de taxas elevadas, nos solos brasileiros sob cultivo com a soja, de transferência horizontal de genes simbióticos entre as estirpes desse sorogrupo e outras estirpes de *B. japonicum*, *B. elkanii* e mesmo outras espécies de rizóbios nativas dos solos do cerrados (Batista et al., 2007; Barcellos et al., 2007).

Para realizar o panorama genômico da estirpe CPAC 15 foram construídas nove bibliotecas a partir do DNA genômico, e três foram sequenciadas (BJPM01, BJPM07 e BJPM09) sendo somente as bibliotecas BJPM01 e BJPM09 validadas. A representatividade no genoma, o tamanho esperado do inserto, a ausência de contaminação e a presença excessiva de vetores foram alguns dos critérios utilizados para a validação destas bibliotecas. Para a construção dessas bibliotecas foram utilizados os vetores PUC18 (para as bibliotecas BJPM01 e BJPM07), e TOPO (para a biblioteca BJPM09). A qualidade e as leituras obtidas em cada uma dessas bibliotecas constam no Quadro 1.

A biblioteca BJPM07 não foi validada, pois apresentou baixo número de colônias transformantes, desse modo, foi realizada a leitura de uma única placa (BJPM0701g e BJPM0701b). Na biblioteca BJPM01, foram enviadas 14 placas, sendo quase todos os clones seqüenciados com os oligonucleotídeos “g” (“primer reverse”) e “b” (“primer forward”), resultando no depósito de 756.118 pares de bases (pb) de boa qualidade e sem vetores, correspondendo a 28,76% de todas as bases depositadas. Na biblioteca BJPM09 foram realizadas as leituras de 12 placas, a maioria também em ambas as direções, que resultaram em 604.309 pb de boa qualidade e sem vetores, correspondendo a 31,81% de todas as bases depositadas (Quadro 1).

Neste estudo, foram realizadas 4.328 leituras de clones, que resultaram em 3.683 leituras de boa qualidade, com poucas bases de vetores ($\leq 10\%$), representando mais de 85% do total das leituras. Apenas 645 leituras apresentaram mais de 10% de bases de vetores, representando 14,9%. A leitura desses clones resultou no depósito de 4.576.417 pb, que corresponderiam a 0,51 vezes o tamanho do genoma, estimado em 9.100.000 de pb (tamanho do genoma da estirpe USDA 110 de *B. japonicum*, Kaneko et al., 2002), com 41,19% das bases com boa qualidade (phred ≥ 20) e 29,42% de bases com ótima qualidade (phred ≥ 30). Ainda ocorreram 7,90 % de bases de vetores, sendo a média do tamanho das leituras de 1.057,40 pb, e o comprimento médio das leituras de boa qualidade de 472,91 pb (Quadro 2).

Com base no estudo de Viprey et al. (2000), o seqüenciamento da CPAC 15 foi realizado para que cobrisse de 10 a 15% do genoma. Conforme já comentado, foram obtidas leituras que corresponderiam a 51% do genoma, contudo, pela natureza da biblioteca "shotgun", em que vários fragmentos correspondem a seqüências idênticas, a cobertura real obtida foi de 13,17% (Quadro 2).

A montagem das leituras da CPAC 15 resultou em 1.106 “contigs”, com número de leituras variando de 1 a 16 por “contig” (Quadro 3). Para a grande maioria dos

“contigs” (1.086), correspondendo a 98,19% do total, houve de uma a seis leituras, (dados não mostrados), e somente vinte (1,08%), contêm de 7 a 16 leituras (Quadro 3).

Os “contigs” podem conter leituras em uma única direção, ou leituras ancoradas em ambas as extremidades, geradas pela amplificação dos oligonucleotídeos “b” e “g”, criando um consenso que, associado com uma nota de qualidade igual ou superior a 20 (um erro em 100), aplicada pelo programa “phred”, confere uma formação mais criteriosa destes “contigs”. Como exemplo, dois “contigs” formados por mais de uma leitura são apresentados na Fig. 1.

O “software” SABIA (System of Automated Bacterial Integrated Annotation), desenvolvido pelo LNCC - Laboratório Nacional de Computação Científica – é formado por programas capazes de interagir com diversos bancos de dados, contendo um sistema integrado de ferramentas de bioinformática, todas com acessos diretos a esses bancos. Entre estes bancos podemos citar: Glimmer, Genemark, tRNAScan-SE, Interpro, KEGG, COG, PSORT, GO, UniProt-Swiss-Prot e os programas da família BLAST (Almeida et al., 2004b) (Fig. 7). Todas as possíveis CDSs geradas durante o seqüenciamento que formam um “contig”, são preditas pelos programas Glimmer (<http://www.tigr.org/software/glimmer>) e Gene-Mark (<http://opal.biology.gatech.edu/GeneMark>), e os tRNAs são detectados pelo programa tRNAscan (<http://opal.biology.gatech.edu/GeneMark/>). Cada CDS é submetida a comparações com vários bancos de dados, com a finalidade de se encontrar a melhor homologia.

Para a identificação de uma CDS e sua possível função, inicialmente se torna necessária à identificação de seqüências transcricionais reguladoras, como a região de ligação do ribossomo (RBS). A busca de homologia das CDS é realizada através da comparação de seqüências de nucleotídeos e de aminoácidos (BLASTN e BLASTP), contra seqüências de bancos de dados primários, bem como a identificação de domínios de proteína através do InterPro (<http://www.ebi.ac.uk/interpro/>), classificações funcionais pelos bancos KEGG (<http://www.genome.ad.jp/kegg>) e COG (<http://www.ncbi.nlm.nih.gov/COG/>), localização de proteínas, PSORT (<http://psort.nibb.ac.jp/>), identificação de possíveis transportadores de membrana, TCDB (<http://tcd.b.ucsd.edu/tcdb/>) e da ontologia de genes GO (<http://www.ebi.ac.uk/interpro/>) (Almeida et al., 2004b).

Após esta anotação automática, cada CDS foi anotada manualmente, com base em todas as informações disponibilizadas pelo programa SABIA. Somente após a análise cuidadosa destas informações é que as CDSs foram anotadas manualmente, como de função conhecida, hipotéticas e hipotéticas conservadas.

Das 1.371 CDSs anotadas, 729, ou 53% do total, foram classificadas como apresentando uma função conhecida, 312, ou 23% do total como hipotéticas conservadas (podem envolver CDSs que indicam um domínio funcional, até CDSs sem função conhecida, mas também presentes em outros organismos) e 330, ou 24% do total, como hipotéticas (= ORFans), presentes apenas no genoma estudado (Fig. 2).

A classificação funcional das CDSs pelo banco de dados COG (Tatusov et al., 1997) mostrou genes putativos em quase todas as classes; foi possível a classificação de 797 (58,1%) CDSs da estirpe CPAC 15 de *B. japonicum* dentre as 1.371 CDSs identificadas (Quadro 4).

Considerando 31 CDSs classificadas na categoria J de COG — genes relacionados à tradução — a média do conteúdo de GC de CPAC 15 foi de 64,86%, e levemente inferior (64,66%) se somente os genes ribossomais forem considerados. O conteúdo de GC é, portanto, muito próximo ao observado na estirpe USDA110 – 64,1% (Kaneko et al., 2002). Vinte e seis dessas 31 CDSs apresentaram maior similaridade pelo KEGG com a USDA110, incluindo todas as proteínas ribossômicas: *rimJ* e *M.* (16S), *rpsG*, *K*, *L*, *M*, *P* e *S* (30S), *rplB* e *V* (50S), e a única exceção foi *rpmJ* (50S), a qual apresentou maior similaridade com *Bradyrhizobium* sp. estirpe BTAi1.

Como observado nos genomas de *B. japonicum* USDA 110 (Kaneko et al., 2002), *Sinorhizobium meliloti* (Capela et al., 2001; Galibert et al., 2001), *Mesorhizobium loti* (Kaneko et al., 2000) e *R. etli* (González et al., 2006), ocorreu uma representatividade maior nas categorias COG de metabolismo e transporte de aminoácidos, metabolismo e transporte de carboidratos e transcrição (COG E, G, e K, respectivamente), com a existência de diversas CDSs a elas relacionadas (Fig. 3)(Quadro 4).

Em todos os rizóbios analisados a função de, aproximadamente, um terço das CDSs está relacionada às categorias de transporte (E, F, G, I, P e Q), sendo que, em *S. meliloti*, cerca de 40% de todas as proteínas putativas não puderam ser classificadas nas categorias do COG (Galibert et al. 2001). Já para *M. loti* MAFF303099 (Kaneko et al., 2000) e *B. japonicum* (Kaneko et al., 2002), cerca um terço das CDSs também não puderam ser classificadas nesse banco de dados (Quadro 5).

Em uma análise mais detalhada, a comparação entre as estirpes USDA 110 e a CPAC 15 de *B. japonicum* indicou uma proporção muito próxima na porcentagem de CDSs encontradas nas diferentes categorias de COG (Fig. 4). Essa também é uma forte evidência da validação da proposta de Viprey et al. (2000), de conhecimento de um genoma a partir do seqüenciamento parcial do mesmo, com a cobertura de 10 a 15%. Essa constatação é

relevante, particularmente para países como o Brasil, reconhecidos pela megabiodiversidade, mas com recursos financeiros escassos para o seqüenciamento de diversos genomas.

Quando comparadas com o banco de dados de KEGG, 1.071 CDSs apresentaram similaridade com outros genomas seqüenciados, das quais 729, ou seja, 52,98% do total têm uma função conhecida (Fig.5).

Ainda em relação ao banco KEGG, houve similaridade das seqüências obtidas entre a CPAC 15 e outros 71 organismos presentes nesse banco de dados, sendo que o organismo que demonstrou maior número de seqüências similares foi a estirpe USDA 110 de *B. japonicum*, com 883 CDSs, representando 64,40%, seguido das estirpes BTAi1 e ORS278 de *Bradyrhizobium* sp., com 24 (1,75%) e 21 (1,63%), respectivamente, do total de CDSs neste banco (Fig.5).

Das CDSs que mostraram uma alta similaridade pelo KEGG com a estirpe BTAi1, onze tem função conhecida. Cabe aqui comentar que cinco proteínas hipotéticas provavelmente estão relacionadas com funções de exportação e função desconhecida (BS00867, BS05486, BS07987, BS09519, BS09942), e a mais alta similaridade com BTAi1 foi também encontrada para a proteína oxidase SoxC (BS08303), e a proteína anti-restrição *ardC* (BS03887), também presente em USDA 110. Contudo, muitas proteínas hipotéticas e hipotéticas conservadas (BS00395, BS00867, BS2617, BS04194, BS09519, BS09549, BS10675) foram detectadas exclusivamente na estirpe BTAi1.

As CDSs que apresentaram alta similaridade com a estirpe ORS278 incluíram uma proteína hipotética conservada que pode ser uma possível proteína de secreção (BS09243), e outra com uma provável função de sinalização na translocação (BS01360), bem como CDSs com outras funções, como: regulação da transcrição (BS06484), transporte (BS05774, BS06397), entre outras. Em relação a outros rizóbios já seqüenciados, mencionando os genes mais interessantes e com maior similaridade, estão as epóxi-hidrolases em *R. leguminosarum* (Ephx2) (BS05001) e em *M. loti* (BS10402), cujas proteínas desempenham um papel relevante no metabolismo de xenobióticos, pela degradação potencial de epóxidos tóxicos. Entre as CDSs com mais alta similaridade de KEGG com *R. etli* pode-se destacar uma glicosiltransferase envolvida na biogênese da parede celular (BS02706), e proteínas regulatórias similares encontradas em *Sinorhizobium* (BS04319, BS10828).

Em relação à similaridade de KEGG com outros organismos, 16 CDSs apresentaram maior semelhança com a estirpe X14 de *Nitrobacter hambuensis*, cinco delas de função conhecida. Cabe aqui ressaltar que algumas das CDSs de função conhecida podem estar relacionadas à adaptação ambiental: a proteína de choque térmico “cold shock”

(BS01664), encontrada em outras *Nitrobacter* e, também, em *B. japonicum* USDA110; RpoD/sigma-70 (BS03456), que mostrou alta similaridade em BLASTP com *Rhodopseudomonas*, uma porina seletiva para carboidratos altamente conservada, OprB (BS05388), a proteína de secreção HlyD (BS07181), que age como uma bomba de extrusão ("multidrug resistance efflux pump") e um regulador de transcrição geral para a família AraC (BS07214).

Conforme já comentado, existem fortes evidências de que os genomas das estirpes SEMIA 566 e CPAC 15 apresentem plasticidade elevada, particularmente em relação à ilha simbiótica, com capacidade elevada de transferência horizontal (Batista et al., 2007; Barcellos et al., 2007). A plasticidade elevada no genoma da estirpe USDA 110 foi evidenciada pela presença de 167 transposases putativas (Kaneko et al., 2002), representando 2,0% de todas as CDSs. Na CPAC 15, quatorze transposases, quase todas provavelmente relacionadas a seqüências de inserção, foram relatadas (BS00533, BS01481, BS05219, BS05636, BS06782, BS08880, BS09497, BS10333, BS10425, BS10432, BS11791, BS11875, BS11879) representando 1,02% do número total de CDSs. Com exceção de uma transposase (BS11791), as demais apresentaram alta similaridade com aquelas da USDA 110 e o conteúdo de GC esteve entre 57,5 e 63,2%.

Ainda em relação a genes adquiridos por transferência horizontal, conforme constatado com bastante evidência nas estirpes de *Bradyrhizobium* sp. BTai1 e ORS278 (Giraud et al., 2007), há diversas indicações de genes exógenos também na CPAC 15. Conforme já comentado, considerando os genes ribossomais, o conteúdo de CG da CPAC 15 foi estimado em 64,66%. Na análise das 1.371 CDSs válidas, 82,28% destas mostraram conteúdo de GC entre 59 e 69%, e 10,79 % entre 49 e 59%, o que poderia representar algum evento de transferência horizontal; além disso, 1,46% das CDSs apresentaram um conteúdo muito baixo de GC, entre 22 e 49% e em 5,4% os valores foram elevados, entre 69 a 79%, portanto, um forte indicativo de que esses genes foram adquiridos de outros organismos.

Outro ponto a ser comentado no genoma da estirpe CPAC 15 é o da existência de genes parálogos. A duplicação de genes é considerada um pré-requisito importante para a inovação gênica, que pode facilitar a adaptação a mudanças ambientais (Gevers et al., 2004). Nos microrganismos existem frações de genes que são homólogos dentro do mesmo genoma, ou seja, genes pertencentes à mesma família, que têm um ancestral comum. Isto pode acontecer devido a dois fatores: 1) duplicação de genes, sendo denominados como genes parálogos e 2) transferência horizontal de genes, recebendo a denominação de genes xenólogos. Esta duplicação depende do número de seqüências

repetidas em um genoma, as quais estão amplamente distribuídas em RNAr, tranposases, seqüências de inserção (IS) e elementos de repetição (REP) (Gevers et al., 2004)

Junto com a transferência horizontal e a perda de genes, a duplicação gênica é considerada um importante processo que molda os genomas procarióticos, tendo como conseqüência uma divergência funcional, que é um importante passo evolucionário na diversidade e no repertório funcional do organismo. Em uma análise realizada com 106 genomas, com a construção de um banco de dados de genes parálogos, revelou a existência de um número significativo desses genes, e sua classificação funcional demonstrou um enriquecimento preferencial nas classes funcionais que estão envolvidas na transcrição, metabolismo e mecanismos de defesa, segundo o banco de dados COG (Gevers et al., 2004).

Das 1.371 CDSs identificadas na CPAC 15, 67 foram classificadas como parálogas, representando 4,54% do número total de CDSs anotadas. Após a classificação, houve a formação de 32 grupos de parálogos, sendo compostos por diferentes quantidades de CDSs (duas, três ou quatro), sendo a maioria dos grupos formada por duas CDSs parálogas (28 no total). Esses grupos foram comparados em relação ao banco de dados de grupos de proteínas ortólogas, COG, e o maior número de CDSs parálogas foi detectado na classe funcional E, relacionada ao transporte e metabolismo de aminoácidos, e na classe I, relacionada com transporte e metabolismo de lipídios (dados não mostrados).

Foram anotadas diversas CDSs que codificam alguns dos genes relacionados com a nodulação e fixação biológica do N₂ (FBN); todas essas CDSs apresentaram maior similaridade (BLASTP) com a estirpe USDA 110 de *B. japonicum* (Quadro 6). Em relação à fixação de N₂, foi anotado o gene *nifE*, responsável pela regulação da FBN, em nível de transcrição, atuando na biossíntese do cofafor FeMo da dinitrogenase. Embora o *nifE* não codifique diretamente a dinitrogenase, papel que cabe ao operon *nifHDK*, alguns estudos comparativos demonstraram alguma homologia entre os polipeptídeos da nitrogenase (produtos dos genes *nifDK*), e os polipeptídeos dos genes de *nifE* e *nifN*. Desse modo, existe uma homologia considerável, em diversos domínios protéicos contendo resíduos de cisteína, entre os produtos dos genes *nifD* e *nifE* e entre os produtos dos genes *nifK* e *nifN*. Esta afirmação pode propor um provável caminho evolutivo de origem comum destes genes.

Ainda em relação à fixação do N₂, também foram detectados os genes *fixP*, *fixQ*, *fixR* e *fixO*, que desempenham um papel essencial no mecanismo de regulação da fixação (Quadro 6). Em *B. japonicum*, o operon *fixNOQP* desempenha um papel essencial no mecanismo de regulação da transcrição dos genes *nif*. Outros experimentos sugerem que o gene *fixN* é uma hemo-beta-subunidade de ligação da cobre oxidase, o *fixO* é uma monômero

do citocromo *c*, o *fixQ*, um polipeptídeo de 54 aminoácidos, e o *fixP* um dímero do citocromo *c*, todos localizados na membrana. Algumas destas proteínas são especificamente induzidas por condições limitadas de oxigênio, podendo ser responsáveis, também, pelo mecanismo de oxidação no bacteróide formado (Preising et al., 1993).

Genes de nodulação como o *nodB* (envolvido na geração de moléculas que estimulam a mitose nos protoplastos da planta hospedeira), *nodD2* (que regula a expressão dos genes *nodABCFE*, que codificam outras proteínas de nodulação), *nodW* (participa da regulação da transcrição de genes que estão envolvidos no processo de nodulação), também foram identificados (Quadro 6). Os genes *nod* essenciais à nodulação, *nodABC*, compartilham um grau elevado de homologia, podendo ser trocados entre espécies de *Rhizobium* e *Bradyrhizobium* (e outros rizóbios), sem que percam suas propriedades funcionais (Fisher et al., 1985), sendo por isso também denominados como genes *nod* “comuns” (Horvath et al., 1987); todos são considerados essenciais à nodulação, pois os mutantes em um desses três genes apresentam o fenótipo Nod⁻, isto é, não formam nódulos (Marvel et al., 1985). A única exceção foi revelada recentemente, pela análise do genoma das estirpes BTAi1 e ORS278 de *Bradyrhizobium* sp., que, surpreendentemente, não apresentam os genes *nodABC* (Giraud et al., 2007)..

Em muitas espécies de rizóbio, como em *Sinorhizobium meliloti*, os genes *nodABC* estão localizados no mesmo operon (Long, 1989). No cromossomo de *Bradyrhizobium* sp. (*Parasponia*) (Scott, 1986) e em *B. japonicum* (Banfalvi et al., 1988), dois novos genes foram detectados, *nodK* e *nodY*, que são co-regulados com *nodABC*, formando os operons *nodKABC* e *nodYABC*, respectivamente. Já em *R. leguminosarum* bv. phaseoli, o *nodA* está separado de *nodBC* por 20 kb (Vásquez et al., 1991).

Outro gene identificado na estirpe CPAC 15 de bastante interesse em relação à nodulação foi o *nopP*, relacionado ao sistema de secreção do tipo III, TTSS, que produz proteínas secretoras envolvidas na especificidade hospedeira (Ausmees, et al., 2004) (Quadro 6). Em *Rhizobium* sp. estirpe NGR234, algumas análises confirmaram que a secreção de *nopP* é dependente de flavonóides e do mecanismo TTSS, e que mutantes deletérios de *nopP* formaram mais nódulos nas raízes de *F. congesta*, em contrapartida, a eficiência desta nodulação é menor (Ausmees, et al., 2004).

O gene *ntrX*, envolvido na regulação dos genes de fixação, também foi detectado no genoma parcial da estirpe CPAC 15. O gene *ntrX* faz parte de um operon com o gene *ntrY*, constituindo um sistema regulatório de dois componentes, envolvido na fixação e no metabolismo do nitrogênio, existindo uma homologia grande com proteínas regulatórias

como a NtrC. A expressão do operon *ntrYX* foi reprimida em um mutante para *ntrC* na presença de nitrato, sugerindo uma possível interação entre *ntrYX/ntrBC* em *A. caulinodans*, que teve a expressão do gene *nifA* inativada, quando mutado para estes genes (Pawlowski, et al., 1991).

A anotação manual da estirpe CPAC 15 também permitiu detectar outras proteínas envolvidas no ciclo metabólico do N, envolvendo a assimilação, redução e fixação de formas nitrogenadas (Fig. 5). Entre estas proteínas estão a glutaminase 1 e a glutaminase sintetase II, codificadas pelos genes *glsA1* e *glnII*, respectivamente, que participam do metabolismo do glutamato. Também foi anotado o gene *norB*, que codifica uma redutase para o óxido nítrico, sendo um intermediário da desnitrificação. Análises de seqüências em *B. japonicum* USDA110, sugerem a existência de um cluster gênico formado por quatro genes, *norCBQD*, sendo que os dois primeiros são genes estruturais do cluster *nor*. As proteínas codificadas por esse “cluster” estão relacionadas às redutases da respiração do óxido nítrico de outros denitrificadores e mutantes para *norB* e *norC* mostraram a importância desses genes para o crescimento em condições microaeróbicas, na presença de nitrato (Mesa et al., 2002).

Podem-se citar, ainda, a d-aminoácido dehidrogenase, que atua na desaminação oxidativa de D-aminoácidos, e a histidina amônia-liase, codificada pelo gene *hutH*, que participa do primeiro passo da via de degradação da histidina.

De todos os compostos que atingem o solo, os xenobióticos merecem atenção especial, pois seu desaparecimento, persistência ou transformação indicam sua eficácia como produto e seu perigo potencial para a microbiota e a qualidade do ambiente (Fay et al., 1997). Desse modo, evidencia-se a importância do isolamento e identificação de microrganismos com potencial de degradação de compostos xenobióticos, incluindo os mais diversos grupos de moléculas.

O seqüenciamento parcial da estirpe CPAC 15 identificou CDSs que codificam para enzimas que participam de quase todas as vias de degradação de xenobióticos. Das 729 CDSs válidas, 5,76 % (42 CDSs) estavam relacionadas à degradação de xenobióticos. (Fig. 6). Destas, a maior porcentagem (17%), estava relacionada à degradação do benzoato via ligação da coenzima A, e 10% via hidroxilação. O benzoato é um composto intermediário da degradação de bifenis policlorinados (PCBs) que, por sua vez, são compostos organoclorados com várias aplicações industriais, como óleos hidráulicos, ou em sistemas de refrigeração, transformadores, capacitores e outros equipamentos elétricos, e isolantes térmicos. Devido a sua elevada estabilidade química e térmica, bem como por ser pouco solúvel em água e lipofílico, é considerado altamente tóxico à maioria dos organismos,

e também persiste muitos anos à degradação, sofrendo processos de bioacumulação e causando sérios problemas ambientais (Safe, 1994). Em estudos recentes com *Burkholderia xenovorans* LB400 (anteriormente classificada como *Pseudomonas*), uma estirpe potencialmente degradadora de PCBs, as vias de degradação do benzoato foram melhor caracterizadas, sugerindo que diferentes vias são utilizadas, dependendo da fase de crescimento e do tipo de substrato utilizado (Denef et, al., 2005). A enzima aldeído desidrogenase (ADH), que participa da via de degradação do bifenil em *Burkholderia xenovorans* LB400, é responsável pela conversão do acetaldeído em ácido acético, e sofre controle direto pelas concentrações de oxigênio durante o crescimento, estando também presente na via de degradação do 1,2-dicloroetano, em *B.japonicum* CPAC 15.

Pode-se concluir, pela descrição do genoma parcial da estirpe CPAC 15, que a cobertura de pouco mais de 10% do genoma permitiu a detecção de diversos genes putativos importantes, relacionados às capacidades simbiótica e saprofítica da bactéria, bem como genes com potencial de aplicação biotecnológica. Foi possível constatar, ainda, que a estirpe CPAC 15 foi bastante distinta da USDA 110, justificando a necessidade de seqüenciamento de outras estirpes dessa espécie, ou seja, reafirmando a importância de estudos de “pan-genoma”. Além disso, houve indicação de um grande número de genes na CPAC 15 adquiridos de outras espécies por transferência horizontal, representando um verdadeiro mosaico de genes. Este estudo inicial da estirpe CPAC 15, estabelecida em quase 22 milhões de hectares no Brasil, pode ser de grande importância para entender o seu comportamento nos solos brasileiros, bem como para prever o comportamento de simbioses da soja nas próximas décadas.

5 CONCLUSÕES

- ✓ A comparação entre as estirpes de *B. japonicum* USDA 110 e CPAC 15 demonstrou uma proporção muito próxima na porcentagem de CDSs encontradas nas diferentes categorias de COG.
- ✓ A comparação das CDSs da CPAC 15 com o banco de dados KEGG indicou maior similaridade com um amplo número de microrganismos, incluindo diversas classes funcionais.
- ✓ A plasticidade elevada dos genomas de rizóbios foi confirmada na estirpe CPAC 15, pela presença de elementos de inserção, transposases e genes parálogos.
- ✓ O panorama genômico permitiu a identificação de genes putativos relacionados à capacidade saprofítica, à competitividade e ao processo de fixação biológica do nitrogênio.
- ✓ Foi realizada a validação do método utilizado por Viprey et al. (2000), confirmando a possibilidade de conhecimento geral de um genoma a partir do seqüenciamento parcial do mesmo, com a cobertura de 10 a 15%, diminuindo custos e ampliando as possibilidades de estudos de genomas de microrganismos.

REFERÊNCIAS

- ALMEIDA, L.G.; PAIXÃO, R.; SOUZA, R.C.; COSTA, G.C.; ALMEIDA, D.F.; VASCONCELOS, A.T.R. A new set of bioinformatics tools for genome projects. **Genetics and Molecular Research**, v.3, p.26-52, 2004a.
- ALMEIDA, L.G.; PAIXÃO, R.; SOUZA, R.C.; COSTA, G.C.; BARRIENTOS, F.J.; SANTOS, M.T.; ALMEIDA, D.F.; VASCONCELOS, A.T. A System for Automated Bacterial (genome) Integrated Annotation—SABIA. **Bioinformatics**, v.20, n.16, p.2832–2833, 2004b.
- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p.403-10, 1990.
- APWEILER, R.; ATTWOOD, T.K.; BAIROCH, A. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. **Bioinformatics**, v.16, p.1145-50, 2000.
- ARNOLD, D.; BALKAN, L.; HUMPHREYS, R. L.; MEIJER, S.; SADLER, L. **Machine translation: an introductory guide**. Manchester-Oxford: NCC Blackwell, 1994.
- AUSMEES, N.; KOBAYASHI, H.; DEAKIN, W.J.; MARIE, C.; KRISHNAN, H.B.; BROUGHTON, W.J.; PERRET, X. Characterization of NopP, a type III secreted effector of *Rhizobium* sp. strain NGR234. **Journal of Bacteriology**, v.186, p.4774-4780, 2004.
- AVALOS, J.; CORROCHANO, L.M.; BRENNER, S. Genomic organization of the fungus *Phycomyces*. **Gene**, v.174, p.43-50, 1996.
- BAEV, N.; ENDRE, G.; PETROVICS, G.; BANFALVI, Z.; KONDOROSI, A. Six nodulation genes of *nod* box locus 4 in *Rhizobium meliloti* are involved in nodulation signal production: *nodM* codes for D-glucosamine synthetase. **Molecular and General Genetics**, v. 228, p.113–124, 1991.
- BANFALVI, Z.; SAKANYAN, V.; KONCZ, C.; DUSHA I.; KONDOROSI, A. Location of nodulation and nitrogen fixation genes on a high molecular weight plasmid of *Rhizobium meliloti*. **Molecular and General Genetics**, v.184, p.318-325, 1981.
- BARRAN, L.R.; RITCHOT, N.; BROMFIELD, E.S. *Sinorhizobium meliloti* plasmid pRm1132f replicates by a rolling-circle mechanism. **Journal of Bacteriology**, v.183, p.2704-2708, 2001.

BRENCIC, A.; WINANS, S.C. Detection of and response to signals involved in host-microbe interactions by plant-associated bacteria. **Microbiology and Molecular Biology Reviews**, v. 69, p. 155-194, 2005.

BODDEY, L.H.; HUNGRIA, M. Phenotypic grouping of Brazilian *Bradyrhizobium* strains which nodulate soybean. **Biology and Fertility of Soils**, v.25, p.407-415, 1997.

CAPELA, D.; BARLOY-HUBLER, F.; GOUZY, J. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. **Proceedings of the National Academy of Sciences of U.S.A.**, v.98, p.9877-82, 2001.

CARRARO, D.M.; CAMARGO, A.A.; SALIM, A.C.M.; GONZAGA, L.; COSTA, G.C.; VASCONCELOS, A.T.R.; SIMPSON, A.J.G. Closure of rRNA related gaps in the *Chromobacterium violaceum* genome with the PCR-assisted contig extension (PACE) protocol. **Genetics and Molecular Research**, v.3, p.53-63, 2004.

CONAB (Companhia Nacional de Abastecimento) Safras: Série histórica. Disponível em: <http://www.conab.gov.br/conabweb/download/safra/SojaSerieHist.xls/> acesso em 17 de abril 2007.

CONAB (Companhia Nacional de Abastecimento). Safras: Terceiro levantamento 2003/2004. Disponível em: <<http://www.conab.gov.br/download/safra/safra20032004Lev03.pdf>>. Acesso em 28 de abril de 2004.

CREGAN, P.; ZHU, Y.; SONG, Q.; NELSON, R. Sequence variation, haplotype analysis and linkage disequilibrium in cultivated and wild soybean populations. In: INTERNATIONAL CONFERENCE ON LEGUME GENOMICS AND GENETICS: TRANSLATION TO CROP IMPROVEMENT. Netherlands: Springer, 2002. p. 25-42.

CREGAN, P.B.; KEYSER, H.H.; SADOWSKY, M.J. Soybean genotype restricting nodulation of a previously unrestricted serocluster 123 bradyrhizobia. **Crop Science**, v.29, p.307-312, 1989.

DELCHER, A.L.; HARMON, D.; KASIF, S.; WHITE, O.; SALZBERG, S.L. Improved microbial gene identification with GLIMMER. **Nucleic Acids Research**, v.27, p.4636-4641, 1999.

DENEF, V.; PARK, J.; TSOI, T.; ROUILLARD, J.; ZHANG, H.; WIBBENMEYER, J.; VERSTRAETE, W.; GULARI, E.; HASHSHAM, S.A.; TIEDJE, J.M. Biphenyl and benzoate metabolism in a genomic context: outlining genome-wide metabolic networks in *Burkholderia xenovorans* LB400. **Applied and Environmental Microbiology** v.70, p. 4961–4970, 2004.

DJORDJEVIC, M.A.; SCHOFIELD, P.R.; RIDGE, R.W.; MORRISON, N.A.; BASSAM, B.J.; PLAZINSKI, J.; ROLFE, B.G. *Rhizobium* nodulation genes involved in root hair curling (Hac) are functionally conserved. **Plant Molecular Biology**, v.4, p.147-160, 1985.

DROZDOWICZ, A. Bactérias do solo. In: VARGAS, M.A.T.; HUNGRIA, M., eds. **Biologia dos solos do cerrados**, Planaltina: EMBRAPA-CPAC, 1997. p.17-67.

D'UTRA, G. Cultura do feijão chinês. Soja. **Jornal do Agricultor**, v.VII, p.185-188, 1882.

D'UTRA, G. Nova cultura experimental da soja (*Soja hispida* Moench). **Boletim do Instituto Agrônômico de Campinas**, V.X, p.582-587, 1899.

EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v.8, p.186-194, 1998.

EWING, B.; HILLIER, L.; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v.8, p.186-194, 1998

FERREIRA, M.C.; HUNGRIA, M. Recovery of soybean inoculant strains from uncropped soils in Brazil. **Field Crops Research**, v.79, p.139-152, 2002.

FISHER, R.F.; EGELHOFF, T.T.; MULLIGAN, J.T.; LONG, S.R. Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. **Genes and Development**, v.2, p.282-293, 1988.

FLEISCHMANN, R.D.; ADAMS, M.D.; WHITE, O.; CLAYTON, R.A.; KIRKNESS, E.F.; KERLAVAGE, A.R.; BULT, C.J.; TOMP, J.F.; DOUGHERTY, B.A.; MERRICK, J.M. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. **Science**, v. 269, p. 496-512, 1995.

FREIBERG, C.; FELLAY, R.; BAIROCH, A.; BROUGHTON, W.J.; ROSENTHAL, A.; PERRET, X. Molecular basis of symbiosis between *Rhizobium* and legumes. **Nature**. v.387, n.6631, p.394-401, 1997

FREIRE, J.R.J.; KOLLING, J.; VIDOR, C.; PEREIRA, J.S.; KOLLING, I.G.; MENDES, N.G. Sobrevivência e competição por sítios de nodulação de estirpes de *Rhizobium japonicum* na cultura da soja. **Revista Brasileira de Ciência do Solo**, v.7, p.47-53, 1983.

GALIBERT, F.; FINAN, T.M.; LONG, S.R.; PÜHLER, A.; ABOLA, A.P.; AMPE, F.; BARLOY-HUBLER, F.; BARNETT, M.J.; BECKER, A.; BOISTARD, P.; BOTHE, G.; BOUTRY, M.; BOWSER, L.; BUHRMESTER, J.; CADIEU, E.; CAPELA, D.; CHAIN, P.; COWIE, A.; DAVIS, R.W.; DRÉANO, S.; FEDERSPIEL, N.A.; FISHER, R.F.; GLOUX, S.; GODRIE, T.; GOFFEAU, A.; GOLDING, B.; GOUZY, J.; GURJAL, M.; HERNANDEZ-LUCAS, I.; HONG, A.; HUIZAR, L.; HYMAN, R.W.; JONES, T.; KAHN, D.; KAHN, M.; KALMAN, S.; KEATING, D.H.; KISS, E.; KOMP, C.; LELAURE, V.; MASUY, D.; PALM, C.; PECK, M.C.; POHL, T.M.; PORTETELLE, D.; PURNELLE, B.; RAMSPERGER, U.; SURZYCKI, R.; THÉBAULT, P.; VANDENBOL, M.; VORHÖLTER, F.J.; WEIDNER, S.; WELLS, D.H.; WONG, K.; YEH, K.C.; BATUT, J. The composite genome of the legume symbiont *Sinorhizobium meliloti*. **Science**, v.293, p.668-672, 2001.

GIRAUD, E.; MOULIN, L.; VALLENET, D.; BARBE, V.; CYTRYN, E.; AVARRE, J.C.; JAUBERT, M.; SIMON, D.; CARTIEAUX, F.; PRIN, Y.; BENA, G.; HANNIBAL, L.; FARDOUX, J.; KOJADINOVIC, M.; VUILLET, L.; LAJUS, A.; CUVEILLER, S.; ROUY, Z.; MANGENOT, S.; SEGURENS, B.; DOSSAT, C.; FRANCK, W.L.; CHANG, W.S.; SAUNDERS, E.; BRUCE, D.; RICHARDSON, P.; NORMAND, P.; DREYFUS, B.; PIGNOL, D.; STACEY, G.; EMERICH, D.; VERMÉGLIO, A.; MÉDIGUE, C.; SADOWSKY, M. Legumes symbioses: absence of *nod* genes in photosynthetic bradyrhizobia. **Science**, v.316, p.1307-1312, 2007

GONZÁLEZ, V.; BUSTOS, P.; RAMYREZ-ROMERO, M.A.; MEDRANO-SOTO, A.; SALGADO, H.; HERNANDEZ-GONZÁLEZ, I.; HERNÁNDEZ-CELIS, J.C.; QUINTERO, V.; MORENO-HAGELSIEB, G.; GIRARD, L.; RODRÍGUEZ, O.; FLORES, M.; CEVALLOS, M.A.; VIDES, J.C.; ROMERO, D.; DÁVILA, G. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments **Genome Biology**, v.4, n.R36, 2003

GONZÁLEZ, V.; SANTAMARÍA, R.I.; BUSTOS, P.; HERNÁNDEZ-GONZÁLEZ, I.; MEDRANO-SOTO, A.; HAGELSIEB, G.M.; JANGA, S.C.; RAMÍREZ, M.A.; JIMÉNEZ-JACINTO, V.; COLLADO-VIDES, J.; DÁVILA, G. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. **Proceedings of the National Academy of Sciences of U.S.A.**, v. 103, n.10, p.3834-3839, 2006.

GORDON, D.; ABAJIAN, C.; GREEN, P. *Consed*: a graphical tool for sequence finishing. **Genome Research**, v.8, p.195-202, 1998.

GÖTTFERT, M.; RÖTHLISBERGER, S.; KÜNDIG, C.; BECK, C.; MARTY, R.; HENNECKE, H. Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. **Journal of Bacteriology**, v.183, n.4, p.1405-1412, 2001.

HALBLEIB, C.M.; LUDDEN, P.W. Regulation of biological nitrogen fixation. **Journal of Nutrition**, v.130, n.5, p.1081-1084, 2000.

HAM, G.E.; CALDWELL, V.B.; JOHNSON, H.W. Evaluation of *Rhizobium japonicum* inoculants in soil containing naturalized populations of rhizobia. **Agronomy Journal**, v.63, p.301-303, 1971.

HORVATH, B.; KONDOROSI, E.; JOHN, M.; SCHMIDT, J.; TÖRÖK, I.; GYÖRGYPAL, Z.; BARABAS, I.; WIENEKE, U.; SCHELL, J.; KONDOROSI, A.. Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. **Cell**, v.46, p.335-343, 1986.

HUNGRIA, M.; VARGAS, M.A.T. Environmental factors impacting N₂ fixation in legumes grown in the tropics, with an emphasis on Brazil. **Field Crops Research**, v.65, p.151-164, 2000.

HUNGRIA, M.; VARGAS, M.A.T. Exploring the microbial diversity and soil management practices to optimize the contribution of soil microorganisms to plant nutrition. In: STACEY, G.; MULLIN, B.; GRESSHOFF, P.M., eds. **Biology of plant-microbe interactions**. St. Paul: International Society of Molecular Plant-Microbe Interactions. (Proceedings of the 8th International Symposium on Molecular Plant-Microbe Interactions), 1996. p.493-496.

HUNGRIA, M.; BODDEY, L.H.; SANTOS, M.G.; VARGAS, M.A.T. Nitrogen fixation capacity and nodule occupancy by *Bradyrhizobium japonicum* and *B. elkanii* strains. **Biology and Fertility of Soils**, v.27, p.393-399, 1998.

HUNGRIA, M.; CAMPO, R.J.; MENDES, I.C. **Fixação biológica do nitrogênio na cultura da soja**. Londrina: Embrapa Soja, 2001. 48p. (Embrapa Soja. Circular Técnica, 35; Embrapa Cerrados. Circular Técnica, 13).

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 in plant productivity**. Houston, Texas: Studium Press, LLC. 2006. p. 43-93.

HUNGRIA, M.; FRANCHINI, J.C.; CAMPO, R.J.; GRAHAM, P.H. The importance of nitrogen fixation to soybean cropping in South America. In: WERNER, D.; NEWTON, W.E., eds. **Nitrogen fixation in agriculture, forestry, ecology, and the environment**. Dordrecht: Springer, 2005. p.25-42.

HUNGRIA, M.; VARGAS, M.A.T.; ANDRADE, D.S.; CAMPO, R.J.; CHUEIRE, L.M.O.; FERREIRA, M.C.; MENDES, I.C. Fixação biológica do nitrogênio em leguminosas de grãos. In: SIQUEIRA, J.O.; MOREIRA, F.M.S.; LOPES, A.S.; GUILHERME, L.R.G.; FAQUIN, V.; FURTINNI NETO, A.E.; CARVALHO, J.G., eds. **Soil fertility, soil biology and plant nutrition interrelationships**. Viçosa: SBCS, Lavras: UFLA/DCS, 1999. p.597-620.

HUNGRIA, M.; VARGAS, M.A.T.; ARAUJO, R.S. Fixação biológica do nitrogênio em feijoeiro. In: VARGAS, M. A. T.; HUNGRIA, M. eds. **Biologia dos solos dos Cerrados**, Planaltina: EMBRAPA-CPAC, 1997. p.189-295.

HUNGRIA, M.; VARGAS, M.A.T.; SUHET, A.R.; PERES, J.R.R. Fixação biológica do nitrogênio em soja. In: ARAUJO, R.S.; HUNGRIA, M., eds. **Microrganismos de importância agrícola**. Brasília: EMBRAPA-SPI, p.9-89. 1994

HYMOWITZ, T. On the domestication of the soybean. **Economic Botany**, v.24, p.408-421, 1970.

HYMOWITZ, T.; NEWELL, C.A. Taxonomy, speciation, domestication, dissemination, germplasm resources and variation in the genus *Glycine*. In: SUMMERFIELD, R.J.; BUNTING, A.H., eds. **Advances in legume science**. Kew/U.K.: Royal Botanic Gardens., 1980. p.251-264.

KANEHISA, M.; GOTO, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v.28, p.29-34, 2000.

KANEKO, T.; NAKAMURA, Y.; SATO, S.; ASAMIZU, E.; KATO, T.; SASAMOTO, S.; WATANABE, A.; IDESAWA, K.; ISHIKAWA, A.; KAWASHIMA, K.; KIMURA, T.; KISHIDA, Y.; KIYOKAWA, C.; KOHARA, M.; MATSUMOTO, M.; MATSUNO, A.; MOCHIZUKI, Y.; NAKAYAMA, S.; NAKAZAKI, N.; SHIMPO, S.; SUGIMOTO, M.; TAKEUCHI, C.; YAMADA, M.; TABATA, S. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. **DNA Research**. v.7, p.331-338, 2000a.

KANEKO, T.; NAKAMURA, Y.; SATO, S.; ASAMIZU, E.; KATO, T.; SASAMOTO, S.; WATANABE, A.; IDESAWA, K.; ISHIKAWA, A.; KAWASHIMA, K.; KIMURA, T.; KISHIDA, Y.; KIYOKAWA, C.; KOHARA, M.; MATSUMOTO, M.; MATSUNO, A.; MOCHIZUKI, Y.; NAKAYAMA, S.; NAKAZAKI, N.; SHIMPO, S.; SUGIMOTO, M.; TAKEUCHI, C.; YAMADA, M.; TABATA, S. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti* (Supplement). **DNA Research**. v.7, p. 381-406, 2000b.

KANEKO, T.; NAKAMURA, Y.; SATO, S.; MINAMISAWA, K.; UCHIUMI, T.; SASAMOTO, S.; WATANABE, A.; IDESAWA, K.; IRIGUCHI, M.; KAWASHIMA, K.; KOHARA, M.; MATSUMOTO, M.; SHIMPO, S.; TSURUOKA, H.; WADA, T.; YAMADA, M.; TABATA, S. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. **DNA Research**, v.9, p. 189-197, 2002.

KANG, U.G.; SOMASEGARAN, P.; HOBEN, H.J.; BOHLOOL, B.B. Symbiotic potential, competitiveness, and serological properties of *Bradyrhizobium japonicum* indigenous to Korean soils. **Applied and Environmental Microbiology**, v.57, p.1038-1045, 1991.

KEYSER, H.H.; CREGAN, P.B. Nodulation and competition for nodulation of selected soybean genotypes among *Bradyrhizobium japonicum* serogroup 123 isolates. **Applied and Environmental Microbiology**, v.53, p.2631-2635, 1987.

KIM, C.W.; MARKIEWICZ, P.; LEE, J.J.; SCHIERLE, C.F.; MILLER, J.H. Studies of the hyperthermophile *Thermotoga maritima* by random sequencing of cDNA and genomic libraries. Identification and sequencing of the *trpEG* (D) operon. **Journal of Molecular Biology**, v.231, p.960-981, 1993.

KRAUSE A.; DOERFEL A.; GÖTTTFERT M. Mutational and transcriptional analysis of the type III secretion system of *Bradyrhizobium japonicum*. **Molecular Plant Microbe Interaction**, v.15, p.1228-1235. 2002.

KONDOROSI, E.; GYURIS, J.; SCHMIDT, J.; JOHN, M.; DUDA, E.; HOFFMAN, J.; SCHELL, J.; KONDOROSI, A. Positive and negative control of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. **EMBO Journal**, v.5, p.1331-1340, 1989.

LONG, S. R. *Rhizobium*-legume nodulation: life together in the under-ground. **Cell**, v.56, p.203-214, 1989.

MARVEL, D.J.; KULDAU, G.; HIRSCH, A.; RICHARDS, E.; TORREY, J.G.; AUSUBEL, F.M. Conservation of nodulation genes between *Rhizobium meliloti* and a slow-growing *Rhizobium* strain that nodulates a non legume host. **Proceedings of the National Academy of Sciences of the U.S.A.**, v.82, p.5841-5845, 1985.

MENDES, I.C.; HUNGRIA, M.; VARGAS, M.A.T. Establishment of *Bradyrhizobium japonicum* and *B. elkanii* strains in a Brazilian Cerrado oxisol. **Biology and Fertility of Soils**, v.40, p.28-35, 2004.

MESA, S.; VELASCO, L.; MANZANERA, M.E.; DELGADO, M.J.; BEDMAR, E.J. Characterization of the *norCBQD* genes, encoding nitric oxide reductase, in the nitrogen fixing bacterium *Bradyrhizobium japonicum*. **Microbiology**, v.148, p. 3553-3560, 2002.

MORSE, W.J. History of soybean production. In: MARKLEY, K.L., ed. **Soybeans and soybean products**. New York: Interscience Publ. Inc. v.1, 1950. p.3-59.

MORGANTE, P.G., Fixação biológica e assimilação de nitrogênio, 2003. Disponível em <<http://www.ciagri.usp.br/~lazaropp/FisioVegGrad/MetNitro.htm>> acesso em 10/03/07.

MOAWAD, H.A.; ELLIS, W.R.; SCHMIDT, E.L. Rhizosphere response as a factor in competition among three serogroups of indigenous *Rhizobium japonicum* for nodulation of field-grown soybeans. **Applied and Environmental Microbiology**, v.47, p.607-612, 1984.

MULDER, N.J.; APWEILER, R.; ATTWOOD, T.K.; BAIROCH, A.; BARRELL, D.; BATEMAN, A.; BINNS, D.; BISWAS, M.; BRADLEY, P.; BORK, P.; BUCHER, P.; COPLEY, R.R.; COURCELLE, E.; DAS, U.; DURBIN, R.; FALQUET, L.; FLEISCHMANN, W.; GRIFFITHS-JONES, S.; HAFT, D.; HARTE, N.; HULO, N.; KAHN, D.; KANAPIN, A.; KRESTYANINOVA, M.; LOPEZ, R.; LETUNIC, I.; LONSDALE, D.; SILVENTOINEN, V.; ORCHARD, S.E.; PAGNI, M.; PEYRUC, D.; PONTING, C.P.; SELENGUT, J.D.; SERVANT, F.; SIGRIST C.J.; VAUGHAN, R.; ZDOBNOV, E.M. The InterPro Database, 2003 brings increased coverage and new features, Hinxton. **Nucleic Acids Research**, v.31, p.315-318. 2003.

MULLIGAN, J.T.; LONG, S.R. Induction of *Rhizobium meliloti nodC* expression by plant exudate requires *nodD*. **Proceedings of National Academy of Sciences of U.S.A.**, v.82, p.6609-6613, 1985.

NAKAI, K.; KANEHISA, M. Expert system for predicting protein localization sites in Gram-negative bacteria. **Proteins: Structure, Function, and Genetics**, v.11, p.95-110, 1991.

NCBI (National Center for Biotechnology Information) – <<http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html>> acesso em 10 de julho de 2007a.

NCBI (National Center for Biotechnology Information) - <http://www.ncbi.nlm.nih.gov/genomes/static/eub_g.html> acesso em 17 de julho de 2007b.

NISHI, C.Y.M.; BODDEY, L.H.; VARGAS, M.A.T.; HUNGRIA, M. Morphological, physiological and genetic characterization of two new *Bradyrhizobium* strains recently recommended as Brazilian commercial inoculants for soybean. **Symbiosis**, v.20, p.147-162, 1996.

PASCHOAL, A. R., ZERILLO, M. M., DE ALMEIDA, L.G.P., CUNHA, O L., DE VASCONCELOS, A. T. R. , MONTEIRO-VITORELLO, C. B. 'GINGA - Graphical Interface for comparative Genome Analysis'. In: INTELLIGENT SYSTEMS FOR MOLECULAR BIOLOGY, ISMB 2006, Fortaleza. ISMB 2006: Fortaleza, Poster Session.

PAWLOWSKI, K.; KLOSSE, U.; DE BRUIJIN, F.J. Characterization of a novel *Azorhizobium caulinodans* ORS571 two-component regulatory system, NtrY/NtrX, involved in nitrogen fixation and metabolism. **Molecular and General Genetics**, v.231, p.124-138, 1991.

PERES, J.R.R.; VIDOR, C. Seleção de estirpes de *Rhizobium japonicum* e competitividade por sítios de infecção nodular em cultivares de soja. **Agronomia Sul Riograndense**, v.16, p.205-219, 1980.

PICHON, M. E. P.; JOURNET, A.; DEDIEU, F.; de BILLY, G.; TRUCHET, T.; BARKER, D.G. *Rhizobium meliloti* elicitors of transgenic alfafa. **The Plant Cell**, v. 4, p. 1199-1211, 1992.

PREISIG, O.; ANTHAMATTEN, D.; HENNECKE, D. Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. **Proceedings of the National Academy of Sciences of the U.S.A.**, v.8, p. 3309-13, 1993.

POSTGATE, J.R. **The fundamentals of nitrogen fixation**. Cambridge: Cambridge Univ. Press., 1982. 252 p.

RIVILLA, R.; DOWNIE, J.A. Identification of a *Rhizobium leguminosarum* gene homologous to *nodT* but located outside the symbiotic plasmid. **Gene**, v.144, p.87-91, 1994.

SAFE, S.H. Polychlorinated biphenyls (PCBs): environmental impact, biochemical and toxic responses and implications for risk assessment. **Critical Reviews in Toxicology**, v.24, p.87-149, 1994.

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular cloning: a laboratory manual**, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1989.

SCHULTZE, M., KONDOROSI, E. Regulation of symbiotic root nodule development. **Annual Review of Genetics**, v.32, p. 33-57, 1998.

SCHWEDOCK, J.; LONG, S.R. Nucleotide sequence and protein products of two new nodulation genes of *Rhizobium meliloti*, *nodP* and *nodQ*. **Molecular Plant-Microbe Interactions**, v.2, p.181-194, 1989.

SEMU, E.; HUME, D.J. Effects of inoculation and fertilizer N levels on N₂ fixation and yields of soybeans. **Canadian Journal of Plant Science**, v.59, p.1129-1137, 1979.

SIMPSON, A.J.G. Genome sequencing network. **Nature Reviews**, v.2, p.979-983, 2001.

SCHMIDT, E.L.; ZIDWICK, M.J.; ABEBE, H. *Bradyrhizobium japonicum* serocluster 123 and diversity among member isolates. **Applied and Environmental Microbiology**, v.51, p.1212-1215, 1986.

SCOTT, K.K. Conserved nodulation genes from the non-legume symbiont *Bradyrhizobium* sp. (Parasponia). **Nucleic Acids Research**, v.14, p.2905-2919, 1986.

STREETER, J.G. Failure in inoculant rhizobia to overcome the dominance of indigenous strains for nodule formation. **Canadian Journal of Microbiology**, v.40, p.513-522, 1994.

SURIN, B.P.; DOWNIE, J.A. Characterization of the *Rhizobium leguminosarum* genes *nodLMN* involved in efficient host specific nodulation. **Molecular Microbiology**, v.2, p.173-183, 1988

TATUSOV, R.; GALPERIN, M.; NATALE, D.; KOONIN, E. The COG database: a tool for genome-scale analysis of protein functions and evolution. **Nucleic Acids Research**, v.28, p.33-36, 2000.

THE UNIPROT CONSORTIUM. The Universal Protein Resource (UniProt). **Nucleic Acids Research**, v.35, p.193-197, 2007.

VANCE, C.P. Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. **Plant Physiology**, v. 127, p. 390-397, 2001.

VAN KESSEL C.; HARTLEY, C.G. Agricultural management of grain legumes: has it led to an increase in nitrogen fixation? **Field Crops Research**, v.65, p.165-181, 2000.

VARGAS, M.A.T.; HUNGRIA, M. Fixação biológica do nitrogênio na cultura da soja. In: VARGAS, M.A.T.; HUNGRIA, M., eds. **Biologia dos solos dos Cerrados**. Planaltina: EMBRAPA-CPAC. 1997. p.297-360.

VARGAS, M.A.T.; HUNGRIA, M. Legume inoculation in Cerrado Soils. In: PEDROSA, F.O.; HUNGRIA, M.; YATES, M.G.; NEWTON, W.E., eds. **Nitrogen fixation: from molecules to crop productivity**. Dordrecht: Kluwer., 2000. p.587-588.

VARGAS, M.A.T.; MENDES, I.C.; SUHET, A.R.; PERES, J.R.R. **Duas novas estirpes de rizóbio para a inoculação da soja**. Planaltina: EMBRAPA-CPAC, 1992. 3p. (EMBRAPA-CPAC. Comunicado Técnico, 62).

VASCONCELOS, A.T.R.; ALMEIDA, D.F.; HUNGRIA, M.; GUIMARÃES, C.T.; ANTÔNIO, R.V.; ALMEIDA, F.C.; ALMEIDA, L.G.P.; ALMEIDA, R.; ALVES-GOMES, J.A.; ANDRADE, E.M.; ARARIPE, J.; ARAÚJO, M.F.F.; ASTOLFI-FILHO, S.; AZEVEDO, V.; BAPTISTA, A.J.; BATAUS, L.A.M.; BATISTA, J.S.; BELÓ, A.; VAN DEN BERG, C.; BOGO, M.; BONATTO, S.; BORDIGNON, J.; BRIGIDO, M.M.; BRITO, C.A.; BROCCHI, M.; BURITY, H.A.; CAMARGO, A.A.; CARDOSO, D.D.P.; CARNEIRO, N.P.; CARRARO, D.M.; CARVALHO, C.M.B.; CASCARDO, J.C.M.; CAVADA, B.S.; CHUEIRE, L.M.O.; CRECZYNSKI-PASA, T.B.; CUNHA-JUNIOR, N.C.; FAGUNDES, N.; FALCÃO, C.L.; FANTINATTI, F.; FARIAS, I.P.; FELIPE, M.S.S.; FERRARI, L.P.; FERRO, J.A.; FERRO, M.I.T.; FRANCO, G.R.; FREITAS, N.S.A.; FURLAN, L.R.; GAZZINELLI, R.T.; GOMES, E.A.; GONÇALVES, P.R.; GRANGEIRO, T.B.; GRATTAPAGLIA, D.; GRISARD, E.C.; HANNA, E.S.; JARDIM, S.N.; LAURINO, J.; LEOI, L.C.T.; LIMA, L.F.A.; LOUREIRO, M.F.; LYRA, M.C.C.P.; MADEIRA, H.M.F.; MANFIO, G.P.; MARANHÃO, A.Q.; MARTINS, W.S.; MAURO, S.M.Z.; MEDEIROS, S.R.B.; MEISSNER, R.V.; MOREIRA, M.A.M.; NASCIMENTO, F.F.; NICOLÁS, M.F.; OLIVEIRA, J.G.; OLIVEIRA, S.C.; PAIXÃO, R.F.C.; PARENTE, J.A.; PEDROSA, F.O.; PENA, S.D.J.; PEREIRA, J.O.; PEREIRA, M.; PINTO, L.S.R.C.; PINTO, L.S.P.; PORTO, J.I.R.; POTRICH, D.P.; RAMALHO-NETO, C.E.; REIS, A.M.M.; RIGO, L.U.; RONDINELLI, E.; SANTOS, E.B.P.; SANTOS, F.R.; SCHNEIDER, M.P.C.; SEUANEZ, H.N.; SILVA, A.M.R.; SILVA, A.L.C.; SILVA, D.W.; SILVA, R.; SIMÕES, I.C.; SIMON, D.; SOARES, C.M.A.; SOARES, R.B.A.; SOUZA, E.M.; SOUZA, K.R.L.; SOUZA, R.C.; STEFFENS, M.B.R.; STEINDEL, M.; TEIXEIRA, S.R.; URMENYI, T.; VETTORE, A.; WASSEM, R.; ZAHA, A. AND SIMPSON, A.J.G. The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability. **Proceedings of the National Academy of Sciences of U.S.A.**, v.100, p.11660-11665, 2003.

VASCONCELOS, A.T.R.; FERREIRA, H.B.; BIZARRO, C.V.; BONATTO, S.L.; CARVALHO, M.O.; PINTO, M.P.; ALMEIDA, D.F.; ALMEIDA, L.G.P.; ALMEIDA, R.; ALVES-FILHO, L.; ASSUNÇÃO, E.N.; AZEVEDO, V.A.C.; BOGO, M.R.; BRIGIDO, M.M.; BROCCHI, M.; BURITY, H.A.; CAMARGO, A.A.; CAMARGO, S.S.; CAREPO, M.S.; CARRARO, D.M.; CASCARDO, J.C.M.; CASTRO, L.A.; CAVALCANTI, G.; CHEMALE, G.; COLLEVATTI, R.G.; CUNHA, C.W.; DALLAGIOVANNA, B.; DAMBRÓS, B.P.; DELLAGOSTIN, O.A.; FALCÃO, C.; FANTINATTI-GARBOGGINI, F.; FELIPE, M.S.S.; FIORENTIN, L.; FRANCO, G.R.; FREITAS, N.S.A.; FRÍAS, D.; GRANGEIRO, T.B.; GRISARD, E.C.; GUIMARÃES, C.T.; HUNGRIA, M.; JARDIM, S.N.; KRIEGER, M.A.; LAURINO, J.P.; LIMA, L.F.A.; LOPES, M.I.; LORETO, E.L.S.; MADEIRA, H.M.F.; MANFIO, G.P.; MARANHÃO, A.Q.; MARTINKOVICS, C.T.; MEDEIROS, S.R.B.; MOREIRA, M.A.M.; NEIVA, M.; RAMALHO NETO, C.E.; NICOLÁS, M.F.; OLIVEIRA, S.C.; PAIXÃO, R.F.C.; PEDROSA, F.O.; PENA, S.D.J.; PEREIRA, M.; PEREIRA-FERRARI, L.; PIFFER, I.; PINTO, L.S.; POTRICH, D.P.; SALIM, A.C.M.; SANTOS, F.R.; SCHMITT, R.; SCHNEIDER, M.P.C.; SCHRANK, A.; SCHRANK, I.; SCHUCK, A.F.; SEUANEZ, H.N.; SILVA, D.W.; SILVA, R.; SILVA, S.C.; SOARES, C.M.A.; SOUZA, K.R.L.; SOUZA, R.C.; STAAS, C.C.; STEFFENS, M.B.; TEIXEIRA, S.M.R.; URMENYI, T.P.; VAINSTEIN, M.H.; ZUCCHERATO, L.W.; SIMPSON, A.J.G.; ZAHA, A. Swine and poultry pathogens: the complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. **Journal of Bacteriology**, v.187, n.16, p.5568-5577, 2005.

VÁZQUEZ, M.; DÁVALOS, A.; DE LAS PEÑAS, A.; SÁNCHEZ, F.; QUINTO, C. Novel organization of the common nodulation genes in *Rhizobium leguminosarum* bv. *phaseoli* strains. **Journal of Bacteriology**, v.173, p.1250-1258, 1991.

VIPREY, V.; ROSENTHAL, A.; BROUGHTON, W.K.; PERRET, X. Genetic snapshots of the *Rhizobium* species NGR234 genome. **Genome Biology**, v.1, p.1-17, 2000. (<<http://genomebiology.com/2000/1/6/research/0014.1>>).

WEBER, D.F.; KEYSER, H.H.; URATSU, S.L. Serological distribution of *Bradyrhizobium japonicum* from U.S. soybean production areas. **Agronomy Journal**, v.81, p.786-789, 1989.

YOUNG, J.P.W.; CROSSMAN, L.C.A.; JOHNSTON, W.B.; THOMSON, N.R.; GHAZOU, Z.F.; HULL, K.H.; WEXLER, M.; CURSON, A.R.J.; TODD, J.D.; POOLE, P.S.; MAUCLINE, T.H.; EAST, A.K.; QUAIL, M.A.; 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.; PARKHILL, J. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. **Genome Biology**, v.7, n.4, art.R34, 2006

LISTAS DE QUADROS E FIGURAS

Parâmetros	Bibliotecas		Última montagem
	BJPM01	BJPM09	BJPM01+BJPM09
Número de leituras	2.268	1.970	4.328
Número de leituras sem vetor	1.787	1.731	3.683
Número de "contigs"	558	489	1.106
Número de "contigs" com duas	415	372	1.072
Número de "singleton" isolados	817 (37,17% das leituras)	570 (28,93% das leituras)	1.439
Número de bases depositadas (pares de bases, pb) (excluindo vetor, incluindo bases de baixa qualidade)	2.470.301	1.783.874	4.576.417
Número de bases de qualidade ≥ 20 e % em relação as bases depositadas	756.118 (28,76%)	604.309 (31,81%)	2.046.740 (41,19%)
Número de bases do vetor e % em relação as bases	158.931 (6,04%)	115.648 (6,09%)	392.353 (7,90%)
Número de leituras "g" + "b"	827	136	1.747
Número somente de "b"	301	56	358
Número somente de "g"	173	210	406
Tamanho médio dos clones (pb)	1102,17	999,94	1.057,40

QUADRO 1 – Parâmetros analisados no seqüenciamento das bibliotecas do tipo "shotgun" da estirpe CPAC 15 de *Bradyrhizobium japonicum*.

Parâmetro	Valores
Tamanho estimado do genoma (pb)	9.100.000*
Número total de leituras	4.328
Numero total de leituras sem vetores (<=10% vetor)	3.683
Número de leituras com 10-80% de bases de vetores	489
Número de leituras com mais de 80% de bases de vetores	156
Nº de bases depositadas (pb) excluindo vetor, incluindo bases de baixa qualidade	4.576.417 (100%)
Cobertura do genoma com as bases depositadas	0,51 vezes (51%)
Cobertura do genoma real	1.184.994 (13,17% do tamanho estimado do genoma)
Numero de bases com qualidade >=20 (pb) (% do total de bases depositadas)	2.046.740 (41,19%)
Numero de bases com qualidade >=30 (pb) (% do total de bases depositadas)	1.461.814 (29,42%)
No. de bases de vetores (pb)	392.353(7,90%)
Comprimento médio (pb)	1.057,40
Tamanho medio das leituras com qualidade>=20 (pb)	473
Numero total de "contigs"	1.371
Tamanho medio dos "contigs" (pb)	1071,42
Cobertura por "contigs" phrap (pb)	1.149.042
Numero médio de leituras em um "contig"	2,49
Qualidade media das bases nos "contigs" (phrap)	18,14
Cobertura do genoma, considerando-se a soma dos "contigs" (pb) e em %	1.340.335 (14,89%)

* Estimado segundo o genoma de *B. japonicum* estirpe USDA 110

QUADRO 2 – Parâmetros relativos ao panorama genômico obtido para a estirpe CPAC 15 de *Bradyrhizobium japonicum*

Contig	Tamanho do contig	Contig size - trimmed(bp)	Número de leituras
1	1.261	701	1
2	1.344	599	1
3	1.174	504	1
4	1.000	407	1
5	1.068	434	1
6	1.224	321	1
7	652	570	1
8	916	702	1
9	856	597	1
10	1.269	493	1
Continuação...			
1.096	2.787	2.607	8
1.097	3.381	3.259	8
1.098	2.491	2.364	9
1.099	2.385	2.232	9
1.100	1.391	1.238	9
1.101	1.411	1.260	10
1.102	2.297	2.275	11
1.103	2.439	2.359	13
1.104	3.068	2.966	14
1.105	2.068	1.902	14
1.106	3.054	2.937	16
	1.340.335	1.149.042	2.889
Número total de leituras: leituras em contigs (2.889) + leituras únicas (1439) = 4.328			

QUADRO 3 – Exemplo da formação, tamanho e número de leituras dos "contigs" obtidos no seqüenciamento parcial do genoma da estirpe CPAC 15 de *B. japonicum*.

COG Categoria funcional	Número de CDS
J – Translation	31
A - RNA processing and modification	0
K – Transcription	63
L - Replication, recombination and repair	27
B - Chromatin structure and dynamics	0
D - Cell cycle control, mitosis and meiosis	2
Y - Nuclear structure	0
V - Defense mechanisms	12
T - Transduction mechanisms	30
M - Cell wall/membrane biogenesis	29
N - Cell motility	5
Z – Cytoskeleton	0
W - Extracellular structures	0
U - Intracellular trafficking and secretion	9
O - Posttranslational modification, protein turnover, chaperones	17
C - Energy production and conversion	43
G - Carbohydrate transport and metabolism	57
E – Amino acid transport and metabolism	105
F - Nucleotide transport and metabolism	3
H - Coenzyme transport and metabolism	25
I - Lipid transport and metabolism	47
P - Inorganic ion transport and metabolism	53
Q - Secondary metabolites biosynthesis, transport and catabolism	45
R - General function prediction only	130
S - Function unknown	64
Total	797

QUADRO 4 – Classificação funcional das CDSs da estirpe CPAC 15 de *B. japonicum* pelo banco de dados COG ("Clusters of Orthologous Groups of proteins").

		<i>B. japonicum</i> USDA 110	<i>M. loti</i> MAFF303099	<i>M. loti</i> BNC1	<i>R. etli</i> CFN 42	<i>R. leguminosarum</i> bv. viciaeae 3841	<i>S. meliloti</i> 1021	<i>B. japonicum</i> CPAC15 **
Proteínas no cromossomo		-	-	-	-	-	-	
Classificadas no COG		5.362	4.580	-	2.859	-	3.341	797
Não classificadas no COG		2.955	2.163	-	1.176	-	2.490	574
Não analisadas no COG		-	-	4.064	-	4699	-	
Nº total de proteínas no cromossomo		8.317	0	4.064	4.035	4699	5.831	1.371
Classificação	Código							
	J Translation	162	161	-	143	-	140	31
	A RNA processing and modification	0	0	-	0	-	0	0
	K Transcription	459	495	-	287	-	222	63
	L Replication, recombination and repair	286	240	-	132	-	167	27
	B Chromatin structure and dynamics	0	0	-	0	-	0	0
	D Cell cycle control, mitosis and meiosis	34	29	-	24	-	25	2
	Y Nuclear structure	-	-	-	-	-	-	0
	V Defense mechanisms	95	60	-	41	-	30	12
	T Signal transduction mechanisms	385	189	-	180	-	101	30
	M Cell wall/membrane biogenesis	309	285	-	222	-	133	29
	N Cell motility	137	48	-	66	-	49	5
	Z Cytoskeleton	1	0	-	0	-	0	0
	W Extracellular structures	0	1	-	0	--	1	0
	U Intracellular trafficking and secretion	121	90	-	50	-	54	9
	O Posttranslational modification, protein turnover, chaperones	221	174	-	126	-	115	17
	C Energy production and conversion	457	300	-	182	-	163	43
	G Carbohydrate transport and metabolism	327	449	-	270	-	192	57
	E Amino acid transport and metabolism	710	644	-	320	-	332	105
	F Nucleotide transport and metabolism	83	81	-	79	-	71	3

H	Coenzyme transport and metabolism	185	171	-	100	-	104	25
I	Lipid transport and metabolism	395	268	-	132	-	117	47
P	Inorganic ion transport and metabolism	323	237	-	144	-	137	53
Q	Secondary metabolites biosynthesis, transport and catabolism	300	204	-	94	-	77	45
R	General function prediction only	809	709	-	403	-	369	130
S	Function unknown	488	431	-	273	-	233	64
Número total*		6.287	0	-	3.268	-	2.832	1.371
Proteínas no plasmídeo								
Plasmid	Não classificado		320	317	175	156	1.294	-
Plasmid	Não classificado		209	113	163	132	1.570	-
Plasmid	Não classificado			49	232	305	-	-
Plasmid	Não classificado			-	336	442	-	-
Plasmid	Não classificado			-	455	634	-	-
Plasmid	Não classificado			-	567	781	-	-
				-		-	-	-
Número total de proteínas no genoma		8.317	7.272	4.543	5.963	7.151	6.205	-
RNAs		56	57	72	59	61	65	13

* O número é maior porque algumas proteínas são classificadas em duas classes de COG

** Valor correspondete a 13% do genoma

QUADRO 5 – Distribuição de CDS de *B. japonicum* CPAC 15 nas categorias funcionais de COG em comparação com alguns rizóbios já seqüenciados

Identificação da CDS	Organismo com maior similaridade	Nome do gene	Produto do BLAST	Query Coverage (%)	Subject Coverage (%)	Identidade (%)	e-value	Tamanho
BS11185	<i>Bradyrhizobium japonicum</i> USDA 110	<i>fixO</i>	protein FixO (partial)/cb-type cytochrome c oxidase subunit II (partial)	100	11	96	6e-07	28*
BS11156	<i>Bradyrhizobium japonicum</i> USDA 110	<i>fixP</i>	protein FixP/cb-type cytochrome c oxidase subunit III /cbb3 oxidase, subunit III	66	18	91	2e-40	134
BS11179	<i>Bradyrhizobium japonicum</i> USDA 110	<i>fixQ</i>	protein FixQ , cytochrome oxidase subunit, small membrane protein/cbb3 oxidase, subunit IV	100	100	94	2e-21	165
BS01532	<i>Bradyrhizobium japonicum</i> USDA 110	<i>fixR</i>	protein FixR	100	56	98	1e-82	159
BS03636	<i>Bradyrhizobium japonicum</i> USDA 110	<i>nifE</i>	nitrogenase iron-molybdenum cofactor biosynthesis protein NifE	100	61	99	0.0	338
BS08735	<i>Bradyrhizobium japonicum</i> USDA 110	<i>nodB</i>	NodB protein-related/chitooligosaccharide deacetylase	95	100	100	4e-38	82
BS11187	<i>Bradyrhizobium japonicum</i> USDA 110	<i>nodD2</i>	transcriptional regulator protein NodD2	93	76	31	7e-19	241
BS01652	<i>Bradyrhizobium japonicum</i> USDA 110	<i>nodW</i>	nodulation protein NodW, two-component regulatory system/NwsB	100	100	61	3e-65	224
BS00587	<i>Bradyrhizobium japonicum</i> USDA 110	<i>nopP</i>	Effector protein nopP (Nodulation outer protein P)	100	100	100	7e-71	128
BS00369	<i>Bradyrhizobium japonicum</i> USDA 110	<i>norB</i>	nitric oxide reductase subunit B	100	10	97	9e-20	50
BS01726	<i>Bradyrhizobium japonicum</i> USDA 110	<i>ntrX</i>	nitrogen assimilation regulatory protein NtrX	100	17	98	9e-38	81
BS05510	<i>Bradyrhizobium japonicum</i> USDA 110	<i>ntrY</i>	nitrogen regulation protein NtrY/two-component sensor histidine kinase	89	40	97	e-180	381

* Apesar do parâmetro estabelecido para análise do tamanho das CDSs ser no mínimo de 50 aminoácidos, o tamanho do gene *fixO* é de somente 54 aminoácidos, possibilitando que apenas 28 aminoácidos dêem uma excelente cobertura, possibilitando a caracterização desta CDS.

QUADRO 6 – Genes relacionados à fixação biológica do nitrogênio encontrados no panorama genômico de *B. japonicum* estirpe CPAC 15

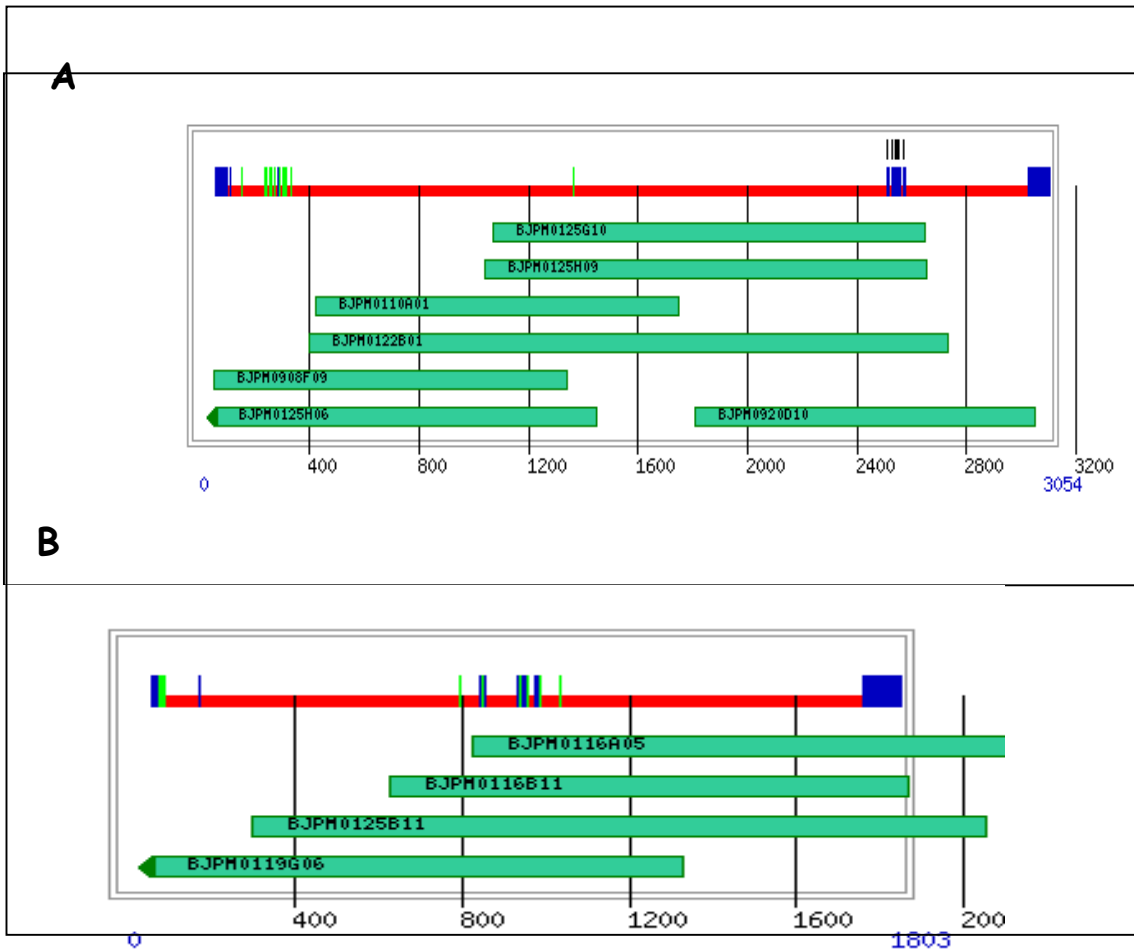


FIGURA 1 – Exemplos de montagem de “contigs”. A) “Contig” formado por 14 leituras (retângulos em verde). Cada retângulo está ancorado com duas leituras “b” e “g”. B) Exemplo de “contig” com oito leituras. A linha vermelha indica a região de consenso das leituras e as linhas azuis e verde claro, o valor dos alinhamentos.

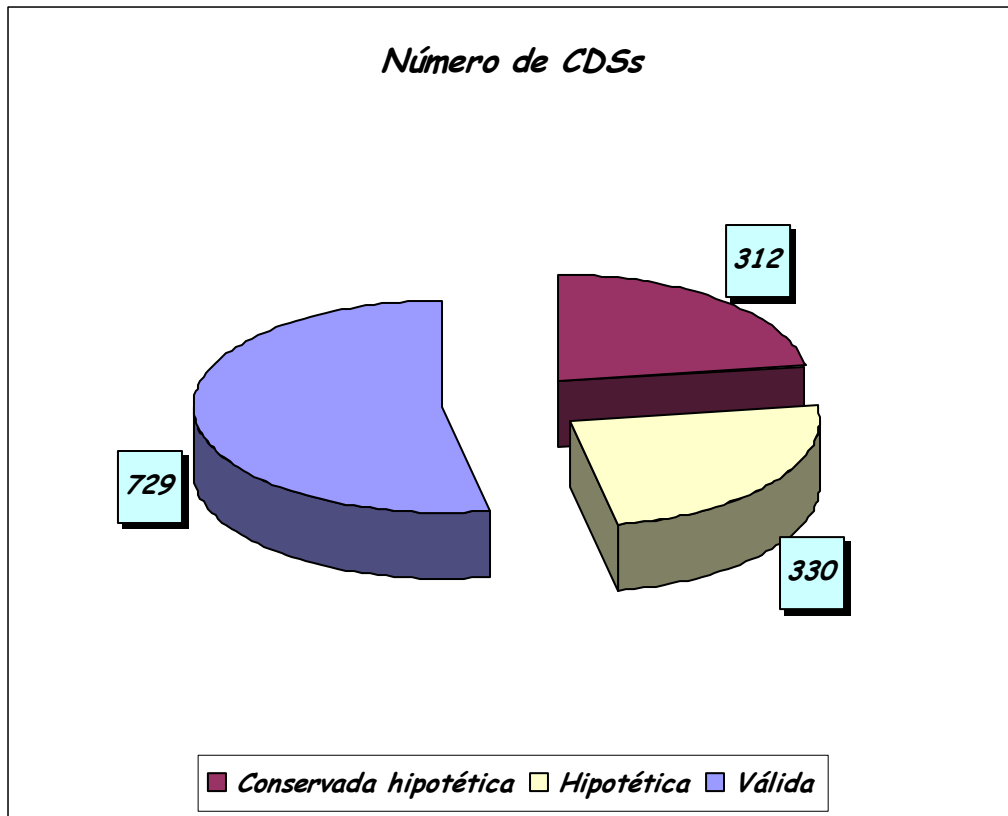


FIGURA 2 – Distribuição das "coding sequences" (CDS) no genoma parcial de *B. japonicum* estirpe CPAC 15.

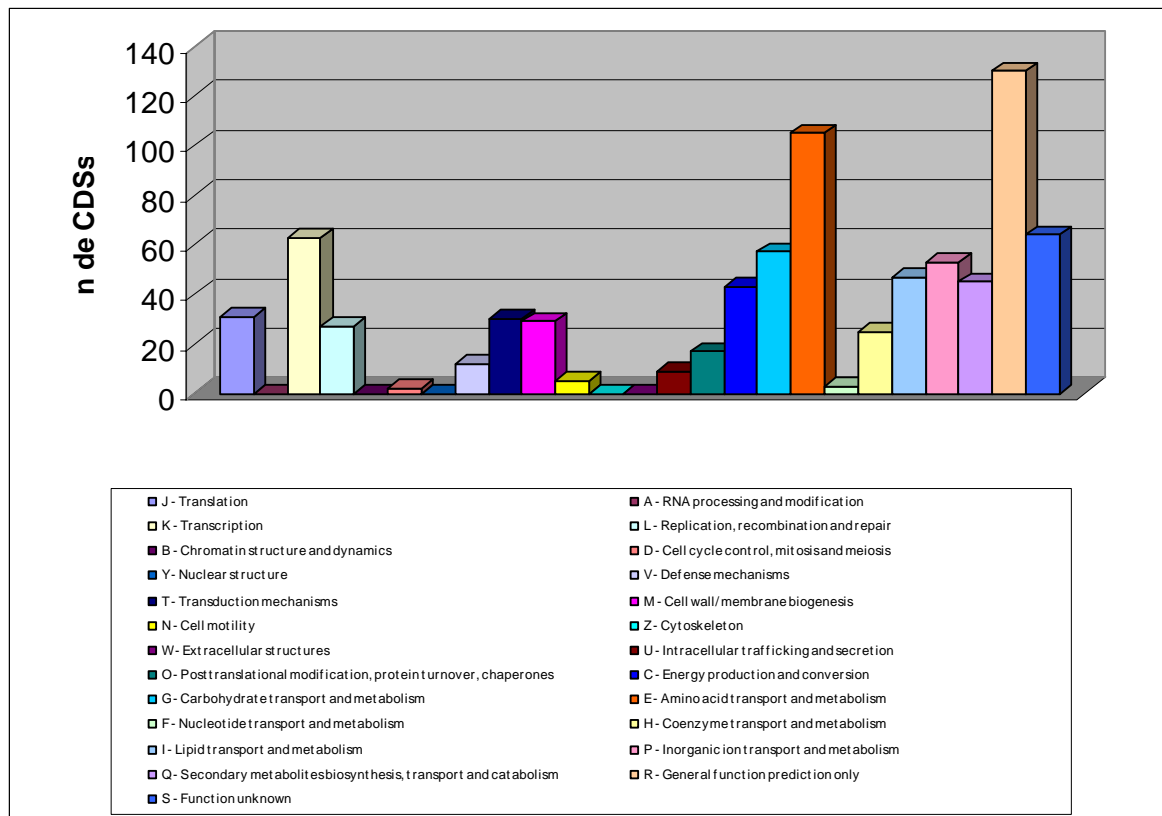


FIGURA 3 – Distribuição percentual das CDSs da estirpe CPAC 15 de *B. japonicum* segundo a base de dados COG.

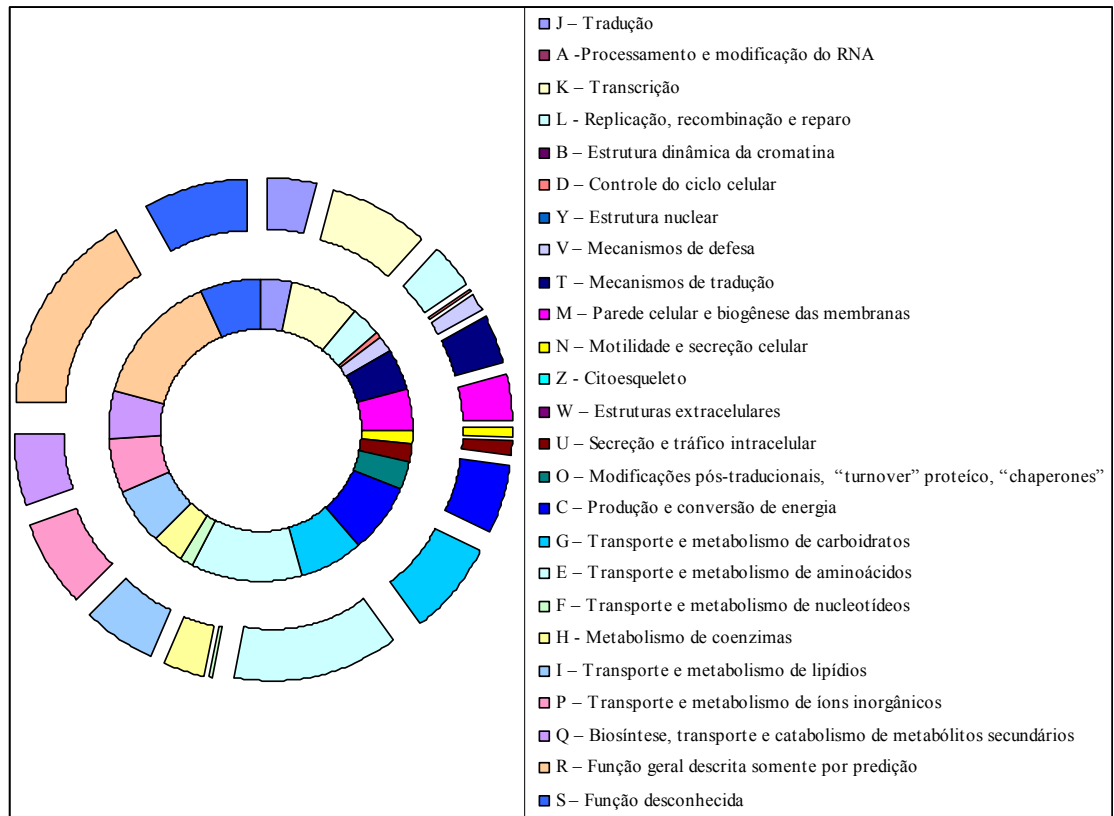


FIGURA 4 – Comparação entre as estirpes de *B. japonicum* quanto à proporção da porcentagem de CDSs encontradas nas diferentes categorias de COG. O círculo externo representa a estirpe CPAC 15 (13% do genoma), e o círculo interno a USDA 110 (100% do genoma).

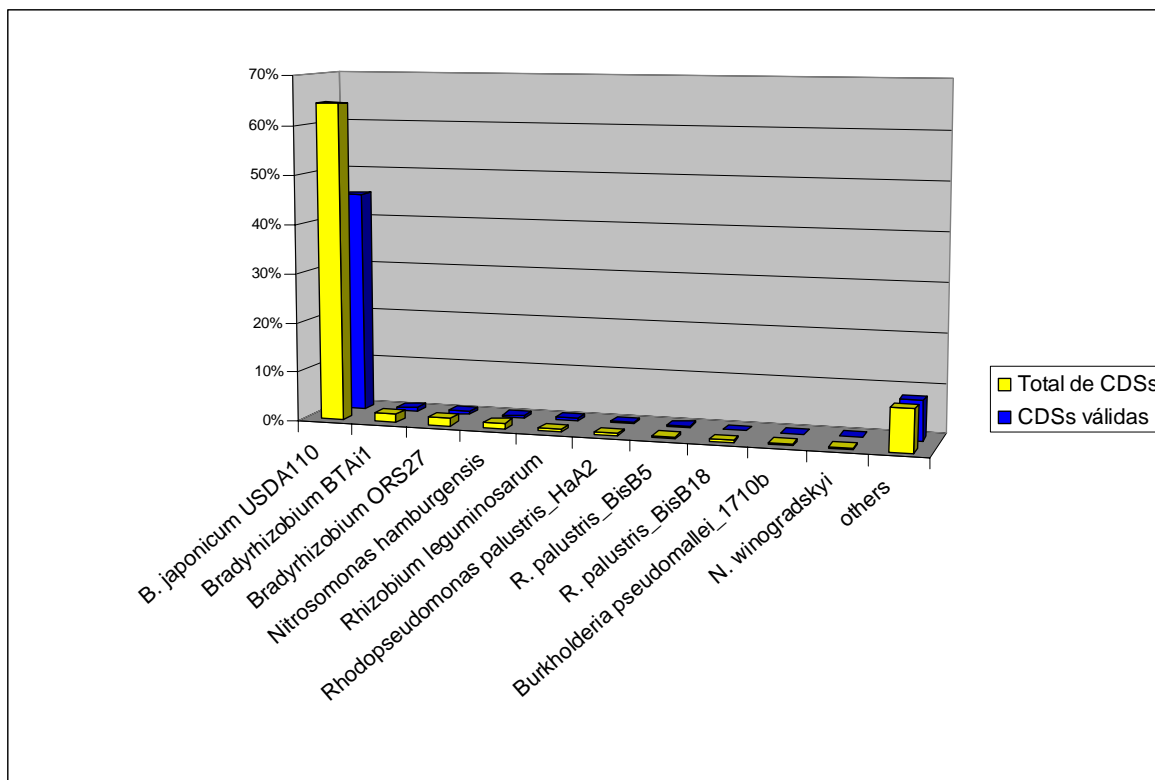


FIGURA 5 – Distribuição percentual de CDSs com função conhecida com similaridade a organismos encontrados no banco de dados KEGG.

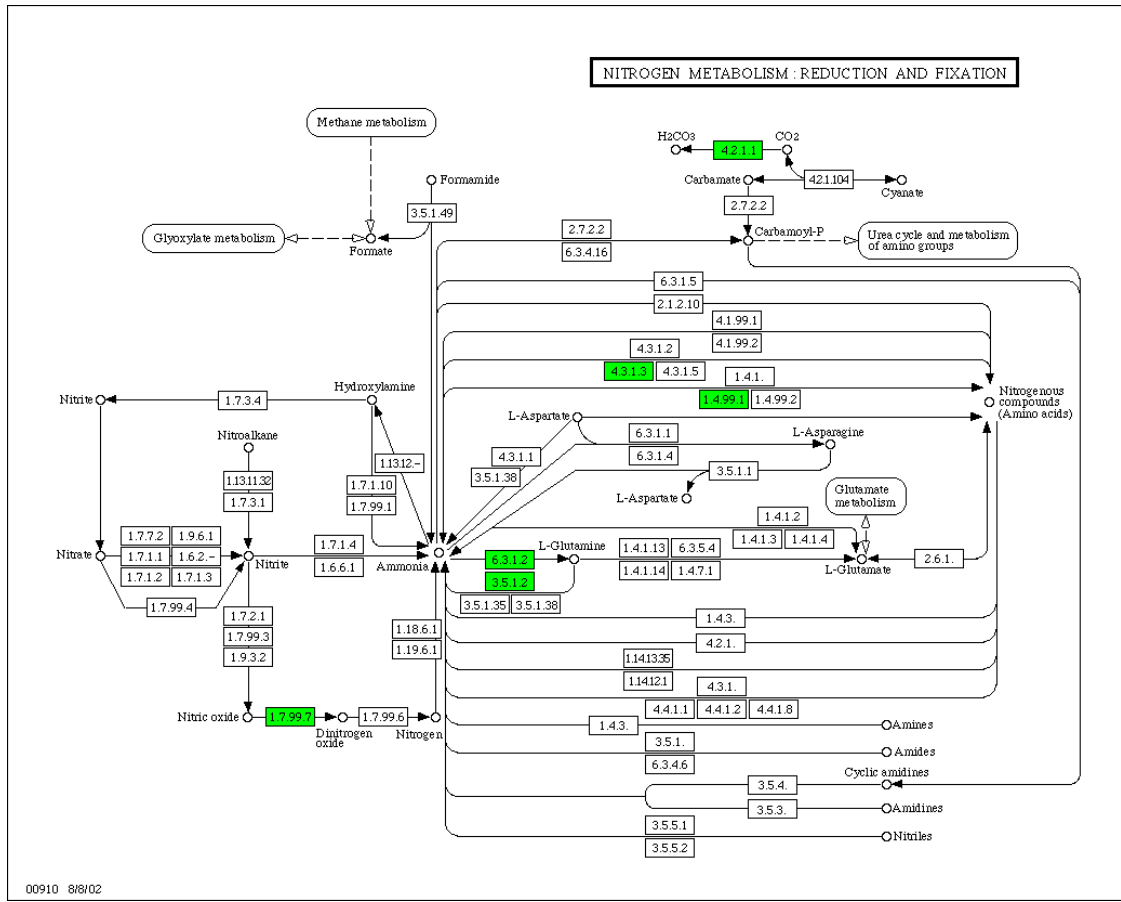


FIGURA 6 – Vias metabólicas de assimilação, redução e fixação do nitrogênio. Em verde as CDSs encontradas no genoma da estirpe CPAC 15, como *norB*, *glsA1* (glutaminase 1), *glnII* (glutamine synthetase II) e *norB* (nitric oxide reductase subunit B). CDSs representadas nos quadros verdes: 4.2.1.1 - carbonic anhydrase (*icfA*); 4.3.1.3 - histidine ammonia-lyase/histidase (*hutH*); 1.4.99.1 – dehydrogenase; 6.3.1.2 - glutamine synthetase II/glutamate-ammonia ligase II/GSII (*glnII*); 3.5.1.2 - glutaminase 1 (*glsA1*); 1.7.99.7 - nitric oxide reductase subunit B (*norB*).

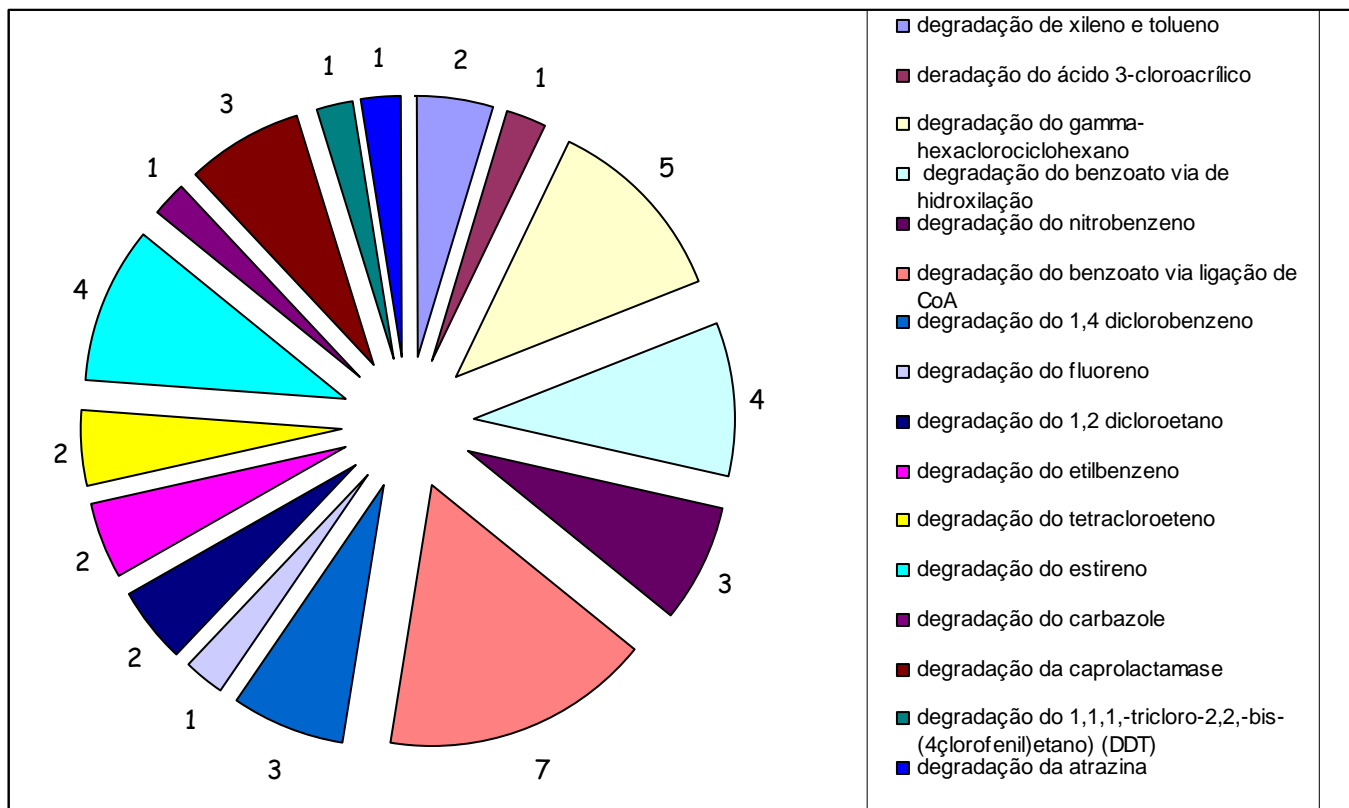



FIGURA 7 – Presença de CDSs na maioria das vias de degradação de xenobióticos.

System for Automated Bacterial Integrated Annotation - SABIA
Bradyrhizobium japonicum SEMIA 5079 - GENOME PROJECT

- * Annotation page
- * General info
- * ORF Table
- * New ORFs in current assembly
- * Changed ORFs in the last assembly
- * Requested ORFs (Automatic annotation)
- * Categories classification
- * Kegg functional classification
- * List of organisms by Kegg
- * Metabolic pathways
- * COG functional classification
- * List of tRNA
- * Contigs maps
- * Search
- * Pick a Sequence
- * ORF Finder
- * Find motifs
- * Blast
- * GC Skew
- * Overlap report
- * Show Operons
- * EC numbers present in organism
- * Paralogous classification
- * View a region
- * Repeated Gene Names
- * Insert mRNA, rRNAs and Frameshift
- * List mRNA, rRNAs and Frameshift
- * Valid ORFs with Frameshift
- * ORFs with assembly problems
- * TCDB reports
- Valid ORFs
- Conserved Hypothetical ORFs
- Hypothetical ORFs
- * Annotation Results
- User Reports
- Partial Results

Links e opções para análises e comparações

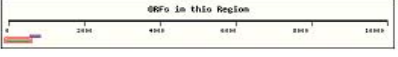


Bradyrhizobium japonicum SEMIA 5079 genome project


Annotation page

Functional annotation | InterPro | GO (Gene Ontology) | Blast | Psort | Annotation | Search

This ORF is valid !



ORF BS02509 alignments (identified by colors)



ORF information

ORF ID	BS02509 (frame -3)	Origin	Glimmer and GeneMark (Contig 285) (Old New)
Position and sequences	complement(1...636) (636 bp) (212 aa)	Upstream -	extragenic region
Molecular weight	23574.96	Theoretical pI	6.98
Optional start codon	6 found	Nucleotides	A (18.24%) C (33.95%) G (32.55%) T (15.25%)
Percent CG	66.51%	percentage	G (32.55%) T (15.25%)
Percent AT	33.49%	Percent AT	33.49%
Overlaps	left of BS02516 (4 bps)		
Transcriptional regulation			
RBS	New start position	Stop position	RBS pattern
	636	1	---
			RBS position
			0
			New start codon
			ATG
			Shift
			0
			Old start codon
			ATG
			Old start position
			636
Promoter	No results found !		

Comparações entre bancos de dados

FIGURA 8 – Análise de uma CDS na página de anotação do programa SÁBIA
<http://www.bnf.lncc.br/bj-bin/annotation/annotation.cgi?gene=BS02509>